JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 8, NUMBER 3 MARCH 1967

Singular Potentials and the $Z = 0$ Condition

H. H. ALY
Physics Department, American University of Beirut, Beirut, Lebanon

H. J. W. MÜLLER
Institut Für Theoretische Physik, der Universität München, Germany

(Received 7 July 1966)

We consider an interaction Lagrangian consisting of renormalizable and nonrenormalizable terms. After a brief discussion of the vertex function and the boundary conditions determining the renormalization constant corresponding to a pseudoscalar bound state, we derive an explicit expression for the nonrelativistic S-matrix for scattering by the potential $1/r^4$. These results are then used to evaluate Z. The explicit expressions obtained contain a strong cut-off dependence which cannot be factorized out. However, a specific value of the cut-off reduces the equation to a well-known eigenvalue problem, so that in this case a discrete spectrum of bound states may be obtained.

I. INTRODUCTION

CONSIDERABLE interest has recently been devoted to the study of singular potentials at small distances, since the infinities appearing in nonrenormalizable quantum field theory are believed to have some analogy with those arising in the perturbation expansion for nonrelativistic scattering by highly singular potentials. The Bethe–Salpeter equation, which gives a covariant description of the relativistic two-body problem, has also been widely discussed in this context; in fact, as Bastai *et al.*[1] have pointed out, there is a marked similarity between the behavior of the solutions of the BS equation near the origin and those of the radial Schrödinger equation. It is clear, therefore, that a more detailed knowledge of the characteristic features of nonrelativistic scattering by highly singular potentials could shed some light on the difficulties encountered in relativistic scattering theory.

Various authors[2] have pointed out that the Jost function in potential theory is the nonrelativistic analog of the constant Z which renormalizes the vertex between elementary and composite particles in a field-theoretical Lagrangian. The vanishing of this constant Z is equivalent to the condition of the bound state, thus yielding a relationship between the masses of the interacting particles and their coupling constants. Several classes of interactions have recently been considered by Furlan and Mahoux[3] in order to examine the characteristics of the resulting eigenvalue equations. In particular, they have shown that, for super-renormalizable and renormalizable interactions, the $Z_1 = 0$ condition is found to be either independent of a cut-off or else the cut-off can be factorized out. However, in the case of nonrenormalizable interactions, the situation is not as clear.

In the present paper, we consider an interaction

[1] A. Bastai, L. Bertocchi, S. Fubini, G. Furlan, and M. Tonin, Nuovo Cimento **30**, 1512, 1532 (1963).

[2] For a list of references, see W. Güttinger, Nuovo Cimento **36**, 968 (1965); L. Bertocchi, M. McMillan, E. Predazzi, and M. Tonin, *ibid.* **31**, 1352 (1964).

[3] G. Furlan and G. Mahoux, Nuovo Cimento **36**, 215 (1965).

Lagrangian consisting of renormalizable and nonrenormalizable terms. In Sec. II, we briefly discuss the vertex function and obtain the boundary conditions determining the renormalization constant corresponding to a pseudoscalar bound state. In Sec. III, we derive an explicit expression for the nonrelativistic S-matrix for scattering by the potential r^{-4}. In Sec. IV, we use these results to evaluate the renormalization constant Z. The explicit expressions obtained contain a strong cut-off dependence which cannot be factorized out. Finally, we discuss the physical significance of this cut-off.

II. RENORMALIZABLE AND NONRENORMALIZABLE INTERACTIONS

We consider the following mixture of a renormalizable and a nonrenormalizable interaction as an example illustrating interactions which, in nonrelativistic potential theory, correspond to highly singular potentials:

$$\mathfrak{L}_1 = g'\bar{\psi}(\gamma_5, 1)\psi\phi + g\bar{\psi}(\gamma_5, 1)\psi\phi^2. \tag{2.1}$$

Here, the ψ's represent spin-$\frac{1}{2}$ fields (mass m) and the ϕ's represent massless bosons. The two couplings $(\gamma_5, 1)$ represent Dirac matrices for pseudoscalar and scalar interactions, respectively. The first of the interactions (2.1) is renormalizable, whereas the second is not. We now want to discuss the case, where the spinors ψ form a pseudoscalar bound state of mass zero by the exchange of one or two of the bosons ϕ. Then in ladder approximation, the integral equation for the renormalized vertex function $\Gamma(p) = \gamma_5\Gamma(p^2)$ is represented by the diagram shown in Fig. 1, where $p = \frac{1}{2}(p_1 - p_2)$ is the relative 4-momentum of the two spin-$\frac{1}{2}$ particles, and where, for convenience, we have set $q = p_1 + p_2 = 0$. Thus, the mass of the composite particle (taken on the mass shell) is fixed, and quantization yields a relationship between g, g', and a quantum number n, say (and a cut-off, as we see later). The scalar and pseudoscalar interactions $(1, \gamma_5)$ in (2.1) may be termed as attractive and repulsive, respectively, since they correspond to attractive and repulsive potentials. $\Gamma(p^2)$, the scalar part of the vertex function, is easily seen to satisfy the integral equation

$$\Gamma(p^2) = Z - \frac{\lambda'\epsilon}{(2\pi)^4}\int \frac{d^4p'\Gamma(p'^2)}{(p'^2 + m^2)(p - p')^2} - \frac{\lambda\epsilon}{(2\pi)^4}\int \frac{d^4p'\Gamma(p'^2)}{(p'^2 + m^2)}\int \frac{d^4k}{k^2(k + p' - p)^2}, \tag{2.2}$$

where Z is the vertex renormalization constant, $\lambda = g^2/(2\pi)^4m^2$, $\lambda' = g'^2/m^2$, and $\epsilon = \pm 1$ corresponding to the $(\gamma_5, 1)$ coupling.

Now, except in the case of scalar–particle interactions, the solution of (2.2) does not decrease sufficiently rapidly at high energies due to an essential singularity at infinity, so that the introduction of a cut-off in the integration becomes inevitable. These difficulties have been discussed by various authors,[2–4] and arguments have been suggested to justify the following procedure.

We perform a Fourier transformation of Eq. (2.2) [divided by $(p^2 + m^2)$] and use the Wick rotation[5] of the time component of x so that we finally work in terms of a pure Euclidean metric. Furthermore, we use the integrals

$$\left.\begin{aligned} &\int (e^{iqx}/q^2)\,dq = (2\pi/x)^2, \\ &\int_0^\pi e^{ipy\cos\alpha}\sin^2\alpha\,d\alpha = (\pi/ipy)I_1(ipy), \\ &\int \frac{e^{ipy}\,d^4p}{(p^2 + m^2)} = (4\pi^2/y)mK_1(my), \end{aligned}\right\} \tag{2.3}$$

where I_1 and K_1 are modified Bessel functions (the volume element d^4p being given by

$$p^3\sin^2\alpha\sin\varphi\,dp\,d\alpha\,d\varphi\,d\theta$$

in four-dimensional spherical co-ordinates) and expand $K_1(m|y - x|)/m|y - x|$ in terms of Gegenbauer polynomials.[6] The orthogonality properties of these polynomials then yield the following integral equation:

$$\psi(y) = ZmK_1(my) - \int_0^\infty dx\left(\frac{\lambda\epsilon}{x^3} + \frac{\lambda'\epsilon}{x}\right)\psi(x) \times [\theta(y - x)K_1(my)I_1(mx) + \theta(x - y)K_1(mx)I_1(my)], \tag{2.4}$$

where

$$\psi(y) = yF(y^2),$$
$$\frac{\Gamma(p^2)}{p^2 + m^2} = \frac{1}{(2\pi)^2}\int e^{-ipx}F(x^2)\,dx. \tag{2.5}$$

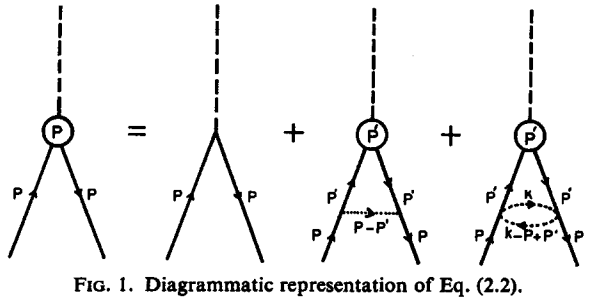


FIG. 1. Diagrammatic representation of Eq. (2.2).

[4] F. T. Hadjioannou, Nuovo Cimento **35**, 570 (1965).
[5] G. C. Wick, Phys. Rev. **96**, 1124 (1954).
[6] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill Book Co., Inc., New York, 1955), Vols. I, II.

Differentiating (2.4) (and using well-known properties of the modified Bessel functions, such as their recurrence relations and asymptotic behavior), we obtain a differential equation and two boundary conditions:

$$\left(\frac{d^2}{dy^2} + \frac{1}{y}\frac{d}{dy} - m^2 - \frac{\lambda'\epsilon + 1}{y^2} - \frac{\lambda\epsilon}{y^4}\right)\psi(y) = 0, \quad (2.6)$$

$$\lim_{y\to\infty}[m^{-1}(d/dy)\psi(y) + \psi(y)] = 0,$$
$$\lim_{y\to 0} y[y(d/dy)\psi(y) - \psi(y)] = -2Z. \quad (2.7)$$

This Z is not to be confused with the argument z of Sec. III below.

Since, in effect, the vertex function represents the bound-state wavefunction of the composite particle,[2] the interactions (2.1) may be regarded as being equivalent to potentials of the form y^{-2}, y^{-4} in the nonrelativistic Schrödinger equation. The first of these leads to a renormalizable interaction as discussed by Furlan and Mahoux.[3] Due to the mathematical complexity of the solutions, however, the potential y^{-4} has only recently been studied in some detail.[7]

In the following section, we show that, at least in the interesting case of reasonably high energies, these difficulties are only apparent, and a simple expression for the S-matrix may be derived.

III. SCATTERING BY THE POTENTIAL r^{-4}

For convenience, we consider the radial Schrödinger equation for all partial waves l and a repulsive potential, i.e.,

$$\varphi''(r) + [k^2 - l(l+1)/r^2 - \gamma^2/r^4]\varphi(r) = 0. \quad (3.1)$$

We next introduce the substitutions

$$\varphi(r) = r^{\frac{1}{2}}\psi(r), \quad \Lambda = (l + \tfrac{1}{2})^2,$$
$$\chi = r/r_a = e^{z'}, \quad z' = z + \tfrac{1}{4}i\pi. \quad (3.2)$$

Then, $\psi(r)$ satisfies a differential equation identical with (3.1), except for an additional term $r^{-1}(d\psi/dr)$. We also choose the parameter r_a such that

$$\left.\begin{aligned} k^2 r_a^2 = \gamma^2/r_a^2 = g^2, \\ h^2 \equiv ig^2. \end{aligned}\right\} \quad (3.3)$$

and set

Then, $\psi(z)$ satisfies the modified Mathieu equation

$$d^2\psi/dz^2 - [\Lambda - 2h^2\cosh 2z]\psi(z) = 0. \quad (3.4)$$

We observe that this equation is invariant under the interchanges

$$h \to \pm h, \quad z \to -z, \quad (3.5a)$$

$$h \to \pm ih, \quad z \to -z \mp \tfrac{1}{2}i\pi, \quad \text{i.e.,} \quad z' \to -z'. \quad (3.5b)$$

The corresponding invariant points are (a) $z = 0$, (b) $z = \mp\frac{1}{4}i\pi$, $z' = 0$.

Now, in an earlier investigation[8] it was shown that, for large values of h^2, two linearly independent solutions of Eq. (3.4) may be written as

$$\left.\begin{aligned} \psi_0(z, h) &= A(z)\exp(-2ih\sinh z), \\ \bar{\psi}_0(z, h) &= \bar{A}(z)\exp(+2ih\sinh z), \end{aligned}\right\} \quad (3.6)$$

where $A(z)$, $\bar{A}(z)$ are known functions of z. The following asymptotic behavior of these solutions is easily verified [cf. (3.2)]

$$\exp[\mp 2ih\sinh z] \underset{x\to 0}{\sim} \exp[\mp g/x] = \exp[\mp\gamma/r],$$
$$\exp[\mp 2ih\sinh z] \underset{x\to+\infty}{\sim} \exp[\mp igx] = \exp[\mp ikr]. \quad (3.7)$$

Clearly by (3.5) and (3.6), we also have the following pair of solutions

$$\psi_\mp(z; h) = \psi_0(-z - \tfrac{1}{2}i\pi; \pm ih) \quad (3.8)$$

with the property

$$\psi_\mp(z; h) \underset{r\to\infty}{\sim} e^{\mp ikr}. \quad (3.9)$$

The S-matrix may then be derived as follows. Writing

$$(r^{\frac{1}{2}})\psi(z; h) = A\{\psi_-(z; h) - S(h)\psi_+(z; h)\}(r^{\frac{1}{2}}) \quad (3.10)$$

and taking the limit $r \to \infty$, we see that (3.10) is the standard definition of the S-matrix. The solution regular at $r = 0$ (i.e., $x = 0$) is

$$(r^{\frac{1}{2}})\psi(z; h) = B\psi_0(z; h)(r^{\frac{1}{2}}). \quad (3.11)$$

Thus, to obtain the S-matrix, all we need is the analytic continuation of (3.11) to (3.10), i.e., [cf. (3.5)] at $z_0 = -\frac{1}{4}i\pi$. Hence, equating (3.10), (3.11), and their derivatives at this point, we have

$$S(h) = \left.\frac{\begin{array}{l}\psi_0(z; h)(d/dz)\psi_-(z; h) \\ \quad - \psi_-(z; h)(d/dz)\psi_0(z; h)\end{array}}{\begin{array}{l}\psi_0(z; h)(d/dz)\psi_+(z; h) \\ \quad - \psi_+(z; h)(d/dz)\psi_0(z; h)\end{array}}\right|_{z_0}$$

$$= \left.\frac{\begin{array}{l}\psi_0(y; h)(d/dy)\psi_0(y; +ih) \\ \quad + \psi_0(y; +ih)(d/dy)\psi_0(y; h)\end{array}}{\begin{array}{l}\psi_0(y; h)(d/dy)\psi_0(y; -ih) \\ \quad + \psi_0(y; -ih)(d/dy)\psi_0(y; h)\end{array}}\right|_{y=-\frac{1}{4}i\pi}$$

$$= \left.\frac{(d/dz)[\psi_0(z; h)\psi_0(z; +ih)]}{(d/dz)[\psi_0(z; h)\psi_0(z; -ih)]}\right|_{z=-\frac{1}{4}i\pi}. \quad (3.12)$$

[7] R. M. Spector, J. Math. Phys. **5**, 1185 (1964); L. Bertocchi, S. Fubini, and G. Furlan, Nuovo Cimento **35**, 633 (1965); H. H. Aly and H. J. W. Müller, J. Math. Phys. **7**, 1 (1966).

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

It is now necessary to specify the function $A(z)$ in solution (3.6); this function is given by[8]

$$\begin{aligned}A(z;h) &= A_q(z) + (2^7h)^{-1}\{P_1(1)A_{q+4} + P_1(-1)A_{q-4}\} \\ &\quad + (2^7h)^{-2}\{P_2(2)A_{q+8} + P_2(1)A_{q+4} \\ &\quad + P_2(-1)A_{q-4} + P_2(-2)A_{q-8}\} + \cdots \\ &= \sum_{p=0}^{\infty}\sum_{j=-p}^{p}(2^7h)^{-p}P_p(q;j)A_{q+4j}(z), \\ &\qquad (j \neq 0 \text{ for } p > 0), \quad (3.13)\end{aligned}$$

where q is related to the Floquet parameter of the Mathieu equation.[8] The coefficients P are known[9] or can easily be calculated. They depend on q. The functions A_q are given by

$$A_{q+4j}(z) = \frac{\cos^{\frac{1}{2}(q+4j-1)}(\frac{1}{4}\pi - \frac{1}{2}iz)}{\cos^{\frac{1}{2}(q+4j+1)}(\frac{1}{4}\pi + \frac{1}{2}iz)}. \quad (3.14)$$

It is readily seen that $A_q(z)$ has the following limits:

$$\begin{aligned}&\text{for } r \to 0: \quad A_q(z) \sim 2(r/r_a)^{\frac{1}{2}} \exp[-\tfrac{1}{4}i\pi(q + \tfrac{1}{2})], \\ &\text{for } r \to \infty: \quad A_q(z) \sim 2(r_a/r)^{\frac{1}{2}} \exp[\tfrac{1}{4}i\pi(q + \tfrac{1}{2})].\end{aligned} \quad (3.15)$$

This yields precisely the behavior to be expected of the solution φ of (3.1). Now, in calculating S, it is incorrect to substitute the full solution (3.13) into (3.6) and thence into (3.12), since the phase shift is determined by the *asymptotic* behavior of the solutions, i.e., by their behavior for $r \to \infty$, or in the present case [due to the symmetry properties (3.8)] by their behavior for $r \to 0$. Thus, S is obtained by substituting into (3.12) the dominant terms of ψ_0 for $r \to 0$ and then setting $z = -\frac{1}{2}i\pi$, i.e., $z' = 0$ or $r = r_a$. So we set

$$\begin{aligned}\psi_0(z, h) &\simeq e^{-\gamma/r}\sum_{p=0}^{\infty}{\sum_{j=-p}^{p}}' P_p(q(h); j)/(2^7e^{\frac{1}{4}i\pi}g)^p 2(r/r_a)^{\frac{1}{2}} \\ &\quad \times e^{-\frac{1}{4}i\pi}[q(h) + 4j + \tfrac{1}{2}] \quad (r \to 0), \quad (3.16)\end{aligned}$$

where the dash on the sum implies $j \neq 0$ for $p > 0$. We have also included an h-dependence in q, since q has to be determined from the secular equation of the Mathieu equation, i.e., as the solution of a function $F(\Lambda, q, h) = 0$. We now have

$$\begin{aligned}&\psi_0(z; h)\psi_0(z; \pm ih)/4e^{-\frac{1}{4}i\pi}[q(h) + q(\pm ih) + 1] \\ &\quad \simeq (r/r_a)e^{-(\gamma/r)(1\pm i)}\sum_{p_1=0}^{\infty}{\sum_{j_1=-p_1}^{p_1}}'\sum_{p_2=0}^{\infty}{\sum_{j_2=-p_2}^{p_2}}'(-1)^{j_1+j_2} \\ &\quad \times \frac{P_{p_1}[q(h); j_1]P_{p_2}[q(\pm ih); j_2]}{(2^7e^{\frac{1}{4}i\pi}g)^{p_1}(\mp 2^7e^{-\frac{1}{4}i\pi}g)^{p_2}} \\ &\quad = (r/r_a)e^{-(\gamma/r)(1\pm i)}\sum_{j=0}^{\infty}(2^7g)^{-j}Q_j^{(\pm)}[q(h)], \quad (3.17)\end{aligned}$$

where

$$\begin{aligned}Q_j^{(\pm)}[q(h)] &= \sum_{m=0}^{j}\sum_{j_1=-m}^{m}{\sum_{j_2=m-j}^{j-m}}'(-1)^{j_1+j_2}\frac{P_m[q(h); j_1]}{(e^{\frac{1}{4}i\pi})^m} \\ &\quad \times \frac{P_{j-m}[q(\pm ih); j_2]}{(\mp e^{-\frac{1}{4}i\pi})^{j-m}} \quad (3.18)\end{aligned}$$

and

$$Q_0^{(\pm)}[q(h)] = 1.$$

Clearly,

$$\begin{aligned}&\frac{(d/dz)[\psi_0(z; h)\psi_0(z; \pm ih)]_{z=-\frac{1}{4}i\pi}}{4e^{-\frac{1}{4}i\pi}[q(h) + q(\pm ih) + 1]} \\ &\quad = r_a\frac{d}{dr}\left[\frac{r}{r_a}e^{-(\gamma/r)(1\pm i)}\sum_{j=0}^{\infty}\frac{Q_j^{(\pm)}[q(h)]}{(2^7g)^j}\right]_{r=r_a} \\ &\quad = e^{-g(1\pm i)}[1 + g(1 \pm i)]\sum_{j=0}^{\infty}\frac{Q_j^{(\pm)}[q(h)]}{(2^7g)^j}. \quad (3.19)\end{aligned}$$

Now, the S-matrix assumes the simple form

$$\begin{aligned}S = e^{-2ig}\left[\frac{1 + i + 1/g}{1 - i + 1/g}\right] &\cdot \frac{\sum_{j=0}^{\infty}Q_j^{(+)}[q(h)]/(2^7g)^j}{\sum_{j=0}^{\infty}Q_j^{(-)}[q(h)]/(2^7g)^j} \\ &\times \exp(-\tfrac{1}{4}i\pi)[q(+ih) - q(-ih)]. \quad (3.20)\end{aligned}$$

We have already remarked that q is a function of h; in fact, our solutions ψ are solutions of the modified Mathieu equation only if also[8,10]

$$\begin{aligned}\Lambda(h^2) &= -2h^2 + 2hq - \tfrac{1}{8}(q^2 + 1) + O(1/h), \\ &\qquad [h = +(h^2)^{\frac{1}{2}}]. \quad (3.21)\end{aligned}$$

This equation determines the (as yet) unknown parameter q. For nonintegral values of q, the parameter h in (3.21) has to be taken as the positive square root of h^2.[11] It then follows that, if we solve (3.21) for q, q is also a function of h^2; in fact,

$$\begin{aligned}&q(h^2) = 4h[2 \pm \sqrt{3}] + O(1/h), \\ &q(+h) = q(-h), \quad h = +(h^2)^{\frac{1}{2}}.\end{aligned} \quad (3.22)$$

[Note: Λ of Eq. (3.21) is the Λ of both solutions proportional to $\exp(\pm 2hi \sinh z)$]. Since $q(\pm ih)$ is to be understood as $q(h)$ with h replaced by $\pm ih$, it follows that these two functions must also be identical. So their exponentials in (3.20) give no contribution.

Unfortunately, an expansion of q [cf. (3.22)] in falling powers of h does not appear to be given in the literature, whereas expansions in rising powers of h^2 are well known.[10] Also, the unitarity of the quotient

[9] K. M. Case, Phys. Rev. **80**, 797 (1950).

[10] J. Meixner and F. W. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen* (Springer-Verlag, Berlin, 1954).

[11] See Ref. 10, pp. 133, 139.

of sums in (3.20) cannot be easily seen. However, substitution of (3.22) into the coefficients[8]

$$P_1(q, 1) = (q + 1)(q + 3),$$
$$P_1(q, -1) = -(q - 1)(q - 3)$$

shows that to 0(1) the terms in the numerator are the same as those in the denominator and are real (so these terms "preserve" unitarity and do not contribute to the phase shift).

Equation (3.21) exhibits the characteristic difference between the behavior of the phase shift δ for potentials of centrifugal or Coulomb type and that for the singular potentials. We observe that, in the limit $k \to \infty$ (i.e., $g \to \infty$)

$$S \sim e^{-2ig}(1 + i)/(1 - i) = e^{\frac{1}{2}i\pi - 2ig}, \tag{3.23}$$

so that the phase shift δ has the behavior

$$\delta = \tfrac{1}{4}\pi - g + O(1/g) \sim \tfrac{1}{4}\pi - (\gamma k)^{\frac{1}{2}} \quad (k^2 > 0). \tag{3.24}$$

This behavior is completely different from that of the regular potentials, where the phase shift decreases to zero as $k \to \infty$. Furthermore, in the case of the regular potentials, high energies imply small coupling parameters, so that the Born expansion leads to the correct high-energy behavior. However, in the case of our singular potential, we see from Eq. (3.3) that high energies correspond to large coupling constants, so even from this point of view one would expect the Born expansion to diverge; this is in fact the case, as is well known. The result (3.24) agrees with that of Bertocchi *et al.*[7]

IV. THE $Z = 0$ CONDITION FOR COMPOSITE SYSTEMS

We first calculate Z by the conditions (2.7). Which is the solution $\psi(y)$ we have to use? Since $m^2 = -k^2$, we have $m = ik$, and (3.9) shows that

$$\psi_{\mp}(r; h) \underset{r\to\infty}{\sim} e^{\mp mr}. \tag{4.1}$$

Clearly, ψ_+ is ruled out by the first of the conditions (2.7). We now want to know the behavior of $r[r\psi'(r) - \psi(r)]$ for $r \to 0$, so we have to re-express $\psi_-(y; h)$ as a linear combination of regular and irregular solutions at $y = 0$. The regular solution is again $\psi_0(z; h)$; the irregular solution may be written $\psi_0(z; -h)$. Thus, the two solutions to be matched at the point $z = -\frac{1}{2}i\pi$ are (choosing an appropriate normalization)

$$\left.\begin{aligned} \psi(z; h) &= \psi_-(z; h) \\ \psi(z; h) &= \psi_0(z; h) + c\psi_0(z; -h) \end{aligned}\right\}. \tag{4.2}$$

Then, proceeding as in Sec. III, we obtain

$$c = -\left.\frac{\frac{d}{dz}[\psi_0(z; +ih)\psi_0(z; h)]}{\frac{d}{dz}[\psi_0(z; +ih)\psi_0(z; -h)]}\right|_{z=-\frac{1}{4}i\pi},$$

$$c = +e^{-2g}\left[\frac{(1 + i) + 1/g}{(1 - i) - 1/g}\right]\frac{\sum_{j=0}^{\infty}\bar{Q}^{(+)}[q(h)]}{\sum_{j=0}^{\infty}\bar{Q}^{(-)}[q(h)]}, \tag{4.3}$$

where

$$\bar{Q}_j^{(\pm)}[q(h)] = \sum_{m=0}^{j}\sum_{j_1=-m}^{m}{}' \sum_{j_2=m-j}^{j-m}{}' \times \frac{(-)^{j_1+j_2}P_m[q(\pm h); j_1]P_{j-m}[q(+ih); j_2]}{(\pm e^{\frac{1}{4}i\pi})^m(-e^{-\frac{1}{4}i\pi})^{j-m}} \times (2^7 g)^{-j}.$$

We also have [cf. (3.6), (3.7), (3.13), and (3.14)]

$$\psi_0(Z; \pm h) = e^{\mp 2ih\sinh z}\sum_{p=0}^{\infty}\sum_{j=-p}^{p}(\pm 2^7 h)^{-p} \times P_p[q(\pm h); j]A_{q+4j}^{(z)} \underset{r\to 0}{\sim} (r)^{\frac{1}{2}}e^{\mp\gamma/r}A_{\pm}, \tag{4.4}$$

where

$$A_{\pm} = 2[m/\gamma]^{\frac{1}{4}}e^{-\frac{1}{4}i\pi q(\pm h)}\sum_{p=0}^{\infty}\sum_{j=-p}^{p}{}'\frac{(-)^j P_p[q(\pm h); j]}{[\pm 2^7 i(m\gamma)^{\frac{1}{2}}]^p}. \tag{4.5}$$

Then

$$\psi(z; h) \underset{r\to 0}{\sim} (r)^{\frac{1}{2}}[A_+e^{-\gamma/r} + CA_-e^{+\gamma/r}] \sim CA_-(r)^{\frac{1}{2}}e^{+\gamma/r}, \tag{4.6}$$

and Z becomes (for the repulsive case)

$$Z = \lim_{r\to 0}\tfrac{1}{2}CA_-[\gamma + \tfrac{1}{2}r](r)^{\frac{1}{2}}e^{+\gamma/r}. \tag{4.7}$$

In order to obtain the corresponding results for the attractive case, i.e., an attractive potential $1/r^4$, we replace γ in (3.1) by $i\gamma$ and g^2 by $i\gamma k$. Then

$$\psi_0(z; \pm h) \underset{r\to 0}{\sim} (r)^{\frac{1}{2}}e^{\mp i\gamma/r}B_{\pm}, \tag{4.8}$$

where

$$B_{\pm} = 2[m/\gamma]^{\frac{1}{4}}e^{-\frac{1}{4}i\pi[q(\pm ih)+\frac{1}{2}]} \times \sum_{p=0}^{\infty}\sum_{j=-p}^{p}\frac{(-1)^j P_p[q(\pm ih); j]}{[\pm 2^7 e^{\frac{3}{4}i\pi}(\gamma m)^{\frac{1}{2}}]^p}. \tag{4.9}$$

Then

$$\psi(z; h) \underset{r\to 0}{\sim} (r)^{\frac{1}{2}}[B_+e^{-i\gamma/r} + CB_-e^{+i\gamma/r}] \tag{4.10}$$

and Z becomes

$$Z = \lim_{r\to 0} -\tfrac{1}{2}(r)^{\frac{1}{2}}[(\tfrac{1}{2}r + i\gamma)B_+e^{-i\gamma/r} + (\tfrac{1}{2}r - i\gamma)CB_-e^{i\gamma/r}]. \tag{4.11}$$

These results reaffirm the conclusions of Furlan and Mahoux.[3] These conclusions state that, in the case of nonrenormalizable interactions, the cut-off is a built-in parameter of the $Z = 0$ condition, and, in general, there is no way to factorize it out. This implies, of course, that the compositeness condition gives a continuous bound-state spectrum, which is unphysical. The same difficulty would arise if we tried to calculate the S-matrix—which we cannot do with the above results, since the energy was taken to be zero (cf. Sec. II, where q was taken equal to zero); it would also be strongly cut-off dependent.

However, Case[9] pointed out that the main difference between regular and singular potentials (i.e., potentials of the form r^{-n} for $n < 2$ or $n > 2$, respectively) is that the latter require the introduction of some other parameter, in addition to the potential parameters, in order to ensure orthogonality of the eigenfunctions, which would otherwise be too numerous and hence overcomplete.

But, representing an actual scattering process by a two-point system with an interaction which becomes infinite as the separation approaches zero, means necessarily an oversimplification. Thus, even for the proton which is expected to have a finite radius, the interaction with an electron is not strictly of Coulomb-type at $r = 0$. However, the eigenvalues are still correctly given by those obtained from the vanishing of the wavefunction with $r = 0$ as a boundary condition. In other words, the eigenvalues are essentially independent of exactly when or how the Coulomb law breaks down near $r = 0$.

Now, since the above results are unphysical for arbitrary values or r, one could ask which specific value of the cut-off r would yield a normal eigenvalue equation involving a quantum number describing the various states of the system. The potential discussed in Sec. II does indicate a possible answer. It is easily seen that several transformations reduce the corresponding radial Schrödinger equation to the proper periodic Mathieu equation, for which the eigenvalues are well known. These eigenvalues, however, are determined by the condition that the eigensolutions be periodic functions of the independent variable. So the question is under what conditions would these eigenvalues also be eigenvalues of the original Schrödinger equation. Clearly, the condition is that $\psi(iZ)$ must satisfy the same periodicity condition. This, in particular, implies that ψ vanish at a point, where $iZ = p(h, q)$, say or

$$r = r_a\, e^{-ip(h,q)}, \quad r_a = (\gamma/k)^{\frac{1}{2}}, \tag{4.12}$$

p being a complex function of h. A value of r such as this is related to the extra parameter introduced by Case.[9] It implies essentially a hard-core boundary condition. We see, therefore, that if we choose the cut-off as a particular function of k and γ, then the equation yields a discrete eigenvalue spectrum, and the difficulties of the continuum are removed.

Of course, the singular nature of the potential would have been avoided. Still we see that r satisfies very plausible limiting conditions, i.e.,

$$\lim_{|k|\to\infty} r = 0; \quad \lim_{|\gamma|\to 0} r = 0.$$

In other words, at extremely high energies or for very weak coupling, the cut off vanishes and the scattering particles approach the actual scattering center.

Non-Usual Topologies on Space-Time and High-Energy Scattering

D. ATKINSON AND M. B. HALPERN
Department of Physics, University of California, Berkeley, California

(Received 21 June 1966)

Motivated by recently observed deviations from quantum electrodynamical theory, we study the possibility that our notions of space-time may need revision at small distances. In this work, we wish to call attention to certain techniques which are available for studying different space-time structures within the framework of topology. Our main effort is in the consideration of a non-usual topology on space-time in which is embedded an elementary length. By working separately in each n-particle subspace, the embedding is done in an inhomogeneous Lorentz invariant way, and we avoid any lattice structure in space-time. Particles in this topology are in general extended structures, and we find the surprising feature that, at high energies, the topology enhances backward and large-angle scattering. From these preliminary investigations, we are not as yet able to make more than qualitative comparison with experiment. Along the way, we have the opportunity to remark on ways of embedding an intrinsic breakdown of certain invariances (e.g., parity) in the topology of space-time.

INTRODUCTION

IT has often been suspected that our common notions concerning the structure of space-time may break down for extremely small intervals. Indeed, it would be *a priori* surprising if these notions were correct at all distances. A recent large-angle pair-production experiment[1] in electrodynamics has raised the possibility that, even at presently available energies, we may in fact be entering into such a regime of breakdown. This disturbing notion finds indirect support, perhaps, in the various unsuccessful *ad hoc* modifications of quantum electrodynamics[2] (stimulated by the pair-production experiment), within the framework of a usual space-time structure. Although one may still hope to find an explanation of the data in essentially conventional terms, we have been motivated by the present situation to look into possible ways of changing the structure of space-time.

The structure of a space is most naturally studied through its topology.[3] We have in mind here the use of topology to describe the local or microscopic structure of a space,[4] rather than its global properties[5] (such as torsion, macroscopic connectedness, etc.).

A serious difficulty, encountered when a non-usual topology is taken for a space, lies in setting up a dynamics. In particular, analytic procedures with which the physicist is familiar have been defined on the usual topology, and any change of topology necessitates a new set of operators, function theory, etc. As physicists, this seriously constrains our ability to construct dynamics on any but the simplest non-usual topologies. Our theories are by no means as complete as the usual theory: we content ourselves with simple dynamical statements on some relatively straightforward non-usual topologies. Even then, we probably raise more questions than we can answer. Nonetheless, we hope that our considerations at least call attention to the fact that there is a well-defined framework (topology) within which to consider changing the presumed structure of space-time; and we hope that our ideas, even if (a) incorrect or (b) premature (or both) may lead to more satisfactory ones in the future.

Among the many simple topologies one might study, we, for definiteness, set ourselves the task of building one in which there is embedded an elementary length, λ, but whose open-set structure is invariant under the full inhomogeneous Lorentz group.[6] Previous attempts[7] at the incorporation of an elementary length in space-time have involved a lattice structure for the manifold, and a consequent loss of translation, rotation, and Lorentz invariance. We avoid this lattice structure by prescribing (different) non-usual topologies on each n-particle subspace ($n \geq 2$). Most of our discussion concerns the two-particle subspace. In this case, we introduce the elementary length by taking a coarse[8] topology only on the space of the difference of the two particle coordinates. This means that we

[1] R. B. Blumenthal, D. C. Ehn, W. L. Faissler, P. M. Joseph, L. J. Lanzerotti, F. M. Pipkin, and D. G. Stairs, Phys. Rev. **144**, 1199 (1966). A momentum-transfer of 6 BeV/c corresponds to a Compton wavelength of 3×10^{-15} cm for the internal electron. If we believe that this experiment is just beginning to show the effects of the non-usual topology, then we might take this wavelength as a very rough indication (more likely an upper limit) for the elementary distance λ.

[2] F. E. Low, Phys. Rev. Letters **14**, 238 (1965); N. M. Kroll, CERN Preprint (1966); E. L. Lomon, CERN Preprint (1966).

[3] J. L. Kelley, *General Topology* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1955); N. Bourbaki, *Eléments de mathématique* (Hermann & Cie, Paris, 1948), Vol. III, Pt. 1, Chap. 8; see especially Chap. 9 for a discussion of the pseudo-metrization of topological spaces.

[4] E. C. Zeeman [J. Math. Phys. **5**, 490 (1964); Cambridge University Preprint (1965)] has already considered a non-usual topology for space-time. [See, also in this connection, D. B. Wolf, Preprint, Computer Associates, London (1965).] Zeeman's and our approaches are basically different, but a marriage between them can be contrived, as briefly noted in the text.

[5] See, for example, D. Finkelstein and C. W. Misner, Ann. Phys. (N.Y.) **6**, 230 (1959); U. Enz, Phys. Rev. **131**, 1392 (1963).

[6] In principle, of course, one would be willing to tolerate a breakdown of Lorentz invariance to order λ, if λ were small enough.

[7] Three recent papers by A. Das [J. Math. Phys. **7**, 45, 52, 61 (1966)] give an adequate referencing of existing theories with an elementary length.

[8] A coarse topology is one that contains fewer than the usual open sets, while a fine topology contains more.

are, in some sense, giving up the ability to specify the relative coordinate of the two particles beyond a certain accuracy. However, the topology induced on either one-particle subspace is strictly finer than usual.

Following Zeeman,[4] we define a trajectory to be a continuous map from some parameter space into our topological space. In this sense, all ordinary trajectories are excluded by the topology, and various interesting alternate possibilities arise. We find that, generally speaking, these trajectories imply that the two particles are extended in space over a minimum distance λ.

At the dynamical level, our job is to find equations of motion in the two-particle subspace. The difficulty involved in setting up operators on the space is the following: In one direction in our two-particle space the natural "derivative" is a difference operator, while in a perpendicular direction it is the usual differential operator. The operators along some intermediate ray are some unfamiliar "combination" of these two familiar operators. We avoid some of these problems by making the simple assumption that the equations of motion are separable between the sum and difference coordinate variables. In this way, the equations involve only differential and difference operators. It should be emphasized, however, that this procedure picks out only one from a large class of possible dynamics on the topology. These other dynamics would involve the use of the "fine" operators along an arbitrary ray; we do not have a great deal to say about them in this work.

We confine our considerations of classical mechanics on the new topologies to an analysis of the possible trajectories; then we proceed directly to quantum mechanics. In a simple momentum, energy, and probability conserving formalism, we infer the high-energy scattering of the theory. The surprising result is that at high energies the topology induces an extra effective "potential" between the interacting particles, which serves to enhance backward and high-energy scattering! This is certainly suggestive of the results of Blumenthal *et al.*; but we have not yet calculated this effect quantitatively enough to allow more than a qualitative comparison with the data. Another interesting and surprising feature of the high-energy scattering is the presence of very high mass, but long-lived resonances. In fact, the higher the mass, the narrower the resonance. Certainly, there would be no mechanism in ordinary field theory or S-matrix theory to generate such particles.

We mention that the presence of the elementary length (and the corresponding damped high-energy behavior of the transformation functions to be discussed below) allows, in principle, the elimination of ultra-violet divergences in the theory. The detailed discussion of such a problem would require the choice of a particular theory on the topology (analogous to a choice of a particular Lagrangian in the usual topology); but we are content here in general with the deduction of what seems to be the high-energy behavior of any theory on the topology. By the same token, we do not discuss "intermediate energy" scattering on the topology, as this would also be "theory-dependent."

Along the way in our discussion, we have the opportunity to present various other topologies in which are embedded intrinsic violations of certain discrete symmetries, e.g., time reversal and/or parity invariance. In particular, we mention a topology in which, at a pre-dynamical level, some particles violate parity, and others do not. These topologies may be of some interest in their own right with regard to embedding certain features of the weak interactions in space–time itself.

The order of our presentation is as follows: In Sec. A, we consider a non-usual topology, and the possible particle trajectories, first in one spatial dimension, where we take care to retain translational invariance, then in three dimensions, where we must also keep rotational invariance, and finally in $1 + 3$ space–time, where we complete the embedding of the elementary length in an inhomogeneous Lorentz invariant manner. By building up in stages, we have illustrated the difficulties involved in incorporating an elementary length in more and more complicated spaces, and with progressively more stringent symmetry requirements. In Sec. B, we develop the simple separable dynamics, again in first one, then three, and finally in four dimensions. At the end of Sec. B, we discuss very briefly some of the problems involved in formulating a full field theory, allowing for the creation and destruction of particles.

A. TOPOLOGIES AND TRAJECTORIES ON THE TWO-PARTICLE SUBSPACE

For the purposes of orientation and simplicity, we first discuss intrinsically nonrelativistic topologies, taking a non-usual topology only on space, and leaving the usual topology on time. Things are in fact much simpler in one spatial dimension, and we can learn much in this simple case, which we accordingly discuss before going on to three dimensions. After that, we turn to a relativistic topologization, in which time and space are kept on an equal footing.

1. Nonrelativistic One-Dimensional Motion

Consider the space of two identical particles, located somewhere on a line, with coordinates x_1 and x_2. Ordinarily one assumes the usual (Euclidean)

topology on the two-dimensional space $x_1 \otimes x_2$, and on a time parameter, t. In this section, we want to study another topology for this space, one which contains an elementary length, λ, but which in no way implies an unaesthetic lattice structure on the line itself. In particular, the topology is completely translationally invariant.

We define the non-usual topology by the base[9]:

$$B^{(1)}_{x_1 \otimes x_2} = \{X, x : a < X < b, n\lambda \leq x < (n+1)\lambda\}, \quad \text{(A1)}$$

where $X = (x_1 + x_2)/2$, $x = x_1 - x_2$, a and b are any real numbers, and n is any positive or negative integer. As we have already indicated, λ is to be the elementary length in the theory. In words, we are taking, as the topology of $x_1 \otimes x_2$, the product of the usual topology on the sum variable X, and an apparently coarse topology (actually one that is strictly incomparable with the usual topology) on the difference variable x. We refer to this topological space as $(x_1 \otimes x_2, \lambda)^{(1)}$, the superscript indicating that the individual particle spaces are one-dimensional. We can only separate two points (in the Hausdorff sense) if we can cover each of them with disjoint open sets. The coarseness of the x topology indicates that we are giving up, in some sense to be discussed below, the ability to specify the distance between the particles more accurately than λ.

A question of paramount importance is of course the induced topology[10] on the space of an individual particle. (This is the topology that should be compared with the usual situation.) One sees immediately that a base for the induced topology on, for example, the x_1 space, for fixed x_2, is

$$B^{(1)}_{x_1(x_2)} = \{x_1 : a < x_1 < b, x_2 + n\lambda \leq x_1 < b\}, \quad \text{(A2)}$$

where a, b, and n are defined as in Eq. (A1). We call this topological space $(x_1, \lambda; x_2)^{(1)}$. This notation emphasizes that the induced topology on x_1 is parametrized by x_2, indicating a pre-dynamical linkage between the two particles. Note that this topology is strictly finer (contains more open sets) than the usual topology. Despite the fact that, for fixed x_2, there is a set of "preferred" points—in that an open set[11] extending to the right (or to the left, for that matter) from such a point may or may not contain the point—this topology is translationally invariant. This is because, in any translation, both particles are moved by the same amount, so that the preferred points in the x_1 space move also.

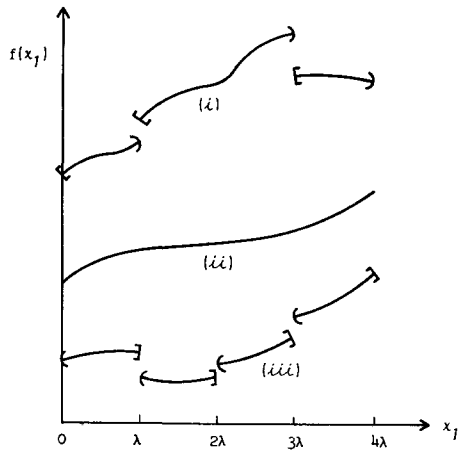


FIG. 1. Continuous functions in $(x_1, \lambda; 0)^{(1)}$.

The fineness of the $(x_1, \lambda; x_2)^{(1)}$ topology is mirrored in the enlarged set of continuous functions[12] on $(x_1, \lambda; x_2)^{(1)}$ into a space with the usual topology. Some single-valued continuous functions on $(x_1, \lambda; 0)^{(1)}$ into $\{f(x_1), U\}$ are shown in Fig. 1. This set is greater than the usual set of continuous functions, since it includes functions which may have arbitrary discontinuities (in the usual topology) at the preferred points $n\lambda$.

Note that $(x_1 \otimes x_2, \lambda)^{(1)}$ is complicated from the topologist's point of view because, although it is normal and regular, it is non-Hausdorff,[13] and is not even T_0! A space with an elementary length is non-Hausdorff in general (because in a Hausdorff space one can "distinguish" between any two points by means of disjoint open sets). By the same token, the space is not metrizable; but, as we see below, it is pseudo-metrizable. On the other hand, the topologies induced on the single particle spaces are Hausdorff and metrizable.

Possible Classical Trajectories in $(x_1 \otimes x_2, \lambda)^{(1)}$

Following Zeeman, we take, as a natural definition of a trajectory in the topological space, a continuous map of a finite interval in (τ, U) into $(x_1 \otimes x_2, \lambda)^{(1)}$, where τ is some invariant parameter and U is the usual topology. For these nonrelativistic topologies, the ordinary time will suffice as the parameter. (We discuss more restrictive definitions of trajectories below.) Note that this is the inverse of the prescription

[9] A subset B of a topology τ is a base for τ if each member of τ is the union of members of B.

[10] By the induced topology on (say) x_1, we mean the relative topology on the space x_1 with respect to $(x_1 \otimes x_2, \lambda)^{(1)}$. The topology on a subset Y of X relative to the topological space (X, τ) is defined to be the family of all intersections of members of τ with Y (see Ref. 3).

[11] An alternate base for $(x_1, \lambda; x_2)^{(1)}$ would be the usual open intervals, plus the points $x_2 + n\lambda$.

[12] A continuous function of one space into another is a mapping of the first (or domain) space into the second (or range) space such that the inverse map of any open set in the range space is open in the domain space.

[13] $(x_1 \otimes x_2, \lambda)^{(1)}$ is the topological product of the Hausdorff (X, U) and the non-Hausdorff (x, λ), so that our "information loss" is in the distance between the two particles.

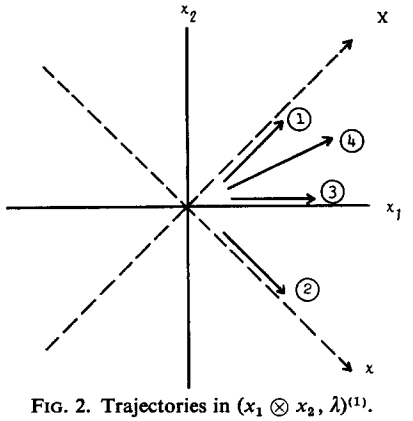


FIG. 2. Trajectories in $(x_1 \otimes x_2, \lambda)^{(1)}$.

giving the continuous functions on $(x_1 \otimes x_2, \lambda)^{(1)}$ *into* a function space with the usual topology.

In general, one can find trajectories in any direction in the $x_1 \otimes x_2$ plane. Various of these are shown in Fig. 2. Trajectory (1) corresponds to both particles' moving with equal velocity, keeping at a constant distance from one another. Trajectories of this sort, in which the particles never change their relative distance "see" the usual topology (on X), and the particle motion appears quite ordinary. Trajectory (2) corresponds to two particles viewed in their center-of-mass system, with the coordinate origin midway between the particles. The topology "seen" by these trajectories is coarse. Trajectory (3) corresponds to the second particle's remaining fixed while the first particle moves. The topology seen by the moving particle is fine. The topology seen on a trajectory like (4)—with both particles moving—is in general fine. Note that, except for peculiar cases like (1), trajectories see non-usual induced topologies.

A representative sampling of the kinds of possible trajectories [corresponding to (3) in Fig. 2] in x_1, for fixed x_2, is given in Fig. 3. Trajectories like that shown in Fig. 3(*i*) correspond to particle two's remaining fixed while particle one jumps forward over each preferred line. In another frame, related by a Galilean

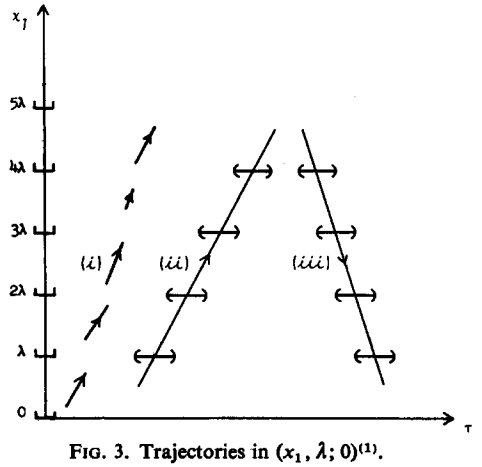


FIG. 3. Trajectories in $(x_1, \lambda; 0)^{(1)}$.

transformation, particle 2 would move quite normally. (The topology allows one particle to move in an ordinary fashion, but not both.) Such a trajectory could be excluded by the requirement that the continuous map into $(x_1 \otimes x_2, \lambda)^{(1)}$ have an inverse in its range. That is, one would want to be able to tell at what time (or times) any particular position was occupied. With this requirement, we immediately exclude any one-to-one map: That is, the particles automatically have an extent in space. For example, "barb" trajectories like (*ii*) (*iii*) do have an inverse and are perhaps of some interest.

Trajectory (*ii*) corresponds to particle one's moving from left to right (while particle two remains fixed at the origin). As it approaches a preferred point, an image of the particle suddenly appears at the preferred point, and remains there while the particle itself passes through the point. The image survives after this for some time, and then it dies. Actually the images can live as long as one likes, but presumably the longer-lived ones would be less physical, and one would therefore want the dynamics to keep the image lifetimes short (i.e., to pick only such a sub-set of the trajectories). Trajectory (*iii*) is the time or parity reversed counterpart of (*ii*). Since any barb trajectory does have a time and/or parity reversed counterpart, it is evident that, unless the dynamics further limits the set of trajectories, the theory of a particle which moves (barb-like) past a fixed particle would be time reversal and parity invariant. (This applies also, of course, to all Galilean transformed configurations.) We see below that there also exist trajectories on this space with no mirror image and hence an intrinsic parity violation.

Other interesting barb-like trajectories can be found by mapping directly into x, X. Some representative maps into $(x, \lambda)^{(1)}$ are shown in Fig. 4. Trajectory (*i*) is the analog of the barbs in x_1—except that this time the trajectory corresponds to two particles moving toward one another. Whenever $x = n\lambda$, an image of each particle is suddenly created, and the two images persist for some time, after which they die simultaneously. Trajectory (*ii*) also involves barb-like motion on the part of both particles. In this case, when the distance between the two particles approaches $n\lambda$, an image appears ahead of each particle and moves backwards, passing through the parent particle when the relative coordinate is $n\lambda$, a short time after which each image dies. Note that all this is taking place in a momentum-conserving manner (since the images always co-exist and move in opposite directions with the same speed). In general, we find momentum-conserving trajectories most easily by mapping into x, X.

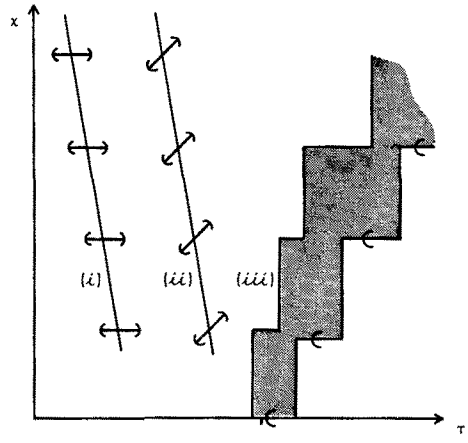


FIG. 4. Trajectories in $(x, \lambda)^{(1)}$.

However, barb trajectories are not bicontinuous between $(x, \lambda)^{(1)}$ and (τ, U)—bicontinuity meaning continuity in either direction between the two topological spaces. If we had the usual topology on x (instead of the λ-topology), then "ordinary" trajectories would be bicontinuous. One might therefore consider bicontinuity to be a reasonable further requirement on a physical trajectory, thus excluding these barbs [note, however, that the barbs on x_1 for fixed x_2 are in fact bicontinuous between $(x_1, \lambda; x_2)^{(1)}$ and (τ, U)]. Bicontinuous trajectories on $(x, \lambda)^{(1)}$ are of the form shown in Fig. 4(*iii*). The horizontal line-segments which are marked C must be taken to include the points that the line designates, while all other boundary lines do not do so. Note that bicontinuity has eliminated even $1:n$ mappings into $(x, \lambda)^{(1)}$ (where n is a finite number, or even $\aleph_0$). The bicontinuous mappings are $1:\aleph_1$ in both directions (where $\aleph_1$ is the continuum infinity). Thus each time, τ, maps into an interval, and the trajectories correspond to two particles, each of which has a minimum spatial extent of $\frac{1}{2}\lambda$. The innermost edges of the two particles are always $n\lambda$ apart, for some integer n. For the two particles to move at all (say toward one another) each must simultaneously extend a "pseudo-pod" of length $\frac{1}{2}n'\lambda$ toward the other. Then each particle pulls its "tail" (again simultaneously) into the region occupied by its pseudo-pod. Thus the motion is amoeba-like past the preferred points.

Note that the space $(x, \lambda)^{(1)}$ is inherently parity breaking [due to the one-sided $\leq$ in Eq. (A1), open sets do not map into open sets under parity reversal]. This fact is apparent in the twin amoeba trajectories: when particle one is to the left of particle two, then the left boundary of the former and right boundary of the latter are included in the respective particles. The mirror image of this configuration would be with particle two to the left of particle one, while the former includes its left and the latter its right boundary—but this configuration is not allowed! The reason for this is that particle one must always include its left boundary, and particle two its right boundary. Thus a theory of these particles would be intrinsically parity-breaking. Such trajectories may be of interest if one wished to embed some of the features of the weak interactions directly in space–time. It is curious to note that, contrary to the case for the twin amoeba trajectories, parity need not be violated for amoeba-like trajectories in x_1, with x_2 constant. This is because these amoeba-like particles can be built to include their boundaries on either side. In three and four dimensions, we want to preserve rotational and Lorentz invariance, and we only consider a parity-invariant topological space.

It is worth mentioning in passing, however, that an intrinsic parity violation can be built into space–time, without any elementary length. For example, if one takes, on a one-particle subspace, the topology defined by the base

$$B = \{x : a \leq x < b\} \tag{A3}$$

for all real a, b, then one finds that the trajectories are of necessity extended in space, although the extension can be made as small as one pleases. If one picks the subset of trajectories that are bicontinuous, then the particles always include their left-hand, but not their right-hand boundaries. These trajectories then have no allowed mirror images. One might imagine that a theory could be set up in which neutrino trajectories, for example, were required to be bicontinuous between a parameter space and the topological space defined by Eq. (A3), while this requirement was not made for particles which interact strongly or electromagnetically. The way to break time-reversal invariance would be, of course, to take one of these one-sided topologies on the time-axis.

Finally, we mention that our topology implies an "action at a distance." For example, imagine two (twin) amoeboid particles, at rest, on opposite sides of the galaxy. By moving one particle a small distance, by means of conventional forces, during which it periodically emits and retracts pseudopods, we can cause the "fixed" particle to emit pseudopods simultaneously. (The "motion" of this particle never carries its center more than $\frac{1}{2}\lambda$ from its initial position.) Action at a distance is not surprising in a nonrelativistic theory. We see below, however, that this characteristic can be preserved in a fully relativistic treatment.

2. Nonrelativistic Three-Dimensional Motion

The problem in three-dimensional motion is to propose a topology that contains an elementary length but is, at the same time, consistent with both translational and rotational invariance. The topology we

wish to study is a natural generalization of the λ-topology on the one-dimensional $x_1 \otimes x_2$. It is defined by the base:

$$B^{(3)}_{x_1 \otimes x_2} = \{\mathbf{X}, \mathbf{x}: a < |\mathbf{X}| < b \text{ and all usual open cones in } \mathbf{X}, \; n\lambda \leq |\mathbf{x}| < (n+1)\lambda \text{ and all usual open cones in } \mathbf{x}\}, \quad \text{(A4)}$$

where we have kept the same definitions of $\mathbf{X}$, $\mathbf{x}$ in terms of $\mathbf{x}_1$, $\mathbf{x}_2$ as in Eq. (A1), and where a, b are any positive real numbers, n any positive integer. Note particularly that in the relative topology of x, the point $x = 0$ is open.

We denote this (six-dimensional) topological space by $(x_1 \otimes x_2, \lambda)^{(3)}$. As before, the space is not Hausdorff, nor T_0. The reasons for this pathology are similar to those for $(x_1 \otimes x_2, \lambda)^{(1)}$, namely that, in taking a non-Hausdorff topology on x, we are giving up some ability to distinguish distances between the particles. Again we find that the space is only pseudo-metrizable. In that the coarsening is radial about either particle, the topology is rotationally invariant, and, of course, translationally invariant also.

The topology induced on the three-dimensional subspace of one particle is again strictly finer than the usual three-dimensional Euclidean topology. We can specify it by means of a base. For fixed x_2, a base for the relative topology on x_1 is the set of usual open ϵ-spheres, centered about any point in the x_1 space, together with each point $\mathbf{x}_1$ that satisfies $|\mathbf{x}_1 - \mathbf{x}_2| = n\lambda$, $n = 0, 1, 2, \cdots$. Hence this induced topology is the usual topology, plus a set of "preferred points," just as in the one-dimensional case. In fact, the induced topology on any straight line, in the x_1 plane, running out of the point $x_1 = x_2$ is the topology of the positive x_1 axis in the one-dimensional example.

The (enlarged) set of continuous functions on $(x_1 \otimes x_2; \lambda)^{(3)}$ into (τ, U) is easily seen to contain, in addition to the usual U-continuous functions functions which may be U-discontinuous across the "preferred" spheres $|\mathbf{x}_1 - \mathbf{x}_2| = n\lambda$.

Trajectories in $(x_1 \otimes x_2, \lambda)^{(3)}$

Again we take Zeeman's definition of a trajectory, as given above. A representative sampling of allowed trajectories that have inverses in their range is shown in Fig. 5. As in $(x_1, \lambda; x_2)^{(1)}$, so here the usual trajectories are in general excluded. Trajectory (i) is in fact usual: so long as a particle does not cross a preferred sphere, its trajectory is entirely ordinary. Note that, in another frame, particle two, originally at rest, will also appear to move normally. Trajectory (ii) is a normal barb trajectory, with images appearing on the preferred spheres. In another frame, particle

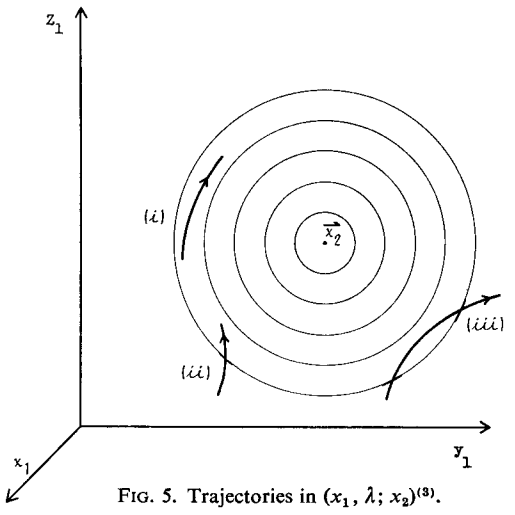


FIG. 5. Trajectories in $(x_1, \lambda; x_2)^{(3)}$.

two would move normally, while particle one would be barbed. The only situations in which both particles move normally are those in which the separation is never equal to an integral number of elementary lengths. Note also that the x_1 trajectories cannot be U-discontinuous in the angular variables about x_2.

In this topology, the barbed trajectories do not intrinsically violate parity invariance. In fact, we have not been able to find a rotationally invariant way of incorporating parity violation into a topology. Moreover, since the topology on the time-axis is still usual, there is no intrinsic violation of time-reversal invariance.

As in one dimension, the requirement of bicontinuity between the parameter space and $(x_1 \otimes x_2, \lambda)^{(3)}$ excludes all the barbed trajectories. Bicontinuous trajectories are easily generated by mapping into x and X. These have the general form shown in Fig. 6 (in $\mathbf{x}$ space). The trajectories correspond to a pair of amoeboid particles with a minimal radial spread (along their line of centers) of $\frac{1}{2}\lambda$ each. The angular extent of one particle about another can be made as small as one pleases. Motion along the line of centers

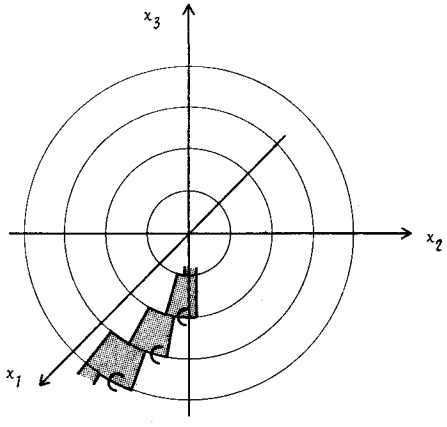


FIG. 6. A bicontinuous trajectory in $(x, \lambda)^{(3)}$.

is accomplished by means of simultaneous, radial pseudo-pods, much as in the one-dimensional case. The angular motion of one particle about the other is entirely normal.

3. Four-Dimensional Relativistic Motion

The problem in four dimensions is to embed the elementary length in a way consistent with full inhomogeneous Lorentz invariance. This involves changing the topology along the time as well as the space axes. This enables one to break time-reversal invariance; but we choose not to do this: our topology is parity and time-reversal invariant.

We specify the non-Hausdorff space $(x_1 \otimes x_2, \lambda)^{(4)}$ by the base

$$B^{(4)}_{x_1 \otimes x_2} = \Big\{ X, x\text{: all usual open } \epsilon\text{-spheres in } X, \begin{pmatrix} x^2 \geq 0, n\lambda \leq x^2 < (n+1)\lambda \\ x^2 \leq 0, n\lambda \leq -x^2 < (n+1)\lambda \end{pmatrix} \text{ and all usual open hyperbolic cones in } x \Big\}. \quad \text{(A5)}$$

Here we use the quadratic form $x^2 = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2$. As before, n is any positive integer. With this topology, we are sacrificing some knowledge of the interval between two particle "observations." It is clear that we have retained explicit Lorentz invariance in our choice of open sets.

Suppose we fix x_2^μ and consider the induced topology in x_1^μ space. A base consists in all the usual open ϵ-spheres, together with all the points lying on the "preferred hyperboloids" $(x_1 - x_2)^2 = n\lambda$, $n = -1, 0, 1, 2, \cdots$.

Note that, as it stands, the induced topology on the light cone ($x^2 = 0$) is very coarse. In particular, in a $1 + 1$ subspace, the only nontrivial open sets on the light cone would be the four quadrants of the light cone and the point $x_1 = x_2$. This means that one could not separate points on a light ray sent between observers at x_1 and x_2. In the full four-dimensional space, however, it is easily seen that x_1 can distinguish (by open sets) the direction of a light ray to x_2, but cannot distinguish points on a given ray. The induced topology on the rest of the topological space $(x_1, \lambda; x_2)^{(4)}$—that is, away from the light cone $x^2 = 0$ is strictly finer than usual, so that in particular, a light ray from any point x_1 aimed *away* from x_2 travels over a very fine topology. This is also the case for general particle trajectories.

Actually, it is a simple matter to refine the topology on the light cone as far as one wishes, without essentially changing the induced topology on other subspaces.[14] In this paper, we leave the question of the desirability of refining the light-cone topology as an open question.

The induced topology of most physical interest is on any subspace $x_1^{(0)} = x_2^{(0)}$ (i.e., the two particles are usually considered at the same time). One sees easily that this subspace is, for all times $t = x_1^{(0)} = x_2^{(0)}$, simply $(x_1 \otimes x_2, \lambda)^{(3)}$. Thus we have succeeded in embedding our three-dimensional topological space in space–time in a relativistically invariant way.

It is clear from the base $B^{(4)}_{x_1 \otimes x_2}$ that we have embedded into the space some information about light-cone structure. If one wished to complete the job of embedding the light-cone structure into space–time, one would need to consider the ideas of Zeeman, according to which the topology is refined as far as is possible, consistent with the requirement that the relative topology on any time or space axis is usual. There seems no reason why this thoroughgoing refinement could not be combined with our notion of embedding an elementary length in a two-particle subspace. One would simply define the topology on $x_1 \otimes x_2$ to be the finest consistent with the usual topology on X and an interval-coarsened topology on x, considered along any space or time axis. We do not consider this idea any further here.

Relativistic Trajectories

For simplicity, we limit ourselves to mapping functions that are bicontinuous between the space $(x_1 \otimes x_2, \lambda)^{(4)}$ and (τ, U), where τ is some invariant parameter, for example, the proper time of one of the two particles. This means that in the variable x, with $t_1 - t_2 = 0$, the trajectories are amoeba-like, just as we found above in $(x_1 \otimes x_2, \lambda)^{(3)}$. For identical particles, one would want to require the spread of each particle along the line connecting them to be the same in the center-of-mass frame.[15] In this frame, the minimum radial spread of each particle is $\frac{1}{2}\lambda$.[16] It follows that the minimum radial spread of one particle, in its rest frame, is $\frac{1}{2}\lambda\gamma$, where $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}} > 1$, and v is the velocity of the center-of-mass frame relative to this rest frame. Hence the spread of, say, the target particle in its rest frame increases (without bound) as the momentum of the incident particle increases. Again we see that the topology is coupling the particles at a pre-dynamical level.

Consider now particle one (say) in its rest frame. Suppose we attach synchronized clocks to different

[14] Although if the light cone topology were so fine as to be discrete, this would add certain single point open sets to the relative topologies on lines in $x_1 \otimes x_2$ that intersect $(x_1 - x_2)^2 = 0$.

[15] As in the one- and three-dimensional cases, one can find trajectories for which one particle is not extended in space, while the other is. This situation would persist in any Lorentz frame.

[16] The angular spread of one particle about another can be made as small as one pleases. Note that the existence of a radial, but not an angular, spread in one frame ensures the same thing in all frames: in special relativity, straight lines always transform into straight lines.

parts of the (extended) particle. Then, in a moving frame we find that the particle is Lorentz contracted, and also that the leading edge is younger (earlier proper time) than the trailing edge. Of course, this must happen in any relativistic theory of extended particles. For example, it is presumably true for the dressed particles of ordinary field theory, but it is to be noted that in our case, even the bare particles in a field theory would in general be extended.

Note that, as mentioned in the one-dimensional nonrelativistic case, we still have action-at-a-distance (in the same sense as above), even though the open set structure is completely Lorentz invariant. Certainly, as we see in the section on dynamics, we can write equations of motion on the topology which are explicitly frame-independent. In this connection, it is important to remember that there is no contradiction between the Lorentz group and information transfer faster than light. Indeed, the Lorentz group admits of spacelike representation.[17] Given (say) any "particle" of spacelike mass, one can always find Lorentz frames in which its velocity is infinite. (In our case these frames are the center-of-mass frames.) This of course is action-at-a-distance. Although causality is broken in our theory, the violation is in general only over intervals of order λ.

Finally, we note that the topological space $(x_1 \otimes x_2, \lambda)^{(4)}$ is explicitly time-reversal and parity invariant (in the sense that open sets map into open sets under these transformations). This can be phrased, as above, in terms of trajectories and mirror trajectories, etc. An example of a relativisitic topology which would intrinsically break time-reversal invariance can be defined by the base

$$\begin{aligned}\mathfrak{B} = \{X, x:\ &\text{usual open } \epsilon\text{-spheres in } X,\\ &\text{usual open } \epsilon\text{-spheres in } x \text{ for } x^2 < 0,\\ &\text{and for } \quad x^2 \geq 0,\\ &\qquad n\lambda \leq x^2 < (n+1)\lambda, \quad x_0 > 0,\\ &\qquad n\lambda < x^2 \leq (n+1)\lambda, \quad x_0 < 0,\\ &\text{together with usual open hyperbolic cones}\}. \qquad \text{(A6)}\end{aligned}$$

In this space, the trajectories involving spatial extent for both particles would violate time-reversal invariance, while the trajectories for which one particle was a point particle would always go over into allowed time-reversed trajectories. [This is much like the situation with parity in $(x_1 \otimes x_2, \lambda)^{(1)}$.]

B. QUANTUM DYNAMICS ON THE NON-USUAL TOPOLOGIES

Our order of presentation in this section parallels that of Sec. A. After a brief discussion of the concept of metrization (explicitly only for the one-dimensional case), we discuss dynamics first in the one spatial dimensional case, then in higher dimensions. The metrization for each higher dimensional case is left to the corresponding section on dynamics.

1. Metrization

To study the dynamics of the particles on these topologies one must first seek a metric (or pseudometric) on the spaces. In general, each of the spaces discussed above is in fact only pseudometrizable—that is, if $d(x, y)$ is the "best distance function" available on the space (compatible with its topology),[18] then it is always necessary that for some distinct points x, y, $d(x, y) = 0$. As in Sec. A, we discuss first the one-dimensional motion and then work up to more dimensions.

What is a metric for $(x_1, \lambda; x_2)^{(1)}$? Since this topological space contains the continuum of points x_1, the ordinary Euclidean metric is certainly a metric. It is not, however, the metric which metrizes the space, because the metric topology would be the usual topology, whereas $(x_1, \lambda; x_2)^{(1)}$ is strictly finer than usual. [The metric topology associated with a metric $d(x, y)$ is that topology defined by the base of open ϵ-spheres $d(x, y) < \epsilon$.] We want to use a metric which metrizes the space, because this metric corresponds to making maximal use of our open sets.[19] An ϵ-parametrized class of metrics with this property is ($\epsilon > 0$)

$$d(x_1, \bar{x}_1) = \begin{cases} |x_1 - \bar{x}_1| \text{ if neither } x_1 \text{ nor } \bar{x}_1 \\ \qquad\qquad \text{equals } n\lambda + x_2, \\ |x_1 - \bar{x}_1| + \epsilon \text{ if either } x_1 \text{ or } \bar{x}_1 \\ \qquad\qquad \text{equals } n\lambda + x_2, \\ 0 \text{ whenever } x_1 = \bar{x}_1. \end{cases} \qquad \text{(B1)}$$

To see what this class of metrics means physically, imagine the following thought experiment: An observer riding on the first particle (x_1) watches the second particle (at point x_2) as he approaches it. Because x_2 is a preferred point, the observer sees the particle at a distance $|x_1 - x_2| + \epsilon$. As x_1 approaches x_2, the observer sees the distance decreasing to ϵ, then the distance changes abruptly to zero, and then to ϵ in the other direction, again abruptly. That is, the observer finds it impossible to approach the second particle smoothly.[20] This is, of course, essentially the statement that the continuous functions on the space into (t, U) may have U-discontinuities at the preferred points. In general the induced topology along any

[17] E. Wigner, (Lecture) Seminar on Theoretical Physics (Trieste, 1962), p. 64.

[18] That is, the (pseudo-) metric topology should be the λ-topology.

[19] Since $(x_1 \otimes x_2, \lambda)^{(1)}$ is a T_1-space, we are guaranteed that it is metrizable.

[20] This emphasizes the difference between finitely curving a space (with the usual topology), and changing the topology.

direction in $x_1 \otimes x_2$ except (1) and (2) of Fig. 2 is fine and can be metrized in a similar fashion.

The space $(x, \lambda)^{(1)}$ cannot be metrized. However, a pseudo-metric which pseudo-metrizes the space is[21]

$$d(x, \bar{x}) = \lambda[x/\lambda] - \lambda[\bar{x}/\lambda], \tag{B2}$$

where $[x]$ is the number theory function meaning the largest integer not greater than x. In particular, the pseudo-metric distance between two points that lie within the same preferred interval, say $n\lambda \leq x < (n+1)\lambda$, is zero. However, the triangular inequality is never violated, and $d(x, \bar{x})$ is truly a pseudo-metric.

We find it convenient to set up a dynamics which is separable (in a sense to be explained below) in x, X, since adequate operators are to hand (difference and differential operators, respectively). This is only a very simple way to proceed, but we envisage that a more thorough-going dynamics, freed from any such artificial separability constraint, would in general use some unfamiliar "combination" of differential and difference operators (such as, e.g., a U-discontinuity operator, which, when operating on continuous functions, would be different from the null operator in most directions). We do not discuss a classical mechanics on the topology, but, rather, go directly to a quantum mechanics. At this level we verify that our simple dynamics seems to describe quantized amoeboid trajectories.

2. One-Dimensional Quantum Mechanical Motion

We want to build a quantum mechanical description of the scattering of two particles of momentum and energy (E_1, p_1) and (E_2, p_2), respectively, on the topological space $(x_1 \otimes x_2, \lambda)^{(1)}$. The easiest way to do this is to suppose that the wavefunction can be written as the product

$$\psi(x_1 x_2, t; p_1 p_2) = \Psi(X)\psi(x)\exp\{(i/\hbar)(E_1 + E_2)t\}. \tag{B3}$$

As ordinarily, we demand that the wavefunctions be continuous on their respective topologies into the usual topology. That is, $\Psi(X)$ can be taken in the usual form

$$\Psi(X) = \exp\{-(i/\hbar)PX\}, \quad P \equiv p_1 + p_2, \tag{B4}$$

whereas $\psi(x)$ is a block-type function.

Our job now is to set up a dynamics on x. First we must seek a suitable momentum operator on the space. With the pseudo-metric discussed above, we clearly cannot define an ordinary differential operator. The closest analog to the usual correspondence $p \to -i\hbar(d/dx)$ that we can define on the space is $p \to -i\hbar\mathfrak{D}$, where $\mathfrak{D}$ is the symmetric difference operator, such that for any continuous function $f(x)$

$$\mathfrak{D}f(x) = [f(x + \lambda) - f(x - \lambda)]/2\lambda. \tag{B5}$$

This momentum is Hermitian with respect to the inner product

$$(\psi_a, 0\psi_b) = \int_{-\infty}^{+\infty} \psi^*(x)0(x)\psi(x)\,dx = \lambda \sum_{n=-\infty}^{\infty} \psi^*(n\lambda)0(n\lambda)\psi(n\lambda). \tag{B6}$$

Note that, with this inner product, the right and left difference operators are not Hermitian.

If we define a position operator, q, with the diagonal representation

$$q = \lambda[x/\lambda], \tag{B7}$$

then we find the following Lie algebra involving p and q

$$[q, p] = i\hbar\mathbf{a}, \quad [\mathbf{a}, q] = \frac{i\lambda^2}{\hbar}p, \quad [p, \mathbf{a}] = 0, \tag{B8}$$

where $\mathbf{a}$, proportional to the commutator of p and q, is an Hermitian averaging operator. It is defined in the q-diagonal representation by

$$\mathbf{a}f(x) = \tfrac{1}{2}[f(x + \lambda) + f(x - \lambda)]. \tag{B9}$$

Note that, as $\lambda \to 0$, $\mathbf{a} \to 1$ (unity) and we recover the usual relations of quantum mechanics. In that the operators p, $\mathbf{a}$ are nonlocal (they couple functional values over a range 2λ), one expects the particles to be spread in a way that involves both $\hbar/m$ and λ.[22]

Suppose that $|q'\rangle$ is an eigenket of q with the eigenvalue q', and $|p'\rangle$ a ket in the dual (momentum) space with eigenvalue p'. Then it is easy to show that

$$\langle p' \mid q' \rangle = g(p') \exp\left\{\frac{iq'}{\lambda}\sin^{-1}\left(\frac{\lambda p'}{\hbar}\right)\right\}, \tag{B10a}$$

where

$$g(p') = \pi^{\frac{1}{2}}\left(\frac{\hbar^2}{\lambda^2} - p'^2\right)^{-\frac{1}{4}} \tag{B10b}$$

is a normalization function which guarantees the unitarity of this transformation function:

$$\langle p' \mid p'' \rangle = \sum_{n=-\infty}^{\infty} \langle p' \mid n\lambda\rangle\langle n\lambda \mid p''\rangle = \delta(p' - p''), \tag{B11a}$$

$$\langle m\lambda \mid n\lambda\rangle = \int_{-\infty}^{\infty} \langle m\lambda \mid p\rangle\, dp \langle p \mid n\lambda\rangle = \delta_{mn}. \tag{B11b}$$

The cut structure of $\langle p' \mid q' \rangle$ and the evaluation of these sums are discussed in the Appendix.

One can guarantee that $\langle p' \mid q' \rangle$ is also the wavefunction of a freely moving particle of energy[23]

[21] The pseudo-metric function (B1) is not bicontinuous between its range and domain spaces. (The Euclidean metric is bicontinuous on the usual topology.) If one wanted to define a distance function, $\hat{d}(x, \bar{x})$, (*on* the usual topology) that was bicontinuous between the range and domain spaces, one would in fact be forced to use a multi-valued function, which is essentially an amoeboid trajectory (see Fig. 4) turned on its side. $\hat{d}(x, \bar{x})$ is not a metric or pseudometric (since, for example, the triangular inequality cannot be unambiguously satisfied—although the violation is always only of order λ). It is curious to note also that had we taken the base for x slightly differently, e.g., $(n-1)\lambda \leq x \leq n\lambda$, then the space would not even be pseudo-metrizable.

[22] Raising and lowering operators for q are $s_\pm = a \mp i\lambda p/\hbar$ which generate the eigenvalues $n\lambda$, $n = \cdots -1, 0, +1, 2, \cdots$ starting from $q = 0$. These are closely related to the right and left difference operators. With $s_\pm$, one sees immediately that the Lie algebra (B8) is reducible. In fact, $[q, s_\pm] = \pm\lambda s_\pm$, $[s_+ s_-] = 0$.

[23] We are using the relative coordinate language; remember that $p = \frac{1}{2}(p_1 - p_2)$.

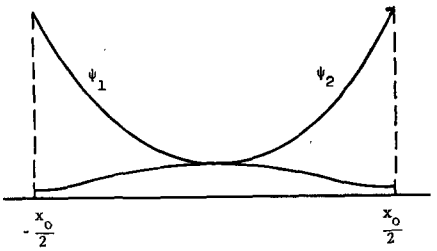


FIG. 7. Two-particle scattering above the critical energy.

$E = p'^2/2m$, m being the reduced mass, by taking as the Schrödinger equation

$$(p^2/2m)\psi = -(\hbar^2/2m)\mathfrak{D}^2\psi = E\psi. \quad \text{(B12)}$$

If one adds the time-dependence appropriate to the difference variable, it is easily seen that, with (essentially) usual assumptions at $x = \pm\infty$, there is overall probability conservation.

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} dx\, |\psi(t, x)|^2 = 0. \quad \text{(B13)}$$

This also holds if a time-independent potential is added to the Schrödinger equation (B12).

However, it is not possible to define a probability current that is meaningful to distances of order λ.[24] That is, although there is never any probability loss or gain for regions large compared with λ, one cannot watch the probability flux too carefully.

We turn now to a discussion of the potential-free motion of the two particles, in particular to a discussion of the function $\langle p' \mid q'\rangle = \psi_{p'}(q')$. For all momenta $|p'| < \hbar/\lambda = p_c$ the particles move past one another freely, as in the usual topology. Above the critical energy $E_c = p_c^2/2m$, $\psi_{p'}(q')$ has an exponential growth or decay in q'. To keep probability conservation, we retain only the decay (in regions allowing arbitrarily large q'). For example, imagine two accelerators, at a distance x_0 from each other, directing beams of particles of equal energy at one another (we choose this center-of-mass experiment purely for simplicity). For clarity, suppose that all ordinary interactions between the particles are negligible (this is actually a high-energy approximation). At energies $E < E_c$, the particles move past one another freely, so the transmission coefficient is unity. For $E > E_c$, one finds backward scattering! In fact, the reflection and transmission coefficients depend on the location of the accelerators and detection apparatus (assumed located at the accelerators). The wavefunctions of the two beams[25] are shown in Fig. 7. (The continuous curves of the figure should of course be the step-type functions appropriate to x. For simplicity we forget this in the subsequent figures.) The absence of exponential tails outside $-\frac{1}{2}x_0 < x < \frac{1}{2}x_0$ is the boundary condition that the detecting apparatus stops any particle which reaches it. That is, the wavefunction of each beam decays exponentially after it leaves its accelerator. This is strikingly analogous to barrier penetration in ordinary quantum mechanics. Our situation is like having a strong repulsive potential of magnitude

$$V = \frac{1}{2m}\left[p^2 + \left\{\frac{\hbar}{\lambda}\sinh^{-1}\left(\frac{\lambda p}{\hbar}\right)\right\}^2\right] > E \quad \text{(B14)}$$

between the two particles. (Actually this "topological potential" fills the whole of space.) Thus, holding x_0 constant but increasing the beam momentum, one finds fewer and fewer particles from accelerator 1 reaching accelerator 2, and vice versa. The same thing happens for constant p as the distance of separation (x_0) increases. Thus, in this topology, experiments at very high energy become "configuration dependent" —observers close to the scattering region may observe different scattering patterns than more distant observers. All this is probability conserving, just as in the case of barrier penetration; that is, any particle from accelerator 1 not reaching accelerator 2 is found in the reflected beam detected at 1. Also, because the "barrier" is a function only of $|x_1 - x_2|$, the scattering is momentum conserving, just as in ordinary quantum mechanics. Note, however, that at supercritical energies the back-scattering will take place independent of the distance of separation of the two particles. (The topology is effectively propagating information at infinity velocity.) This is action-at-a-distance again, just as discussed in Sec. A.

We can say a few things qualitatively about scattering in the presence of an ordinary potential [to be added to Eq. (B12)]. We do not study any particular potentials (although the Schrödinger difference equation is in general no more difficult to solve than the corresponding differential equation), but it is interesting to note the qualitative effect of attractive and repulsive potentials on the "topological scattering." Consider a scattering in which the free wave approaches an attractive potential step with a somewhat sub-critical momentum; on the far side of the step the momentum $p = [2m(E - V)]^{\frac{1}{2}}$ is larger than on the near side. If it exceeds p_c, the effect is repulsion, or, more accurately, backward scattering of the attractive step. This occurs for a small step in ordinary potential theory, but in our topology, the repulsive effect is enhanced by taking the attraction stronger! On the other hand, in the vicinity of a repulsive potential, the topological scattering is not set in until $E = p_c^2/2m + V > p_c^2/2m$.

[24] This sort of difficulty is common to any nonlocal theory. See, for example, P. Kristenson and C. Møller, Dan. Mat. Fys. Medd. 27, (1952), and C. Bloch, *ibid.*, 27, (1952).

[25] We assume for clarity that the particles of each accelerator are distinguishable, so that we can talk of individual wavefunctions.

We also note that our quantum mechanics does indeed seem to describe quantized amoeboid trajectories (rather than quantized barbs or quantized discrete trajectories). To see this, one tries to construct the tightest possible wave packets. Clearly, these cannot be tightened below λ, regardless of how large a momentum spread is allowed. Moreover, reattaching the correct time dependence, we see that the "particle" is always present (i.e., does not disappear and reappear rapidly, as it would for a discrete trajectory).

Finally, we mention that the wave equation (B12), taken together with (B3) and (B4), does not allow the calculation of a one-particle wavefunction or wave equation (say, for x_1 independent of x_2). The topology has inextricably interwoven the two particles. This peculiarity will carry over into four dimensions.

3. Three-Dimensional Quantum Mechanical Motion

In the three-dimensional case, we again separate off the (ordinary) dynamics in X as above, and pseudometrize the difference variable $\mathbf{x}$ in analogy with the one-dimensional situation:

$$d(\mathbf{x}, \mathbf{x}') = [(x_1 - x_1')^2 + (x_2 - x_2')^2 + (x_3 - x_3')^2]^{\frac{1}{2}},$$
$$\left.\begin{aligned} x_1 &= \lambda[r/\lambda] \sin\theta \cos\varphi \\ x_2 &= \lambda[r/\lambda] \sin\theta \sin\varphi \\ x_3 &= \lambda[r/\lambda] \cos\theta \end{aligned}\right\} \begin{array}{l}(r, \theta, \varphi) \text{ are the} \\ \text{spherical polar} \\ \text{coordinates of } \mathbf{x},\end{array} \tag{B15}$$

and similarly for the primed quantities.

It turns out that it is not possible to find a set of three commuting momenta in this space. For example, define

$$p_{x_3} \equiv \frac{\hbar}{i}\left[\cos\theta \mathfrak{D}_\imath - \frac{1}{\imath}\sin\theta\frac{\partial}{\partial\theta}\right], \quad \imath \equiv \lambda\left[\frac{r}{\lambda}\right], \tag{B16}$$

where $\mathfrak{D}_\imath$ is the symmetric difference operator

$$\mathfrak{D}_\imath f(\imath, \theta, \varphi) = \begin{cases} [f(\imath + \lambda, \theta, \varphi) - f(\imath - \lambda, \theta, \varphi)]/2\lambda & \imath \neq 0, \\ [f(\imath + \lambda, \theta, \varphi) - f(\imath, \theta, \varphi)]/\lambda & \imath = 0 \end{cases} \tag{B17}$$

(and similarly for $p_{x_1}p_{x_2}$ taking $\partial/\partial r \to \mathfrak{D}_\imath$ in the usual spherical polar expressions). Although these operators are Hermitian with respect to the usual inner product, they do not commute. Physically, this is because they are not really generators of orthogonal translations. (This in turn is because the r translation cannot be infinitesimal.) On the other hand these "momenta" commute for very large r and can be thus used to classify "plane" waves[26] according to a momentum vector $\mathbf{p}$, at least very far away from the target.

One can build parallel translation operators on the space in the following manner: Define D_{x_1}, the translation operator in x_1, as that operator which, having changed r by λ, readjusts θ, φ so that x_2, x_3 are in the end unchanged.[27] One finds that D_{x_1} depends on x_2 and x_3, which means that it fails to commute with D_{x_2}, D_{x_3} (defined in a similar way). Physically, this means that, for example, a translation in x_1 followed by a translation in x_2 is not equivalent to the operations in the opposite order—i.e., the difference space x is curved. The noncommutativity of D_{x_1}, D_{x_2} is illustrated in Fig. 8.

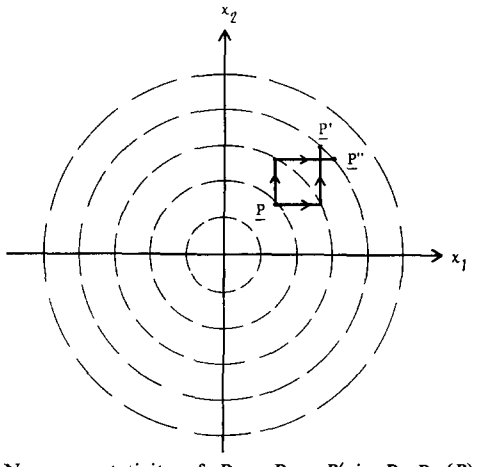


FIG. 8. Noncommutativity of D_{x_1}, D_{x_2}. P' is $D_{x_2}D_{x_1}(P)$, while P'' is $D_{x_1}D_{x_2}(P)$.

We take, as the coarsened form of the free Schrödinger equation on the difference variable ($p = \hbar k$),

$$\mathfrak{D}_\imath^2\chi + \frac{1}{\imath^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial\chi}{\partial\theta} + \frac{1}{\imath^2\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\chi + k^2\chi = 0, \tag{B18}$$

where $\chi = \imath\psi$. Note that, although the energy $E = p^2/2m$ is well defined, the solutions are not (except at very large r) eigenfunctions of p_{x_i}. Thus, although we can classify each wave according to its (asymptotic) momentum, we cannot, in the interaction region, resolve E accurately into the sum of the squares of any momenta. This is a general feature of spaces with noncommuting momenta. There are of course other ways of coarsening the Schrödinger equation, but the qualitative results we extract from the dynamics are independent of the particular dynamics we choose. With essentially the usual boundary conditions on the wavefunction, one can show that overall probability is conserved with this wave equation.

We turn now to solutions of the wave equation. It is easy to find spherical waves in the topology. The waves

$$\chi(\imath) = \exp\left\{\pm\frac{i\imath}{\lambda}\sin^{-1}\left(\frac{p\lambda}{\hbar}\right)\right\}, \quad E = \frac{p^2}{2m} \tag{B19}$$

are exact solutions of (B18). Thus, for super-critical

[26] We see the form these take below.

[27] These operators do not in general approach the usual translation operators for large r. That is, for example, far out on the x_3 axis, the only allowed translations in x_1 are very large. Moreover, these operators are not Hermitian.

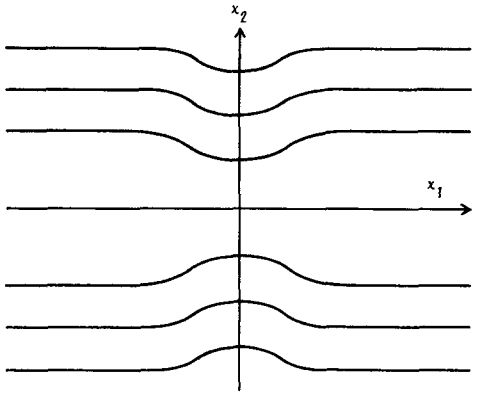


FIG. 9. Bending of flux-lines in the three-dimensional topological potential.

momenta, waves trying to approach the scattering center will tend to be reflected, and waves attempting to radiate from the center will tend to be reflected back towards the center.[28]

Of more physical interest are the analogs of plane waves on the topology. These are much more difficult to obtain. Towards this end, we guess a solution to Eq. (B19) of the form

$$\chi(\imath, \theta, \varphi) = \imath \exp\left\{\frac{i\imath}{\lambda} \sin^{-1}[\lambda p f(\theta)]\right\}. \quad \text{(B20)}$$

This function solves Eq. (B18) up to terms of order $(\lambda/\imath)$ with the proviso that

$$[(d/d\theta)f(\theta)]^2 = (1 - f^2)(1 - \lambda^2 k^2 f^2). \quad \text{(B21)}$$

Thus f is an inverse elliptic function. We content ourselves with a discussion of f in certain energy ranges.

To second order in λk, one finds

$$\psi(r, \theta, \varphi) \underset{r \gg \lambda}{\sim} \exp\left\{\frac{ir}{\lambda} \sin^{-1}(\lambda k \cos\theta)\right\}, \quad \text{(B22)}$$

which becomes more and more like a plane wave as λk tends to zero. If we loosely define a current[29] as proportional to $\psi^* \tilde{\nabla} \psi - \psi \tilde{\nabla} \psi^*$, where $\tilde{\nabla}$ is a coarsened gradient in spherical coordinates [as in Eq. (B16)], then one finds that the flux lines corresponding to the almost plane-wave solution (B22) bend into the scattering center as shown in Fig. 9, but straighten out again as usual for very large r. Physically, this is like having a potential which is attractive for large $|x|$ and repulsive for small $|x|$, balanced in such a way that there is no scattering in the absence of a real potential. In the presence of such an ordinary potential, it is clear that scattering would be enhanced in the lower partial waves[30] (since the effective impact parameter of each flux line is reduced by the "topological potential").

[28] This latter situation is reminiscent of a bound-state wave-function—the binding being done by the topological potential.

[29] Recall that probability currents cannot be believed over distances of order λ.

[30] As we see below, the breakup into partial waves is essentially normal.

At supercritical energies the solution (B22) is also valid so long as θ is in the cone $|\theta - \pi| \ll 1/\lambda k$. In this range the $\sin^{-1}$ is imaginary and we find damping in this cone. That is, part of the incoming beam is back-scattered through the angles in this cone. Presumably this cone of backward (and large angle) scattering first appears at the critical energy with zero solid angle along the backward z axis, the solid angle of the back-scattering increasing as the energy increases. We have not yet been able to calculate the exact dependence of the solid angle on energy. Finally we note that, as the energy goes to infinity, the space becomes more and more opaque to the scattering wave, which is back-scattered completely—just as in the one-dimensional case.

At this point in our discussion, it is worth considering a variant of the present model. In particular, suppose our non-usual topology extended only out to a radius (say) b in x space. (This could roughly simulate an elementary length which decreased to zero for larger particle separations.) In terms of the topological potential, we would be chopping off most or all of its long-range attraction, leaving only the repulsive core. In this case, there would be scattering in the absence of an ordinary potential; the scattering would be like that from a soft repulsive sphere of radius b and hardness proportional to $(\lambda k)^2$.[31] The overall scattering effect is shown in Fig. 10. We have in mind a smooth joining of flux tangents at the sphere boundary, but we do not go into the difficult questions associated with a rigorous embedding of the non-usual sphere in the usual topology.

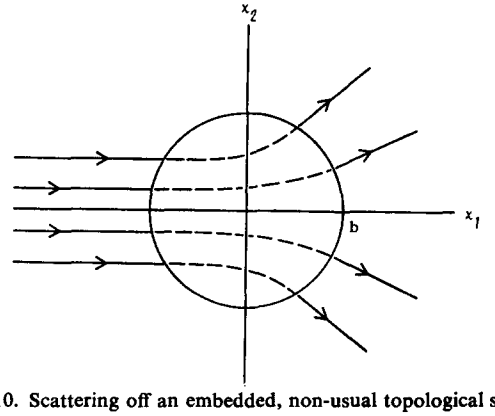


FIG. 10. Scattering off an embedded, non-usual topological sphere.

As the scattering energy increases further, the bending effect becomes more pronounced, until finally, at the critical energy, a cone of large angle scattering opens up. (The cone begins purely backward as above, but subtends a larger solid angle with increasing energy.) At still higher energies, the embedded sphere

[31] We learn more about the structure of this topological potential in the discussion below on partial wave analysis.

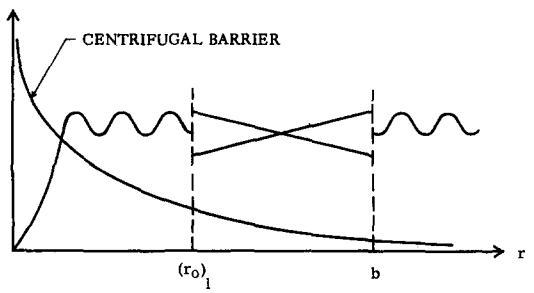


FIG. 11. High-energy scattering off the centrifugal barrier in the lth partial-wave.

becomes totally opaque, large angle-scattering everything that hits it. As usual for hard-sphere scattering, there will be a forward diffraction peak as well at high energies.

We can discuss some other interesting properties of this model in terms of the individual partial waves. Because our coarsening of the space is spherically symmetric, and the angular part of the Schrödinger equation is unchanged, we can expand the wavefunction in the usual Legendre polynomials:

$$\psi(r, \theta) = \sum_{l=0}^{\infty} (2l + 1)P_l(\cos\theta)\psi_l(r). \qquad \text{(B23)}$$

The (free) partial wave equation for $\chi_l(\imath) = \imath\psi_l(\imath)$ is

$$\mathfrak{D}_\imath^2\chi_l(\imath) + [k^2 - l(l + 1)/\imath^2]\chi_l(\imath) = 0. \qquad \text{(B24)}$$

If we keep the topology throughout all $\imath$, the asymptotic solutions for large $\imath$ are

$$\chi_l(\imath) \underset{\imath\to\infty}{\sim} \exp\{\pm(i\imath/\lambda)\sin^{-1}(\lambda k)\}. \qquad \text{(B25)}$$

The partial wave equation is very much like the one-dimensional wave equation discussed above—but this time with a repulsive centrifugal barrier. It is interesting to discuss the interplay of this barrier with the topology. For this it is convenient to go to the embedded case. At super-critical energies, the free solutions will look much as in Fig. 11. That is, the wave (of energy $E > E_c$) propagates freely inward to b, at which point it senses a repulsive thick spherical shell. This shell extends into roughly

$$(\imath_0)_l = [l(l + 1)\hbar^2/2m(E - E_0)]^{\frac{1}{2}} \qquad \text{(B26)}$$

at which point the momentum p becomes again less than critical and the wave propagates freely (again) until it hits the centrifugal barrier itself—after this it decays into the barrier as usual. As the energy increases at fixed l, the spherical shell becomes thicker and higher, the height growing with E essentially as in Eq. (B14). At fixed energy, the shell is thicker for smaller l. Thus this effective potential induced on the topology is both energy-dependent and nonlocal.[32]

A very curious feature of the shell is its hollow center, in which resonances (of energy) can easily be trapped. (Actually, they are trapped between the inside of the topological potential barrier and the centrifugal barrier.) The really peculiar thing about these topological resonances is their long lifetime: Because the shell is thicker and tougher at higher energies, the higher mass a resonance has, the longer it lives! Certainly there is no mechanism in ordinary dynamics which could produce such a particle.

[32] In the sense that the partial waves are decoupled, the induced potential is still central.

4. Relativistic Quantum Mechanical Motion

Analogously to the one- and three-dimensional cases, we pseudo-metrize the difference variable x, in such a way that the interval x^2 is discretized. Moreover, we for the most part content ourselves with checking that the equations in a $1 + 1$ space (one space and one time dimension) still yield the back-scattering effect.

A serious difficulty in a naive time quantization is always that probability may leak in time. For example, if one quantizes a single particle's time (t) with the "Schrödinger" equation

$$i\hbar\mathfrak{D}_t\psi = E\psi, \qquad \text{(B27)}$$

then one finds the solutions decaying in time for $E > E_c$. This is highly undesirable, and we want to check that it is not occurring in our "difference-interval" quantization.

We study the scattering of two free bosons of mass μ on $(x_1 \otimes x_2, \lambda)^{(2)}$. To get a suitable wave equation, we first factor the solution of the Klein–Gordon equation[33] on the usual topology

$$(\square_1^2 + \square_2^2 + 2\mu^2)\varphi(x_1, x_2) = 0 \qquad \text{(B28)}$$

into a product of a center-of-mass factor $\exp\{i\mathbf{P}\cdot X\}$ and a solution of the equation in the relative coordinate:

$$(\square_x^2 + \mu^2 - \tfrac{1}{4}s)\varphi(x) = 0, \qquad \text{(B29)}$$

where $s = (p_1 + p_2)^2$. Our main interest is in the system as it appears to some observer whose time is t_0, i.e., $t_1 = t_2 = t_0$ and $t_1 - t_2 = 0$. Thus we specialize for the moment to x^2 spacelike. In particular, for spatial x positive, the equation in the difference variable can be written

$$\frac{1}{\sigma}\frac{\partial}{\partial\sigma}\sigma\frac{\partial\varphi(\sigma,\theta)}{\partial\sigma} - \frac{1}{\sigma^2}\frac{\partial^2}{\partial\theta^2}\varphi(\sigma,\theta) + (\tfrac{1}{4}s - \mu^2)\varphi(\sigma,\theta) = 0, \qquad \text{(B30)}$$

where $\sigma = (-x^2)^{\frac{1}{2}}$ and $\theta = \tanh^{-1}(x_0/x_1)$. We can separate off the angular dependence by the "partial wave analysis"

$$\varphi(\sigma,\theta) = \frac{1}{\sigma^{\frac{1}{2}}}\int_0^\infty dn\, e^{-n|\theta|}R_n(\sigma);\ R_n''(\sigma) + \left[(\tfrac{1}{4}s - \mu^2) - \frac{n^2 + \frac{5}{4}}{\sigma^2}\right]R_n(\sigma) = 0, \qquad \text{(B31)}$$

[33] Our quadratic form is $p^2 = p_0^2 - |\mathbf{p}|^2 = \mu^2$.

where a "prime" means differentiation with respect to σ. With this form in mind, we are in a position to write a suitable generalization of the 2 particle Klein–Gordon equation on our space. We define $\varphi(x_1x_2)$ as the product of $\exp\{iPX\}$ and $\varphi(\sigma, \theta)$ as shown in (B30), but this time with R_n satisfying

$$\mathfrak{D}_\sigma^2 R_n(\sigma) + [(\tfrac{1}{4}s - \mu^2) - (n^2 + \tfrac{5}{4})/\sigma^2]R_n(\sigma) = 0. \quad \text{(B32)}$$

This is of the form of the coarsened partial wave Schrödinger equations (B24) in three dimensions—but with

$$\text{"}k^2\text{"} \sim \tfrac{1}{4}s - \mu^2, \quad \text{"}l(l+1)\text{"} \sim n^2 + \tfrac{5}{4}$$

we have in mind the same procedure for each of the other three orbits about the light cone; that is, coarsen only the "partial wave" equations.

As in the three-dimensional case, we find that there is a critical point in Eq. (B32) for increasing s. In fact (ignoring for the moment the "centrifugal barrier"), the propagation is no longer free for energies s greater than

$$s_c = 4[\mu^2 + (\hbar c/\lambda)^2]. \quad \text{(B33)}$$

Physically, this means that the aforementioned observer sees at any time t_0 a system wavefunction like Fig. 7—i.e., the particles are back-scattering off each other's topological potential—just as in the non-relativistic cases. This is not surprising—as we pointed out in Sec. A, the topology induced on this observer's subspace at any time is exactly $(x_1 \otimes x_2, \lambda)^{(3)}$. In particular, the "bounce" of the two particles is simultaneous in the center-of-mass frame. In this frame then the "forces of topological repulsion" propagate at infinite velocity (although in other frames there is in general a time lag between the bounce of one and the bounce of the other). As mentioned in Sec. A, there is no contradiction between the Lorentz group and action-at-a-distance. Of course, the "centrifugal barrier" introduces the same qualitative features discussed in the three-dimensional section. Note that the behavior of the system with t_0 is entirely independent of $t_1 - t_2$. Thus, our coarsening of the difference variable has avoided any loss of probability with t_0.[34]

5. Many Free Particles and Interactions

Thus far we have treated only the two-particle subspace with our non-usual topology. The analog of our relativistic topology on the n-particle subspace would be to coarsen the topology on some subset of difference variables. Requiring that all $\frac{1}{2}n(n-1)$ difference variables are coarse would be too restrictive: For large n, there would be in general only one available configuration for the particles (collinear) even at low energies. This precludes a coarsening that puts all n particles on an equal footing.

In electrodynamics one could certainly, for example, coarsen the topology between every electron and positron that were created together. In this way, every electron (or positron) would have a "memory"—showing up only when the relative momentum of the pair was supercritical: If one could obtain, say, the electron of a pair created on the other side of the galaxy, then, by raising the electron's momentum above critical (relative to the positron), one could affect the positron. (For example, firing the electron towards the positron, they would scatter off one another at a large angle.) In this way, one could transfer information over large distances instantaneously. Notice that these long distance effects could be avoided by cutting off the topology at a relativistic radius analogous to b in the three-dimensional case.

In the case of the recent pair production experiments, one would obtain enhancement of large angle production due to the "topological repulsion" between the electron and the positron. On the other hand, there would be no effect on ordinary electron–positron scattering (unless they were originally created together). Putting the new topology only between e^+e^- and not between $\mu^+\mu^-$ could distinguish between these two situations, leaving the μ case normal.

Because the new topology is always on a two-particle subspace, and one-particle wave equations (say, for x_1 independent of x_2) cannot be found, it does not seem possible to formulate a one-particle-equation-of-motion type of field theory, or for that matter, to write down a Lagrangian in any simple sense. It seems to us that the simplest way to construct a Lorentz-invariant theory allowing for particle creation would be in terms of the generalized unitarity equations of axiomatic field theory.[35] For example, consider the equations for the retarded functions. These equations are ordinarily written in terms of difference variables. One could then use the coarsened Klein–Gordon operators (discussed above) in the (appropriate electron-positron) difference variables. For $\Delta_+(u-v)$ one would want to use the solution to our coarsened difference Klein–Gordon equation

$$\Delta_+(x) \sim \int_{c_+'} d^4k \, \frac{\phi_k(x)}{k^2 - \mu^2}, \quad \text{(B34)}$$

[34] It is curious to note that in the analogous partial-wave equations inside the light cone, the energy term $(\frac{1}{4}s - \mu^2)$ appears with the opposite sign. In that the effective "energy" is always negative, there is no critical point in this region. If there had been a critical energy, it would have meant that, above this point, the appearance of a particle in some small spatial region would herald the rapid appearance of the second particle in that region—i.e., the topology would have simulated some sort of powerful attractive force that drags one particle after the other. This, however, is *not* what happens.

[35] See, for example, K. Nishijima, Phys. Rev. **119**, 485 (1960); **122**, 298 (1961); **124**, 255 (1961).

where c_+' is a counterclockwise circle about $k_0 = +(\mathbf{k}^2 + \mu^2)^{\frac{1}{2}}$ for $x_0 > 0$, and a clockwise circle about $k_0 = -(\mathbf{k}^2 + \mu^2)^{\frac{1}{2}}$ for $x_0 < 0$—with the proviso that when a pole meets a branch point of $\phi_k(x)$, it moves onto that side of the cut with exponentially decreasing behavior. Thus one can begin the usual iterative solution of these equations. Of course there will be the usual equal-time ambiguities at each order—which can be used as usual to specify the "interaction." To obtain the scattering amplitudes from the retarded functions, one would want to use the usual formulas, only this time replacing appropriate pairs of Fourier transform factors by solutions of our coarsened Klein–Gordon equation. Detailed discussion of such a program is beyond the scope of the present work. However, it should be emphasized that, whatever the interaction chosen, at ultra-high energies the elastic back-scattering will dominate, in that it will in general prevent the particles from reaching the interaction region.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge helpful conversations with Dr. L. Shepley, Dr. A. Goldhaber, and Dr. R. Moore, and with Professor S. Coleman. One of the authors (M. B. H.) would like to thank D. B. Wolf for a stimulating conversation about Zeeman's topology.

This research was supported in part by the Air Force Office of Scientific Research Office of Aerospace Research, United States Air Force, under Grant AF-AFOSR-232-66.

APPENDIX

The transformation function from configuration to momentum space [Eq. (B10a)], can be written

$$\langle p \mid n\lambda \rangle = \frac{\{(1 - \lambda^2 p^2/\hbar^2)^{\frac{1}{2}} + i(\lambda p/\rangle\hbar\}^n}{\pi^{\frac{1}{2}}(\hbar^2/\lambda^2 - p^2)^{\frac{1}{4}}}.$$

This function is defined to be cut for $(-\infty < p \leq -\hbar/\lambda)$ and $(\hbar/\lambda \leq p < \infty)$, and the first ("physical") sheet is specified by

$$-\pi < \arg\{(1 - \lambda^2 p^2/\hbar^2)^{\frac{1}{2}} + i(\lambda p)/\hbar\} < \pi.$$

It can be seen easily that, for $n > 0$, $\langle p \mid n\lambda \rangle$ tends to zero as $p \to \pm\infty + i\epsilon$, $\epsilon > 0$, while it is unbounded as $p \to \pm\infty - i\epsilon$. For $n < 0$, the same statements hold if the sign of ϵ is changed. Hence we make the rule that, whenever, an integral over $-\infty < p < \infty$ of an integrand involving $\langle p \mid n\lambda \rangle$ has to be performed, we take the integration contour just above/below the cuts in the p-plane for $n>/<0$. With this prescription it is easy to demonstrate Eqs. (B11).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Modification of the Ehrenfest Model in Statistical Mechanics

EDWIN D. BANTA
General Atronics Corporation, Philadelphia, Pennsylvania
(Received 3 November 1965)

The Ehrenfest model has been used to explain the "irreversibility" of thermodynamics and statistical mechanics. The modification described in this paper allows transitions to occur in both directions between the two "boxes" at each step of the model procedure. The equilibrium probability distribution is given in the form of a finite product, or in an iterated form particularly suitable for machine calculation. The analysis is illustrated by a simple model of an ionization–recombination process.

THE Ehrenfest model[1] of heat exchange between two bodies has been used successfully by Kac[2] in discussing the relationship between "irreversibility" of thermodynamic laws and statistical mechanics. In this model two boxes represent two bodies in thermal contact, and their temperatures are represented by a number of balls contained in each box. At successive intervals of time, a single ball moves from one box to the other according to a probabilistic law: the probability of transition is simply the ratio of the present number of balls in the box to the total number of balls in both boxes. Since the sum of these two ratios is unity, there is always exactly one transition at each step of the model procedure.

This note is concerned with a slightly more complicated set of transition probabilities in which transitions occur both ways between the two boxes at each step of the model procedure. Furthermore, the transition probabilities are not necessarily the ratios of the number of balls in each box. To illustrate, let $p_{12}(k)$ be the probability of a ball going from box 1 to box 2 at a given step of the procedure when there are exactly k balls in box 1 prior to the transfer. Similarly, let $p_{21}(k)$ be the probability of a ball going from box 2 to box 1 at a given step of the procedure when there are exactly k balls in box 1 (*not* box 2) prior to the

[1] P. Ehrenfest and T. Ehrenfest, *Z. Physik* **8**, 311 (1907).
[2] M. Kac, *Am. Math. Mon.*, **54**, 369 (1947).

where c_+' is a counterclockwise circle about $k_0 = +(\mathbf{k}^2 + \mu^2)^{\frac{1}{2}}$ for $x_0 > 0$, and a clockwise circle about $k_0 = -(\mathbf{k}^2 + \mu^2)^{\frac{1}{2}}$ for $x_0 < 0$—with the proviso that when a pole meets a branch point of $\phi_k(x)$, it moves onto that side of the cut with exponentially decreasing behavior. Thus one can begin the usual iterative solution of these equations. Of course there will be the usual equal-time ambiguities at each order—which can be used as usual to specify the "interaction." To obtain the scattering amplitudes from the retarded functions, one would want to use the usual formulas, only this time replacing appropriate pairs of Fourier transform factors by solutions of our coarsened Klein–Gordon equation. Detailed discussion of such a program is beyond the scope of the present work. However, it should be emphasized that, whatever the interaction chosen, at ultra-high energies the elastic back-scattering will dominate, in that it will in general prevent the particles from reaching the interaction region.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge helpful conversations with Dr. L. Shepley, Dr. A. Goldhaber, and Dr. R. Moore, and with Professor S. Coleman. One of the authors (M. B. H.) would like to thank D. B. Wolf for a stimulating conversation about Zeeman's topology.

This research was supported in part by the Air Force Office of Scientific Research Office of Aerospace Research, United States Air Force, under Grant AF-AFOSR-232-66.

APPENDIX

The transformation function from configuration to momentum space [Eq. (B10a)], can be written

$$\langle p \mid n\lambda \rangle = \frac{\{(1 - \lambda^2 p^2/\hbar^2)^{\frac{1}{2}} + i(\lambda p/\rangle\hbar\}^n}{\pi^{\frac{1}{2}}(\hbar^2/\lambda^2 - p^2)^{\frac{1}{4}}}.$$

This function is defined to be cut for $(-\infty < p \leq -\hbar/\lambda)$ and $(\hbar/\lambda \leq p < \infty)$, and the first ("physical") sheet is specified by

$$-\pi < \arg\{(1 - \lambda^2 p^2/\hbar^2)^{\frac{1}{2}} + i(\lambda p)/\hbar\} < \pi.$$

It can be seen easily that, for $n > 0$, $\langle p \mid n\lambda \rangle$ tends to zero as $p \to \pm\infty + i\epsilon$, $\epsilon > 0$, while it is unbounded as $p \to \pm\infty - i\epsilon$. For $n < 0$, the same statements hold if the sign of ϵ is changed. Hence we make the rule that, whenever, an integral over $-\infty < p < \infty$ of an integrand involving $\langle p \mid n\lambda \rangle$ has to be performed, we take the integration contour just above/below the cuts in the p-plane for $n>/<0$. With this prescription it is easy to demonstrate Eqs. (B11).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Modification of the Ehrenfest Model in Statistical Mechanics

EDWIN D. BANTA
General Atronics Corporation, Philadelphia, Pennsylvania
(Received 3 November 1965)

The Ehrenfest model has been used to explain the "irreversibility" of thermodynamics and statistical mechanics. The modification described in this paper allows transitions to occur in both directions between the two "boxes" at each step of the model procedure. The equilibrium probability distribution is given in the form of a finite product, or in an iterated form particularly suitable for machine calculation. The analysis is illustrated by a simple model of an ionization–recombination process.

THE Ehrenfest model[1] of heat exchange between two bodies has been used successfully by Kac[2] in discussing the relationship between "irreversibility" of thermodynamic laws and statistical mechanics. In this model two boxes represent two bodies in thermal contact, and their temperatures are represented by a number of balls contained in each box. At successive intervals of time, a single ball moves from one box to the other according to a probabilistic law: the probability of transition is simply the ratio of the present number of balls in the box to the total number of balls in both boxes. Since the sum of these two ratios is unity, there is always exactly one transition at each step of the model procedure.

This note is concerned with a slightly more complicated set of transition probabilities in which transitions occur both ways between the two boxes at each step of the model procedure. Furthermore, the transition probabilities are not necessarily the ratios of the number of balls in each box. To illustrate, let $p_{12}(k)$ be the probability of a ball going from box 1 to box 2 at a given step of the procedure when there are exactly k balls in box 1 prior to the transfer. Similarly, let $p_{21}(k)$ be the probability of a ball going from box 2 to box 1 at a given step of the procedure when there are exactly k balls in box 1 (*not* box 2) prior to the

[1] P. Ehrenfest and T. Ehrenfest, *Z. Physik* **8**, 311 (1907).
[2] M. Kac, *Am. Math. Mon.*, **54**, 369 (1947).

transfer. Thus, box 1 can exhibit a net increase of m balls ($m = -1, 0, 1$) after the transfer. Let $P(m; k)$ be the probability of a net increase of m balls in box 1 when there were exactly k balls in box 1 prior to the transfer. By simple enumeration, it is found that

$$\begin{aligned} P(-1; k) &= p_{12}(k) \cdot [1 - p_{21}(k)], \\ P(0; k) &= p_{12}(k) \cdot p_{21}(k) \\ &\quad + [1 - p_{12}(k)] \cdot [1 - p_{21}(k)], \\ P(1; k) &= p_{21}(k) \cdot [1 - p_{12}(k)]. \end{aligned} \tag{1}$$

Now, let $U(k; n)$ be the probability that box 1 contains exactly k balls after n steps of the model procedure. Again, by simple enumeration, it is found that $U(k; n)$ must satisfy the recurrence relationship

$$\begin{aligned} U(k; n+1) &= P(1; k) \cdot U(k-1; n) + P(0; k) \\ &\quad \cdot U(k; n) + P(-1; k) \cdot U(k+1; n). \end{aligned} \tag{2}$$

The appearance of this equation is improved through the substitution

$$\begin{aligned} U(k; n) &= W(k; n) \prod_{j=1}^{k} p_{21}(j-1) \\ &\quad \cdot [1 - p_{12}(j-1)]/p_{12}(j) \cdot [1 - p_{21}(j)], \end{aligned} \tag{3}$$

whence Eq. (2) becomes

$$\begin{aligned} [W(k; n+1) - W(k; n)] &= p_{12}(k) \\ &\quad \cdot [1 - p_{21}(k)][W(k-1; n) - W(k; n)] + p_{21}(k) \\ &\quad \cdot [1 - p_{12}(k)][W(k+1; n) - W(k; n)]. \end{aligned} \tag{4}$$

A general solution of Eq. (4) has not yet been found; however, in the limit as $n \to \infty$ (leading to a steady-state solution), it is clear that $W(k; n) \to W_0$ is a solution. This implies that the steady-state probability of k balls in box 1 is given by

$$\begin{aligned} U(k) &= \prod_{j=1}^{k} p_{21}(j-1) \cdot [1 - p_{12}(j-1)] \\ &\quad \cdot W_0/p_{12}(j) \cdot [1 - p_{21}(j)]. \end{aligned} \tag{5}$$

[The dependence upon n is omitted since $U(k)$ is the steady-state solution.] The constant W_0 must be chosen so that the sum of the $U(k)$'s is unity. Notice also that Eq. (5) can be rewritten as

$$\begin{aligned} U(k) &= p_{21}(k-1) \cdot [1 - p_{12}(k-1)] \\ &\quad \cdot U(k-1)/p_{12}(k) \cdot [1 - p_{21}(k)]. \end{aligned} \tag{6}$$

This form of the solution is particularly suited to machine evaluation; in this case the constant W_0 is determined through the initial value $U(0)$ necessary to make the sum of the $U(k)$'s equal unity.

Finally, as a simple although somewhat artificial example which leads to an analytic solution consider the following: let N atoms be contained in a vessel, and let them be subjected to ionizing radiation. The probability of ionization in a small time interval is proportional to the number of un-ionized atoms in existence at that time; at the same time, however, recombinations are occurring, and the probability of a recombination depends upon the number of ions existing at that time. In general the ionization and recombination rates are not the same. To relate this example to the mathematical model, let box 1 represent the ions and let box 2 represent the un-ionized atoms. Now, assume that recombinations occur faster than ionizations, so that if $M < N$ ions exist at a given time, it is certain that a recombination occurs in the previously defined short-time interval. Clearly, this implies that

$$p_{12}(k) = k/M, \quad p_{21}(k) = 1 - k/N. \tag{7}$$

(Remember that k refers to the number of ions in both definitions.)

Upon substitution of Eqs. (7) into Eq. (5), it is found that

$$\begin{aligned} U(k) &= W_0 \cdot \prod_{j=1}^{k} (N+1-j)(M+1-j)/j^2 \\ &= W_0 \binom{N}{k} \binom{M}{k}. \end{aligned} \tag{8}$$

In addition, it is found that the moment generating function[3] is given by

$$G(t) = W_0 \cdot \sum_{k=0}^{\infty} U(k) \cdot t^k = W_0 \cdot {}_2F_1(-M, N; 1; t), \tag{9}$$

where ${}_2F_1(\ ,\ ;\ ;\ )$ is the hypergeometric function[4]; but, by a well-known formula concerning these functions, it is found that

$$G(1) = W_0 \binom{M+N}{N} = 1. \tag{10}$$

[This last equality follows from the fact that $G(1)$ is the sum of all the individual probabilities and thus unity.] Therefore,

$$G(t) = {}_2F_1(-M, -N; 1; t) \Big/ \binom{N+M}{N}. \tag{11}$$

Again, by well-known formulas, it is easy to find the mean and variance of k as

$$\begin{aligned} \bar{k} &= MN/(M+N), \\ \sigma_k^2 &= N^2M^2/[(N+M)^2(N+M-1)]. \end{aligned} \tag{12}$$

Notice that if $M \ll N$ these formulas indicate that the number of ions is concentrated at the maximum possible number, M, with very little dispersion. On the other hand, if $N = M$, the mean number of ions and un-ionized atoms are identical; the dispersion about this mean is inversely proportional to the number of atoms, N.

[3] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946).

[4] E. T. Whittaker and G. N. Watson, *A Course in Modern Analysis* (The Macmillan Company, New York, 1944).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Thermodynamically Equivalent Hamiltonian Method in Nonequilibrium Statistical Mechanics

M. D. GIRARDEAU
Institute of Theoretical Science and Department of Physics, University of Oregon, Eugene, Oregon

(Received 4 March 1966)

The equilibrium statistical mechanics of the Bardeen–Cooper–Schrieffer model of superconductivity, as well as that of a wide class of similar models, can be evaluated exactly by the "thermodynamically equivalent Hamiltonian" method of Bogoliubov, Zubarev, Tserkovnikov, and Wentzel. It has been pointed out by Wentzel that this method can be extended to certain weakly nonequilibrium situations. It is shown here that the method allows an exact evaluation of the nonequilibrium statistical mechanics of the following situations: (a) temporal evolution of the statistical expectation value of an observable $O(\mathbf{r})$ whose initial deviation from equilibrium is *spatially localized*, but *not necessarily small*; (b) temporal evolution of the statistical expectation value of an observable O due to a perturbation V which is *spatially localized*, but *not necessarily small*.

1. INTRODUCTION

IT was shown long ago by Bogoliubov, Zubarev, and Tserkovnikov[1] that, for the model Hamiltonian of the BCS theory[2] of superconductivity, the equilibrium thermodynamic functions can be evaluated exactly in the thermodynamic (infinite-system) limit. Their method was subsequently extended by Wentzel[3] to a wide class of model Hamiltonians of the general structure

$$H = \sum_{k\lambda} (E_{k\lambda} b_{k\lambda} + E^*_{k\lambda} b^\dagger_{k\lambda}) + \Omega^{-1} \sum_{k\lambda,k'\lambda'} J_{k\lambda,k'\lambda'} b^\dagger_{k\lambda} b_{k'\lambda'}. \tag{1}$$

Here Ω is the volume of the system (which eventually $\to \infty$), k refers to linear momentum as in the BCS Hamiltonian, each $b_{k\lambda}$ is bilinear in Fermi or Bose annihilation and creation operators, and λ takes on a finite (volume-independent) set of values. It is assumed that

$$E_{k\lambda} = O(1), \quad J_{k\lambda,k'\lambda'} = O(1), \tag{2}$$

where $O(1)$ denotes a volume-independent quantity.

We briefly review the method of BZT and Wentzel here, in order to provide a foundation for its generalization to nonequilibrium statistical mechanics. The method relies on a linearization of the interaction terms $b^\dagger_{k\lambda} b_{k'\lambda'}$ through the identity

$$b^\dagger_{k\lambda} b_{k'\lambda'} = (b^\dagger_{k\lambda} - \eta^*_{k\lambda})(b_{k'\lambda'} - \eta_{k'\lambda'}) + (\eta^*_{k\lambda} b_{k'\lambda'} + \eta_{k'\lambda'} b^\dagger_{k\lambda}) - \eta^*_{k\lambda}\eta_{k'\lambda'}, \tag{3}$$

where the $\eta_{k\lambda}$ are c-number parameters to be determined. In this way, H is separated into a Hamiltonian H_0 quadratic in annihilation and creation operators, plus a residual Hamiltonian H':

$$H = H_0 + H',$$
$$H_0 = U + \sum_{k\lambda} (G_{k\lambda} b_{k\lambda} + G^*_{k\lambda} b^\dagger_{k\lambda}),$$
$$H' = \Omega^{-1} \sum_{k\lambda,k'\lambda'} J_{k\lambda,k'\lambda'} (b^\dagger_{k\lambda} - \eta^*_{k\lambda})(b_{k'\lambda'} - \eta_{k'\lambda'}) \tag{4}$$

with

$$U = -\Omega^{-1} \sum_{k\lambda\, k'\lambda'} J_{k\lambda,k'\lambda'} \eta^*_{k\lambda} \eta_{k'\lambda'},$$
$$G_{k\lambda} = E_{k\lambda} + \Omega^{-1} \sum_{k'\lambda'} J_{k'\lambda',k\lambda} \eta^*_{k'\lambda'}. \tag{5}$$

The exact thermodynamic potential[4] of the system described by H is

$$F = -\beta^{-1} \ln \mathrm{Tr}\, e^{-\beta H}. \tag{6}$$

One now observes that, since H_0 is only bilinear in single-particle annihilation and creation operators[5] a_k and $a^\dagger_k$ (each $b_{k\lambda}$ or $b^\dagger_{k\lambda}$ is bilinear in the a and $a^\dagger$ operators), H_0 can be brought into diagonal form

$$H_0 = E_0 + \sum_k \epsilon_k \alpha^\dagger_k \alpha_k \tag{7}$$

by a *linear*[6] canonical transformation to new Fermi or Bose operators α_k, $\alpha^\dagger_k$, which have the physical significance of quasiparticle annihilation and creation operators. It is assumed that the bilinear operators $b_{k\lambda}$, $b^\dagger_{k\lambda}$ are constructed from the a_k, $a^\dagger_k$ operators in

[1] N. N. Bogoliubov, D. N. Zubarev, and Yu. A. Tserkovnikov, Dokl. Akad. Nauk SSSR **117**, 788 (1957) [English transl.: Soviet Phys.—Doklady **2**, 535 (1957)]; this paper is referred to as BZT herein.

[2] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).

[3] G. Wentzel, Phys. Rev. **120**, 1572 (1960).

[4] The term $-\mu N$ is included in H, and the trace in (6) runs over states belonging to all eigenvalues of the total particle-number N. Thus F here is the grand potential.

[5] In the Fermi case the a_k and $a^\dagger_k$ also carry spin indices; these are suppressed since they do not enter the argument in any essential way.

[6] Hence tractable.

such a way that when transformed into the quasiparticle representation, they have the general structure[7]

$$b_{k\lambda} = c_{k\lambda} + \sum_{k' \in \mathcal{S}_k} \sum_{k'' \in \mathcal{S}_k} (d_{k\lambda,k'k''} \alpha^\dagger_{k'} \alpha_{k''} + e_{k\lambda,k'k''} \alpha_{k'} \alpha_{k''} + f_{k\lambda,k'k''} \alpha^\dagger_{k'} \alpha^\dagger_{k''}), \tag{8}$$

where the coefficients c, d, e, f are all volume-independent, and where, for each fixed k, $\mathcal{S}_k$ is a set of a finite, volume-independent number of wave vectors; e.g., for the BCS Hamiltonian $\mathcal{S}_k = \{k, -k\}$. Then, applying statistical-mechanical perturbation theory to the evaluation of the trace (6) in powers of the perturbation H', BZT and Wentzel showed that the contribution of H' to the thermodynamic potential F is only $O(1)$ (finite in the thermodynamic limit $\Omega \to \infty$ with μ fixed) to all orders in H' provided that $\eta_{k\lambda}$ in (3)–(5) is chosen as the thermal average[8] of $b_{k\lambda}$ in the ensemble of H_0:

$$\eta_{k\lambda} = \langle b_{k\lambda} \rangle_0 = c_{k\lambda} + \sum_{k' \in \mathcal{S}} d_{k\lambda,k'k'} \langle \alpha^\dagger_{k'} \alpha_{k'} \rangle_0 . \tag{9}$$

Here

$$\langle O \rangle_0 \equiv \mathrm{Tr}\,(O e^{-\beta H_0})/\mathrm{Tr}\, e^{-\beta H_0}. \tag{10}$$

Then

$$F = -\beta^{-1} \ln \mathrm{Tr}\, e^{-\beta H_0} + O(1). \tag{11}$$

The term $O(1)$ is completely negligible in the thermodynamic limit since $-\beta^{-1} \ln \mathrm{Tr}\, e^{-\beta H_0}$, the thermodynamic potential of H_0, is proportional to the volume Ω; for this reason, Wentzel calls H_0 the "thermodynamically equivalent Hamiltonian". The reason for the negligibility of the contributions of H' is that with the choice (9), H' becomes bilinear in *fluctuations* $b_{k\lambda} - \langle b_{k\lambda} \rangle_0$; these fluctuations are negligible under the conditions (2), (8).[9] Since H_0 represents a system of noninteracting (but temperature-dependent) quasiparticles, all quasiparticle interactions are contained in H'. Thus, another way of stating the results is that quasiparticle interactions are negligible in the thermodynamic limit to all orders in H' *provided that the quasiparticles are defined properly* [so that (9) is satisfied].

The exact thermodynamic equivalence of H_0 and H does not extend to arbitrary nonequilibrium situations, nor is there any reason to expect it to. Nevertheless, since H' is completely negligible in equilibrium, an obvious continuity argument indicates that the effects of H' should be small near equilibrium. This was first pointed out by Wentzel,[10] who showed that quasiparticle interactions could be ignored in a first-order calculation of the momentum transferred to the system by a weak force center dragged through the system with constant velocity.

In Secs. 2–4 we consider the slightly different problem of the time evolution of the statistical expectation value, $\langle O(\mathbf{r}, t) \rangle$, of a position-dependent observable $O(\mathbf{r})$ due to an initial *spatially localized* deviation from equilibrium, $\langle O(\mathbf{r}, 0) \rangle - \langle O(\mathbf{r}, 0) \rangle_{\mathrm{eq}}$, in the absence of any external driving force. We prove that in evaluating $\langle O(\mathbf{r}, t) \rangle$, all effects of H' are negligible [$O(\Omega^{-1})$ as $\Omega \to \infty$]. This result is proved to all orders in H' *and to all orders in the departure from equilibrium*. The conclusion thus goes beyond the continuity argument which led us to expect the result to hold for *small* departures from equilibrium. The physical reasons for the more general result are not clearly understood at present, but we emphasize that the proof only goes through for *localized* deviations from equilibrium.

In Sec. 5 we study the temporal evolution of the statistical expectation value, $\langle O(t) \rangle$, of an observable O due to a *spatially localized* perturbation V, given that the system was in equilibrium before V was turned on. It is again found that H' is negligible in the thermodynamic limit, to all orders in H' *and in the perturbation* V. It is emphasized that the spatial localization of the deviation from equilibrium or of the external perturbation is quite essential in the proofs. Also, the operator O in Sec. 5 must not have any delta-function singularities in its matrix elements in momentum space. This restriction rules out the total linear momentum but not the momentum density; this somewhat paradoxical distinction is discussed with reference to the orders of noncommuting limits.

In Sec. 6 we comment briefly on the dangers of a perturbation-theoretic proof "to all orders," and in Sec. 7 applications and unsolved problems are suggested.

Before proceeding with the analysis, we wish to say a few words concerning motivation. It might be objected that, since our results only apply to Hamiltonians of generalized BCS type, they are not applicable to real physical systems. It is certainly true that such Hamiltonians are highly simplified relative to real physical systems. Nevertheless, since exactly soluble models are even rarer in nonequilibrium statistical mechanics than in the equilibrium

[7] This form of the requirements on the $b_{k\lambda}$ is somewhat more general than that employed by BZT and Wentzel and is more convenient for our subsequent analysis.

[8] Since H_0 is diagonal in the α, $\alpha^\dagger$ representation, one has $\langle \alpha_{k'} \alpha_{k''} \rangle_0 = \langle \alpha^\dagger_{k'} \alpha^\dagger_{k''} \rangle_0 = 0$, $\langle \alpha^\dagger_{k'} \alpha_{k''} \rangle_0 = \delta_{k'k''} (\exp \beta \epsilon_{k'} \pm 1)^{-1}$.

[9] We do not give further details of the BZT–Wentzel proof here, since they are clear from the generalization carried out in Sec. 3.

[10] G. Wentzel, in *Werner Heisenberg und die Physik unserer Zeit* (Friederick Vieweg und Sohn, Braunschweig, Germany, 1961), p. 189.

case, we feel that the solubility of localized nonequilibrium problems for such Hamiltonians may prove illuminating. In this connection, it should be noted that the operator $O(\mathbf{r})$ whose time evolution is investigated in Secs. 2–4, or the external perturbation V and observable O in Sec. 5, need not have the simplified pairing structure typical of the interaction terms of the generalized BCS Hamiltonian. Furthermore, the validity of our method does not rest on any assumption of small distortion of the equilibrium ensemble. We find, instead, that problems involving large *local* distortions of the ensemble can be dealt with by the method of the thermodynamically equivalent Hamiltonian.

2. INITIAL-VALUE PROBLEM IN NONEQUILIBRIUM STATISTICAL MECHANICS

We employ Jaynes' formulation[11] of nonequilibrium statistical mechanics. Suppose that at time $t = 0$ not only the mean energy, but also the statistical average values of certain other observables F_j, are known. Then the statistical expectation value, at time t, of any observable 0 is given by[11]

$$\langle O(t)\rangle = \frac{\mathrm{Tr}\left[O(t)\exp\left(-\beta H - \sum_j \lambda_j F_j\right)\right]}{\mathrm{Tr}\exp\left(-\beta H - \sum_j \lambda_j F_j\right)}, \quad (12)$$

where $O(t)$ is the Heisenberg operator

$$O(t) = e^{itH}Oe^{-itH} \quad (13)$$

and the Lagrange multipliers λ_j and β are determined by the known initial statistical expectation values $\langle F_j(0)\rangle$ and $\langle H(0)\rangle$. We are interested in the case that the only initial data consist of the values of the total Hamiltonian H and of a position-dependent operator $O(\mathbf{r}) = O(\mathbf{r}, 0)$ at all positions $\mathbf{r}$. Then the expectation value of $O(\mathbf{r})$ at any other time t is given by[12]

$$\langle O(\mathbf{r}, t)\rangle = \frac{\mathrm{Tr}\left\{O(\mathbf{r}, t)\exp\left[-\beta H - \int d^3r'\lambda_0(\mathbf{r}')O(\mathbf{r}')\right]\right\}}{\mathrm{Tr}\exp\left[-\beta H - \int d^3r'\lambda_0(\mathbf{r}')O(\mathbf{r}')\right]}. \quad (14)$$

In order to evaluate (14), it is convenient to define a nonequilibrium partition function Ξ and thermodynamic potential W by

$$\Xi \equiv e^{-\beta W} \equiv \mathrm{Tr}\exp\left[-\beta H - \int_{-\infty}^{\infty} dt \int d^3r\lambda(\mathbf{r}, t)O(\mathbf{r}, t)\right]. \quad (15)$$

Then $\langle O(\mathbf{r}, t)\rangle$ can be expressed as a functional derivative:

$$\langle O(\mathbf{r}, t)\rangle = [\delta(\beta W)/\delta\lambda(\mathbf{r}, t)]_{\lambda(\mathbf{r},t)=\lambda_0(\mathbf{r})\delta(t)}, \quad (16)$$

where the subscript denotes that $\lambda(\mathbf{r}, t)$ is to be set equal to $\lambda_0(\mathbf{r})\delta(t)$ after the differentiation. The multipliers[13] β and $\lambda_0(\mathbf{r})$ are determined from the known initial data $\langle H(0)\rangle$ and $\langle O(\mathbf{r}, 0)\rangle$:

$$\begin{aligned}[\partial(\beta W)/\partial\beta]_{\lambda(\mathbf{r},t)=\lambda_0(\mathbf{r})\delta(t)} &= \langle H(0)\rangle,\\ [\delta(\beta W)/\delta\lambda(\mathbf{r}, 0)]_{\lambda(\mathbf{r},t)=\lambda_0(\mathbf{r})\delta(t)} &= \langle O(\mathbf{r},0)\rangle.\end{aligned} \quad (17)$$

We next apply a standard form of statistical mechanical perturbation theory to the evaluation of the trace in (15), obtaining

$$\begin{aligned}\Xi = (\mathrm{Tr}\, e^{-\beta H})\Bigg[1 + \sum_{j=1}^{\infty}(-1)^j \int_{-\infty}^{\infty}\cdots\int dt_1\cdots dt_j \\ \times \int\cdots\int d^3r_1\cdots d^3r_j\lambda(\mathbf{r}_1, t_1)\cdots\lambda(\mathbf{r}_j, t_j) \\ \times \int_0^1 ds_1\int_0^{s_1} ds_2\cdots\int_0^{s_{j-1}} ds_j \\ \times \langle O(\mathbf{r}_1, t_1 - i\beta s_1)\cdots O(\mathbf{r}_j, t_j - i\beta s_j)\rangle_{\mathrm{eq}}\Bigg],\end{aligned} \quad (18)$$

where $\langle\ \rangle_{\mathrm{eq}}$ denotes a thermal-equilibrium average:

$$\langle O\rangle_{\mathrm{eq}} \equiv \mathrm{Tr}\,(Oe^{-\beta H})/\mathrm{Tr}\, e^{-\beta H}. \quad (19)$$

With the use of (13) and cyclic invariance of the trace, one finds

$$\Xi = \Xi_{\mathrm{eq}}(1 + \Delta_1 + \Delta_2 + \cdots) \quad (20)$$

with

$$\Xi_{\mathrm{eq}} = \mathrm{Tr}\, e^{-\beta H} \quad (21)$$

and

$$\begin{aligned}\Delta_1 &= -\int_{-\infty}^{\infty} dt\int d^3r\lambda(\mathbf{r}, t)\langle O(\mathbf{r}, 0)\rangle_{\mathrm{eq}},\\ \Delta_2 &= \frac{1}{2}\int_0^1 ds \iint_{-\infty}^{\infty} dt_1\, dt_2 \\ &\quad \times \iint d^3r_1\, d^3r_2\lambda(\mathbf{r}_1, t_1)\lambda(\mathbf{r}_2, t_2) \\ &\quad \times \langle O(\mathbf{r}_1, 0)O(\mathbf{r}_2, t_2 - t_1 + i\beta s)\rangle_{\mathrm{eq}}.\end{aligned} \quad (22)$$

[11] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); Phys. Rev. **108**, 171 (1957); in *Statistical Physics: Vol. 3. Brandeis Summer Institute, 1962* (W. A. Benjamin, Inc., New York, 1963). Also unpublished work of E. T. Jaynes presented in an informal seminar at the University of Oregon.

[12] The extension to the case where more than one such observable $O(\mathbf{r})$ is known initially, or $O(\mathbf{r})$ is a vector or tensor instead of a scalar, is obvious (just extend the meaning of $\mathbf{r}$) and is not treated explicitly.

[13] β reduces to $1/kT$ in the equilibrium limit $\lambda_0(\mathbf{r}) \to 0$, but does not, strictly speaking, have that significance when $\lambda_0 \not\equiv 0$.

The definition (15) of W then gives

$$W = W_0 + W_1 + W_2 + \cdots, \quad W_0 = -\beta^{-1} \ln \Xi_{eq},$$
$$W_1 = -\beta^{-1}\Delta_1, \qquad W_2 = -\beta^{-1}(\Delta_2 - \tfrac{1}{2}\Delta_1^2),$$
$$\cdot \quad \cdot \quad \cdot \tag{23}$$

Finally, evaluation of the functional derivative in (16) gives

$$\langle O(\mathbf{r}, t)\rangle = \langle O(\mathbf{r}, 0)\rangle_{eq} - \frac{1}{2}\int_0^1 ds \int d^3r_1 \lambda_0(\mathbf{r}_1)[\langle O(\mathbf{r}, 0)O(\mathbf{r}, t + i\beta s)\rangle_{eq} + \langle O(\mathbf{r}, t - i\beta s)O(\mathbf{r}_1, 0)\rangle_{eq}] + \langle O(\mathbf{r}, 0)\rangle_{eq}\int d^3r_1 \lambda_0(\mathbf{r}_1)\langle O(\mathbf{r}_1, 0)\rangle_{eq} + O(\lambda_0^2), \tag{24}$$

where $O(\lambda_0^2)$ denotes terms bilinear in λ_0. Since $\langle 0\rangle \to \langle 0\rangle_{eq}$ as $\lambda_0 \to 0$, the neglected terms in (24) are of second and higher orders in the departure from equilibrium, i.e., in $\langle 0\rangle - \langle 0\rangle_{eq}$.

Defining a kernel

$$K(\mathbf{r}, t; \mathbf{r}_1, 0) = \langle O(\mathbf{r}, 0)\rangle_{eq}\langle O(\mathbf{r}_1, 0)\rangle_{eq} - \frac{1}{2}\int_0^1 ds[\langle O(\mathbf{r}_1, 0)O(\mathbf{r}, t + i\beta s)\rangle_{eq} + \langle O(\mathbf{r}, t - i\beta s)O(\mathbf{r}_1, 0)\rangle_{eq}], \tag{25}$$

we see that (24) can be written as

$$\langle O(\mathbf{r}, t)\rangle = \langle O(\mathbf{r}, 0)\rangle_{eq} + \int K(\mathbf{r}, t; \mathbf{r}_1, 0)\lambda_0(\mathbf{r}_1)\, d^3r_1 + O(\lambda_0^2). \tag{26}$$

The function $\lambda_0(\mathbf{r}_1)$ is to be determined by inversion of the integral transform

$$\int K(\mathbf{r}, 0; \mathbf{r}_1, 0)\lambda_0(\mathbf{r}_1)\, d^3r_1 = \langle O(\mathbf{r}, 0)\rangle - \langle O(\mathbf{r}, 0)\rangle_{eq}, \tag{27}$$

given the initial (in general nonequilibrium) statistical expectation value $\langle O(\mathbf{r}, 0)\rangle$ of the observable $O(\mathbf{r})$.

The second term in the expression for K [the only term if the mean value of $O(\mathbf{r})$ vanishes in equilibrium] is a space–time correlation function of a type already familiar in the theory of transport.[14] Actually, the transport kernel can be expressed purely in terms of the true correlation function even in the case that $O(\mathbf{r})$ has nonzero mean value in thermal equilibrium. Defining a true correlation function $\langle\ \rangle_{eq}^{corr}$ by subtracting off the uncorrelated part:

$$\langle O(\mathbf{r}, t)O(\mathbf{r}', t')\rangle_{eq}^{corr} = \langle O(\mathbf{r}, t)O(\mathbf{r}', t')\rangle_{eq} - \langle O(\mathbf{r}, t)\rangle_{eq}\langle O(\mathbf{r}', t')\rangle_{eq} \tag{28}$$

and noting that $\langle O(\mathbf{r}, t)\rangle_{eq} = \langle O(\mathbf{r}, 0)\rangle_{eq}$ by cyclic invariance of the trace, one can rewrite (25) in the form

$$K(\mathbf{r}, t; \mathbf{r}_1, 0) = -\frac{1}{2}\int_0^1 ds[\langle O(\mathbf{r}_1, 0)O(\mathbf{r}, t + i\beta s)\rangle_{eq}^{corr} + \langle O(\mathbf{r}, t - i\beta s)O(\mathbf{r}_1, 0)\rangle_{eq}^{corr}]. \tag{29}$$

In conclusion, we note that the expansions (20), (23), (24) are only useful if the initial deviation from equilibrium is spatially localized, i.e., of finite (Ω-independent) range. Only in that case can one expect[15] $\lambda_0(\mathbf{r})$ to be of finite range. On the other hand, if $\langle O(\mathbf{r}, 0)\rangle - \langle O(\mathbf{r}, 0)\rangle_{eq}$ were appreciable throughout the volume of the system, one expects that the same would be true of $\lambda_0(\mathbf{r})$. Then the integral over $\mathbf{r}_1$ would be proportional to the volume,[16] the term $O(\lambda_0^2)$ proportional to Ω^2, etc. In such a case the expansion would be useless; to obtain meaningful results one would have to replace the expansion (20) of the partition function by a linked-cluster expansion for the thermodynamic potential W. The analysis then becomes much more involved; we therefore do not consider nonlocalized deviations from equilibrium here.

3. THERMODYNAMIC EQUIVALENCE OF H_0 FOR LINEAR TRANSPORT OF ONE-PARTICLE OBSERVABLES

We now assume that $O(\mathbf{r})$ is a one-particle operator of the general structure

$$O(\mathbf{r}) = \Omega^{-1}\sum_{kk'} O_{kk'}(\mathbf{r})a_k^\dagger a_{k'} \tag{30}$$

with volume-independent matrix elements $O_{kk'}(\mathbf{r})$. We furthermore assume that the canonical transformation from the a, $a^\dagger$ representation to the α, $\alpha^\dagger$ representation [in which H_0 of Eq. (7) is diagonal] is of the form

$$a_k = \sum_{k' \in \mathcal{S}_k}(u_{kk'}\alpha_{k'} + v_{kk'}\alpha_{k'}^\dagger), \tag{31}$$

where $\mathcal{S}_k$ is the same as in (8), and the $u_{kk'}$ and $v_{kk'}$ are volume-independent. We show that, subject to the

[14] See, e.g., R. Zwanzig, in *Annual Reviews of Physical Chemistry* (Annual Reviews, Inc., Palo Alto, 1965), Vol. 16, pp. 67 ff.

[15] It is easy to show by Fourier transformation of (27) that $\lambda_0(\mathbf{r})$ will be of finite range if $\langle O(\mathbf{r}, 0)\rangle - \langle O(\mathbf{r}, 0)\rangle_{eq}$ is, provided that K is translationally invariant, i.e., $K(\mathbf{r}, 0; \mathbf{r}', 0) = K_0(\mathbf{r} - \mathbf{r}')$. This restriction on K can probably be weakened without affecting the conclusion.

[16] Since the systems for which the thermodynamically-equivalent Hamiltonian is useful (e.g., superconductors, superfluids, and ferromagnets) have long-range order in equilibrium, the kernel K will in general not vanish as $|\mathbf{r} - \mathbf{r}_1| \to \infty$.

assumptions (2), (8), and (31), the contributions of H' to the transport kernel K [Eq. (25)], evaluated to all orders in H', are only $O(\Omega^{-1})$, hence negligible in the thermodynamic limit $\Omega \to \infty$. In other words, H_0 is also a thermodynamically equivalent Hamiltonian for linear transport theory. In order to show this, one must go beyond the original BZT[1]–Wentzel[3] argument, since that argument only applies to the equilibrium thermodynamic potential. In evaluating the transport kernel (25), H' enters in two ways: in the factor $e^{-\beta H} = e^{-\beta(H_0+H')}$ occurring in the definition of the equilibrium average $\langle\ \rangle_{\rm eq}$, and in the evaluation of the Heisenberg operators (13). On the other hand, in the BZT–Wentzel proof for the equilibrium case, H' only enters in the factor $e^{-\beta H}$.

We consider first the term $\langle O(\mathbf{r},0)\rangle_{\rm eq}\langle O(\mathbf{r}_1,0)\rangle_{\rm eq}$ in K; in this term the Heisenberg operators are evaluated at $t=0$, so that H' enters only through $e^{-\beta H}$. Our proof for this term will thus be only a minor extension of the BZT–Wentzel proof. Using the perturbation expansion

$$e^{-\beta H} = e^{-\beta H_0}\Big[1+\sum_{j=1}^{\infty}(-i)^j\int_0^{-i\beta}dt_1\int_0^{t_1}dt_2\cdots \times\int_0^{t_{j-1}}dt_j H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)\Big], \quad (32)$$

where

$$O^{(0)}(t) \equiv e^{itH_0}Oe^{-itH_0}, \quad (33)$$

the definition (19) of $\langle\ \rangle_{\rm eq}$, and cyclic invariance of the trace, one finds[17]

$$\langle O^{(0)}(\mathbf{r},0)\rangle_{\rm eq} = \Big[\langle O^0(\mathbf{r},0)\rangle_0+\sum_{j=1}^{\infty}(-i)^j\int_0^{-i\beta}dt_1\cdots \times\int_0^{t_{j-1}}dt_j\langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)O^{(0)}(\mathbf{r},0)\rangle_0\Big] \times\Big[1+\sum_{j=1}^{\infty}(-i)^j\int_0^{-i\beta}dt_1\cdots\int_0^{t_{j-1}} \times dt_j\langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)\rangle_0\Big]^{-1}, \quad (34)$$

where $\langle O\rangle_0$ is defined by (10). Since H_0 is bilinear in single-particle annihilation and creation operators, one can apply Matsubara's theorem[18] to decompose the averages $\langle\ \rangle_0$ into sums of products of contractions, each such contraction being the average $\langle\ \rangle_0$ of an operator bilinear in single-particle annihilation and creation operators. In this way one sees that

$$\langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)O^{(0)}(\mathbf{r},0)\rangle_0 = \langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)\rangle_0\langle O^{(0)}(\mathbf{r},0)\rangle_0 + \langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)O^{(0)}(\mathbf{r},0)\rangle_0', \quad (35)$$

where the prime on $\langle\ \rangle_0'$ implies the omission of all terms in which $O^{(0)}(\mathbf{r},0)$ is self-contracted, such terms being already included in $\langle\ \rangle_0\langle\ \rangle_0$.

The expression $\langle\ \rangle_0'$ can be further reduced by noting that, according to (8) and (9),

$$b_{k\lambda}-\eta_{k\lambda} = \sum_{k'\in\mathfrak{S}_k}\sum_{k''\in\mathfrak{S}_k}[d_{k\lambda,k'k''}(\alpha_{k'}^\dagger\alpha_{k''}-\delta_{k'k''}\langle\alpha_{k'}^\dagger\alpha_{k'}\rangle_0) + e_{k\lambda,k'k''}\alpha_{k'}\alpha_{k''}+f_{k\lambda,k'k''}\alpha_{k'}^\dagger\alpha_{k''}^\dagger]. \quad (36)$$

But since

$$\langle\alpha_{k'}^\dagger\alpha_{k''}\rangle_0 = \delta_{k'k''}\langle\alpha_{k'}^\dagger\alpha_{k'}\rangle_0, \qquad \langle\alpha_{k'}\alpha_{k''}\rangle_0 = \langle\alpha_{k'}^\dagger\alpha_{k''}^\dagger\rangle_0 = 0 \quad (37)$$

by (7), it follows on application of Matsubara's theorem that, for any two operators A and B,

$$\langle A(b_{k\lambda}-\eta_{k\lambda})B\rangle_0 = \langle Ab_{k\lambda}'B\rangle_{0,\rm cross}. \quad (38)$$

Here

$$b_{k\lambda}' \equiv \sum_{k'\in\mathfrak{S}_k}\sum_{k''\in\mathfrak{S}_k}(d_{k\lambda,k'k''}\alpha_{k'}^\dagger\alpha_{k''}+e_{k\lambda,k'k''}\alpha_{k'}\alpha_{k''} + f_{k\lambda,k'k''}\alpha_{k'}^\dagger\alpha_{k''}^\dagger), \quad (39)$$

and the subscript "cross" on $\langle\ \rangle_0$ implies omission of all terms in which $b_{k\lambda}'$ is self-contracted in the expansion of $\langle Ab_{k\lambda}'B\rangle_0$ according to Matsubara's theorem. Introducing the definitions (4) and (30) of H' and $O(\mathbf{r})$, and applying the lemma (38) and its Hermitian conjugate repeatedly, one finds that

$$\langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)O^{(0)}(\mathbf{r},0)\rangle_0' = \Omega^{-(j+1)} \times\sum_{kk'}\sum_{k_1\lambda_1\cdots k_j\lambda_j}\sum_{k_1'\lambda_1'\cdots k_j'\lambda_j'}O_{kk'}(\mathbf{r})J_{k_1\lambda_1,k_1'\lambda_1'}\cdots \times J_{k_j\lambda_j,k_j'\lambda_j'}\langle b_{k_1\lambda_1}^{\dagger\prime(0)}(t_1)b_{k_1'\lambda_1'}^{\prime(0)}(t_1)\cdots b_{k_j\lambda_j}^{\dagger\prime(0)} \times(t_j)b_{k_j'\lambda_j'}^{\prime(0)}(t_j)a_k^\dagger a_{k'}\rangle_0'', \quad (40)$$

where the double prime implies omission of all terms in which either the operator $a_k^\dagger a_{k'}$ or any of the $b'^{(0)}$ or $b^{\dagger\prime(0)}$ operators is self-contracted.

Finally, noting (39) and (31) and recalling that $\mathfrak{S}_k$, for each k, is a *finite* (volume-independent) set of wave vectors, we see that each expression $\langle\ \rangle_0''$ in (40) vanishes unless not more than j of the $(2j+2)$ vectors $k_1\cdots k_j, k_1'\cdots k_j', kk'$ are independent; the complete proof is given in the Appendix. Thus, on converting k-sums to integrals by the prescription

$$\sum_k \to (2\pi)^{-3}\Omega\int d^3k,$$

[17] By (13) and (33), $O(\mathbf{r},0) = O^{(0)}(\mathbf{r},0) = O(\mathbf{r})$. Nevertheless, the notation $O^{(0)}$ has been employed in order to make the generalization to the case $t \neq 0$ more transparent.

[18] T. Matsubara, Progr. Theoret. Phys. (Kyoto) 14, 351 (1955). A complete proof of the theorem, in the form which we employ, is given by C. Bloch and C. DeDominicis, Nucl. Phys. 7, 459 (1958).

one obtains only j factors of the volume Ω from the summations, so that[19]

$$\langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)O^{(0)}(\mathbf{r},0)\rangle_0' = O(\Omega^{-1}). \quad (41)$$

Finally, substituting (41) into (35) and (35) into (34), one finds that[20]

$$\langle O(\mathbf{r},0)\rangle_{\rm eq} = \langle O^{(0)}(\mathbf{r},0)\rangle_0 + O(\Omega^{-1}), \quad (42)$$

from which it follows that no error in evaluating the term $\langle O(\mathbf{r},0)\rangle_{\rm eq}\langle O(\mathbf{r}_1,0)\rangle_{\rm eq}$ in K [Eq. (25)] is made in the thermodynamic limit if one replaces H by H_0.

The proof for the more complicated correlation-function terms in K can now be constructed by analogy. By (13), (19), and the cyclic invariance of the trace, one has

$$\langle O(\mathbf{r}, t-i\beta s)O(\mathbf{r}_1,0)\rangle_{\rm eq} = \langle O(\mathbf{r},0)O(\mathbf{r}_1,-t+i\beta s)\rangle_{\rm eq}. \quad (43)$$

Thus it is sufficient to consider correlation functions of the form $\langle O(\mathbf{r},0)O(\mathbf{r}',t)\rangle_{\rm eq}$ in evaluating (25). Making use of the formulas

$$e^{it(H_0+H')} = \exp_-\left[i\int_0^t dt' H'^{(0)}(t')\right]e^{itH_0},$$
$$e^{-it(H_0+H')} = e^{-itH_0}\exp_+\left[-i\int_0^t dt' H'^{(0)}(t')\right], \quad (44)$$

where $\exp_+$ and $\exp_-$ are the positively (ordinary) and negatively time-ordered exponentials, one finds[21] with (19), (13), (33), and (32), and cyclic invariance of the trace

$$\langle O(\mathbf{r},0)O(\mathbf{r}',t)\rangle_{\rm eq} = \left[\sum_{j=0}^{\infty}(-i)^j\int_0^{-i\beta}dt_1\cdots\int_0^{t_{j-1}}dt_j\langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j) \times O^{(0)}(\mathbf{r},0)U_-(t)O^{(0)}(\mathbf{r}',t)U_+(t)\rangle_0\right] \times\left[\sum_{j=0}^{\infty}(-i)^j\int_0^{-i\beta}dt_1\cdots\int_0^{t_{j-1}}dt_j\langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)\rangle_0\right]^{-1}, \quad (45)$$

where

$$U_-(t) = \exp_-\left[i\int_0^t dt' H'^{(0)}(t')\right],$$
$$U_+(t) = \exp_+\left[-i\int_0^t dt' H'^{(0)}(t')\right]. \quad (46)$$

As in (35), upon applying Matsubara's theorem and noting that

$$U_-(t)U_+(t) = 1, \quad (47)$$

one finds that

$$\langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)O^{(0)}(\mathbf{r},0)U_-(t)O^{(0)}(\mathbf{r}',t)U_+(t)\rangle_0 = \langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)\rangle_0\langle O^{(0)}(\mathbf{r},0)O^{(0)}(\mathbf{r}',t)\rangle_0 + \langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)O^{(0)}(\mathbf{r},0)U_-(t) \times O^{(0)}(\mathbf{r}',t)U_+(t)\rangle_0', \quad (48)$$

where the prime on $\langle\ \rangle_0'$ now implies omission of all terms in which *neither* $O^{(0)}(\mathbf{r},0)$ *nor* $O^{(0)}(\mathbf{r}',t)$ is connected to operators in U_-, U_+, or any of the $H'^{(0)}$ factors by one or more contractions. When U_- and U_+ are expanded, the expression $\langle\ \rangle_0'$ in (48) becomes a sum of expressions of the form

$$\langle H'^{(0)}(t_1)\cdots H'^{(0)}(t_j)O^{(0)}(\mathbf{r},0)H'^{(0)}(t_l')\cdots \times H'^{(0)}(t_1')O^{(0)}(\mathbf{r}',t)H'^{(0)}(t_1'')\cdots H'^{(0)}(t_m'')\rangle_l'. \quad (49)$$

By a derivation paralleling that of (40) one then finds

$$(49) = \Omega^{-(j+l+m+2)}\sum\cdots\sum[O(1)]\langle b^{\dagger\prime(0)}_{k_1\lambda_1}b'^{(0)}_{k_1'\lambda_1'}\cdots \times b^{\dagger\prime(0)}_{k_j\lambda_j}b'^{(0)}_{k_j'\lambda_j'}a^\dagger_k a_{k'}b^{\dagger\prime(0)}_{p_l\mu_l}b'^{(0)}_{p_l'\mu_l'}\cdots \times b^{\dagger\prime(0)}_{p_1\mu_1}b'^{(0)}_{p_1'\mu_1'}a^\dagger_p a_{p'}b^{\dagger\prime(0)}_{q_1\nu_1}b'^{(0)}_{q_1'\nu_1'}\cdots \times b^{\dagger\prime(0)}_{q_m\nu_m}b'^{(0)}_{q_m'\nu_m'}\rangle_0'', \quad (50)$$

where $\sum\cdots\sum$ denotes summation over all $(2j+2l+2m+4)$ k-vectors appearing in $\langle\ \rangle_0''$, $[O(1)]$ is a product of factors $O_{kk'}(\mathbf{r})$, $O_{pp'}(\mathbf{r}')$, and J's, all of which are volume-independent, and the inessential time arguments (which lead only to phase factors) have been omitted. The double prime on $\langle\ \rangle_0''$ now implies omission of all terms in which any $b'^{(0)}$ or $b^{\dagger\prime(0)}$ is self-contracted, as well as the omissions implied by the single prime in (48).

A further reduction can be made by noting that contractions in which *either* $a_k^\dagger a_{k'}$ *or* $a_p^\dagger a_{p'}$ is self-contracted are included in (50); on the other hand, contractions in which *both* $a_k^\dagger a_{k'}$ and $a_p^\dagger a_{p'}$ are self-contracted are excluded by the definition of $\langle\ \rangle_0'$ in (48), being already included in $\langle O^{(0)}(\mathbf{r},0)O^{(0)}(\mathbf{r}',t)\rangle_0$. Suppose, e.g., that $a_p^\dagger a_{p'}$ is self-contracted. Then the remaining factor (which excludes self-contraction of $a_k^\dagger a_{k'}$) is of the same general structure as (40) with j replaced by $j+l+m$. Hence by the previous argument, it vanishes unless not more than $(j+l+m)$ of the $(2j+2l+2m+2)$ vectors $k_1\cdots k_j$, $k_1'\cdots k_j'$, $p_1\cdots p_l$, $p_1'\cdots p_l'$, $q_1\cdots q_m$, $q_1'\cdots q_m'$, kk' are independent. Furthermore, $\langle a_k^\dagger a_{k'}\rangle_0$ vanishes unless $\mathcal{S}_k$ and $\mathcal{S}_{k'}$ overlap, so that k and k' are not independent. One thus has $(j+l+m+1)$ free k-summations, giving a factor $\Omega^{j+l+m+1}$. When pre-multiplied by the

[19] This result fails in the exceptional case of a system of bosons below the Bose–Einstein condensation temperature, since contractions involving the zero-momentum single-particle state then give additional factors of the volume. In order to apply the formalism to such a system, the zero-momentum mode would have to be eliminated first by methods which are well known, but will not be discussed here.

[20] Note that the denominator of (34) is only $O(1)$ by the BZT–Wentzel proof, which involves essentially the same arguments as ours.

[21] The $j=0$ term is the denominator of (45) is 1, and that in the numerator is $\langle O^{(0)}(\mathbf{r},0)U_-(t)O^{(0)}(\mathbf{r}',t)U_+(t)\rangle_0$.

factor $\Omega^{-(j+l+m+2)}$ in (48), one obtains a negligible contribution $O(\Omega^{-1})$.

The only contributions remaining are of the form $\langle\ \rangle_0'''$, where the triple prime denotes omission of terms in which any $b'^{(0)}$ or $b^{\dagger\prime(0)}$, or *either* of $a_k^\dagger a_{k'}$ or $a_p^\dagger a_{p'}$, is self-contracted, as well as omission of terms in which $a_k^\dagger a_{k'}$ is contracted with $a_p^\dagger a_{p'}$ leaving no connection to the other factors in (50). By topological analysis similar to that employed in proving (41), one then shows that all expressions $\langle\ \rangle_0'''$ are $O(\Omega^{j+l+m+1})$ or smaller, so that (49), and hence the term $\langle\ \rangle_0'$ in (48), are $O(\Omega^{-1})$. The structures of all contributions to $\langle\ \rangle_0'''$ of $O(\Omega^{j+l+m+1})$ are shown in Fig. 5 in the Appendix; all other contributions are still smaller. One thus concludes by (48) and (45) that

$$\langle O(\mathbf{r},0)O(\mathbf{r}',t)\rangle_{\text{eq}} = \langle O^{(0)}(\mathbf{r},0)O^{(0)}(\mathbf{r}',t)\rangle_0 + O(\Omega^{-1}). \tag{51}$$

Substitution of (51), (43), and (42) into the expression (25) for the transport kernel K gives

$$\begin{aligned} K(\mathbf{r},t;\mathbf{r}_1,0) = {} & \langle O^{(0)}(\mathbf{r},0)\rangle_0\langle O^{(0)}(\mathbf{r}_1,0)\rangle_0 \\ & - \frac{1}{2}\int_0^1 ds[\langle O^{(0)}(\mathbf{r}_1,0)O^{(0)}(\mathbf{r},t+i\beta s)\rangle_0 \\ & + \langle O^{(0)}(\mathbf{r},t-i\beta s)O^{(0)}(\mathbf{r}_1,0)\rangle_0] \\ & + O(\Omega^{-1}). \end{aligned} \tag{52}$$

Thus, no error is made in the transport kernel K in the thermodynamic limit ($\Omega\to\infty$) if H is replaced by H_0 in evaluating K. This is what we wished to prove: the quasiparticle interaction Hamiltonian H' has *no* effect on *linear* transport in the thermodynamic limit.

4. GENERALIZATION TO NONLINEAR TRANSPORT AND TWO-PARTICLE OBSERVABLES

The terms beyond Δ_2 in the expansion (20) lead to terms in $\langle O(\mathbf{r},t)\rangle$ which depend in a nonlinear way on the initial value $\langle O(\mathbf{r},0)\rangle - \langle O(\mathbf{r},0)\rangle_{\text{eq}}$; they become important when the initial deviation from equilibrium is not small. Extending (20), (22), and (23) to the next order, one finds

$$W_3 = -\beta^{-1}(\Delta_3 - \Delta_1\Delta_2 + \tfrac{1}{3}\Delta_1^3), \tag{53}$$

with

$$\begin{aligned} \Delta_3 = {} & -\frac{1}{3}\int_0^1 ds\int_{1-s}^1 ds'\int_{-\infty}^{\infty}\!\!\int\!\!\int dt_1\,dt_2\,dt_3 \\ & \times\int\!\!\int\!\!\int d^3r_1\,d^3r_2\,d^3r_3\lambda(\mathbf{r}_1,t_1)\lambda(\mathbf{r}_2,t_2)\lambda(\mathbf{r}_3,t_3) \\ & \times\langle O(\mathbf{r}_1,0)O(\mathbf{r}_2,t_2-t_1+i\beta s) \\ & \times O(\mathbf{r}_3,t_3-t_1+i\beta s')\rangle_{\text{eq}}. \end{aligned} \tag{54}$$

Applying (16) to (23), one finds, with (53) and (54),

$$\langle O(\mathbf{r},t)\rangle = \langle O(\mathbf{r},0)\rangle_{\text{eq}} + \langle O(\mathbf{r},t)\rangle_1 + \langle O(\mathbf{r},t)\rangle_2 + \cdots, \tag{55}$$

where $\langle\ \rangle_1$ is the linear term of Secs. 2 and 3:

$$\langle O(\mathbf{r},t)\rangle_1 = \int K(\mathbf{r},t;\mathbf{r}_1,0)\lambda_0(\mathbf{r}_1)\,d^3r_1, \tag{56}$$

and $\langle\ \rangle_2$ is the first nonlinear[22] correction:

$$\begin{aligned} \langle O(\mathbf{r},t)\rangle_2 = {} & -\tfrac{1}{2}\langle O(\mathbf{r},0)\rangle_{\text{eq}}\int_0^1 ds \\ & \times\iint d^3r_1\,d^3r_2\lambda_0(\mathbf{r}_1)\lambda_0(\mathbf{r}_2)\langle O(\mathbf{r}_1,0)O(\mathbf{r}_2,i\beta s)\rangle_{\text{eq}} \\ & - \langle O(\mathbf{r},t)\rangle_1\int d^3r_1\lambda_0(\mathbf{r}_1)\langle O(\mathbf{r}_1,0)\rangle_{\text{eq}} \\ & + \frac{1}{3}\int_0^1 ds\int_{1-s}^1 ds'\iint d^3r_1\,d^3r_2\lambda_0(\mathbf{r}_1)\lambda_0(\mathbf{r}_2) \\ & \times[\langle O(\mathbf{r},t)O(\mathbf{r}_1,i\beta s)O(\mathbf{r}_2,i\beta s')\rangle_{\text{eq}} \\ & + \langle O(\mathbf{r}_1,0)O(\mathbf{r},t+i\beta s)O(\mathbf{r}_2,i\beta s')\rangle_{\text{eq}} \\ & + \langle O(\mathbf{r}_1,0)O(\mathbf{r}_2,i\beta s)O(\mathbf{r},t+i\beta s')\rangle_{\text{eq}}]. \end{aligned} \tag{57}$$

It has already been shown in Sec. 3 that only errors of order Ω^{-1} are introduced in evaluating $\langle O(\mathbf{r},0)\rangle_{\text{eq}}$, $\langle O(\mathbf{r},t)O(\mathbf{r}',t')\rangle_{\text{eq}}$, and $\langle O(\mathbf{r},t)\rangle_1$ if H is replaced by H_0; hence it only remains to be shown that the same is true of the triple correlation functions $\langle OOO\rangle_{\text{eq}}$. The reasoning in Sec. 3 and the Appendix is easily extended to show that this is, indeed, true. In fact, the reasoning is easily extended to the general term $\langle O(\mathbf{r},t)\rangle_{n-1}$ in (55), for any finite n; such a term involves n-fold correlation functions, i.e., n factors of O. Noting that the product of n O-operators must contain at least two linkages to H' factors [cf. Fig. 5], one finds that if all chains[23] are of minimum possible length, then there will be exactly n O-chains and $(m-1)$ H'-chains, where n is the number of factors of O, m the number of factors of H', an O-chain is one terminated by labeled sets $\mathcal{S}_k$, $\mathcal{S}_{k'}$, as in Figs. 1–5, and an H'-chain contains no such labeled sets. One thus has $(n+m-1)$ free k-summations, giving a factor Ω^{n+m-1}. When pre-multiplied by the factor $\Omega^{-(n+m)}$ coming from the explicit Ω^{-1} in each O and H' factor, one obtains a contribution of $O(\Omega^{-1})$. Thus, to all finite orders in λ_0 (i.e., in the departure from equilibrium) and in H', one concludes that only errors of order Ω^{-1} are made in evaluating $\langle O(\mathbf{r},t)\rangle$ [Eq. (16)] if H is replaced by H_0

[22] Being quadratic in λ_0, $\langle\ \rangle_2$ is quadratic in the departure from equilibrium.

[23] Our terminology and reasoning should be clear from reference to the Appendix; the details of the proof are left to the reader.

and $O(\mathbf{r}, t)$ by $O^{(0)}(\mathbf{r}, t)$ in (15), provided that the initial deviation from equilibrium,

$$\langle O(\mathbf{r}, 0)\rangle - \langle O(\mathbf{r}, 0)\rangle_{\text{eq}},$$

is spatially localized.

We have so far taken $O(\mathbf{r})$ to be a single-particle operator [Eq. (30)]. There are, however, physically interesting operators which, though labeled by a single position variable $\mathbf{r}$, are nevertheless two-particle or several-particle operators, or at least contain portions which are. One example is the energy (Hamiltonian) density operator $\mathcal{H}(\mathbf{r})$, which is important in heat transport. Let $\mathcal{U}(\mathbf{r})$ be the two-particle interaction-density operator, i.e., the contribution of two-particle interactions to $\mathcal{H}(\mathbf{r})$. Then $\mathcal{U}(\mathbf{r})$ has the general structure

$$\mathcal{U}(\mathbf{r}) = \Omega^{-2} \sum_{k_1 k_2 k_3 k_4} \mathcal{U}_{k_1 k_2 k_3 k_4}(\mathbf{r}) a_{k_1}^\dagger a_{k_2}^\dagger a_{k_4} a_{k_3} \tag{58}$$

with volume-independent matrix elements $\mathcal{U}_{k_1 k_2 k_3 k_4}(\mathbf{r})$. It is easy to see that our proofs still go through for such an operator; $\mathcal{U}(\mathbf{r})$ behaves as the product of two one-particle operators (30). Thus the only essential property of $O(\mathbf{r})$ is its locality, together with the spatial localization of $\langle O(\mathbf{r}, 0)\rangle - \langle O(\mathbf{r}, 0)\rangle_{\text{eq}}$.

5. TRANSPORT DUE TO EXTERNAL FIELDS

Our analysis has thus far been restricted to cases in which the temporal evolution of $\langle O(t)\rangle$ is due not to any external field absent in thermal equilibrium, but instead is a result of the noncommutativity of O with H together with an initial localized deviation, $\langle O(0)\rangle - \langle O(0)\rangle_{\text{eq}}$, from equilibrium. An equally interesting case is that in which the temporal evolution of $\langle O(t)\rangle$ is due to the presence of some perturbation Hamiltonian V not contained in H. We call V the "external field", although in applications it might be entirely due to internal influences, e.g., fixed impurity centers in a superconductor. Thus "external" means merely "not contained in H." A time-independent external field will be of the general form

$$V = \Omega^{-1} \sum_{kk'} \int d^3r V_{kk'}(\mathbf{r}) a_k a_{k'}, \tag{59}$$

where $V_{kk'}(\mathbf{r})$ is volume-independent.

Let us consider, in particular, the case that V is not present for $t < 0$, and is turned on instantaneously at $t = 0$. Then for $t \geq 0$, the statistical expectation value $\langle O(t)\rangle$ of any observable O is given by

$$\langle O(t)\rangle = \text{Tr}\,[\rho(0) O(t)], \tag{60}$$

where $\rho(0)$ corresponds to thermal equilibrium in the absence of V:

$$\rho(0) = e^{-\beta H}/\text{Tr}\, e^{-\beta H}, \tag{61}$$

and O is propagated for $t > 0$ with the full Hamiltonian $H + V$:

$$O(t) = e^{it(H+V)} O e^{-it(H+V)}. \tag{62}$$

Let us make a perturbation expansion of $\langle O(t)\rangle$ in powers of V:

$$\begin{aligned}\langle O(t)\rangle &= \left\langle \exp_- \left[i \int_0^t dt' V_{\text{eq}}(t') \right] O_{\text{eq}}(t) \right. \\ &\qquad \left. \times \exp_+ \left[-i \int_0^t dt' V_{\text{eq}}(t') \right] \right\rangle_{\text{eq}} \\ &= \langle O(0)\rangle_{\text{eq}} + \langle O(t)\rangle_1 + \cdots, \end{aligned} \tag{63}$$

where

$$\langle O(t)\rangle_1 = i \int_0^t dt' \langle [V_{\text{eq}}(t' - t), O]\rangle_{\text{eq}}, \tag{64}$$

$\langle\ \rangle_{\text{eq}}$ is defined by (19), and

$$O_{\text{eq}}(t) = e^{itH} O e^{-itH}. \tag{65}$$

A further expansion in powers of H', as in (45), gives

$$\begin{aligned}&\langle [V_{\text{eq}}(t' - t), O]\rangle_{\text{eq}} \\ &= \left[\sum_{j=0}^{\infty} (-i)^j \int_0^{-i\beta} dt_1 \cdots \int_0^{t_{j-1}} dt_j \langle H'^{(0)}(t_1) \cdots \right. \\ &\qquad \left. \times H'^{(0)}(t_j) [U_-(t' - t) V^{(0)}(t' - t) U_+(t' - t), O]\rangle_0 \right] \\ &\qquad \times \left[\sum_{j=0}^{\infty} (-i)^j \int_0^{-i\beta} dt_1 \cdots \int_0^{t_{j-1}} dt_j \right. \\ &\qquad \left. \times \langle H'^{(0)}(t_1) \cdots H'^{(0)}(t_j)\rangle_0 \right]^{-1}, \end{aligned} \tag{66}$$

where all quantities are defined as in (45). Application of Matsubara's theorem as in (48) gives

$$\begin{aligned}&\langle [V_{\text{eq}}(t' - t), O]\rangle_{\text{eq}} - \langle [V^{(0)}(t' - t), O]\rangle_0 \\ &= \left[\sum_{j=0}^{\infty} (-i)^j \int_0^{-i\beta} dt_1 \cdots \int_0^{t_{j-1}} dt_j \langle H'^{(0)}(t_1) \cdots \right. \\ &\qquad \left. \times H'^{(0)}(t_j) [U_-(t' - t) V^{(0)}(t' - t) U_+(t' - t), O]\rangle_0' \right] \\ &\qquad \times \left[\sum_{j=0}^{\infty} (-i)^j \int_0^{-i\beta} dt_1 \cdots \int_0^{t_{j-1}} dt_j \right. \\ &\qquad \left. \times \langle H'^{(0)}(t_1) \cdots H'^{(0)}(t_j)\rangle_0 \right], \end{aligned} \tag{67}$$

where the prime on $\langle\ \rangle_0'$ implies omission of all terms in which *none* of the annihilation and creation operators in O, U_-, U_+, or $V^{(0)}$ are contracted with those in any of the factors $H'^{(0)}(t_1) \cdots H'^{(0)}(t_j)$.

In order to proceed with the analysis, it is now necessary to make definite assumptions about the form of the operator O whose time evolution is being investigated. Let us first suppose that O is a single-particle operator of the same general structure as V, i.e.,

$$O = \Omega^{-1} \sum_{kk'} O_{kk'} a_k^\dagger a_{k'} \tag{68}$$

with volume-independent $O_{kk'}$. Then by an analysis almost identical with that used in proving (51), one finds that the right-hand side of (67) is $O(\Omega^{-1})$ *provided that the perturbation V is localized* in the sense that $V_{kk'}(\mathbf{r})$ in (59) has finite (Ω-independent) range with respect to some fixed center or with respect to a finite[24] (Ω-independent) set of such centers. On the other hand, it is easy to see[25] that, subject to the same conditions

$$\langle [V^{(0)}(t'-t), O]\rangle_0 = O(1) \tag{69}$$

the analysis can be extended to all orders in V, just as that in Sec. 4 was extended to all orders in λ_0. One thus concludes that, for *localized* perturbations V, $\langle O(t)\rangle$ can be evaluated with negligible error in the thermodynamic limit by replacing H by H_0 in (60)–(63). The analysis can be generalized to show that the conclusion is also true for two-particle operators of the general structure of (58) (with the $\mathbf{r}$ dependence omitted).

On the other hand, our formalism is not easily extended to the case of nonlocalized perturbations V, i.e., cases in which $V_{kk'}(\mathbf{r})$ in (59) is of infinite range with respect to a finite (Ω-independent) set of centers or of finite range with respect to a volume-proportional set of scattering centers.[26] In such a case, the expansion (63) becomes useless since successive terms involve higher and higher powers of Ω. To treat such a situation one would have to use linked-cluster perturbation theory for the generalized thermodynamic potential, rather than perturbation theory for the trace. It is almost certainly not true that for nonlocalized perturbations V, the quasiparticle-interaction (H') effects on the evolution of O are negligible to all orders in V. On the other hand, a continuity argument suggests that the effects of H' might be small for a V which, though nonlocalized, is "small" in some sense. However, we cannot draw any definite conclusions about such a case here.

Even for localized V, the effects of H' on the evolution of *all* observables O is certainly not negligible. Consider, e.g., the case

$$O = \sum_k O_k a_k^\dagger a_k, \tag{70}$$

where O_k is volume-independent. Although this is formally of the form (68) with $O_{kk'} = \Omega\delta_{kk'}$, we assumed in (68) that $O_{kk'}$ is Ω-independent; thus (70) is not of the form assumed before, and requires separate investigation. One in fact finds that, for O of the form (70), the right-hand side of (67) is $O(1)$ for localized perturbations V. On the other hand, it is easy to see that (69) still holds with O of the form (70).[25] Thus in this case, the contributions of H' to the time evolution of $\langle O(t)\rangle$ are of the same order as those of H_0, i.e., the thermodynamically equivalent Hamiltonian method fails.

An illuminating example is provided by taking O in (68) to be the momentum-density operator

$$P(\mathbf{r}) = \psi^\dagger(\mathbf{r})(\nabla/i)\psi(\mathbf{r}) = \Omega^{-1}\sum_{kk'} k' e^{-i(k-k')\cdot\mathbf{r}} a_k^\dagger a_{k'} \tag{71}$$

in a free-particle representation. Then one can calculate the temporal evolution of $P(\mathbf{r})$ due to a scattering center[27] V exactly in the thermodynamic limit ($\Omega \to \infty$) by ignoring the quasiparticle interactions H'. As expected physically, the effect of V on $P(\mathbf{r})$ is $O(1)$ (Ω-independent), whereas that of H' is only $O(\Omega^{-1})$. Suppose, on the other hand, we try to calculate the temporal evolution of the *total* momentum

$$P = \int d^3r P(\mathbf{r}) = \sum_k k a_k^\dagger a_k. \tag{72}$$

Then the effect of H' on $\langle P(t)\rangle$ is not negligible, being $O(1)$ just as is the main contribution (the effect of V via H_0). This seems somewhat paradoxical, since the evolution $\langle P(t)\rangle$ can be calculated indirectly by first calculating $\langle P(\mathbf{r}, t)\rangle$ and then integrating over $\mathbf{r}$; H' is negligible in evaluating $\langle P(\mathbf{r}, t)\rangle$. Note, however, that these two different methods of calculating $\langle P(t)\rangle$ differ in the order of performing the limit $\Omega \to \infty$ and the $\mathbf{r}$-integration. Apparently these two operations may not be interchanged with impunity. If the *exact* $\langle P(\mathbf{r}, t)\rangle - \langle P(\mathbf{r}, 0)\rangle$ is known to have finite range with respect to the center of V, then the right answer for $\langle P(t)\rangle$ must be obtained by first calculating $\langle P(\mathbf{r}, t)\rangle$ with neglect of H' in the limit $\Omega \to \infty$, and then integrating over $\mathbf{r}$, since the integral of the $O(\Omega^{-1})$ contribution of H' over a *finite* range is still $O(\Omega^{-1})$.

6. INADEQUACIES OF A PROOF "TO ALL ORDERS"

Just as is the case for all proofs "to all orders," our proofs are not rigorous since the following possibilities have not been excluded: (a) the series in the numerator or denominator of (34), (45), or (66) might fail to converge, or (b) the effect of H' might be nonanalytic and hence not admit any such series expansion. It is, however, reassuring that Bogoliubov has presented an intricate, but apparently rigorous,

[24] Note that this rules out the case of a nonzero *density* of scatterers.

[25] Self-contractions of O cancel due to the commutator; otherwise the left-hand side of (69) would be $O(\Omega)$.

[26] This would be the case for a nonzero *density* of scatterers.

[27] This is essentially the case considered by Wentzel (Ref. 10).

proof[28] which confirms the conclusions of the original BZT proof[1] "to all orders." Since our proof is almost identical with the latter, it seems not unreasonable to expect that its conclusion is also correct.

7. DISCUSSION; SUGGESTED APPLICATIONS; UNSOLVED PROBLEMS

We have shown that the BZT–Wentzel method of the "thermodynamically equivalent Hamiltonian" can be extended to the following nonequilibrium situations: (a) temporal evolution of the statistical expectation value, $\langle O(\mathbf{r}, t)\rangle$, of an observable $O(\mathbf{r})$ due to an initial deviation from equilibrium, $\langle O(\mathbf{r}, t)\rangle - \langle O(\mathbf{r}, 0)\rangle$, which is *spatially localized* but not necessarily small; (b) temporal evolution of the statistical expectation value, $\langle O(t)\rangle$, of an observable O due to a perturbation V which is *spatially localized*, but not necessarily small. In order that the proofs go through, it is necessary that $O(\mathbf{r})$ and O are "well-behaved" in the sense that their matrix elements do not contain delta functions in momentum space. It is noted that this requirement excludes, e.g., the case that O is the total linear momentum, but does not exclude the momentum density.

There are a number of interesting problems which can be investigated with this formalism. Examples which come to mind immediately are transport of spin, energy, and current in superconductors and ferromagnets. We hope to report on such investigations in the future.

The more difficult question of the influence of quasiparticle interactions on nonlocalized transport ought to be investigated, in view of its importance to problems of thermal and electrical conductivity. Such investigations require linked-cluster perturbation expansions or their equivalent.

ACKNOWLEDGMENT

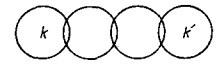
This research was supported in part by the National Science Foundation (GP 4367).

APPENDIX

We wish to show that each expression $\langle\ \rangle_0''$ in (40) vanishes unless not more than j of the $2j+2$ vectors $k_1 \cdots k_j$, $k_1' \cdots k_j'$, kk' are summed over independently.

Note first that according to (39), (33), and (7), $b_{k\lambda}'^{(0)}(t)$ differs from $b_{k\lambda}'$ only by phase factors; thus the time argument can be ignored in discussing the topology of contractions. Now, consider the simplest case $j = 1$. If $a_k^\dagger$ and $a_{k'}$ are both contracted with $b_{k_1\lambda_1}'^\dagger$ or both with $b_{k_1'\lambda_1'}'$, we get no contribution since the remaining factor $b_{k_1'\lambda_1'}'$ or $b_{k_1\lambda_1}^{\dagger\prime}$ is not allowed to be self-contracted according to the definition of $\langle\ \rangle_0''$. If $a_k^\dagger$ is contracted with, e.g., $b_{k_1\lambda_1}^{\dagger\prime}$ and $a_{k'}$ with $b_{k_1'\lambda_1'}'$, then we only get a contribution when $\mathcal{S}_k$ overlaps $\mathcal{S}_{k_1}$ and $\mathcal{S}_{k'}$ overlaps $\mathcal{S}_{k_1'}$. But then $\mathcal{S}_{k_1}$ must overlap $\mathcal{S}_{k_1'}$ in order to get a nonzero contraction of the remaining two α and $\alpha^\dagger$ operators. This situation is illustrated schematically in Fig. 1; $\mathcal{S}_k$ and $\mathcal{S}_{k_1}$ are labeled by k and k', whereas $\mathcal{S}_{k_1}$ and $\mathcal{S}_{k_1'}$ are unlabeled. It is clear that only one free k-summation remains, giving a

FIG. 1. Topology of nonzero contractions for the case $j = 1$.



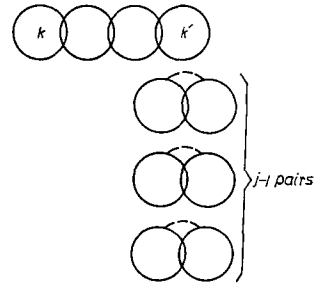


FIG. 2. Topology of nonzero contractions in which $a_k^\dagger$ and $a_{k'}$ are contracted with different $b^{\dagger\prime}$ or b' factors.

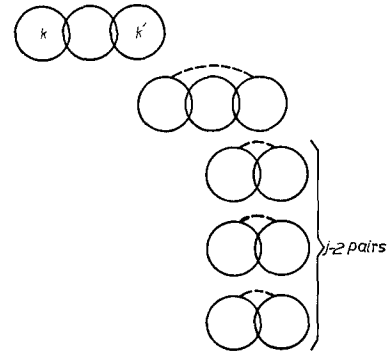


FIG. 3. Topology of nonzero contractions in which $a_k^\dagger$ and $a_{k'}$ are contracted with the same $b^{\dagger\prime}$ or b' factor.

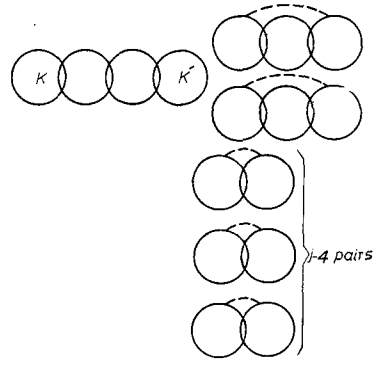


FIG. 4. An example of a contraction with chains of longer than optimal length.

[28] N. N. Bogoliubov, Physica, Suppl. 26, 1 (1960). A nonrigorous proof different from that of Ref. 1 was given by N. N. Bogoliubov, D. N. Zubarev, and Yu. A. Terserkovnikov, Zh. Eksperim. i Teor. Fiz. 39, 120 (1960) [English transl.: Soviet Phys.—JETP 12, 88 (1961)].

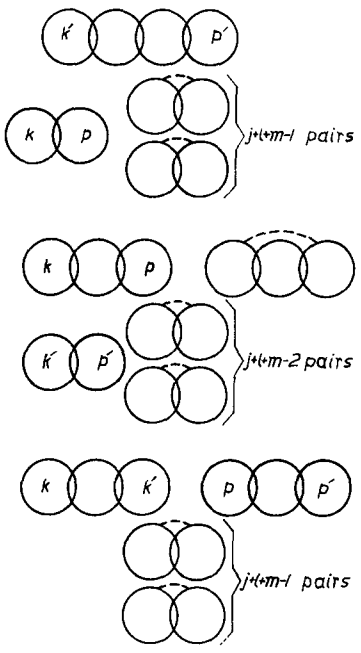


FIG. 5. Topology of all contributions to $\langle\ \rangle_0'''$ of order $\Omega^{j+l+m+1}$. The labels k, k', p, p' may be permuted subject to the restrictions that S_k and $S_{k'}$ may not overlap, and S_p and $S_{p'}$ may not overlap.

single factor Ω. Thus, because of the explicit factor $\Omega^{-(1+1)}$ in (40), we get a net $O(\Omega^{-1})$ contribution.

The terms with $j > 1$ in which $a_k^\dagger$ and $a_{k'}$ are contracted with different $b^{\dagger\prime}$ or b' operators have a similar structure, denoted by Fig. 2. They differ from Fig. 1 only in that the remaining $(2j-2)$ $b^{\dagger\prime}$ and b' operators are paired in order to avoid self-contractions. The dashed line connecting the two members of each such pair denotes one contraction between a β operator in one $b^{\dagger\prime}$ or b' and a $\beta^\dagger$ operator in the other; the other contraction is implied by the overlap of the circles. There are also nonzero terms with $j > 1$ in which $a_k^\dagger$ and $a_{k'}$ are contracted with the *same* $b^{\dagger\prime}$ or b' factor; in order to avoid self-contractions of $b^{\dagger\prime}$ or b' factors, these will have the structure shown in Fig. 3. It is clear that both Fig. 2 and Fig. 3 have j free k-summations, giving a factor Ω^j. When premultiplied by the explicit factor $\Omega^{-(j+1)}$ in (40), they give net contributions of $O(\Omega^{-1})$.

Finally, there are nonzero contractions for $j > 2$ in which there are some chains of longer than optimal length. All such contractions have fewer than j free k-summations, and hence give contributions to (40) which are even smaller than the already-negligible contributions $O(\Omega^{-1})$ already considered. An example is given in Fig. 4.

The topology of all contributions to $\langle\ \rangle_0'''$ [see Eq. (50) and subsequent discussion] of $O(\Omega^{j+l+m+1})$ is shown in Fig. 5. All other contributions are $O(\Omega^{j+l+m})$ or smaller, due to the presence of chains of greater than optimal length.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Local Weights Which Determine Area, and the Ising Model

GLEN BAXTER
Department of Statistics, Purdue University, Lafayette, Indiana

(Received 4 May 1966)

A group G of local weights is constructed which assigns to closed paths in the square lattice the enclosed area and the number of turns of the tangent vector (mod 2) to the path. Special cases of this group have been used previously in explicit evaluations of the partition function for the Ising model in 2-dimensions. Properties of G are examined to cast light on the combinatorial approach to the Ising problem developed by Kac and Ward, Feynman, and Sherman. It is shown that their method breaks down in the general case.

1. INTRODUCTION

LET $\mathfrak{L}$ denote the two-dimensional square lattice in the plane with vertices (m, n), where m and n are integers, and with unit-long-segment edges connecting vertices in the horizontal and vertical directions. By a path, we mean a well-defined sequence of directed edges in $\mathfrak{L}$ touching terminal to initial such that two consecutive edges are not the same. This means that a path can never reverse direction at a point. It can *back up*, however, provided only that the initial point of the first edge touches the terminal point of the second edge (see Fig. 1).

In a recent paper on the Ising model by the author,[1] there was constructed a group G of "local" weights for $\mathfrak{L}$ which assigned to a closed path in $\mathfrak{L}$ a weight from

[1] Glen Baxter, J. Math. Phys. 6, 1015 (1965).

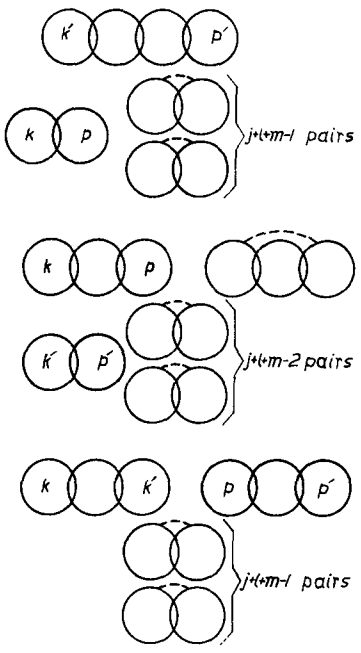


FIG. 5. Topology of all contributions to $\langle\ \rangle_0'''$ of order $\Omega^{j+l+m+1}$. The labels k, k', p, p' may be permuted subject to the restrictions that S_k and $S_{k'}$ may not overlap, and S_p and $S_{p'}$ may not overlap.

single factor Ω. Thus, because of the explicit factor $\Omega^{-(1+1)}$ in (40), we get a net $O(\Omega^{-1})$ contribution.

The terms with $j > 1$ in which $a_k^\dagger$ and $a_{k'}$ are contracted with different $b^{\dagger\prime}$ or b' operators have a similar structure, denoted by Fig. 2. They differ from Fig. 1 only in that the remaining $(2j-2)$ $b^{\dagger\prime}$ and b' operators are paired in order to avoid self-contractions. The dashed line connecting the two members of each such pair denotes one contraction between a β operator in one $b^{\dagger\prime}$ or b' and a $\beta^\dagger$ operator in the other; the other contraction is implied by the overlap of the circles. There are also nonzero terms with $j > 1$ in which $a_k^\dagger$ and $a_{k'}$ are contracted with the *same* $b^{\dagger\prime}$ or b' factor; in order to avoid self-contractions of $b^{\dagger\prime}$ or b' factors, these will have the structure shown in Fig. 3. It is clear that both Fig. 2 and Fig. 3 have j free k-summations, giving a factor Ω^j. When premultiplied by the explicit factor $\Omega^{-(j+1)}$ in (40), they give net contributions of $O(\Omega^{-1})$.

Finally, there are nonzero contractions for $j > 2$ in which there are some chains of longer than optimal length. All such contractions have fewer than j free k-summations, and hence give contributions to (40) which are even smaller than the already-negligible contributions $O(\Omega^{-1})$ already considered. An example is given in Fig. 4.

The topology of all contributions to $\langle\ \rangle_0'''$ [see Eq. (50) and subsequent discussion] of $O(\Omega^{j+l+m+1})$ is shown in Fig. 5. All other contributions are $O(\Omega^{j+l+m})$ or smaller, due to the presence of chains of greater than optimal length.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Local Weights Which Determine Area, and the Ising Model

GLEN BAXTER
Department of Statistics, Purdue University, Lafayette, Indiana

(Received 4 May 1966)

A group G of local weights is constructed which assigns to closed paths in the square lattice the enclosed area and the number of turns of the tangent vector (mod 2) to the path. Special cases of this group have been used previously in explicit evaluations of the partition function for the Ising model in 2-dimensions. Properties of G are examined to cast light on the combinatorial approach to the Ising problem developed by Kac and Ward, Feynman, and Sherman. It is shown that their method breaks down in the general case.

1. INTRODUCTION

LET $\mathfrak{L}$ denote the two-dimensional square lattice in the plane with vertices (m, n), where m and n are integers, and with unit-long-segment edges connecting vertices in the horizontal and vertical directions. By a path, we mean a well-defined sequence of directed edges in $\mathfrak{L}$ touching terminal to initial such that two consecutive edges are not the same. This means that a path can never reverse direction at a point. It can *back up*, however, provided only that the initial point of the first edge touches the terminal point of the second edge (see Fig. 1).

In a recent paper on the Ising model by the author,[1] there was constructed a group G of "local" weights for $\mathfrak{L}$ which assigned to a closed path in $\mathfrak{L}$ a weight from

[1] Glen Baxter, J. Math. Phys. 6, 1015 (1965).

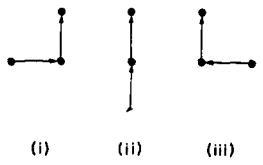


FIG. 1. The three path configurations at a lattice point.

which one could determine the change in the argument of the tangent vector (mod 2) in one traversal of the path and the number of enclosed squares (mod 2). These weights were used to evaluate explicitly the partition function for the two-dimensional Ising model with a certain imaginary external magnetic field. In order to evaluate the partition function for the two-dimensional Ising model with arbitrary external magnetic field, one would have to determine for closed paths the number of enclosed squares *in general* and not just mod 2. This raises the question: Can we construct a weight group G of local weights which assigns to a path a weight (an element of G) from which we can recapture the exact number of enclosed squares? In other words, is area a "global" property which allows itself to be analyzed "locally?" Our goal here is to show that it is.

The area enclosed inside a simple closed nonintersecting path is very easy to describe. But what is the area of a path which intersects itself? In Fig. 2, we have drawn two paths which "enclose" the same two squares. At first glance it seems clear that any group G of local weights must assign area weight 2 to both paths. However, we see in Sec. 2 that, necessarily, the possible groups G assign area weight 2 to path (a) and area weight 0 to path (b). That is, the group G which we construct assigns a positive area weight to a simple closed path with a clockwise sense and an equal negative area weight to the same path traversed in the opposite sense. Figure-eight paths like path (b) of Fig. 2 are assigned weights which come from the difference of the "clockwise" and "counterclockwise" portions of the path. Only in the case of area (mod 2) is the area assigned to path (a) the same as that assigned to path (b) in Fig. 2.

The comments of the preceding paragraph indicate difficulty in using the more general weight groups G to solve the Ising model with nonzero external magnetic field. The combinatorial approach to the Ising problem is based on a "figure-eight" cancellation. However, since it is not true that the assigned areas are the same for paths (a) and (b) of Fig. 2, the weights do not cancel as desired. In Sec. 4, we discuss the Ising model with an eye toward understanding the difficulty just mentioned. As yet, we have been unable to find any specific information on the two-dimensional Ising model with arbitrary external magnetic field using our general area groups G.

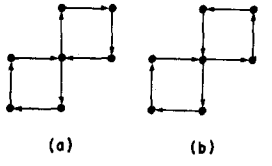


FIG. 2. Two closed paths enclosing the same lattice squares.

In Sec. 2, we construct the group G of local weights which assigns to a simple closed path traversed in a clockwise sense the enclosed area [or enclosed area (mod r)] and which also assigns to any closed path the change of the argument of the tangent vector (mod 2). Properties of this group are examined in Sec. 3, where we also look at specific examples with $r = 1, 2, 3$. We show that the area (mod r) groups G_r are all finite, and that they have the order $8r^3$. As we said earlier, some combinatorial aspects of the Ising problem are discussed in Sec. 4.

2. CONSTRUCTION OF THE GROUP G

Let $\mathfrak{L}$ be the lattice as described in Sec. 1. For any path which does not double back on itself at a vertex, there are three path configurations that are possible as illustrated in Fig. 1 We assign a local weight $\hat{\alpha}$, β, α to the one-step motion along the direction shown for the configurations (i), (ii), (iii), respectively. The weights for backing up are assigned the inverses of weights for the forward motion. For any path we assign a weight as follows:

> Let $v_1, \cdots, v_n$ be the vertices of the path and let $\alpha_1, \cdots, \alpha_n$ be the corresponding weights assigned to the path configuration at these vertices according to the above. The weight of the path is then $w = \alpha_1 \alpha_2 \cdots \alpha_n$.

The question which we posed in the introduction can now be re-phrased. Can we find a group G (resp. G_r) generated by $\hat{\alpha}$, β, and α such that the weight assigned to any simple closed path traversed in a clockwise sense gives the enclosed area [resp. area (mod r)] and the change of the argument of the tangent vector (mod 2) to the path? To describe the area, we use an element b (an element of G, if it exists) and to describe the "twist" of the tangent vector we use an element a (again an element of G, if it exists). That is, for a simple closed path the weight assigned should have the form ab^j, where j is the number of enclosed units of area. Since any closed path can be described in a variety of manners using different directed edges as

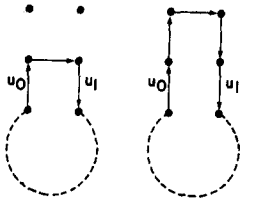


FIG. 3. Two closed paths which differ only by inclusion or non-inclusion of sides from a particular square.

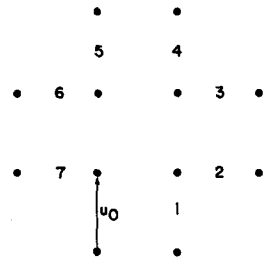


FIG. 4. The seven positions for the directed edge of the closed path as the path leaves the central square.

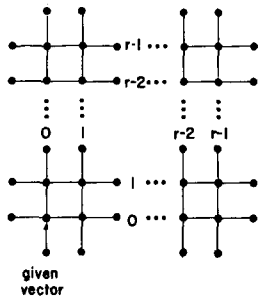


FIG. 5. The possible positions for the ending vector for simple paths whose weights will be different in G_r.

the start, it follows that a and b must lie in the center of G. Also $a^2 = I$, but as yet b has no finite order. If we want simply the enclosed area (mod r) we would require that $b^r = I$. We propose to find necessary conditions for the groups G and G_r. From these we find that groups G and G_r do actually exist.

Let us look at Fig. 3. There we have shown two paths which differ only in the fact that one of them, namely the second, contains one more area unit than does the other. If there is a group G of weights, then we can cancel out the common weight associated with that part of both paths lying between u_1 and u_0 and deduce the relation $b\alpha^2 = \beta\alpha^2\beta$. Actually, there are seven relations like this one which can be deduced, one for each of seven different path pairs corresponding to different positions of u_1 as shown in Fig. 4. Adding the obvious condition $\alpha^4 = ab$ to the others, we get a list of eight conditions

$$\begin{aligned} &\text{(a)}\quad b\alpha^2 = \beta\alpha^2\beta, \quad \text{(e)}\quad b\alpha\hat{\alpha}^2\alpha = \beta^2,\\ &\text{(b)}\quad b\alpha\beta = \beta\alpha^2\hat{\alpha}, \quad \text{(f)}\quad b\alpha\hat{\alpha}^2\beta = \beta\hat{\alpha},\\ &\text{(c)}\quad b\alpha\hat{\alpha}\alpha = \beta\alpha\beta, \quad \text{(g)}\quad b\alpha\hat{\alpha}^3\alpha = \hat{\alpha},\\ &\text{(d)}\quad b\alpha\hat{\alpha}\beta = \beta\alpha\hat{\alpha}, \quad \text{(h)}\quad \alpha^4 = ab. \end{aligned} \tag{1}$$

These conditions can be simplified to the following four:

$$\begin{aligned} &\text{(i)}\quad \alpha^{-1}\beta = \beta^{-1}\hat{\alpha},\\ &\text{(ii)}\quad \hat{\alpha}^{-1}\beta^{-1}\hat{\alpha} = \alpha^{-1}\beta\alpha,\\ &\text{(iii)}\quad b\alpha^2 = \beta\alpha^2\beta,\\ &\text{(iv)}\quad \alpha^4 = ab. \end{aligned} \tag{2}$$

In fact, (i) follows from (a) and (b). From (b) and (c), it follows that $\beta^{-1}\hat{\alpha}\alpha = \hat{\alpha}^{-1}\alpha^{-1}\beta$, and from (i) we get (ii). It is also true that (a)–(h) follow from (i)–(iv). In fact, (a) and (b) follow from (iii) and (i). Reversing the steps just above shows that (i), (ii), and (b) implies (c). Then, (c) and (i) implies (d), etc. Conditions (b) are thus necessary for the existence of the group G. The additional relation

$$\text{(v)}\quad \beta^r = I \tag{3}$$

will be added to the conditions necessary for the existence of G_r.

Theorem: There exists a group G containing elements α, β, $\hat{\alpha}$ such that (2) is satisfied. If $K_r = \{\beta^r, b^r\}$ is the subgroup of G generated by β^r and b^r, we take $G_r = G/K_r$.

Proof: A heuristic discussion aids in understanding the next argument. We are going to think of writing every element of G in the form (possibly non-uniquely) of a multiple a^ib^j of

$$\beta^k \cdot \alpha\beta^{-1} \cdot \beta^m \cdot \beta\alpha^{-1} \cdot (\alpha\beta^{-1})^s \equiv (k, m, s). \tag{4}$$

The element in (4) has a geometric interpretation. The term $\alpha\beta^{-1}$ amounts to a rotation of a vector clockwise about a point through an angle of 90°. The term β^k amounts to a translation of k units along the direction of the directed edge or vector. Thus, (k, m, s) geometrically amounts to a motion from a starting vertex with a given direction to a new vertex displaced k units along the original direction, displaced m units to the right (looking along the original direction), and rotating by s 90° turns clockwise. Thus, we are going to identify elements of G with multiples a^ib^j of weights for simple paths starting with a given vector (directed edge) and ending with another vector. In the case of G_r with condition (v) in addition to (b), the elements are identified with multiples a^ib^j of weights for simple paths starting with a given vector and ending with a vector lying in a square of size r adjacent the original vector (see Fig. 5). The ending vector of the path occupies one of extreme most line segments of this square as pictured in Fig. 5 only if the direction of the ending vector is toward the central square. There are $4r^2$ ending vectors and $2r$ multiples a^ib^j so that we find for G_r a group that has $8r^3$ elements.

Let us continue with the heuristics for a moment longer. Suppose we want the product

$$(k_1, m_1, 0) \cdot (k_2, m_2, s_2), \tag{5}$$

where for simplification we have taken the case where $s_1 = 0$. In Fig. 6(a), we have drawn two paths which represent the combined parts of the two "motions" $(k_1, m_1, 0)$ and (k_2, m_2, s_2). We have also shown in Fig. 6(b) a path given by a single term (k_3, m_3, s_2), which would have the same starting and ending vector. The path in Fig. 6(b) contains m_1k_2 more area units

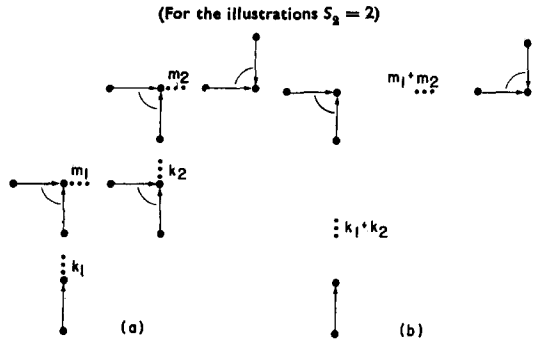


FIG. 6. The composite of two motions along the lattice and a single motion with the same terminal vector.

than does the path in Fig. 6(a). Thus, we would expect a relation

$$(k_1, m_1, 0)(k_2, m_2, s_2) = b^{-m_1k_2}(k_1 + k_2, m_1 + m_2, s_2). \tag{6}$$

By a completely analogous procedure, we can write three other relations depending on the particular value of s_1.

$$\begin{aligned}
&(k_1, m_1, 1)(k_2, m_2, s_2) \\
&\quad = b^{m_1m_2+m_2k_2}(k_1 - m_2, m_1 + k_2, 1 + s_2), \\
&(k_1, m_1, 2)(k_2, m_2, s_2) \\
&\quad = b^{m_1k_2}(k_1 - k_2, m_1 - m_2, 2 + s_2), \\
&(k_1, m_1, 3)(k_2, m_2, s_2) \\
&\quad = b^{-m_1m_2+m_2k_2}(k_1 + m_2, m_1 - k_2, 3 + s_2).
\end{aligned} \tag{7}$$

Finally, we would expect that

$$(k, m, s) = a(k, m, s - 4). \tag{8}$$

We now drop the heuristics and define a set of elements G^0: $a^ib^j(k, m, s)$ which have a multiplication defined by (6), (7), and (8), and

$$\begin{aligned}
&a^{i_1}b^{j_1}(k_1, m_1, s_1)a^{i_2}b^{j_2}(k_2, m_2, s_2) \\
&\quad = a^{i_1+i_2}b^{j_1+j_2}(k_1, m_1, s_1)(k_2, m_2, s_2).
\end{aligned} \tag{9}$$

We intend to show that G^0 is a group. According to (8), we can always reduce the consideration of values of s between 0 and 3. We have first

$$I = (0, 0, 0) \tag{10}$$

and

$$\begin{aligned}
(k, m, 0)^{-1} &= b^{-mk}(-k, -m, 0), \\
(k, m, 1)^{-1} &= \quad (-m, k, -1), \\
(k, m, 2)^{-1} &= b^{-mk}(k, m, -2), \\
(k, m, 3)^{-1} &= \quad (m, -k, -3).
\end{aligned} \tag{11}$$

Finally, we must show associativity. There are exactly 16 different cases to consider according to the values of s_1 and s_2 in the product

$$(k_1, m_1, s_1)(k_2, m_2, s_2)(k_3, m_3, s_3). \tag{12}$$

We consider just one such case and leave the remaining 15 to the interested reader. First, (12) can be written in the form

$$b^t(\cdots, \cdots, s_1 + s_2 + s_3) \tag{13}$$

in perhaps two different ways according to the different ways of forming the product in (12). There is no difficulty in seeing that the k and m terms of (13) are the same in either case. This is because the general relations (6), (7), and (8) show with $a = b = 1$ that (k, m, s) can be identified with a geometrical motion from (0, 0) with direction "up" to (m, k) with a direction s 90° turns clockwise. Thus, the k and m components of (k, m, s) combine in products just as they would if (k, m, s) were identified with a geometrical motion or transformation, and transformations are associative. Thus, we need consider only the two possible values for t in (13).

Consider for $s_1 = s_2 = 1$

$$[(k_1, m_1, 1)(k_2, m_2, 1)](k_3, m_3, s_3). \tag{14}$$

One can compute (14) to show that it has the form

$$b^{m_1m_2+m_2k_2}b^{k_3(m_1+k_2)}(\cdots, \cdots, 2 + s_3). \tag{15}$$

Also

$$(k_1, m_1, 1)[(k_2, m_2, 1)(k_3, m_3, s_3)]$$

can be shown to be equal to

$$b^{m_2m_3+m_3k_3}b^{m_1(m_2+k_3)+(m_2+k_3)(k_2-m_3)}(\cdots, \cdots, 2 + s_3). \tag{16}$$

Comparing the exponents t in (15) and (16) shows that they are the same. Thus, the associativity condition is satisfied in this case. In a similar manner, the associativity condition can be verified in the other 15 cases. We have thus shown that G^0 is a group.

Next, we show that our desired weight group G can be taken to be G^0. We use the fact that (2) is satisfied for the terms

$$\begin{aligned}
\alpha &= (0, 1, 1), \\
\hat{\alpha} &= (0, -1, -1), \\
\beta &= (1, 0, 0).
\end{aligned} \tag{17}$$

To assist in the computation, we note that

$$\begin{aligned}
\alpha^{-1} &= (-1, 0, -1), \\
\hat{\alpha}^{-1} &= (-1, 0, 1), \\
\beta^{-1} &= (-1, 0, 0).
\end{aligned}$$

We have

$$\begin{aligned}
\text{(i)} \quad \alpha^{-1}\beta &= (-1, 0, -1)(1, 0, 0) = (-1, -1, -1) \\
&= (-1, 0, 0)(0, -1, -1) = \beta^{-1}\hat{\alpha}, \\
\text{(ii)} \quad \hat{\alpha}^{-1}\beta^{-1}\hat{\alpha} &= (-1, 0, 1)(-1, 0, 0)(0, -1, -1) \\
&= b(0, -1, 0) \\
&= (-1, 0, -1)(1, 0, 0)(0, 1, 1) = \alpha^{-1}\beta\alpha, \\
\text{(iii)} \quad b\alpha^2 &= b(0, 1, 1)(0, 1, 1) = b(-1, 1, 2) \\
&= (1, 0, 0)(-1, 1, 2)(1, 0, 0) = \beta\alpha^2\beta, \\
\text{(iv)} \quad \alpha^4 &= (0, 1, 1)^4 = b^2(-1, 1, 2)^2 = b(0, 0, 4) \\
&= ab.
\end{aligned} \tag{18}$$

This completes the proof that G can be taken to be G^0. Now, if $b^r = I$, then $\beta^r = I$ is consistent with $b^r = I$ according to (6), (7), and (8). Thus, we can take $G_r = G/\{\beta^r, b^r\}$, which is a group of order $8r^3$.

3. PROPERTIES OF G AND OF G_r

Let N be the subgroup of G of elements of the form $a^i b^j$, $i = 0, 1$, j integral, and let $N_r = N/\{b^r\}$. Some facts are stated below assuming that the weight group is G. Analogous results hold if G is replaced by G_r and N is replaced by N_r.

Fact 1: All closed paths in $\mathfrak{L}$ are assigned weights from N. Simple closed paths traversed in a clockwise sense have a weight ab^k, where k is the enclosed area.

Proof: Any closed path *with no backward steps can* be generated out of a path surrounding a single square by successively adding some or all of the edges of another single square (in general, with the removal of others). But relations (1) were established so that the alteration just described would change the interpretation of the "enclosed" area in just the right way so that an element of N would result for the weight of the new path if the old path had a weight which is an element of N. Of course, the second statement of the theorem follows in a similar way. To be precise, we would need to check that the two paths pictured in Fig. 7 have weights which differ by ab^2, and even *more importantly* to show that a list of eight conditions (plus another like that implied by Fig. 7) analogous to (1) with α replacing $\hat{\alpha}$, $\hat{\alpha}$ replacing α (and b^{-1} replacing b) are valid. These latter relations would arise out of considering counterclockwise motions around simple closed paths just as we consider clockwise motions in Figs. 3 and 4 to establish (1). The fact that b is replaced by b^{-1} in these counterclockwise relations follows from the fact that all of the desired relations are implied by the assumed (i)–(iv). For example, we can show that from (i) we have $\beta\alpha^{-1} = \hat{\alpha}\beta^{-1}$, and premultiplying $b\alpha^2 = \beta\alpha^2\beta$ by this we get $b\beta\alpha = \hat{\alpha}\alpha^2\beta$. This latter equality is the counterpart of (f) in (1). The other counterparts can be obtained by more or less the same argument.

By means of a rather messy argument, we can now remove the condition that the closed path have no backward steps. Basically, we need to show that Fact 1 is valid for closed paths which contain no area. More simply even, it reduces to showing the following. Let P be a path with no backward steps assumed to be not closed, and let Q be a path formed from P by tracing out P, rotating 180°, tracing out P in reverse, and rotating 180° again. Rotating through 180° involves two forward and one backward steps and can be carried out in 6 different ways. We must show that Q has a weight from N. Having shown this, we can then replace the "backward segments" of a closed path by the same segment traced in reverse (forward sense) and alter the weight only by a factor from N. By an inductive argument, we can reduce the problem further to the case where P contains only one step. Even so there are 36 cases to consider. In Fig. 8 we have illustrated one case. The weight assigned to the path 1–2–3–4–5–6–1 is $\beta\hat{\alpha}\alpha^{-1}\beta\alpha\hat{\alpha}^{-1}$, which according to (ii) of (2) is the identity. The other 35 cases as well as the filling in of details is left to the reader.

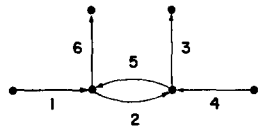


FIG. 8. A path of one step, traced out in both directions with two rotations of 180°.

For the next fact, we start with the assumption that directions are labeled as in Fig. 9, and that our path begins at (0, 0) in direction 0 with the terminal point of the directed edge at (0, 0). We wish to show the following.

Fact 2: All paths which start at a given directed edge and end at another given directed edge have weights which are elements of the same coset of N in G.

We remark that it is not really significant that the path begins in direction state 0, since our weights are not direction oriented, so we stated Fact 2 quite generally. Fact 2 sheds some light on the definition of (k, m, s) given in (4) since (k, m, s) is defined there to be the weight of one of the more convenient paths from one given directed edge to another.

Proof: For the proof of the Fact 2, let $A \sim B$, where A and B are elements from G, meaning that $A = a^i b^j B$ for some i, j. We know from Fact 1 that the weight w of a closed path which is in $\mathfrak{L}$ satisfies $w \sim I$. Let P and Q be two paths starting and ending with the

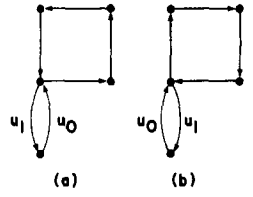


FIG. 7. Two pairs of path segments which differ only in their motion around the central square.

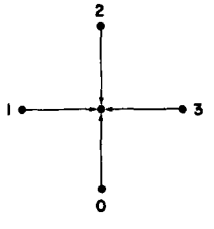


FIG. 9. The labelings of directions at a point.

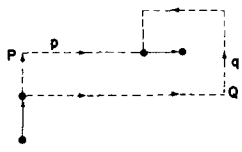


FIG. 10. Two paths starting and ending with the same directed edges.

same directed edges with weights p and q, respectively (see Fig. 10). If we reverse the path P, we get a new path whose weight is denoted by $\hat{p}$. Similarly, the reverse of the path Q has a weight denoted by $\hat{q}$. We take into consideration two closed paths. First, we take P, then rotate clockwise 180° (a three step procedure involving a right turn, a step back, and a right turn), then reverse the path P, and finally rotate clockwise 180°. For the second path, we take Q, rotate clockwise 180°, reverse the path P, and finally rotate clockwise 180°. The weights of these two closed paths are

$$p \cdot \alpha\beta^{-1}\alpha \cdot \hat{p} \cdot \alpha\beta^{-1}\alpha \sim I,$$
$$q \cdot \alpha\beta^{-1}\alpha \cdot \hat{p} \cdot \alpha\beta^{-1}\alpha \sim I. \tag{19}$$

The fact that $p \sim q$ is immediate from the fact that $\sim$ is an equivalence relation. This proves the fact.

Fact 3: Every element of G_r/N_r can be identified with a class of paths starting at (0, 0) with direction state 0 and ending at (m, k), $0 \leq m,\ k \leq r-1$, in direction state s, $0 \leq s \leq 3$.

We now turn our attention to some explicit cases.

Examples:

$n = 1$. In this case $\beta = I$, $b = I$ and, according to (2), $\hat{\alpha} = \alpha^{-1}$ and $\alpha^4 = a$. In this case G_1 is a cyclic group of order 8. This is the weight group which is useful in the combinatorial method for evaluating the partition function of the two-dimensional Ising model with zero external magnetic field.[2] G_1 assigns the "index" ± 1 to any closed path.

$n = 2$. In this case $\beta^2 = b^2 = I$, and the group G_2/N_2 is according to Fact 3 a group of order 16. G_2 is a group of order 64 whose defining relations are

$$\beta^2 = I = \alpha\beta\hat{\alpha}\beta, \quad \alpha^2\hat{\alpha}^2 = b, \quad \alpha\hat{\alpha} = \hat{\alpha}\alpha, \quad \alpha^4 = ab. \tag{20}$$

The author[1] used this group to evaluate the partition function for the Ising model in two dimensions with a particular external magnetic field. The method involves finding via G_2 the enclosed area (mod 2) of a closed path. There is figure-eight cancellation in this case, typified by weights assigned to paths in Fig. 2(a) and 2(b) which differ only by a factor of a.

$n = 3$. In this case G_3/N_3 has 36 elements associated with the four directions at each of the nine points in the diagram of Fig. 5. These elements have the form

$$\beta^k\alpha\beta^{-1}\beta^m\beta\alpha^{-1}(\alpha\beta^{-1})^s, \quad 0 \leq k, \quad m \leq 2, \quad 0 \leq s \leq 3.$$

The group G_3 is of order 216. As yet we have not been able to connect this group with the two-dimensional Ising model. In any case, use of the weights in G_3 would lead to a "modified" enclosed area (mod 3) for any closed path. Reversing the direction of the path would give the inverse weighting and hence the negative area. Figure-eight cancellation, typified by G_3 assigning weights to the paths in Fig. 2(a) and 2(b) which differ only by a factor of a, is *absent* in this case. Path 2(a) is assigned weight ab^2, while Path 2(b) is assigned weight I. Unfortunately, the combinatorial approach to the Ising problem is based in part on figure-eight cancellation. There are also other difficulties encountered in trying to use the weights in G_3 in a combinatorial approach to the Ising problem. These difficulties are discussed in the next section.

We remark in passing that we have been able to use the above ideas in evaluating the partition function for the Kagomé lattice for a particular external magnetic field. In the computation one uses the weight group G_4 for the triangular lattice. Explicit results will be presented elsewhere.

4. STRONG COMBINATORIAL IDENTITIES AND THE ISING MODEL

Beginning with a determinantal identity by Kac and Ward,[2] the combinatorial approach to the Ising problem has been based on what we call a "strong" combinatorial identity. The best way to describe the situation is to repeat the identity of Sherman.[3] A graph K of n vertices is given in the plane with no intersecting edges in which every vertex has an even number of edges touching it. Loops are allowed, and they are counted twice at the vertex in question in calculating the number of touching edges. To each edge i there is assigned an indeterminant d_i particular to that one given edge i. An admissible subgraph K_0 of K is a subgraph of K which also has an even number of edges touching each vertex. To each admissible subgraph, we assign a product χ_{K_0} of the indeterminants d_i associated with the edges in K_0. A closed path (p) in the graph K is a sequence of directed connected edges touching terminal to initial in which no two successive edges are the same and such that the last edge touches the first edge terminal to initial. We do *not* allow backing up. That is, at each vertex the path continues in the direction of the directed edge

[2] M. Kac and J. C. Ward, Phys. Rev. **88**, 1332 (1952).

[3] S. Sherman, J. Math. Phys. **1**, 202 (1960).

going out of the vertex along another edge. The other edge must exist by the fact that the graph K is admissible. To each closed path (p) we assign a weight

$$W(p) = (-1)^{k(p)} \prod_{s=1}^{n(p)} d_{i_s},$$

where $2\pi k(p)$ is the change in the argument of the tangent vector to the path (p), where $n(p)$ is the number of edges in the path (p), and where the d_i are the indeterminants associated with the successive edges of the path (p). The Feynman–Sherman identity states that

$$\sum_{K_0 \subset K} \chi_{K_0} = \exp\left(-\frac{1}{2}\sum_{(p)} \frac{W(p)}{n(p)}\right), \qquad (21)$$

where the summation on the right extends over all closed paths (p) in K and where the summation on the left extends over all admissible subgraphs of K. When $L = \mathfrak{L}$ and $d_i = x$ (independent of i), the left-hand side of (21) can be identified with a constant multiple of the partition function for the Ising model with zero external magnetic field.

We note how "strong" must be the combinatorial cancellation in this formula, since we have assigned a *different* indeterminant to each edge. Cancellation on the right can occur only among paths which in total have the same edges. Thus, for example, if we are dealing with the lattice of the previous sections, then paths of the type pictured in Fig. 11, which appear on the right of (21) but not on the left, must cancel out on the right. Note that the paths in Fig. 11 cannot be subdivided into more basic closed paths. The *only* possibility for cancellation is that the four paths in Fig. 11 have weights which cancel in (21). This is actually the case as is easily seen, since $1/2\pi$ times the change in the argument of the tangent vector $k(p)$ is 1, 0, 1, 0, respectively, for all the paths (p) in Fig. 11.

For computing the partition function in the case of nonzero external magnetic field what is needed is an analog of (21) which gives on the left the enclosed "area" of the graph K_0 in addition to the weighting χ_{K_0}. If there were an analog of (21) for the lattice $\mathfrak{L}$ using the weight group of the previous sections, i.e., a strong combinatorial result, then once again the cancellation of the paths in Fig. 11 would be required. Now, according to the weight group G,the paths in (a), (b), (c), and (d) are assigned weights χ_{d_i} times $-b^2$, I, $-b^{-2}$, I, respectively. Only the cases $b = I$ and $b^2 = I$ give the desired "strong" cancellation. One might ask, why not change the group G so that the desired cancellation takes place? But a simple closed path in $\mathfrak{L}$ with clockwise motion would have to be assigned the weight $(-1)b^{\text{Area}}\Pi d_i$ in any case, and only this was used to construct G. Thus, we cannot find in general an analog of (21) giving strong cancellation and providing us with a formula for computing the partition function of the two-dimension Ising model with nonzero external magnetic field.

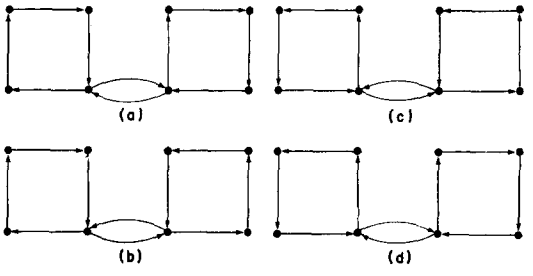


FIG. 11. Four closed paths enclosing the same two lattice squares.

Having disproved the strong combinatorial nature of the general problem, we may now inquire as to how to proceed. Certainly functional-analytical techniques will be required. Assuming that the weight groups constructed in this paper play a role in the solution we can point out one additional feature. Recall that the evaluation of the partition function for the lattice $\mathfrak{L}$ with nonzero external magnetic field can, for low temperatures, be reduced to the evaluation of the generating function

$$Z(\lambda) = \sum_{n,m} g(n; m) x^n \lambda^m$$

for the number $g(n; m)$ of closed graphs in $\mathfrak{L}$ of n sides and having m units of enclosed area. The enclosed area of a closed graph is defined to be the minimal area contained inside a set of simple closed paths whose union is the graph and no two of which have a side in common. The problem is to evaluate $Z(\lambda)$ for λ on the unit circle. If the weight groups G constructed here play a role in the evaluation, then one has to evaluate a function which is symmetric in λ and $\bar{\lambda}$. This is because the group G distinguishes between areas circumscribed clockwise and those circumscribed counterclockwise, assigning numerically equal areas with opposite sign to the two senses of traversal. Thus, λ^m and $\lambda^{-m} = \overline{\lambda^m}$ should appear symmetrically in the problem. We suspect that one should try to evaluate $|Z(\lambda)|^2$. Note that this is exactly what Kac and Ward evaluated by their method in the case that $\lambda = 1$. Also, the Feynman–Sherman identity can be rewritten so as to give $Z(1)^2$ simply by removing the factor $\frac{1}{2}$ in the exponent on the right-hand side. Of course, if one can evaluate $|Z(\lambda)|^2$, then $Z(\lambda)$ can be determined from a Wiener–Hopf factorization.

ACKNOWLEDGMENT

This research was supported in part by the United States Air Force under Grant No. AFOSR 955-65.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

New Characterization of the Ray Representations of the Galilei Group

HERBERT J. BERNSTEIN
University of California, San Diego, La Jolla, California

(Received 29 April 1966)

An elementary derivation of the system of phase factors defining the ray representations (those faithful to within a phase factor) of the Galilei group is presented. The proof employs the group elements themselves. In particular, the operation of conjugation (which corresponds to coordinate transformation) is used extensively to effect the desired result, i.e., in the notation of Levy-Leblond,

$$U(b', \mathbf{a}', \mathbf{v}', R')U(b, \mathbf{a}, \mathbf{v}, R) = \pm\exp[i(\tfrac{1}{2}m)(\mathbf{a}' \cdot R'\mathbf{v} - \mathbf{v}' \cdot R'\mathbf{a} + b\mathbf{v}' \cdot R'\mathbf{v})] \times U(b' + b, \mathbf{a}' + R'\mathbf{a} + b\mathbf{v}', \mathbf{v}' + R'\mathbf{v}, R'R).$$

IN a classic mathematical paper dealing with arbitrary Lie groups, Bargmann[1] has shown that the ray representations of the Galilei group are not all equivalent to vector representations. In addition, only those representations which are *not* equivalent to vector representations are capable of physical interpretation.[2,3] Furthermore, a superselection rule occurs, requiring a well-defined mass for states in nonrelativistic quantum mechanics.[4] This paper[4] presents an elementary derivation of Bargmann's result. Hopefully, the interested physicist will find it easier to follow than the previous works since the use of the group elements themselves (instead of their infinitesimal generators) makes the physical content more apparent. This approach was originally used by Wigner in his work on the Poincaré group.[5] In the interest of brevity and simplicity, mathematical rigor has occasionally been sacrificed, particularly in the continuity arguments. The interested reader may easily supply additional rigor, by arguments similar to those presented in Wigner's paper.[5]

To state the desired result we adopt the notation of Levy-Leblond.[3] An element of the Galilei group is denoted

$$G = (b, \mathbf{a}, \mathbf{v}, R), \tag{1}$$

where b is the time displacement, $\mathbf{a}$ the space translation, $\mathbf{v}$ the pure Galilean transformation, and R the rotation. The element G transforms $\mathbf{x}$ and t (a point in space and a time) into $\mathbf{x}' = R\mathbf{x} + \mathbf{v}t + \mathbf{a}$ and $t' = t + b$. The ray representations obey

$$U(G')U(G) = \omega(G', G)U(G'G) = \exp[i\xi(G', G)]U(G'G), \tag{2}$$

where $U(G)$ is the unitary operator representing G; $\omega(G', G)$ is the phase factor in question; and $\xi(G', G)$ is a real, continuous function of the parameters specifying G and G'. We show that for any real number m, a representation of the covering group of the Galilei group exists such that

$$\xi(G, G') = (\tfrac{1}{2}m)(\mathbf{a}' \cdot R'\mathbf{v} - \mathbf{v}' \cdot R'\mathbf{a} + b\mathbf{v}' \cdot R'\mathbf{v}). \tag{3}$$

[The covering group is simply connected, two-to-one homomorphic to the Galilei group, and is obtained by replacing the rotation subgroup with $SU(2)$.] Except for a possible additional sign ambiguity in $\omega(G', G)$, the result for the Galilei group is the same as that for the covering group.

Although the desired result [Eq. (3)] follows directly from the laws of multiplication of the "vector" operators (accelerations and translations), our proof requires that we first know their transformation properties under rotation. Any standard work on group theory demonstrates that every ray representation of $SU(2)$ is equivalent to a vector representation.[1,6] Thus, if R and Q are pure rotations [or, more accurately, pure $SU(2)$ transformations in the covering group], no phase factor is required in the equation

$$U(R)U(Q) = U(RQ). \tag{4}$$

The operators of translation must now be adjoined to this subgroup.

Let us choose translations in one direction, for example, that of the $\hat{z}$ axis. Rotations about the $\hat{z}$ axis leave these translations invariant. But multiplication in a ray representation could introduce a phase factor ω so we must write

$$U(\alpha, \hat{z})U(a\hat{z})U^{-1}(\alpha, \hat{z}) = \omega(\alpha, a)U(a\hat{z}), \tag{5}$$

where $U(\alpha, \hat{z})$ represents a rotation about $\hat{z}$ through an angle α and $U(a\hat{z})$ a translation by $a\hat{z}$. Conjugating

[1] V. Bargmann, Ann. Math. **59**, 1 (1954).
[2] E. Inönu and E. P. Wigner, Nuovo Cimento **9**, 705 (1952); M. Hammermesh, Ann. Phys. (N.Y.) **9**, 518 (1960).
[3] J.-M. Levy-Leblond, J. Math. Phys. **4**, 776 (1963).
[4] This is the "Bargmann superselection rule" [A. S. Wightman, in *Les Houches 1960 Summer School Proceedings* (Hermann & Cie., Paris, 1960), pp. 159–226].
[5] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[6] See, for example, E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959), Chap. 14.

Eq. (5) with a rotation $(\beta, \hat{z})$ through β rad about the $\hat{z}$ axis, and using Eq. (4), we have

$$\omega(\alpha, a)\omega(\beta, a) = \omega(\alpha + \beta, a). \tag{6}$$

The solution of (6) is $\omega(\theta, a) = \exp[i\chi(a)\theta]$, where χ is a function of a, the length of the translation vector. The representative of a rotation through 2π rad (about *any* axis) is $\pm I$, so $\chi(a)$ must be an integer for all values of a.

Since translations along the $\hat{z}$ axis are a one-parameter subgroup, any representation is equivalent to a vector representation.[1] We therefore have

$$U(\alpha, \hat{z})U(a\hat{z})U^{-1}(\alpha, \hat{z})$$
$$= U(\alpha, \hat{z})U(\tfrac{1}{2}a\hat{z})U^{-1}(\alpha, \hat{z})U(\alpha, \hat{z})U(\tfrac{1}{2}a\hat{z})U^{-1}(\alpha, \hat{z}), \tag{7}$$

which implies $\chi(\frac{1}{2}a) + \chi(\frac{1}{2}a) = \chi(a)$. But $\chi(\frac{1}{2}a)$ must be an integer as well as $\chi(a)$; therefore, $\chi(a)$ must be even. The argument may be repeated any number of times; since the only number which remains integral when repeatedly halved is zero, we conclude that $\chi(a) = 0$, or $\omega(\alpha, a) = 1$.

With this result in hand, all translations are unambiguously defined by conjugating the operators of $\hat{z}$ axis translations with the appropriate rotation. Thus, we define

$$U(aR\hat{z}) \equiv U(R)U(a\hat{z})U^{-1}(R), \tag{8}$$

where R is any rotation. If there is another rotation Q such that $R\hat{z} = Q\hat{z}$, then $Q^{-1}R\hat{z} = \hat{z}$ and Eq. (5), with $\omega = 1$, gives

$$U(Q^{-1}R)U(a\hat{z})U^{-1}(Q^{-1}R) = U(a\hat{z}).$$

Repeated use of Eq. (4) shows

$$U(R)U(a\hat{z})U^{-1}(R) = U(Q)U(a\hat{z})U^{-1}(Q). \tag{9}$$

The operator $U(aR\hat{z})$ is therefore well defined. Furthermore, the operators $U(aR\hat{z})$ form a vector representation of the one-parameter subgroup of translations along $R\hat{z}$, since the translations along $\hat{z}$ form such a representation.

Obviously, everything that has been said of translations applies equally well to the other subgroup of "vector" group elements, the pure Galilean transformations. In other words,

$$U(R)U(\mathbf{a})U^{-1}(R) = U(R\mathbf{a}), \quad U(R)U(\mathbf{v})U^{-1}(R) = U(R\mathbf{v}), \tag{10}$$

for any rotation R and any translation $(\mathbf{a})$ or acceleration $(\mathbf{v})$.

We have arrived at the crux of the derivation, determining the phase factors for the products of translation and acceleration operators. It is convenient to deal with group elements specified by perpendicular vectors first. Let

$$U(\mathbf{a})U(\mathbf{v})U^{-1}(\mathbf{a}) = \exp[i\phi_\perp(\mathbf{a}, \mathbf{v})]U(\mathbf{v}), \tag{11}$$

where $\mathbf{a}$ and $\mathbf{v}$ are mutually perpendicular and each may represent either a pure Galilean transformation *or* a spatial translation. Conjugation of this equation by $U(R)$, where R is a rotation about $\mathbf{a}$ through π rad, gives

$$U(\mathbf{a})U(-\mathbf{v})U^{-1}(\mathbf{a}) = \exp[i\phi_\perp(\mathbf{a}, \mathbf{v})]U(-\mathbf{v}), \tag{12}$$

by Eq. (10). Multiplying Eqs. (12) and (11), we find

$$I = \exp[2i\phi_\perp(\mathbf{a}, \mathbf{v})].$$

Of the two roots of this equation, $\phi_\perp(\mathbf{a}, \mathbf{v}) = \pi$ is ruled out by the continuity of $\phi_\perp$ in $\mathbf{a}$ and the value $\phi_\perp(\mathbf{0}, \mathbf{v}) = 0$. Thus $\phi_\perp = 0$, and any two operators which are specified by perpendicular vectors commute.

Two operators specified by parallel vectors of the same type also commute, for they are members of a one-parameter subgroup. On the other hand, a phase factor might occur in the mixed product; indeed the determination of this phase factor is the essence of the proof. Let

$$U(\mathbf{a})U(\mathbf{v})U(-\mathbf{a}) = \exp[i\phi_\parallel(\mathbf{a}, \mathbf{v})]U(\mathbf{v}), \tag{13}$$

where $\mathbf{a}$ and $\mathbf{v}$ are now parallel; $\mathbf{a}$ is a translation and $\mathbf{v}$ an acceleration. Clearly, we may "rotate" this equation to the $\hat{z}$ axis by conjugation with the appropriate rotation operator. Then

$$U(a\hat{z})U(v\hat{z})U(-a\hat{z}) = \exp[i\phi_\parallel(\mathbf{a}, \mathbf{v})]U(v\hat{z}), \tag{14}$$

where a and v are the lengths of the corresponding vectors. Therefore $\phi_\parallel$ is a function of the lengths of $\mathbf{a}$ and $\mathbf{v}$ only; $\phi_\parallel(\mathbf{a}, \mathbf{v}) = f(a, v)$. Substituting f for $\phi_\parallel$ and conjugating Eq. (14) with a translation along the $\hat{z}$ axis, $U(b\hat{z})$, yields (after equating exponents)

$$f(a, v) + f(b, v) = f[(a + b), v]. \tag{15}$$

The solution of Eq. (15) is $f(a, v) = k(v)a$, where $k(v)$ is a real function of v.

By a similar argument $k(v)$ must also be linear. Multiplying Eq. (14) on the right by $U(a\hat{z})U(-v\hat{z})$ and conjugating the resulting equation with the acceleration $U(u\hat{z})$ and substituting $k(v)a$ for $f(v, a)$ results in

$$k(v)a + k(u)a = k(v + u)a. \tag{16}$$

The solution of Eq. (16) is $k(v) = mv$, where m is any real number. Thus Eq. (13) becomes

$$U(a\hat{n})U(v\hat{n}) = \exp[im(a \cdot v)]U(v\hat{n})U(a\hat{n}) \tag{17}$$

for any unit vector $\hat{n}$.

The general mixed product may now be constructed by analyzing the multiplicand operator into perpendicular and parallel components. The result is the desired one,

$$U(\mathbf{a})U(\mathbf{v}) = \exp[im(\mathbf{a} \cdot \mathbf{v})]U(\mathbf{v})U(\mathbf{a}). \tag{18}$$

To arrive at the complete result, Eq. (3), we must

introduce the time displacement operators. It is clear that a vector representation can be chosen so that if (b) and (b') are elements of the time displacement subgroup, then

$$U(b)U(b') = U(b + b'), \tag{19}$$

because this is a one-parameter group. The conjugation of a time displacement by a rotation must now be shown to be devoid of a phase factor.

Let

$$U(\alpha, \hat{n})U(b)U^{-1}(\alpha, \hat{n}) = \exp[i\phi(\alpha, \hat{n}, b)] \cdot U(b), \tag{20}$$

where $(\alpha, \hat{n})$ stands for a rotation through angle α about the unit vector $\hat{n}$. Conjugation of this equation by a rotation through β about $\hat{n}$ leads to

$$\phi(\alpha, \hat{n}, b) + \phi(\beta, \hat{n}, b) = \phi(\alpha + \beta, \hat{n}, b).$$

Again the solution is linear,

$$\phi(\alpha, \hat{n}, b) = g(\hat{n}, b)\alpha. \tag{21}$$

Using this substitution in Eq. (20) and multiplying by a similar equation in which b' has replaced b, gives (after equating exponents)

$$g(\hat{n}, b)\alpha + g(\hat{n}, b')\alpha = g(\hat{n}, b + b')\alpha. \tag{22}$$

Thus $g(\hat{n}, b) = h(\hat{n})b$, and $\phi(\alpha, \hat{n}, b) = h(\hat{n})b\alpha$. But, as noted above, $U(\hat{n}, 2\pi) = \pm I$, so $\phi(\hat{n}, 2\pi, b) = 2\pi k$, where k is an integer. Again, this forces $h(\hat{n}) = 0$, so

$$U(R)U(b)U^{-1}(R) = U(b) \tag{23}$$

for all time displacements b and rotations R.

Using the trick of conjugation by rotation (through π rad about an axis perpendicular to $\mathbf{a}$) on the equation

$$U(\mathbf{a})U(b)U(-\mathbf{a}) = \exp[i\phi(\mathbf{a}, b)]U(b), \tag{24}$$

where $\mathbf{a}$ is any translation, produces

$$U(-\mathbf{a})U(b)U(\mathbf{a}) = \exp[i\phi(\mathbf{a}, b)]U(b). \tag{25}$$

Multiplying Eq. (24) on the right and Eq. (25) on the left by $U(\mathbf{a})$ yields

$$\begin{aligned} U(\mathbf{a})U(b) &= \exp[i\phi(\mathbf{a}, b)]U(b)U(\mathbf{a}), \\ U(b)U(\mathbf{a}) &= \exp[i\phi(\mathbf{a}, b)]U(\mathbf{a})U(b). \end{aligned} \tag{26}$$

Comparison of the above equations shows that $\exp[i\phi] = \exp[-i\phi]$, and the continuity of ϕ allows only the solution $\phi(\mathbf{a}, b) = 0$. This permits us to define a representative for the combined operations of space and time translations,

$$U(\mathbf{a})U(b) = U(b)U(\mathbf{a}) \equiv U(G), \tag{27}$$

where $G = (b, \mathbf{a}, \mathbf{0}, I)$.

For the conjugation of a time displacement by an acceleration, the group law gives

$$\begin{aligned} U(\mathbf{v})U(b)U(-\mathbf{v}) &= \exp[i\phi(\mathbf{v}, b)]U(b, \mathbf{v}b, \mathbf{0}, I) \\ &= \exp[i\phi(\mathbf{v}, b)]U(b)U(\mathbf{v}b), \end{aligned} \tag{28}$$

where $U(\mathbf{v})$ is again an acceleration, $U(\vec{\mathbf{v}}b)$ a translation, and $U(b, \mathbf{v}b, \mathbf{0}, I)$ is a combined time and space translation. Rotating through π rad, this time about an axis perpendicular to $\mathbf{v}$, we have

$$U(-\mathbf{v})U(b)U(\mathbf{v}) = \exp[i\phi(\mathbf{v}, b)]U(b)U(-\mathbf{v}b). \tag{29}$$

Sufficient manipulation of the last two equations produces

$$\exp[2i\phi(\mathbf{v}, b)]U(\mathbf{v})U(b) = U(-\mathbf{v}b)U(\mathbf{v})U(\mathbf{v}b)U(b). \tag{30}$$

Here the connection with the previous results [Eq. (18)] becomes apparent. Substituting into the right-hand side of Eq. (30), we have

$$\exp[2i\phi(\mathbf{v}, b)]U(\mathbf{v})U(b) = \exp[im(-\mathbf{v}b \cdot \mathbf{v})]U(\mathbf{v})U(b). \tag{31}$$

Again continuity allows only one solution for ϕ,

$$\begin{aligned} \phi(\mathbf{v}, b) &= (\tfrac{1}{2}m)(-\mathbf{v}b \cdot \mathbf{v}), \\ U(\mathbf{v})U(b) &= \exp[-i(\tfrac{1}{2}m)(\mathbf{v}b \cdot \mathbf{v})]U(b)U(b\mathbf{v})U(\mathbf{v}). \end{aligned} \tag{32}$$

Together with Eq. (18), this suggests defining the operator $U(\mathbf{a}, \mathbf{v})$, which corresponds to the group element $(\mathbf{0}, \mathbf{a}, \mathbf{v}, I)$, as

$$\begin{aligned} U(\mathbf{a}, \mathbf{v}) &\equiv \exp[-i(\tfrac{1}{2}m)(\mathbf{a} \cdot \mathbf{v})]U(\mathbf{a})U(\mathbf{v}) \\ &= \exp[i(\tfrac{1}{2}m)(\mathbf{a} \cdot \mathbf{v})]U(\mathbf{v})U(\mathbf{a}). \end{aligned} \tag{33}$$

Thus, Eq. (31) takes the simple form

$$U(\mathbf{v})U(b) = U(b)U(b\mathbf{v}, \mathbf{v}). \tag{34}$$

The final result is now at hand. Defining

$$U(b, \mathbf{a}, \mathbf{v}, R) \equiv U(b)U(\mathbf{a}, \mathbf{v})U(R), \tag{35}$$

we may easily check that the desired limits of

$$\begin{aligned} U(b, 0, 0, I) &= U(b), \quad U(0, \mathbf{a}, 0, I) = U(\mathbf{a}), \\ U(0, 0, \mathbf{v}, I) &= U(\mathbf{v}), \quad U(0, 0, 0, R) = U(R), \end{aligned} \tag{36}$$

all occur. Thus, the definition is consistent, and all the multiplications occurring in the definition are free from phase factors. The product of two operators, $U(b', \mathbf{a}', \mathbf{v}', R')$ and $U(b, \mathbf{a}, \mathbf{v}, R)$, may be easily shown to be

$$\begin{aligned} &U(b', \mathbf{a}', \mathbf{v}', R')U(b, \mathbf{a}, \mathbf{v}, R) \\ &= \exp[i(\tfrac{1}{2}m)(\mathbf{a}' \cdot R'\mathbf{v} - \mathbf{v}' \cdot R'\mathbf{a} + b\mathbf{v}' \cdot R'\mathbf{v})] \\ &\quad \times U(b + b', \mathbf{a}' + \mathbf{v}'b + R'\mathbf{a}, R'\mathbf{v} + \mathbf{v}', R'R). \end{aligned} \tag{37}$$

ACKNOWLEDGMENTS

The author would like to thank Professor F. R. Halpern for his suggestion of the problem, encouragement and discussions during its execution, and careful reading of the final manuscript.

This work was supported in part by the United States Atomic Energy Commission.

Time-Dependent Green's Function for Electromagnetic Radiation in a Conducting Moving Medium: Nonrelativistic Approximation

IOANNIS M. BESIERIS†
Case Institute of Technology, Cleveland, Ohio

(Received 10 June 1966)

The electromagnetic radiation emanating from a source immersed in a linear, homogeneous, conducting medium moving with a uniform velocity v with respect to the rest frame of the source distribution is investigated. It is shown that in the nonrelativistic limit, that is for v/c_v and $v/c_m \ll 1$,where c_v is the velocity of light in free space and c_m denotes the phase velocity of a wave in the medium considered at rest, the electromagnetic field intensities can be expressed in terms of a pair of scalar and vector potentials by specifying a modified Lorentz condition. The time-dependent Green's function associated with the hyperbolic partial differential equations satisfied by these potentials is determined explicitly.

I. INTRODUCTION

IN view of its relevance to astrophysical and spatial studies, there has recently been a renewal of interest in the subject of electrodynamics of moving material media. Critically reviewed and reformulated by Tai,[1,2] the fundamental work of Minkowski[3] and Sommerfeld[4] has been used to examine the problem of electromagnetic radiation in a homogeneous, isotropic moving medium in the nonrelativistic approximation case.[5,6] The same problem, but without restrictions as to the velocity of the medium, was approached by Tai[7] from an operational point of view introduced originally by Levine and Schwinger,[8] whereas Lee and Papas[9] followed a four-vector covariant formulation. More recently, Tai[10] examined the first-order theory of the electromagnetics of moving anisotropic media, and Lee and Lo[11] solved for the radiation in a moving uniaxially anisotropic medium. The aforementioned investigations are restricted to lossless media and time-harmonic variations of the source distribution. Collier and Tai,[12,13] however, have presented a brief discussion of plane-wave propagation in lossy media. Lastly, Compton[14] has determined the time-dependent Green's function for a lossless, isotropic medium.

The purpose of this paper is to examine the electromagnetic radiation resulting from sources of arbitrary time dependence in a homogeneous, isotropic, conducting medium of infinite extent. The material is assumed to be moving at a uniform velocity v with respect to the rest frame of the source distribution.

Because no solution of this problem could be found in the literature, it was thought worthwhile to find the modification of the character of the radiation due to the presence of conductivity. To avoid excessive difficulties in the ensuing development, only the nonrelativistic approximation situation is considered.

It is determined first that the electromagnetic field intensities referred to the laboratory coordinate system are expressible in terms of a pair of scalar and vector potential functions satisfying symmetric hyperbolic partial differential equations of the second order with respect to time and the space coordinates. This is made possible by invoking a generalized Helmholtz theorem, and specifying a new type of Lorentz gauge.

Ordinarily, one would solve for the time-dependent Green's function associated with the potential equations by using both time and space Fourier transformations. Instead of following this classical approach, however, we introduce an alternative method which is based on the fact that there exists a relation between the fundamental solution of a radiation problem and that of a corresponding Cauchy initial-value problem. In addition to its being ideally suited for *bona fide* initial-value problems, it is believed that this technique is "operationally" easier to apply, especially when dealing with simple, single partial differential equations, or small systems of partial

† Present Address: Bell Telephone Laboratories Inc., Whippany, N.J.

[1] C. T. Tai, Proc. IEEE, **52**, 685 (1964).

[2] C. T. Tai, University of Michigan, Radiation Laboratory Report No. RL-310 (1965).

[3] H. Minkowski, Nachr. kgl. Ges. Wiss. Göttingen **1**, 53 (1908).

[4] A. Sommerfeld, *Electrodynamics* (Academic Press Inc., New York, 1964), pp. 280–290.

[5] R. T. Compton, Jr., and C. T. Tai, Ohio State University, Antenna Laboratory Report No. 1691-3 (1964).

[6] R. T. Compton, Jr., and C. T. Tai, Trans. IEEE Antennas Propagation, **AP13**, 574 (1965).

[7] C. T. Tai, Trans. IEEE Antennas Propagation, **AP13**, 322 (1965).

[8] H. Levine and J. Schwinger, in *Symposium on the Theory of Electromagnetic Waves* (Interscience Publishers, Inc., New York, 1951), pp. 355–391.

[9] K. S. H. Lee and C. H. Papas, J. Math. Phys. **5**, 1668 (1964).

[10] C. T. Tai, Radio Sci. **69D**, 407 (1965).

[11] S. W. Lee and Y. T. Lo, Radio Sci. **1**, 313 (1966).

[12] J. R. Collier and C. T. Tai, Trans. IEEE Antennas Propagation, **AP12**, 375 (1965).

[13] J. R. Collier and C. T. Tai, Trans. IEEE Microwave Theory and Techniques, **MTT13**, 441 (1965).

[14] R. T. Compton, Jr., J. Math. Phys. **7**, 2145 (1966).

differential equations of the first order with respect to time.

II. THE MAXWELL–MINKOWSKI EQUATIONS FOR A CONDUCTING MEDIUM

Let K and K' denote two inertial systems in relative motion. We identify the primed coordinate system as being at rest with respect to a homogeneous, isotropic, conducting medium of infinite extent which moves with a uniform velocity $\mathbf{v}$ relative to the laboratory system K.

As measured by an observer in the laboratory system, the electromagnetic fields must satisfy Maxwell's equations

$$\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t, \tag{1a}$$

$$\nabla \times \mathbf{H} = \partial \mathbf{D}/\partial t + \mathbf{J}_f + \mathbf{J}, \tag{1b}$$

$$\nabla \cdot \mathbf{D} = \rho_f + \rho, \tag{1c}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{1d}$$

where $\mathbf{E}$, $\mathbf{H}$ signify, respectively, the electric and magnetic field intensities, $\mathbf{D}$, $\mathbf{B}$ the electric and magnetic displacements, ρ_f, $\mathbf{J}_f$ the free charge and current densities and, finally, ρ, $\mathbf{J}$ the externally applied charge and current distribution densities, all referred to the mks system of units.

Due to the invariance of the fundamental laws of physics in the light of the special theory of relativity, Maxwell's equations in the K' frame can be written down as follows:

$$\nabla' \times \mathbf{E}' = \partial \mathbf{B}'/\partial t', \tag{2a}$$

$$\nabla' \times \mathbf{H}' = \partial \mathbf{D}'/\partial t' + \mathbf{J}_f' + \mathbf{J}', \tag{2b}$$

$$\nabla' \cdot \mathbf{D}' = \rho_f' + \rho', \tag{2c}$$

$$\nabla' \cdot \mathbf{B}' = 0. \tag{2d}$$

It is assumed that the constitutive relations in K' are given by

$$\mathbf{D}' = \epsilon' \mathbf{E}', \tag{3a}$$

$$\mathbf{B}' = \mu' \mathbf{H}', \tag{3b}$$

$$\mathbf{J}_f' = \sigma' \mathbf{E}'. \tag{3c}$$

The electric and magnetic permittivities and the conductivity are taken to be independent of time and the space coordinates.

If $\mathbf{v}$ is small compared with the speed of light, the following first-order relativistic transformations apply:

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B}, \tag{4a}$$

$$\mathbf{D}' = \mathbf{D} + c^{-2}\mathbf{v} \times \mathbf{H}, \tag{4b}$$

$$\mathbf{H}' = \mathbf{H} - \mathbf{v} \times \mathbf{D}, \tag{4c}$$

$$\mathbf{B}' = \mathbf{B} - c^{-2}\mathbf{v} \times \mathbf{E}, \tag{4d}$$

$$\mathbf{J}_f' = \mathbf{J}_f - \rho_f \mathbf{v}, \tag{4e}$$

$$\rho_f' = \rho_f - c^{-2}\mathbf{v} \cdot \mathbf{J}_f. \tag{4f}$$

In view of the brevity of the relaxation time, it may be assumed that the free charge density ρ_f' is zero inside the conducting medium. Bearing this restriction in mind, by substituting Eqs. (4a)–(4f) into Eqs. (3a)–(3c) and neglecting terms of the order of $v/c_{v,m}$, we obtain the constitutive relationships

$$\mathbf{D} = \epsilon' \mathbf{E} + \mathbf{\Lambda} \times \mathbf{H}, \tag{5a}$$

$$\mathbf{B} = \mu' \mathbf{H} - \mathbf{\Lambda} \times \mathbf{E}, \tag{5b}$$

$$\mathbf{J}_f = \sigma'(\mathbf{E} + \mu' \mathbf{v} \times \mathbf{H}), \tag{5c}$$

$$\rho_f = c^{-2}\mathbf{v} \cdot \mathbf{J}_f, \tag{5d}$$

where $\mathbf{\Lambda} = (\mu'\epsilon' - \mu_0\epsilon_0)\mathbf{v}$, ϵ_0, μ_0 being, respectively, the electric and magnetic permittivities of free space.

Maxwell's equations in K assume now the "definite" form

$$\nabla \times \mathbf{E} = -(\partial/\partial t)(\mu' \mathbf{H} - \mathbf{\Lambda} \times \mathbf{E}), \tag{6a}$$

$$\nabla \times \mathbf{H} = (\partial/\partial t)(\epsilon' \mathbf{E} + \mathbf{\Lambda} \times \mathbf{H}) + \sigma'(\mathbf{E} + \mu' \mathbf{v} \times \mathbf{H}) + \mathbf{J}, \tag{6b}$$

$$\nabla \cdot (\epsilon' \mathbf{E} + \mathbf{\Lambda} \times \mathbf{H}) = \sigma' c^{-2} \mathbf{v} \cdot (\mathbf{E} + \mu' \mathbf{v} \times \mathbf{H}) + \rho, \tag{6c}$$

$$\nabla \cdot (\mu' \mathbf{H} - \mathbf{\Lambda} \times \mathbf{E}) = 0. \tag{6d}$$

These expressions, commonly known as the *Maxwell–Minkowski equations*, may be rearranged into the following more convenient form:

$$\mathbf{D}_0 \times \mathbf{E} = -(\partial/\partial t)\mu' \mathbf{H}, \tag{7a}$$

$$\mathbf{D}_0 \times \mathbf{H} = (\partial/\partial t)\epsilon' \mathbf{E} + \sigma' \mathbf{E} + \sigma' \mu' \mathbf{v} \times \mathbf{H} + \mathbf{J}, \tag{7b}$$

$$\mathbf{D}_0 \cdot (\epsilon' \mathbf{E}) = \epsilon' \mu' \sigma' \mathbf{v} \cdot \mathbf{E} + \rho + \mathbf{\Lambda} \cdot \mathbf{J}, \tag{7c}$$

$$\mathbf{D}_0 \cdot (\mu' \mathbf{H}) = 0, \tag{7d}$$

in which $\mathbf{D}_0$ represents the differential operator $\nabla - \mathbf{\Lambda}(\partial/\partial t)$.

III. SCALAR AND VECTOR POTENTIALS

It is shown in this section that the electromagnetic field intensities can be given in terms of appropriately defined scalar and vector potentials satisfying a generalized Lorentz condition.

In view of the identities

$$\mathbf{D}_0 \cdot \mathbf{D}_0 \times \mathbf{F} = 0, \tag{8a}$$

$$\mathbf{D}_0 \times \mathbf{D}_0 \phi = 0, \tag{8b}$$

which hold for all twice differentiable scalar and vector functions ϕ and $\mathbf{F}$, a generalized Helmholtz theorem can be formulated such that for any vector field $\mathbf{C}$, there exists a scalar field ϕ and a vector field $\mathbf{F}$, satisfying the equation[14]

$$\mathbf{C} = \mathbf{D}_0 \phi + \mathbf{D}_0 \times \mathbf{F}. \tag{9}$$

On the basis of Eq. (7d) and the above remarks, a vector potential $\mathbf{A}$ is defined by

$$\mathbf{H} = (1/\mu')\mathbf{D}_0 \times \mathbf{A}. \tag{10}$$

This result, in conjunction with Eq. (7a), suggests that

$$\mathbf{E} = -\mathbf{D}_0\psi - \partial\mathbf{A}/\partial t, \tag{11}$$

where ψ is a suitably chosen scalar potential function. Substituting these expressions for $\mathbf{E}$ and $\mathbf{H}$ into the second of the Maxwell–Minkowski equations, one finds that

$$\begin{aligned}\mathbf{D}_0^2\mathbf{A} - \mu'\epsilon'(\partial^2\mathbf{A}/\partial t^2) - \mu'\sigma'(\partial\mathbf{A}/\partial t) \\ = \mathbf{D}_0(\mathbf{D}_0\cdot\mathbf{A}) - \mu'\sigma'\mathbf{v}\times(\mathbf{D}_0\times\mathbf{A}) \\ + \mu'\epsilon'(\partial/\partial t)(\mathbf{D}_0\psi) + \mu'\sigma'\mathbf{D}_0\psi - \mu'\mathbf{J}.\end{aligned} \tag{12}$$

However, since

$$\begin{aligned}\mathbf{v}\times(\mathbf{D}_0\times\mathbf{A}) = \mathbf{D}_0(\mathbf{v}\cdot\mathbf{A}) - (\mathbf{v}\cdot\mathbf{D}_0)\mathbf{A} - (\mathbf{A}\cdot\mathbf{D}_0)\mathbf{v} \\ - \mathbf{A}\times(\mathbf{D}_0\times\mathbf{v}) \simeq \mathbf{D}_0(\mathbf{v}\cdot\mathbf{A}) - \mathbf{v}\cdot\nabla\mathbf{A}\end{aligned}$$

and

$$\mathbf{D}_0^2\mathbf{A} \simeq [\nabla^2 - 2(\partial/\partial t)\mathbf{\Lambda}\cdot\nabla]\mathbf{A}$$

to first order in $v/c_{v,m}$, it follows that

$$\begin{aligned}[\nabla^2 - 2(\partial/\partial t)\,\mathbf{\Lambda}\cdot\nabla - \mu'\epsilon'\,(\partial^2/\partial t^2) - \mu'\sigma'\,(\partial/\partial t) \\ - \mu'\sigma'\mathbf{v}\cdot\nabla]\mathbf{A} = \mathbf{D}_0[\mathbf{D}_0\cdot\mathbf{A} + \mu'\epsilon'\,(\partial\psi/\partial t) \\ + \mu'\sigma'\psi + \mu'\sigma'\mathbf{v}\cdot\mathbf{A}] - \mu'\mathbf{J}.\end{aligned} \tag{13}$$

Similarly, from Eqs. (7c) and (11), it can be shown that

$$\begin{aligned}\mathbf{D}_0^2\psi - \mu'\sigma'\mathbf{v}\cdot(\mathbf{D}_0\psi) + (\mathbf{D}_0 - \mu'\sigma'\mathbf{v})\cdot(\partial\mathbf{A}/\partial t) \\ = -(\rho + \mathbf{\Lambda}\cdot\mathbf{J})/\epsilon'\end{aligned} \tag{14}$$

or, since

$$\mathbf{D}_0^2 \simeq [\nabla^2 - 2(\partial/\partial t)\mathbf{\Lambda}\cdot\nabla]$$

and, furthermore,

$$\mathbf{v}\cdot(\mathbf{D}_0\psi) \simeq \mathbf{v}\cdot\nabla\psi$$

to order $(v/c_{v,m})^2$,

$$\begin{aligned}[\nabla^2 - 2(\partial/\partial t)\mathbf{\Lambda}\cdot\nabla - \mu'\sigma'\mathbf{v}\cdot\nabla]\psi + (\partial/\partial t) \\ \times(\mathbf{D}_0\cdot\mathbf{A} - \mu'\sigma'\mathbf{v}\cdot\mathbf{A}) = -(\rho + \mathbf{\Lambda}\cdot\mathbf{J})/\epsilon'.\end{aligned} \tag{15}$$

If $\mathbf{A}$ and ψ are chosen to satisfy the generalized Lorentz condition

$$\mathbf{D}_0\cdot\mathbf{A} + \mu'\epsilon'\partial\psi/\partial t + \mu'\sigma'\psi - \mu'\sigma'\mathbf{v}\cdot\mathbf{A} = 0, \tag{16}$$

it is seen immediately that the potential functions obey the following partial differential equations:

$$\begin{aligned}[\nabla^2 - 2(\partial/\partial t)\,\mathbf{\Lambda}\cdot\nabla - \mu'\epsilon'(\partial^2/\partial t^2) - \mu'\sigma'(\partial/\partial t) \\ - \mu'\sigma'\mathbf{v}\cdot\nabla]\mathbf{A} = -\mu'\mathbf{J},\end{aligned} \tag{17a}$$

$$\begin{aligned}[\nabla^2 - 2(\partial/\partial t)\,\mathbf{\Lambda}\cdot\nabla - \mu'\epsilon'(\partial^2/\partial t^2) - \mu'\sigma'(\partial/\partial t) \\ - \mu'\sigma'\mathbf{v}\cdot\nabla]\psi = -(1/\epsilon')(\rho + \mathbf{\Lambda}\cdot\mathbf{J}).\end{aligned} \tag{17b}$$

Let us for convenience assume that $\mathbf{v} = v\boldsymbol{\alpha}_z$. This condition does not constitute a serious restriction, since a coordinate transformation of the final result can be used to treat the more general case. To solve for the potentials under this assumption, it is customary to define the time-dependent Green's function $G(\mathbf{r}, t/\mathbf{r}', t')$ as the solution of the equation

$$\begin{aligned}[\nabla^2 - \epsilon'\mu'(\partial^2/\partial t^2) - \mu'\sigma'(\partial/\partial t) - 2\Lambda(\partial^2/\partial t\partial z) \\ - \mu'\sigma' v(\partial/\partial z)]G(\mathbf{r}, t/\mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t'), \\ t \geq t';\ \Lambda = |\mathbf{\Lambda}|.\end{aligned} \tag{18}$$

The Green's function should satisfy the causality condition; namely, $G \equiv 0$ for $t < t'$. The electromagnetic potentials are given in terms of G in the following fashion[15]:

$$\mathbf{A}(\mathbf{r}, t) = -\mu'\int_{E_4}\mathbf{J}(\mathbf{r}', t')G(\mathbf{r}, t/\mathbf{r}', t')\,d\mathbf{r}'\,dt', \tag{19a}$$

$$\begin{aligned}\psi(\mathbf{r}, t) = -\frac{1}{\epsilon'}\int_{E_4}[\rho(\mathbf{r}', t') + \mathbf{\Lambda}\cdot\mathbf{J}(\mathbf{r}', t')] \\ \times G(\mathbf{r}, t/\mathbf{r}', t')\,d\mathbf{r}'\,dt'.\end{aligned} \tag{19b}$$

It has already been established that $\mathbf{E}$ and $\mathbf{H}$ can be derived from the potentials [cf. Eqs. (10) and (11)].

IV. RELATION BETWEEN RADIATION AND INITIAL-VALUE PROBLEMS

Ordinarily, when dealing with a radiation problem, one determines the time-dependent Green's function using a combination of temporal and threefold spatial Fourier transformations and residue theory. It is shown here, however, that in certain cases it is simpler to examine first the associated Cauchy initial-value problem. This alternative technique is described briefly in the following two subsections and is illustrated in Sec. V by exhibiting a solution to the radiation problem (18).

A. The Riemann Matrix

An elementary solution to the partial differential equation

$$L\psi(\mathbf{r}, t) = f(\mathbf{r}, t) \tag{20}$$

is the time-dependent Green's function $G(\mathbf{r}, t/\mathbf{r}', t')$ satisfying the equation

$$LG(\mathbf{r}, t/\mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t') \tag{21}$$

and the causality condition, as pointed out in the previous section. On the other hand, if L is considered to be a second-order partial differential operator with respect to time, the fundamental solution of the

[15] The integration in Eq. (19b) extends over a four-dimensional Euclidean space containing the space coordinates and time.

Cauchy initial-value problem

$$L\Phi(\mathbf{r}, t) = 0, \tag{22a}$$

$$\Phi(\mathbf{r}, t)]_{t=t'} = 0, \tag{22b}$$

$$(\partial/\partial t)\Phi(\mathbf{r}, t)]_{t=t'} = g(\mathbf{r}, t), \tag{22c}$$

is defined for $t \geq t'$ by the equation

$$LH(\mathbf{r}, t/\mathbf{r}', t') = 0 \tag{23a}$$

with H satisfying the initial conditions

$$H]_{t=t'} = 0, \tag{23b}$$

$$\partial H/\partial t]_{t=t'} = \delta(\mathbf{r} - \mathbf{r}'). \tag{23c}$$

G is related to H as follows[16]:

$$G(\mathbf{r}, t/\mathbf{r}', t') = \begin{cases} 0 & , \quad t < t', \\ H(\mathbf{r}, t/\mathbf{r}', t'), & t \geq t'. \end{cases} \tag{24}$$

It is essential for the development of the new approach to the solution of the radiation problem that we formulate the associated Cauchy initial-value problem as a system of partial differential equations of the first order with respect to time. Let us therefore write

$$(\partial/\partial t)u(\mathbf{r}, t) = \mathsf{P}u(\mathbf{r}, t), \tag{25a}$$

where, in general, $u(\mathbf{r}, t)$ is the (column) matrix representation of an n-component vector function, and P is an $n \times n$ matrix whose entries are polynomials in the differential operators with respect to the spatial coordinates.[17] With this problem we associate the initial condition

$$u(\mathbf{r}, t)]_{t=t'} = u_0(\mathbf{r}, t'). \tag{25b}$$

It is our intent here to determine a solution $u(\mathbf{r}, t)$ which satisfies the initial conditions and depends continuously on the initial data for $t \geq t'$. It is assumed that the Cauchy problem is well posed so that the solution is unique and sufficiently differentiable.[18]

The n-fold spatial Fourier transforms are introduced next:

$$F \cdot u(\mathbf{r}, t) \equiv w(\mathbf{s}, t) = \int_{E_n} e^{-i\mathbf{r}\cdot\mathbf{s}} u(\mathbf{r}, t)\, d\mathbf{r}, \tag{26a}$$

$$F^{-1} \cdot w(\mathbf{s}, t) = u(\mathbf{r}, t) = \frac{1}{(2\pi)^n} \int_{E_n} e^{i\mathbf{r}\cdot\mathbf{s}} w(\mathbf{s}, t)\, d\mathbf{s}. \tag{26b}$$

The integration is over an n-dimensional Euclidean space. Taking the Fourier transform of the system we obtain

$$(\partial/\partial t)w(\mathbf{s}, t) = \mathsf{P}(\mathbf{s})w(\mathbf{s}, t). \tag{27a}$$

Note that the Fourier-transformed initial condition becomes

$$F \cdot u_0(\mathbf{r}, t') \equiv w_0(\mathbf{s}, t'). \tag{27b}$$

We remark here that the problem of the correctness of the Cauchy problem (22) is equivalent to the correctness of the problem (27). On the basis of this remark, a unique solution exists for $t \geq t'$ and can be written in the form

$$w(\mathbf{s}, t) = \mathsf{Q}(\mathbf{s}, t, t')w_0(\mathbf{s}, t'), \tag{28}$$

where

$$\mathsf{Q}(\mathbf{s}, t, t') = \exp\,[\mathsf{P}(\mathbf{s})(t - t')]. \tag{29}$$

A solution of this type is written down by analogy to the scalar case. The matrix exponential in Eq. (29) is defined by means of the infinite series

$$e^{\underline{P}t} = \mathsf{I} + \mathsf{P}t + \cdots + \frac{1}{n!}\mathsf{P}^n t^n + \cdots. \tag{30}$$

This matrix series exists for all P for any fixed value of t, and for all t for any fixed value of P.[19]

Operating with the inverse Fourier transform on both sides of Eq. (28), we arrive at

$$\begin{aligned} u(\mathbf{r}, t) &= F^{-1} \cdot w(\mathbf{s}, t) \\ &= [F^{-1} \cdot \mathsf{Q}(\mathbf{s}, t, t')] * [F^{-1} \cdot w_0(\mathbf{s}, t')] \\ &= \int_{E_n} R(\mathbf{r}, t/\mathbf{r}', t')u_0(\mathbf{r}', t')\, d\mathbf{r}'. \end{aligned} \tag{31}$$

The matrix function

$$\mathsf{R}(\mathbf{r}, t/\mathbf{r}', t') = \frac{1}{(2\pi)^n} \int_{E_n} \exp\,[i(\mathbf{r} - \mathbf{r}') \cdot \mathbf{s}\mathsf{I} + \mathsf{P}(\mathbf{s})(t - t')]\, d\mathbf{s} \tag{32}$$

is known as the *Riemann matrix* of the initial-value problem. Its connection with the scalar fundamental solution $H(\mathbf{r}, t/\mathbf{r}', t')$ introduced earlier in this section is given in Sec. V. The Riemann matrix satisfies the system

$$\partial\mathsf{R}/\partial t = \mathsf{PR} \tag{33a}$$

with the initial condition

$$\mathsf{R}]_{t=t'} = \delta(\mathbf{r} - \mathbf{r}')\mathsf{I}. \tag{33b}$$

Both of these statements are easily verifiable.

B. Sylvester's Interpolation Formula

It is clear from the preceding subsection that a significant task for the determination of the Riemann

[16] I. M. Gel'fand and G. E. Shilov, *Generalized Functions*, translated by E. Saletan (Academic Press Inc., New York, 1964), Vol. I, pp. 204–205.

[17] In general, $\mathsf{P} = \mathsf{P}(\partial/\partial x_\nu, x_\nu, t)$. In the subsequent work, however, we shall restrict the discussion to homogeneous, linear media whose characteristic parameters are independent of time.

[18] V. M. Borok, Am. Math. Soc. Transl. **2**, 285 (1957).

[19] R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill Book Company, Inc., New York, 1960), pp. 159–169.

matrix is to expand the matrix exponential

$$\mathsf{Q}(\mathbf{s}, t, t') = \exp[\mathsf{P}(\mathbf{s})(t - t')]$$

in a simple form. Towards this goal we now state a basic property of functions of matrices.

Let P be an $n \times n$ matrix which can be diagonalized by means of a similarity transformation, namely,

$$\mathsf{D} = \mathsf{SPS}^{-1}. \tag{34}$$

Furthermore, it is assumed that the eigenvalues $\lambda_j, j = 1, 2, \cdots, n$, of P are distinct. Although these are serious restrictions, the following result is sufficient for the illustration to be given in the next section.

If $f(\lambda)$ and $f(\mathsf{P})$ denote corresponding analytic scalar and matrix functions, it can be proven that

$$f(\mathsf{P}) = \sum_{j=1}^{n} f(\lambda_j)\mathsf{A}_j. \tag{35}$$

The constituent idempotent matrices A_j are defined by the expression

$$\mathsf{A}_j = \mathsf{S}^{-1}\epsilon_{jj}\mathsf{S} \tag{36}$$

in which ϵ_{jj} is obtained from an $n \times n$ null matrix by replacing the (jj) entry with 1.

One usually needs the eigenvectors of P in order to find the matrices A_j. This is because the A_j's are defined in terms of S, the matrix whose columns are the eigenvectors of P. Nevertheless, it is possible to determine the constituent idempotents directly as polynomials in P, without knowing the eigenvectors. Specifically, we have

$$\mathsf{A}_k = q_k(\mathsf{P}). \tag{37}$$

The interpolatory polynomial $q_k(\lambda)$ of degree $n - 1$ in λ is given by[20]

$$q_k(\lambda) = \prod_{\substack{j=1 \\ j \neq k}}^{n} \frac{\lambda - \lambda_j}{\lambda_k - \lambda_j}. \tag{38}$$

Finally, by virtue of Eqs. (35) and (37), we obtain *Sylvester's interpolation formula*

$$f(\mathsf{P}) = \sum_{k=1}^{n} f(\lambda_k)q_k(\mathsf{P}). \tag{39}$$

In particular, we write

$$\mathsf{Q}(\mathbf{s}, t, t') = \sum_{k=1}^{n} e^{\lambda_k}\left[\prod_{\substack{j=1 \\ j \neq k}}^{n} \frac{\mathsf{P}(\mathbf{s})(t - t') - \lambda_j \mathsf{I}}{\lambda_k - \lambda_j}\right]. \tag{40}$$

If the characteristic roots of P are not distinct, or P cannot be diagonalized by a collineatory transformation, the above result no longer applies. For a complete discussion of the general case the reader should consult Refs. 18 and 21.

V. THE SOLUTION FOR THE TIME-DEPENDENT GREEN'S FUNCTION

Consider the Cauchy problem

$$[\nabla^2 - \epsilon'\mu'(\partial^2/\partial t^2) - \mu'\sigma'(\partial/\partial t) - 2\Lambda(\partial^2/\partial z \partial t) - \mu'\sigma'(\partial/\partial z)]\psi(\mathbf{r}, t) = 0, \tag{41a}$$

$$\psi(\mathbf{r}, t)]_{t=t'} = 0, \tag{41b}$$

$$(\partial/\partial t)\psi(\mathbf{r}, t)]_{t=t'} = g(\mathbf{r}, t'). \tag{41c}$$

Its solution can be simplified considerably by the substitution

$$\Phi(\mathbf{r}, t) = \psi(\mathbf{r}, t) \exp(\alpha z + \beta t). \tag{42}$$

If the scalars α, β are chosen so that

$$\alpha = -\tfrac{1}{2}(\mu'\sigma' v) + \frac{\mu'\sigma'\Lambda(1 + v\Lambda)}{2(\mu'\epsilon' + \Lambda^2)}$$

and

$$\beta = \frac{\mu'\sigma'(1 + v\Lambda)}{2(\mu'\epsilon' + \Lambda^2)},$$

$\Phi(\mathbf{r}, t)$ satisfies the "semicanonical" equation[22]

$$[\nabla^2 - \epsilon'\mu'(\partial^2/\partial t^2) - 2\Lambda(\partial^2/\partial t \partial z) + q^2]\Phi(\mathbf{r}, t) = 0 \tag{43a}$$

with the modified initial conditions

$$\Phi(\mathbf{r}, t)]_{t=t'} = 0, \tag{43b}$$

$$(\partial/\partial t)\Phi(\mathbf{r}, t)]_{t=t'} = g(\mathbf{r}, t') \exp(\alpha z + \beta t') \equiv h(\mathbf{r}, t'). \tag{43c}$$

In Eq. (43a), $q^2 \equiv \mu'\epsilon'\beta^2 + 2\Lambda\alpha\beta - \alpha^2$.

We now apply the notions and results of the previous section to the Cauchy problem for the scalar function $\Phi(\mathbf{r}, t)$. First, we convert it into a system of two first-order partial differential equations with respect to time by defining

$$u_1 = \Phi, \quad u_2 = \partial\Phi/\partial t, \quad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}. \tag{44}$$

We may therefore write

$$(\partial/\partial t)u(\mathbf{r}, t) = \mathsf{P}u(\mathbf{r}, t), \tag{45}$$

where

$$\mathsf{P} = \begin{bmatrix} 0 & 1 \\ (\nabla^2 + q^2)/\mu'\epsilon' & -(2\Lambda/\mu'\epsilon')(\partial/\partial z) \end{bmatrix}.$$

[20] J. S. Frame, IEEE Spectrum **1**, 102 (1964).

[21] F. R. Gantmacher, *The Theory of Matrices*, translated by K. A. Hirsch (Chelsea Publishing Company, New York, 1959), Vol. I, pp. 89–129.

[22] The second-order partial differential equation would be in the "canonical" form if the term involving the mixed derivatives with respect to z and t were absent. See R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1962), Vol. II, pp. 180–184.

A threefold spatial Fourier transformation yields the relation

$$(\partial/\partial t)w(\mathbf{s}, t) = \mathsf{P}(\mathbf{s})w(\mathbf{s}, t) \tag{46}$$

in which

$$\mathsf{P}(s) = \begin{bmatrix} 0 & 1 \\ (-s^2 + q^2)/\mu'\epsilon' & -(2\Lambda/\mu'\epsilon')is_z \end{bmatrix}; \qquad s^2 = s_x^2 + s_y^2 + s_z^2.$$

A solution of the characteristic equation

$$\det\,[\mathsf{P}(\mathbf{s})t - \lambda \mathsf{I}] = 0 \tag{47}$$

results in

$$\lambda_{1,2} = (-\omega_0 \pm i\omega_1)t, \tag{48}$$

where

$$\omega_0 = (\Lambda/\mu'\epsilon')is_z,$$

$$\omega_1 = (\mu'\epsilon')^{-\frac{1}{2}}[s_x^2 + s_y^2 + s_z^2(1 + \Lambda^2/\mu'\epsilon') - q^2]^{\frac{1}{2}}$$

for the eigenvalues of $\mathsf{P}(s)t$. The interpolatory polynomial is given by

$$q(\lambda) = \frac{\lambda - \lambda_2}{\lambda_1 - \lambda_2}\, e^{\lambda_1} + \frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1}\, e^{\lambda_2}$$

$$= e^{-\omega_0 t}\left(\lambda \frac{\sin \omega_1 t}{\omega_1 t} + \omega_0 \frac{\sin \omega_1 t}{\omega_1} + \cos \omega_1 t\right). \tag{49}$$

Hence, by Eq. (40),

$$\mathsf{Q}'(\mathbf{s}, t, t') = e^{-\omega_0 \tau}\left[\frac{\sin \omega_1 \tau}{\omega_1}\,\mathsf{P}(s) + \left(\omega_0 \frac{\sin \omega_1 \tau}{\omega_1} + \cos \omega_1 \tau\right)\mathsf{I}\right] \tag{50}$$

in which $\tau = t - t'$.

By reason of the definitions in Eq. (44) and the fact that we are involved with a second-order differential equation [cf. Eq. (43)] in this particular discussion, we need only be concerned with the (12) entry of Q', viz.,

$$(\mathsf{Q})_{12} = e^{-\omega_0 \tau}(\sin \omega_1 \tau/\omega_1). \tag{51}$$

The corresponding term in the Riemann matrix is found by taking the inverse Fourier transform. Thus,[23]

$$(\mathsf{R}')_{12} = F^{-1} \cdot (\mathsf{Q}')_{12}$$

$$= \frac{1}{(2\pi)^3}\int_{E_3} e^{i\mathbf{s}\cdot(\mathbf{r}-\mathbf{r}')}\left(e^{-\omega_0 \tau}\frac{\sin \omega_1 \tau}{\omega_1}\right) d\mathbf{s} \tag{52a}$$

or, more explicitly,

$$(\mathsf{R}')_{12} = \frac{1}{(2\pi)^3}\int_{E_3} e^{i\mathbf{s}\cdot\mathbf{R}} \times \left(e^{-i\Lambda u^2 s_z \tau}\frac{\sin u\tau(s_x^2 + s_y^2 + s_z^2/b^2 - q^2)^{\frac{1}{2}}}{u(s_x^2 + s_y^2 + s_z^2/b^2 - q^2)^{\frac{1}{2}}}\right) d\mathbf{s}, \tag{52b}$$

[23] Q' and R' are used in connection with the scalar function $\Phi(\mathbf{r}, t)$. The corresponding unprimed quantities are referred to $\psi(\mathbf{r}, t)$.

where the following abbreviations are used:

$$u = (\mu'\epsilon')^{-\frac{1}{2}}, \quad b = (1 + \Lambda^2/\mu'\epsilon')^{-\frac{1}{2}}, \quad \mathbf{R} = \mathbf{r} - \mathbf{r}'.$$

Let $\mathbf{s}$ and $\mathbf{R}$ undergo the following linear transformation:

$$\mathbf{s}_0 = \mathsf{A} \cdot \mathbf{s}, \tag{53a}$$

$$\mathbf{R}_0 = \mathsf{A}^{-1} \cdot \mathbf{R}, \tag{53b}$$

where the matrix representation of A is given by

$$\mathsf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/b \end{bmatrix}$$

and

$$\mathbf{s}_0 = \sum_{j=1}^{3} s_j \mathbf{a}_j, \quad \mathbf{R}_0 = \sum_{j=1}^{3} x_j \mathbf{a}_j.$$

We should recall at this stage the general equivalence relation

$$(\mathsf{R}')_{12} = \frac{1}{(2\pi)^3}\int_{E_3} e^{i\mathbf{s}\cdot\mathbf{R}} F(\mathbf{s})\, d\mathbf{s}$$

$$= \frac{1}{(2\pi)^3}\int_{E_3} e^{i\mathbf{s}_0\cdot\mathbf{R}_0} F(\mathsf{A}^{-1} \cdot \mathbf{s}_0)(\det \mathsf{A})^{-1}\, d\mathbf{s}_0, \tag{54}$$

which is easily shown to hold for the above linear transformation.[24] As a consequence of these results we have

$$(\mathsf{R}')_{12} = \frac{b}{(2\pi)^3}\int_{E_3} e^{i\mathbf{s}_0\cdot\mathbf{R}_1}\frac{\sin u\tau(s_0^2 - q^2)^{\frac{1}{2}}}{u(s_0^2 - q^2)^{\frac{1}{2}}}\, d\mathbf{s}_0. \tag{55}$$

In this equation, $\mathbf{R}_1 = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + (x_3 - \gamma\tau)\mathbf{a}_3$, $\gamma = u^2 b\Lambda$. This integral can be brought into a more manageable form by considering a spherical coordinate system and choosing the polar axis in the $\mathbf{R}_1$ direction. Then,

$$(\mathsf{R}')_{12} = \frac{b}{(2\pi)^3}\int_0^{\infty}\int_0^{\pi}\int_0^{2\pi} e^{is_0 R_1 \cos\theta} \times \frac{\sin u\tau(s_0^2 - q^2)^{\frac{1}{2}}}{u(s_0^2 - q^2)^{\frac{1}{2}}}\, s_0^2 \sin\theta\, ds_0\, d\theta\, d\phi$$

$$= \frac{2b}{(2\pi)^2}\int_0^{\infty}\frac{\sin s_0 R_1}{R_1}\frac{\sin u\tau(s_0^2 - q^2)^{\frac{1}{2}}}{u(s_0^2 - q^2)^{\frac{1}{2}}}\, s_0\, ds_0$$

$$= -\frac{2b}{(2\pi)^2}\frac{1}{R_1}\frac{d}{dR_1} I(R_1, \tau), \tag{56}$$

where

$$I(R_1, \tau) = \int_0^{\infty}\cos s_0 R_1 \frac{\sin u\tau(s_0^2 - q^2)^{\frac{1}{2}}}{u(s_0^2 - q^2)^{\frac{1}{2}}}\, ds_0 \tag{57}$$

[24] G. Birkhoff and S. MacLane, *A Brief Survey of Modern Algebra* (The Macmillan Company, New York, 1962), pp. 212–254.

or,[25] since the integrand is an even function of s_0,

$$I(R_1, \tau) = \frac{1}{2}\int_{-\infty}^{\infty} e^{is_0R_1}\frac{\sin u\tau(s_0^2 - q^2)^{\frac{1}{2}}}{u(s_0^2 - q^2)^{\frac{1}{2}}}\,ds_0$$
$$= \frac{\pi}{2}J_0[q(R_1^2 - u^2\tau^2)^{\frac{1}{2}}]$$
$$\times\int_0^{\tau}[\delta(R_1 + u\tau') + \delta(R_1 - u\tau')]\,d\tau'. \quad (58)$$

By direct differentiation, it follows from Eqs. (56) and (58) that[26]

$$(\mathrm{R}')_{12} = \frac{b}{4\pi u}\frac{1}{R_1}\delta(R_1 - u\tau)$$
$$+ \frac{1}{4\pi(R_1^2 - u^2\tau^2)^{\frac{1}{2}}}J_1[q(R_1^2 - u^2\tau^2)^{\frac{1}{2}}]$$
$$\times 1_+(R_1 - u\tau), \quad (59)$$

$1_+(x)$ designates the Heaviside unit step function defined as

$$1_+(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

As a result of Eqs. (31), (41), (43), and (44) we may write

$$\Phi(\mathbf{r}, t) = (\mathrm{R}')_{12} * h(\mathbf{r}, t')$$
$$= \int_{E_3}[\mathrm{R}'(\mathbf{r}, t/\mathbf{r}', t')]_{12}\frac{\partial\psi(\mathbf{r}', t)}{\partial t}\Big]_{t=t'}$$
$$\times \exp(\alpha z' + \beta t')\,d\mathbf{r}'. \quad (60)$$

On the other hand, we recall that

$$\psi(\mathbf{r}, t) = \exp(-\alpha z - \beta t)\Phi(\mathbf{r}, t)$$
$$= \int_{E_3}\exp[-\alpha(z - z') - \beta(t - t')]$$
$$\times [\mathrm{R}'(\mathbf{r}, t/\mathbf{r}', t')]_{12}\frac{\partial\psi(\mathbf{r}', t)}{\partial t}\Big]_{t=t'}d\mathbf{r}'$$
$$= \int_{E_3}[\mathrm{R}(\mathbf{r}, t/\mathbf{r}', t')]_{12}\frac{\partial\psi(\mathbf{r}', t)}{\partial t}\Big]_{t=t'}d\mathbf{r}'. \quad (61)$$

This yields immediately the relationship

$$(\mathrm{R})_{12} = \exp[-\alpha(z - z') - \beta(t - t')](\mathrm{R}')_{12}. \quad (62)$$

On the basis of Eqs. (24), (59), and (62), we finally assert that

$$G(\mathbf{r}, t/\mathbf{r}', t') = 0 \quad (63a)$$

for $t < t'$, and

$$G(\mathbf{r}, t/\mathbf{r}', t') = u^2(\mathrm{R})_{12} = u^2\exp[-\alpha(z - z') - \beta\tau]$$
$$\times\left(\frac{b}{4\pi u}\frac{1}{R_1}\delta(R_1 - u\tau) - \frac{b}{4\pi}\frac{1}{(R_1^2 - u^2\tau^2)^{\frac{1}{2}}}\right.$$
$$\left.\times J_1[q(R_1^2 - u^2\tau^2)^{\frac{1}{2}}]1_+(R_1 - u\tau)\right) \quad (63b)$$

for $t \geq t'$. In this equation,

$$R_1 = \{(x - x')^2 + (y - y')^2 + b^2$$
$$\times [(z - z') - (\gamma/b)(t - t')]^2\}^{\frac{1}{2}}. \quad (64)$$

Before an attempt is made to give an interpretation of the above solution, we consider the following interesting special cases:

(1) If $v = 0$ and $\sigma' \neq 0$,

$$G(\mathbf{r}, t/\mathbf{r}', t') = u^2\exp\left(-\frac{\sigma'}{2\epsilon'}\tau\right)$$
$$\times\left(\frac{1}{4\pi u}\frac{1}{R}\delta(R - u\tau) - \frac{1}{4\pi}\frac{1}{(R^2 - u^2\tau^2)^{\frac{1}{2}}}\right.$$
$$\left.\times J_1[\tfrac{1}{2}\sigma'(\mu'/\epsilon')^{\frac{1}{2}}(R^2 - u^2\tau^2)^{\frac{1}{2}}]1_+(R - u\tau)\right) \quad (65)$$

is the solution of the problem

$$[\nabla^2 - \mu'\epsilon'(\partial^2/\partial t^2) - \mu'\sigma'(\partial/\partial t)]\,G(\mathbf{r}, t/\mathbf{r}', t')$$
$$= \delta(\mathbf{r} - \mathbf{r}')\delta(t - t'), \quad t \geq t'. \quad (66)$$

It is also very interesting to note here that if the factor $\exp[-(\sigma'\tau)/(2\epsilon')]$ were absent in Eq. (65), the remaining expression would correspond to the time-dependent Green's function for the three-dimensional Klein–Gordon equation of relativistic quantum mechanics, viz.,

$$[\nabla^2 - \mu'\epsilon'(\partial^2/\partial t^2) + q^2]\,G(\mathbf{r}, t/\mathbf{r}', t')$$
$$= \delta(\mathbf{r} - \mathbf{r}')\delta(t - t'), \quad t \geq t'. \quad (67)$$

A solution to this equation has been obtained by means of contour integration in the Ref. 27.

(2) If $v \neq 0$ and $\sigma' = 0$,

$$G(\mathbf{r}, t/\mathbf{r}', t') = (bu/4\pi)R_1^{-1}\,\delta(R_1 - u\tau) \quad (68)$$

satisfies the equation

$$[\nabla^2 - \mu'\epsilon'(\partial^2/\partial t^2) - 2\Lambda(\partial^2/\partial t\partial z)]\,G(\mathbf{r}, t/\mathbf{r}', t')$$
$$= \delta(\mathbf{r} - \mathbf{r}')\delta(t - t'), \quad t \geq t', \quad (69)$$

which is the nonrelativistic approximation of Compton's result (cf. Ref. 14). The reader is also referred to a recent communication by Unz and Chawla.[28]

[25] G. F. D. Duff and D. Naylor, *Differential Equations of Applied Mathematics* (John Wiley & Sons, Inc., New York, 1966), p. 412.

[26] Since the following solution applies only for $t \geq t'$, there is no contribution from the term involving the Dirac delta function $\delta(R_1 + u\tau)$.

[27] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, pp. 854–857.

[28] B. R. Chawla and H. Unz, Proc. IEEE **54**, 307 (1966).

(3) Lastly, if $v = 0$ (or $\mu'\epsilon' = \mu_0\epsilon_0$) and $\sigma' = 0$,

$$G(\mathbf{r}, t/\mathbf{r}', t') = (u/4\pi)\, R^{-1}\, \delta(R - u\tau) \qquad (70)$$

is the well-known solution of the simple wave equation

$$[\nabla^2 - u^{-2}(\partial^2/\partial t^2)]\, G(\mathbf{r}, t/\mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t'), \quad t \geq t'. \qquad (71)$$

VI. CONCLUSIONS

The part of the Green's function $G(\mathbf{r}, t/\mathbf{r}', t')$ [cf. Eq. (63b)] containing the Dirac δ function can be interpreted as an expanding wavefront which arrives at $R_1 = u\tau$ diminished by the geometrical factor $1/R_1$ and modified by the exponential term

$$\exp\,[-\alpha(z - z') - \beta\tau].$$

R_1, as given in Eq. (64), can be taken as the "radial" distance between the point $(x', y', z' + \gamma\tau/b)$ and the observation point $\mathbf{r}(x, y, z)$ with a scaling of the z axis dimensions by the factor b^2 which, in turn, is associated with the Lorentz contraction along this axis.

Whereas for a "stationary" medium ($v = 0$ or $\epsilon'\mu' = \epsilon_0\mu_0$) the expanding wavefronts are spheres centered at the spatial position of the source, in the more general problem under consideration here, apart from the multiplicative factors $1/R_1$ and

$$\exp\,[-\alpha(z - z') - \beta\tau],$$

the wavefronts obey the equation

$$R_1 = u\tau. \qquad (72)$$

It is quite easy to show that, for constant τ, the wavefronts are spheroidal surfaces with semiaxes $u\tau$, $u\tau$, and $u\tau/b$ along the directions of the x, y, and z axes, respectively. Since $b > 1$, it is seen that $u\tau/b < a\tau$; hence the wavefronts are oblate spheroids with respect to the z direction.[29] The wavefront center $(x', y', z' + \gamma\tau/b)$ moves along the z direction with a speed γ/b. Inasmuch as $\gamma = u^2 b\Lambda$, it follows that

$$\gamma/b = u^2\Lambda = (1 - n^{-2})v < v; \quad n = (\mu'\epsilon'/\mu_0\epsilon_0)^{\frac{1}{2}}.$$

Thus, the center of the spheroid cannot keep up with the medium. Furthermore, since $u\tau/b > \gamma\tau/b$ for the nonrelativistic approximation case, the wavefronts enclose the source point; that is the source radiates in all directions. This excludes the important phenomenon of Čerenkov radiation which takes place in the arbitrary-velocity case if $nv/c > 1$ (cf. Ref. 14).

The effect of a pulse at a distance R_1 and at a time τ after its onset vanishes for $R_1 > u\tau$, that is as long as the wave initiated by the pulse has not had sufficient time to reach the observation point r. At $R_1 = u\tau$, the original pulse arrives, diminished by the geometrical factor $1/R_1$. The wave then leaves in its wake a residue or tail which persists for an infinite time at points which have been traversed by the wavefronts. This contribution is represented by the second part of $G(\mathbf{r}, t/\mathbf{r}', t')$ in Eq. (63b). The entire solution is, of course, attenuated exponentially in the z direction. Furthermore, it subsides exponentially with respect to time.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor R. E. Collin and Professor R. T. Compton, Jr., for their encouragement and helpful suggestions.

This work was supported by the Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract AF 19(628)1699.

[29] For an arbitrary velocity v, $b = [(n^2 - \beta^2)/n^2(1 - \beta^2)]^{\frac{1}{2}}$, where $\beta = v/c$. In the nonrelativistic limit, i.e., for $\beta \ll 1$ and $n\beta = v/u \ll 1$, b approaches unity from above.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Linear Representation of Spinors by Tensors

CARLO CERCIGNANI
Istituto di Scienze Fisiche-Universita di Milano, Milano, Italy

(Received 1 November 1965)

A linear representation of spinors in n-dimensional space by tensors is proposed. In particular, in three-dimensional space a set composed by a scalar and a vector is associated to any two-component spinor, while in four-dimensional space the set corresponding to a four-component spinor is composed by a scalar, a pseudoscalar, a vector, a pseudovector, and an antisymmetrical tensor of second order. The resulting formalism is then applied to Schrödinger's and Dirac's equations. In three-dimensional space it turns out that the proposed procedure automatically assigns an intrinsic magnetic moment to an electron in a magnetic field without introducing any relativistic ideas or *ad hoc* assumptions. In four-dimensional space we can write the Dirac equation in a generally covariant fashion, without introducing new concepts with respect to the usual tensor analysis. The zero-mass Dirac equation splits into two sets of equations, describing respectively the neutrino and the photon. The possible bearing of the proposed approach upon the theories of elementary particles is briefly discussed.

1. INTRODUCTION

IT is well known that one of the strongest objections which can be raised against the theories which aim to describe quantum phenomena in the frame of a classical space–time geometry consists in the difficulty of deriving the concept of spinor out of the concept of tensor.[1,2] As a matter of fact, the inverse procedure of deriving vectors and tensors out of spinor fields[3–6] has been sometimes suggested.

It is also well known that tensor quantities can be constructed by means of spinors. However, these tensors are quadratic in the spinor components and the tensor equations equivalent to the Dirac equation lose the important property of linearity, at least explicitly. The quadratic character of the mentioned tensors also suggested the idea of spinor as "square root" of a vector.[2]

The aim of this paper is to exhibit a systematic procedure to transcribe any (linear) spinor equation in pure tensor terms. The advantages of this transcription are not merely formal (among other things the resulting equations are generally covariant), but could turn out to be essential for (1) the above-mentioned unitary theories; (2) any theory, where the linear equations of quantum mechanics are assumed to be an approximation (which ceases to hold at small distances) of a nonlinear theory.[7,8] In fact, in such a case the spinor and tensor descriptions could no longer be equivalent, and neglecting one of them could mean discarding useful alternatives.

The problem of a tensor transcription of spinor equations has been recently considered by Klauder.[9] In our opinion, however, the solution given by Klauder is not satisfactory, because arbitrary tensors devoid of any physical significance appear in his tensor equations. The choice of these tensors is equivalent to the choice of an explicit representation of the Dirac matrices in the usual spinor description; here we propose circumventing this particular difficulty that Klauder encountered by introducing a formalism which appears to be interesting. To be precise we exhibit a method which leads to tensor equations free of any quantity to be chosen arbitrarily. In Sec. 2 we describe the simple rule to transcribe an arbitrary spinor equation in a n-dimensional space into a system of tensor equations. In Sec. 3 an application to the three-dimensional case (Schrödinger–Pauli equation) is made; in Sec. 4 the much more interesting case of space–time is considered. Finally, in Sec. 5 the possible objections to the abandonment of the usual description in favor of the presently proposed one are discussed.

2. FUNDAMENTAL RULES

The spinor equations in a n-dimensional space make use of square matrices of order 2^N, i.e., with 2^{2N} elements, where N is determined by n; it is well known that $N = 1$ for $n = 3$, $N = 2$ for $n = 4$. The algebra of these matrices is a Clifford algebra, which can be thought of as generated by n elements (this is not the only possible choice of the generators, but is convenient from our standpoint). We denote such

[1] W. Pauli, *Teoria della Relativita* (Boringhieri, Torino, Italy, 1958), p. 317.
[2] J. A. Wheeler, *Geometrodynamics* (Academic Press Inc., New York, 1962), p. 89.
[3] P. Jordan, Ergeb. Exakt. Naturw. 7, 158 (1928); Z. Physik. **105**, 229 (1937).
[4] P. Jordan and R. de L. Kronig, Z. Physik., **100**, 569 (1936).
[5] R. de L. Kronig, Physica 3, 1120 (1936).
[6] L. de Broglie, W. Heisenberg, and H. A. Kramers, in *L. de Broglie, Physicien et Penseur* (Paris, 1953).
[7] L. de Broglie, *Introduction to the Vigier Theory of Elementary Particles* (Elsevier Publishing Company, Amsterdam, 1963).
[8] W. Heisenberg, Rev. Mod. Phys. **29**, 269 (1957).
[9] J. R. Klauder, J. Math. Phys. **5**, 1 (1964).

generators by γ_i. They are assumed to satisfy the anticommutation rules

$$\gamma_i\gamma^j + \gamma^j\gamma_i = \delta_i^j \quad (i = j = 1, 2, \cdots, n), \quad (1)$$

where the upper index γ's are related to the lower index ones by means of the fundamental tensor g_{ij} according to the following relations:

$$\gamma^j = g^{ji}\gamma_i; \quad \gamma_i = g_{ij}\gamma^j \quad (i, j = 1, 2, \cdots, n). \quad (2)$$

The above matrices are assumed to operate either on one-column matrices at their right or on one-row matrices at their left: these four elements matrices are usually referred to as spinors,[10] once a precise transformation rule when going from a Cartesian reference system to another is associated with them. Usually only orthogonal Cartesian coordinates are adopted and the γ's are treated as invariants. However, if general coordinates are considered, it is more convenient to treat "spinors" as invariants and the four matrices γ_i as components of a vector [see Ref. 11 for a general review of the literature].

In any case, in a given reference frame, a linear spinor equation can be written as follows:

$$L\psi = 0, \quad (3)$$

where L is a square matrix whose elements are operators (e.g., differential operators) operating on the components of the one-column matrix ψ. Of course, a row-by-column product between L and ψ is understood.

The matrix L can be written as a linear combination of those elements of the Clifford algebra which generate the whole algebra additively. These elements are the γ's and suitable combinations of multiple products of the γ's. The products are chosen in such a way as to form a linearly independent set together with the γ's. Under this respect the algebra is treated as a 2^{2N}-dimensional space, $S_{2^{2N}}$.

In the three-dimensional case the mentioned elements are simply the γ's and the identity 1 of the algebra, while in the case of space–time we have to consider a set of 16 elements, which are conveniently individuated as follows:

$$\begin{aligned} &1, \\ &\gamma_\alpha, \\ &\gamma_{\alpha\beta} = \tfrac{1}{2}[\gamma_\alpha\gamma_\beta - \gamma_\beta\gamma_\alpha], \\ &\gamma_{\alpha\beta\gamma} = \tfrac{1}{3}[\gamma_\alpha\gamma_{\beta\gamma} + \gamma_\beta\gamma_{\gamma\alpha} + \gamma_\gamma\gamma_{\alpha\beta}], \\ &\gamma_{\alpha\beta\gamma\delta} = \tfrac{1}{4}[\gamma_\alpha\gamma_{\beta\gamma\delta} - \gamma_\beta\gamma_{\gamma\delta\alpha} + \gamma_\gamma\gamma_{\delta\alpha\beta} - \gamma_\delta\gamma_{\alpha\beta\gamma}] \\ &\qquad (\alpha, \beta, \gamma, \delta = 0, 1, 2, 3). \end{aligned} \quad (4)$$

It is evident that each element endowed with two or more indexes is completely antisymmetric in these indexes; therefore it is easily verified that we have exactly $2^4 = 16$ linearly independent elements.

If γ_α is assumed to transform as the αth component of a vector, $\gamma_{\alpha\beta}$ transforms as an antisymmetric tensor of second order, $\gamma_{\alpha\beta\gamma}$ as an antisymmetric tensor of third order (i.e., as a pseudovector in space–time) and $\gamma_{\alpha\beta\gamma\delta}$ as an antisymmetric tensor of fourth order (i.e., as a pseudoscalar).

Generally speaking, in a n-dimensional space γ_J denotes any one of the above-mentioned generators; accordingly a capital letter as an index denotes a set of different indexes of the previous kind (lower case letters). Such capital letters can in turn be considered as numerical indexes, varying from 1 to 2^{2N}, provided that an arbitrary numeration of the above sets of indexes is introduced.

Therefore L can be written as follows:

$$L = L_J\gamma_J \quad (J = 1, \cdots, 2^{2N}), \quad (5)$$

where by L_J we denote scalar operators (e.g., differential operators) of the same kind as the components of the matrix L.

Let us note now that also the spinor ψ can be considered as an element of the Clifford algebra; it suffices, e.g., to write in place of the original one-column matrix a square matrix having $2^N - 1$ columns of zeros and the remaining column equal to the column which constituted the original spinor. In such a way we can also write

$$\psi = \psi_J\gamma_J \quad (J = 1, 2, \cdots, 2^{2N}). \quad (6)$$

As a consequence, Eq. (3) can be written as follows:

$$(L\psi)_J\gamma_J = 0, \quad (7)$$

where the quantities $(L\psi)_J$ are linear combinations of the quantities $L_I\psi_K$, which can be determined through the multiplication table of the algebra. Thanks to the linear independence of the γ_J, Eq. (7) immediately gives

$$(L\psi)_J = 0. \quad (8)$$

Now, if the generators of the vector space $S_{2^{2N}}$, γ_J, have been chosen in such a way as to have tensor properties in the n-dimensional space (this is always possible), the above equations are grouped immediately to give tensor equations. This important result seems to be partially overshadowed by the circumstance that the number of the required components is to be increased (from 2^N to 2^{2N}) and the new components, though more numerous, can be expressed as linear combinations of the old ones.

3. THE CASE OF $n = 3$

In this section we consider in some detail the transcription of the Schrödinger–Pauli equation,

[10] From the standpoint of the theory of the algebras, a spinor can be described as an ideal of the considered Clifford algebra. [M. Riesz, Lund University Math Seminar **12**, 241 (1954).]

[11] W. L. Bade and H. Jehle, Rev. Mod. Phys. **25**, 714 (1953).

which regulates the quantum mechanics of a nonrelativistic electron in a magnetic field. In three-dimensional space we conform to the usual notations and write σ_i in place of γ_i.

As we have already noticed, in the three-dimensional case any element of the Clifford algebra can be written as follows:

$$a_0 + a_i\sigma^i \quad (i = 1, 2, 3), \tag{9}$$

where, if the whole expression is treated as an invariant and the σ's as the components of a vector, a_0 is a scalar and $a \equiv (a_1, a_2, a_3)$ is a vector.

The translation of the usual Schrödinger–Pauli equation in the new language is straightforward and is not effected explicitly here. Instead, we give a direct derivation of this equation from classical mechanics through the usual rules of correspondence between classical observables and quantum operators. This derivation is very interesting, since it shows that an accurate use of the quantum-classical analogy leads one to automatically foresee the existence of spin, which is usually considered as an *ad hoc* assumption or a consequence of introducing relativistic invariance.

We note that the classical Hamiltonian is characterized by the presence of the square of a vector (the momentum $\mathbf{p}$ in absence of magnetic field, $\mathbf{p} - e\mathbf{A}/c$ in the general case). Therefore we need a definition of the square of a vector when its components are not c numbers, but q numbers. This gives no trouble if the components of the vector commute. However, the rule that the square of a vector is the sum of the squares of its orthogonal Cartesian components is usually given even in the case of noncommuting components. This rule does not appear to be consistent with the vector concept, which is primarily synthetic and only subordinately analytic. Such conception of vectors is not correctly appreciated even in the ordinary vector calculus, which is only formally synthetic. On the contrary, the concept of vector as a whole is completely expressed in the concepts of quaternion algebra, according to the views of R. W. Hamilton, or, alternatively, of the Clifford algebra corresponding to $n = 3$, which is isomorphic to the quaternion algebra (on the complex field). Classically we can write a vector in the following form:

$$\mathbf{v} = \sigma_i v^i = \sigma^i v_i . \tag{10}$$

Then, according to the multiplication rules of the σ's and taking into account that the components of $\mathbf{v}$ commute, the square of $\mathbf{v}$ is given by

$$\begin{aligned} \mathbf{v}^2 = (\sigma^i v_i)(\sigma_j v^j) &= \sigma^i\sigma_j v_i v^j \\ &= \tfrac{1}{2}(\sigma_i\sigma^j + \sigma^j\sigma_i)v_i v^j \\ &= \delta_i^j v_j v^i = v_i v^i, \end{aligned} \tag{11}$$

i.e., it is the same as obtained by operating on the components according to usual rules. But, if the vector components are noncommuting operators, we get

$$\begin{aligned} \mathbf{v}^2 &= \sigma^i\sigma_j v_i v^j \\ &= \tfrac{1}{2}(\sigma^i\sigma_j + \sigma_j\sigma^i)v_i v^j + \tfrac{1}{2}(\sigma^i\sigma_j - \sigma_j\sigma^i)v_i v^j \\ &= \delta_j^i v_i v^j + i\epsilon^i_{\cdot jk}\sigma^k v_i v^j \\ &\equiv v_i v^i + \frac{i}{2}\epsilon_{ijk}[v_i, v_j]\sigma^k. \end{aligned} \tag{12}$$

Here we have taken into account that the following relation holds ($n = 3$):

$$\tfrac{1}{2}(\sigma^i\sigma_j - \sigma_j\sigma^i) = i\epsilon^i_{\cdot jk}\sigma^k, \tag{13}$$

where $\epsilon^i_{\cdot jk}$ is the completely antisymmetric tensor of Ricci and Levi–Civita. In Eq. (12) $[v_i, v_j]$ denotes, as usual, the commutator between v_i and v_j.

It appears that the result given by Eq. (13) is not only different from the result which can be obtained by operating on the components, but is not even a scalar, being a more general quaternion or element of the considered Clifford algebra.

In the case of interest, where $\mathbf{v} = \mathbf{p} - e\mathbf{A}/c$ with $\mathbf{p} = -i\hbar$ grad, we have

$$\begin{aligned} -\hbar^2\Delta_2 + 2i\hbar eA_i\frac{\partial}{\partial x_i} + i\hbar\frac{e}{c}\frac{\partial A_i}{\partial x_i} \\ + \frac{e^2}{c^2}A^2 - \frac{e\hbar}{c}H_k\sigma^k. \end{aligned} \tag{14}$$

Here $H_k = \frac{1}{2}\epsilon_{ijk}(\partial A_j/\partial x_i - \partial A_i/\partial x_j)$ is the magnetic field corresponding to the vector potential $\mathbf{A}$. It follows that the Hamiltonian operator $v^2/2m + U$ is given by

$$\begin{aligned} -\frac{\hbar^2}{2m}\Delta_2 + \frac{i\hbar e}{m}A_i\frac{\partial}{\partial x_i} + \frac{i\hbar e}{2mc}\frac{\partial A_i}{\partial x_i} \\ + \frac{e^2}{2mc^2}A^2 - \frac{e\hbar}{2mc}H_k\sigma^k + U. \end{aligned} \tag{15}$$

It is easily verified that Eq. (15) gives just the operator which could be obtained from the usual Pauli–Schrödinger equation by means of the rules given in Sec. 2. Therefore the present procedure turns out to automatically ascribe a magnetic moment $\mu_0 = e\hbar/2mc$ to a nonrelativistic particle of charge e and mass m, as a mere consequence of the commutation rules.

In order to clarify the procedure, let us write the Schrödinger equation separating the scalar and the vector parts of the quaternion equation. If $\mathcal{H}_0$ denotes the scalar part of the Hamiltonian given by

Eq. (15),[12] we easily find

$$\mathcal{H}_0\psi_0 - \mu_0\mathbf{H}\cdot\boldsymbol{\psi} = i\hbar(\partial\psi_0/\partial t), \tag{16a}$$

$$\mathcal{H}_0\boldsymbol{\psi} - \mu_0\mathbf{H}\psi_0 - i\mu_0\mathbf{H}\times\boldsymbol{\psi} = i\hbar(\partial\boldsymbol{\psi}/\partial t), \tag{16b}$$

where ψ_0 is the scalar part of the quaternion wave-function and $\boldsymbol{\psi}$ its vector part. In the above equations a dot denotes the ordinary scalar product, $\times$ the ordinary vector product.

From Eqs. (16) we can easily obtain well-known results in the case of an uniform magnetic field. In such a case, in fact, if we assume that $\boldsymbol{\psi}$ is not parallel to $\mathbf{H}$, the component of Eq. (16b) along the normal to the plane individuated by $\mathbf{H}$ and $\boldsymbol{\psi}$ gives

$$\mathbf{H}\times\boldsymbol{\psi} = 0, \tag{17}$$

i.e., against our assumption, $\boldsymbol{\psi}$ is parallel to $\mathbf{H}$. Then, if we put $\boldsymbol{\psi} = \mathbf{H}\varphi_0/H$, where H denotes the absolute value of the magnetic field, we easily get

$$\mathcal{H}\psi_0 - \mu_0 H\varphi_0 = i\hbar(\partial\psi_0/\partial t), \tag{18a}$$

$$\mathcal{H}_0\varphi_0 - \mu_0 H\psi_0 = i\hbar(\partial\varphi_0/\partial t). \tag{18b}$$

It follows that φ_0 and ψ_0 are proportional to each other: $\varphi_0 = \lambda\psi_0$ and, in order that Eqs. (18a) and (18b) can be contemporarily satisfied, $\lambda = \pm 1$. Then ψ_0 satisfies

$$\mathcal{H}_0\psi_0 \pm \mu_0 H\psi_0 = i\hbar(\partial\psi_0/\partial t). \tag{19}$$

Hence the well-known result of the splitting of energy levels follows. We note that this result has been obtained immediately in a form which is valid in any reference frame without using the single components. The above treatment is open to criticism; as a matter of fact, it seems puzzling that a magnetic moment $\mu_0 = e\hbar/2mc$ is ascribed to a nonrelativistic particle of charge e and mass m, while, e.g., the charged π meson has zero spin. Different answers can be given to this objection. Firstly we note that the known charged particles with spin different from $\frac{1}{2}$ can be hardly regarded as nonrelativistic. Secondly, since the properties of the particles should depend on the fields with which they can interact, the exclusion of the nuclear field could be of vital importance for the result we have found. Finally, since we have replaced the ordinary unit vectors of rectangular coordinates by the σ_j's and this is not a strict consequence of the discussion on vectors given at the beginning of this section, one could suggest[13] that this replacement is a matter of experiment, i.e., the fact that the procedure works is an *a posteriori* proof that the algebra of the σ_j's underlies the theory of nonrelativistic electrons.

[12] This scalar part of the Hamiltonian coincides with the Hamiltonian which is obtained according to the usual rules without introducing spin with an *ad hoc* assumption.

[13] This opinion was expressed by an anonymous referee, whose observations suggested the present discussion.

4. THE CASE $n = 4$

In this section we consider in some detail the transcription of the Dirac equation which regulates the quantum relativistic mechanics of an electron. We neglect the electromagnetic fields, since the extension of the equation in absence of fields to the equation in presence of fields is trivial, following from well-known rules based on the principle of gauge invariance.

The mentioned transcription is easily effected, provided that the following multiplication table is taken into account:

$$\begin{aligned}
\gamma^\mu\gamma_\beta &= \delta^\mu_\beta + \gamma^\mu_{\cdot\beta}, \\
\gamma^\mu\gamma_{\alpha\beta} &= \gamma^\mu_{\cdot\alpha\beta} + \gamma_\beta\delta^\mu_\alpha - \gamma_\alpha\gamma^\mu_\beta, \\
\gamma^\mu\gamma_{\alpha\beta\gamma} &= \gamma^\mu_{\cdot\alpha\beta\gamma} + \gamma_{\alpha\beta}\delta^\mu_\gamma + \gamma_{\beta\gamma}\delta^\mu_\alpha + \gamma_{\gamma\alpha}\delta^\mu_\beta, \\
\gamma^\mu\gamma_{\alpha\beta\gamma\delta} &= \gamma_{\beta\gamma\delta}\gamma^\mu_\alpha - \gamma_{\gamma\delta\alpha}\delta^\mu_\beta + \gamma_{\delta\alpha\beta}\delta^\mu_\beta - \gamma_{\alpha\beta\gamma}\delta^\mu_\beta \\
&\qquad (\alpha, \beta, \gamma, \delta = 0, 1, 2, 3).
\end{aligned} \tag{20}$$

We obtain

$$\begin{aligned}
\psi^\beta{}_{/\beta} &= i\lambda\psi, \\
\psi_{/\beta} + 2\psi^{\mu\beta}{}_{/\mu} &= i\lambda\psi^\beta, \\
\tfrac{1}{2}[\psi^{\beta/\alpha} - \psi^{\alpha/\beta}] + 3\psi^{\alpha\beta\mu}{}_{/\mu} &= i\lambda\psi^{\alpha\beta}, \\
\tfrac{1}{3}[\psi^{\alpha\beta/\delta} + \psi^{\beta\delta/\alpha} + \psi^{\delta\alpha/\beta}] + 4\psi^{\alpha\beta\mu\delta}{}_{/\mu} &= i\lambda\psi^{\delta\alpha\beta}, \\
\tfrac{1}{4}[\psi^{\alpha\beta\gamma/\mu} - \psi^{\beta\gamma\mu/\alpha} + \psi^{\gamma\mu\alpha/\beta} - \psi^{\mu\alpha\beta/\gamma}] &= i\lambda\psi^{\mu\alpha\beta\gamma} \\
&\qquad (\alpha, \beta, \gamma, \delta, \mu = 0, 1, 2, 3),
\end{aligned} \tag{21}$$

where $\lambda = mc/\hbar$.

It is to be noted that the above equations are automatically covariant with respect to general coordinate transformations. After a careful review of the literature, we have found that the above tensor transcription of the Dirac equation had already been found by Lanczos as early as 1929,[14] as a consequence of a quaternion treatment of the Dirac equation. However, it turns out that this work remained unnoticed in the subsequent papers dealing with the covariant form of the Dirac equation.

Now two cases are to be distinguished, according to whether λ is equal to or different from zero. In the second case, we can express ψ, $\psi_{\alpha\beta}$, $\psi_{\alpha\beta\gamma\delta}$ in terms of ψ_α and $\psi_{\alpha\beta\gamma}$ through the first, third, and fifth equations of system (21), and eliminating the former quantities from the second and fourth equation gives

$$\Box\psi_\alpha + \lambda^2\psi_\alpha = 0; \quad \Box\psi_{\alpha\beta\gamma} + \lambda^2\psi_{\alpha\beta\gamma} = 0, \tag{22}$$

i.e., the vector ψ_α and the pseudovector $\psi_{\alpha\beta\gamma}$ satisfy a Proca–Yukawa equation. Equations (22) are the only restrictions on ψ_α and $\psi_{\alpha\beta\gamma}$. Alternatively, we can eliminate ψ_α and $\psi_{\alpha\beta\gamma}$ in favor of ψ, $\psi_{\alpha\beta}$, $\psi_{\alpha\beta\gamma\delta}$ and find

$$\Box\psi + \lambda^2\psi = 0; \quad \Box\psi_{\alpha\beta} + \lambda^2\psi_{\alpha\beta} = 0; \quad \Box\psi_{\alpha\beta\gamma} + \lambda^2\psi_{\alpha\beta\gamma} = 0, \tag{23}$$

[14] C. Lanczos, Z. Physik. **57**, 447, 474, 484 (1929).

while ψ_α and $\psi_{\alpha\beta\gamma}$ are given by the second and fourth equations of system (21).

If $\lambda = 0$, then the first, third, and fifth equations of system (21) decouple completely from the second and fourth ones. The latter become

$$\psi^{\alpha\beta}{}_{/\alpha} + \tfrac{1}{2}\psi^{/\beta} = 0,$$
$$\psi^{\alpha\beta/\delta} + \psi^{\beta\delta/\alpha} + \psi^{\delta\alpha/\beta} + 12\psi^{\alpha\beta\gamma\delta}{}_{/\gamma} = 0$$
$$(\alpha, \beta, \gamma, \delta = 0, 1, 2, 3), \quad (24)$$

and are therefore a generalization of the Maxwell system (which can be obtained by putting $\psi = 0$, $\psi^{\alpha\beta\gamma\delta} = 0$). The remaining equations can in turn be written as follows:

$$\psi^{\beta}{}_{/\beta} = 0; \quad \tfrac{1}{2}[\psi^{\beta/\alpha} - \psi^{\alpha/\beta}] + 3\psi^{\alpha\beta\mu}{}_{/\mu} = 0.$$
$$\psi^{\alpha\beta\gamma/\mu} - \psi^{\beta\gamma\mu/\alpha} + \psi^{\gamma\mu\alpha/\beta} - \psi^{\mu\alpha\beta/\gamma} = 0$$
$$(\alpha, \beta, \gamma, \mu = 0, 1, 2, 3). \quad (25)$$

It is easily seen that the system of Eqs. (25) is equivalent to the Pauli–Yang–Lee equation for the neutrino. In fact, the second equation of system (25) can be written as follows:

$$3\varphi_{\delta\mu\alpha}{}^{/\alpha} + \tfrac{1}{2}(\varphi_{\mu/\delta} - \varphi_{\delta/\mu}) = 0$$
$$(\alpha, \delta, \mu = 0, 1, 2, 3), \quad (26)$$

where

$$\varphi_{\delta\mu\alpha} = (i/6)\epsilon_{\delta\mu\alpha\beta}\psi^{\beta}; \quad \varphi^{\mu} = i\epsilon_{\alpha\beta\gamma\mu}\psi^{\alpha\beta\gamma}.$$

Here $\epsilon_{\delta\mu\alpha\beta}$ is the four-dimensional Ricci–Levi-Civita tensor with components 0, $\pm(-g)^{\frac{1}{2}}$ ($g = \det \|g_{ik}\|$).

We see that the φ's and the ψ's satisfy the same equations; accordingly they can be assumed to be proportional as follows:

$$\psi_\mu = C\varphi_\mu = iC\epsilon_{\alpha\beta\gamma\mu}\psi^{\alpha\beta\gamma}, \quad (27a)$$
$$\psi_{\delta\mu\alpha} = C\varphi_{\delta\mu\alpha} = \tfrac{1}{6}iC\epsilon_{\delta\mu\alpha\beta}\psi^{\beta}. \quad (27b)$$

But substituting Eq. (27b) into Eq. (27a) gives

$$\psi_\mu = -\tfrac{1}{6}C^2\epsilon_{\alpha\beta\gamma\mu}\epsilon^{\alpha\beta\gamma\delta}\psi_\delta = \tfrac{1}{6}C^2 6\delta^{\delta}_{\mu}\psi_\delta = C^2\psi_\mu, \quad (28)$$

i.e.,

$$C = \pm 1, \quad (29)$$

and consequently

$$\psi_\mu = \pm i\epsilon_{\alpha\beta\gamma\mu}\psi^{\alpha\beta\gamma}; \quad \psi_{\delta\mu\alpha} = \pm\tfrac{1}{6}i\epsilon_{\delta\mu\alpha\beta}\psi^{\beta}. \quad (30)$$

The equations satisfied by ψ_μ can now be written as follows:

$$\psi^{\alpha}{}_{/\alpha} = 0, \quad \psi^{\beta/\alpha} - \psi^{\alpha/\beta} + i\epsilon^{\alpha\beta\gamma\delta}\psi_{\delta/\gamma} = 0, \quad (31)$$

where the first equation is the transcription of the first and third equations of system (25), which now coincide because of Eq. (30). Only the upper sign of Eq. (30) has been retained, since if ψ^α solves system (31), $\bar{\psi}^\alpha$ clearly solves the analogous system corresponding to the alternative choice. Here the bar denotes complex conjugation.

In order to show that system (31) is equivalent to the Pauli–Yang–Lee equation for neutrino and antineutrino, we note that the latter can be written[15] by equating to zero the result of operating on a spinor with the operator

$$(\partial/\partial x_0) + \sigma^k(\partial/\partial x_k). \quad (32)$$

Introducing the vector transcription discussed in Sec. 3 gives

$$[(\partial/\partial x_0) + \sigma^k(\partial/\partial x_k)](\psi^0 + \sigma_r\psi^r) = 0. \quad (33)$$

By means of the multiplication table of the σ's and in particular Eq. (13), Eq. (33) can be rewritten as follows:

$$\psi^0{}_{/0} + \psi^k{}_{/k} = 0, \quad \psi^{0/k} + \psi^k{}_{/0} + i\epsilon_{rsk}\psi_{r/s} = 0$$
$$(r, k, s = 1, 2, 3). \quad (34)$$

These equations have tensor character in three-dimensional space, but can be immediately rewritten in the four-dimensional formalism by introducing a vector $\psi^\alpha \equiv (\psi^0, \psi^1, \psi^2, \psi^3)$. The result coincides with system (31) provided we take into account that the second equation of this system, if verified for the values $\alpha = 1, 2, 3$, $\beta = 0$ of the indexes, is true for any couple of values of α and β, thanks to its duality invariance.

We note also that the second equation of system (31) is invariant under the gauge transformation

$$\psi_\alpha \to \psi_\alpha + \varphi_{/\alpha}; \quad (35)$$

accordingly the first equation of system (31) appears as a gauge normalization condition.

Finally we mention that system (21) can be obviously considered as the consequence of a stationary action principle with the following Lagrangian density:

$$\mathfrak{L} = \frac{1}{2i}\{1!\,(\bar{\psi}^\beta\psi^{/\beta} - \bar{\psi}_\beta\psi^{/\beta}) + 2!\,(\bar{\psi}_{\alpha\beta}\psi^{\beta/\alpha} - \psi_{\alpha\beta}\bar{\psi}^{\beta/\alpha})$$
$$+ 3!\,(\bar{\psi}_{\alpha\beta\gamma}\psi^{\alpha\beta/\gamma} - \psi_{\alpha\beta\gamma}\bar{\psi}^{\alpha\beta/\gamma})$$
$$+ 4!\,(\bar{\psi}_{\alpha\beta\gamma\delta}\psi^{\beta\gamma\delta/\alpha} - \psi_{\alpha\beta\gamma\delta}\bar{\psi}^{\beta\gamma\delta/\alpha})$$
$$- \lambda(\bar{\psi}\psi + 1!\,\bar{\psi}^\beta\psi_\beta + 2!\,\bar{\psi}_{\alpha\beta}\psi^{\alpha\beta}$$
$$+ 3!\,\bar{\psi}_{\alpha\beta\gamma}\psi^{\alpha\beta\gamma} + 4!\,\bar{\psi}_{\alpha\beta\gamma\delta}\psi^{\alpha\beta\gamma\delta})\}. \quad (36)$$

In particular, if $\lambda = 0$, systems (24) and (25), though decoupled, can be deduced from a joint variational principle.

5. DISCUSSION

The procedure which has been introduced and illustrated in the previous sections can meet with many objections which are presently examined.

A first objection concerns the considerable increasing of the number of components to be introduced. On the other hand, this circumstance is counterbalanced by the properties that the equations

[15] T. D. Lee and C. N. Yang, Phys. Rev. **104**, 254 (1956); **105**, 1119(L) (1957).

are symmetrical with respect to all the space–time coordinates (this is not true if one writes the components of a spinor equation) and are automatically covariant without introducing any new concept with respect to the usual tensor analysis (in particular the Christoffel symbols are enough to introduce covariant differentiation, without any need for the Fock–Ivanenko matrices or equivalent devices).

In connection with the increased number of unknowns, the question can be raised whether the number of possible solutions for a given problem can increase. The answer is negative, if only completely defined problems are considered. This means that if we have only one solution in the original spinor formulation, we can have possibly more than one solution in the tensor version, but all these have the same eigenvalues, give the same quadratic quantities, etc. Therefore it appears that this larger arbitrariness plays, in a certain sense, the same role as the indeterminacy in the explicit representation of the γ's in the usual spinor formulation. This does not mean that the multiplicity depends on some explicit representation of the γ_J. On the contrary, the origin of the multiplicity clearly lies in the lack of uniqueness of the decomposition of the spinor field ψ into a superposition of γ_J [Eq. (6)]. In fact, the representation of a spinor as a single column of a square matrix is not the only acceptable one. We could have taken any square matrix ψ such that it satisfies the matrix equation

$$\psi^2 - \psi(\mathrm{Sp}\ \psi) = 0, \tag{37}$$

where Sp ψ denotes the spur of ψ. This equation in turn implies a quadratic relation between the tensor components associated with the spinor individuated by the matrix ψ. The discussion of these quadratic relations leads us into another subject, which is not pursued further here. In fact, this matter seems to be especially important in the study of interacting particles, for which the present formalism could turn out to be very helpful.

In the frame of the matter considered in this paper, the advantage of the proposed point of view consists in the possibility of exploiting the undeniably more intuitive concepts of vector and tensor in place of spinors, in order to choose, among the possible equivalent solutions, that one endowed with the symmetry features of the physical system under investigation. This circumstance was exploited, e.g., in the three-dimensional case (Sec. 3), where the vector ψ was treated as it had a real unit vector, although ψ itself has complex components. In such a way we singled out a solution where the mentioned unit vector coincides with the unit vector of the magnetic field. But one can be easily convinced that there are infinitely many solutions with a nonreal "unit vector." These solutions are destitute of any significant symmetry, but are essentially equivalent to the explicitly found solution.

Another advantage of the proposed approach is that it appears to be the only consistent realization of the standpoint that the γ's transform as vector components, without encountering the objection that this standpoint violates the spirit if not the letter of the relativity idea.[11] This objection arises in the usual theory with a smaller number of components since the Dirac equation, when written in full, is not the same in all Lorentz frames.

A second objection which immediately arises is suggested by the connection between the tensor character of a field and the spin of the associated particle. Of course, a precise description of this question requires to study the field not free but interacting with other fields. We restrict ourselves merely to note that inconsistencies cannot arise on this point since the present description is just a transcription of the usual theory in different terms; therefore differences in physically significant results cannot be present, provided that the definitions of the different physical quantities are consistently transcribed.

Concerning the suggestions that the present treatment can give for subsequent developments, we note the noticeable fact that, from a single Dirac equation with zero mass (and a single action principle), one obtains the wave equations of the two known massless particles, i.e., photon and neutrino. This fact suggests the possibility that from a single equation (with a nonlinear self-interaction term) one can derive the wave equations of both massive and massless particles. Of course, this is not a new idea; however, the present formalism seems able to suggest a limited number of possible choices for the nonlinear term. In this frame the multiplication of the number of required components could be important in order to explain the existence of particles very similar but distinguishable as electron and muon or ν_e and ν_μ neutrinos.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Stability of Matter. I

FREEMAN J. DYSON AND A. LENARD*
The Institute for Advanced Study, Princeton, New Jersey

(Received 3 October 1966)

The stability problem of a system of charged point particles is discussed, and a number of relevant theorems are proven. The total energy of a system of N particles has a negative lower bound proportional to $N^{\frac{5}{3}}$ when no assumption is made on the statistics of the particles. When all particles belong to a fixed number of fermion species, a lower bound exists proportional to N.

1. INTRODUCTION

IN a recent paper, Fisher and Ruelle[1] raised the question: Is a quantum-mechanical system of electrical point charges stable? By stability they mean: There exists a lower bound for the total energy proportional to the total number of particles. In this and a following paper we address ourselves to this problem, by proving with rigorous analysis a number of theorems which are relevant to it.

The question of why matter is stable was very much the center of attention of physicists during the years after the discovery by Rutherford that matter consists of positive and negative point particles interacting by Coulomb forces, and before the establishment of wave mechanics. The origin of quantum theory, starting with Planck's work, is intimately bound up with this question. Planck's quantization of the radiation oscillators and Bohr's quantization of orbits in atoms served to stop the energy in matter from disappearing into the bottomless sink of the classical radiation field. In 1925 wave mechanics provided a quantitative solution to this problem. It became clear that an atom with a nuclear charge Ze and Z electrons of charge $-e$ could not have an energy state lower than $-Z^2\text{Ry}$, where $\text{Ry} = me^4/2\hbar^2$ is the natural atomic energy unit, the Rydberg, formed from the fundamental constants m, e, and $\hbar$.

This solved the problem of stability for single atoms. However, matter in bulk consists of a very large number of particles, positively and negatively charged, attracting and repelling each other by the Coulomb force. The effects of the Coulomb force are manifold and subtle, and often cooperative. They include such diverse phenomena as chemical binding, metallic cohesion, Van der Waals forces, superconductivity, superfluidity, and (in all likelihood) biology. The stability problem for matter in bulk is not a simple one. We need to understand why all these subtle effects have in common a saturation property, so that the binding energy per particle remains always bounded.

The empirical stability of matter does not depend on non-Coulombian forces (nuclear forces, magnetic dipole interactions, retardation and relativistic effects, radiative corrections). These contribute very small corrections to the binding energies of atoms and molecules. We are therefore justified in adopting the point of view that "matter" is a collection of point charges, interacting only through Coulomb forces, and subject to the laws of nonrelativistic quantum mechanics. If stability for this mathematical model is understood, stability for real matter is understood too.

We now give a formal definition of stability. Let the Hamiltonian operator of $N \geq 2$ charged particles be

$$H_N = \sum_{j=1}^{N}\left(-\frac{\hbar^2}{2m_j}\Delta_j\right) + \sum\sum_{1\leq i<j\leq N}\frac{e_ie_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (1.1)$$

Here we use the standard notation; the charges e_j may have either sign. We write

$$E_{\min}(N, e, m) = \text{Inf}\,(\psi, H_N\psi), \quad (1.2)$$

where the infimum is taken with respect to all N-particle wavefunctions $\psi = \psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$ normalized according to $(\psi, \psi) = 1$, all values of the masses satisfying

$$0 < m_j \leq m, \quad (1.3)$$

and all values of the charges satisfying

$$-e \leq e_j \leq e. \quad (1.4)$$

If there is a numerical constant A such that for all N

$$E_{\min} > -AN\,\text{Ry}, \quad (\text{Ry} = me^4/2\hbar^2), \quad (1.5)$$

we say that the system is *stable*.

In this definition, we have not mentioned the

* On leave of absence from the Plasma Physics Laboratory, Princeton University, Princeton, New Jersey. Present address: Department of Mathematics, Indiana University, Bloomington, Indiana.

[1] M. E. Fisher and D. Ruelle, J. Math. Phys. 7, 260 (1966).

statistics of the particles. Fisher and Ruelle in their paper[1] conjecture stability "with perhaps the restriction that either the positive or the negative particles obey Fermi statistics." The complete statement of stability or instability therefore involves a specification of the statistics of the particles. In that case the constant A may depend on the number and kind (in the sense of statistics) of different particle species.

The recent consideration of the stability problem[1] arose in connection with the need to establish a mathematically rigorous basis for statistical mechanics. Statistical mechanics makes physical sense only if thermodynamic quantities such as the energy, entropy, etc. are extensive, i.e., proportional (asymptotically for a large system) to the number of particles. Thus, stability in the sense (1.5) is necessary for the definition of a finite free energy per particle. The investigations of Ruelle[2] and Fisher[3] were restricted to models with short-range forces only. Thus, our investigation of the stability problem for Coulomb systems may be regarded as a necessary first step in establishing a rigorous statistical mechanics based on Coulomb forces alone, a challenging and difficult task.

2. STATEMENT OF RESULTS

Quite simple arguments suffice to give lower bounds for the energy of a system of charged particles, provided we do not require these bounds to be good for large N. We begin by stating two theorems of this nature. They are superseded by later theorems, and are only interesting because of the simplicity of their proofs.

Theorem 1: Under the hypotheses (1.3) and (1.4) we have

$$E_{\min} \geq -\tfrac{1}{8}N^2(N-1)\ \text{Ry}. \tag{2.1}$$

This is the result of Fisher and Ruelle.[1] For the sake of completeness, we reproduce their proof.

The following theorem, whose proof is slightly more difficult, is a refinement of Theorem 1 for $N > 5$, and it holds under the same hypotheses.

Theorem 2:

$$E_{\min} > -[N(N-1)/\sqrt{2}]\ \text{Ry}. \tag{2.2}$$

Both of these theorems give lower bounds which are far too low (except for small values of N). Our first nontrivial result is a further improvement which comes much closer to the truth.

Theorem 3:

$$E_{\min} > -AN^{\frac{5}{3}}\ \text{Ry}, \tag{2.3}$$

where $A < 52$ is an absolute constant.

Again, we assume inequalities (1.3) and (1.4) of the Introduction, but no assumption is made on particle statistics.

In connection with these theorems the question arises, what is the best possible result of this type? We believe that it is

$$E_{\min} > -AN^{\frac{7}{5}}\ \text{Ry}. \tag{2.4}$$

To prove that the exponent $\frac{7}{5}$ cannot be decreased it is sufficient to exhibit states ψ_N of N particles such that for some constant A'

$$(\psi_N, H_N\psi_N) < -A'N^{\frac{7}{5}}\ \text{Ry}. \tag{2.5}$$

Because the inequality (2.5) states an upper bound for the energy, conventional variational techniques are adequate for proving it. The result (2.5) is suggested by both a simple heuristic argument and by a detailed calculation based on the work of Foldy[4] and others.[5] Since we are interested in lower bounds for which new techniques must be used, we do not discuss the derivation of (2.5) in this paper but refer the interested reader to the lectures one of us held at the Summer Physics Institute of Brandeis University in 1966.[6] We find later that an improvement from (2.3) to (2.4) would necessitate going in an essential way beyond the techniques of the present work.

While (2.5) indicates that a Coulomb system without any restriction on particle statistics is unstable, the following result shows the importance of the exclusion principle for stability.

Theorem 4: Suppose that N particles whose masses and charges satisfy (1.3) and (1.4) belong to $q \geq 1$ distinct species of fermions. Then

$$E_{\min} > -Aq^{\frac{2}{3}}N\ \text{Ry}, \tag{2.6}$$

where $A < 500$ is an absolute constant. Briefly, a system whose particles belong to a fixed number of Fermion species is stable.

In counting the number of species, each spin state of a type of particle must be counted separately, for the antisymmetry of the spatial wavefunction holds only

[2] D. Ruelle, Helv. Phys. Acta. **36**, 183; **36**, 789 (1963).
[3] M. E. Fisher, Arch. Ratl. Mech. Anal. **17**, 377 (1964).
[4] L. L. Foldy, Phys. Rev. **124**, 649 (1961).
[5] M. Girardeau and G. Arnowitt, Phys. Rev. **113**, 755 (1959); M. Girardeau, *ibid.* **127**, 1809 (1962); J. M. Stephen, Proc. Phys. Soc. (London) **79**, 994 (1962); W. H. Bassichis and L. L. Foldy, Phys. Rev. **133**, A935 (1964); W. H. Bassichis, *ibid.* **134**, A543 (1964). Another paper concerned with the stability problem, with a point of view closer to ours is: E. Teller, Rev. Mod. Phys. **34**, 627 (1962).
[6] F. J. Dyson (to be published).

between particles of the same type and spin quantum number. Note that the constants A appearing in (2.3), (2.4), (2.6), and (2.7) below are not the same.

Theorem 4 falls short in two ways of what we need in a theorem establishing the stability of matter. First, it ought not require that all particles be fermions. The statistics of the nuclei are irrelevant to stability. Therefore the hypothesis that only particles of one sign of charge (say negative) are fermions should be sufficient. Second, it is an empirical fact that all chemical binding and cohesive energies are determined by the Rydberg constant $\mathrm{Ry} = me^4/2\hbar^2$ formed with the electron mass and not the nuclear mass. Stability should be independent of the nuclear mass and should persist even if the nuclear mass is taken infinite. Both of these defects are removed in our final theorem.

Theorem 5: Let N negatively charged particles belong to q different fermion species. Let their masses and charges be subject to (1.3) and (1.4), respectively. Let an arbitrary number of positively charged particles be subject to the sole restriction (1.4) on their charges, their statistics and their masses being arbitrary. Then

$$E_{\min} > -Aq^{\frac{2}{3}}N \,\mathrm{Ry}, \tag{2.7}$$

where A is an absolute constant.

In this theorem there are no unnecessary hypotheses. However, its proof is longer and more difficult than those of the others. In this paper we prove only Theorems 1–4 and delay Theorem 5 to a separate paper. It turns out that the proof of Theorem 5 requires all the preliminary results needed for the proofs of the earlier theorems, and a number of additional ones besides. Because of its fundamental significance, it would be desirable to simplify the proof of Theorem 5. We hope that this is possible by using ideas different from ours.

We may remark that the dependence of Theorems 4 and 5 on the number q of fermion species is probably not the best possible. The results stated should hold with the exponent $\frac{2}{3}$ replaced by $\frac{2}{5}$. For some discussion of this point the reader is referred to Ref. 6.

3. PROOFS OF THEOREMS 1 AND 2

The following simple argument is due to Fisher and Ruelle.[1] Write the Hamiltonian (1.1) in the form

$$\begin{aligned} H_N &= \sum_{1\le i<j\le N}\sum\Big[-\frac{\hbar^2}{2m_i(N-1)}\Delta_i \\ &\quad -\frac{\hbar^2}{2m_j(N-1)}\Delta_j+\frac{e_ie_j}{|\mathbf{r}_i-\mathbf{r}_j|}\Big] \\ &= \sum_{1\le i<j\le N}\sum H_{ij}. \end{aligned} \tag{3.1}$$

The operator H_{ij} is the Hamiltonian of a two-particle system with charges e_i, e_j and masses $m_i(N-1)$, $m_j(N-1)$. We have then

$$E_{\min} = \mathrm{Inf}\,(\psi, H_N\psi) \ge \sum_{1\le i<j\le N}\sum \mathrm{Inf}\,(\psi, H_{ij}\psi), \tag{3.2}$$

$$\mathrm{Inf}\,(\psi, H_{ij}\psi) = \begin{cases} -\dfrac{(N-1)m_im_j}{m_i+m_j}\dfrac{e_i^2e_j^2}{2\hbar^2} & (e_ie_j<0), \\ 0 & (e_ie_j\ge 0). \end{cases} \tag{3.3}$$

Among the pairs (i,j) there are at most $\frac{1}{4}N^2$ for which $e_ie_j<0$, and for these

$$\frac{(N-1)m_im_j}{m_i+m_j}\frac{e_i^2e_j^2}{2\hbar^2} \le \frac{(N-1)me^4}{4\hbar^2} = \frac{N-1}{2}\,\mathrm{Ry}. \tag{3.4}$$

This proves Theorem 1.

The proof of Theorem 2 is slightly more complicated. We now write

$$H_N = \sum_{1\le i<j\le N}\sum H_{ij} + \sum_{1\le i<j\le N}\sum H'_{ij}, \tag{3.5}$$

where

$$H_{ij} = -\frac{\hbar^2}{2m_i(N-1)}\Delta_i - \frac{\hbar^2}{2m_j(N-1)}\Delta_j + \frac{e_ie_j}{|\mathbf{r}_i-\mathbf{r}_j|}e^{-\mu|\mathbf{r}_i-\mathbf{r}_j|} \tag{3.6}$$

and

$$H'_{ij} = (e_ie_j/|\mathbf{r}_i-\mathbf{r}_j|)(1-e^{-\mu|\mathbf{r}_i-\mathbf{r}_j|}), \tag{3.7}$$

and μ is a positive number. We need a lemma which asserts that a particle in a Yukawa potential cannot have negative energy if the range of the potential is short enough.

Lemma 1: The one-particle Hamiltonian

$$H = -(\hbar^2/2m)\Delta - (e^2/r)e^{-\mu r} \tag{3.8}$$

is nonnegative if

$$\mu\hbar^2/me^2 \ge \sqrt{2}. \tag{3.9}$$

Thus if we choose

$$\mu = (N-1)me^2/\sqrt{2}\hbar^2 \tag{3.10}$$

all H_{ij} are nonnegative operators. For the second sum in (3.5) we write

$$\sum_{1\le i<j\le N}\sum H'_{ij} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{e_ie_j}{|\mathbf{r}_i-\mathbf{r}_j|} \times (1-e^{-\mu|\mathbf{r}_i-\mathbf{r}_j|}) - \tfrac{1}{2}\mu\sum_{j=1}^{N}e_j^2. \tag{3.11}$$

By Fourier transformation the double sum may be written in the manifestly positive form

$$\frac{1}{2\pi^2}\int d^3k\left(\frac{1}{k^2}-\frac{1}{k^2+\mu^2}\right)\left|\sum_{j=1}^{N}e_je^{i\mathbf{k}\cdot\mathbf{r}_j}\right|^2 > 0. \tag{3.12}$$

Hence we have from (3.10)

$$E_{\min} > -\tfrac{1}{2}\mu Ne^2 = -[N(N-1)/\sqrt{2}]\ \text{Ry}. \quad (3.13)$$

It remains to prove Lemma 1. We write the energy in momentum representation

$$(\psi, H\psi) = \frac{\hbar^2}{2m}\int d^3k k^2\,|\tilde{\psi}(\mathbf{k})|^2 - \frac{e^2}{2\pi^2}\int d^3k \int d^3k' \frac{\tilde{\psi}^*(\mathbf{k})\tilde{\psi}(\mathbf{k}')}{\mu^2 + |\mathbf{k}-\mathbf{k}'|^2}, \quad (3.14)$$

where $\tilde{\psi}(\mathbf{k})$ is the Fourier transform of the wave-function. By the Schwarz inequality we have

$$\left|\int d^3k \int d^3k' \frac{\tilde{\psi}^*(\mathbf{k})\tilde{\psi}(\mathbf{k}')}{\mu^2 + |\mathbf{k}-\mathbf{k}'|^2}\right| \le \int d^3k k^2\,|\tilde{\psi}(\mathbf{k})|^2\, J^{\frac{1}{2}} \quad (3.15)$$

with

$$J = \int d^3k \int d^3k' \frac{1}{|\mathbf{k}|^2\,|\mathbf{k}'|^2(\mu^2 + |\mathbf{k}-\mathbf{k}'|^2)^2} = 2\pi^4/\mu^2. \quad (3.16)$$

Therefore,

$$(\psi, H\psi) \ge \left(\frac{\hbar^2}{2m} - \frac{e^2}{2\pi^2}J^{\frac{1}{2}}\right)\int d^3k k^2\,|\tilde{\psi}(\mathbf{k})|^2 \ge 0, \quad (3.17)$$

when the condition

$$\hbar^2/2m \ge (e^2/2\pi^2)\,J^{\frac{1}{2}}, \quad (3.18)$$

which is the same as (3.9), is fulfilled. This proves Lemma 1 and Theorem 2.

4. A THEOREM OF ELECTROSTATICS

We begin to work toward the proof of Theorems 3–5 by a simple consideration of electrostatics. We obtain a lower bound on the Coulomb energy of an arbitrary finite system of point charges. The resulting inequality is one of the essential tools for all that follows.

Let $\mathbf{r}_i$ $(i = 1, 2, \cdots, N)$ be points in space at which there are charges e_i. Let a_i be arbitrary positive numbers and let S_i be the spheres $|\mathbf{r} - \mathbf{r}_i| = a_i$. Suppose that each of the charges e_i is distributed uniformly over the corresponding surface S_i. This results in a surface distribution of charges, where the element of surface $d\sigma$ carries the charge $e_i d\sigma/4\pi a_i^2$ if $d\sigma$ is on S_i. If $\mathbf{E} = \mathbf{E}(\mathbf{x})$ is the electric field at the point $\mathbf{x}$, produced by this charge distribution, we have for the total energy

$$\frac{1}{8\pi}\int d^3x\,|\mathbf{E}|^2 = \frac{1}{2}\sum_{i=1}^{N}\frac{e_i}{4\pi a_i^2}\int_{S_i} d\sigma_x \times \sum_{j=1}^{N}\frac{e_j}{4\pi a_j^2}\int_{S_j} d\sigma_y \frac{1}{|\mathbf{x}-\mathbf{y}|}. \quad (4.1)$$

The double surface integral depends only on the distances $|\mathbf{r}_i - \mathbf{r}_j|$ between the centers of the spheres S_i and S_j, and on their radii a_i and a_j. For two spheres S_a and S_b, of radii a and b, respectively, whose centers are at a distance r, we write

$$\int_{S_a}\frac{d\sigma_x}{4\pi a^2}\int_{S_b}\frac{d\sigma_y}{4\pi b^2}\frac{1}{|\mathbf{x}-\mathbf{y}|} = \frac{1}{r} - \Delta(r, a, b). \quad (4.2)$$

This defines the function Δ. One finds

$$\Delta(r, a, b) = \begin{cases} \dfrac{1}{r} - \min\left(\dfrac{1}{a}, \dfrac{1}{b}\right) & (0 < r \le |a-b|), \\ \dfrac{(a+b-r)^2}{4abr} & (|a-b| \le r \le a+b), \\ 0 & (a+b \le r). \end{cases} \quad (4.3)$$

Δ is positive and monotone decreasing with r in the interval $(0, a+b)$, zero beyond it.

Let us write

$$W(\mathbf{r}, e) = \sum_{1\le i<j\le N}\sum \frac{e_i e_j}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (4.4)$$

$$U(\mathbf{r}, e, a) = -\sum_{1\le j\le N}\frac{e_j^2}{2a_j} + \sum_{1\le i<j\le N}\sum e_i e_j \Delta(|\mathbf{r}_i - \mathbf{r}_j|, a_i, a_j). \quad (4.5)$$

Theorem 6: $W(\mathbf{r}, e) > U(\mathbf{r}, e, a)$.

The proof consists in merely observing that the total electrostatic field energy (4.1) is positive, and then rewriting the right-hand side in terms of the notation (4.2). Note that whenever

$$a_i + a_j \le |\mathbf{r}_i - \mathbf{r}_j| \quad (1 \le i < j \le N), \quad (4.6)$$

we have

$$U(\mathbf{r}, e, a) = -\sum_{j=1}^{N}\frac{e_j^2}{2a_j}, \quad (4.7)$$

and the inequality $W > U$ is specially simple. The inequality in this form was used by Onsager in a little known paper[7] in which he established an additive lower bound for the Coulomb energy of a system in which the particles are assumed to possess hard cores. Indeed, if it is required that

$$|\mathbf{r}_i - \mathbf{r}_j| \ge 2a \quad (1 \le i < j \le N) \quad (4.8)$$

for some fixed positive a, one has

$$W > -N(e^2/2a) \quad (e = \max|e_i|). \quad (4.9)$$

This observation was also made by Fisher and Ruelle.[1]

In our work where there are no *a priori* given hard

[7] L. Onsager, J. Phys. Chem. 43, 189 (1939).

cores it is essential to keep the a_j variable. Indeed, the power of Theorem 6 lies largely in the freedom with which the a_j may be chosen.

One useful choice is $a_j = \frac{1}{2}R_j$, where

$$R_j = \min_{(1\le i\le N, i\neq j)} |\mathbf{r}_i - \mathbf{r}_j| \,. \tag{4.10}$$

Then evidently (4.6) is fulfilled. Thus we have

Theorem 7:

$$W > -\sum_{j=1}^{N} \frac{e_j^2}{R_j} \,. \tag{4.11}$$

In this paper we use Theorem 6 only in the form of Theorem 7. The right-hand side of (4.11) may be interpreted physically as the potential energy of a fictitious system in which each particle is attracted by a Coulomb force *to its nearest neighbor alone*. The fictitious system always has a potential energy lower than the real Coulomb system, and—what is most essential—the number of terms out of which this fictitious potential energy is made up is N and not of the order of N^2 as for the true energy.

5. PROOF OF THEOREM 3

Let $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N$ be N distinct points in space. For a fixed j we write $R_{j1}, R_{j2}, \cdots, R_{jN-1}$ for the $N-1$ distances $|\mathbf{r}_j - \mathbf{r}_1|, |\mathbf{r}_j - \mathbf{r}_2|, \cdots, |\mathbf{r}_j - \mathbf{r}_N|$ arranged in increasing order. Thus R_{j1} [the same as R_j defined by (4.10) above] is the distance between $\mathbf{r}_j$ and its nearest neighbor among the other points, R_{j2} is the distance between it and its second nearest neighbor, and so on. Conventionally we define $R_{jl} = \infty$ for $l \geq N$. The R_{jl} are well-defined functions of the N variable points $R_{jl} = R_{jl}(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$.

Suppose we consider a quantum-mechanical system of N particles in a state described by the wavefunction $\psi = \psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$ normalized in the usual way

$$(\psi, \psi) = \int \cdots \int d^{3N}r \, |\psi|^2 = 1. \tag{5.1}$$

Let us introduce the following quantities:

$$K_l = \frac{1}{N} \int \cdots \int d^{3N}r \, |\psi|^2 \sum_{j=1}^{N} R_{jl}^{-1} \quad (l = 1, 2, \cdots). \tag{5.2}$$

By definition of the R_{jl} we have

$$K_1 \geq K_2 \geq K_3 \geq \cdots \geq 0, \tag{5.3}$$

and $K_l = 0$ for $l \geq N$. The K_l have the dimension of an inverse distance; K_l^{-1} is a measure of the typical linear dimension of regions which contain $l+1$ (but no more) particles.

The quantity K_1 is particularly important in connection with the inequality (4.11). From it we immediately see the following fact: If the charges of a finite system of particles satisfy (1.4), then the total Coulomb energy satisfies the inequality

$$(\psi, W\psi) > -Ne^2K_1. \tag{5.4}$$

On the other hand, if the masses satisfy (1.3) we have for the kinetic energy

$$(\psi, T\psi) = \sum_{j=1}^{N} \frac{\hbar^2}{2m_j} \int \cdots \int d^{3N}r \, |\nabla_j \psi|^2 \geq N \frac{\hbar^2}{2m} t, \tag{5.5}$$

where

$$t = \frac{1}{N} \sum_{j=1}^{N} \int \cdots \int d^{3N}r \, |\nabla_j \psi|^2. \tag{5.6}$$

Thus the total energy satisfies

$$(\psi, H_N\psi) > N[(\hbar^2/2m)t - e^2K_1]. \tag{5.7}$$

Our aim is to derive an inequality involving both t and K_1 which allows the establishment of a lower bound on the right-hand side of (5.7) independent of both.

We begin by deriving an upper bound on the cumulative sum

$$\sum_{l=1}^{k} K_l \tag{5.8}$$

in terms of K_{k+1} and t. By definition, the sum (5.8) may be written out in detail as follows:

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} \int d^3r_i \int d^3r_j \sum_{P} \int_{\text{in}} d^3r_{\alpha_1} \cdots \int_{\text{in}} d^3r_{\alpha_k} \int_{\text{out}} d^3r_{\beta_1} \cdots \int_{\text{out}} d^3r_{\beta_{N-2-k}} \sum_{l=1}^{k} \frac{|\psi|^2}{|r_{\alpha_l} - r_i|}. \tag{5.9}$$

P is a partition of the set of $N-2$ integers $\{1, 2, \cdots, i-1, i+1, \cdots, j-1, j+1, \cdots, N\}$ into two sets $\{\alpha_1, \cdots, \alpha_k\}$ and $\{\beta_1, \cdots, \beta_{N-2-k}\}$, one containing k integers, the other $N-2-k$ integers (k being fixed). The sum over P runs over all such partitions. The phrase "in" under the integration signs means that the domain of integration is

$$|\mathbf{r}_{\alpha_l} - \mathbf{r}_i| < |\mathbf{r}_j - \mathbf{r}_i| \quad (l = 1, 2, \cdots, k), \tag{5.10}$$

while "out" means the opposite

$$|\mathbf{r}_{\beta_l} - \mathbf{r}_i| \geq |\mathbf{r}_j - \mathbf{r}_i| \quad (l = 1, 2, \cdots, N-2-k). \tag{5.11}$$

In other words: $\mathbf{r}_j$ is the $(k+1)$st nearest neighbor, and $\mathbf{r}_{\alpha_1}, \mathbf{r}_{\alpha_2}, \cdots, \mathbf{r}_{\alpha_k}$ are (in some order) the first, second, $\cdots$, (k)th nearest neighbors of the point $\mathbf{r}_i$. We now make use of the following.

Lemma 2: For any positive λ, and any complex valued function $\Psi(\mathbf{r})$, having continuous derivatives and defined in the sphere Ω: $|\mathbf{r}| \leq b$,

$$\int_\Omega d^3r \frac{|\Psi|^2}{|\mathbf{r}|} < \left(\frac{1}{\lambda} + \frac{3}{2b}\right) \int_\Omega d^3r\, |\Psi|^2 + \frac{\lambda}{4} \int_\Omega d^3r\, |\nabla\Psi|^2. \tag{5.12}$$

The proof of Lemma 2 is given later. The inequality (5.12) is applied to the integration over the variable $\mathbf{r}_{\alpha_l}$ (to be carried out first). The sphere Ω is given by (5.10), with the radius $b = |\mathbf{r}_j - \mathbf{r}_i|$ and center $\mathbf{r}_i$. It follows that an upper bound for (5.8) is obtained if we replace the integrand in (5.9) by

$$|\psi|^2 \left(\frac{1}{\lambda} + \frac{3}{2} \frac{1}{|\mathbf{r}_j - \mathbf{r}_i|}\right) + \frac{\lambda}{4} |\nabla_{\alpha_l}\psi|^2. \tag{5.13}$$

For the first two terms the sum over l in (5.9) gives merely k equal integrals, so that for these one obtains

$$(k/\lambda) + \tfrac{3}{2}kK_{k+1}. \tag{5.14}$$

The gradient term may be rewritten

$$\frac{\lambda}{4N} \sum_{i=1}^{N} \int \cdots \int d^{3N}r \sum_\alpha{}' |\nabla_\alpha\psi|^2, \tag{5.15}$$

where the prime on the summation sign indicates that only those values of α are to be included in the sum for which $\mathbf{r}_\alpha$ is the (l)th nearest neighbor of $\mathbf{r}_i$ with $1 \leq l \leq k$. (The set of these values of α is a function of the integration variables $\mathbf{r}_1, \cdots, \mathbf{r}_N$, of course.) This, in turn, may be written

$$\frac{\lambda}{4N} \sum_{\alpha=1}^{N} \int \cdots \int d^{3N}r M_{\alpha k} |\nabla_\alpha\psi|^2, \tag{5.16}$$

where $M_{\alpha k} = M_{\alpha k}(\mathbf{r}_1, \cdots, \mathbf{r}_N)$ is the number of those $\mathbf{r}_i$ to which $\mathbf{r}_\alpha$ is the (l)th nearest neighbor with $1 \leq l \leq k$.

Lemma 3: For any finite set of points $\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N\}$ and $\alpha = 1, 2, \cdots, N$,

$$M_{\alpha k} < (4\pi/\omega)k < 15k, \tag{5.17}$$

where

$$\omega = 2\pi(1 - \cos \tfrac{1}{6}\pi) = \pi(2 - \sqrt{3}) \tag{5.18}$$

is the solid angle inside a circular cone of half-angle $\frac{1}{6}\pi$.

This Lemma is a purely geometrical fact which is proved later. Since trivially $M_{\alpha k} \leq N - 1$, we have now the upper bound for (5.16)

$$\tfrac{1}{4}\lambda t \min \{N - 1, (4\pi/\omega)k\} \tag{5.19}$$

in the notation (5.6). Thus we have obtained the following inequalities

$$\sum_{l=1}^{k} K_l < \frac{3k}{2} K_{k+1} + \frac{k}{\lambda} + \frac{\lambda t}{4} \min \left\{N - 1, \frac{4\pi}{\omega} k\right\} \quad (k = 1, 2, \cdots, N - 1). \tag{5.20}$$

For $k = N - 1$ one has to set $K_N = 0$ (the proof makes use of Lemma 2 with Ω all space and $b = \infty$).

Lemma 4: Let the sequence of nonnegative numbers $x_1, x_2, \cdots$ satisfy

$$\sum_{l=1}^{k} x_l < a_k x_{k+1} + b_k \quad (k = 1, 2, \cdots), \tag{5.21}$$

where the coefficients a_k and b_k are nonnegative. Then

$$x_1 \leq A_k x_{k+1} + B_k \quad (k = 1, 2, \cdots), \tag{5.22}$$

where

$$A_k = a_k \prod_{j=1}^{k-1} \frac{a_j}{1 + a_j} \tag{5.23}$$

and

$$B_k = \sum_{l=1}^{k-1} \frac{b_l}{1 + a_l} \prod_{j=1}^{l-1} \frac{a_j}{1 + a_j} + b_k \prod_{j=1}^{k-1} \frac{a_j}{1 + a_j}. \tag{5.24}$$

In the last two equations empty sums are interpreted as zero and empty products as unity. The proof is given later.

We use Lemma 4 to eliminate $K_2, K_3, \cdots, K_k$ from (5.20) and obtain a single inequality which involves only K_1, K_{k+1}, and t. Indeed, (5.20) is precisely of the form (5.21) with

$$a_k = \tfrac{3}{2}k \tag{5.25}$$

and

$$b_k = (k/\lambda) + \tfrac{1}{4}\lambda t \min \{N - 1, (4\pi/\omega)k\}. \tag{5.26}$$

We have then

$$A_k = \prod_{j=1}^{k} \frac{3j}{3j - 1} = \frac{k!\, \Gamma(\frac{2}{3})}{\Gamma(k + \frac{2}{3})}. \tag{5.27}$$

A simple upper bound on A_k is obtained by noting

$$[3j/(3j - 1)]^3 < (3j + 1)/(3j - 2), \tag{5.28}$$

so that

$$A_k < (3k + 1)^{\frac{1}{3}}. \tag{5.29}$$

In particular

$$A_{N-1} < (3N)^{\frac{1}{3}}. \tag{5.30}$$

The computation of B_k is more complicated due to the two different analytic expressions involved in (5.26). We temporarily ignore this complication and set simply

$$b_k = k[(1/\lambda) + (\pi/\omega)\lambda t]. \tag{5.31}$$

(The inequality will be somewhat worse but the calculation is easier.) One finds with (5.25) and (5.31) the identity

$$B_k = 2(A_k - 1)[(1/\lambda) + (\pi/\omega)\lambda t]. \tag{5.32}$$

In particular, using (5.30)

$$B_{N-1} < (3N)^{\frac{1}{3}}[(2/\lambda) + (2\pi/\omega)\lambda t]. \tag{5.33}$$

We now write down the inequality which follows from (5.20) by Lemma 4 for the case $k = N - 1$

$$K_1 < (3N)^{\frac{1}{3}}[(2/\lambda) + (2\pi/\omega)\lambda t], \tag{5.34}$$

or equivalently (since λ is arbitrary)

$$K_1 < 4(3N)^{\frac{1}{3}}(\pi t/\omega)^{\frac{1}{2}}. \tag{5.35}$$

We are now ready to complete the proof of Theorem 3. From (5.7) and (5.35) we have

$$(\psi, H_N\psi) > N(\hbar^2/2m)t - 4e^2N(3N)^{\frac{1}{3}}[(\pi/\omega)t]^{\frac{1}{2}} \geq -AN^{\frac{5}{3}}\,\mathrm{Ry} \tag{5.36}$$

with

$$A = (16\pi/\omega)3^{\frac{2}{3}} = 124.2\cdots. \tag{5.37}$$

The last inequality in (5.36) arises by minimizing with respect to t.

A lower value of A can be obtained by using (5.26) instead of (5.31) to compute B_{N-1}. We find $A < 52$. However, the exponent $\frac{5}{3}$ cannot be improved. The latter originates in the factor $\frac{3}{2}$ in (5.25) and that goes back to the factor $\frac{3}{2}$ on the right-hand side of (5.12) in Lemma 2. The inequality in (5.12) can be made to approach equality with arbitrary precision, as the example $\Psi = \text{const}$ and

$$\int_\Omega d^3r\,\frac{|\Psi|^2}{|\mathbf{r}|} = \frac{3}{2b}\int_\Omega d^3r\,|\Psi|^2 \tag{5.38}$$

shows. It is clear that no constant larger than $\frac{3}{2}$ would do.

It is also easy to see that as long as we use not the true Coulomb energy W but rather the lower estimate given by Theorem 7, it is impossible to improve on the exponent $\frac{5}{3}$ of Theorem 3. For we can exhibit a sequence of states ψ_N such that

$$(\psi_N, [T + U]\psi_N) \sim -AN^{\frac{5}{3}}\,\mathrm{Ry} \tag{5.39}$$

as $N \to \infty$. Take wavefunctions of the form

$$\psi_N(\mathbf{r}_1, \cdots, \mathbf{r}_N) = \prod_{j=1}^{N} u_\Lambda(\mathbf{r}_j), \tag{5.40}$$

where $u_\Lambda(\mathbf{r})$ is a smooth wave packet of spatial extent Λ. The energy is about

$$N\left\{\frac{\hbar^2}{2m}\frac{1}{\Lambda^2} - e^2\frac{N^{\frac{1}{3}}}{\Lambda}\right\}, \tag{5.41}$$

because in the absence of correlations the nearest-neighbor distance is about the mean interparticle distance $\Lambda N^{-(\frac{1}{3})}$. If N is taken large and $\Lambda = \Lambda(N)$ is taken to minimize (5.41) one obtains (5.39). Therefore *a significant improvement over our Theorem 3 can be achieved only by giving up the use of Theorem 7.*

6. PROOFS OF LEMMAS 2, 3, AND 4

In order to complete the proof of Theorem 3 we now have to prove the three lemmas used in the last section.

We begin with Lemma 2. Suppose first that Ω is an arbitrary region and $V(r)$ an arbitrary potential. The ground-state energy ϵ of a particle of mass $(2\hbar^2/\lambda)$ in this potential is defined by

$$\epsilon = \mathrm{Inf}\left\{\int_\Omega d_3r(\tfrac{1}{4}\lambda\,|\nabla\Psi|^2 + V\,|\Psi|^2)\Big/\int_\Omega d_3r\,|\Psi|^2\right\}, \tag{6.1}$$

where the infimum is taken over all wavefunctions Ψ defined in Ω. No boundary condition is imposed on Ψ, but the minimizing Ψ satisfies the "natural" condition

$$(\mathbf{n}\cdot\nabla)\Psi = 0 \tag{6.2}$$

on the boundary of Ω. The eigenvalue equation for ϵ is

$$-\tfrac{1}{4}\lambda\nabla^2\Psi + V\Psi = \epsilon\Psi. \tag{6.3}$$

Since the minimizing Ψ is positive and nonzero in Ω, we may introduce the vector

$$\boldsymbol{\omega} = -(\nabla\Psi/\Psi),$$

so that (6.3) becomes

$$\epsilon = V + \tfrac{1}{4}\lambda(\mathrm{div}\ \boldsymbol{\omega} - \boldsymbol{\omega}^2). \tag{6.4}$$

Taking the gradient of (6.4), we find

$$(\boldsymbol{\omega}\cdot\nabla)\boldsymbol{\omega} = \tfrac{1}{2}\nabla^2\boldsymbol{\omega} + (2/\lambda)\nabla V, \tag{6.5}$$

an equation identical with the Navier–Stokes equation for steady flow of a fluid with velocity $\boldsymbol{\omega}$ and with kinematical viscosity equal to $\frac{1}{2}$. We do not pursue this peculiar hydrodynamical analogy any further (see note added in proof). Integrating (6.4) over the volume Ω, we obtain

$$\epsilon = \langle V\rangle_{\mathrm{av}} - \tfrac{1}{4}\lambda\langle\boldsymbol{\omega}^2\rangle_{\mathrm{av}}, \tag{6.6}$$

where $\langle\ \rangle_{\mathrm{av}}$ denotes an average over Ω, and the term in (div $\boldsymbol{\omega}$) has vanished by virtue of (6.2).

We apply this analysis to the special case of a Coulomb potential

$$V(r) = -r^{-1}$$

in a spherical shell Ω defined by $a \leq |r| \leq b$. In this case

$$\langle V\rangle_{\mathrm{av}} = -\tfrac{3}{2}[(b^2 - a^2)/(b^3 - a^3)] > -3/2b. \tag{6.7}$$

The conclusion (5.12) of Lemma 2 states that

$$\epsilon > -(3/2b) - (1/\lambda) \tag{6.8}$$

for the spherical region $|r| < b$. If (6.8) holds for the shell $a \leq |r| \leq b$, then Lemma 2 follows by taking

the limit $a \to 0$. By (6.6) and (6.7), we have only to prove

$$\langle \boldsymbol{\omega}^2 \rangle_{\text{av}} \leq (4/\lambda^2) \tag{6.9}$$

for the spherical shell Ω.

For a spherically symmetrical Ω, the ground-state Ψ is spherically symmetric, and the vector $\boldsymbol{\omega}$ is parallel to r. We denote by ω the component of the vector $\boldsymbol{\omega}$ in the radial direction. Then (6.5) becomes

$$\omega'' + 2\omega'[(1/r) - \omega] + (2/r^2)[(2/\lambda) - \omega] = 0, \tag{6.10}$$

where the prime denotes differentiation with respect to r, and the boundary condition (6.2) gives

$$\omega(a) = \omega(b) = 0. \tag{6.11}$$

If $\omega(r)$ were ever negative in $a \leq r \leq b$, there would be at least one minimum with

$$\omega'' \geq 0, \quad \omega' = 0, \quad \omega < 0,$$

which contradicts (6.10). If $\omega(r)$ were ever greater than $(2/\lambda)$, there would be at least one maximum with

$$\omega'' \leq 0, \quad \omega' = 0, \quad \omega > (2/\lambda),$$

again contradicting (6.10). Therefore

$$0 \leq \omega(r) \leq (2/\lambda) \quad \text{for} \quad a \leq |r| \leq b, \tag{6.12}$$

which proves (6.9) and also Lemma 2.

Lemma 3 deals with a geometrical property of a finite set of points in space. Let this set be $\{\mathbf{r}_0, \mathbf{r}_1, \cdots, \mathbf{r}_n\}$. We distinguish a point, $\mathbf{r}_0$ say, and attach an *index* to each of the rest of them. The index of $\mathbf{r}_i$ is said to be the integer l if $\mathbf{r}_0$ is the (l)th nearest neighbor of $\mathbf{r}_i$ in the given set. Let now $k \geq 1$ be fixed, and define a certain subset, say $\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m\}$, consisting of all those points whose indices do not exceed k. We want to prove

$$m \leq (4\pi/\omega)k, \tag{6.13}$$

with ω defined in (5.18).

Let C_θ be the circular cone with vertex at $\mathbf{r}_0$, half-angle $\frac{1}{6}\pi$ and axis pointing in the direction θ. Let $\nu = \nu(\theta)$ be the number of points among $\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m\}$ which are inside C_θ. We have

$$\int d\Omega_\theta \nu(\theta) = m\omega, \tag{6.14}$$

where the integration is over the solid angle element formed by varying θ. Thus (6.13) follows if we show

$$\nu(\theta) \leq k. \tag{6.15}$$

Let now θ be fixed, and suppose for the sake of definiteness that out of $\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m\}$ the first ν points are inside C_θ. If $\nu = 0$ or 1, (6.15) is true trivially, so we may suppose $\nu \geq 2$. We choose the notation so that

$$|r_0 - \mathbf{r}_1| \leq |\mathbf{r}_0 - \mathbf{r}_2| \leq \cdots \leq |\mathbf{r}_0 - \mathbf{r}_\nu|. \tag{6.16}$$

Take an i $(1 \leq i \leq \nu - 1)$ and consider the triangle $(\mathbf{r}_0, \mathbf{r}_i, \mathbf{r}_\nu)$. Since the angle at $\mathbf{r}_0$ is $\leq \frac{1}{3}\pi$, the largest angle of this triangle must be either at $\mathbf{r}_i$ or at $\mathbf{r}_\nu$. But the latter is excluded because $|\mathbf{r}_0 - \mathbf{r}_i| \leq |\mathbf{r}_0 - \mathbf{r}_\nu|$ and in a triangle the largest side occurs opposite the largest angle. Thus the angle at $\mathbf{r}_i$ is largest and so, by the same principle,

$$|\mathbf{r}_\nu - \mathbf{r}_i| \leq |\mathbf{r}_\nu - \mathbf{r}_0|. \tag{6.17}$$

Since this is true of $i = 1, 2, \cdots, \nu - 1$, $\mathbf{r}_0$ cannot be less than the (ν)th nearest neighbor of $\mathbf{r}_\nu$ or, in other words, the index of $\mathbf{r}_\nu$ is at least ν. By assumption this index does not exceed k, therefore $\nu \leq k$ which is what was to be shown. This completes the proof of Lemma 3.

We may remark that the numerical factor $4\pi/\omega = 8 + 4\sqrt{3} = 14.928\cdots$ in Lemma 3 is close to the best possible (if indeed not the best). To see this we display a set of $n = 12k + 1$ points such that $12k$ of them possess the index k. Choose one point $\mathbf{r}_0$ at the center of a regular icosahedron and the rest of them, $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_n$ in groups of k very close to the 12 vertices. Since the edge of an icosahedron exceeds the distance of its center from its vertices, the center $\mathbf{r}_0$ is the (k)th nearest neighbor to all $12k$ points. Thus the best constant of Lemma 3 must be ≥ 12.

To prove Lemma 4 we choose a fixed $k \geq 2$ (the case $k = 1$ is trivial), and define coefficients h_l as follows

$$\begin{cases} h_l = \dfrac{1}{1 + a_l} \displaystyle\prod_{j=1}^{l-1} \frac{a_j}{1 + a_j} \quad (l = 1, 2, \cdots, k - 1), \\ h_k = \displaystyle\prod_{j=1}^{k-1} \frac{a_j}{1 + a_j}. \end{cases} \tag{6.18}$$

These quantities satisfy

$$\sum_{j=1}^{k} h_j = 1,$$

$$\sum_{j=l}^{k} h_j = a_{l-1} h_{l-1} \quad (l = 2, 3, \cdots, k). \tag{6.19}$$

Now multiply the first k of the inequalities (5.21) by $h_1, h_2, \cdots, h_k$ respectively and add (this is valid because $h_l \geq 0$). The inequality which results is just (5.22) with A_k and B_k given by (5.23) and (5.24).

7. PROOF OF THEOREM 4

We now assume all particles are fermions and that they fall into $q \geq 1$ groups so that the exclusion

principle holds between particles of the same group. No assumption is made about the number in each group except that the total number is $N \geq q+1$.

We make use of the antisymmetry of the wavefunction of identical fermions only by the application of the following inequality.

Lemma 5: Let $\Psi = \Psi(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_\nu)$ be a function of $\nu \geq 2$ space points having continuous first derivatives, antisymmetric with respect to interchange of any two points, and defined with all points in a sphere Ω of radius λ. Then

$$\int_\Omega d^{3\nu}x \sum_{i=1}^{\nu} |\nabla_i \Psi|^2 \geq (\nu - 1)\frac{\xi^2}{\lambda^2}\int_\Omega d^{3\nu}x\, |\Psi|^2, \quad (7.1)$$

where $\xi = 2.082$ is the smallest positive root of the equation

$$(d^2/dx^2)(\sin x/x) = 0. \quad (7.2)$$

For simplicity take $\nu = 2$ (the proof for $\nu \geq 3$ is analogous). Expand

$$\Psi = \Psi(\mathbf{x}, \mathbf{y}) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} C_{n,m} u_n(\mathbf{x}) u_m(\mathbf{y}) \quad (7.3)$$

in terms of the complete orthonormal set of eigenfunctions $\{u_n(\mathbf{x})\}$ defined by the eigenvalue problem

$$-\Delta_x u_n(\mathbf{x}) = \epsilon_n u_n(\mathbf{x}) \quad (7.4)$$

with the boundary conditions

$$\frac{\partial}{\partial r} u_n(\mathbf{x}) = 0 \quad \text{for} \quad r = |\mathbf{x}| = \lambda. \quad (7.5)$$

One finds

$$\int_\Omega d^3x \int_\Omega d^3y\, |\nabla_x \Psi|^2 = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \epsilon_n |C_{n,m}|^2 = \frac{1}{2}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} (\epsilon_n + \epsilon_m)|C_{n\,m}|^2 \quad (7.6)$$

because the antisymmetry of Ψ implies $C_{n,m} = -C_{m,n}$. Also

$$\int_\Omega d^3x \int_\Omega d^3y\, |\Psi|^2 = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} |C_{n,m}|^2. \quad (7.7)$$

The ratio of (7.6) to (7.7) is smallest when $C_{n,m} \neq 0$ only for those two values $n \neq m$ for which ϵ_n and ϵ_m are the two lowest eigenvalues. There is one s-state eigenvalue $\epsilon_0 = 0$ with $u_0(\mathbf{x}) = (\frac{4}{3}\pi\lambda^3)^{-\frac{1}{2}}$, and three degenerate p-state eigenvalues

$$\epsilon_1 = \epsilon_2 = \epsilon_3 = \xi^2/\lambda^2. \quad (7.8)$$

The remaining ϵ_j all lie higher than (7.8). This completes the proof of Lemma 5.

The proof of Theorem 4 is based on an inequality which is important in its own right. It involves only t and K_{p-1} for some $p \geq q+1$. It depends purely on the antisymmetry of the wavefunction and has nothing to do with the Coulomb problem as such.

Theorem 8: For a system of $N \geq q+1$ fermions belonging to $q \geq 1$ species

$$(43/\xi^2)[p/(p-q)]t \geq K_{p-1}^2 \quad (7.9)$$

for $q+1 \leq p \leq N$.

The proof of Theorem 8 is based on Lemma 5. Before proving it we show how Theorem 4 is derived from Theorem 8. Since for $q = 1$ the physical content of Theorem 4 is vacuous, we may assume $q \geq 2$. From (5.20) and Lemma 4 we derive, using (5.29) and (5.32),

$$K_1 < (3p)^{\frac{1}{3}}[K_{p-1} + (2/\lambda) + (2\pi/\omega)\lambda t]. \quad (7.10)$$

The inequality (7.9) may be rewritten in the alternate form

$$\mu\frac{43}{8\xi^2}\frac{p}{p-q}t - K_{p-1} + \frac{2}{\mu} \geq 0, \quad (7.11)$$

where μ is an arbitrary positive number. From (7.10) and (7.11) we eliminate K_{p-1}, obtaining

$$K_1 < (3p)^{\frac{1}{3}}\left[\left(\frac{2\pi}{\omega}\lambda + \frac{43}{8\xi^2}\frac{p}{p-q}\mu\right)t + 2\left(\frac{1}{\lambda} + \frac{1}{\mu}\right)\right]. \quad (7.12)$$

Comparing with (5.34) we observe that (7.12) is a weaker inequality when $p > N$. Thus we may ignore the restriction $p \leq N$ given in Theorem 8 and choose

$$p = 2q \quad (7.13)$$

for any $N \geq q+1$. Finally, λ and μ are chosen to minimize the right-hand side of (7.12). This results in the following.

Corollary to Theorem 8: Under the conditions given in the theorem,

$$K_1 < Aq^{\frac{1}{3}}t^{\frac{1}{2}}, \quad (7.14)$$

with the constant

$$A = 2\cdot 6^{\frac{1}{3}}\left[\left(\frac{4\pi}{\omega}\right)^{\frac{1}{2}} + \left(\frac{43}{2\xi^2}\right)^{\frac{1}{2}}\right] = 22\,.\,2\cdots. \quad (7.15)$$

The proof of Theorem 4 is now completed by using (7.14) in (5.7)

$$\begin{aligned}(\psi, H_N\psi) &> N[(\hbar^2/2m)t - e^2K_1] \\ &> N[(\hbar^2/2m)t - Aq^{\frac{1}{3}}e^2t^{\frac{1}{2}}] \\ &\geq -A^2q^{\frac{2}{3}}N\ \text{Ry},\end{aligned} \quad (7.16)$$

with $A^2 < 500$ by (7.15).

8. PROOF OF THEOREM 8

We begin by introducing an arbitrary length λ and writing

$$t = \left(\frac{4\pi\lambda^3}{3}\right)^{-1} \int d^{3N}r \, \frac{1}{N} \sum_{j=1}^{N} |\nabla_j \psi|^2 \int_{|r_j - y| \le \lambda} d^3y. \quad (8.1)$$

If the order of the integrations over the $\mathbf{r}_j$ and over $\mathbf{y}$ is interchanged, this becomes

$$t = \left(\frac{4\pi\lambda^3}{3}\right)^{-1} \frac{1}{N} \int d^3y \sum_P \int_{\text{out}} d^3r_{j_1} \cdots \int_{\text{out}} d^3r_{j_{N-\nu}} \times \int_{\text{in}} d^3r_{i_1} \cdots \int_{\text{in}} d^3r_{i_\nu} [|\nabla_{i_1}\psi|^2 + \cdots + |\nabla_{i_\nu}\psi|^2]. \quad (8.2)$$

The summation is over all partitions P of the set of subscripts $\{1, 2, \cdots, N\}$ into two parts $\{i_1, i_2, \cdots, i_\nu\}$ and $\{j_1, j_2, \cdots, j_{N-\nu}\}$. The phrase "in" under an integral sign signifies that the corresponding integration variable is restricted to lie inside the sphere of radius λ around the center $\mathbf{y}$, while "out" means the opposite restriction.

We now drop all terms from the sum over P which do not satisfy

$$p \le \nu \le N, \quad (8.3)$$

where p is an arbitrary integer satisfying

$$q + 1 \le p \le N. \quad (8.4)$$

Consider now a particular P and the particles labeled $i_1, i_2, \cdots, i_\nu$ which are inside the sphere of radius λ around the center $\mathbf{y}$. Let $\nu_1, \nu_2, \cdots, \nu_q$ be the numbers among them which belong to the first, second, $\cdots$, (q)th species respectively. We apply Lemma 5 to the integration over the ν_1 variables belonging to the first species, then over the ν_2 variables belonging to the second species, and so on. Since

$$\sum_{s=1}^{q} (\nu_s - 1) = \nu - q \ge \frac{p-q}{p} \nu \quad (8.5)$$

under the restriction (8.3), we obtain

$$t \ge \left(\frac{4\pi\lambda^3}{3}\right)^{-1} \frac{1}{N} \frac{\xi^2}{\lambda^2} \sum_P{}' \frac{p-q}{p} \nu \int d^3y \int_{\text{out}} d^3r_{j_1} \cdots \int_{\text{out}} d^3r_{j_{N-\nu}} \int_{\text{in}} d^3r_{i_1} \cdots \int_{\text{in}} d^3r_{i_\nu} |\psi|^2. \quad (8.6)$$

The prime on the summation sign stands for the restriction of the sum to terms P for which $\nu = \nu(P)$ satisfies (8.3). We now restore the original order of the integration variables. This gives

$$t \ge \left(\frac{4\pi\lambda^3}{3}\right)^{-1} \frac{1}{N} \frac{\xi^2}{\lambda^2} \frac{p-q}{p} \int d^{3N}r \, |\psi|^2 \sum_{\nu=p}^{N} \nu \int_{\Omega_\nu} d^3y. \quad (8.7)$$

Here $\Omega_\nu = \Omega_\nu(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$ is the set of points $\mathbf{y}$ such that the inequality $|\mathbf{r}_i - \mathbf{y}| \le \lambda$ is true for exactly ν (and not more) values of the subscript i. We find it convenient to rewrite (8.7) somewhat differently

$$\sum_{\nu=p}^{N} \nu \int_{\Omega_\nu} d^3y = \sum_{i=1}^{N} \int_{\Xi_i} d^3y. \quad (8.8)$$

Ξ_i is a set of points $\mathbf{y}$, defined by the condition that $|\mathbf{r}_i - \mathbf{y}| \le \lambda$, and *at least* $p - 1$ *more* inequalities of the same type $|\mathbf{r}_j - \mathbf{y}| \le \lambda$ $(j \ne i)$ hold. The identity (8.8) is verified easiest after its intuitive content is grasped in terms of simple examples.

The next step is to obtain a lower bound for the volume of Ξ_i (as a function of the $\mathbf{r}_1, \cdots, \mathbf{r}_N$). It is at this point that we introduce the $(p-1)$st nearest-neighbor distance $R_{i,p-1}$ of the point $\mathbf{r}_i$. When $R_{i,p-1} \ge \lambda$ we write

$$\int_{\Xi_i} d^3y \ge 0. \quad (8.9)$$

Let then $R_{i,p-1} < \lambda$. There are precisely $p - 1$ values of j $(j \ne i)$ such that

$$|\mathbf{r}_j - \mathbf{r}_i| \le R_{i,p-1}. \quad (8.10)$$

Consider the set Ξ_i' of $\mathbf{y}$ satisfying

$$|\mathbf{y} - \mathbf{r}_i| \le \lambda - R_{i,p-1}. \quad (8.11)$$

Ξ_i' is a sphere of volume

$$\int_{\Xi_i'} d^3y = \frac{4\pi}{3} (\lambda - R_{i,p-1})^3. \quad (8.12)$$

For any $\mathbf{y}$ inside it and any j satisfying (8.11) one has $|\mathbf{y} - \mathbf{r}_j| \le \lambda$, which shows that Ξ_i' is a subset of Ξ_i. So we have

$$\int_{\Xi_i} d^3y \ge \frac{4\pi}{3} (\lambda - R_{i,p-1})^3. \quad (8.13)$$

We now take (8.13) and (8.9) into (8.8) and (8.7). This gives

$$t \ge \frac{\xi^2}{\lambda^2} \frac{p-q}{p} \frac{1}{N} \int d^{3N}r \, |\psi|^2 \sum_{i=1}^{N} \max\left\{0, \left(1 - \frac{R_{i\,p-1}}{\lambda}\right)^3\right\}. \quad (8.14)$$

This inequality holds for any positive λ. We average it over all values of λ in the interval $0 \le \lambda \le a$. We have

$$\frac{1}{a} \int_0^a \frac{d\lambda}{\lambda^2} \max\left\{0, \left(1 - \frac{R}{\lambda}\right)^3\right\} = \frac{1}{4aR}\left(1 - \frac{R}{a}\right) \max\left\{0, \left(1 - \frac{R}{a}\right)^3\right\} \ge (1/4aR) - (1/a^2). \quad (8.15)$$

Therefore (8.14) implies

$$t \geq \xi^2 \frac{p-q}{p}\left(\frac{K_{p-1}}{4a} - \frac{1}{a^2}\right). \tag{8.16}$$

The best value of a is $8(K_{p-1})^{-1}$, yielding

$$t \geq \frac{\xi^2}{64}\frac{p-q}{p} K_{p-1}^2. \tag{8.17}$$

This completes the proof of Theorem 8, except that 64 appears instead of the coefficient 43 on the left side of (7.9).

We have succeeded in deducing (8.17) with the coefficient 43, starting from (8.14). This requires only elementary but complicated manipulations which we do not present here.[8] In mathematical terms the problem is the following. Given some probability distribution function $F(t)$ on the positive real axis [F is nondecreasing and $F(0) = 0$, $F(\infty) = 1$], such that

$$\int_0^x (x-t)^3\, dF(t) \leq C^2 x^5 \tag{8.18}$$

for all positive x, where C is a constant. Write

$$K = \int_0^\infty \frac{1}{t}\, dF(t). \tag{8.19}$$

What is the best possible inequality of the type

$$K^2 \leq \alpha C^2? \tag{8.20}$$

The argument above shows $\alpha \leq 64$. Our more elaborate argument gives $\alpha \leq 43$. It is easy to see that the best α cannot be less than 40. For if $F(t) = \min\{1, 10C^2t^2\}$, then $K^2 = 40C^2$. To determine the best α is an amusing problem, but it would give only a trivial numerical improvement of Theorems 8 and 4.

9. SMOOTH BACKGROUND CHARGE

Theorem 4 can be generalized by adding a smooth external charge distribution to the N fermions. The particles now interact not only with each other but also with the field produced by this background charge. Let $\rho(\mathbf{x})$ be the charge density producing the external field. The Hamiltonian is now

$$\bar{H}_N = \sum_{j=1}^N \left(-\frac{\hbar^2}{2m_j}\Delta_j\right) + \sum\sum_{1\leq i<j\leq N} \frac{e_i e_j}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^N e_i \int d_3x \frac{\rho(\mathbf{x})}{|\mathbf{x}-\mathbf{r}_i|} + \frac{1}{2}\int d^3x \int d^3y \frac{\rho(\mathbf{x})\rho(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}. \tag{9.1}$$

The last term is a C number, the self-energy of the background charge. We assume that it is finite.

[8] For details, see Ref. 6.

Theorem 9: Suppose N particles satisfy the conditions of Theorem 4 and are subject to an external field generated by a smooth charge density with finite self-energy. Then

$$E_{\min} > -A(2q)^{\frac{2}{3}}N \text{ Ry}, \tag{9.2}$$

where A is the same constant as in Theorem 4.

To prove this we consider a fictitious system consisting of $2N$ particles, N of them having the given masses m_i and charges e_i, and the other N of them having the same masses m_i but opposite charges $-e_i$. The total number of species is $2q$. Let H'_{2N} denote the Hamiltonian of this system, which includes the kinetic energy and the Coulomb energy due to the interactions between all $2N$ charges. Consider now the energy of this system in a state Ψ defined by

$$\Psi(\mathbf{r}_1, \cdots, \mathbf{r}_{2N}) = \psi(\mathbf{r}_1, \cdots, \mathbf{r}_N)\psi(\mathbf{r}_{N+1}, \cdots, \mathbf{r}_{2N}). \tag{9.3}$$

It is

$$(\Psi, H'_{2N}\Psi) = 2(\psi, H_N\psi) - \int d^{6N}r\, |\psi(\mathbf{r}_1 \cdot \cdots, \mathbf{r}_N)|^2 \times |\psi(\mathbf{r}_{N+1}, \cdots, \mathbf{r}_{2N})|^2 \sum_{i=1}^N \sum_{j=N+1}^{2N} \frac{e_i e_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{9.4}$$

Here by H_N we mean the Hamiltonian (1.1), i.e., the energy of the first N particles alone. Theorem 4 asserts that

$$(\Psi, H'_{2N}\Psi) > -A2N(2q)^{\frac{2}{3}} \text{ Ry}. \tag{9.5}$$

We compare this with the expectation value of the operator $\bar{H}_N$ given by (9.1) in the state ψ.

$$(\psi, \bar{H}_N\psi) = (\psi, H_N\psi) + \int d^{3N}r\, |\psi(\mathbf{r}_1, \cdots, \mathbf{r}_N)|^2 \int d_3x \times \sum_{i=1}^N \frac{e_i\rho(\mathbf{x})}{|\mathbf{r}_i - \mathbf{x}|} + \frac{1}{2}\int d^3x \int d^3y \frac{\rho(\mathbf{x})\rho(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}. \tag{9.6}$$

Therefore we have

$$(\psi, \bar{H}_N\psi) - \tfrac{1}{2}(\Psi, H'_{2N}\Psi) = \frac{1}{2}\int d^3x \int d^3y \frac{\rho'(\mathbf{x})\rho'(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}, \tag{9.7}$$

where

$$\rho'(\mathbf{x}) = \rho(\mathbf{x}) + \int d^{3N}r\, |\psi|^2 \sum_{i=1}^N e_i\delta(\mathbf{r}_i - \mathbf{x}). \tag{9.8}$$

The integral on the right-hand side of (9.7) is nonnegative. Therefore

$$(\psi, \bar{H}_N\psi) \geq \tfrac{1}{2}\Psi, H(_{2N}'\Psi), \tag{9.9}$$

and comparing this with (9.5) the conclusion of Theorem 9 follows. Equality can occur in (9.9) only when $\rho' = 0$ identically, that is when the given background charge density exactly cancels the charge density $\int d^{3N}r\, |\psi|^2 \sum_i e_i \delta(\mathbf{r}_i - \mathbf{x})$ of the particles.

In this proof it is essential that we included the last term in (9.1), the self-energy of the background charge, in the definition of the Hamiltonian $\bar{H}_N$. Thus it is impossible to think of $\rho(\mathbf{x})$ as the (singular) charge density of a certain number of fixed point charges, for in that case the self-energy is infinite and Theorem 9 is vacuous. This consideration shows that our Theorem 5 is a significantly deeper result than Theorem 9, because it asserts the stability of a system of charged fermions in the field of fixed point charges where the energy, by definition, does not contain any self-energy term.

Note added in proof: For a deeper discussion of (6.5) see E. Nelson, Phys. Rev. **150**, 1079 (1966).

ACKNOWLEDGMENTS

The second-named author wishes to thank Professor J. R. Oppenheimer for extending to him the hospitality of the Institute during his year of absence from the Plasma Physics Laboratory of Princeton University.

Financial support from the National Science Foundation is also gratefully acknowledged.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Translational Invariance Properties of a One-Dimensional Fluid with Forces of Finite Extent

MICHAEL H. COOPERSMITH AND HARVEY S. LEFF
Department of Physics, Case Institute of Technology, Cleveland, Ohio

(Received 13 August 1966)

The translational invariance properties of a one-dimensional fluid with finite range forces are investigated. For N particles in the interval $[0, L]$, with a two-body interaction potential $w(x) = 0$ for $x \geq R$, we find the following: (a) If $w(x)$ has a hard core of diameter d and $R \leq 2d$, each n-particle distribution function $D_n(x_1, \cdots, x_n)$ is translationally invariant if and only if $L > 2(N - n)R$ and $x_1, \cdots, x_n$ lie in $[(N - n)R, L - (N - n)R]$. (b) For arbitrary finite values of R, with or without a hard core, the above conditions are sufficient for translational invariance of the D_n. These conditions hold for all temperatures.

I. INTRODUCTION

IN a recent paper[1] (referred to as I), translational invariance properties for a finite one-dimensional hard-core fluid were established. It was found that, for densities less than half the close packing density, there exists a *central* region in which the one-, two-, $\cdots$, N-particle distribution functions are translationally invariant. It is the main purpose of this paper to extend these results to one-dimensional systems with arbitrary forces of finite extent, R.

In I, use was made of the fact that, for systems with nearest-neighbor interactions, the n-particle distribution functions, D_n, are expressible in terms of the configurational partition function. For pure hard cores (no attractive forces), this function is well known, and its precise form was used explicitly throughout the investigation of paper I. In order to extend the investigation to a general class of potentials of finite extent, we employ a method which expresses derivatives of distribution functions in terms of other distribution functions. These expressions are in the form of recursion relations which lead to the translational invariance properties of the n-particle distribution functions. The bulk of this paper deals with the derivation of these recursion relations. Once obtained, the translational invariance properties are immediately established using mathematical induction.

The main result is that, for N particles contained in the interval $[0, L]$, where $L > 2(N - n)R$, there exist *central* regions, $[(N - n)R, L - (N - n)R]$, in which the functions D_n for $n = 1, \cdots, N$ are translationally invariant. It is rigorously established that, for nearest-neighbor potentials, these translational invariance properties *do not* hold outside the *central* regions and evidence that this is also true for potentials of arbitrary extent is presented. An interesting

[1] H. S. Leff and M. H. Coopersmith, J. Math. Phys. **8**, 306 (1967).

and comparing this with (9.5) the conclusion of Theorem 9 follows. Equality can occur in (9.9) only when $\rho' = 0$ identically, that is when the given background charge density exactly cancels the charge density $\int d^{3N}r\, |\psi|^2 \sum_i e_i \delta(\mathbf{r}_i - \mathbf{x})$ of the particles.

In this proof it is essential that we included the last term in (9.1), the self-energy of the background charge, in the definition of the Hamiltonian $\bar{H}_N$. Thus it is impossible to think of $\rho(\mathbf{x})$ as the (singular) charge density of a certain number of fixed point charges, for in that case the self-energy is infinite and Theorem 9 is vacuous. This consideration shows that our Theorem 5 is a significantly deeper result than Theorem 9, because it asserts the stability of a system of charged fermions in the field of fixed point charges where the energy, by definition, does not contain any self-energy term.

Note added in proof: For a deeper discussion of (6.5) see E. Nelson, Phys. Rev. **150**, 1079 (1966).

ACKNOWLEDGMENTS

The second-named author wishes to thank Professor J. R. Oppenheimer for extending to him the hospitality of the Institute during his year of absence from the Plasma Physics Laboratory of Princeton University.

Financial support from the National Science Foundation is also gratefully acknowledged.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Translational Invariance Properties of a One-Dimensional Fluid with Forces of Finite Extent

MICHAEL H. COOPERSMITH AND HARVEY S. LEFF
Department of Physics, Case Institute of Technology, Cleveland, Ohio

(Received 13 August 1966)

The translational invariance properties of a one-dimensional fluid with finite range forces are investigated. For N particles in the interval $[0, L]$, with a two-body interaction potential $w(x) = 0$ for $x \geq R$, we find the following: (a) If $w(x)$ has a hard core of diameter d and $R \leq 2d$, each n-particle distribution function $D_n(x_1, \cdots, x_n)$ is translationally invariant if and only if $L > 2(N - n)R$ and $x_1, \cdots, x_n$ lie in $[(N - n)R, L - (N - n)R]$. (b) For arbitrary finite values of R, with or without a hard core, the above conditions are sufficient for translational invariance of the D_n. These conditions hold for all temperatures.

I. INTRODUCTION

IN a recent paper[1] (referred to as I), translational invariance properties for a finite one-dimensional hard-core fluid were established. It was found that, for densities less than half the close packing density, there exists a *central* region in which the one-, two-, $\cdots$, N-particle distribution functions are translationally invariant. It is the main purpose of this paper to extend these results to one-dimensional systems with arbitrary forces of finite extent, R.

In I, use was made of the fact that, for systems with nearest-neighbor interactions, the n-particle distribution functions, D_n, are expressible in terms of the configurational partition function. For pure hard cores (no attractive forces), this function is well known, and its precise form was used explicitly throughout the investigation of paper I. In order to extend the investigation to a general class of potentials of finite extent, we employ a method which expresses derivatives of distribution functions in terms of other distribution functions. These expressions are in the form of recursion relations which lead to the translational invariance properties of the n-particle distribution functions. The bulk of this paper deals with the derivation of these recursion relations. Once obtained, the translational invariance properties are immediately established using mathematical induction.

The main result is that, for N particles contained in the interval $[0, L]$, where $L > 2(N - n)R$, there exist *central* regions, $[(N - n)R, L - (N - n)R]$, in which the functions D_n for $n = 1, \cdots, N$ are translationally invariant. It is rigorously established that, for nearest-neighbor potentials, these translational invariance properties *do not* hold outside the *central* regions and evidence that this is also true for potentials of arbitrary extent is presented. An interesting

[1] H. S. Leff and M. H. Coopersmith, J. Math. Phys. **8**, 306 (1967).

feature is that the translational invariance of D_n does not require a hard core and depends only upon the finite extent of the potential. Furthermore, the results are *independent* of the temperature.

In Sec. II, we establish a notation convention and recall several useful formulas from paper I. Section III introduces the general method of this paper and the major results of paper I are rederived. In Sec. IV, these results are extended to the case of general nearest-neighbor forces. Section V is devoted to the case of arbitrary forces of finite extent. A discussion of the results is contained in Sec. VI.

II. NOTATION AND REVIEW

Consider a one-dimensional system of $N+1$ particles contained in the interval $[0, L]$. (We use $N+1$ instead of N for algebraic convenience.) Assume the particles interact according to a two-body potential energy $w(x)$, where x is the interparticle separation. The configurational partition function for the system is

$$Z(L, N+1) = (N+1)! \times \int_{\mathcal{O}_L} \cdots \int \exp\left[-\beta \sum_{i<j}^{N+1} w(x_{ij})\right] \prod_{k=1}^{N+1} dx_k. \quad (1)$$

Here $x_{ij} = |x_i - x_j|$, β is the inverse temperature and $\mathcal{O}_L$ is the ordered domain $0 \le x_1 \le \cdots \le x_{N+1} \le L$. The n-particle distribution functions

$$D_n^{(N+1)}(x_1, \cdots, x_n \mid L)$$

are defined by[2]

$$D_n^{(N+1)}(x_1, \cdots, x_n \mid L) = \frac{(N+1)!}{(N+1-n)!} \times \frac{1}{Z(L, N+1)} \times \int_0^L \cdots \int_0^L \exp\left[-\beta \sum_{i<j}^{N+1} w(x_{ij})\right] \prod_{k=n+1}^{N+1} dx_k \quad (2)$$

if $0 \le x_i \le L$ for $i = 1, \cdots, n$ and are identically zero otherwise.

If $w(x)$ contains a hard-core part with diameter d and is zero for $x \ge 2d$, then, according to I, the ordered n-particle distribution functions have a particularly useful form

$$D_n^{(N+1)}(x_1 \le \cdots \le x_n \mid L) = \frac{1}{n!\, Z(L, N+1)} \times \sum_{(N_i)}{}' \frac{(N+1)!}{\Pi_i N_i!} \tilde{Z}(x_1, N_1) \tilde{Z}(L - x_n, N_{n+1}) \times \prod_{k=2}^{n} \bar{Z}(x_k - x_{k-1}, N_k). \quad (3)$$

[2] This notation is slightly more detailed than that in paper I, indicating both N and L dependence.

This function is normalized to

$$\binom{N+1}{n}$$

over the ordered domain $\mathcal{O}_L$. The prime indicates that the summation over the set $\{N_1, \cdots, N_{n+1}\}$ is to be carried out under the constraint

$$\sum_{j=1}^{n+1} N_j = N + 1 - n.$$

When $n = 1$, the product of $\bar{Z}$ functions does not appear. $\tilde{Z}$ and $\bar{Z}$ are modified configurational integrals which are defined by

$$\tilde{Z}(x, n) = n! \int_{\mathcal{O}_x} \cdots \int \exp\left[-\beta \sum_{i<j}^{n} w(x_{ij}) - \beta \sum_{i=1}^{n} w(|x_i - x|)\right] \prod_{k=1}^{n} dx_k, \quad (4a)$$

$$\bar{Z}(x, n) = n! \int_{\mathcal{O}_x} \cdots \int \exp\left[-\beta \sum_{i<j}^{n} w(x_{ij}) - \beta \sum_{i=1}^{n} w(|x_i - x|) - \beta \sum_{i=1}^{n} w(x_i)\right] \prod_{k=1}^{n} dx_k. \quad (4b)$$

For the special case when $w(x)$ is a pure hard-core potential, one has (see Appendix B of I)

$$Z(x, n) = [x - (n-1)d]^n, \quad \text{for} \quad x \ge (n-1)d, \quad (5a)$$

$$\tilde{Z}(x, n) = [x - nd]^n, \quad \text{for} \quad x \ge nd, \quad (5b)$$

$$\bar{Z}(x, n) = [x - (n+1)d]^n, \quad \text{for} \quad x \ge (n+1)d. \quad (5c)$$

These functions are identically zero outside the indicated domains. It should be emphasized that Eq. (3) holds only for the case of nearest-neighbor forces and *requires* that $w(x)$ contain a hard-core part.[3]

III. ONE-DIMENSIONAL PURE HARD-CORE FLUID

In this section, we rederive the main results of I without resort to the explicit evaluation of the partition function. Specifically, we prove that, for N hard-core particles of length d in the one-dimensional interval $[0, L]$, where $L > 2(N-n)d$, the distribution function $D_n^{(N)}(x_1, \cdots, x_n \mid L)$ is translationally invariant for $(N-n)d < x_i < L - (N-n)d$ and $i = 1, 2, \cdots, n$. To do this, we first look at the derivative of $D_1^{(N+1)}(x \mid L)$ with respect to x. From Eq. (3), we have

$$D_1^{(N+1)}(x \mid L) = \frac{(N+1)}{Z(L, N+1)} \times \sum_{n=\sigma(x)}^{\gamma(x)} \binom{N}{n} \tilde{Z}(x, n) \bar{Z}(L - x, N - n), \quad (6)$$

[3] The concept of nearest-neighbor forces is meaningful only if the potential has a hard-core part. The requirement of a hard core is dropped in Sec. V.

where $\sigma(x)$ and $\gamma(x)$ are given by[4]

$$\gamma(x) = \begin{cases} k & \text{for } kd < x < (k+1)d; \\ & k = 0, \cdots, N-1, \\ N & \text{for } x > Nd, \end{cases} \quad (7a)$$

$$\sigma(x) = \begin{cases} k & \text{for } L - Nd + (k-1)d < x < L \\ & \quad - (N-k)d; \; k = 1, \cdots, N, \\ 0 & \text{for } x < L - Nd. \end{cases} \quad (7b)$$

The derivative is then given by

$$\frac{\partial}{\partial x} D_1^{(N+1)}(x \mid L) = \frac{(N+1)}{Z(L, N+1)} \sum_{n=\sigma(x)}^{\gamma(x)} \binom{N}{n} \times \left[\tilde{Z}(L-x, N-n) \frac{\partial}{\partial x} \tilde{Z}(x, n) + \tilde{Z}(x, n) \frac{\partial}{\partial x} \tilde{Z}(L-x, N-n) \right] + \text{terms proportional to } \partial\gamma/\partial x \text{ and } \partial\sigma/\partial x. \quad (8)$$

If we now restrict ourselves to values of x which are not integral multiples of d, then the terms involving the derivatives of $\gamma(x)$ and $\sigma(x)$ vanish. This restriction is of no consequence since this set of points has measure zero.

It is necessary now to evaluate the derivative of $\tilde{Z}$. For the pure hard-core case, this can be done easily since we have an explicit expression for $\tilde{Z}$ as given by Eq. (5b). For $x \geq nd$,

$$(\partial/\partial x)\tilde{Z}(x, n) = n(x - nd)^{n-1} = n\tilde{Z}(x - d, n - 1). \quad (9)$$

However, this equation can be derived *without* explicit use being made of Eq. (5b). To do this, we go back to the original definition of $\tilde{Z}$ as given by Eq. (4a) applied to the nearest-neighbor case.

$$\tilde{Z}(x, n) = n! \int_{\mathcal{O}_x} \cdots \int \exp\left[-\beta \sum_{i=1}^{n-1} w(x_{i+1} - x_i) - \beta w(x - x_n) \right] \prod_{k=1}^{n} dx_k, \quad (10)$$

where $w(x)$ is the hard-core potential given by

$$w(x) = \begin{cases} \infty & \text{for } x < d, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Differentiating Eq. (10), we find

$$\begin{aligned} \frac{\partial}{\partial x} \tilde{Z}(x, n) &= n! \int_{\mathcal{O}_x} \cdots \int \exp\left[-\beta \sum_{i=1}^{n-1} w(x_{i+1} - x_i) \right] \times \left\{ \frac{\partial}{\partial x} \exp\left[-\beta w(x - x_n) \right] \right\} \prod_{k=1}^{n} dx_k \\ &= n \int_0^x dx_n \left\{ \frac{\partial}{\partial x} \exp\left[-\beta w(x - x_n) \right] \right\} \times \tilde{Z}(x_n, n-1) \\ &= n\tilde{Z}(x - d, n - 1) \end{aligned} \quad (12)$$

[4] Here, we are explicitly using the fact that $\tilde{Z}(x, n) \equiv 0$ for $x < nd$.

in agreement with Eq. (9). In Eq. (12), the integrand of $\tilde{Z}$ evaluated at $x_n = x$ vanishes because of the Boltzmann factor $\exp[-\beta w(0)]$. The last step of Eq. (12) is true since

$$\partial/\partial x \exp\left[-\beta w(x - x_n) \right] = \delta(x - x_n - d),$$

where $\delta(x)$ represents the Dirac delta function. Similarly, we have

$$\begin{aligned} \frac{\partial}{\partial x} \tilde{Z}(L - x, N - n) &= (N-n) \int_x^L dx_1 \left\{ \frac{\partial}{\partial x} \exp\left[-\beta w(x_1 - x) \right] \right\} \times \tilde{Z}(L - x_1, N - 1 - n) \\ &= -(N-n)\tilde{Z}(L - x - d, N - 1 - n). \end{aligned} \quad (13)$$

We now proceed to use Eqs. (12) and (13) in Eq. (8).

$$\begin{aligned} \frac{\partial}{\partial x} D_1^{(N+1)}(x \mid L) &= \frac{(N+1)N}{Z(L, N+1)} \sum_{n=\sigma(x)}^{\gamma(x)} \left[\binom{N-1}{n-1} \tilde{Z}(x - d, n-1) \times \tilde{Z}(L - d - (x-d), N - 1 - (n-1)) - \binom{N-1}{n} \tilde{Z}(x, n) \tilde{Z}(L - d - x, N - 1 - n) \right] \\ &= \frac{(N+1)Z(L-d, N)}{Z(L, N+1)} \times [D_1^{(N)}(x - d \mid L - d) - D_1^{(N)}(x \mid L - d)]. \end{aligned} \quad (14)$$

The last step follows from a change of variables $(n - 1 \to n)$ in the first summation. As in I, the binomial coefficient takes care of any error in the limits on the summations. Equation (14) is a recursion relation which expresses the derivative of a $D_1^{(N+1)}$ in terms of a difference of two $D_1^{(N)}$'s. We now observe that $D_1^{(1)}(x \mid L)$ is a nonzero constant for $0 \leq x \leq L$ and vanishes otherwise. Therefore, from Eq. (14) it follows that

$$\begin{aligned} \frac{\partial}{\partial x} D_1^{(2)}(x \mid L) &= \frac{2Z(L-d, 1)}{Z(L, 2)} [D_1^{(1)}(x - d \mid L - d) - D_1^{(1)}(x \mid L - d)] \\ &= 0 \end{aligned} \quad (15)$$

for $d < x < L - d$; i.e., $D_1^{(2)}(x \mid L)$ is constant in the interval $[d, L - d]$. Furthermore, outside of the interval $[d, L - d]$, $D_1^{(2)}(x \mid L)$ is clearly *not* constant

since

$$\frac{\partial}{\partial x} D_1^{(2)}(x \mid L) = \begin{cases} \dfrac{2Z(L-d,1)}{Z(L,2)} D_1^{(1)}(x-d \mid L-d), \\ \qquad \text{for} \quad L-d < x < L, \\ -\dfrac{2Z(L-d,1)}{Z(L,2)} D_1^{(1)}(x \mid L-d), \\ \qquad \text{for} \quad 0 < x < d. \end{cases} \tag{16}$$

These results can now be used to show that $D_1^{(3)}(x \mid L)$ is constant in the interval $[2d, L-2d]$ but is *not* constant outside this interval. Working up step by step, one eventually arrives at the results

$$D_1^{(N)}(x \mid L)$$

is constant for

$$(N-1)d < x < L-(N-1)d$$

and

$$L > 2(N-1)d,$$

with

$$N \geq 1. \tag{17a}$$

$$D_1^{(N)}(x \mid L)$$

is *not* constant for

$$x < (N-1)d, \quad x > L-(N-1)d$$

and

$$L > (N-1)d,$$

with

$$N > 1. \tag{17b}$$

A formal proof by induction follows. Assume that $D_1^{(N)}(x \mid L)$ is independent of x in the interval $[(N-1)d, L-(N-1)d]$ for $L > 2(N-1)d$ and $N = r$. By Eq. (14), $D_1^{(r)}(x \mid L)$ independent of x in the interval $[(r-1)d, L-(r-1)d]$ for $L > 2(r-1)d$ implies that $D_1^{(r+1)}(x \mid L)$ is independent of x in the interval $[rd, L-rd]$ for $L > 2rd$. Since $D_1^{(1)}(x \mid L)$ is independent of x in the interval $[0, L]$ for $L > 0$, we have by induction that $D_1^{(N)}(x \mid L)$ is *independent* of x for $L > 2(N-1)d$ and all $N \geq 1$ when $(N-1)d < x < L-(N-1)d$. To prove (17b), assume that $D_1^{(N)}(x \mid L)$ is dependent on x for $L > (N-1)d$ and $N = r$ when $x < (r-1)d$ or $x > L-(r-1)d$. Again using Eq. (14), $D_1^{(r)}(x \mid L)$ dependent on x for $L > (r-1)d$ when $x < (r-1)d$ or $x > L-(r-1)d$ implies that $D_1^{(r+1)}(x \mid L)$ is dependent on x for $L > rd$ when $x < rd$ or $x > L-rd$. Since $D_1^{(2)}(x \mid L)$ is dependent on x for $L > d$ when $x < d$ or $x > L-d$, we have by induction that $D_1^{(N)}(x \mid L)$ is *dependent* on x for $L > (N-1)d$ and all $N > 1$ when $x < (N-1)d$ or $x > L-(N-1)d$. This completes the proof of (17a) and (17b).

Due to the existence of Eq. (3), we may use the translational invariance properties of $D_1^{(N)}$ to determine the corresponding properties of $D_n^{(N)}$. Combining Eqs. (3) and (6), we have

$$\begin{aligned} &D_n^{(N)}(x_1 \leq \cdots \leq x_n \mid L) \\ &= \frac{1}{n!\, Z(L,N)} \sum{}'' \frac{N!}{\eta!\, N_2! \cdots N_n!} \prod_{k=2}^{n} Z(x_k - x_{k-1}, N_k) \\ &\quad \times \sum_{N_1=0}^{\eta} \binom{\eta}{N_1} \bar{Z}(x_1, N_1) \bar{Z}(L-(x_n-x_1)-x_1, \eta - N_1) \\ &= \frac{1}{n!\, Z(L,N)} \sum{}'' \frac{N!\, Z(L-(x_n-x_1), \eta+1)}{(\eta+1)!\, N_2! \cdots N_n!} \\ &\quad \times \left\{ \prod_{k=2}^{n} \bar{Z}(x_k - x_{k-1}, N_k) \right\} D_1^{(\eta+1)}(x_1 \mid L-(x_n-x_1)). \end{aligned} \tag{18}$$

The doubly primed summation is over $N_2, \cdots, N_n$ and $\eta = N_1 + N_{n+1}$, with the constraint $N_2 + \cdots + N_n + \eta = N - n$. It is clear that when $D_1^{(\eta+1)}$ is independent of x_1, then $D_n^{(N)}$ depends *only* upon $(x_2 - x_1), \cdots, (x_n - x_{n-1})$. The condition for this independence is $\eta d \leq x_1 \leq L-(x_n-x_1)-\eta d$. Since this must hold for $\eta = 0, 1, \cdots, N-n$ and $x_n - x_1 \geq 0$, each term of Eq. (18) is x_1-independent if $(N-n)d \leq x_1 \leq \cdots \leq x_n \leq L-(N-n)d$ when $L > 2(N-n)d$. Furthermore, it follows that if one or more of the x_i is outside $[(N-n)d, L-(N-n)d]$ then for fixed values of the nearest-neighbor spacings, $D_n^{(N)}(x_1, \cdots, x_n \mid L)$ is a nonconstant function of x_1. For further details the reader is referred to paper I.

IV. ONE-DIMENSIONAL NEAREST-NEIGHBOR FLUID

Since the form of the one-particle distribution function [Eq. (6)] remains the same when we add nearest-neighbor forces to the hard core, we are tempted to derive an equation similar to Eq. (14), that is, a recursion relation for the D_1's. We now show that it is possible to do this and thus extend the results of Sec. III to include all nearest-neighbor forces.

We start with Eq. (8) which is correct for nearest-neighbor forces. Using Eqs. (12) and (13) and making the variable changes $z = x - x_n$ and $z = x_1 - x$, we have

$$\frac{\partial}{\partial x} \bar{Z}(x,n) = n \int_0^x dz \left\{ \frac{\partial}{\partial z} \exp\left[-\beta w(z)\right] \right\} \bar{Z}(x-z, n-1), \tag{19a}$$

$$\begin{aligned} &\frac{\partial}{\partial x} \bar{Z}(L-x, N-n) \\ &\quad = -(N-n) \int_0^{L-x} dz \left\{ \frac{\partial}{\partial z} \exp\left[-\beta w(z)\right] \right\} \\ &\qquad \times \bar{Z}(L-x-z, N-1-n), \end{aligned} \tag{19b}$$

where $R(<2d)$ is the range of the potential defined by

$$w(x) = \begin{cases} \infty, & x < d, \quad (20a) \\ w(x) > -\infty, & d < x < R, \quad (20b) \\ 0, & x > R. \quad (20c) \end{cases}$$

Using Eqs. (19), Eq. (8) can be written, for non-integral values of x/d, as

$$\frac{\partial}{\partial x} D_1^{(N+1)}(x \mid L)$$

$$= \frac{(N+1)N}{Z(L, N+1)} \sum_{n=\sigma(x)}^{\gamma(x)} \left[\binom{N-1}{n-1} \int_0^x dz \left\{ \frac{\partial}{\partial z} \exp[-\beta w(z)] \right\} \tilde{Z}(x-z, n-1) \tilde{Z}(L-z-(x-z), N-1-(n-1)) \right.$$

$$\left. - \binom{N-1}{n} \int_0^{L-x} dz \left\{ \frac{\partial}{\partial z} \exp[-\beta w(z)] \right\} \tilde{Z}(x, n) \tilde{Z}(L-z-x, N-1-n) \right]$$

$$= \frac{(N+1)}{Z(L, N+1)} \left[\int_0^x dz \left\{ \frac{\partial}{\partial z} \exp[-\beta w(z)] \right\} Z(L-z, N) D_1^{(N)}(x-z \mid L-z) \right.$$

$$\left. - \int_0^{L-x} dz \left\{ \frac{\partial}{\partial z} \exp[-\beta w(z)] \right\} Z(L-z, N) D_1^{(N)}(x \mid L-z) \right]$$

$$= \frac{(N+1)}{Z(L, N+1)} \int_0^R dz \left\{ \frac{\partial}{\partial z} \exp[-\beta w(z)] \right\} Z(L-z, N) [D_1^{(N)}(x-z \mid L-z) - D_1^{(N)}(x \mid L-z)]. \tag{21}$$

The last step holds if $R < x < L - R$. We can now proceed by induction as before. Since the argument is identical with the previous one for the pure hard-core case, we omit the details. It is only necessary to replace d by R in the induction proof of Sec. III.[5] The result is that

$$D_1^{(N)}(x \mid L)$$

is constant for

$$(N-1)R < x < L - (N-1)R$$

and

$$L > 2(N-1)R,$$

with

$$N \geq 1. \tag{22a}$$

$D_1^{(N)}(x \mid L)$ is *not* constant for

$$x < (N-1)R, \quad x > L - (N-1)R$$

and

$$L > (N-1)R,$$

with

$$N > 1. \tag{22b}$$

Furthermore, due to Eq. (18), the translational invariance properties of $D_1^{(N)}(x \mid L)$ again ensure that $D_n^{(N)}(x_1, \cdots, x_n \mid L)$ will be translationally invariant for $(N-n)R < x_1 < \cdots < x_n < L - (N-n)R$ when $L > 2(N-n)R$ and will not be translationally invariant when one or more of the x_i lies outside this interval.

[5] Strictly speaking, the upper limit of Eq. (21) should be $R + \epsilon$ where ϵ is an arbitrarily small positive number. Then, for the pure hard-core system $R = d$ and the delta function peak is within the domain of integration. For this case, Eq. (21) reduces to Eq. (14).

V. LONG RANGE FORCES: TRANSLATIONAL INVARIANCE OF $D_n^{(N)}$

Since the key to the results of Sec. IV lies in Eq. (21), we are tempted to develop a similar recursion relation when the force range is of arbitrary, but finite extent. Specifically, we consider the case where R in Eq. (20c) is arbitrary and the condition (20a) is unnecessary. The most direct approach is to begin with $D_1^{(N+1)}(x \mid L)$, as given by Eq. (2) for $n = 1$. Differentiation of this *unordered* multiple integral leads to[6]

$$\frac{\partial D_1^{(N+1)}(x \mid L)}{\partial x} = -\beta \left(\int_0^x dz w'(z) D_2^{(N+1)}(x-z, x \mid L) - \int_0^{L-x} dz w'(z) D_2^{(N+1)}(x, x+z \mid L) \right). \tag{23}$$

If $R < x < L - R$ then both integrals run from zero to R.[7] One then sees that translational invariance of $D_2^{(N+1)}$ over some range of its variables implies translational invariance of $D_1^{(N+1)}$ over a corresponding range of x. Similarly, an examination of

$$\frac{\partial}{\partial x} D_2^{(N+1)}(x, x+z \mid L)$$

[6] Here, it is convenient to explicitly use the form

$$(\partial/\partial x) \exp[-\beta w(x)] = -\beta w'(x) \exp[-\beta w(x)].$$

In the preceding sections this was not the case.

[7] For the case of nearest-neighbor forces, this equation becomes a recursion relation for $D_1^{(N+1)}$ being identical with Eq. (21). This can be seen using Eq. (3) applied to $D_2^{(N+1)}(x-z, x \mid L)$ with $0 \leq z \leq 2d$. Since $\bar{Z}(z, n) = \exp[-\beta w(z)] \delta_{0,n}$, Eq. (3) yields

$$D_2^{(N+1)}(x-z, x \mid L) = (N+1) Z(L-z, N) \exp[-\beta w(z)] \times D_1^{(N)}(x-z \mid L-z) / Z(L, N+1).$$

shows that the translational invariance of $D_3^{(N+1)}$ implies this property for $D_2^{(N+1)}$. One might hope to continue this process for $(\partial/\partial x)D_n^{(N+1)}$ with $n = 1, 2, \cdots, N$. However, for $n = 3$, one already finds that the integrand, which is a sum of $D_4^{(N+1)}$ functions, does not vanish in an obvious way. Furthermore, this process appears to become untractable as n becomes large. These difficulties lead one to adopt a slightly different approach.

In Secs. III and IV we first discussed the translational invariance properties of $D_1^{(N+1)}$. We then used these to determine the translational invariance properties of $D_n^{(N+1)}$ for $2 \leq n \leq N$. Since Eq. (3) is not valid for the case of long range forces it is not *a priori* clear that the translational invariance of $D_1^{(N+1)}$ implies this property for $D_n^{(N+1)}$. The method of the preceding paragraph, if tractable, would establish a chain leading *from* the translational invariance of $D_{N+1}^{(N+1)}$ *to* the translational invariance of $D_1^{(N+1)}$. The method which we adopt retains this desirable feature and, in fact, allows us to relate the translational invariance properties of $D_{N-n}^{(N+1)}$ to those of $D_{N-n}^{(N-k)}$, where $0 \leq k \leq n$.

To begin, we define the functions

$$g_n^{(N+1)}(x_1, \cdots, x_n \mid L)$$

by

$$g_n^{(N+1)}(x_1, \cdots, x_n \mid L) \equiv Z(L, N+1)D_n^{(N+1)}(x_1, \cdots, x_n \mid L) \quad (24)$$

for $n = 1, \cdots, N+1$. Note that

$$g_{N+1}^{(N+1)}(x_1, \cdots, x_{N+1} \mid L) = (N+1)! \exp\left[-\beta \sum_{i<j}^{N+1} w(x_{ij})\right]$$

is translationally invariant for *all* values of $x_1, \cdots, x_{n+1}$ in $[0, L]$.[8] We denote the set $x_1, \cdots, x_j$ by X_j and write

$$g_N^{(N+1)}(X_N \mid L) = \int_0^L dy\, g_{N+1}^{(N+1)}(X_N, y \mid L). \quad (25)$$

To determine the translational invariance properties of $g_N^{(N+1)}$ (and thus of $D_N^{(N+1)}$), we investigate $(\partial/\partial\epsilon)g_N^{(N+1)}(X_N + \epsilon \mid L)$ for $\epsilon \to 0$, where $X_N + \epsilon$ represents $x_1 + \epsilon, \cdots, x_N + \epsilon$. Using the translational invariance property of $g_{N+1}^{(N+1)}$, one easily sees that[9]

$$\lim_{\epsilon\to 0} \frac{\partial}{\partial\epsilon} g_N^{(N+1)}(X_N + \epsilon \mid L) = \lim_{\epsilon\to 0} \frac{\partial}{\partial\epsilon} \int_{-\epsilon}^{L-\epsilon} dy\, g_{N+1}^{(N+1)}(X_N, y \mid L) = g_{N+1}^{(N+1)}(X_N, 0 \mid L) - g_{N+1}^{(N+1)}(X_N, L \mid L). \quad (26)$$

[8] Recall that if one or more of the $\{x_i\}$ lies outside $[0, L]$, $g_{N+1}^{(N+1)}$ vanishes.

[9] Equation (26) may be interpreted as signifying a translation of the container, holding the particles fixed. The fact that the ϵ-dependence can be removed from the integrand and isolated in the limits of integration is crucial to the present method.

It follows that if the set X_N lies in $[R, L-R]$ then each term of Eq. (26) reduces to $(N+1)g_N^{(N)}(X_N \mid L-R)$ and the right-hand side vanishes.[10] Thus, for X_n in $[R, L-R]$, $g_N^{(N+1)}(X_N \mid L)$ is translationally invariant. If X_N does not lie in this interval then the two terms on the right are in general unequal and $g_N^{(N+1)}$ is *not* invariant under translations of the set X_N.

It is now desirable to generalize this discussion to $g_{N-1}^{(N+1)}, g_{N-2}^{(N+1)}, \cdots, g_1^{(N+1)}$. The general situation is far more complicated because, after the differentiation $(\partial/\partial\epsilon)g_{N-n}^{(N+1)}$, there still remain n integrations of $g_{N+1}^{(N+1)}$ and one must prove that these vanish under appropriate restrictions on the set X_{N-n}. Thus, we consider

$$\lim_{\epsilon\to 0} \frac{\partial}{\partial\epsilon} g_{N-n}^{(N+1)}(X_{N-n} + \epsilon \mid L) = \frac{1}{(n+1)!} \lim_{\epsilon\to 0} \frac{\partial}{\partial\epsilon} \int_{-\epsilon}^{L-\epsilon} \cdots \times \int_{-\epsilon}^{L-\epsilon} g_{N+1}^{(N+1)}(X_{N-n}, y_1, \cdots, y_{n+1} \mid L) \prod_{i=1}^{n+1} dy_i = g_{N-n+1}^{(N+1)}(X_{N-n}, 0 \mid L) - g_{N-n+1}^{(N+1)}(X_{N-n}, L \mid L). \quad (27)$$

In order to show that this expression vanishes under certain restrictions on the X_{N-n}, we decompose the integrals involved on the right-hand side of Eq. (27). For $n = 1$, for example, we write

$$g_N^{(N+1)}(X_{N-1}, 0 \mid L) = \int_0^R dy\, g_{N+1}^{(N+1)}(X_{N-1}, 0, y \mid L) + \int_R^L dy\, g_{N+1}^{(N+1)}(X_{N-1}, 0, y \mid L). \quad (28)$$

If no member of the set X_{N-1} lies to the left of $2R$, then

$$g_N^{(N+1)}(X_{N-1}, 0 \mid L) = \binom{N+1}{2} g_1^{(2)}(0 \mid R) g_{N-1}^{(N-1)}(X_{N-1} - R \mid L - R) + (N+1) g_{N-1}^{(N)}(X_{N-1} - R \mid L - R). \quad (29a)$$

Similarly, using the decomposition

$$\int_0^L = \int_0^{L-R} + \int_{L-R}^L,$$

we find that

$$g_N^{(N+1)}(X_{N-1}, L \mid L) = (N+1) g_{N-1}^{(N)}(X_{N-1} \mid L-R) + \binom{N+1}{2} g_1^{(2)}(R \mid R) g_{N-1}^{(N-1)}(X_{N-1} \mid L-R) \quad (29b)$$

[10] If the set X_N lies to the right of R, $g_{N+1}^{(N+1)}(X_N, 0 \mid L) = (N+1)g_N^{(N)}(X_N - R \mid L - R)$. If the set X_N lies to the left of $L-R$, $g_{N+1}^{(N+1)}(X_N, L \mid L-R) = (N+1)g_N^{(N)}(X_N \mid L-R)$. These expressions are equal since $g_N^{(N)}$ is manifestly translationally invariant. Since X_N is restricted to $[R, L-R]$ one *could* denote the domain in the $g_N^{(N)}$ functions by $L - 2R$, rather than $L - R$. The precise label is irrelevant for our purposes.

when no member of the set X_{N-1} lies to the right of $L - 2R$. Combining Eqs. (27) and (29) we see that $g_{N-1}^{(N+1)}(X_{N-1} \mid L)$ is translationally invariant if the set X_{N-1} lies in $[2R, L - 2R]$.

We may gain an understanding of the integral decomposition by a simple graphical notation. We represent the two integrands of Eq. (28) by Fig. 1. Here, one sees that in the first term, y interacts with the fixed point at zero, but not with the set X_{N-1}. In the second term, the reverse situation holds. The fact that y cannot *simultaneously* interact with the set X_{N-1} *and* the fixed point accounts for the factorization which occurs in Eqs. (29). This provides a clue as to how to proceed with Eq. (27) for arbitrary n. Symbolically, we write

$$g_{N-n+1}^{(N+1)}(X_{N-n}, 0 \mid L) = \frac{1}{n!}\left[\int_0^L dy\right]^n g_{N+1}^{(N+1)}(X_{N-n}, 0, y_1, \cdots, y_n \mid L) \quad (30)$$

using an obvious notation. Now, let $\mathbf{I}_j$ and $\mathbf{I}_n$ be integral operators defined by

$$\mathbf{I}_j \equiv \int_{jR}^{(j+1)R} dy, \quad \text{for} \quad j = 0, \cdots, n-1, \quad (31a)$$

$$\mathbf{I}_n \equiv \int_{nR}^{L} dy. \quad (31b)$$

Since $g_{N+1}^{(N+1)}(X_{N-n}, 0, y_1, \cdots, y_n \mid L)$ is symmetric under permutations of the y_i variables, we may combine Eqs. (30) and (31), using the multinomial expansion, as follows:

$$\begin{aligned} g_{N-n+1}^{(N+1)}(X_{N-n}, 0 \mid L) &= \frac{1}{n!}(\mathbf{I}_0 + \mathbf{I}_1 + \cdots + \mathbf{I}_n)^n \\ &\quad \times g_{N+1}^{(N+1)}(X_{N-n}, 0, y_1, \cdots, y_n \mid L) \\ &= \sum_{\{l_i\}}{}' \frac{\mathbf{I}_0^{l_0}\mathbf{I}_1^{l_1}\cdots\mathbf{I}_n^{l_n}}{l_0!\, l_1! \cdots l_n!} g_{N+1}^{(N+1)}(X_{N-n}, 0, y_1, \cdots, y_n \mid L). \end{aligned} \quad (32)$$

It is understood using this notation that each integral operator acts on a *different* y_i variable. The primed summation signifies that $l_0, l_1, \cdots, l_n$ each run from zero to n under the constraint

$$\sum_{i=0}^{n} l_i = n.$$

The utility of Eq. (32) lies in the fact that, for any configuration of the $\{l_i\}$, *at least one* l_i must be zero since there are $n + 1$ intervals and n integration variables. Graphically, this means that the integrand of Eq. (32) consists of a sum of terms, each of which has *at least one gap* of length R separating two groupings of the y_i variables. For example, if $l_m = 0$, the configuration may be represented by Fig. 2 if the set X_{N-n} lies to the right of $(n + 1)R$. This condition and the occurrence of gaps allows a factorization of the $g_{N+1}^{(N+1)}$ function which is essential to the translational invariance proof.

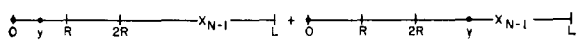


FIG. 1. Graphical representation of the two integrands of Eq. (28).

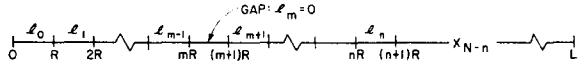


FIG. 2. Graphical representation of a typical term of the summation in Eq. (32). The set X_{N-n} is assumed to lie to the right of $(n + 1)R$, and $l_m = 0$ where $m < n$.

To continue, we represent the summation in Eq. (32) as a sum over the various possible gaps, as indicated by Eq. (33).

$$\sum_{\{l_i\}}{}' = \sum_{l_0=0}{}' + \sum_{\substack{l_1=0 \\ l_0>0}}{}' + \sum_{\substack{l_2=0 \\ l_0, l_1>0}}{}' + \cdots + \sum_{\substack{l_n=0 \\ l_0, l_1, \cdots, l_{n-1}>0}}{}' . \quad (33)$$

In each sum, the prime means that

$$\sum_{i=0}^{n} l_i = n.$$

The other indicated restrictions are necessary to prevent the overcounting of 2-, 3-, $\cdots$, $(n-1)$-gap configurations. An analysis of the first three summations enables one to see that a useful pattern emerges.

$$\sum_{l_0=0}{}' = (N+1) g_{N-n}^{(N)}(X_{N-n} - R \mid L - R), \quad (34)$$

$$\begin{aligned} \sum_{\substack{l_1=0 \\ l_0>0}}{}' &= \sum_{l_0=1}^{n} \binom{N+1}{l_0+1} \frac{\mathbf{I}_0^{l_0}}{l_0!} g_{l_0+1}^{(l_0+1)}(0, y_1, \cdots, y_{l_0} \mid R) \\ &\quad \times \sum_{l_2, \cdots, l_n} \frac{(n-l_0)!}{(n-l_0)!} \frac{\mathbf{I}_2^{l_2}\cdots\mathbf{I}_n^{l_n}}{l_2!\cdots l_n!} \\ &\quad \times g_{N-l_0}^{(N-l_0)}(X_{N-n}, y_{l_0+1}, \cdots, y_n \mid L - 2R) \\ &= \sum_{l_0=1}^{n} \binom{N+1}{l_0+1} g_1^{(l_0+1)}(0 \mid R) \\ &\quad \times g_{N-n}^{(N-l_0)}(X_{N-n} - 2R \mid L - 2R). \end{aligned} \quad (35)$$

In Eq. (35), the summation over $l_2, \cdots, l_n$ is performed with $l_2 + \cdots + l_n = n - l_0$. The binomial coefficient

$$\binom{N+1}{l_0+1}$$

arises from the identity

$$\begin{aligned} g_{N+1}^{(N+1)}(X_{N-n}, 0, y_1, \cdots, y_n \mid L) &= \binom{N+1}{l_0+1} g_{l_0+1}^{(l_0+1)}(0, y_1, \cdots, y_{l_0} \mid R) \\ &\quad \times g_{N-l_0}^{(N-l_0)}(X_{N-n}, y_{l_0+1}, \cdots, y_n \mid L - 2R), \end{aligned} \quad (36)$$

which holds when a gap separates $(y_1, \cdots, y_{l_0})$ and $(y_{l_0+1}, \cdots, y_n, X_{N-n})$; i.e., when $l_1 = 0$. For the third summation we have

$$\sum_{\substack{l_2=0 \\ l_0, l_1 > 0}}' = \sum_{\substack{l_0+l_1=k=2 \\ l_0, l_1 > 0}}^{n} \binom{N+1}{k+1} \frac{\mathbf{I}_0^{l_0}}{l_0!} \frac{\mathbf{I}_1^{l_1}}{l_1!} g_{k+1}^{(k+1)}(0, y_1, \cdots, y_k \mid 2R)$$
$$\times \sum_{l_3+\cdots+l_n=n-k} \frac{(n-k)!}{(n-k)!} \frac{\mathbf{I}_3^{l_3} \cdots \mathbf{I}_n^{l_n}}{l_3! \cdots l_n!}$$
$$\times g_{N-k}^{(N-k)}(X_{N-n}, y_{k+1}, \cdots, y_n \mid L - 3R)$$
$$= \sum_{\substack{l_0+l_1=k=2 \\ l_0, l_1 > 0}}^{n} \binom{N+1}{k+1} \frac{\mathbf{I}_0^{l_0}}{l_0!} \frac{\mathbf{I}_1^{l_1}}{l_1!} g_{k+1}^{(k+1)}(0, y_1, \cdots, y_k \mid 2R)$$
$$\times g_{N-n}^{(N-k)}(X_{N-n} - 3R \mid L - 3R). \tag{37}$$

It is now clear that Eq. (32) is expressible in the form

$$g_{N-n+1}^{(N+1)}(X_{N-n}, 0 \mid L)$$
$$= \sum_{j=0}^{n} \sum_{\substack{l_0+\cdots+l_{j-1}=k=j \\ l_0, \cdots, l_{j-1} > 0}}^{n} C_{jk}(l_0, \cdots, l_{j-1})$$
$$\times g_{N-n}^{(N-k)}(X_{N-n} - (j+1)R \mid L - (j+1)R), \tag{38}$$

with

$$C_{jk}(l_0, \cdots, l_{j-1}) \equiv \binom{N+1}{k+1} \frac{\mathbf{I}_0^{l_0} \cdots \mathbf{I}_{j-1}^{l_{j-1}}}{l_0! \cdots l_{j-1}!}$$
$$\times g_{k+1}^{(k+1)}(0, y_1, \cdots, y_k \mid jR). \tag{39}$$

The $j = 0$ term in Eq. (38) is understood to be given by (34).

It is possible to decompose the integrals in

$$g_{N-n+1}^{(N+1)}(X_{N-n}, L \mid L)$$

using the following operators

$$\int_0^L dy = (\mathbf{J}_1 + \mathbf{J}_2 + \cdots + \mathbf{J}_n), \tag{40}$$

where

$$\mathbf{J}_j \equiv \int_{L-(j+1)R}^{L-jR} dy \quad \text{for} \quad j = 0, 1, \cdots, n-1 \tag{41a}$$

and

$$\mathbf{J}_n \equiv \int_0^{L-nR} dy. \tag{41b}$$

The analysis is identical to that above and leads to the following expression [as before, the $j = 0$ term is understood to be $(N+1)g_{N-n}^{(N)}(X_{N-n} \mid L - R)$].

$$g_{N-n+1}^{(N+1)}(X_{N-n}, L \mid L)$$
$$= \sum_{j=0}^{n} \sum_{\substack{l_0+\cdots+l_{j-1}=k=j \\ l_0, \cdots, l_{j-1} > 0}}^{n} B_{jk}(l_0, \cdots, l_{j-1})$$
$$\times g_{N-n}^{(N-k)}(X_{N-n} \mid L - (j+1)R) \tag{42}$$

with

$$B_{jk}(l_0, \cdots, l_{j-1}) = \binom{N+1}{k+1} \frac{\mathbf{J}_0^{l_0} \cdots \mathbf{J}_{j-1}^{l_{j-1}}}{l_0! \cdots l_{j-1}}$$
$$\times g_{k+1}^{(k+1)}(L, y_1, \cdots, y_k \mid jR) \tag{43}$$

when the set X_{N-n} lies to the left of $L - (n+1)R$. Making the transformation of variables $y_i' = L - y_i$ for $i = 1, 2, \cdots, k$ and utilizing the reflection symmetry of $g_{k+1}^{(k+1)}$ about $\frac{1}{2}L$ it is obvious that[11]

$$B_{jk}(l_0, \cdots, l_{j-1}) = C_{jk}(l_0, \cdots, l_{j-1}). \tag{44}$$

Therefore, the combination of Eqs. (27), (38), (42), and (44) yields the result

$$\lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} g_{N-n}^{(N+1)}(X_{N-n} + \epsilon \mid L)$$
$$= \sum_{j=0}^{n} \sum_{\substack{l_0+\cdots+l_{j-1}=k=j \\ l_0, \cdots, l_{j-1} > 0}}^{n} C_{jk}(l_0, \cdots, l_{j-1})$$
$$\times [g_{N-n}^{(N-k)}(X_{N-n} - (j+1)R \mid L - (j+1)R)$$
$$- g_{N-n}^{(N-k)}(X_{N-n} \mid L - (j+1)R)] \tag{45}$$

for $(n+1)R \leq x_i \leq L - (n+1)R$, $i = 1, \cdots, (N-n)$ and $n = 1, \cdots, (N-1)$. In the $j = 0$ term the second sum is ignored, $k = 0$, and C_{00} is taken to be $N + 1$. Equation (45) is the desired recursion relation representing $g_{N-n}^{(N+1)}$ in terms of differences of $g_{N-n}^{(M)}$'s with $N - n \leq M \leq N$. The proof of translational invariance is again by induction. Assume that $g_n^{(n+m)}(X_n \mid L)$ is translationally invariant for any n, $0 \leq m \leq r$, and L such that $mR < x_i < L - mR$, $i = 1, \cdots, n$. By Eq. (45), this implies that

$$g_n^{(n+r+1)}(X_n \mid L)$$

is translationally invariant for $(r+1)R < x_i < L - (r+1)R$. But $g_n^{(n)}(X_n \mid L)$ is manifestly translationally invariant for any n and L such that $0 \leq x_i \leq L$, $i = 1, \cdots, n$. Therefore, by induction, $g_n^{(n+m)}(X_n \mid L)$ is translationally invariant for any n, all $m \geq 0$, and L such that $mR < x_i < L - mR$. Replacing m by $N - n$ and using D_n functions, we have as our major result:

$$D_n^{(N)}(x_1, \cdots, x_n \mid L)$$

is translationally invariant if

$$L > 2(N-n)R$$

and

$$(N-n)R \leq x_i \leq L - (N-n)R \quad \text{for} \quad i = 1, \cdots, n$$

and

$$n = 1, \cdots, N$$

with

$$N \geq 1. \tag{46}$$

It should be emphasized that this result holds for all potentials of finite extent, with or without a hard core.

[11] Note that the intervals $[0, jR]$ and $[L - jR, L]$ are both denoted simply as jR in the $g_{k+1}^{(k+1)}$ functions. The specific interval involved is clear from its context.

The result (46) gives sufficient conditions for translational invariance. It appears to be very difficult to establish whether or not these conditions are also necessary. Some evidence exists indicating that the stated conditions *are* necessary, and we are content to state this evidence with the hope that a general proof ultimately will be found.

First, consider Eq. (26) and the remarks which follow it. Clearly, the condition that the set X_N lies in $[R, L - R]$ is necessary and sufficient for $g_N^{(N+1)}$ to be translationally invariant for arbitrary temperatures. Next, we consider $g_{N-n}^{(N+1)}(X_{N-n} \mid L)$ where $x_2, \cdots, x_{N-n}$ lie in $[(n+1)R, L-(n+1)R]$, but x_1 lies in $[nR, (n+1)R]$. Going back to the preceding analysis, one finds that for $n \geq 1$

$$\lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} g_{N-n}^{(N+1)}(X_{N-n} + \epsilon \mid L) = \binom{N+1}{n+1} g_{N-n}^{(N-n)}(X_{N-n} \mid L - nR) \times \mathbf{I}_0 \cdots \mathbf{I}_{n-1} g_{n+1}^{(n+1)}(0, y_1, \cdots, y_n \mid nR) \times \{\exp[-\beta w(x_1 - y_n)] - 1\}. \quad (47)$$

Performing the integrals over $y_1, \cdots, y_{n-1}$, one has

$$\lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} g_{N-n}^{(N+1)}(X_{N-n} + \epsilon \mid L) = \binom{N+1}{n+1} g_{N-n}^{(N-n)}(X_{N-n} \mid L - nR) \times \int_{(n-1)R}^{nR} dy_n F(y_n)\{\exp[-\beta w(x_1 - y_n)] - 1\} \quad (48)$$

with

$$F(y_n) \equiv \mathbf{I}_0 \cdots \mathbf{I}_{n-2} g_{n+1}^{(n+1)}(0, y_1, \cdots, y_n \mid nR). \quad (49)$$

In general, Eq. (48) does not vanish for all values of β, which shows that $g_{N-n}^{(N+1)}$ is *not* translationally invariant if $x_2, \cdots, x_{N-n}$ lie in $[(n+1)R, L-(n+1)R]$, but x_1 lies in $[nR, (n+1)R]$.

In order to demonstrate that the stated translational invariance properties are *necessary*, one would have to show that

$$\lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} g_{N-n}^{(N+1)}$$

is nonzero for *any* configuration where one or more of the x_i lies outside the interval $[(n+1)R, L-(n+1)R]$. Although we have not been able to prove this, the above examples suggest that the conditions in Eq. (46) are indeed necessary.

VI. DISCUSSION

We have seen that it is not necessary for the length of a one-dimensional system to become infinite in order to have translational invariance of the n-particle distribution functions. It is a remarkable fact that the container walls have no effect on the property of translational invariance as long as they are a *sufficient* distance apart. The method which we have used in Sec. V provides a particularly transparent physical interpretation of this distance. $D_n^{(N)}(X_n \mid L)$ is translationally invariant when each member of the set X_n is far enough from the walls so that it cannot interact with a particle fixed at either wall, through a chain of interactions with the remaining particles. That is, translational invariance of $D_n^{(N)}$ is assured when $x_1, \cdots, x_n$ are all *at least* a distance $(N-n)R$ from either wall.

The translational invariance properties are strictly geometrical in nature and do not depend in any way upon the temperature of the system. Furthermore, these properties do not require a hard core and depend only upon the finite extent of the potential. They therefore hold even for systems which may be unstable in the thermodynamic limit.

On the basis of the physical picture in the first paragraph of this section one might expect three-dimensional systems to have the *same* translational invariance properties as one-dimensional systems. If this is indeed so, then one intuitively expects the deviations from translational invariance outside the central regions to be exceedingly small in three dimensions. To see this, consider a box of volume L^3 with $N = 10^{22}$ and $R = 10^{-8}$ cm. The above conditions for translational invariance would require that $L > 2 \times 10^{14}$ cm, corresponding to a density of $\sim 10^{-21}$ particles/cm^3! On the basis of known homogeneity properties for real fluids, this suggests that either (a) less stringent conditions than $(N-n)R < x_i < L - (N-n)R$ are *necessary* for the translational invariance of $D_n^{(N)}$ in three dimensions, or (b) if these conditions *are* necessary then *approximate* translational invariance must exist for realistic densities. This open question poses a challenging problem for future study.

ACKNOWLEDGMENT

This work was supported in part by the United States Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Symmetry Group of the Hydrogen Atom

R. J. FINKELSTEIN
Department of Physics, University of California, Los Angeles

(Received 31 May 1966)

It is shown how the complete dynamics of the hydrogen atom is related to the three-dimensional rotation group.

1. INTRODUCTION

THE angular momentum, **L**, is an integral of the motion and gives rise to a $(2l+1)$-fold degeneracy for every spherically symmetric potential. The Coulomb potential has in addition a second vector integral of the motion, **V**, the Runge–Lenz vector, and a total degeneracy N^2, where N is the principal quantum number. These two integrals, **L** and **V**, were related to the six generators of 0_4 by Fock[1] and by Bargmann.[2] More recent papers have discussed the relation of this problem to larger symmetry groups. Here, on the other hand, we base our treatment entirely on the group 0_3.

Our method depends on the fact that the Schrödinger equation in the momentum representation may be interpreted as an integral equation on the group space of 0_3; the connection with 0_4 arises because the group of motions of this space is just 0_4. Here, however, we discuss the intrinsic geometry of the group space instead of embedding it in a space of higher dimensions as is usually done. Therefore our entire treatment is based on 0_3. One may then say that the dynamics and symmetries of this problem, which determines the magic numbers (N^2) of atomic physics, are both determined by a single group.

2. MOMENTUM REPRESENTATION AND SPHERICALLY SYMMETRIC POTENTIAL

We first consider a spherically symmetric attractive potential which is not necessarily Coulomb. In the momentum representation, we have

$$\left(E - \frac{\mathbf{p}^2}{2m}\right)\varphi(\mathbf{p}) = \int \tilde{V}(\mathbf{p}-\mathbf{p}')\varphi(\mathbf{p}')\,d\mathbf{p}', \quad (2.1)$$

where $\tilde{V}(\mathbf{p})$ is the Fourier transform of the potential with the property

$$\tilde{V}(\mathbf{p}) = \tilde{V}(|\mathbf{p}|). \quad (2.2)$$

Let

$$E = -p_0^2/2m. \quad (2.3)$$

Then p_0 is real for bound states and imaginary for scattering states, and

$$G^{-1}(\mathbf{p})\varphi(\mathbf{p}) = \frac{1}{E}\int \tilde{V}(\mathbf{p}-\mathbf{p}')\varphi(\mathbf{p}')\,d\mathbf{p}', \quad (2.4)$$

where

$$G(\mathbf{p}) = p_0^2/(p^2 + p_0^2) \quad (2.4a)$$

is the propagator.

Equation (2.4) may be written as an integral equation on the group space of 0_3 by using **p** itself to parameterize a rotation. Let the spin representation of the rotation **w** be

$$D^{\frac{1}{2}}(\mathbf{w}) = \exp(\tfrac{1}{2}i\mathbf{w}\boldsymbol{\sigma}) \quad (2.5a)$$

$$= (p_0 + i\boldsymbol{\sigma}\mathbf{p})/(p_0 - i\boldsymbol{\sigma}\mathbf{p}). \quad (2.5b)$$

In terms of **p** the invariant volume element in group space is (as shown in paragraph 4)

$$d\tau = g^{\frac{1}{2}}\,d\mathbf{p}, \quad (2.6a)$$

where the group metric is simply related to the propagator G of (2.4),

$$g^{\frac{1}{2}} = G^3. \quad (2.6b)$$

Then (2.4) becomes

$$\Phi(\mathbf{p}) = \frac{1}{E}\int U(\mathbf{p},\mathbf{p}')\Phi(\mathbf{p}')\,d\tau', \quad (2.7)$$

where

$$\Phi(\mathbf{p}) = G^{-2}(\mathbf{p})\varphi(\mathbf{p}), \quad (2.7a)$$

$$U(\mathbf{p},\mathbf{p}') = G^{-1}(p)\tilde{V}(|\mathbf{p}-\mathbf{p}'|)G^{-1}(p'). \quad (2.7b)$$

The kernel $U(p,p')$ is now symmetric in view of (2.2). Since $U(p,p')$ is real and symmetric, the eigenvalues E are real. The transformation just made is of course valid for any spherically symmetric potential.

The function $g^{\frac{1}{2}}$ is positive definite for bound states $(E<0)$. For scattering states $(E>0)$, $g^{\frac{1}{2}}$ may become negative.

3. GEOMETRY OF ROTATION GROUP

Every three-dimensional rotation may be represented by a vector **w** giving the magnitude and axis of the rotation and, therefore, by a point in a sphere of radius π. This spherical ball is the group space and has a well-defined geometrical structure

[1] V. Fock, Z. Physik **98**, 145 (1935).
[2] V. Bargmann, Z. Physik **99**, 576 (1936).

which may be characterized by its metric and connection. It may be mapped into an unbounded three-dimensional continuum, but its total volume must, of course, remain finite.

Since this space has constant torsion and therefore absolute parallelism, its metric and connection may be derived from either of two sets of absolutely parallel triad fields $\lambda_\mu^{(l)}(\pm)$ as follows,[3,4]

$$g_{\alpha\beta} = \sum_l \lambda_\alpha^{(l)}(\pm)\lambda_\beta^{(l)}(\pm), \tag{3.1}$$

$$L_{\alpha\beta}^{\mu} = \sum_l \lambda_{(l)}^{\mu}(\pm)\frac{\partial\lambda_\alpha^{(l)}(\pm)}{\partial x^\beta}, \tag{3.2}$$

where the $(\pm)$ means that there are in fact two parallels at P' to a vector at P, and one may describe the space equally well by using either the right $(+)$ or the left $(-)$ parallel triads. [Here $\lambda_{(l)}^{\mu}(\pm)$ means the reciprocal triad.]

In terms of $\lambda_{(l)}^{\mu}$, one defines the displacement operators,

$$X_l(\pm) = \frac{1}{i}\lambda_{(l)}^{\mu}(\pm)\partial_\mu, \tag{3.3}$$

which have the commutative properties,

$$[X_i(\pm), X_j(\pm)] = \mp(2i/R_0)\epsilon_{ijk}X_k(\pm), \tag{3.4a}$$

$$[X_i(\pm), X_j(\mp)] = 0, \tag{3.4b}$$

where R_0 is the radius of curvature of the space. The Casimir operator has the simple geometrical interpretation,

$$\sum_i X_i(\pm)^2 = -\Delta, \tag{3.5}$$

where Δ is the Laplace–Beltrami operator,

$$\Delta = \frac{1}{g^{\frac{1}{2}}}\frac{\partial}{\partial a^\mu}g^{\frac{1}{2}}g^{\mu\nu}\frac{\partial}{\partial a^\nu}, \tag{3.5a}$$

and a^μ is any coordinate system in group space. The matrix elements of the irreducible representation are solutions of the differential equations,

$$X^2(\pm)D_{mm'}^j(a) = [4j(j+1)/R_0^2]D_{mm'}^j(a), \tag{3.6a}$$

$$X_3(+)D_{mm'}^j(a) = (2m/R_0)D_{mm'}^j(a), \tag{3.6b}$$

$$X_3(-)D_{mm'}^j(a) = (2m'/R_0)D_{mm'}^j(a), \tag{3.6c}$$

and satisfy

$$\int \bar{D}_{mn}^j(a)D_{m'n'}^{j'}(a)g^{\frac{1}{2}}\,da = \delta^{jj'}\delta_{mm'}\delta_{nn'}\frac{V}{d_j}, \tag{3.7}$$

where $d_j = 2j+1$ is the dimensionality of the representation and

$$\int g^{\frac{1}{2}}\,da = V \tag{3.8}$$

is the invariant volume of the group space.

The eigenfunctions of (3.6) also satisfy the following integral equation[5]:

$$D_{mn}^j(a) = \frac{d_j}{\mu^{2j}}\int K_\mu(a, a')D_{mn}^j(a')\,d\tau_{a'}, \tag{3.9}$$

where

$$d\tau_a = g^{\frac{1}{2}}\,da, \tag{3.9a}$$

$$K_\mu(a, a') = \frac{1}{V}\sum_{jmn} D_{mn}^j(a)\bar{D}_{mn}^j(a')\mu^{2j} \tag{3.10}$$

$$= \frac{1}{V}\sum_j \frac{\sin(j+\frac{1}{2})w}{\sin\frac{1}{2}w}\mu^{2j} \tag{3.10a}$$

$$= \frac{1}{V}\frac{1}{1-2\mu\cos\frac{1}{2}w+\mu^2}, \tag{3.10b}$$

where w is the magnitude of the rotation $R^{-1}R'$ connecting the points a and a', and $|\mu| \leq 1$.

The two triads of displacement operators $X_i(\pm)$ are the generators of motions which carry the group manifold into itself while preserving the metric and connection. Both the left- and right-hand triads generate 0_3 and the complete group of motions is $0_4 = 0_3 \times 0_3$.

4. STEREOGRAPHIC COORDINATES

Let us put

$$D(a) = [1 + \tfrac{1}{2}iR(a)]/[1 - \tfrac{1}{2}iR(a)], \tag{4.1}$$

where $R(a)$ is Hermitian. Choosing the spin representation $D^{\frac{1}{2}}(a)$, one may introduce stereographic coordinates as follows:

$$R = \boldsymbol{\sigma}\mathbf{r}/R_0, \tag{4.2}$$

where R_0 may be shown to be the radius of the group space. In this coordinate system, the metric and connection take the following form[4,5]:

(a) metric

$$g_{\alpha\beta} = G^2\delta_{\alpha\beta}, \tag{4.3}$$

$$g^{\frac{1}{2}} = G^3; \tag{4.3a}$$

(b) symmetric part of connection

$$L_{(\alpha\beta)}^{\mu} = (\delta_{\mu\alpha}\partial_\beta + \delta_{\mu\beta}\partial_\alpha - \delta_{\alpha\beta}\partial_\mu)\ln G; \tag{4.4a}$$

(c) torsion

$$L_{[\alpha\beta]}^{\mu} = \frac{2g^{\frac{1}{2}}}{R_0}g^{\mu\sigma}\epsilon_{\sigma\alpha\beta}. \tag{4.4b}$$

[3] Most of the geometrical background referred to here may be found in L. P. Eisenhart, *Continuous Groups of Transformations* (Dover Publications, Inc., New York, 1961); see particularly Chap. V.

[4] R. Finkelstein, J. Math. Phys. **1**, 440 (1960).

[5] R. Finkelstein, J. Math. Phys. **7**, 1632 (1963).

In these formulas the scale factor G is

$$G^{-1}(r) = 1 + \tfrac{1}{4}(r^2/R_0^2), \tag{4.5}$$

and the invariant group volume is

$$\int g^{\frac{1}{2}}\, d\mathbf{r} = 2\pi^2 R_0^3. \tag{4.6}$$

The fundamental triads with which we began are[4]

$$\lambda_\alpha^l(\pm) = G^2[(2 - G^{-1})\delta_\alpha^l + \tfrac{1}{2}(1/R_0^2)r^l r_\alpha \mp \epsilon_{\alpha t}^l (R/R_0)^t], \tag{4.7}$$

and the corresponding displacement operators are

$$X_l(\pm) = d_l - (1/4R_0^2)A_l \pm (1/R_0)L_l, \tag{4.8}$$

where

$$d_l = (1/i)(\partial/\partial r_l), \tag{4.8a}$$

$$L_l = r_j\, d_k - r_k\, d_k, \tag{4.8b}$$

$$A_l = r^2\, d_l - 2r_l\theta, \tag{4.8c}$$

$$\theta = r^s\, d_s. \tag{4.8d}$$

To terms of order $1/R_0$ the X_i are helicity operators; the terms of order $1/R_0^2$ are present in an Einstein space without torsion and may be interpreted in terms of an acceleration operator.[5]

The kernel of the integral equation is now expressed directly in terms of r with the aid of the relation

$$(1/R_0^2)G(r)(\mathbf{r} - \mathbf{r}')^2 G(r') = 4\sin^2(\tfrac{1}{4}w). \tag{4.9}$$

Therefore,

$$K_\mu(r, r') = \frac{1}{V}\left(\frac{\mu^{-1}}{(\mu^{\frac{1}{2}} - \mu^{-\frac{1}{2}})^2 + R_0^{-2}G(r)(\mathbf{r} - \mathbf{r}')^2 G(r')}\right). \tag{4.10}$$

5. CORRESPONDENCE BETWEEN MOMENTUM SPACE AND GROUP SPACE

By Eq. (2.5b) we may parameterize a rotation with the components of the momentum itself. By comparing (2.5b) with (4.1), one establishes the following correspondence:

$$\mathbf{p}/p_0 = \tfrac{1}{2}\mathbf{r}/R_0 \tag{5.1}$$

between momentum space and group space described in stereographic coordinates $\mathbf{r}$. Let us put

$$p_0 = 2R_0, \tag{5.1a}$$

$$\mathbf{p} = \mathbf{r}. \tag{5.1b}$$

Then the energy $-p_0^2/2m$ determines the radius (R_0) of the group space and

$$G^{-1}(r) = G^{-1}(p) = 1 + \mathbf{p}^2/p_0^2. \tag{5.2}$$

The invariant volume element is by (4.3a)

$$d\tau = g^{\frac{1}{2}}\, d\mathbf{p} = G^3\, d\mathbf{p}, \tag{5.3}$$

as assumed in (2.6). The invariant group volume is then

$$V = 2\pi^2 R_0^3 \tag{5.4a}$$

$$= \tfrac{1}{4}\pi^2 p_0^3. \tag{5.4b}$$

The main point now is that the integral equation, (3.9), on the group space is the same as Schrödinger's equation in the momentum representation when the potential is of the Coulomb type. For then (in the attractive case)

$$\tilde{V}(\mathbf{p} - \mathbf{p}') = -(e^2/2\pi^2\hbar)[1/(\mathbf{p} - \mathbf{p}')^2], \tag{5.5}$$

and therefore,

$$U(p, p') = -\frac{e^2}{2\pi^2\hbar}\frac{1}{(\mathbf{p} - \mathbf{p}')^2}\frac{1}{G(p)}\frac{1}{G(p')}. \tag{5.5a}$$

By comparing with (4.10), one has the result

$$U(p, p') = -\frac{e^2}{2\pi^2\hbar}\frac{V}{R_0^2}K_1(p, p') = -\frac{e^2}{\hbar}\frac{p_0}{2}K_1(p, p'). \tag{5.6}$$

Then (3.9) becomes

$$\Phi_E(p) = -\frac{1}{E}\frac{e^2 p_0}{2\hbar}\int K_1(p, p')\Phi_E(p')\, d\tau'. \tag{5.7}$$

The eigenvalues are by (3.9)

$$-e^2 p_0/2E\hbar = d_j \tag{5.8}$$

or

$$p_0 = me^2/\hbar\, d_j, \tag{5.8a}$$

$$E = -(e^4/2m\hbar^2)(1/d_j^2), \tag{5.8b}$$

which is the Balmer formula, where d_j is the principal quantum number N.

The eigenfunctions are

$$\Phi_E(p) = D^j_{mn}(p). \tag{5.8c}$$

The total degeneracy is $d_j^2 = N^2$ since both m and n run from $-j$ to $+j$.

In the general case ($\mu \neq 1$) we may again construct an integral equation for $\Phi(p)$ by defining

$$\tilde{V}(p, p') = G(p)U(p, p')G(p') = -\frac{e^2 p_0}{2\hbar}G(p)K(p, p')G(p') \tag{5.9}$$

$$= -\frac{2e^2}{\pi^2\hbar}\left[\frac{\mu^{-1}}{(\mu^{\frac{1}{2}} - \mu^{-\frac{1}{2}})^2 p_0^2 G(p)^{-1}G(p')^{-1} + 4|\mathbf{p} - \mathbf{p}'|^2}\right]. \tag{5.9a}$$

This is not the Fourier transform of a central symmetric potential in configuration space. Nevertheless, if we regard the dynamical problem as given in the momentum representation, we may say that the

particle behaves as if it were moving in a Yukawa potential with a velocity-dependent mass or range, with the limiting value for $\mathbf{p} = \mathbf{p}' = 0$ given by

$$(\mu^{\frac{1}{2}} - \mu^{-\frac{1}{2}})^2.$$

On the other hand, for large values of p and p' and a fixed value of momentum transfer,

$$\tilde{V}(p, p') \sim [\mu^{-1}/(\mu^{\frac{1}{2}} - \mu^{-\frac{1}{2}})^2]G(p)G(p') \to 0.$$

The spectrum is now determined by

$$me^2/\hbar p_0 = d_j/\mu^{2j}$$

or

$$E(N) = \mu^{2(N-1)}E(N)_{\text{Balmer}}. \tag{5.10}$$

The wavefunctions are the same as in the case $\mu = 1$, but they are differently correlated with the energy. For small μ there is effectively one bound state ($N = 1$); the rest of the bound states are crowded into a small interval lying just below zero energy; thus the true continuum is extended into a quasi-continuum lying just below zero.

6. HAMILTONIAN

It appears most natural to formulate this problem in the momentum representation and to ignore the Hamiltonian, since the natural operators are the six displacement operators $X_i(\pm)$, which do not include the Hamiltonian. In order to connect with the usual formulation, however, let us express the Hamiltonian in terms of the $X_i(\pm)$. We have

$$E_j = -\epsilon\mu^{2j}/(2j + 1)^2, \tag{6.1}$$

where $-\epsilon$ is the lowest Balmer energy and where $2j + 1$ may be regarded as an eigenvalue associated with the integral equation, (3.9), or with the differential equation, (3.6a).

It follows from (6.1) and (3.6a) that the energy operator which works on the bound states of the Coulomb potential may be expressed as follows:

$$H = -\epsilon/[R_0^2 X(\pm)^2 + 1]. \tag{6.2}$$

Since R_0 also depends on the energy, the relation (6.2) may be re-expressed in the following operator form:

$$\tfrac{1}{2}X^2(\pm) = (H + \epsilon)/mH^2, \tag{6.3}$$

where m is the mass. Therefore the Hamiltonian commutes with $X_i(+)$ and $X_i(-)$.

The Hamiltonian also commutes with the six operators (U_{lm}, U_{k4}) defined as follows:

$$U_{lm} = \tfrac{1}{2}[X_k(+) - X_k(-)], \tag{6.4a}$$

$$U_{k4} = \tfrac{1}{2}[X_k(+) + X_k(-)]. \tag{6.4b}$$

From (4.8) we see that U_{lm} is, except for p_0, the ordinary angular momentum,

$$U_{lm} = p_0^{-1}L_k,$$

and that

$$U_{k4} = d_k - p_0^{-2}A_k.$$

According to Bargmann,[2] the Runge–Lenz vector is

$$V_k = (i\hbar/2e^2m)p_0G^2U_{4k}G^{-2},$$

and in three-dimensional notation,

$$\mathbf{V} = (1/2e^2m)[\mathbf{L} \times \mathbf{p} - \mathbf{p} \times \mathbf{L}] + \mathbf{r}/r.$$

It follows that $\mathbf{L}$ and $\mathbf{V}$ both commute with the Hamiltonian defined by (6.3).

The (U_{jk}, U_{i4}) together generate 0_4 while the $X_i(\pm)$ are the screw displacement operators which carry the group space of 0_3 into itself. Therefore the angular momentum and Runge–Lenz vectors are also simply related to the left- and right-handed screw motions which carry the group space of 0_3 into itself.

7. ANGULAR MOMENTUM STATES

The angular momentum, like the Hamiltonian, is not included in the original set of six generators $X_i(\pm)$. The functions $D^j_{mm'}$ are therefore not eigenfunctions of the angular momentum; in fact, the index j labels the energy and the set $D^j_{mm'}(m, m' = -j \cdots + j)$ spans the complete manifold belonging to that energy and therefore includes all values of the angular momentum from $l = 0$ up to and including $l = N - 1$. If we wish to obtain eigenvalues of the angular momentum, we must form the linear combination,

$$\varphi_{NLM}(\mathbf{p}) = \sum_{m,m'} C(NLM; jmm')D^j_{mm'}(\mathbf{p})G^2(p),$$

such that φ_{NLM} has the correct angular dependence, namely,

$$\varphi_{NLM}(\mathbf{p}) = \Pi_{NL}(p)G^2(p)\,Y_{LM}(\theta, \varphi). \tag{7.1a}$$

Therefore,

$$\Pi_{NL}(p)Y_{LM}(\theta, \varphi) = \sum_{m,m'} C(NLM; jmm')D^j_{mm'}(p). \tag{7.1b}$$

The right side, since there is no sum over j, must also satisfy (3.6a) but not (3.6b) and (3.6c). In fact, if we can write (7.1), these states must be eigenfunctions of (L^2, L_z), where

$$\mathbf{L} = \tfrac{1}{2}p_0[\mathbf{X}(+) - \mathbf{X}(-)].$$

These states may be determined by either differential or integral equations as follows.

(a) *Differential Equations:* Let

$$\Pi_{NL}(p)\,Y_{LM}(\theta, \varphi) = \chi_{NLM}(p). \tag{7.2}$$

Then

$$\Delta\chi_{NLM} = -[4j(j + 1)/R_0^2]\chi_{NLM}$$
$$= -[2(N^2 - 1)N^2/\epsilon m]\chi_{NLM}. \tag{7.3}$$

If we express Δ in stereographic coordinates, we obtain

$$\Delta = G^{-2}[\nabla^2 - 2(G/p_0^2)\mathbf{p}\nabla],$$

where

$$\nabla^2 = \sum \frac{\partial^2}{\partial p_i^2}$$
$$= \frac{1}{p^2}\frac{\partial}{\partial p}p^2\frac{\partial}{\partial p} + \frac{1}{p^2}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right),$$

$\mathbf{p}\nabla = p(\partial/\partial p)$.

Hence,

$$\left[\frac{1}{x^2}\frac{d}{dx}x^2\frac{d}{dx} - \frac{2x}{1+x^2}\frac{d}{dx} + \frac{L(L+1)}{x^2} + \frac{4(N^2-1)}{(1+x^2)^2}\right]\Pi = 0, \quad (7.4a)$$

where

$$x = p/p_0 = Np/(-2m\epsilon)^{\frac{1}{2}}.$$

The solution of (7.4a) leads to the Gegenbauer functions $C_N^\nu(x)$ as follows[6]:

$$\Pi_{NL}(p) = \left[\frac{2}{\pi}\frac{(N-L-1)!}{(N+L)!}\right]^{\frac{1}{2}} N^2 2^{2(L+1)} \times \frac{L!\,N^L p^L}{(N^2p^2+1)^{L+2}} C_{N-L+1}^{L+1}\left(\frac{N^2p^2-1}{N^2p^2+1}\right), \quad (7.5)$$

where $C_N^\nu(x)$ is the coefficient of h^N in the expansion of $(1-2hx+x^2)^{-\nu}$. The Fourier transforms of the $G^2(p)\Pi_{NL}(p)$ are the associated Laguerre functions.

The S-state solutions are spherically symmetric and, therefore, can only be the character functions,

$$\chi^j = \sum_m D^j_{mm} = \frac{\sin(2j+1)(\frac{1}{2}w)}{\sin(\frac{1}{2}w)}, \quad (7.6)$$

where w is the angle of rotation or

$$\cos\tfrac{1}{2}w = (p_0^2 - p^2)/(p_0^2 + p^2). \quad (7.7)$$

The χ^j satisfy the integral equation,

$$\chi^j(a) = \frac{d_j}{\mu^{2j}}\int K_\mu(a, a')\chi^j(a')\,d\tau', \quad (7.8)$$

and the orthogonality relations,

$$\int \chi^j(a)\chi^{j'}(a)\,d\tau_a = V\delta^{jj'}. \quad (7.9)$$

If one puts

$$\chi = \cos\tfrac{1}{2}w, \quad (7.10a)$$
$$U_{2j}(x) = \chi^j(p), \quad (7.10b)$$

then $U_{2j}(x)$ are the Tschebyscheff polynomials which satisfy by (7.9) and (7.10), the orthogonality relations,

$$\int_{-1}^{1} U_{2j}(x)U_{2j'}(x)(1-x^2)^{\frac{1}{2}}\,dx = \delta^{jj'}\tfrac{1}{2}\pi. \quad (7.11)$$

The momentum amplitude of an s state may be very simply expressed in terms of the momentum itself,

$$\varphi^j(p) = G^2(p)\chi^j(p),$$
$$\varphi_N(p) = -\frac{i}{4}\frac{p_0}{p}\left[\left(\frac{p_0+ip}{p_0-ip}\right)^N - \left(\frac{p_0-ip}{p_0+ip}\right)^N\right]. \quad (7.12)$$

Of course the $\chi^j(p)$ are the same as $\Pi_{NO}(p)$ and must also satisfy (7.4) for $L = 0$.

(b) *Integral equation.* We again make the ansatz (7.1)—but now in (5.7). Then

$$\Pi_{NL}(p)Y_{LM}(\theta, \varphi) = \frac{1}{E}\frac{e^2R_0}{\hbar}\int K_1(p, p')\Pi_{NL}(p')Y_{LM}(\theta', \varphi')\,d\tau',$$

where

$$K_1(p, p') = \frac{1}{\pi^2 p_0}\frac{1}{G(p)(|\mathbf{p}-\mathbf{p}'|)^2G(p')}.$$

Choose $\mathbf{p}$ along the z axis and $Y_{LO}(\theta, \varphi) = 1$. Then

$$\Pi_{NL}(p) = \frac{1}{2\pi^2}\frac{e^2}{\hbar E}\int \frac{\Pi_{NL}(p')[g(p')]^{\frac{1}{2}}p'^2\,dp'}{G(p')G(p)} \times \int \frac{Y_{LO}(\theta', \varphi')\,d\mu'\,d\varphi'}{p^2+p'^2-2pp'\mu}$$
$$= \frac{1}{2\pi^2}\frac{e^2}{\hbar E}\int \frac{\Pi_{NL}(p')[g(p')]^{\frac{1}{2}}p'^2\,dp'}{G(p')G(p)} \times \left[\frac{2\pi}{pp'}Q_L\left(\frac{p^2+p'^2}{2pp'}\right)\right]$$

or

$$\Pi_{NL}(p) = \frac{e^2}{\pi\hbar E}\int_0^\infty k(p, p')\Pi_{NL}(p')\,d\tau_p', \quad (7.13)$$

where

$$d\tau_p = [g(p)]^{\frac{1}{2}}p^2\,dp, \quad (7.13a)$$

$$k(p, p') = \frac{1}{pp'}\frac{1}{G(p)G(p')}Q_L\left(\frac{p^2+p'^2}{2pp'}\right)$$
$$= \frac{p_0^2+p^2}{p_0^2p}\frac{p_0^2+p'^2}{p_0^2p'}Q_L\left(\frac{p^2+p'^2}{2pp'}\right). \quad (7.13b)$$

The solution of (7.13) for general L are again the Gegenbauer functions while the special case $L = 0$ leads to the Tschebyscheff functions.

8. GENERAL COORDINATE SYSTEM

The fundamental correspondence between the Schrödinger equation in the momentum representation

[6] H. Bethe and E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic Press Inc., New York, 1957).

and an integral equation on the group space has so far been discussed in a particular coordinate system (the stereographic system). However, since this integral equation has been written in a generally covariant form, it is possible to introduce other coordinate systems in the group space. For example, we may parameterize the rotations by Eulerian angles (α, β, γ) and adopt these coordinates in group space, and therefore in momentum space.

The differential equations are now

$$X^2(\pm)D^N_{MM'} = -[(N^2-1)/R_0^2]D^N_{MM'}, \quad (8.1)$$

$$X_3(+)D^N_{MM'} = (2M/R_0)D^N_{MM'}, \quad (8.2+)$$

$$X_3(-)D^N_{MM'} = (2M'/R_0)D^N_{MM'}, \quad (8.2-)$$

where

$$R_0^2 = \tfrac{1}{2}(m\epsilon/N^2).$$

In order to express either the integral or differential equation in a particular coordinate system, one needs a general method for going from an arbitrary parameterization of the group to the metric and connection of the group space. This may be done as follows.[5]

Let $U(a)$ be some representation of the group, where a is an arbitrary coordinate system in the group space. The fundamental triad fields may be expressed as follows:

$$\lambda^i_\alpha(\pm) = R_0 \operatorname{Tr} \Lambda_\alpha(\pm)\sigma^i, \quad (8.3)$$

where

$$\Lambda_\alpha(+) = -i(\partial U/\partial a_\alpha)U^{-1}, \quad (8.4+)$$

$$\Lambda_\alpha(-) = -iU^{-1}(\partial U/\partial a_\alpha) = i(\partial U^{-1}/\partial a_\alpha)U. \quad (8.4-)$$

It is then possible with the aid of (3.1) and (3.2) to calculate the metric and connection and also to obtain

$$g_{\alpha\beta} = R_0^2 \operatorname{Tr} \Lambda_\alpha \Lambda_\beta, \quad (8.5)$$

$$L^\mu_{\alpha\beta} = R_0^2 \operatorname{Tr} \Lambda^\mu \partial_\beta \Lambda_\alpha. \quad (8.6)$$

We may illustrate these formulas with Eulerian coordinates. Then

$$U = U_3(\tfrac{1}{2}\alpha)U_2(\tfrac{1}{2}\beta)U_3(\tfrac{1}{2}\gamma), \quad (8.7)$$

where

$$U_m(\theta) = \exp(i\sigma_m\theta). \quad (8.7a)$$

We find

$$\partial U/\partial\alpha = (\tfrac{1}{2}i\sigma_3)U, \quad (8.8a)$$

$$\partial U/\partial\beta = U_3(\tfrac{1}{2}\alpha)U_2(\tfrac{1}{2}\beta)U_3(-\tfrac{1}{2}\gamma)(\tfrac{1}{2}i\sigma_3), \quad (8.8b)$$

$$\partial U/\partial\gamma = U(\tfrac{1}{2}i\sigma_3), \quad (8.8c)$$

and therefore

$$\Lambda_\alpha(+) = \tfrac{1}{2}\sigma_3, \quad (8.9a)$$

$$\Lambda_\beta(+) = U_3(\alpha)(\tfrac{1}{2}\sigma_2), \quad (8.9b)$$

$$\Lambda_\gamma(+) = U(\tfrac{1}{2}\sigma_3)U^{-1}. \quad (8.9c)$$

To obtain $\Lambda_\mu(-)$, calculate the right side for U^{-1} instead of U. By (8.3) one finds

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & \cos\beta \\ 0 & 1 & 0 \\ \cos\beta & 0 & 1 \end{pmatrix}, \quad (8.10a)$$

$$g^{\mu\nu} = \frac{1}{R_0^2 \sin^2\beta}\begin{pmatrix} 1 & 0 & -\cos\beta \\ 0 & \sin^2\beta & 0 \\ -\cos\beta & 0 & 1 \end{pmatrix}, \quad (8.10b)$$

$$g^{\frac{1}{2}} = \tfrac{1}{8}R_0^3 \sin\beta. \quad (8.11)$$

Finally,

$$\nabla = (1/g^{\frac{1}{2}})\partial_\mu g^{\frac{1}{2}}\partial^\mu$$

$$= \frac{1}{\sin^2\beta}\left(\frac{\partial^2}{\partial\alpha^2} + \frac{\partial^2}{\partial\gamma^2} - 2\cos\beta\frac{\partial^2}{\partial\alpha\partial\gamma}\right) + \frac{\partial^2}{\partial\beta^2} + \cot\beta\frac{\partial}{\partial\beta}, \quad (8.12)$$

$$X_3(+) = \lambda^\mu_3(+)\partial_\mu = (2/R_0)(\partial/\partial\alpha), \quad (8.13)$$

$$X_3(-) = \lambda^\mu_3(-)\,\partial_\mu = (2/R_0)(\partial/\partial\gamma), \quad (8.14)$$

so that (8.1), (8.2−) may be expressed in Eulerian coordinates.

These equations have been separated in other coordinate systems by Wenger,[7] who has also shown that the Eulerian system has a geometrical interpretation in terms of cylindrical coordinates.

The different possible coordinate systems are most simply related through the spin representation. Thus if we write

$$\exp(\tfrac{1}{2}i\mathbf{w}\boldsymbol{\sigma}) = \frac{1 + \frac{1}{2}i\boldsymbol{\sigma}\mathbf{r}/R_0}{1 - \frac{1}{2}i\boldsymbol{\sigma}\mathbf{r}/R_0} \quad (8.15)$$

$$= \begin{pmatrix} \cos\tfrac{1}{2}\beta\exp[\tfrac{1}{2}i(\alpha+\gamma)] & \sin\tfrac{1}{2}\beta\exp[\tfrac{1}{2}i(\alpha-\gamma)] \\ -\sin\tfrac{1}{2}\beta\exp[\tfrac{1}{2}i(\gamma-\alpha)] & \cos\tfrac{1}{2}\beta\exp[-\tfrac{1}{2}i(\alpha+\gamma)] \end{pmatrix}, \quad (8.16)$$

then the $\mathbf{w}$, $\mathbf{r}$, and (α, β, γ) are Riemannian, stereographic, and Eulerian coordinates, respectively. Then we find, for example, the following relation between the Riemannian and Eulerian coordinates:

$$(w_3/w)\sin\tfrac{1}{2}w = \cos\tfrac{1}{2}\beta\sin\tfrac{1}{2}(\alpha+\gamma), \quad (8.17a)$$

$$(w_\perp/w)\sin\tfrac{1}{2}w = \sin\tfrac{1}{2}\beta, \quad (8.17b)$$

where

$$\cos\tfrac{1}{2}w = \cos\tfrac{1}{2}\beta\cos\tfrac{1}{2}(\alpha+\gamma). \quad (8.17c)$$

Let us finally note that the invariant kernel of the integral equation now takes the following form, instead of (4.10),

$$K_\mu(a, a') = \frac{1}{1 + \mu^2 - 2\mu\cos\frac{1}{2}\beta\cos\frac{1}{2}(\alpha+\gamma)}. \quad (8.18)$$

[7] D. Wenger, J. Math. Phys. (to be published).

9. SYMMETRIC TOP

The correspondence between the Eulerian angles and the components of $\mathbf{p}$ are determined by (8.15) and (8.16) as follows:

$$\begin{pmatrix} \exp[\frac{1}{2}i(\alpha+\gamma)]\cos\frac{1}{2}\beta & \exp[\frac{1}{2}i(\alpha-\gamma)]\sin\frac{1}{2}\beta \\ -\exp[-\frac{1}{2}i(\alpha-\gamma)]\sin\frac{1}{2}\beta & \exp[-\frac{1}{2}i(\alpha+\gamma)]\cos\frac{1}{2}\beta \end{pmatrix} = \begin{pmatrix} p_0^2 - p^2 + 2ip_0p_3 & 2ip_0(p_1 - ip_2) \\ 2ip_0(p_1 + ip_2) & p_0^2 - p^2 - 2ip_0p_3 \end{pmatrix} \frac{1}{p_0^2 + p^2}. \tag{9.1}$$

Therefore,

$$\sin\tfrac{1}{2}\beta = 2p_0p_\perp/(p_0^2 + p^2), \tag{9.2a}$$

$$\tan\tfrac{1}{2}(\alpha - \gamma) = p_1/p_2, \tag{9.2b}$$

$$\tan\tfrac{1}{2}(\alpha + \gamma) = 2p_0p_3/(p_0^2 - p^2). \tag{9.2c}$$

Note that the character of the spin representation is

$$\chi^{\frac{1}{2}}(w) = 2\cos\tfrac{1}{2}w = 2\cos\tfrac{1}{2}(\alpha+\gamma)\cos\tfrac{1}{2}\beta;$$

Therefore

$$\cos\tfrac{1}{2}w = (p_0^2 - p^2)/(p_0^2 + p^2), \tag{9.3}$$

where w is the magnitude of the rotation.

In the Eulerian system we may solve the differential equations (8.12), (8.13), (8.14) simply by $D^j_{mn}(\alpha\beta\gamma)$, where m and n are the eigenvalues of the operators $X_3(+)$ and $X_4(-)$. But the vector operators $\mathbf{X}(+)$ and $\mathbf{X}(-)$ obey just the same commutation rules as the angular momenta of a top with respect to body-fixed and space-fixed axes. Therefore $D^j_{mn}(\alpha\beta\gamma)$ may be interpreted as the state function of a symmetric top where m and n are the quantum numbers giving the z component of angular momentum with respect to body-fixed and space-fixed axes. These functions may be expressed in the familiar form

$$D^j_{mn}(\alpha\beta\gamma) = \exp(im\alpha)\, d^j_{mn}(\beta)\exp(in\gamma),$$

where

$$d^j_{mn}(\beta) = \left[\frac{(j+m)!\,(j-m)!}{(j+n)!\,(j-n)!}\right]^{\frac{1}{2}} \times (\cos\tfrac{1}{2}\beta)^{m+n}(\sin\tfrac{1}{2}\beta)^{m-n}P^{m-n,\,m+n}_{j-m}(\cos\beta).$$

The orthogonality statements (3.7) now factor into the familiar exponential relations and the corresponding equation for the Jacobi polynomials $P^{(\mu,\nu)}_m$. That is, from

$$\frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi} \bar{D}^j_{mn}(\alpha\beta\gamma)D^{j'}_{m'n'}(\alpha\beta\gamma)\sin\beta\, d\alpha\, d\beta\, d\gamma = \delta_{mm'}\delta_{nn'}\delta^{jj'}\, d_j^{-1},$$

one finds the following orthogonality relations:

$$\int_{-1}^{1}(1-x)^\mu(1+x)^\nu P^{(\mu,\nu)}_n(x)P^{(\mu,\nu)}_m(x)\, dx = \frac{2^{\mu+\nu+1}}{2n+\mu+\nu+1}\frac{\Gamma(n+\mu+1)\Gamma(n+\nu+1)}{\Gamma(n+1)\Gamma(n+\mu+\nu+1)}\delta_{nm},$$

as well as the corresponding relations for the exponential functions.

The problems of the top and the Coulomb field are then reciprocal in the sense that momentum space and configuration space are interchanged. The correlation of the states with the energy is of course different in the two cases.

10. CONFIGURATION SPACE

Momentum space is natural for the Coulomb problem since it is isomorphic to the group space. In momentum space the natural operators are the generators $X_i(+)$ and $X_i(-)$, in terms of which

$$\mathbf{L} = \tfrac{1}{2}p_0[\mathbf{X}(+) - \mathbf{X}(-)],$$

$$\mathbf{V} = (\hbar p_0/4e^2m)G^2[\mathbf{X}(+) + \mathbf{X}(-)]G^{-2},$$

where $\mathbf{L}$, $\mathbf{V}$ are the angular momentum and the Runge–Lenz vectors. If one diagonalizes L^2, L_3 one gets Gegenbauer functions in momentum space and Laguerre functions in configuration space; in this case the natural coordinate systems in both momentum and configuration space are spherical. On the other hand, if one diagonalizes $X_3(+)$ and $X_3(-)$, one obtains

$$\varphi(p) \sim G^2D^j_{mn}, \quad V_3\varphi(p) \sim (m + m')\varphi(p).$$

In this case one is led to the Wigner functions in momentum space. The corresponding natural coordinate system is (α, β, γ) since

$$X_3(+) \sim \partial/\partial\alpha, \quad X_3(-) \sim \partial/\partial\gamma.$$

In configuration space the corresponding natural coordinate system is parabolic and the corresponding functions are the confluent hypergeometric functions.[2]

The statements of this paragraph as well as of Secs. 6, 7, and 9 relate known results[8] to our use of the geometry of 0_3.

ACKNOWLEDGMENTS

We wish to thank the referee of this paper for having pointed out to us that the general potential ($\mu \neq 1$) has been discussed in a University of Chicago preprint.[9]

This work was supported in part by the National Science Foundation.

[8] See the review papers of M. Bander and C. Itzykson [Rev. Mod. Phys. **38**, 330, 346 (1966)].

[9] M. Luming and E. Predazzi (EFINS 66-24). Our own results including the exactly soluble potential were obtained in August (1965).

Variational Theorem for Reduced Density Matrices

HSIAO S. KIANG
Department of Chemistry, Indiana University, Bloomington, Indiana

(Received 6 June 1966)

This paper discusses a theorem concerning the variational description of the eigenfunctions and eigenvalues of the two complementary reduced density matrices for a many-particle system in a bound state.

IN this paper, we discuss a theorem concerning the variational description of the eigenfunctions and eigenvalues of the two complementary reduced density matrices for a many-particle system in a bound state. This theorem is the following:

Theorem: Given a bound state $|\psi\rangle$ of an N-particle system, any state vector $|p\rangle$ of a m-particle subsystem ($m < N$), for which the functional $\|\langle\psi \mid p\rangle\|$ is stationary, is an eigenvector of the reduced density operator D_p associated with the subsystem. The stationary $\|\langle\psi \mid p\rangle\|$ gives the corresponding eigenvalue to which the eigenvector belongs. $\langle\psi \mid p\rangle$ corresponds to an eigenvector belonging to the same eigenvalue of the reduced density operator D_q associated with the (N–m)-particle subsystem. The operators D_p and D_q in the Schrödinger representation correspond to the Dirac density matrices.[1]

Proof: Consider the N-particle system to be composed of two interacting subsystems, each consisting of m and N–m particles and each with associated Hilbert spaces E_p and E_q, respectively. Take the Hilbert space E for the system to be the tensor product of E_p and E_q($E = E_p \otimes E_q$).

Consider the stationary conditions of the functional

$$S = \langle pq \mid \psi\rangle$$

of the space E, where $|\psi\rangle$ is a given unit vector, and $|p\rangle$ and $|q\rangle$ are some arbitrary unit vectors in E_p and E_q. The quantity S is, of course, the amplitude of finding the state $|\psi\rangle$ of the N-particle system in the state $|pq\rangle$.

By means of the method of Lagrange multipliers, the stationary condition of S reads

$$\delta S + \lambda_p \delta\langle p \mid p\rangle + \lambda_q \delta\langle q \mid q\rangle = 0, \tag{1}$$

where λ_p and λ_q are the Lagrangian undetermined constants.

Upon calculating the variations and rearranging terms,

$$\langle\delta q|\,[\langle p \mid \psi\rangle + \lambda_q |q\rangle] + \langle q|\,\lambda_q\,|\delta q\rangle + \langle\delta p|[\langle q \mid \psi\rangle + \lambda_p\,|p\rangle] + \langle p|\,\lambda_p\,|\delta p\rangle = 0. \tag{2}$$

Since $|\delta p\rangle$ and $|\delta q\rangle$ are arbitrary, one may replace the kets $|\delta p\rangle$ by $i|\delta p\rangle$ and $|\delta q\rangle$ by $i|\delta q\rangle$, and the bras $\langle\delta p|$ by $-i\langle\delta p|$ and $\langle\delta p|$ by $-i\langle\delta q|$, where $i = (-1)^{\frac{1}{2}}$. Taking a linear combination of the new equation with (2), results the nontrivial equation

$$\langle\delta q|\,[\langle p \mid \psi\rangle + \lambda_q|q\rangle] + \langle\delta p|\,[\langle q \mid \psi\rangle + \lambda_p|\,p\rangle] = 0, \tag{3}$$

where $\langle\delta p|$ and $\langle\delta q|$ are also independent of each other. It turns out that condition (1) satisfies the two following simultaneous equations:

$$\langle p \mid \psi\rangle + \lambda_p|q\rangle = 0, \tag{4}$$

$$\langle q \mid \psi\rangle + \lambda_q|p\rangle = 0. \tag{5}$$

With the aid of the normalization constraints on $|p\rangle$ and $|q\rangle$, the two Lagrangian undetermined constants λ_p and λ_q are now determined to be equal to the negative of the stationary S. Hence, we write both as $-\lambda$ and Eqs. (4) and (5) become

$$\langle p \mid \psi\rangle = \lambda|q\rangle, \tag{6}$$

$$\langle q \mid \psi\rangle = \lambda|p\rangle. \tag{7}$$

The above two equations were previously[2–4] derived from the theory of integral equations and are valid only when the kernel of the integral operator is symmetric in a finite-dimensional space. In this case, it is known that the eigenvectors $|p_i\rangle$, $|q_i\rangle$ belonging to the eigenvalue $|\lambda_i|^2$ of the respective density operators D_p and D_q associated with the two subsystems satisfy this theorem. In fact, one may alternatively adopt a variational definition of the reduced density matrices by taking the totality of the set of independently admissible solutions from (6) and (7) to construct the two complementary reduced density operators

$$\mathsf{D}_p = \sum_i |\lambda_i|^2\,|p_i\rangle\langle p_i|,$$

$$\mathsf{D}_q = \sum_i |\lambda_i|^2\,|q_i\rangle\langle q_i|. \qquad \text{Q.E.D.}$$

[1] P. A. M. Dirac, Proc. Cambridge Phil. Soc. **27**, 240 (1930).

[2] B. C. Carlson and J. M. Keller, Phys. Rev. **121**, 659 (1961).

[3] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).

[4] T. Ando, Rev. Mod. Phys. **35**, 690 (1963).

It may be noted that, due to the following identity,

$$\langle \psi - pq \mid \psi - pq \rangle = \langle \psi \mid \psi \rangle + \langle pq \mid pq \rangle - 2 \operatorname{Re} \langle pq \mid \psi \rangle,$$

the solutions of $|pq\rangle$ stationary to S also give the stationary solutions to the mean-square deviation of $|pq\rangle$ from $|\psi\rangle$. A discussion in terms of the latter quantity was given by Coleman.[3] Much of the motivation for the present work, however, came from earlier discussion by Löwdin and Shull[5] of the overlap properties in connection with the first-order density matrix.

The theorem presented in this paper may be applied to the direct computation of the largest eigenvalues of reduced density matrices by a simple iteration procedure. We have, in particular, applied the variational generating equations (6) and (7), of these eigenvalue problems to the natural expansion of a many-electron wavefunction. A detailed discussion of this work with illustrative examples will be represented elsewhere.[6]

ACKNOWLEDGMENTS

The author wishes to express his sincere gratitude to his advisor, Professor H. Shull, and to Professor S. Hagstrom for helpful discussions.

This paper is a part of the author's Ph.D. Thesis submitted to the Graduate School of the Indiana University. This research was in part supported by the National Science Foundation and by the U.S. Air Force through the Office of Scientific Research.

[5] P. O. Löwdin and H. Shull, Phys. Rev. **101**, 1730 (1956).

[6] H. S. Kiang, Phys. Rev. (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Relativistic Partial Wave Analysis of Electron Scattering

C. R. FISCHER AND R. AKERIB
Queens College of the City University of New York, Flushing, New York

(Received 14 July 1966)

The partial wave series for the scattering amplitude for high-energy electron scattering is not uniformly convergent. The singularity responsible for the nonuniform convergence containing terms in $(\sin \frac{1}{2}\theta)^{2i\alpha-2}$, $(\sin \frac{1}{2}\theta)^{2i\alpha-1}$, and $(\sin \frac{1}{2}\theta)^{2i\alpha}$ is separated from the rest of the series so that an accurate partial wave analysis may be carried out for any scattering angle.

I. INTRODUCTION

FOR high-energy electron scattering the long-range nature of the Coulomb interaction causes the partial wave series for the amplitude to converge slowly. In some calculations[1,2] this difficulty was overcome by calculating $f(\theta) = f_{\text{Coul}}(\theta) + [f(\theta) - f_{\text{Coul}}(\theta)]$, where $f_{\text{Coul}}(\theta)$ is the point Coulomb scattering amplitude. This method is analogous to the one used to calculate nonrelativistic nuclear scattering when the Coulomb interaction is present.[3] In both instances the series represented by the term $f(\theta) - f_{\text{Coul}}(\theta)$ converges. However, for relativistic scattering of Dirac particles the point Coulomb amplitude is not available in closed form, although an accurate evaluation can be made.[4,5] Another method was developed by Yennie *et al.*[6] who considered the expansion of the function $(1 - \cos\theta)^n f(\theta)$. This method works well except at small angles where many partial waves are required.

Herman *et al.*[2] have suggested that the singular part of the scattering amplitude may be explicitly separated and summed so that a reliable partial wave calculation can be carried out at any scattering angle. Hetherington[7] attempted to do this for the Klein–Gordon equation. He separated the nonrelativistic point

[1] L. R. B. Elton, Proc. Phys. Soc. (London) **A63**, 1115 (1950).

[2] R. Herman, B. C. Clark, and D. G. Ravenhall, Phys. Rev. **132**, 414 (1963).

[3] See for example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), 2nd ed., Sec. 20.

[4] J. H. Bartlett and R. E. Watson, Proc. Am. Acad. Arts Sci. **74**, 53 (1940).

[5] See, for example, R. L. Gluckstern and S. R. Lin, J. Math. Phys. **5**, 1954 (1964); W. A. McKinley and H. Feshbach, Phys. Rev. **74**, 1759 (1948).

[6] D. R. Yennie, D. G. Ravenhall, and R. M. Wilson, Phys. Rev. **95**, 500 (1954).

[7] J. H. Hetherington, Ph.D. thesis, University of Illinois (1960) (unpublished); J. H. Hetherington, J. Math. Phys. **4**, 357 (1963).

It may be noted that, due to the following identity,

$$\langle \psi - pq \mid \psi - pq \rangle = \langle \psi \mid \psi \rangle + \langle pq \mid pq \rangle - 2 \operatorname{Re} \langle pq \mid \psi \rangle,$$

the solutions of $|pq\rangle$ stationary to S also give the stationary solutions to the mean-square deviation of $|pq\rangle$ from $|\psi\rangle$. A discussion in terms of the latter quantity was given by Coleman.[3] Much of the motivation for the present work, however, came from earlier discussion by Löwdin and Shull[5] of the overlap properties in connection with the first-order density matrix.

The theorem presented in this paper may be applied to the direct computation of the largest eigenvalues of reduced density matrices by a simple iteration procedure. We have, in particular, applied the variational generating equations (6) and (7), of these eigenvalue problems to the natural expansion of a many-electron wavefunction. A detailed discussion of this work with illustrative examples will be represented elsewhere.[6]

ACKNOWLEDGMENTS

The author wishes to express his sincere gratitude to his advisor, Professor H. Shull, and to Professor S. Hagstrom for helpful discussions.

This paper is a part of the author's Ph.D. Thesis submitted to the Graduate School of the Indiana University. This research was in part supported by the National Science Foundation and by the U.S. Air Force through the Office of Scientific Research.

[5] P. O. Löwdin and H. Shull, Phys. Rev. **101**, 1730 (1956).

[6] H. S. Kiang, Phys. Rev. (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Relativistic Partial Wave Analysis of Electron Scattering

C. R. FISCHER AND R. AKERIB
Queens College of the City University of New York, Flushing, New York

(Received 14 July 1966)

The partial wave series for the scattering amplitude for high-energy electron scattering is not uniformly convergent. The singularity responsible for the nonuniform convergence containing terms in $(\sin \frac{1}{2}\theta)^{2i\alpha-2}$, $(\sin \frac{1}{2}\theta)^{2i\alpha-1}$, and $(\sin \frac{1}{2}\theta)^{2i\alpha}$ is separated from the rest of the series so that an accurate partial wave analysis may be carried out for any scattering angle.

I. INTRODUCTION

FOR high-energy electron scattering the long-range nature of the Coulomb interaction causes the partial wave series for the amplitude to converge slowly. In some calculations[1,2] this difficulty was overcome by calculating $f(\theta) = f_{\text{Coul}}(\theta) + [f(\theta) - f_{\text{Coul}}(\theta)]$, where $f_{\text{Coul}}(\theta)$ is the point Coulomb scattering amplitude. This method is analogous to the one used to calculate nonrelativistic nuclear scattering when the Coulomb interaction is present.[3] In both instances the series represented by the term $f(\theta) - f_{\text{Coul}}(\theta)$ converges. However, for relativistic scattering of Dirac particles the point Coulomb amplitude is not available in closed form, although an accurate evaluation can be made.[4,5] Another method was developed by Yennie *et al.*[6] who considered the expansion of the function $(1 - \cos\theta)^n f(\theta)$. This method works well except at small angles where many partial waves are required.

Herman *et al.*[2] have suggested that the singular part of the scattering amplitude may be explicitly separated and summed so that a reliable partial wave calculation can be carried out at any scattering angle. Hetherington[7] attempted to do this for the Klein–Gordon equation. He separated the nonrelativistic point

[1] L. R. B. Elton, Proc. Phys. Soc. (London) **A63**, 1115 (1950).

[2] R. Herman, B. C. Clark, and D. G. Ravenhall, Phys. Rev. **132**, 414 (1963).

[3] See for example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), 2nd ed., Sec. 20.

[4] J. H. Bartlett and R. E. Watson, Proc. Am. Acad. Arts Sci. **74**, 53 (1940).

[5] See, for example, R. L. Gluckstern and S. R. Lin, J. Math. Phys. **5**, 1954 (1964); W. A. McKinley and H. Feshbach, Phys. Rev. **74**, 1759 (1948).

[6] D. R. Yennie, D. G. Ravenhall, and R. M. Wilson, Phys. Rev. **95**, 500 (1954).

[7] J. H. Hetherington, Ph.D. thesis, University of Illinois (1960) (unpublished); J. H. Hetherington, J. Math. Phys. **4**, 357 (1963).

Coulomb amplitude and a term containing a factor $[\sin\frac{1}{2}\theta]^{-2i\lambda-1}$, where $\lambda = Z_1Z_2\alpha/\beta$ and α and β are the fine structure constant and v/c. However, his series is not absolutely convergent, since it still contains a singular factor $[\sin\frac{1}{2}\theta]^{-2i\lambda}$ due to his neglect of the phase factor which increases logarithmically with l in his $\mathcal{O}(1/l)$ term.

We have separated and summed in closed form all the singular parts of the series for $f(\theta)$ for high-energy electron scattering in the limit of zero electron mass. The series to be evaluated is then absolutely convergent. It should be noted that the resulting form $f(\theta) = f_{\text{sing}}(\theta) +$ absolutely convergent series contains of course the same singularities as $f_{\text{Coul}}(\theta)$, the difference being in the nonsingular parts, i.e., $f_{\text{sing}}(\theta)$ is a closed form while $f_{\text{Coul}}(\theta)$ is not.[4]

II. CALCULATION

In the high-energy limit the Dirac scattering amplitude is[6]

$$f(\theta) = \frac{1}{2i\kappa}\sum_{k=1}^{\infty} ke^{2i\eta_k}[P_k(\cos\theta) + P_{k-1}(\cos\theta)], \quad (1)$$

where η_k, the phase shift, is the sum of a "nuclear" part δ_k and the relativistic Coulomb phase shift χ_k given by

$$ke^{2i\chi_k} = \frac{\Gamma(\rho_k + 1 - i\alpha)}{\Gamma(\rho_k + i\alpha)}\, e^{i\pi(k-\rho_k)}.$$

Here $\alpha = Ze^2/(\hbar c)$ and $\rho_k = (k^2 - \alpha^2)^{\frac{1}{2}}$. The δ_k fall off rapidly with increasing k and only the Coulomb phase shifts need be considered.

To expand the factor $ke^{2i\chi_k}$ in inverse powers of k, we write

$$e^{i\pi(k-\rho_k)} = 1 + \frac{i\pi\alpha^2}{2k} - \frac{\pi^2\alpha^4}{8k^2} + \mathcal{O}\left(\frac{1}{k^3}\right),$$

and we also expand $\Gamma(\rho_k + 1 - i\alpha)/\Gamma(\rho_k + i\alpha)$ in a Taylor series about $\rho = k$. Employing the asymptotic form for $\psi(z) = \Gamma'(z)/\Gamma(z)$ we obtain

$$\frac{\Gamma(\rho_k + 1 - i\alpha)}{\Gamma(\rho_k + i\alpha)} = \frac{\Gamma(k + 1 - i\alpha)}{\Gamma(k + i\alpha)} \times \left[1 - \frac{1}{2}\frac{\alpha^2(1 - 2i\alpha)}{k^2} + \mathcal{O}\left(\frac{1}{k^3}\right)\right].$$

Combining the results we obtain

$$ke^{2i\chi_k} = \frac{\Gamma(k + 1 - i\alpha)}{\Gamma(k + 1 + i\alpha)}\left[k + i\alpha(1 + \pi\alpha) + \frac{\alpha^2}{k}\left(i\alpha - \tfrac{1}{2}\pi\alpha - \frac{1}{2} - \frac{\pi^2\alpha^2}{8}\right) + \mathcal{O}\left(\frac{1}{k^2}\right)\right]. \quad (2)$$

When Eq. (2) is substituted into Eq. (1), $f(\theta)$ may be rewritten in analogy with the nonrelativistic case as follows:

$$f(\theta) = f_1(\theta) + f_2(\theta) + f_3(\theta) + \frac{1}{2i\kappa}\sum_{k=1}^{\infty}\left\{ke^{2i(\delta_k+\chi_k)} - \frac{\Gamma(k + 1 - i\alpha)}{\Gamma(k + 1 + i\alpha)} \times \left[k + i\alpha(1 + \tfrac{1}{2}\pi\alpha) + \frac{\alpha^2}{k - i\alpha} \times \left(i\alpha - \tfrac{1}{2}\pi\alpha - \frac{1}{2} - \frac{\pi^2\alpha^2}{8}\right)\right]\right\} \times [P_k(\cos\theta) + P_{k-1}(\cos\theta)], \quad (3)$$

where

$$f_1(\theta) = \frac{1}{2i\kappa}\sum_{k=1}^{\infty} k\frac{\Gamma(k + 1 - i\alpha)}{\Gamma(k + 1 + i\alpha)} \times [P_k(\cos\theta) + P_{k-1}(\cos\theta)],$$

$$f_2(\theta) = \frac{\alpha}{2\kappa}(1 + \tfrac{1}{2}\pi\alpha)\sum_{k=1}^{\infty}\frac{\Gamma(k + 1 - i\alpha)}{\Gamma(k + 1 + i\alpha)} \times [P_k(\cos\theta) + P_{k-1}(\cos\theta)],$$

$$f_3(\theta) = \frac{\alpha^2}{2\kappa}\left(i\alpha - \tfrac{1}{2}\pi\alpha - \frac{1}{2} - \frac{\pi^2\alpha^2}{8}\right)\sum_{k=1}^{\infty}\frac{1}{k - i\alpha} \times \frac{\Gamma(k + 1 - i\alpha)}{\Gamma(k + 1 + i\alpha)}[P_k(\cos\theta) + P_{k-1}(\cos\theta)].$$

The series in Eq. (3) is absolutely convergent since its terms are $\mathcal{O}(1/k^2)$.

It is possible to separate the nonrelativistic point Coulomb amplitude from the first term, but we do not do this since all of the above sums may be evaluated by the same method as follows. Labeling the sums which appear in $f_1(\theta)$, $f_2(\theta)$, and $f_3(\theta)$, $\sum_1$, $\sum_2$, and $\sum_3$, respectively, we find by employing the definition of the beta function

$$\sum_1 = \frac{1}{\Gamma(2i\alpha)}\sum_{k=1}^{\infty} k[P_k(z) + P_{k-1}(z)] \times \int_0^1 t^{k-i\alpha}(1 - t)^{2i\alpha-1}\, dt, \quad (4)$$

where $z = \cos\theta$. The integral may be made to converge by adding ϵ to the exponents. Using

$$\sum_{k=0}^{\infty} t^k P_k(z) = \frac{1}{(1 - 2tz + t^2)^{\frac{1}{2}}}$$

and

$$t\frac{\partial}{\partial t}\left[t\sum_{k=0}^{\infty} t^k P_k(z)\right] = \sum_{k=1}^{\infty} kt^k P_{k-1}(z)$$

we find after some algebra

$$\sum_1 = \frac{(1 + z)4^{i\alpha}}{\Gamma(2i\alpha)}\int_0^1 \frac{[(1 - t)^2/4t]^{i\alpha}t\, dt}{\{[(1 - t)^2/4t] + \frac{1}{2}(1 - z)\}^{\frac{3}{2}}(4t)^{\frac{3}{2}}}. \quad (5)$$

Making the substitution $u = (1-t)^2/4t$ and factoring out appropriate powers of z, we obtain

$$\sum_1 = \frac{4^{i\alpha-1}\cos^2\frac{1}{2}\theta}{\Gamma(2i\alpha)}\Big[(\sin\tfrac{1}{2}\theta)^{i\alpha-2}\int_0^\infty u^{i\alpha-\frac{1}{2}}(1+u)^{-\frac{3}{2}}\,du - (\sin\tfrac{1}{2}\theta)^{i\alpha-1}\int_0^\infty u^{i\alpha}(1+u)^{-\frac{3}{2}}(1+u\sin^2\tfrac{1}{2}\theta)^{-\frac{1}{2}}\,du\Big]. \tag{6}$$

Recognizing the first integral as a beta function and the second as a hypergeometric function which is then written in terms of hypergeometric functions with argument $\sin^2(\frac{1}{2})\theta$ to obtain a form useful for small θ, and using the duplication formula for the gamma functions, the result is

$$\sum_1 = \frac{i\alpha\cos^2\frac{1}{2}\theta\Gamma(1-i\alpha)}{(\sin\frac{1}{2}\theta)^{2-2i\alpha}\Gamma(1+i\alpha)} \times\Big\{1-\frac{\sin\frac{1}{2}\theta}{4}\frac{\Gamma(1+i\alpha)}{\Gamma(1-i\alpha)}\frac{\Gamma(\frac{1}{2}-i\alpha)}{\Gamma(\frac{1}{2}+i\alpha)} \times F(\tfrac{1}{2},1+i\alpha,\tfrac{1}{2}+i\alpha;\sin^2\tfrac{1}{2}\theta)+\frac{(\cos\frac{1}{2}\theta)^{1-2i\alpha}}{2i\alpha-1} \times F(\tfrac{3}{2},1-i\alpha,\tfrac{3}{2}-i\alpha;\sin^2\tfrac{1}{2}\theta)\Big]\Big\}, \tag{7}$$

where $F(a, b, c; z)$ is the hypergeometric function.

The singular part may be explicitly exhibited by writing out the hypergeometric series to obtain

$$\sum_1 = \frac{i\alpha\cos^2\frac{1}{2}\theta}{(\sin\frac{1}{2}\theta)^{2-2i\alpha}}\frac{\Gamma(1-i\alpha)}{\Gamma(1+i\alpha)} \times\Big[1-\frac{\Gamma(1+i\alpha)}{\Gamma(1-i\alpha)}\frac{\Gamma(\frac{1}{2}-i\alpha)}{\Gamma(\frac{1}{2}+i\alpha)}\sin\tfrac{1}{2}\theta + (\cos\tfrac{1}{2}\theta)^{1-2i\alpha}\frac{\sin\frac{1}{2}\theta}{1-2i\alpha}+\cdots\Big]. \tag{8}$$

We see that $\sum_1$ contains the singularity associated with nonrelativistic point Coulomb scattering plus lower-order singular terms in $\sin\frac{1}{2}\theta$.

The second sum $\sum_2$ may be evaluated in the same way provided an infinitesimal is added to the exponent in the denominator of the first integral in the expression analogous to Eq. (6). The result is

$$\sum_2 = -\frac{\Gamma(1-i\alpha)}{\Gamma(1+i\alpha)}-(\sin\tfrac{1}{2}\theta)^{2i\alpha}-\frac{2i\sinh\pi\alpha}{\pi} \times\frac{\Gamma(\frac{1}{2}-i\alpha)}{\Gamma(\frac{1}{2}+i\alpha)}\cos^2\tfrac{1}{2}\theta(\sin\tfrac{1}{2}\theta)^{2i\alpha-1} \times F(\tfrac{1}{2},1+i\alpha,\tfrac{1}{2}-i\alpha;\sin^2\tfrac{1}{2}\theta) +\frac{2\Gamma(\frac{1}{2}-i\alpha)}{\pi^{\frac{1}{2}}\Gamma(1+i\alpha)}(\cos\tfrac{1}{2}\theta)^{1-i\alpha} \times F(\tfrac{3}{2},1-i\alpha,{}_2-i\alpha,\sin^2\tfrac{1}{2}\theta). \tag{9}$$

Again writing out the hypergeometric series to determine the nature of the singularity at $\theta = 0°$, we find

$$\sum_2 = -\frac{\Gamma(1-i\alpha)}{\Gamma(1+i\alpha)}-(\sin\tfrac{1}{2}\theta)^{2i\alpha}+(\sin\tfrac{1}{2}\theta)^{2i\alpha-1} \times\Big(2\cos^2\tfrac{1}{2}\theta\frac{\Gamma(\frac{1}{2}-i\alpha)}{\Gamma(\frac{1}{2}+i\alpha)} +\frac{(\cos\frac{1}{2}\theta)^{3-2i\alpha}}{(i\alpha-\frac{1}{2})}\frac{\Gamma(1-i\alpha)}{\Gamma(1+i\alpha)}\Big)+\cdots. \tag{10}$$

We evaluate $\sum_3$ by the same method used for $\sum_1$ and $\sum_2$, and we find

$$\sum_3 = \frac{\Gamma(-i\alpha)}{\Gamma(1+i\alpha)}[-1+(\sin\tfrac{1}{2}\theta)^{2i\alpha}]. \tag{11}$$

The scattering amplitude may now be calculated for small θ by means of Eq. (3) which is exact, since the term by term subtraction is compensated for by adding $f_1(\theta), f_2(\theta), f_3(\theta)$ back into $f(\theta)$.

III. KLEIN–GORDON CASE

Hetherington's series for the Klein–Gordon case may be made absolutely convergent by subtracting the series

$$S=\sum_{k=0}^{\infty}\frac{1}{2k+1}\frac{\Gamma(k+1+i\lambda)}{\Gamma(k+1-i\lambda)}P_k(\cos\theta),$$

which is summed as follows:

$$S=\int_0^1 du\int_0^1 v^{i\lambda}(1-v)^{-1-2i\lambda}\sum_{k=0}^{\infty}(vu^2)^kP_k(\cos\theta)\,dv =\int_0^1 v^{i\lambda}(1-v)^{-1-2i\lambda}I(v,\theta)\,dv, \tag{12}$$

where

$$I(v,\theta)=\int_0^1\frac{du}{(1-2vu^2\cos\theta+v^2u^4)^{\frac{1}{2}}}.$$

If we let $u = 1/x$ in $I(v, \theta)$ we get

$$I(v,\theta)=(1/v^{\frac{1}{2}})K(\cos\tfrac{1}{2}\theta),$$

where $K(\cos\frac{1}{2}\theta)$ is the complete elliptic integral of the first kind. Substituting back into Eq. (12) yields

$$S=\frac{\Gamma(\frac{1}{2}+i\lambda)\Gamma(2-i\lambda)}{\Gamma(\frac{1}{2}-i\lambda)}K(\cos\tfrac{1}{2}\theta).$$

ACKNOWLEDGMENT

One of the authors (C. R. F.) is indebted to Dr. L. Maximon for several valuable discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Mixed Irreducible Representations for U_3

JORGE FLORES*†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

(Received 14 September 1966)

In this paper we study some properties of irreducible representations of the unitary group in three dimensions U_3 with positive and negative indices. These representations are useful for the group theoretical classification of particle-hole states in nuclear shell theory, as well as in elementary particle physics. We show how the rules for reducing the direct product of two given representations should be modified when we include positive and negative indices and we use these rules to obtain an algebraic expression for the irreducible representations contained in the direct product.

1. INTRODUCTION

WE consider in this paper the construction of irreducible basis for the group U_3, which are expressed in terms of covariant and contravariant vectors. The irreducible representations IR carried by these polynomial basis are mixed, in the sense that they include positive as well as negative indices. The problem arises naturally when we consider the problem of constructing and classifying particle-hole wavefunctions in nuclear shell theory[1] or elementary particle states containing both particles and anti-particles.

In Sec. 2 we consider the group U_3 as a subgroup of a six-dimensional orthogonal group and we use some results recently derived by Chacón[2] for this group, to determine the highest-weight polynomial for mixed representations. We then show that not all Young diagrams with positive and negative components correspond to IR of U_3.

In Sec. 3 the reduction of the direct product of two given IR of U_3 is considered. We use the mixed representations to derive an algebraic expression for the irreducible components of the direct product.[3] To accomplish this, we express Littlewood rules[4] in a different form, which is convenient when we deal with mixed representations. Finally, in the Appendix, the equivalence of the modified with the usual Littlewood rules is shown.

2. MIXED IRREDUCIBLE REPRESENTATIONS FOR THE UNITARY GROUP U_3

We consider in this section the construction of a polynomial basis for irreducible representations of the group U_3 which contain the components of two types of three-dimensional vectors; one of them,

$$x_j^t \quad (j = 1, 2, 3), \tag{2.1}$$

transforms according to the three-dimensional unitary matrices which form the group U_3; the other vector,

$$\zeta_t^j \quad (j = 1, 2, 3), \tag{2.2}$$

transforms according to the contragradient representation to x_j^t of U_3. In both (2.1) and (2.2) j is the index affected by U_3 transformations and the index t is used to differentiate among different vectors and takes, for the most general representation of U_3, three values also, say 1, 2, and 3.

It proves convenient for our purposes to introduce the vectors (2.1) and (2.2) using a different notation, making the identification

$$x_j^t \to y_m^t, \quad \zeta_t^j \to y_{-m}^t \quad (m, t = 1, 2, 3), \tag{2.3}$$

whereby we define the vector y_m^t, remembering that a negative index m indicates the component of a contravariant vector.

In terms of this new vector y_m^t, we define the following operators:

$$\Lambda_m^{m'} = \sum_{t=1}^{3} (y_m^t p_t^{m'} - y_{-m'}^t p_t^{-m}) \equiv \mathfrak{C}_m^{m'} - \mathfrak{C}_{-m'}^{-m}, \tag{2.4}$$

where

$$p_t^{m'} = \partial/\partial y_m^t \quad \text{and} \quad p_t^{-m} = \partial/\partial y_{-m}^t. \tag{2.5}$$

By using well-known commutation relations (and taking into account that y_m^t and y_{-m}^t, with $m > 0$, are independent of each other) it is a simple matter to show that the operators defined in Eq. (2.4) are the elements of the Lie algebra corresponding to a rotational group in six dimensions. From the same commutation rules we can see that the set of operators $\Lambda_m^{m'}$, with m and $m' > 0$, define in turn a Lie algebra, this time for a unitary group in three dimensions; this latter is isomorphic to the group U_3 introduced above.

* This work was supported by the U.S. Atomic Energy Commission.

† On leave of absence from Instituto de Física, Universidad de México and Comisión Nacional de Energía Nuclear, Mexico.

[1] J. Flores and M. Moshinsky, Nucl. Phys. (to be published).

[2] E. Chacón, Ph.D. thesis, University of Mexico (1966).

[3] N. Mukunda and L. Pandit, Progr. Theoret. Phys. (Kyoto) **34**, 46 (1965).

[4] D. E. Littlewood, *The Theory of Group Characters* (Oxford University Press, New York, 1940).

In order to construct irreducible polynomials, we follow Moshinsky[5] and introduce the invariant operators,

$$C_s^{s'} = \sum_{m=1}^{3} y_m^s p_{s'}^{-m}, \quad \Gamma_{s'}^{s} = \sum_{m=1}^{3} y_{-m}^{s} p_{s'}^{-m},$$

$$D_{ss'} = \sum_{m=1}^{3} p_s^m p_{s'}^{-m}, \quad D^{+ss'} = \sum_{m=1}^{3} y_m^s y_{-m}^{s'}. \tag{2.6}$$

Using the same procedure as with the operators (2.4) one can show that the set given in Eq. (2.6) define a unitary group in six dimensions, which we call the complementary group to U_3. With these generators we can characterize, as shown below, the irreducible basis for U_3. The basis is given as a polynomial function, expressed in terms of y_m^t fulfilling the following invariant conditions,[5]

$$C_s^{s'} P(y_m^t) = 0, \quad \Gamma_s^{s'} P(y_m^t) = 0 \quad (s < s'), \tag{2.7a}$$

$$D_{ss'} P(y_m^t) = 0, \quad \forall\, s, s', \tag{2.7b}$$

$$C_s^s P(y_m^t) = h_s P(y_m^t), \quad \Gamma_s^s P(y_m^t) = k_s P(y_m^t) \quad (s = 1, 2, 3). \tag{2.7c}$$

In this set of equations $C_s^{s'}$, $\Gamma_s^{s'}$, with $s < s'$, and $D_{ss'}$, for all values of s and s', form the set of raising generators[5] of the complementary group to U_3. The meaning of both sets of numbers (h_s) and (k_s) becomes clear below.

If, besides conditions (2.7), we impose on $P(y_m^t)$ the following restrictions:

$$\Lambda_m^{m'} P(y_m^t) = 0 \quad (0 < m < m') \tag{2.8a}$$

and

$$\Lambda_m^m P(y_m^t) = \lambda_m P(y_m^t) \quad (m = 1, 2, 3), \tag{2.8b}$$

the polynomial $P(y_m^t)$ will be the highest-weight function for an IR of U_3, with indices $(\lambda_1\, \lambda_2\, \lambda_3)$. One should notice that $P(y_m^t)$ is a function of y_m^t and thus of both covariant and contravariant vectors. In case P depends only on y_m^t, $m > 0$, Eqs. (2.7) reduce to those previously used by Moshinsky.[6]

We now obtain a solution for Eqs. (2.7) and (2.8), showing that it is unique. If we consider Eqs. (2.7a) and (2.7c) only, the polynomial

$$P = (\Delta_1^1)^{h_1-h_2}(\Delta_{12}^{12})^{h_2-h_3}(\Delta_{123}^{123})^{h_3}(\Delta_{-3}^1)^{k_1-k_2}(\Delta_{-3-2}^{12})^{k_2-k_3} \times (\Delta_{-3-2-1}^{123})^{k_3} Z\left(\frac{\Delta_m^1}{\Delta_1^1}, \frac{\Delta_{1m}^{12}}{\Delta_{12}^{12}}, \frac{\Delta_{-m}^1}{\Delta_{-3}^1}, \frac{\Delta_{-3-m}^{12}}{\Delta_{-3-2}^{12}}\right) = \left(\frac{P}{Z}\right) \times Z \quad (m = 1, 2, 3) \tag{2.9}$$

is a solution for these equations.[6] In (2.9) we have used the notation

$$\Delta_m^s = y_m^s, \quad \Delta_{-m}^s = y_{-m}^s,$$

$$\Delta_{m_1 \cdots m_j}^{1 \cdots j} = \sum_{Pm} (-1)^{P_m} P_m \Delta_{m_1}^1 \Delta_{m_2}^2 \cdots \Delta_{m_j}^j, \tag{2.10}$$

where P_m stands for a permutation over the indices m, $(-1)^{P_m}$ being the parity of the permutation. The last relation in Eq. (2.9) defines (P/Z), i.e., the factor of the function Z in $P(y_m^t)$. This function Z depends on the variables explicitly indicated in (2.9) as its arguments, and is only restricted by the condition that when we multiply this function Z by the polynomial (P/Z), a polynomial $P(y_m^t)$ in y_m^t is obtained. It is possible, therefore, that Z is not a polynomial in y_m^t.

We now introduce conditions (2.8a). We apply the raising generators of U_3,

$$\Lambda_2^3, \Lambda_1^3, \text{ and } \Lambda_1^2, \tag{2.11}$$

to the polynomial defined in Eq. (2.9). When acting on (P/Z), the generators (2.11) give a vanishing result. We are therefore left only with the application of operators (2.11) to the function Z. In order to do this, we follow Chacón[2] and introduce new variables, defined as follows,

$$I_1 = \sum_m \Delta_{-m}^1 \Delta_{1m}^{12},$$

$$I_2 = \sum_m \Delta_m^1 \Delta_{-3-m}^{12},$$

$$I_3 = \sum_m \Delta_m^1 \Delta_{-m}^1, \tag{2.12}$$

which have the property of commuting with operators (2.11). By using these defining relations, we can replace some of the quotients appearing as arguments of Z in Eq. (2.9), obtaining a new function Z', which is now a function of the ratios

$$\frac{\Delta_3^1}{\Delta_1^1}, \frac{\Delta_{-1}^1}{\Delta_{-3}^1}, \frac{\Delta_{-2}^1}{\Delta_{-3}^1}, \frac{I_1}{\Delta_{-3}^1 \Delta_{12}^{12}}, \frac{I_1}{\Delta_1^1 \Delta_{-3-2}^{12}}, \frac{I_3}{\Delta_1^1 \Delta_{-3}^1}. \tag{2.13}$$

Equation

$$\Lambda_1^3 P(y_m^t) = 0$$

then implies that Z' is not a function of $\Delta_{-1}^1/\Delta_{-3}^1$. In a similar way, $\Lambda_2^3 P = 0$ and $\Lambda_1^2 P = 0$ tell us that Z' is not a function of $\Delta_{-2}^1/\Delta_{-3}^1$ and Δ_3^1/Δ_1^1, respectively. As a consequence of this, Z' is a function of the last three ratios indicated in (2.13).

We can now expand Z' in a power series, obtaining the following expression for $P(y_m^t)$:

$$P(y_m^t) = \sum_{n_1 n_2 n_3} A_{n_1 n_2 n_3} (\Delta_1^1)^{h_1-h_2-n_1-n_3} (\Delta_{12}^{12})^{h_2-h_3-n_1} \times (\Delta_{123}^{123})^{h_3} (\Delta_{-3}^1)^{k_1-k_2-n_1-n_3} (\Delta_{-3-2}^{12})^{k_2-k_3-n_2} \times (\Delta_{-3-2-1}^{123})^{k_3} I_1^{n_1} I_2^{n_2} I_3^{n_3}, \tag{2.14}$$

[5] M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).
[6] M. Moshinsky, Nucl. Phys. **31**, 384 (1962).

where the dummy indices n_i are restricted by the condition that $P(y^t_m)$ is a polynomial function in y^t_m. This leads, in particular, to the restriction that n_1 are nonnegative integers.

Using the form (2.14) for $P(y^t_m)$, conditions (2.8b) now yield the relations

$$\begin{aligned} \lambda_1 &= h_1 - k_3 - n_2 - n_3, \\ \lambda_2 &= h_2 - k_2 - n_1 + n_2, \\ \lambda_3 &= h_3 - k_1 + n_1 + n_3, \end{aligned} \tag{2.15}$$

and we are left with Eq. (2.7b) only, to determine the coefficient $A_{n_1 n_2 n_3}$.

As can be seen from its definition, $D_{ss'}$ is a second-order operator, and it is difficult to apply. It is at this point that our analysis is restricted to the three-dimensional group U_3, since we are forced to apply $D_{ss'}$ directly to $P(y^t_m)$. This is not simply done for the general case of U_r. In any case, Eq. (2.7b) determines uniquely the values of n_i, giving after a lengthy calculation the values

$$n_i = 0. \tag{2.16}$$

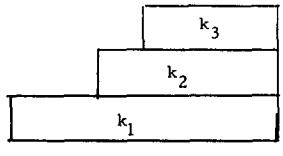
This shows that the highest weight polynomial $P(y^t_m)$ is determined uniquely and given by

$$\begin{aligned} P(y^t_m) = (\Delta^1_1)^{h_1-h_2}(\Delta^{12}_{12})^{h_2-h_3}(\Delta^{123}_{123})^{h_3}(\Delta^1_{-3})^{k_1-k_2} \\ \times (\Delta^{12}_{-3-2})^{k_2-k_3}(\Delta^{123}_{-3-2-1})^{k_3}. \end{aligned} \tag{2.17}$$

We can now see the meaning of both sets of numbers (h_i) and (k_i). Let us assume, for the moment, that $P(y^t_m)$ is a function of y^t_m, with $m > 0$, only. In this case $k_i = 0$, $i = 1, 2, 3$ and $\lambda_i = h_i$. If we now assume $P(y^t_m)$ to be a function of the contravariant vectors y^t_m, $m < 0$, only, we have $h_i = 0$ and

$$\lambda_1 = -k_3, \quad \lambda_2 = -k_2, \quad \lambda_3 = -k_1. \tag{2.18}$$

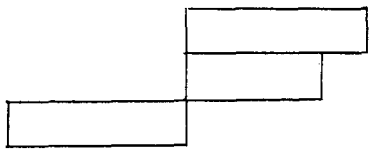
In this latter case, we have IR for the group U_3, with indices $(\lambda_1\, \lambda_2\, \lambda_3)$ which are *negative* numbers. We could represent (2.18) by a Young diagram of the form



, (2.19)

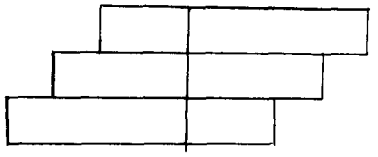
and we could call the blocks forming this diagram, negative blocks or antiblocks.

In the general case P is a function of y^t_m with both negative and positive m. For example, we could have an IR with h_1 and h_2 different from zero and $h_3 = 0$, and $k_i = 0$ except k_1. In this case we have a diagram, representing $(\lambda_1\, \lambda_2\, \lambda_3)$, of the form



. (2.20)

We could ask ourselves, what is the meaning of a Young diagram with negative as well as positive blocks in the same row? In other words, what is the meaning of a Young diagram of the following type,



. (2.21)

We call these diagrams *virtual diagrams*, and show presently that they do not correspond to IR of U_3. In order to do this, consider the second-order Casimir operator of the group U_3, defined as

$$\begin{aligned} \Phi &= \sum_{m\,m'} \Lambda^{m'}_m \Lambda^m_{m'} \\ &= \sum_m (\Lambda^m_m)^2 + 2\Lambda^{m'}_m \Lambda^m_{m'} + \sum_{m m'} (\Lambda^m_m - \Lambda^{m'}_{m'}), \end{aligned} \tag{2.22}$$

where, in the last equation we have expanded the double summation and have used the commutation rules satisfied by $\Lambda^{m'}_m$.[7] If we use the definitions (2.4) and (2.6), we can readily see that Φ can also be expressed as

$$\Phi = \sum_{s,s'} C^{s'}_s C^s_{s'} + \sum_{s,s'} \Gamma^{s'}_s \Gamma^s_{s'} + 2 \sum_{s<s'} D^{+ss'} D_{ss'}. \tag{2.23}$$

By expanding the right-hand side of this equation as was done in Eq. (2.22), and equating the resulting expression to (2.22), we obtain the operator identity,

$$\begin{aligned} \sum_m (\Lambda^m_m)^2 + 2 \sum_{m>m'} \Lambda^{m'}_m \Lambda^m_{m'} + \sum_{m<m'} (\Lambda^m_m - \Lambda^{m'}_{m'}) \\ = \sum_s (C^s_s)^2 + 2 \sum_{s>s'} C^{s'}_s C^s_{s'} + \sum_{s<s'} (C^s_s - C^{s'}_{s'}) \\ + \sum_s (\Gamma^s_s)^2 + 2 \sum_{s>s'} \Gamma^{s'}_s \Gamma^s_{s'} + \sum_{s<s'} (\Gamma^s_s - \Gamma^{s'}_{s'}) \\ + 2 \sum_{s<s'} D^{+ss'} D_{ss'}. \end{aligned} \tag{2.24}$$

Acting with this operator identity on the polynomial given in Eq. (2.17), we get

$$h_1 k_3 + h_2 k_2 + h_3 k_1 = 0. \tag{2.25}$$

Since h_i and k_i are nonnegative integers, we conclude from Eq. (2.25) that

$$h_1 k_3 = h_2 k_2 = h_3 k_1 = 0, \tag{2.26}$$

proving with this that diagrams as (2.21) do not correspond to an irreducible representation of U_3.

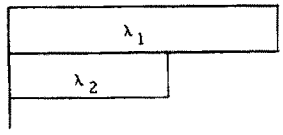
We can understand the content of Eq. (2.26) in a fairly simple way, if instead of dealing with the group U_3, we consider its unimodular subgroup SU_3. As is well known,[5] there is a one-to-one correspondence between the IR of U_3 and those of SU_3. Furthermore, it is also well known that complete columns (i.e., columns formed by three blocks) can be eliminated

[7] M. Moshinsky, in *Physics of Many Particle Systems*, E. Meron, Ed. (Gordon and Breach Science Publishers, Inc., New York, 1965).

when considering Young diagrams for SU_3. This has to do with the fact that the determinants

$$\Delta^{123}_{123} \quad \text{and} \quad \Delta^{123}_{-3-2-1} \tag{2.27}$$

are invariant functions of y^t_m, with respect to SU_3 transformations. The reference line can then be displaced at will, and the most general IR of SU_3 is characterized by two indices $(\lambda_1 \lambda_2)$, the Young diagram is of the form



. (2.28)

From a diagram such as (2.28) we obtain another one, looking like the diagram (2.20), by making a displacement of the reference line to the right. And vice versa, moving the vertical reference line in (2.20) to the left k_1 positions, we get a two-row diagram of the general form (2.28).

We can obtain the highest-weight polynomial corresponding to a displaced diagram by multiplication of the original highest-weight polynomial with an invariant function of y^t_m with respect to SU_3 transformations [i.e., determinants (2.27)]. The multiplication process, therefore, does not change the SU_3 IR carried by the polynomials.

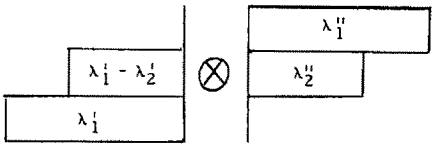
We can now understand condition (2.26): If we displace the reference line in a diagram whose indices violate conditions (2.26), a non-allowed Young diagram for the unimodular group is obtained.

Using the fact that we can displace the vertical reference line, we discuss in the next section the reduction of two given IR of SU_3.

3. REDUCTION OF THE DIRECT PRODUCT OF TWO IR OF SU_3

We now use the negative indices representations introduced in the previous section to find an algebraic expression for the irreducible components of the direct product of two given IR of SU_3. The same expression has been found by Mukunda and Pandit[3] using an entirely different approach. Our procedure is based directly upon Littlewood's rules,[4] conveniently modified to introduce mixed representations. The procedure is generalizable to other unitary groups, but the algebraic expressions obtained are much more complicated, so as to cease to be useful from a practical point of view. We therefore restrict ourselves to the SU_3 group.

Suppose we want to find out which IR of SU_3 are contained in the direct product $(\lambda'_1 \lambda'_2) \times (\lambda''_1 \lambda''_2)$. Using the fact mentioned at the end of Sec. 1, we can make a translation of the vertical reference line by λ'_1 blocks to the right, in the Young diagram corresponding to $(\lambda'_1 \lambda'_2)$. We can then represent the direct product, graphically, as

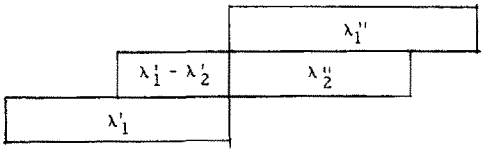


. (3.1)

At the end of the process we turn back the reference line to its original position, moving it the same λ'_1 positions to the left.

We now state the rules to reduce the direct product indicated in (3.1). We show in the Appendix the equivalence of this set of rules to the well-known Littlewood rules. Although the argument is given there for the group SU_3 only, we have shown the same result to be true in general for SU_r, with arbitrary r.

In order to reduce the direct product (3.1), first of all superpose both diagrams appearing in (3.1) to obtain the (virtual) diagram



. (3.2)

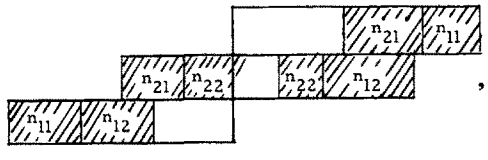
As has been shown before, this is not an allowed Young diagram (except for the particular cases with $\lambda'_1 = \lambda'_2$ or $\lambda''_2 = 0$). To obtain the indices of the IR contained in the direct product, we make all possible contractions (i.e., annihilation of one positive with one negative block in the diagram) in diagram (3.2), one at a time, in such a way that the following conditions are satisfied:

(1) An allowed Young diagram is obtained [which means, as a matter of fact, that a large enough number of contractions has been made in (3.2), so as to obtain a nonvirtual diagram from it].

(2) No two blocks in the same negative row (or column) are to be contracted with two positive blocks in the same positive column (or row).

(3) The order of the contractions of negative (positive) blocks in the same row with some positive (negative) blocks in any row or column is immaterial.

Applying these set of rules to the case indicated in (3.2), we get a diagram of the type



, (3.3)

where we have indicated by n_{ij} the number of blocks in negative row i (counting from bottom to top) contracted with a corresponding number of blocks in the positive row j (counting from top to bottom, in the positive component of the diagram). The resulting diagram has been left blank in (3.3). Proceeding in this form we have taken rule (3) into account, since we have made no distinction among contractions differing only in the order in which they were made.

In diagram (2.3) we have made the further assumption

$$\lambda_1' - \lambda_2' \leq \lambda_2'', \tag{3.4}$$

which implies lack of generality. However, if the opposite inequality to (3.4) holds, the analysis is similar, and is not given here. Using (3.4) we have completely annihilated the negative blocks in the second row, in such a way to get a nonvirtual diagram, as required by rule (1). In other words, we have assumed that

$$n_{21} + n_{22} = \lambda_1' - \lambda_2'. \tag{3.5}$$

If rule (2) is to be satisfied, the following inequality should hold:

$$0 \leq n_{11} \leq \lambda_2', \tag{3.6}$$

in order to avoid contractions of two blocks in the same negative column with two blocks in the first positive row. By applying the same rule, we get another restriction on the nonnegative integer n_{11}, i.e.,

$$n_{11} \leq \lambda_1'' - \lambda_2''. \tag{3.7}$$

Using (3.7) together with (3.6) we conclude that n_{11} is restricted by

$$n_{11} \leq \min(\lambda_2', \lambda_1'' - \lambda_2''). \tag{3.8}$$

Considering now the second negative row, the inequality

$$\lambda_2'' - n_{12} \leq \lambda_1'' - (n_{11} + n_{21}) \tag{3.9}$$

holds, since otherwise we would have contracted two blocks in this negative row with two positive blocks in the same column.

At this point we have two alternatives,

$$n_{21} \geq n_{12}; \quad n_{21} = n_{12} + r \quad (r \geq 0), \tag{3.10a}$$

$$n_{21} < n_{12}; \quad n_{21} = n_{12} - s \quad (s < 0). \tag{3.10b}$$

We consider case (a) first. By construction, we have

$$n_{21} = n_{12} + r \leq \lambda_1' - \lambda_2' \leq \lambda_2'', \tag{3.11}$$

which is equivalent to

$$0 \leq r \leq (\lambda_1' - \lambda_2') - n_{12}. \tag{3.12}$$

On the other hand, using inequality (3.10a) in the relation (3.9), we get

$$n_{11} + r \leq \lambda_1'' - \lambda_2''.$$

This latter, together with (3.12), implies the following restriction on the integer r,

$$0 \leq r \leq \min(\lambda_1' - \lambda_2' - n_{12}, \lambda_1'' - \lambda_2'' - n_{11}). \tag{3.13}$$

We see finally, from the diagram (3.3), that the resulting mixed U_3 tableaux have components $(\bar{\lambda}_1 \bar{\lambda}_2 \bar{\lambda}_3)$ given by

$$\begin{aligned} \bar{\lambda}_1 &= \lambda_1'' - n_{21} - n_{11}, \\ \bar{\lambda}_2 &= \lambda_2'' - n_{22} - n_{12}, \\ \bar{\lambda}_3 &= -(\lambda_1' - n_{11} - n_{12}), \end{aligned}$$

where $\bar{\lambda}_1$ and $\bar{\lambda}_2$ are positive numbers and $\bar{\lambda}_3$ is a negative integer. Returning the reference line to its original position, i.e., moving it λ_1' positions to the left, we obtain the indices $(\lambda_1 \lambda_2)$ for the irreducible components of the direct product. These are given by

$$\begin{aligned} \lambda_1 &= \lambda_1' + \lambda_1'' - 2n_{21} - 2n_{11} - r, \\ \lambda_2 &= \lambda_2' + \lambda_2'' - n_{21} - n_{11} + r, \end{aligned} \tag{3.14}$$

with n_{11}, n_{21} and r restricted by the conditions given in (3.8), (3.11), and (3.13), respectively. Once these conditions are fulfilled, rules (1), (2), and (3) are satisfied, and Eq. (3.14) gives the possible irreducible indices in the decomposition of the direct product.

If we now consider case (3.10b) and follow exactly the same steps, we arrive in this case, at the following expressions for $(\lambda_1 \lambda_2)$

$$\begin{aligned} \lambda_1 &= \lambda_1' + \lambda_1'' - 2n_{21} - 2n_{11} - s, \\ \lambda_2 &= \lambda_2' + \lambda_2'' - n_{21} - n_{11} - 2s, \end{aligned} \tag{3.15}$$

where n_{11} and n_{21} are restricted as in the previous case, but now the possible values for the positive integer s are such that the inequality

$$0 \leq s \leq \min(\lambda_2' - n_{11}, \lambda_2'' - n_{21}) \tag{3.16}$$

holds; this is the analogous relation to (3.13) for case (b).

As a final remark, we notice that (3.14) and (3.15) represent different IR of SU_3 contained in the reduction of $(\lambda_1' \lambda_2') \times (\lambda_1'' \lambda_2'')$, as is clear from the way they have been derived, cases (a) and (b) correspond to a different set of contractions.

ACKNOWLEDGMENTS

The author is deeply indebted to Professor M. Moshinsky for many interesting suggestions and discussions. The author is also indebted to E. Chacón, for supplying some of his results prior to publication.

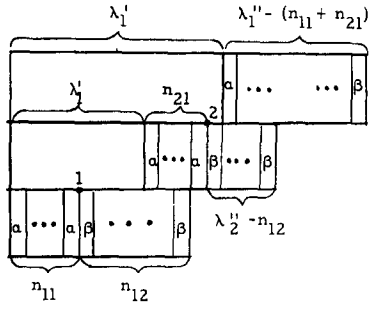
APPENDIX

We show in this Appendix the equivalence between the rules enunciated in Sec. 3 and the usual Littlewood rules[4] for reducing the direct product of two IR of SU_3. By exactly the same argument we have proved this equivalence in the general case for the group SU_r.

We restrict the proof given here to those cases which satisfy the inequality (3.4), although the proof can be given in the other case in a similar way.

In order to show the equivalence between both sets of rules, we notice that when the vertical reference line is returned to its original position, any place in the diagram where a negative block is present, does not contain a block in the diagram obtained at the end of the process and vice versa. Having this in mind, we label the positive blocks in the first row of (3.3) with the letter α and, at the same time, we use this letter to label the negative blocks contracted with these blocks [i.e., those blocks indicated in (3.3) by n_{11} and n_{21}]. In a similar way, we use the label β for the positive and negative blocks, indicated in diagram (3.3) by n_{12} and n_{22}.

Using this labeling, we obtain the following diagram at the end of the process:



(A1)

One can see from (A1) that the number of blocks labeled by α is equal to λ_1'' and that the number of blocks labeled by β is equal to λ_2'', as required by Littlewood rules. Furthermore, the alphabetical order from left to right and from top to bottom is obtained automatically from our procedure, agreeing with Littlewood.[4]

We now show that our second rule (Sec. 3) is equivalent to the two following Littlewood rules:

If a diagram such as (A1) corresponds to an irreducible component of the direct product of two IR of SU_3, then

(a) no two blocks labeled with the same letter are to be placed in the same column;

(b) the number of blocks labeled by α should be greater or equal to the number of blocks labeled by β at any position in the diagrams when counting them from right to left and from top to bottom.

As can be seen from diagram (A1) there exists the possibility of placing blocks with the same label in a given column [and, therefore, of violating rule (a)] at two positions only [marked as 1 and 2 in (A1)]. We consider point 1 first, and we see that the condition imposed by rule (a) on n_{ij} is

$$0 \leq n_{11} \leq \lambda_2', \tag{A2}$$

which is identical to inequality (3.6). Regarding point 2, the condition is

$$0 \leq n_{11} + n_{12} \leq \lambda_2' + n_{21}, \tag{A3}$$

which is equivalent to

$$\lambda_1' - \lambda_2' - n_{21} \leq \lambda_1' - (n_{11} + n_{21}).$$

Inequalities (A2) and (A3) can be obtained from diagram (3.3) by requiring that no two blocks in the same negative column are contracted with blocks in the first and second positive row, respectively.

We consider rule (b) now; we could violate this rule only at points 1 and 2 in diagram (A1), as before. We restrict our attention to these two points only, obtaining the relation

$$(\lambda_2'' - n_{12}) + n_{12} \leq \lambda_1'' - (n_{11} + n_{21}) + n_{21}, \tag{A4}$$

which is identical to (3.7). Analogously, at point 2, rule (b) imposes the following restriction over n_{ij},

$$\lambda_2'' - n_{12} \leq \lambda_1'' - (n_{11} + n_{21}), \tag{A5}$$

which is identical to (3.9).

We then see that rule (b) is equivalent to the condition that no two blocks in the same positive column should be contracted with two blocks in the same negative row. This proves the equivalence of Littlewood's rules (a) and (b) with our rule (2).

The set of rules we have used here, have the two following nice features:

(1) From a practical point of view, they can be used instead of the usual prescription, when one of the diagrams in the direct product has a smaller number of negative blocks (when the reference line is displaced to the right in order to obtain a diagram with negative indices only) than the number of positive blocks it had with the reference line in its original position.

(2) From a theoretical point of view, they show that Littlewood rules (a) and (b) have a very simple meaning. This can be seen, if we consider the highest weight tensor corresponding to each one of the irreducible components of the direct product, each of these tensors being formed by homogeneous linear combinations of the vector components x_j^t and ζ_t^j, introduced in Sec. 2, and classified by their symmetry with respect to the exchange of the indices j and t.[1] If rule 2 is violated, the corresponding highest weight tensor is identically zero, having to do with the fact that the contraction of two symmetrical with two antisymmetrical indices leads to a vanishing function, and showing that the corresponding diagrams are not contained in the reduction of the direct product.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Density Fluctuations of a Fluid

GERALD L. JONES AND ROBERT E. KENNEDY
Department of Physics and the Radiation Laboratory, University of Notre Dame, Notre Dame, Indiana

(Received 29 April 1966)

A proof is given, in the canonical and the grand canonical formalism, to show that the density fluctuations in a macroscopic region of a fluid are large in the two-phase states and small in the one-phase states. The definition of large and small density fluctuations is one used previously by Dobrushin in connection with the Ising model. The density fluctuations at the critical point are, using this definition, small if the critical isotherm has no flat portion.

I. INTRODUCTION

THE authors wish to examine the probability of a density fluctuation in a macroscopic region of a fluid. If the fluid is in a one-phase state, the probability of such a fluctuation should be small since we expect every macroscopic subvolume of the total volume to have essentially the same density as the over-all density of the fluid. If the fluid is in a two-phase state, then a macroscopic subvolume may contain varying amounts of the two phases, and one would expect the probability of density fluctuations to be large. The problem is to rigorously and conveniently characterize "large-" and "small-" density fluctuations, and to prove within this characterization the above-mentioned properties. In a recent paper Dobrushin[1] has solved this problem for the lattice gas (Ising model) in any number of dimensions, using the canonical ensemble. He makes use of these results to obtain some estimates of the boundaries of the phase transition regions. In this paper we obtain results identical to Dobrushin's for the characterization of the fluctuations, except that we take the case of a fluid and either the canonical or grand canonical ensembles. For the grand canonical ensemble, the classical and the quantum fluid may be treated in the same way. We treat only the classical fluid in the canonical ensemble. We have not yet been able to use these results to estimate the boundaries of the phase transition region.

In the second section we review some known results concerning the thermodynamic limit which we need later. In the third section we consider the problem using the grand canonical formalism. We take this case first because the pertinent theorem is easier to state and to prove than in the case of the canonical ensemble. In the fourth section we consider the canonical ensemble. Throughout we consider only a three-dimensional gas but none of the arguments depend on the dimensionality of the space.

II. THERMODYNAMIC LIMIT

We are interested only in systems for which a proper thermodynamic limit exists. One takes this limit by considering a sequence of systems, contained in successively larger volumes, but with a fixed density of particles. In order that a proper thermodynamic limit exists and is independent of the shape of the containing region, one must put some conditions on the Hamiltonian of the system and on the sequence of domains which one uses in taking the limit. This problem is most recently and comprehensively treated by Ruelle[2] and by Fisher.[3] We review those results which are necessary for this problem. We use Fisher's notation as much as possible. There is apparently no known set of necessary and sufficient conditions which leads to the properties we want; however, Fisher gives two sets of sufficient conditions of which we pick one to simplify the presentation. The other could also be used. The results listed here are valid for both quantum and classical systems.

The Hamiltonian is of the form

$$H_N = \sum_{i=1}^{N} \frac{p_i^2}{2m} + U_N(\mathbf{r}_1 \cdots \mathbf{r}_N), \tag{1}$$

where $\mathbf{p}_i$ and $\mathbf{r}_i$ are real variables in the classical case and the usual operators in the quantum case. The classical partition function is given by

$$Z(\beta, N, \Omega) = \frac{\Lambda^{-3N}}{N!} \int_\Omega \cdots \int_\Omega \exp(-\beta U_N)\, d\mathbf{r}_1 \cdots d\mathbf{r}_N, \tag{2}$$

where $\Lambda = (\beta h^2/2\pi m)^{\frac{1}{2}}$, $\beta = 1/kT$, and Ω is the region in which the fluid is contained. The quantum partition function is given by

$$Z(\beta, N, \Omega) = \mathrm{Tr}_\Omega \exp(-\beta H_N). \tag{3}$$

[1] R. L. Dobrushin, Dokl. Acad. Nauk SSSR **160**, 1046 (1965) [English transl.: Soviet Phys.—Doklady **10**, 111 (1965)].

[2] D. Ruelle, Helv. Phys. Acta **36**, 183 (1963).

[3] M. E. Fisher, Arch. Ratl. Mech. Anal. **17**, 377 (1964).

The trace is taken over a complete set of N-particle wavefunctions (including internal degrees of freedom which have not been indicated explicitly) which satisfy appropriate boundary conditions and symmetry requirements. We do not notationally distinguish between the classical partition function and the quantum-mechanical partition function for Fermi, Bose, or Boltzmann statistics, because the results quoted in this section are independent of this distinction. In the particular case of Boltzmann statistics, the right-hand side of (3) should be multiplied by $1/N!$. If the region Ω has volume $V(\Omega)$, then the free energy per unit volume is defined by

$$g(\beta, \rho, \Omega) = V(\Omega)^{-1} \ln Z(\beta, N, \Omega)$$

for particle densities $\rho = N/V(\Omega)$, which are integer multiples of $V(\Omega)^{-1}$. For fixed Ω, the definition is extended to all ρ by linear interpolation. If the potential U_N has a hard-core component so that two particles can never get too close together, Z vanishes for densities larger than some finite density ρ_m, which is the maximum allowed density. If there is no hard core, ρ_m may be infinite. The thermodynamic limit is taken by choosing a sequence of domains Ω_j with $V(\Omega_j) \to \infty$ and considering the sequence of functions $g(\beta, \rho, \Omega_j)$ with β and ρ fixed. We make the following assumptions:

(a) The potential $U_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ is stable, that is, there exists a fixed positive W such that

$$U_N(\mathbf{r}_1 \cdots \mathbf{r}_N) \geq -NW \tag{4}$$

for all N and all $\mathbf{r}_1 \cdots \mathbf{r}_N$.

(b) The potential is strongly tempered. This means that the interaction between two groups of particles becomes nonpositive if the groups are separated by a large enough distance. There exists a distance R_0 such that for all N_1, N_2

$$\begin{aligned}\Phi_{N_1,N_2}(\mathbf{r}_1 \cdots \mathbf{r}_{N_1}, \mathbf{r}_1' \cdots \mathbf{r}_{N_2}') \\ = U_{N_1+N_2}(\mathbf{r}_1 \cdots \mathbf{r}_{N_1}, \mathbf{r}_1' \cdots \mathbf{r}_{N_2}') \\ - U_{N_1}(\mathbf{r}_1 \cdots \mathbf{r}_{N_1}) - U_{N_2}(\mathbf{r}_1' \cdots \mathbf{r}_{N_2}') \leq 0,\end{aligned}$$

whenever $|\mathbf{r}_i - \mathbf{r}_j'| \geq R_0$ for all $1 \leq i \leq N_1$, $1 \leq j \leq N_2$. Both Fisher and Ruelle discuss examples of potentials with properties (a) and (b). In particular, a two-body hard-core interaction with a finite range tail satisfies (a) and (b).

(c) The sequence of domains $\{\Omega_j\}$ allowed in taking the thermodynamic limit has the following properties. For all j, Ω_j is a bounded, simply connected domain with volume $V(\Omega_j)$ (which we sometimes shorten to V_j). Let $V(h, \Omega)$ be the volume of the set of points within a distance h of the boundary of Ω and interior to Ω for $h > 0$, exterior to Ω if $h < 0$. We require that $V(h, \Omega_j)V^{-1}(\Omega_j) \to 0$ as $j \to \infty$ for any fixed h. This is Fisher's condition of asymptotic regularity which restricts the rate at which the surface of Ω_j can grow compared to the volume. Finally, we assume that if π_j is the smallest parallelepiped containing Ω_j, then there is a $\delta > 0$ such that $V(\Omega_j)V^{-1}(\pi_j) \geq \delta$ for all j.

Under assumptions (a), (b), and (c), it is proven[3] that

$$\lim_{j \to \infty} g(\beta, \rho, \Omega_j) = g(\beta, \rho), \tag{5}$$

where $g(\beta, \rho)$ is a continuous, convex upward function of ρ in the interval $0 \leq \rho \leq \rho_m$, and the convergence is uniform in any closed subinterval $0 \leq \rho \leq \rho_1 < \rho_m$. If we define

$$Q(\beta, z, \Omega) = \sum_{N=0}^{\infty} (\Lambda^3 z)^N Z(\beta, N, \Omega), \tag{6}$$

$$\pi(\beta, z, \Omega) = V(\Omega)^{-1} \ln Q(\beta, z, \Omega), \tag{7}$$

then the series (6) is absolutely convergent for any z and

$$\lim_{j \to \infty} \pi(\beta, z, \Omega_j) = \pi(\beta, z) \tag{8}$$

for $0 < z < \infty$, where $\pi(\beta, z)$ is a continuous convex function of $\ln z$. Furthermore,

$$\pi(\beta, z) = \max [\rho \ln (\Lambda^3 z) + g(\beta, \rho)], \quad 0 \leq \rho \leq \rho_m, \tag{9}$$

provided the maximum does not occur at $\rho = \rho_m$. Fisher points out that this cannot happen if the pressure diverges to infinity as $\rho \to \rho_m$. Equation (9) implies the equivalence of the canonical and grand canonical formalism. For each positive value of z, there is at least one value of ρ determined by (9). These are all the results we need. It should be pointed out that one can somewhat relax the strong tempering condition[3] (b) if one is willing to put more restrictions on the sequence $\{\Omega_k\}$.

III. DENSITY FLUCTUATIONS IN THE GRAND CANONICAL FORMALISM

In this formalism the probability of finding a system with N particles in it is

$$P_\Omega(\beta, z, N) = (\Lambda z)^N Z(\beta, N, \Omega)/Q(\beta, z, \Omega). \tag{10}$$

Corresponding to a given value of z and β, there is at least one density $\rho(\beta, z)$ given by (9). Let $P_\Omega^\epsilon(\beta, z)$ be the probability that the system has a density differing from $\rho(\beta, z)$ by more than ϵ. Then

$$P_\Omega^\epsilon(\beta, z) = \sum_N P_\Omega(\beta, z, N), \quad |N/V(\Omega) - \rho| > \epsilon. \tag{11}$$

If there is more than one value of ρ corresponding to (β, z), we can consider the above expression for any

of those ρ. For such (β, z) we should distinguish the probabilities corresponding to different ρ, but the results do not depend on such a distinction so we do not explicitly indicate it. Let us look at the sequence $V(\Omega_j)^{-1} \ln P^{\epsilon}_{\Omega_j}(\beta, z) = V_j^{-1} \ln P^{\epsilon}_j(\beta, z)$. This is a sequence of negative numbers; therefore, if it has a limit, the limit must be negative or zero. This lets us characterize the fluctuation as being large if the limit is zero and small if the limit is negative. Before we state Dobrushin's theorem for this case, we need to characterize a one-phase and two-phase state. For our purposes we say that (β, z) corresponds to a one-phase state if there is a unique $\rho(\beta, z)$ determined by the maximization procedure in (9). If there is more than one value of ρ corresponding to (β, z) the state is called two phase. One can show, as we do in Appendix A for completeness, that this definition of a two-phase state is equivalent to the condition that $g(\beta, \rho)$ be a linear function of ρ in some interval, which in turn implies the canonical pressure is constant. The statement which gives the relationship between the density fluctuations and the number of phases follows.

If (β, z) is a one-phase state, then for every $\epsilon > 0$,

$$\lim_{j\to\infty} V_j^{-1} \ln P^{\epsilon}_j(\beta, z)$$

exists and is less than zero. If (β, z) is a two-phase state then there exists $\epsilon > 0$ such that

$$\lim_{j\to\infty} V_j^{-1} \ln P^{\epsilon}_j(\beta, z)$$

exists and is zero.

Proof: Let us consider the two-phase state first. Since $V_j^{-1} \ln P^{\epsilon}_j(\beta, z)$ is always less than zero, we must show that, for large enough j, it is always larger than any pre-assigned negative number. If (β, z) is a two-phase state, there are at least two densities ρ_1 and ρ_2 such that

$$\begin{aligned} 0 \le \rho_1 < \rho_m, \quad 0 \le \rho_2 < \rho_m, \quad \rho_1 \ne \rho_2, \\ \pi(\beta, z) = \rho_1 \ln \Lambda^3 z + g(\beta, \rho_1) \\ = \rho_2 \ln \Lambda^3 z + g(\beta, \rho_2). \end{aligned} \tag{12}$$

Let $\epsilon > 0$, $\epsilon < \rho_1$, $\epsilon < \frac{1}{2}|\rho_1 - \rho_2|$ and let ρ_3 be such that $\max[\rho_1, \rho_2] < \rho_3 < \rho_m$. Since $\rho \ln \Lambda^3 z + g(\beta, \rho)$ is continuous in the closed interval $[0, \rho_3]$, it is uniformly continuous there. Given any $\delta > 0$, we can find $\gamma > 0$ and $\gamma < \epsilon$ such that, whenever $|\rho - \rho'| < \gamma$ and both ρ, ρ' are in the interval $[0, \rho_3]$, we have

$$|\rho \ln \Lambda^3 z + g(\beta, \rho) - \rho' \ln \Lambda^3 z - g(\beta, \rho')| < \tfrac{1}{3}\delta. \tag{13}$$

From the convergence properties quoted in Sec. II, there is an integer j_0 such that for all $j > j_0$ we have

$$|g(\beta, \rho, \Omega_j) - g(\beta, \rho)| < \tfrac{1}{3}\delta \tag{14}$$

for all $0 \le \rho \le \rho_3$ and

$$|\pi(\beta, z, \Omega_j) - \pi(\beta, z)| < \tfrac{1}{3}\delta. \tag{15}$$

Now $P^{\epsilon}_j(\beta, z)$ is greater than any single member of the sum (11). From (10) and (11) we have for a fluctuation from ρ_1

$$P^{\epsilon}_j(\beta, z) \ge (\Lambda^3 z)^N Z(\beta, N, \Omega_j)/Q(\beta, z, \Omega_j) \tag{16}$$

for any N such that

$$|NV_j^{-1} - \rho_1| > \epsilon. \tag{17}$$

For each j choose N_j such that (17) is satisfied and so that

$$|N_j V_j^{-1} - \rho_2| < \gamma. \tag{18}$$

Set $\rho_j = N_j V_j^{-1}$. Then from (13)–(16) we have, for $j > j_0$,

$$\begin{aligned} P^{\epsilon}_j(\beta, z) &\ge \frac{(\Lambda^3 z)^{N_j} \exp\{V_j[g(\beta, \rho_j) - \frac{1}{3}\delta]\}}{\exp\{V_j[\pi(\beta, z) + \frac{1}{3}\delta]\}} \\ &= \exp\{V_j[\rho_j \ln \Lambda^3 z + g(\beta, \rho_j) - \pi(\beta, z) - \tfrac{2}{3}\delta]\} \\ &\ge \exp\{V_j[\rho_2 \ln \Lambda^3 z + g(\beta, \rho_2) - \pi(\beta, z) - \delta]\}. \end{aligned}$$

By (12) the first three terms in the exponent cancel, so for all $j > j_0$, $V_j^{-1} \ln P^{\epsilon}_j(\beta, z) \ge -\delta$, which completes the proof for the two-phase state.

Now suppose (β, z) is a one-phase state of density ρ_1. From the lower bound (4) on the potential energy, one gets the upper bound

$$Z(\beta, N, \Omega) \le \Lambda^{-3N} V(\Omega)^N \exp(\beta N W)/N!,$$

and since $N! > N^N e^{-N}$, we have

$$Z(\beta, N, \Omega) \le [\Lambda^{-3} V(\Omega) N^{-1} \exp(\beta W + 1)]^N. \tag{19}$$

This bound is for a classical system. There exists a similar bound for quantum-mechanical systems,[3] and the rest of the proof is the same in either case. Consider first the case when $\rho_m = \infty$. Choose, for fixed z, ρ_2 so large that $\rho_2 > \rho_1$ and

$$z\rho_2^{-1} \exp(\beta W + 1) \le \tfrac{1}{2}. \tag{20}$$

From (19) and (20) we have

$$\sum_{N > \rho_2 V(\Omega)} (\Lambda^3 z)^N Z(\beta, N, \Omega) \le \sum_{N > \rho_2 V(\Omega)} \tfrac{1}{2}^N < 1. \tag{21}$$

From (10), (11), and (21) we obtain

$$P^{\epsilon}_{\Omega}(\beta, z) \le \left[1 + \sum_{N \in G^{\epsilon}(\Omega)} (\Lambda^3 z)^N Z(\beta, N, \Omega)\right] Q^{-1}(\beta, z, \Omega),$$

where $G^{\epsilon}(\Omega)$ is the set of all integers N such that

$N < \rho_2 V(\Omega)$ and $|NV(\Omega)^{-1} - \rho_1| > \epsilon$. There are less than $\rho_2 V(\Omega)$ terms in the sum so

$$P_\Omega^\epsilon(\beta, z) \leq \frac{[\rho_2 V(\Omega) + 1]}{Q(\beta, z, \Omega)} \max_{N \in G^\epsilon(\Omega)} \{(\Lambda^3 z)^N Z(\beta, N, \Omega)\}, \tag{22}$$

where we have also used the fact that

$$\max_{N \in G^\epsilon(\Omega)} \{(\Lambda^3 z)^N Z(\beta, N, \Omega)\} \geq Z(\beta, 0, \Omega) = 1.$$

It is clear that

$$P_\Omega^\epsilon(\beta, z) \geq Q(\beta, z, \Omega)^{-1} \max_{N \in G^\epsilon(\Omega)} \{(\Lambda^3 z)^N Z(\beta, N, \Omega)\}. \tag{23}$$

From (22) we have

$$V(\Omega)^{-1} \ln P_\Omega^\epsilon(\beta, z) \leq V(\Omega)^{-1} \ln [\rho_2 V(\Omega) + 1] + \max_{N \in G^\epsilon(\Omega)} \{NV(\Omega)^{-1} \ln \Lambda^3 z + g[\beta, NV(\Omega)^{-1}, \Omega]\} - \pi(\beta, z, \Omega), \tag{24}$$

and from (23)

$$V(\Omega)^{-1} \ln P_\Omega^\epsilon(\beta, z) \geq \max_{N \in G^\epsilon(\Omega)} \{NV(\Omega)^{-1} \ln \Lambda^3 z + g[\beta, NV(\Omega)^{-1}, \Omega]\} - \pi(\beta, z, \Omega). \tag{25}$$

Now choose a sequence of regions Ω_j satisfying the regularity conditions (c). Since $g(\beta, \rho, \Omega_j) \to g(\beta, \rho)$, uniformly in the interval $[0, \rho_2]$ and $V_j^{-1} \times \ln [\rho_2 V_j + 1] \to 0$, $\pi(\beta, z, \Omega_j) \to \pi(\beta, z)$, we have from (24)

$$V_j^{-1} \ln P_j^\epsilon(\beta, z) \leq \max_{N \in G^\epsilon(\Omega_j)} [NV_j^{-1} \ln \Lambda^3 z + g(\beta, NV_j^{-1})] - \pi(\beta, z) + \delta_j,$$

where $\delta_j \to 0$. Then

$$V_j^{-1} \ln P_j^\epsilon(\beta, z) \leq \max_{\rho \in G^\epsilon(\rho_1)} [\ln \rho \Lambda^3 z + g(\beta, \rho)] - \pi(\beta, z) + \delta_j, \tag{26}$$

where $G^\epsilon(\rho_1)$ is the set of ρ such that

$$0 \leq \rho \leq \rho_2 \quad \text{and} \quad |\rho - \rho_1| > \epsilon.$$

From (25) we have

$$V_j^{-1} \ln P_j^\epsilon(\beta, z) \geq \max_{N \in G^\epsilon(\Omega_j)} [NV_j^{-1} \ln \Lambda^3 z + g(\beta, NV_j^{-1})] - \pi(\beta, z) - \delta_j.$$

Since $\rho \ln \Lambda^3 z + g(\beta, \rho)$ is uniformly continuous in ρ in the interval $[0, \rho_2]$, it must be true that

$$\left| \max_{N \in G^\epsilon(\Omega_j)} [NV_j^{-1} \ln (\Lambda z) + g(\beta, NV_j^{-1})] - \max_{\rho \in G^\epsilon(\rho_1)} [\rho \ln \Lambda^3 z + g(\beta, \rho)] \right| < \gamma_j,$$

where $\gamma_j \to 0$. Hence,

$$V_j^{-1} \ln P_j^\epsilon(\beta, z) \geq \max_{\rho \in G^\epsilon(\rho_1)} [\rho \ln \Lambda^3 z + g(\beta, \rho)] - \pi(\beta, z) - \gamma_j - \delta_j. \tag{27}$$

Combining (26) and (27) and passing to the limit we find

$$\lim_{j \to \infty} V_j^{-1} \ln P_j^\epsilon(\beta, z) = \max_{\rho \in G^\epsilon(\rho_1)} [\rho \ln \Lambda^3 z + g(\beta, \rho)] - \pi(\beta, z). \tag{28}$$

By (9) the maximum value of the right-hand side of (28) is zero. By assumption of the one-phase state, this maximum is attained only for $\rho = \rho_1$. But ρ_1 is not in $G^\epsilon(\rho_1)$; hence, the limit in (28) must be less than zero.

In the case $\rho_m \neq \infty$, we simply take $\rho_2 = \rho_m$ in the previous calculation. In this case, however, we must assume that the maximum in (28) occurs at a value of ρ less than ρ_m. This is necessary because we can approximate $g(\beta, \rho, \Omega_j)$ by $g(\beta, \rho)$ uniformly in ρ only in a closed subinterval $0 \leq \rho \leq \rho' < \rho_m$.

IV. DENSITY FLUCTUATIONS IN THE CANONICAL FORMALISM

In this section we restrict ourselves to classical systems. Later in this section we need some stronger assumptions on the potentials, too. We want to consider the probability that a macroscopic subvolume of the total volume has a density different from the over-all density. Let us choose a sequence of domains $\{\Omega_j\}$ and a sequence of subdomains $\{\omega_j\}$, where ω_j is in Ω_j. We indicate the set of points in Ω_j but not in ω_j by $\Omega_j - \omega_j$. We assume the sequences $\{\Omega_j\}$, $\{\omega_j\}$, and $\{\Omega_j - \omega_j\}$ are chosen to satisfy the condition (c) in Sec. II which is used in the thermodynamic limit. In particular, we note that for all j, the three domains must be simply connected (Fig. 1). This condition is not necessary for the proof of the thermodynamic limit, but it is used in Ref. 3, and we retain it here. We want ω to be a macroscopic subvolume of Ω, so we require that $V(\omega_j)/V(\Omega_j) \to \alpha$, where $0 < \alpha < 1$.

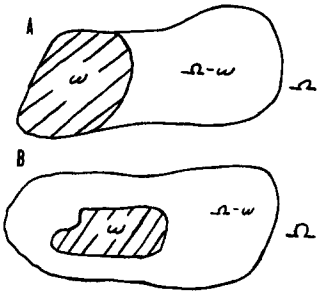


FIG. 1. (A) Ω, ω, $\Omega - \omega$ all simply connected; (B) $\Omega - \omega$ not simply connected.

Given Ω and ω, the probability of finding n particles in ω if the total system has N particles is

$$P_\Omega(\beta, n, \omega, N) = [n!\,(N-n)!\,Z(\beta, N, \Omega)]^{-1} \times \int_\omega \cdots \int_{\Omega-\omega} \exp(-\beta U_N)\, d\mathbf{r}_1 \cdots d\mathbf{r}_n\, d\mathbf{r}_{n+1} \cdots d\mathbf{r}_N, \tag{29}$$

where the first n integrations are over ω and the rest are over $\Omega - \omega$. Let $P^\epsilon_\Omega(\beta, \rho, \omega, N)$ be the probability that the domain ω has a density differing from ρ by more than ϵ.

$$P^\epsilon_\Omega(\beta, \rho, \omega, N) = \sum_{|n-\rho V(\omega)| > \epsilon V(\omega)} P_\Omega(\beta, n, \omega, N). \tag{30}$$

In this section (β, ρ) is a two-phase state if there is a neighborhood of ρ, such that $g(\beta, \rho)$ is linear in ρ in that neighborhood. Otherwise, (β, ρ) is a one-phase state. For notational convenience, we set $P^\epsilon_j(\beta, \rho) = P^\epsilon_{\Omega_j}(\beta, \rho, \omega_j, N_j)$.

The theorem can be stated as follows: Let $N_j V(\Omega_j)^{-1} \to \rho$, then if (β, ρ) is a two-phase state, there is an $\epsilon > 0$ such that

$$\lim_{j\to\infty} V_j(\Omega)^{-1} \ln P^\epsilon_j(\beta, \rho) = 0. \tag{31}$$

If (β, ρ) is a one-phase state, then for any $\epsilon > 0$

$$\lim_{j\to\infty} V(\Omega_j)^{-1} \ln P^\epsilon_j(\beta, \rho) < 0. \tag{32}$$

Let us consider the two-phase state first. Since each term in the sequence (31) is negative, we need only a lower bound on $V_j(\Omega)^{-1} \ln P^\epsilon_j(\beta, \rho)$. We choose ϵ in the following way. Since (β, ρ) is a two-phase state, we can find $\gamma > 0$ such that $g(\beta, \rho')$ is a linear function of ρ' for all $|\rho' - \rho| < \gamma$. Let ρ_1 and ρ_2 be any densities such that $\rho_1\alpha + \rho_2(1-\alpha) = \rho$. We want to choose ϵ so small that we can find ρ_1 such that

$$|\rho_1 - \rho| < \gamma, \quad |\rho_2 - \rho| < \gamma, \quad |\rho_1 - \rho| > \epsilon.$$

This is possible if we choose

$$\epsilon < \min\{\gamma, \gamma[(1-\alpha)/\alpha]\}. \tag{33}$$

Now let n_j be a sequence of integers such that

$$n_j V(\omega_j)^{-1} \to \rho_1.$$

Since $V(\omega_j)V(\Omega_j)^{-1} \to \alpha$ and $N_j V(\Omega_j)^{-1} \to \rho$, we have $(N_j - n_j)[V(\Omega_j) - V(\omega_j)]^{-1} \to \rho_2$. Then we can find a j_0 so that for all $j > j_0$ we have

$$\begin{aligned} &|n_j V(\omega_j)^{-1} - \rho| < \gamma, \\ &|(N_j - n_j)[V(\Omega_j) - V(\omega_j)]^{-1} - \rho| < \gamma, \\ &|n_j V(\omega_j)^{-1} - \rho| > \epsilon. \end{aligned} \tag{34}$$

From (30), $P^\epsilon_j(\beta, \rho)$ is larger than any term in the sum; therefore, we can put $n = n_j$ and write

$$P^\epsilon_j(\beta, \rho) \geq [n_j!\,(N_j - n_j)!\,Z(\beta, N_j, \Omega_j)]^{-1} \times \int_{\omega_j} \cdots \int_{\Omega_j - \omega_j} e^{-\beta U_{N_j}}\, d\mathbf{r}_1 \cdots d\mathbf{r}_{n_j}\, d\mathbf{r}_{n_j+1} \cdots d\mathbf{r}_{N_j}. \tag{35}$$

Now let ω'_j be the set of all points interior to ω_j and distance R_0 from the boundary of ω_j. If we allow the first n_j integrations in (35) to go only over ω'_j, we decrease the value of the integral. But now the interaction between particles in ω'_j and $\Omega_j - \omega_j$ is negative [see Sec. II, assumption (b)]. Therefore, if we replace U_{N_j} by $U_{N_j - n_j} + U_{n_j}$, we decrease the integrand and the resulting integral factors into a product.

$$P^\epsilon_j(\beta, \rho) \geq Z(\beta, n_j, \omega'_j) Z(\beta, N_j - n_j, \Omega_j - \omega_j) \times Z(\beta, N_j, \Omega_j)^{-1}.$$

Hence, if the limits exist,

$$\begin{aligned} \lim_{j\to\infty} V(\Omega_j)^{-1} \ln P^\epsilon_j(\beta, \rho) &\geq [V(\omega'_j)V(\Omega_j)^{-1}] \\ &\times \{V(\omega'_j)^{-1} \ln Z(\beta, n_j, \omega'_j)\} + [V(\Omega_j) - V(\omega_j)]V(\Omega_j)^{-1} \\ &\times \{[V(\Omega_j) - V(\omega_j)]^{-1} \ln Z(\beta, N_j - n_j, \Omega_j - \omega_j)\} \\ &- V(\Omega_j)^{-1} \ln Z(\beta, N_j, \Omega_j). \end{aligned} \tag{36}$$

From assumption (b) in Sec. II, we have

$$V(\omega'_j)V(\omega_j)^{-1} \to 1.$$

By the way we have chosen n_j, we know $n_j V(\omega'_j)^{-1} \to \rho_1$, $(N_j - n_j)[V(\Omega_j) - V(\omega_j)]^{-1} \to \rho_2$, and therefore,

$$\lim_{j\to\infty} V(\Omega_j)^{-1} \ln P^\epsilon_j(\beta, \rho) \geq \alpha g(\beta, \rho_1) + (1-\alpha) g(\beta, \rho_2) - g(\beta, \rho). \tag{37}$$

But the right side of (37) is zero since $g(\beta, \rho)$ is assumed linear in the interval containing ρ_1 and ρ_2. So if the limit exists, it is greater than or equal to zero. But each term in the original sequence is negative; therefore, the limit exists and is zero.

Now let (β, ρ) be a one-phase state. For any j and $\epsilon > 0$, we have from (30)

$$P^\epsilon_j(\beta, \rho) \leq N_j \max_{|n-\rho V(\omega_j)| > \epsilon V(\omega_j)} P_{\Omega_j}(\beta, n, \omega_j, N_j). \tag{38}$$

We want to find an upper bound on $P_{\Omega_j}(\beta, n, \omega_j, N_j)$. To do this we see from (29) that we need a lower bound on the interaction across the surface of ω_j.

To obtain this lower bound we define

$$\Phi(\omega) = \min [U_{N_1+N_2}(\mathbf{r}_1 \cdots \mathbf{r}_{N_1}, \mathbf{r}'_1 \cdots \mathbf{r}'_{N_2}) - U_{N_1}(\mathbf{r}_1 \cdots \mathbf{r}_{N_1}) - U_{N_2}(\mathbf{r}'_1 \cdots \mathbf{r}'_{N_2})], \tag{39}$$

where the minimum is taken over all configurations $(\mathbf{r}_1 \cdots \mathbf{r}_{N_1}, \mathbf{r}'_1 \cdots \mathbf{r}'_{N_2})$ such that $U_{N_1+N_2} \neq +\infty$,

$(\mathbf{r}_1 \cdots \mathbf{r}_{N_1})$ are all in ω, $(\mathbf{r}'_1 \cdots \mathbf{r}'_{N_2})$ are not in ω, and N_1 and N_2 vary from 0 to ∞. We must assume

$$\Phi(\omega_j)V(\omega_j)^{-1} \to 0. \tag{40}$$

In Appendix B we show that (40) holds for the case where U_N is a sum of pair potentials of finite range with hard cores. In fact, for such potentials

$$\Phi(\omega) > KV(b, \omega), \tag{41}$$

where b is the range of the potential, K is a constant, and $V(b, \omega)$ is the volume of those points interior to ω and within b of the boundary of ω. From (29) and (39) we have the upper bound,

$$P_{\Omega_j}(\beta, n, \omega_j, N_j) \leq Z(\beta, n, \omega_j)Z(\beta, N_j - n, \Omega_j - \omega_j) \times Z(\beta, N_j, \omega_j)^{-1} \exp[-\beta\Phi(\omega_j)]. \tag{42}$$

From (38) and (42) we have

$$\begin{aligned} V(\Omega_j)^{-1} \ln P^{\epsilon}_j(\beta, \rho) &\leq V(\Omega_j)^{-1} \ln N_j + \max_{|n-\rho V(\omega_j)| > \epsilon V(\omega_j)} V(\Omega_j)^{-1} \\ &\times \{\ln Z(\beta, n, \omega_j) + \ln Z(\beta, N_j - n_j, \Omega_j - \omega_j)\} \\ &- \ln Z(\beta, N_j, \omega_j) - \beta V(\Omega_j)^{-1}\Phi(\omega_j). \end{aligned} \tag{43}$$

The first term in (43) converges to zero. By virtue of (40) and the assumption $V(\omega_j)V(\Omega_j)^{-1} \to \alpha$, the last term in (43) converges to zero. If we assume the maximum in (43) occurs for each j for $nV(\omega_j)^{-1} \leq \rho_2 < \rho_m$, we can approximate each $\ln Z$ by the corresponding g to obtain, for large j,

$$\begin{aligned} V(\Omega_j)^{-1} \ln P_j(\beta, \rho) &\leq \max_{|nV(\omega_j)^{-1} - \rho| > \epsilon} \{\alpha g[\beta, nV(\omega_j)^{-1}] + (1 - \alpha) \\ &\times g(\beta, [N_j - n][V(\Omega_j) - V(\omega_j)]^{-1})\} \\ &- g(\beta, \rho) + \delta_j, \end{aligned} \tag{44}$$

where $\delta_j \to 0$ as $j \to \infty$. Because $g(\beta, \rho)$ is uniformly continuous, we can set

$$\begin{aligned} V(\Omega_j)^{-1} \ln P^{\epsilon}_j(\beta, \rho) &\leq \max_{|\rho' - \rho| > \epsilon} \{\alpha g(\beta, \rho') + (1 - \alpha)g(\beta, \rho'') - g(\beta, \rho)\} \\ &+ \delta_j + \gamma_j, \end{aligned}$$

where $\gamma_j \to 0$ and ρ'' is defined by $\alpha\rho' + (1 - \alpha)\rho'' = \rho$. Since g is a convex function of ρ, the combination $\alpha g(\beta, \rho') + (1 - \alpha)g(\beta, \rho'') - g(\beta, \rho)$ is less than or equal to zero for all ρ'. If ρ is not in a linear portion of g, then this expression is negative for all $\rho' \neq \rho$. Since ρ' is bounded away from ρ in the maximization, the maximum is negative; hence,

$$\begin{aligned} \lim_{j \to \infty} V(\Omega_j)^{-1} \ln P_j(\beta, \rho) &\leq \max_{|\rho' - \rho| > \epsilon} \{\alpha g(\beta, \rho') + (1 - \alpha)g(\beta, \rho'') - g(\beta, \rho)\} < 0. \end{aligned} \tag{45}$$

We have shown that if the limit on the left exists it is bounded from above by a negative number. To show that the limit exists, we should bound it from below by the same number. We note from (30) that

$$P^{\epsilon}_j(\beta, \rho) \geq \max_{|nV(\omega_j)^{-1} - \rho| > \epsilon} P^{\epsilon}_{\Omega_j}(\beta, n, \omega_j, N_j). \tag{46}$$

Since we need a lower bound on $P^{\epsilon}_{\Omega_j}(\beta, n, \omega_j, N_j)$, we can proceed as in the two-phase case. The calculation is straightforward so we omit it. The result is that one obtains a lower bound of the same form as in (45) and, hence, completes the theorem.

V. CONCLUSIONS

We have applied Dobrushin's idea for characterizing large and small density fluctuations to the case of a fluid described by either canonical or the grand canonical formalism. We have proven the fluctuations are large in the two-phase region and small in the one-phase region. It is interesting to note that by this criterion the density fluctuations at the critical point are small if the critical isotherm has only a point of inflection and not a flat portion.

ACKNOWLEDGMENTS

The first author (G. L. J.) was supported in part by the United States Atomic Energy Commission under Contract AT(11-1)-427. The second author (R. E. K.) was supported by The Radiation Laboratory of the University of Notre Dame, which is operated under contract with the United States Atomic Energy Commission. This is AEC Document No. COO-38-479.

APPENDIX A

We want to show that there is a two-phase state, if and only if $g(\beta, \rho)$ is a linear function of ρ in some interval. Suppose there are two densities $\rho_1 < \rho_2$ which maximize $\rho \ln \Lambda^3 z + g(\beta, \rho)$. Then

$$\begin{aligned} \rho_1 \ln \Lambda^3 z + g(\beta, \rho_1) &= \rho \ln \Lambda^3 z + g(\beta, \rho_2) \\ &\geq \rho \ln \Lambda^3 z + g(\beta, \rho) \\ &\text{for any } \rho. \end{aligned} \tag{A1}$$

Let $0 \leq \alpha \leq 1$ and set $\rho = (1 - \alpha)\rho_1 + \alpha\rho_2$. Then

$$\begin{aligned} \rho \ln \Lambda^3 z + g(\beta, \rho) &\leq \rho_1 \ln \Lambda^3 z + g(\beta, \rho_1) \\ &= (1 - \alpha)\rho_1 \ln \Lambda^3 z + (1 - \alpha)g(\beta, \rho_1) \\ &+ \alpha\rho_2 \ln \Lambda^3 z + \alpha g(\beta, \rho_2) \\ &= \rho \ln \Lambda^3 z + (1 - \alpha)g(\beta, \rho_1) + \alpha g(\beta, \rho_2). \end{aligned}$$

Hence,

$$g(\beta, \rho) \leq (1 - \alpha)g(\beta, \rho_1) + \alpha g(\beta, \rho_2). \tag{A2}$$

But since $g(\beta, \rho)$ is convex upward in ρ,

$$g(\beta, \rho) \geq (1 - \alpha)g(\beta, \rho_1) + \alpha g(\beta, \rho_2). \tag{A3}$$

Equations (A2) and (A3) imply

$$g(\beta, \rho) = (1 - \alpha)g(\beta, \rho_1) + \alpha g(\beta, \rho_2),$$

which means g is linear in ρ for

$$\rho_1 \leq \rho \leq \rho_2.$$

Now consider the converse case. Let ρ_1 maximize $\rho \ln \Lambda^3 z + g(\beta, \rho)$ and suppose $g(\beta, \rho)$ is linear in a neighborhood of ρ_1. If ρ_2 is close enough to ρ_1, then $g(\beta, \rho_2) = g(\beta, \rho_1) + c(\rho_2 - \rho_1)$, where $c = (dg/d\rho)(\beta, \rho_1)$. Since ρ_1 is a point of maximum of $\rho \ln \Lambda^3 z + g(\beta, \rho)$, we must have $\ln \Lambda^3 + c = 0$. Hence,

$$g(\beta, \rho_2) = g(\beta, \rho_1) - (\rho_2 - \rho_1) \ln \Lambda^3 z,$$

and this implies that ρ_2 is also a point of maximum.

If the point ρ_1 is the end point of a linear portion of $g(\beta, \rho)$ and if $(dg/d\rho)(\beta, \rho_1)$ does not exist there, then this proof fails at ρ_1. This cannot happen if the pressure is a continuous function of ρ.

APPENDIX B

Let

$$U_N(\mathbf{r}_1 - \mathbf{r}_N) = \sum_{i<j=2}^{N} \phi(\mathbf{r}_i - \mathbf{r}_j),$$

and suppose

$$\phi(\mathbf{r}) = +\infty \quad \text{if} \quad \mathbf{r} < a, \quad \phi(\mathbf{r}) = 0 \quad \text{if} \quad \mathbf{r} > b.$$

We assume $\phi(\mathbf{r})$ is bounded from below by $-d$, where $d \geq 0$. From (38),

$$\Phi(\omega) = \min \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \phi(\mathbf{r}_i - \mathbf{r}'_j),$$

where the $\mathbf{r}_i$ are in ω and the $\mathbf{r}'_r$ are not, and no two particles are ever closer than a. Now each particle can interact with at most $(8b^3/a^3)$ others without violating the hard-core condition. The most negative each interaction can be is $-d$. Only those particles within b of the boundary of ω can interact with those particles not in ω, and there are at most $V(b, \omega)[\frac{1}{6}\pi a^3]^{-1}$ such particles. Hence,

$$\Phi(\omega) > -(db^3 a^{-6} \tfrac{4}{3}\pi) V(b, \omega).$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Strong Coupling Limit in Potential Theory. I

WILLIAM M. FRANK*
Bar-Ilan University, Ramat Gan, Israel

(Received 16 June 1966)

Analytic properties of the Jost function in g, the coupling constant, are studied for potentials which are $L^{\frac{1}{2}}$, i.e., $\int_0^\infty dr\, |V(r)|^{\frac{1}{2}} < \infty$, and have a negative power or exponential tail for large distances. For the bound state and scattering problems, it is found that the Jost function has exponential order $\frac{1}{2}$ for large g, which implies that the scattering phase shift and the number of bound states in an attractive potential grow like $|g|^{\frac{1}{2}}$ for large g. The latter result is the best possible and is a considerable improvement over earlier estimates.

THIS investigation of strong coupling methods in potential theory is motivated, as is much of the attention given these days to potential theory, by analogies to elementary particle theory and field theory. The desirability of strong coupling methods to study strong and broken interactions of elementary particles is self-evident. Another motivation for this study derives from some recent investigations[1] in field theory which show in the case of some solvable models that the renormalized perturbation solution in field theory represents a function, singular at the origin but analytic at the point at infinity, as a function of g, the coupling constant. With the aim of developing techniques for calculation of a perturbation series in negative powers of g, the corresponding potential theory problem is first considered. Naturally, the problem has its own intrinsic interest both theoretically and practically. From the practical point of view a strong coupling expansion would complement the usual Born expansion whose convergence limitations are well known, in particular for problems of strongly repulsive potentials.[2] For potential theory in fact, a

* Present address: U.S. Naval Ordnance Laboratory, Silver Spring, Maryland.

[1] W. M. Frank, Nuovo Cimento **38**, 1077 (1965).

[2] See, e.g., S. Weinberg, Phys. Rev. **131**, 440 (1963); N. Rotenberg, Ann. Phys. (N.Y.) **21**, 579 (1963); S. Weinberg, preprint "Perturbation Theory for Strongly Repulsive Potentials."

Equations (A2) and (A3) imply

$$g(\beta, \rho) = (1 - \alpha)g(\beta, \rho_1) + \alpha g(\beta, \rho_2),$$

which means g is linear in ρ for

$$\rho_1 \leq \rho \leq \rho_2.$$

Now consider the converse case. Let ρ_1 maximize $\rho \ln \Lambda^3 z + g(\beta, \rho)$ and suppose $g(\beta, \rho)$ is linear in a neighborhood of ρ_1. If ρ_2 is close enough to ρ_1, then $g(\beta, \rho_2) = g(\beta, \rho_1) + c(\rho_2 - \rho_1)$, where $c = (dg/d\rho)(\beta, \rho_1)$. Since ρ_1 is a point of maximum of $\rho \ln \Lambda^3 z + g(\beta, \rho)$, we must have $\ln \Lambda^3 + c = 0$. Hence,

$$g(\beta, \rho_2) = g(\beta, \rho_1) - (\rho_2 - \rho_1) \ln \Lambda^3 z,$$

and this implies that ρ_2 is also a point of maximum.

If the point ρ_1 is the end point of a linear portion of $g(\beta, \rho)$ and if $(dg/d\rho)(\beta, \rho_1)$ does not exist there, then this proof fails at ρ_1. This cannot happen if the pressure is a continuous function of ρ.

APPENDIX B

Let

$$U_N(\mathbf{r}_1 - \mathbf{r}_N) = \sum_{i<j=2}^{N} \phi(\mathbf{r}_i - \mathbf{r}_j),$$

and suppose

$$\phi(\mathbf{r}) = +\infty \quad \text{if} \quad \mathbf{r} < a, \quad \phi(\mathbf{r}) = 0 \quad \text{if} \quad \mathbf{r} > b.$$

We assume $\phi(\mathbf{r})$ is bounded from below by $-d$, where $d \geq 0$. From (38),

$$\Phi(\omega) = \min \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \phi(\mathbf{r}_i - \mathbf{r}'_j),$$

where the $\mathbf{r}_i$ are in ω and the $\mathbf{r}'_r$ are not, and no two particles are ever closer than a. Now each particle can interact with at most $(8b^3/a^3)$ others without violating the hard-core condition. The most negative each interaction can be is $-d$. Only those particles within b of the boundary of ω can interact with those particles not in ω, and there are at most $V(b, \omega)[\frac{1}{6}\pi a^3]^{-1}$ such particles. Hence,

$$\Phi(\omega) > -(db^3 a^{-6} \tfrac{4}{3}\pi) V(b, \omega).$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Strong Coupling Limit in Potential Theory. I

WILLIAM M. FRANK*
Bar-Ilan University, Ramat Gan, Israel

(Received 16 June 1966)

Analytic properties of the Jost function in g, the coupling constant, are studied for potentials which are $L^{\frac{1}{2}}$, i.e., $\int_0^\infty dr\, |V(r)|^{\frac{1}{2}} < \infty$, and have a negative power or exponential tail for large distances. For the bound state and scattering problems, it is found that the Jost function has exponential order $\frac{1}{2}$ for large g, which implies that the scattering phase shift and the number of bound states in an attractive potential grow like $|g|^{\frac{1}{2}}$ for large g. The latter result is the best possible and is a considerable improvement over earlier estimates.

THIS investigation of strong coupling methods in potential theory is motivated, as is much of the attention given these days to potential theory, by analogies to elementary particle theory and field theory. The desirability of strong coupling methods to study strong and broken interactions of elementary particles is self-evident. Another motivation for this study derives from some recent investigations[1] in field theory which show in the case of some solvable models that the renormalized perturbation solution in field theory represents a function, singular at the origin but analytic at the point at infinity, as a function of g, the coupling constant. With the aim of developing techniques for calculation of a perturbation series in negative powers of g, the corresponding potential theory problem is first considered. Naturally, the problem has its own intrinsic interest both theoretically and practically. From the practical point of view a strong coupling expansion would complement the usual Born expansion whose convergence limitations are well known, in particular for problems of strongly repulsive potentials.[2] For potential theory in fact, a

* Present address: U.S. Naval Ordnance Laboratory, Silver Spring, Maryland.

[1] W. M. Frank, Nuovo Cimento **38**, 1077 (1965).

[2] See, e.g., S. Weinberg, Phys. Rev. **131**, 440 (1963); N. Rotenberg, Ann. Phys. (N.Y.) **21**, 579 (1963); S. Weinberg, preprint "Perturbation Theory for Strongly Repulsive Potentials."

strong coupling expansion is really an asymptotic expansion, since quantities such as the Jost function have an essential singularity at infinity. One interesting theoretical result we find is a significantly improved bound on the numbers of bound states in a given potential. This result has also been recently discovered by Calogero.[3] Bound l-wave states are considered first in Sec. I. In Sec. II more specific results are established for the bound S-wave state. In Sec. III scattering states are considered. This article restricts itself to spherically symmetric local potentials which are $L^{\frac{1}{2}}$ and whose absolute square root can be expressed as a Laplace transform in the radial variable. Singular potentials and those which go to zero at large distances faster than any negative linear exponential will be dealt with in separate articles.

The strong coupling limit is studied via function-theoretic techniques. The scattering equation is written as an integral equation and the Jost function can be represented in terms of the familiar power series (in the coupling constant g) for the Fredholm determinant of this integral equation. Some standard results in the theory of entire functions connect the growth of the power series coefficients with the nature of the essential singularity of the power series for large coupling constant. These connections provide the strong coupling behavior of the Jost function whose zeros represent the eigenvalues g for which bound states exist for fixed energy, and whose phase corresponds to the scattering phase shift.

I. BOUND STATES: GENERAL l WAVES

We restrict our considerations to potentials $V(r)$ which are local, central, and are $L^{\frac{1}{2}}$, i.e., which satisfy

$$\int_0^\infty dr\,|V(r)|^{\frac{1}{2}} < \infty$$

(all integrability conditions are to be understood in one dimension). This condition is somewhat novel to potential theory, but is in fact the significant one from a number of points of view. We also demand that $|V(r)|^{\frac{1}{2}}$ be expressible as a Laplace transform

$$|V(r)|^{\frac{1}{2}} = \int_0^\infty d\alpha\sigma(\alpha)e^{-\alpha r}, \tag{1}$$

which means that $V(r)e^{Kr} \to \infty$ for sufficiently large r for some K. Potentials which decrease at large distances faster than any linear exponential will be considered in a separate article.

We consider general l-wave states, and fix our attention on a given l value. $V(r)$ is assumed to be real and for the moment everywhere nonnegative (i.e., generally attractive). The Schrödinger equation describing an l-wave bound state with energy $E = -\hbar^2\mu^2/2m$ reads

$$\left[\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \mu^2 - \frac{l(l+1)}{r^2} + gV(r)\right]\frac{u_l(r)}{r} = 0, \tag{2}$$

with $u_l(r)/r \equiv \psi_l(r)$ the wavefunction, and g is a coupling constant which must be positive for bound states to exist. We introduce the Fourier–Bessel transforms defined by

$$\hat{u}_l(k) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\int_0^\infty dr r^2 j_l(kr)\frac{u_l(r)}{r},$$

$$\frac{u_l(r)}{r} = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\int_0^\infty dk k^2 j_l(kr)\hat{u}_l(k), \tag{3}$$

with $j_l(kr)$ the familiar spherical Bessel function. From Eqs. (2), (3) one finds the integral equation (the subscript l is dropped where the context allows) [$\omega_k = (k^2 + \mu^2)^{\frac{1}{2}}$]

$$\hat{u}(k) = \frac{g}{k^2 + \mu^2}\int_0^\infty dk' k'^2 V_l(k, k')\hat{u}(k') \tag{4}$$

with

$$V_l(k, k') = \frac{2}{\pi}\int_0^\infty dr r^2 V(r) j_l(kr) j_l(k'r) \tag{5}$$

a real symmetric kernel. Setting

$$\phi_l(k) = k\omega_k\hat{u}_l(k), \quad U_l(k, k') \equiv kk' V_l(k, k')/\omega_k\omega_{k'},$$

Eq. (4) becomes the Fredholm eigenvalue equation

$$\phi_l(k) = g\int dk' U_l(k, k')\phi_l(k') \tag{6}$$

with the real symmetric kernel $U_l(k, k')$. $U_l(k, k')$ is L^1 if $\int dr\,|V(r)| < \infty$. Equation (4) is an eigenvalue equation whose eigenvalues g_r are the zeros of the Fredholm determinant,[4] alias the Jost function. The correspondence between the Fredholm determinant and the Jost function has been demonstrated for the scattering problem, but is to be rederived as a basic step in our reasoning, and to establish notation. The Fredholm determinant for Eq. (6) is constructed as

$$\Delta(g) = \sum_{n=0}^\infty \frac{(-g)^n}{n!}\int dk_1 \cdots \int dk_n \det_{1\le i,j\le n} |U_l(k_i, k_j)|, \tag{7}$$

where an evident notation for the determinant is used.

[3] F. Calogero, Commun. Math. Phys. 1, 80 (1965).

[4] R. Jost and A. Pais, Phys. Rev. 82, 840 (1951), Appendix.

The determinant can be transformed (i labels rows, j labels columns)

$$\begin{aligned}
&\det |U_l(k_i, k_j)| \\
&= \det \left| \frac{k_i k_j}{\omega_{k_i}\omega_{k_j}} V_l(k_i, k_j) \right| \\
&= \prod_{m=1}^{n} \left(\frac{k_m}{\omega_{k_m}}\right)^2 \det |V_l(k_i, k_j)| \\
&= \prod_{m=1}^{n} \left(\frac{k_m}{\omega_{k_m}}\right)^2 \det \left| \frac{2}{\pi} \int_0^\infty dr_i r_i^2 V(r_i) j_l(k_j r_i) j_l(k_i r_i) \right| \\
&= \left(\frac{2}{\pi}\right)^n \prod_{m=1}^{n} \left(\frac{k_m}{\omega_{k_m}}\right)^2 \int_0^\infty dr_1 r_1^2 \cdots \int_0^\infty dr_n r_n^2 \\
&\quad \times V(r_1) \cdots V(r_n) j_l(k_1 r_1) \cdots j_l(k_n r_n) \det |j_l(k_j r_i)| \\
&= \int_0^\infty dr_1 \cdots \int_0^\infty dr_n r_1^2 V(r_1) \cdots r_n^2 V(r_n) \\
&\quad \times \det \left| \frac{2}{\pi} \frac{k_j^2}{\omega_{k_j}^2} j_l(k_j r_t) j_l(k_j r_j) \right|. \qquad (8)
\end{aligned}$$

Performance of the k integrations of Eq. (7) inside the determinant leads to

$$\begin{aligned}
&\int_0^\infty dk_1 \cdots \int_0^\infty dk_n \det |U_l(k_i, k_j)| \\
&\quad = \int_0^\infty dr_1 \cdots \int_0^\infty dr_n V(r_1) \cdots V(r_n) \det |\mathfrak{G}_l(r_i, r_l)|, \qquad (9)
\end{aligned}$$

where

$$\mathfrak{G}_l(r, r') \equiv \frac{2}{\pi} \int_0^\infty dk \, \frac{kr j_l(kr) kr' j_l(kr')}{\omega_k^2}. \qquad (10)$$

This quantity is in fact the Green's function of the differential operator

$$\mathfrak{D}_{l,\mu} \equiv (d^2/dr^2) - \mu^2 - [l(l+1)/r^2]$$

corresponding to solutions which vanish at $r = 0$ and $r = \infty$. We can therefore write

$$\mathfrak{G}_l(r,r') = \mu r_< r_> j_l(i\mu r_<) h_l^{(+)}(i\mu r_>), \qquad (11)$$

with $rj_l(i\mu r)$, $rh_l^{(+)}(i\mu r)$ zeros of the differential operator $\mathfrak{D}_{l,\mu}$ which respectively vanish at $r = 0$, $r = \infty$. $j_l(\rho)$, $h_l^{(+)}(\rho)$ are respectively the regular spherical Bessel and the spherical Hankel function, and for large ρ

$$\rho j_l(\rho) \sim \sin(\rho - \tfrac{1}{2}l\pi), \quad \rho h_l^{(+)}(\rho) \sim \exp[i(\rho - \tfrac{1}{2}l\pi)].$$

The quantity

$$D_l(r_1 \cdots r_n) \equiv \det_{1 \le i,j \le n} |\mathfrak{G}_l(r_i, r_j)| \qquad (12)$$

is a symmetric function of all its arguments and is in fact nonnegative.[5] In view of this and Eq. (9), we can write

$$\begin{aligned}
\Delta(g) &= \sum_{n=0}^{\infty} \frac{(-g)^n}{n!} \int_0^\infty dr_1 \cdots \int_0^\infty dr_n V(r_1) \cdots V(r_n) \\
&\quad \times D_l(r_1 \cdots r_n) \\
&= \sum_{n=0}^{\infty} (-g)^n \int_0^\infty dr_1 \int_0^{r_1} dr_2 \cdots \int_0^{r_{n-1}} dr_n \\
&\quad \times V(r_1) \cdots V(r_n) D_l^{>}(r_1 \cdots r_n), \qquad (13)
\end{aligned}$$

where in $D_l^{>}(r_1 \cdots r_n)$ the condition $r_1 \ge r_2 \ge r_3 \cdots \ge r_n$ must be obeyed. Using the representation Eq. (11) one finds, with a little juggling in the determinant $D_l^{>}(r_1 \cdots r_n)$, the recursion relation

$$\begin{aligned}
D_l^{>}(r_1 \cdots r_n) &= \frac{\mu r_1^2 h_l^{(+)}(i\mu r_1)}{h_l^{(+)}(i\mu r_2)} |h_l^{(+)}(i\mu r_2) j_l(i\mu r_1) \\
&\quad - h_l^{(+)}(i\mu r_1) j_l(i\mu r_2)| \, D_l^{>}(r_2 \cdots r_n) \\
&\equiv H_l(r_1, r_2) D_l^{>}(r_2 \cdots r_n), \qquad (14)
\end{aligned}$$

which leads to

$$\begin{aligned}
&D_l^{>}(r_1 \cdots r_n) \\
&\quad = \mu r_n^2 j_l(i\mu r_n) h_l^{(+)}(i\mu r_n) H_l(r_1, r_2) \\
&\qquad \times H_l(r_2, r_3) \cdots H_l(r_{n-1}, r_n) \\
&\quad = H_l(r_1, r_2) H_l(r_2, r_3) \cdots H_l(r_{n-1}, r_n) \\
&\qquad \times H_l(r_n, 0). \qquad (15)
\end{aligned}$$

One can show that for all r, s

$$|H_l(r, s)|/|r - s| \le |H_0(r, s)|/|r - s| \le 1. \qquad (16)$$

Then from Eq. (13)

$$\begin{aligned}
|\Delta(g)| &\le \sum_{n=0}^{\infty} |g|^n \\
&\quad \times \int_0^\infty dr_1 \int_0^{r_1} dr_2 \cdots \int_0^{r_{n-1}} dr_n \, |V(r_1) \cdots V(r_n)| \\
&\quad \times (r_1 - r_2)(r_2 - r_3) \cdots (r_{n-1} - r_n) r_n. \qquad (17)
\end{aligned}$$

We make the variable changes

$$\xi_j = r_j - r_{j+1} \; (j = 1 \cdots n-1), \quad \xi_n = r_n.$$

Then

$$\begin{aligned}
|\Delta(g)| &\le \sum_{n=0}^{\infty} |g|^n \int_0^\infty d\xi \cdots \int_0^\infty d\xi_n \xi_1 \cdots \xi_n V(\xi_n) \\
&\quad \times V(\xi_{n-1} + \xi_n) \cdots V(\xi_1 + \cdots \xi_n) \\
&\equiv \sum_{n=0}^{\infty} |g|^n L_n. \qquad (18)
\end{aligned}$$

We set

$$U(r) \equiv |V(r)|^{\frac{1}{2}} = \int_0^\infty d\alpha \sigma(\alpha) e^{-\alpha r}$$

[5] See, e.g., discussion in W. M. Frank, Ann. Phys. (N.Y.) **29**, 175 (1964), Sec. III.

[$\sigma(\alpha)$ is known as the "Laplace weight" of $U(r)$]. Then

$$L_n = \int_0^\infty d\alpha_1 \cdots d\alpha_{2n} \times \frac{\sigma(\alpha_1)\cdots\sigma(\alpha_{2n})}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+\alpha_3+\alpha_4)^2\cdots(\alpha_1+\alpha_2+\cdots\alpha_{2n-1}+\alpha_{2n})^2}. \tag{19}$$

We use the inequality

$$(\alpha_1+\alpha_2+\cdots\alpha_{2k-1}+\alpha_{2k})^2]^{-1} \leq [(\alpha_1+\alpha_2+\cdots\alpha_{2k-1}) \times (\alpha_1+\alpha_2+\cdots\alpha_{2k-1}+\alpha_{2k})]^{-1} \tag{20}$$

to find

$$L_n \leq \int_0^\infty d\alpha_1 \cdots \int_0^\infty d\alpha_{2n} \times \frac{|\sigma(\alpha_1)\cdots\sigma(\alpha_{2n})|}{[\alpha_1(\alpha_1+\alpha_2)\cdots(\alpha_1+\alpha_2+\cdots\alpha_{2n-1}) \times (\alpha_1+\alpha_2+\cdots\alpha_{2n})]} = \frac{1}{(2n)!}\int_0^\infty d\alpha_1\cdots d\alpha_{2n}\,|\sigma(\alpha_1)\cdots\sigma(\alpha_{2n})| \times \sum_P [\alpha_{P_1}(\alpha_{P_1}+\alpha_{P_2})\cdots(\alpha_{P_1}+\alpha_{P_2}+\cdots\alpha_{P_{2n}})]^{-1}, \tag{21}$$

where the integral in Eq. (21) has been replaced by a symmetrized sum over all permutations P of the $\alpha_1, \alpha_2, \alpha_3, \cdots, \alpha_{2n}$. This quantity sums to a simple expression resulting in

$$L_n \leq \frac{1}{(2n)!}\int_0^\infty d\alpha_1 \frac{|\sigma(\alpha_1)|}{\alpha_1} \times \int_0^\infty d\alpha_2 \frac{|\sigma(\alpha_2)|}{\alpha_2}\cdots\int_0^\infty d\alpha_{2n}\frac{|\sigma(\alpha_{2n})|}{\alpha_{2n}} \equiv \frac{U_1^{2n}}{(2n)!} \tag{22}$$

with the notation

$$U_1 \equiv \int_0^\infty d\alpha \frac{|\sigma(\alpha)|}{\alpha}.$$

We find therefore that

$$|\Delta(g)| \leq \cosh(U_1|g|^{\frac{1}{2}}), \tag{23}$$

which shows $\Delta(g)$ to be an entire function of exponential order[6] at most $\frac{1}{2}$. That the exponential order of $\Delta(g)$ is precisely $\frac{1}{2}$ is demonstrated by providing a lower bound to $\Delta(g)$ for negative real g [where $\Delta(g)$ is positive], which also has exponential order $\frac{1}{2}$. Such a lower bound is readily found in the Fredholm determinant $\Delta^S(g)$ corresponding to a square-well potential $S(r)$ which satisfies $V(r) \geq S(r)$. Clearly since $D(r_1\cdots r_n)$ is nonnegative

$$\Delta(g) \geq \Delta^S(g) \tag{24}$$

for negative real g. That the exponential order of $\Delta^S(g)$ is $\frac{1}{2}$ follows from the consideration that $\Delta^S(g)$ can be expressed in terms of its Weierstrass factorization[7]

$$\Delta^S(g) = \prod_{m=1}^\infty \left(1 - \frac{g}{g_m}\right), \tag{25}$$

where $\Delta^S(g_m) = 0$. The zeros g_m of $\Delta^S(g)$ are just the values of coupling constant which bind a particle in the potential $gS(r)$ with energy $-\hbar^2\mu^2/2m$. If [$\theta(x)$ represents the step function, 0 for negative x, 1 for positive x]

$$S(r) = V_0\theta(b-r),$$

the g_m are solutions to

$$\cot b(V_0 g_m - \mu^2)^{\frac{1}{2}} = -\mu(V_0 g_m - \mu^2)^{-\frac{1}{2}} \tag{26}$$

and clearly for large m

$$g_m \approx (m+\tfrac{1}{2})^2\pi^2/b^2V_0, \tag{27}$$

so that the g_m have exponent of convergence $\frac{1}{2}$. This establishes[8] the exponential order of $\Delta^S(g)$ to be $\frac{1}{2}$. For general $L^{\frac{1}{2}}$ potential which is not restricted to be positive, all the reasoning which leads to Eq. (23) follows if $V(r)$ is replaced by $|V(r)|$ and $U(r) \equiv |V(r)|^{\frac{1}{2}}$. If $V(r)$ is purely repulsive, $\Delta(g)$ would correspond to the Fredholm determinant for the attractive potential with the reversed sign. If $V(r)$ has a finite number of nodes corresponding to alternating regions of attraction and repulsion, eventually becoming (say) positive, a lower bound $S(r)$ to $V(r)$ can be found in the spirit of the previous analysis in the form of a step function with a finite number of steps. The zeros of the corresponding $\Delta^S(g)$ can be estimated in principle for large g. This has not been done, but one expects the eigenvalue equation for large g to involve an almost-periodic function of $(g)^{\frac{1}{2}}$, so that the eigenvalues are "relatively dense,"[9] leading to a set of zeros whose exponent of convergence is again $\frac{1}{2}$.

II. S-WAVE BOUND STATES

In the special case of S waves the kernel $\mathfrak{G}_0(r,s)$ takes the especially simple form

$$\mathfrak{G}_0(r,s) = \frac{2rs}{\pi}\int_0^\infty dk k^2 \frac{j_0(kr)j_0(ks)}{\omega_k^2} = \frac{1}{2\mu}[e^{-\mu|r-s|} - e^{-\mu(r+s)}], \tag{28}$$

[6] An entire function $f(z)$ is said to be of exponential order $\alpha \geq 0$ if α is the greatest lower bound of numbers γ such that one can find a K such that $|f(z)| \leq \exp K|z|^\gamma$. If $f(z)$ is of exponential order α its type τ is defined by $\tau = \limsup_{|z|\to\infty} |z|^{-\gamma}|\ln f(z)|$.

[7] E. C. Titchmarsh, *The Theory of Functions* (Oxford University Press, London, 1939), 2nd ed., Sec. 8.24.

[8] E. C. Titchmarsh, Ref. 7, Secs. 8.25, 8.26.

[9] See, e.g., A. S. Besicovitch, *Almost Periodic Functions* (Cambridge University Press, New York, 1932), p. 1. "A set E of real numbers is 'relatively dense' if there exists a number $l > 0$ such that any interval of length l contains at least one number of E."

and it is not difficult to obtain more accurate estimates on the behavior of $\Delta(g)$. One can show that

$$\begin{aligned} D^{>}(r_1 \cdots r_n) &= \frac{1}{(2\mu)^n} \det_{1\le i,j\le n} |e^{-\mu|r_i - r_j|} - e^{-\mu(r_i + r_j)}| \\ &\quad \times \theta(r_1 - r_2)\theta(r_2 - r_3)\cdots\theta(r_{n-1} - r_n) \\ &= \frac{1}{(2\mu)^n}(e^{-2\mu(r_1 - r_2)} - 1)(e^{-2\mu(r_2 - r_3)} - 1)\cdots \\ &\quad (e^{-2\mu(r_{n-1} - r_n)} - 1)(e^{-2\mu r_n} - 1) \\ &\quad \times \theta(r_1 - r_2)\cdots\theta(r_{n-1} - r_n) \\ &\equiv \prod_{j=1}^{n-1} G(r_j - r_{j+1})G(r_n). \end{aligned} \tag{29}$$

Then with the variable change applied in connection with Eq. (17) one finds that

$$\begin{aligned} \Delta(g) &= \sum_{n=0}^{\infty}(-g)^n \\ &\quad \times \int_0^\infty d\xi_1 \cdots \int_0^\infty d\xi_n G(\xi_1)\cdots G(\xi_n)V(\xi_n) \\ &\quad \times V(\xi_{n-1} + \xi_n)\cdots V(\xi_1 + \cdots \xi_n) \\ &\equiv \sum_{n=0}^{\infty}(-g)^n L_n(\mu). \end{aligned} \tag{30}$$

In terms of $\sigma(\alpha)$, the Laplace weight of $|V(r)|^{\frac{1}{2}}$ one finds as before

$$\begin{aligned} L_n(\mu) \le (-)^n \int_0^\infty d\alpha_1 \cdots \int_0^\infty d\alpha_{2n} |\sigma(\alpha_1)\cdots\sigma(\alpha_{2n})| \\ \times \tilde{G}(\alpha_1 + \alpha_2)\tilde{G}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)\cdots \\ \tilde{G}(\alpha_1 + \alpha_2 + \cdots \alpha_{2n}), \end{aligned} \tag{31}$$

where

$$-\tilde{G}(\alpha) = 1/\alpha(\alpha + 2\mu). \tag{32}$$

The estimates of Sec. I were based on the bound

$$|\tilde{G}(\alpha)| \le 1/\alpha^2,$$

which corresponds to inequality $|G(\xi)| \le |\xi|$, equivalent to Eq. (16). This leads as was seen in Sec. I to the conclusion that $\Delta(g)$ is of exponential order $\frac{1}{2}$ and of type[6]

$$\tau_\Delta \le U_1 \equiv \int_0^\infty d\alpha \frac{|\sigma(\alpha)|}{\alpha}. \tag{33}$$

For $\sigma(\alpha) \ge 0$, $U_1 = \int_0^\infty dr\, |V(r)|^{\frac{1}{2}}$. For the case of finite range $V(r) \ge 0$ with $\sigma(\alpha) \ge 0$ a lower bound on the type can be obtained from the inequality

$$\begin{aligned} b/(\alpha_1 + \cdots + \alpha_{2k})(\alpha_1 + \cdots \alpha_{2k} + \alpha_{2k+1}) \\ \le -\tilde{G}(\alpha_1 + \cdots \alpha_{2k}), \end{aligned} \tag{34}$$

where

$$b = \min(1, \beta/2\mu) \tag{35}$$

and $1/\beta < \infty$ is double the range of $V(r)$; i.e., β is the greatest lower bound of the support $\sigma(\alpha)$. Then from Eq. (31)

$$\begin{aligned} L_n(\mu) \ge (2\mu b)^n \int_0^\infty d\alpha_1 \int_0^\infty d\alpha_{2n} \\ \times \frac{\sigma(\alpha)\cdots\sigma(\alpha_{2n})}{(\alpha_1 + \alpha_2)\cdots(\alpha_1 + \cdots \alpha_{2n})} = \frac{(bU_1^2)^n}{(2n)!}. \end{aligned} \tag{36}$$

This corresponds to the lower bound for the type

$$\tau_\Delta \ge b^{\frac{1}{2}}U_1 = \min[1, (\beta/2\mu)^{\frac{1}{2}}]U_1. \tag{37}$$

For $2\mu < \beta$ the type is exactly U_1.

We now show that if $V(r) \ge 0$, U_1 is in fact exactly the type for $\Delta(g)$ as a function of a possibly complex μ. $\Delta(g) \equiv \Delta(g, \mu)$ as a function of μ is analytic in the entire complex μ plane except for a cut along the negative real axis ending at $\mu = -\frac{1}{2}\beta$, for each value of g. This can be recognized from Eqs. (30)–(32). We consider the quantity

$$\tau(\mu) \equiv \lim_{g\to\infty} \tau_g(\mu) \equiv \lim_{g\to\infty} g^{-\frac{1}{2}} \ln \Delta(g, \mu). \tag{38}$$

g may approach ∞ along any path avoiding zeros of $\Delta(g; \mu)$. In view of what has been proven, $\tau_g(\mu)$ is uniformly bounded for sufficiently large g if the zeros of $\Delta(g; \mu)$ are avoided. Equation (38) defines an analytic function of μ. It has the constant value U_1 along the segment $0 \le 2\mu \le \beta$ and hence has this value everywhere in the cut μ plane. $\tau(\mu)$ is just the aforementioned type and is constant. This is borne out by the WKB approximation.

The interest in the type stems from the fact that it is related to the asymptotic distribution of the zeros of $\Delta(g, \mu)$. From the Hermiticity of the kernel $U_l(k, k')$ it follows that all the zeros of $\Delta(g)$ are real and in fact nonnegative. (We ignore the complications of the case where $g = 0$ is an eigenvalue.) A theorem[10] tells us that an entire function of exponential order $\frac{1}{2}$ and type τ with real positive zeros has an asymptotic distribution of zeros $n(g)$ (the number of zeros of modulus $\le |g|$) given by

$$n(g) \sim (1/\pi)\tau |g|^{\frac{1}{2}}. \tag{39}$$

By applying Jensen's theorem[11] we can find a bound on the number of bound states, i.e., zeros of $\Delta(g, 0)$ determined by the relation

$$\begin{aligned} n(\delta g) \ln \frac{1}{\delta} \le \int_0^g dg' \frac{n(g')}{g'} = \frac{1}{2\pi}\int_0^{2\pi} \ln |\Delta(ge^{i\theta}, 0)| \\ \le |g|^{\frac{1}{2}}\tau. \end{aligned}$$

[10] Reference 7, Sec. 8.64.

[11] See, e.g., E. Hille, *Analytic Function Theory* (Ginn and Company, Boston, Massachusetts, 1962). Vol. II, Theorem 14.1.4, p. 189.

Optimizing δ with the value $\delta = e^{-2}$, we find the inequality for all g

$$n_B \leq \tfrac{1}{2}e\,|g|^{\frac{1}{2}}\int_0^\infty dr\,|V(r)|^{\frac{1}{2}}. \tag{40}$$

The quantity in Eq. (40) is a bound on the number of zeros with modulus $\leq |g|$ of $\Delta(g, 0)$ (at zero energy). Since each bound state in a potential $gV(r)$ passed zero energy at a smaller value of the coupling constant, the number of bound states is just this number of zeros. [For potentials such as the Coulomb potential, bound states exist for arbitrarily weak coupling. However, potentials which are $L^{\frac{1}{2}}$ also satisfy the Bargmann condition $\int dr\,|rV(r)| < \infty$ and do not bind at "zero energy."] The Bargmann inequality[12]

$$n_B \leq \frac{|g|}{2l+1}\int_0^\infty dr r\,|V(r)| \tag{41}$$

or the L^2 bound[13]

$$n_B \leq \frac{|g|^2}{4\pi^2}\int d^3r\int d^3r'\,\frac{|V(r)V(r')|}{|r-r'|} \tag{42}$$

both fail to give the correct growth of the number of bound states with g for large g, and greatly overestimate this quantity. The result Eq. (40) is "best possible" as far as the g-dependent goes, and a criterion of this type for the first bound state was found by Calogero.[3] For high l values the inequality Eq. (41) may give sharper bounds. The integral $\int dr\,|rV(r)|$ converges ever so slightly better than $\int dr\,|V(r)|^{\frac{1}{2}}$ for borderline potentials, such as those which behave for large r like $(r\,lnr)^{-2}$. Potentials which are not $L^{\frac{1}{2}}$ will be discussed in a subsequent article.

Another interesting by-product of these results is the existence of the "square-root kernel" of the integral equation (6). $U_l(k, k')$ as a real symmetric kernel can be expressed in terms of the orthonormal eigenfunctions $\phi_s(k)$ and the positive eigenvalues g_s of Eq. (6) in the form

$$U_l(k, k') = \sum_s \frac{\overline{\phi_s(k')}\phi_s(k)}{g_s}. \tag{43}$$

One can define

$$W(k, k') \equiv \sum_s \frac{\overline{\phi_s(k')}\phi_s(k)}{g_s^{\frac{1}{2}}}, \tag{44}$$

which exists as a square integrable kernel since

$$\sum \frac{1}{g_s} < \infty$$

according to Eq. (39). Moreover, $W(k, k')$ is the square root of $U(k, k')$ in the sense that

$$\int dk'' W(k, k'')W'k'', k') = U(k, k') \tag{45}$$

as is evident from Eq. (44).

It was established that for g large and negative

$$\Delta(g, \mu) = \sum_{n=0}^\infty (-g)^n L_n(\mu) \sim \exp\,[\tau(-g)^{\frac{1}{2}} + o(g^{\frac{1}{2}})]$$
$$= \exp\,[\tau(e^{-i\pi}g)^{\frac{1}{2}} + o(g^{\frac{1}{2}})]. \tag{46}$$

Since all the zeros of $\Delta(g)$ are real and positive, the growth described in Eq. (46) for large $|g|$ is valid[10] for all $0 < \arg g < 2\pi$. It is not difficult to deduce from the asymptotic distribution of the (real positive) zeros of $\Delta(g, \mu)$ as given by Eq. (39), that $\Delta(g, \mu)$ is bounded for positive g, and as a real function of exponential order $\frac{1}{2}$, its behavior for large positive g must be of the form

$$\Delta(g, \mu) \sim p(g, \mu)\sin\,[\tau g^{\frac{1}{2}} + \phi(g, \mu)], \tag{47}$$

where

$$\ln p(g, \mu) = o(g^{\frac{1}{2}}), \quad \phi(g, \mu) = o(g^{\frac{1}{2}}) \tag{48}$$

with $\tau = \int dr\,|V(r)|^{\frac{1}{2}}$ for $V(r)$ nonnegative, and independent of μ. This agrees with the extrapolation from Eq. (46)

III. SCATTERING STATES

The corresponding estimates can be carried out for the scattering problem. The Fredholm determinant in this case can be obtained by reasoning similar to that in Sec. I. The expression for it is, however, well known[4] in a form similar to Eq. (13):

$$\bar{\Delta}(g, k) \equiv \bar{\Delta}(g) = \sum_{n=0}^\infty \frac{(-g)^n}{n!}$$
$$\times \int_0^\infty dr_1 \cdots \int_0^\infty dr_n V(r_1)\cdots V(r_n)\bar{D}_l(r_1 \cdots r_n), \tag{49}$$

where

$$\bar{D}_l(r_1 \cdots r_n) = \det_{1\leq i,j\leq n} |\bar{\mathfrak{G}}_l(r_i, r_j: k)|, \tag{50}$$

$$\mathfrak{G}_l(r, r': k) \equiv -ikr_< r_> j_l(kr_<)h_l^{(+)}(kr_>), \tag{51}$$

where $j_l(kr)$, $h_l^{(+)}(kr)$ are familiar spherical Bessel functions whose significant boundary values are

$$j_l(0) = 0, \quad j_l(kr) \underset{r\to\infty}{\longrightarrow} \sin\,[kr - l(\tfrac{1}{2}\pi)]/kr;$$
$$h_l^{(+)}(kr) \underset{r\to\infty}{\longrightarrow} e^{[-ikr - l(\frac{1}{2}\pi)]}/kr.$$

From symmetry, following notations used earlier,

$$\bar{\Delta}(g) = \sum_{n=0}^\infty (-g)^n \int_0^\infty dr_1 \cdots \int_0^\infty dr_n V(r_1)\cdots V(r_n)$$
$$\times\, \bar{D}_l^>(r_1 \cdots r_n); \tag{52}$$

[12] V. Bargmann, Proc. Natl. Acad. Sci. U.S. **38**, 961 (1952).
[13] J. Schwinger, Proc. Natl. Acad. Sci. U.S. **47**, 122 (1961); C. G. Ghirardi and A. Rimini, J. Math. Phys. **6**, 40 (1965).

also

$$\bar{D}_l^{>}(r_1 \cdots r_n) = \bar{H}_l(r_1, r_2)\bar{H}_l(r_2, r_3) \cdots \bar{H}_l(r_{n-1}, r_n)\bar{H}_l(r_n, 0),$$

where

$$\bar{H}_l(r, s) = -ikr^2[h_l^{(+)}(kr)/h_l^{(+)}(ks)] \times |h_l^{(+)}(ks)j_l(kr) - h_l^{(+)}(kr)j_l(ks)|. \quad (53)$$

Once again

$$|\bar{H}_l(r, s)/(r - s)| \leq 1,$$

so that the very same estimate is obtained as in Eq. (23).

$$|\bar{\Delta}(g, k)| \leq \cosh(|g|^{\frac{1}{2}} U_1). \quad (54)$$

The similarity of Eqs. (23) and (54) is more than coincidental. It is of interest to note the relation between $\bar{\Delta}(g, k)$ of Eq. (49) and $\Delta(g, \mu) \equiv \Delta(g)$ of Eq. (13). Comparison of $\mathfrak{G}_l(r, r' : \mu)$ in Eq. (11) and $\bar{\mathfrak{G}}_l(r, r' : k)$ of Eq. (50) shows that

$$\bar{\mathfrak{G}}_l(r, r' : k) = \mathfrak{G}_l(r, r' : -ik). \quad (55)$$

Then from Eqs. (12), (13), (48), (49)

$$\bar{\Delta}(g, k) = \Delta(g, -ik). \quad (56)$$

From

$$\bar{\tau}(k) \equiv \lim_{g \to \infty} g^{-\frac{1}{2}} \ln \bar{\Delta}(g, k) \quad (57)$$

follows by considerations presented earlier that

$$\bar{\tau}(k) = \tau(-ik) \equiv \tau, \quad (58)$$

For g positive, it is presumed that Eq. (47) can be analytically continued to imaginary μ. Then one would write

$$\Delta(g, ik) \sim p(g, ik) \sin[\tau g^{\frac{1}{2}} + \phi(g, ik)] \quad (59)$$

with Eq. (48) continuing to hold. The functions $p(g, ik)$ $\phi(g, ik)$ will generally be complex. The phase shift $\delta(k, g)$, which is merely the phase of the Jost function $f(k, g) = \bar{\Delta}(g, -k) = \Delta(g, ik)$ is therefore given by

$$\delta(k, g) = \operatorname{Im} \ln \Delta(g, ik). \quad (60)$$

To order $g^{\frac{1}{2}}$, the leading behavior of $\delta(g, k)$ can be shown with the help of Eqs. (60), (59), and (48) to be

$$\delta(k, g) \sim -\tau g^{\frac{1}{2}} \quad (61)$$

independently of k for $k \neq 0$. For strongly repulsive potentials [if one analytically continues in Eq. (46) from positive μ to pure imaginary values as discussed], one obtains no phase shift to order $g^{\frac{1}{2}}$ but rather an enhancement factor[14] $h(g, k)$ whose leading behavior is

$$h(g, k) \sim e^{-\frac{1}{2}\tau g^{\frac{1}{2}}}. \quad (62)$$

Equation (61) is related to Levinson's theorem and is also derivable from it.

[14] M. Goldberger and K. M. Watson, *Collision Theory* (John Wiley & Sons, Inc., New York, 1964), Sec. 6.5.

IV. DISCUSSION

In this article only the leading asymptotic behavior in g is discussed. Subject to the assumptions that $|V(r)|^{\frac{1}{2}}$ is local, $L^{\frac{1}{2}}$, and expressible as a Laplace transform it was shown that the Jost function is of exponential order $\frac{1}{2}$ in g, which implies that the number of bound states and the phase shift for an attractive potential both grow as $|g|^{\frac{1}{2}}$ for large g. If $V(r)$ is singular in the sense of not being $L^{\frac{1}{2}}$, then the exponential order of the Jost function is in fact not $\frac{1}{2}$. These cases will be dealt with separately.

Many of the results of this article can also be obtained by WKB methods which also permit means of calculating further terms in the asymptotic expansion of the Jost function for large g. Such a treatment will appear separately. The WKB method, however, does not of itself permit analytic continuation in g, unless independent knowledge of analytic properties is available. A criterion for the "largeness" of g is set by the condition for validity of the WKB method, namely the slow change of the potential over the de Broglie wavelength in the region of interaction, i.e.,

$$\frac{\partial \lambda}{\partial r} = \frac{\partial}{\partial r}\left[\frac{1}{(k^2) + gV}\right]^{\frac{1}{2}} \approx \frac{\partial}{\partial r}\frac{1}{(gV)^{\frac{1}{2}}} \ll 1. \quad (63)$$

The noncommutativity of the limits $g \to \infty$, $k \to \infty$ should be noted. The present results on exponential order in g are valid for fixed k (in an upper half-plane), and are not affected by a consequent growth of k. If g is fixed and k grows asymptotically large, $\tilde{G}(\alpha) \to 1/\alpha$ and a function of exponential order unity in g is obtained. It is clear from the criterion Eq. (63) that the largeness of g is determined relative to k, and the criterion is not uniform in k. The potential, no matter how strong, must eventually go to zero and yield to the asymptotic behavior modulated by k^2, and its largeness is therefore never uniform in space.

It is seen that, in the strong coupling limit of potential theory, one can effectively neglect the total energy k^2 in comparison with the potential energy in the first approximation. Such a consideration may not carry over to field theories, where the particle acquires a self mass which is large if the coupling is large. It would seem to be a subtler problem to determine the dominating terms in the interaction between renormalized particles.

ACKNOWLEDGMENTS

The author would like to acknowledge a helpful discussion with Dr. P. Martin, and the hospitality of the Weizmann Institute.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Variational Principles and Weighted Averages

RUBIN GOLDSTEIN
Brookhaven National Laboratory, Upton, New York

(Received 4 April 1966)

An examination of two functionals, which are in common use for making variational estimates of weighted averages, reveals that one may be preferred over the other in certain cases. In particular, for a positive-definite self-adjoint operator, the normalization-independent functional always yields the better approximation to the stationary value.

VARIATIONAL principles are particularly convenient for approximating weighted averages. If the solution to a system of equations is unknown, but the quantity of interest is some weighted average of the unknown solution, then rendering an appropriate functional stationary enables one to calculate the quantity of interest to a degree of accuracy which is higher than that of the solution itself.

Two functionals which are currently in common use in variational analyses are

$$F_1 = (s^\dagger, \psi_t) + (\phi_t, s) - (\phi_t, H\psi_t), \tag{1}$$

$$F_2 = (s^\dagger, \psi_t)(\phi_t, s)/(\phi_t, H\psi_t). \tag{2}$$

H is some operator, s and $s^\dagger$ are known ("source") functions, and ψ_t and ϕ_t are unknown ("trial") functions. The scalar product notation is used to denote integration over all the variables of interest.

$$(f, g) = \int f(x)g(x)\,dx. \tag{3}$$

Recently, these functionals have been generalized in one sense or another.[1–3] (For earlier work, the reader is referred to the bibliography of Ref. 1 or 2.) We concentrate on the simpler forms given by Eqs. (1) and (2), with the aim of pointing out that in certain situations one functional may be preferred over the other. Frequently, one or the other of the functionals has been used as a matter of convenience, but it appears that there may be a calculational advantage in using one over the other, depending on the particular situation.

Consider the functional F_1. Assume that the exact solutions ψ and ϕ to some system of equations are known only to some accuracy $\delta\psi$ and $\delta\phi$, respectively. Using the trial functions $\psi_t = \psi + \delta\psi$ and $\phi_t = \phi + \delta\phi$ in Eq. (1), we obtain

$$F_1 = (s^\dagger, \psi) + (\delta\phi, s - H\psi) + (s^\dagger - H^\dagger\phi, \delta\psi) - (\delta\phi, H\delta\psi), \tag{4}$$

where the adjoint operator $H^\dagger$ is defined by the relation,

$$(f, Hg) = (H^\dagger f, g), \tag{5}$$

for all functions f and g. Implicit in this definition are appropriate boundary conditions.

If we require that the first variations of the functional always vanish, then ψ and ϕ must be solutions to

$$H\psi = s, \tag{6}$$

$$H^\dagger\phi = s^\dagger. \tag{7}$$

The functional F_1 is thus stationary about the solutions to Eqs. (6) and (7). The boundary conditions that are to be associated with Eq. (7) are to be adjoint to the boundary conditions that are associated with Eq. (6), in order for the equality (5) to hold.

The stationary value of the functional, which we call I, is given by

$$I = (s^\dagger, \psi). \tag{8}$$

The weighting function in the quantity of interest, I, is to be chosen, therefore, as the source in the adjoint equation, (7). For example, if the quantity of interest is a resonance integral, $s^\dagger$ is taken as the resonance cross section.

Applying Eqs. (6)–(8) to Eq. (4), the trial functional F_1 may be written as

$$F_1 = I\{1 - [(\delta\phi, H\delta\psi)/(\phi, H\psi)]\}. \tag{9}$$

Having required that the first variations of the functional vanish, we see that the stationary value is estimated to second order, though the solutions to the equations of interest are assumed to be known only to first order. The variational principle yields an accurate estimate of I, however, only if the second-order terms are small. In other words, we still have to make fairly good guesses for the trial functions (based on physical considerations, usually) in order to have a good approximation to I.

Consider now the functional F_2. This functional is also stationary about the solutions to Eqs. (6) and

[1] M. D. Kostin and H. Brooks, J. Math. Phys. **5**, 1691 (1964).
[2] G. C. Pomraning, J. Soc. Indust. Appl. Math. **13**, 511 (1965).
[3] D. S. Selengut, Trans. Am. Nucl. Soc. **8**, 485 (1965).

(7), and its stationary value is given by (8). An analysis, similar to that given for F_1, gives for the trial functional F_2,

$$F_2 = I\left(1 - \frac{(\delta\phi, H\delta\psi)}{(\phi, H\psi)} + \frac{(\delta\phi, s)(s^\dagger, \delta\psi)}{(\phi, s)(s^\dagger, \psi)}\right) + O(\delta^3), \tag{10}$$

where the expression $O(\delta^3)$ represents terms of third order (and higher) in $\delta\psi$ and $\delta\phi$, which we neglect in the remaining discussion.

Clearly, the second-order terms in Eqs. (9) and (10) are, in general, different. It may be that, for a given operator and for a particular class of trial functions, the second-order terms of one functional are smaller in magnitude than those of the other. This would make one of the functionals preferred over the other for this class of trial functions. It appears, therefore, that a careful investigation of these second-order terms may be important and useful when one is calculating weighted averages.

These arguments are not limited entirely to weighted averages. They may be extended to a class of eigenvalue problems by use of the following artifice[3]: Suppose that we are interested in finding the eigenvalues to the equation

$$(L - \lambda M)\psi = 0. \tag{11}$$

Consider the auxiliary problems given by Eqs. (6) and (7), but take $H = L$ and choose the sources so as to represent the operator M by taking

$$M = s(s^\dagger, \quad). \tag{12}$$

This is possible, for example, for integral operators with separable kernels. The functional F_2 now becomes

$$F_2 = (\phi_t, M\psi_t)/(\phi_t, L\psi_t), \tag{13}$$

which is stationary about the inverse eigenvalue, $1/\lambda$, to Eq. (11).

Qualitatively, it would appear that, if the two second-order terms of Eq. (10) are of comparable magnitude and the same sign, then, for the same trial functions, F_2 gives a better estimate of I than does F_1. It is difficult to make quantitative statements about completely general operators H, so we consider a special case.

Consider the case when H is a positive-definite self-adjoint operator; $H^\dagger = H$ and $(f, Hf) > 0$. (The arguments are equally valid for negative-definite operators.) We choose the sources $s^\dagger = s$, so that from Eqs. (6) and (7), we have that $\psi = \phi$. For this case, (9) and (10) become

$$F_1 = I\{1 - [(\delta\psi, H\delta\psi)/(\psi, H\psi)]\}, \tag{14}$$

$$F_2 = I\left(1 - \frac{(\delta\psi, H\delta\psi)}{(\psi, H\psi)} + \frac{(s, \delta\psi)^2}{(s, \psi)^2}\right). \tag{15}$$

Now, the signs of the second-order terms are definite. Because of the definiteness of the operator (positive or negative), the term involving H is positive, while the last term in Eq. (15), being a perfect square (everything is assumed real here), is also positive. Thus we have the following inequalities:

$$F_1 < I \quad \text{and} \quad F_1 < F_2. \tag{16}$$

In other words, the functional F_1 gives us a maximum principle bounded by I. But it is still not clear whether F_2 is greater than or less than I. In either case, however, since the two second-order terms in (15) have opposite signs in front of them, if they are of comparable magnitude, F_2 gives a better estimate of I than F_1.

It is possible to show directly that the last term in Eq. (15) is always smaller than the other second-order term, so that, in fact, F_2 also generates a maximum principle. This means that $F_1 \leq F_2 \leq I$.

We show this, however, in an indirect way. The functional F_2 has the feature of being normalization independent; that is to say, substituting $A\psi_t$ as the trial function does not affect the value of the functional. On the other hand, the functional F_1 is affected by this substitution. Since for all A, $F_1(A)$ is bounded from above by I, let us choose that A which maximizes F_1. The solution to $\partial F_1(A)/\partial A = 0$, is $A_0 = (s, \psi_t)/(\psi_t, H\psi_t)$, and the maximum value of F_1 turns out to be F_2; in other words, $F_2 = F_1(A_0) < I$, which was to be proved. The fact that the functional F_2 can be derived from F_1 is not new,[1] but the approach used here reveals its importance in the calculation of weighted averages.

As an illustrative example of this case, consider the diffusion equation in a multiplying subcritical assembly of a one-dimensional slab of width $2a$.

$$(-D\nabla^2 + \Sigma_a - \nu\Sigma_f)\psi = S, \tag{17}$$

where D is the diffusion coefficient, Σ_a, the absorption cross section, is greater than $\nu\Sigma_f$, the multiplication, and S is some external source. We choose the homogeneous boundary conditions

$$\psi(\pm a) = 0. \tag{18}$$

Letting $(\Sigma_a - \nu\Sigma_f)/D = \kappa^2$ and $S/(\Sigma_a - \nu\Sigma_f) = s$, where both κ^2 and s are positive quantities, Eq. (17) becomes $[1 - (1/\kappa^2)\nabla^2]\psi = s$, or

$$H\psi = s, \quad \text{where} \quad H = 1 - (1/\kappa^2)\nabla^2. \tag{19}$$

Note that this H is a positive-definite self-adjoint operator. A reasonable trial function which satisfies the boundary conditions for this problem is

$$\psi_t = \cos(\pi/2a)x. \tag{20}$$

If the source s is constant in space, then the weighted average (s,ψ) is related to the volume averaged flux, $\bar{\psi}$. If the source is a point source, then the weighted average is the flux at the source point. For the case of constant s, the variational approximations to the average flux with the trial function (20) are plotted in Fig. 1 together with the exact solution to the problem. If one changes the amplitude of the trial function, then F_2 remains unchanged, while the curve for F_1 moves up or down but never becomes greater than F_2. The preferred functional is thus F_2.

The above arguments fail for the more general non-self-adjoint case, because the functionals are not maximized. It is possible for F_1 or F_2 to be larger or smaller than I, depending on the magnitude and sign of the second-order terms. For this case, however, it may be possible to show for a given operator and a given class of trial functions that the second-order terms of one functional are smaller than the other. Therefore, an investigation of these second-order terms would seem appropriate, for it may reveal a functional preference in a given class of problems. Furthermore, it may be possible to alter the preference between the functionals in the non-self-adjoint case by altering the amplitudes of the trial functions.

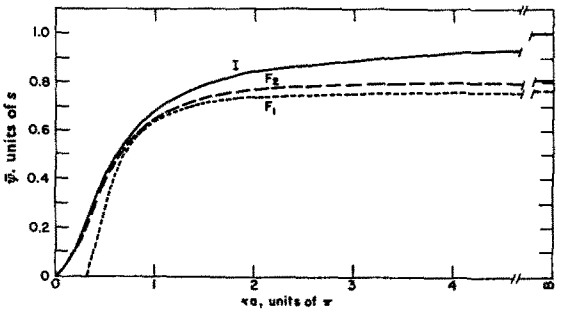


FIG. 1. Variational estimates F_1 and F_2 and the exact solution I for the average flux $\bar{\psi}$ as a function of the slab half-width a.

ACKNOWLEDGMENT

This work was performed under the auspices of the United States Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Calculation of Intersect Distribution Functions in Small Angle X-Ray Scattering

PAUL W. SCHMIDT
Physics Department, University of Missouri, Columbia, Missouri

(Received 23 May 1966)

As has been recently suggested, many subjects in small angle x-ray scattering theory can be discussed by using a function called the intersect distribution function $G(M)$, which gives an average value of the distribution of lines with length M which pass through a point in a particle and which also have both ends lying on the boundary of the particle. Some properties of the intersect distribution function for a plane lamina with a convex boundary are investigated. The calculation is found to require the use of a weighting factor which is expressible in terms of the function generating the boundary of the lamina. The relation between $G(M)$ and the two-dimensional characteristic function is given. The exact intersect distribution function is found for a circle, and an approximate calculation of $G(M)$ is carried out for small M for an arbitrary plane lamina with a convex boundary.

I. INTRODUCTION

IT has recently been pointed out by Porod[1] that many topics in small angle x-ray scattering theory can conveniently be discussed in terms of a function called the intersect distribution function, which up to this time has been rarely used, even though an essentially equivalent function,[2] the line distribution function, was introduced a number of years ago.[3]

The name "intersect" has been applied by Porod to

[1] G. Porod, in *Proceedings of the Small Angle X-Ray Scattering Conference* (Gordon and Breach Science Publishers, Inc., New York, 1967).

[2] A. Guinier, G. Fournet, C. B. Walker, and K. L. Yudowitch, *Small Angle Scattering of X-Rays* (John Wiley & Sons, Inc., New York, 1955), pp. 12–13.

[3] The intersect distribution function $G(M)$ defined below is essentially equivalent to Porod's intersect distribution function. The line distribution function $g(M)$ of Ref. 2 can be obtained from $G(M)$ by the relation $g(M) = (M/\bar{M})G(M)$. The normalizing constant $\bar{M}$ is defined below in Eq. (4).

If the source s is constant in space, then the weighted average (s,ψ) is related to the volume averaged flux, $\bar{\psi}$. If the source is a point source, then the weighted average is the flux at the source point. For the case of constant s, the variational approximations to the average flux with the trial function (20) are plotted in Fig. 1 together with the exact solution to the problem. If one changes the amplitude of the trial function, then F_2 remains unchanged, while the curve for F_1 moves up or down but never becomes greater than F_2. The preferred functional is thus F_2.

The above arguments fail for the more general non-self-adjoint case, because the functionals are not maximized. It is possible for F_1 or F_2 to be larger or smaller than I, depending on the magnitude and sign of the second-order terms. For this case, however, it may be possible to show for a given operator and a given class of trial functions that the second-order terms of one functional are smaller than the other. Therefore, an investigation of these second-order terms would seem appropriate, for it may reveal a functional preference in a given class of problems. Furthermore, it may be possible to alter the preference between the functionals in the non-self-adjoint case by altering the amplitudes of the trial functions.

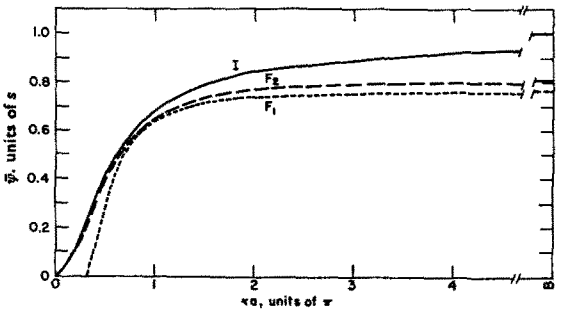


FIG. 1. Variational estimates F_1 and F_2 and the exact solution I for the average flux $\bar{\psi}$ as a function of the slab half-width a.

ACKNOWLEDGMENT

This work was performed under the auspices of the United States Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Calculation of Intersect Distribution Functions in Small Angle X-Ray Scattering

PAUL W. SCHMIDT
Physics Department, University of Missouri, Columbia, Missouri

(Received 23 May 1966)

As has been recently suggested, many subjects in small angle x-ray scattering theory can be discussed by using a function called the intersect distribution function $G(M)$, which gives an average value of the distribution of lines with length M which pass through a point in a particle and which also have both ends lying on the boundary of the particle. Some properties of the intersect distribution function for a plane lamina with a convex boundary are investigated. The calculation is found to require the use of a weighting factor which is expressible in terms of the function generating the boundary of the lamina. The relation between $G(M)$ and the two-dimensional characteristic function is given. The exact intersect distribution function is found for a circle, and an approximate calculation of $G(M)$ is carried out for small M for an arbitrary plane lamina with a convex boundary.

I. INTRODUCTION

IT has recently been pointed out by Porod[1] that many topics in small angle x-ray scattering theory can conveniently be discussed in terms of a function called the intersect distribution function, which up to this time has been rarely used, even though an essentially equivalent function,[2] the line distribution function, was introduced a number of years ago.[3]

The name "intersect" has been applied by Porod to

[1] G. Porod, in *Proceedings of the Small Angle X-Ray Scattering Conference* (Gordon and Breach Science Publishers, Inc., New York, 1967).

[2] A. Guinier, G. Fournet, C. B. Walker, and K. L. Yudowitch, *Small Angle Scattering of X-Rays* (John Wiley & Sons, Inc., New York, 1955), pp. 12–13.

[3] The intersect distribution function $G(M)$ defined below is essentially equivalent to Porod's intersect distribution function. The line distribution function $g(M)$ of Ref. 2 can be obtained from $G(M)$ by the relation $g(M) = (M/\bar{M})G(M)$. The normalizing constant $\bar{M}$ is defined below in Eq. (4).

any line passing through a body and with both ends lying on the boundary of the body. The intersect distribution function $G(M)$ is an average distribution of lines of length M passing through a point and with both ends lying on the particle surface.

Because of the renewed interest in the intersect distribution function we have recently investigated some properties of this function. In these studies, approximate values of the function were calculated for a plane lamina with a convex boundary. In the calculation, a weighted average had to be computed. Since the form of this weighting function was not apparent from other discussions of intersect distribution functions, and since the weighting function can affect other calculations of the intersect distribution function and its properties, the results for the plane lamina are outlined below.

For simplicity, only the plane lamina is considered, although analogous results would be expected for convex three-dimensional bodies.

II. THE WEIGHTING FUNCTION $P_1(\mathbf{p}, \theta)$

For a plane lamina with a convex boundary and uniform electron density, let $\beta_0(r)$ be the two-dimensional characteristic function, which is the analog of the three-dimensional characteristic function[4] $\gamma_0(r)$. By analogy with the three-dimensional case,[4] for a plane lamina

$$A = \int_0^D 2\pi r \beta_0(r)\, dr = A^{-1} \int_A dA \int_0^{2\pi} d\theta \int_0^{R(\mathbf{p},\theta)} r\, dr, \tag{1}$$

where A is the area of the lamina; D, the maximum diameter, is the length of the longest straight line that can be contained in the lamina; and where $\mathbf{p}$ is a vector from a fixed origin to the area element dA. The surface integration extends over the area of the lamina. The vector $\mathbf{p}$ defines a point which is called point $\mathbf{p}$ and which is taken as the origin of a polar coordinate system with coordinates r and θ. The boundary of the lamina is given by the equation $r = R(\mathbf{p}, \theta)$. Equation (1) can be written as

$$A = (2A)^{-1} \int_A dA \int_0^{2\pi} [R(\mathbf{p}, \theta)]^2\, d\theta. \tag{2}$$

Passing through each point $\mathbf{p}$ there is a line with length M, the ends of which lie on the lamina boundary and which is oriented at an angle θ with respect to the reference axis of the polar coordinate system. This line is called an intersect.

Equation (2) can be expressed as

$$A = (2A)^{-1} \int_A dA \int_0^{\pi} d\theta M^2 P_1(\mathbf{p}, \theta), \tag{3}$$

where

$$P_1(\mathbf{p}, \theta) = M^{-2}\{[M - R(\mathbf{p}, \theta)]^2 + [R(\mathbf{p}, \theta)]^2\}.$$

When M replaces θ as a variable of integration, (3) can be put in the form

$$A = \pi/(3\bar{M}) \int_0^D dM M^3 G(M), \tag{4}$$

where

$$G(M) = [3\bar{M}/(2\pi MA)] \sum_{i=1}^{j} \int_A dA\, |d\theta/dM|_i\, P(\mathbf{p}, M), \tag{5}$$

$$\bar{M} = \int_0^D dM M G(M).$$

The weighting factor $P(\mathbf{p}, M)$ is obtained by expressing θ in terms of $\mathbf{p}$ and M in $P_1(\mathbf{p}, \theta)$. In the change of variables from θ to M, the interval of integration must be divided into j subintervals in each of which θ is a single-valued function of M. Therefore, as (5) indicates, $G(M)$ must be expressed as the sum of j integrals with $|d\theta/dM|_i$ being the value of $|d\theta/dM|$ valid in interval i. In (5) the value of M determines the part of the lamina area over which the area integration extends.

One of the main purposes of this note is to emphasize the need for the use of weighting factor $P(\mathbf{p}, M)$ in calculating the intersect distribution $G(M)$ or the line distribution function $g_r(M)$ described by Guinier *et al.*[5] An approximate expression for $P(\mathbf{p}, M)$ for small M is given below for a plane lamina with a smooth convex boundary.

III. THE RELATION BETWEEN $G(M)$ AND $\beta_0(r)$

By two partial integrations, (4) can be written as

$$\pi/(3\bar{M}) \int_0^D dM M^3 G(M) = 2\pi \bar{M}^{-1} \int_0^D r\, dr \int_r^D dM(M - r) G(M).$$

Then by analogy with the three-dimensional case,[6] the two-dimensional characteristic function $\beta_0(r)$ is related to the intersect distribution function $G(M)$ by the equation

$$\beta_0(r) = \bar{M}^{-1} \int_r^D dM(M - r) G(M). \tag{6}$$

Thus

$$\beta_0''(r) = \bar{M}^{-1} G(r). \tag{7}$$

The characteristic function $\beta_0(r)$ and the line distribution function $G(M)$ therefore give very nearly equivalent information. At times, however, one function may be more convenient to use than the other.

[4] Reference 2, pp. 10–19.

[5] Reference 2, p. 12, Footnote 1.

[6] Reference 2, p. 13.

IV. THE INTERSECT DISTRIBUTION FUNCTION FOR A CIRCLE

Equation (5) requires knowledge of the relation between the angle θ and the intersect which has a length M and which passes through point **p**. Let the boundary be represented by the circle

$$x^2 + (R_0 - b - y)^2 = R_0^2, \tag{8}$$

where R_0 is the radius of the circle, and b is the distance from the circle to point **p**. The intersect can be represented by the line

$$y = x \tan \theta.$$

This line intersects the circle at two points, the x coordinates of which are given by the two roots x_1 and x_2 of the quadratic equation

$$x_2^2 \sec^2 \theta - 2x(R_0 - b) \tan \theta + (R_0 - b)^2 - R_0^2 = 0.$$

Since the two points of intersection must be separated by a distance M,

$$M \cos \theta = x_2 - x_1.$$

Therefore for $0 \leq \theta \leq \frac{1}{2}\pi$,

$$\theta = \cos^{-1}[(R_0^2 - \tfrac{1}{4}M^2)^{\frac{1}{2}}/(R_0 - b)], \tag{9}$$

and for $\frac{1}{2}\pi \leq \theta \leq \pi$,

$$\theta = \cos^{-1}[-(R_0^2 - \tfrac{1}{4}M^2)^{\frac{1}{2}}/(R_0 - b)]. \tag{10}$$

Since two expressions are necessary to specify θ as a function of M throughout the interval $0 \leq \theta \leq \pi$, in (5) $j = 2$.

The quantity $R(\mathbf{p}, \theta)$ in (2) is the distance from point **p** to the boundary. Therefore, in (8), $y = R(\mathbf{p}, \theta) \sin \theta$ and $x = R(\mathbf{p}, \theta) \cos \theta$. By substitution of these values of x and y into (8), a quadratic equation for $R(\mathbf{p}, \theta)$ is obtained with the solution

$$R(\mathbf{p}, \theta) = (R_0 - b) \sin \theta + [R_0^2 - (R_0 - b)^2 \cos^2 \theta]^{\frac{1}{2}}.$$

From (9) and (10),

$$M = 2[R_0^2 - (R_0 - b)^2 \cos^2 \theta]^{\frac{1}{2}}.$$

Thus

$$R(\mathbf{p}, \theta) = M/2 + [(R_0 - b)^2 + (M/2)^2 - R_0^2]^{\frac{1}{2}},$$
$$P(\mathbf{p}, M) = 1 + 2M^{-2}[(R_0 - b)^2 - R_0^2]. \tag{11}$$

Points on the boundary of the circle are specified by the arc length t from the point to a reference point. The boundary point with arc length t is called "point t." For the surface integration in (5), the variables b and t are employed. Then

$$dA = (1 - b/R_0)\, db\, dt.$$

The largest value $b_{\max}$ of b is determined by the condition that the values of $\cos \theta$ given by (9) and (10) must satisfy the condition that $|\cos \theta| \leq 1$. Therefore

$$b_{\max} = R_0 - (R_0^2 - \tfrac{1}{4}M^2)^{\frac{1}{2}}.$$

When the above results are substituted in (5), $G(M)$ can be expressed as

$$G(M) = \frac{3\bar{M}}{2\pi M A} \sum_{i=1}^{2} 2\pi R_0 \times \int_0^{b\max} db \left(1 - \frac{b}{R_0}\right) P(\mathbf{p}, M) \left|\frac{d\theta}{dM}\right|_i.$$

Thus

$$G(M) = \bar{M}M/\pi R_0^2(4R_0^2 - M^2)^{\frac{1}{2}}.$$

This result agrees with the intersect distribution for a circle obtained from (7) using the expression for the characteristic function for a circle.[7]

V. THE INTERSECT DISTRIBUTION FUNCTION FOR SMALL M

The results for a circular boundary can be used to find an approximation for $G(M)$ for small M for a smooth convex boundary with arbitrary shape since, when the quantity R in (8)–(11) is replaced by the radius of curvature $R(t)$ at point t, these equations hold approximately at a point t on an arbitrary smooth convex curve. Equation (5) then gives

$$G(M) = L\bar{M}(4\pi A)^{-1}\,\overline{(R^{-2})}M + \cdots, \tag{12}$$

where

$$\overline{R^{-2}} = L^{-1}\int_0^L [R(t)]^{-2}\, dt$$

and L is the total arc length of the boundary. The quantity $\overline{R^{-2}}$ thus is the average of $[R(t)]^{-2}$ over the boundary.

Since[8] $\beta_0(0) = 1$ and $\beta_0'(0) = -L/(\pi A)$, from (6) and (12)

$$\beta_0(r) = 1 - (L/\pi A)r + [L/(24\pi A)]\overline{(R^{-2})}r^3 + \cdots.$$

Kirste and Porod[9] obtained this same expression by a direct calculation of $\beta_0(r)$.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Professor G. Porod for several stimulating discussions which led to the calculations described in this paper, and to Hsin-i Wu and Professor Porod for advice during the preparation of the manuscript.

This work was supported by the National Science Foundation.

[7] G. Porod, Acta Phys. Austriaca **2**, 286 (1948).
[8] P. W. Schmidt and R. Hight, Jr., J. Appl. Phys. **30**, 870 (1959), Eq. (8).
[9] R. Kirste and G. Porod, Kolloid-Z. **184**, 4 (1962), Eq. (23).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Correlations in Ising Ferromagnets. I*

ROBERT B. GRIFFITHS
Physics Department, Carnegie Institute of Technology, Pittsburgh, Pennsylvania

(Received 26 July 1966)

The following results are proved for a system of Ising spins $\sigma_i = \pm 1$ in zero magnetic field coupled by a purely ferromagnetic interaction of the form $-\sum_{i<j} J_{ij}\sigma_i\sigma_j$ with $J_{ij} \geq 0$, for arbitrary crystal lattice and range of interaction: (1) The binary correlation functions $\langle\sigma_k\sigma_l\rangle$ are always nonnegative ($\langle\ \rangle$ denotes a thermal average). (2) For arbitrary i, j, k, and l, $\langle\sigma_i\sigma_j\sigma_k\sigma_l\rangle \geq \langle\sigma_i\sigma_j\rangle\langle\sigma_k\sigma_l\rangle$. Consequences of these results, in particular the second, are: (i) $\langle\sigma_k\sigma_l\rangle$ never decreases if any J_{ij} is increased. (ii) If an Ising model with ferromagnetic interactions exhibits a long-range order, this long-range order increases if additional ferromagnetic interactions are added. This last fact may be used to prove the existence of long-range order in a large class of two- and three-dimensional Ising lattices with purely ferromagnetic interactions of bounded or unbounded range.

I. INTRODUCTION

CONSIDER a finite system of Ising spins $\sigma_i = \pm 1$ with a Hamiltonian

$$\mathcal{H} = -\sum_{i<j}^{N} J_{ij}(\sigma_i\sigma_j - 1), \tag{1}$$

where for every pair $i \neq j$

$$0 \leq J_{ij} = J_{ji} < \infty. \tag{2}$$

That is, all interactions are ferromagnetic, favoring parallel alignment of spins. The thermal average of an operator $\mathcal{O}$ is defined by

$$\langle\mathcal{O}\rangle = \mathrm{Tr}\,[\mathcal{O}\exp(-\beta\mathcal{H})]/Z, \tag{3}$$

where

$$Z = \mathrm{Tr}\,[\exp(-\beta\mathcal{H})] \tag{4}$$

is the partition function, and the inverse temperature $\beta = (kT)^{-1}$ is always positive. As all interactions favor parallel alignment, the following result is not surprising.

Theorem 1: For the system described by (1) and (2) and any pair k, l,

$$\langle\sigma_k\sigma_l\rangle \geq 0. \tag{5}$$

Also it seems intuitively plausible that increasing the ferromagnetic interaction between any pair of spins tends to enhance the tendency of other pairs to line up parallel, a result embodied in Theorem 2.

Theorem 2: For the system described by (1) and (2), and where k, l, m, n denote any four spins (not necessarily all different), the following is true:

$$\beta^{-1}\,\partial\langle\sigma_k\sigma_l\rangle/\partial J_{mn} = \langle\sigma_k\sigma_l\sigma_m\sigma_n\rangle - \langle\sigma_k\sigma_l\rangle\langle\sigma_m\sigma_n\rangle \geq 0. \tag{6}$$

Further, the result (6) still holds when J_{kl} or J_{mn} (of both) is negative (we suppose *all other* J_{ij} are nonnegative).

Section II contains the straightforward proof of Theorem 1 together with definitions and notation useful in discussing Theorem 2. The latter is proved in Sec. III with assistance from two lemmas in Appendix A. An immediate consequence of Theorem 2, with proof in Sec. III, is found in Theorem 3.

Theorem 3: For the system described in (1) and (2), and where k, l, and n denote any three spins, the following relation holds:

$$\langle\sigma_k\sigma_n\rangle \geq \langle\sigma_k\sigma_l\rangle\langle\sigma_l\sigma_n\rangle \tag{7}$$

and it is unnecessary to assume that J_{kl} and J_{ln} are nonnegative.

Some applications of Theorems 2 and 3 to the problem of long-range order in various types of Ising ferromagnets are found in Sec. IV. We hope to present others in a future publication. The principal utility of these theorems seems to lie in applications where the results, just as the theorems themselves, are intuitively very reasonable, but formally difficult to prove. We feel the results merit publication because at the present time the statistical theory of phase transitions, in which the Ising model has played a major role, is seriously restricted by a lack of exact solutions for even relatively simple models. Various approximation methods are of much value, at least in regions removed from the critical point, but there is increasing evidence that they are not adequate to answer many questions of theoretical interest. In the absence of exact solutions (and even if they were available), precise mathematical results may be useful for gaining insight into the behavior of various

* Research supported in part by the National Science Foundation.

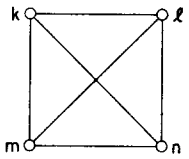


FIG. 1. Complete diagram for a system of 4 spins.

models.[1] We hope our results may make some contribution toward this end.

II. DEFINITIONS, NOTATION, AND THE PROOF OF THEOREM 1

For conceptual purposes it is convenient to represent Ising spins as small circles in a diagram connected with lines or bonds, the bond between spins k and l representing the term $-J_{kl}(\sigma_k\sigma_l - 1)$ in (1). An example with 4 spins is shown in Fig. 1. With each bond we associate a factor (Boltzmann factor)

$$X_{kl} = \exp(-2\beta J_{kl}) \tag{8}$$

representing the contribution of the bond to the partition function when $\sigma_k\sigma_l = -1$. In fact, the partition function is simply a sum of terms which are polynomials in the $\{X_{ij}\}$, with any given X_{kl} occurring to the zeroth or first power. For J_{ij} satisfying (2) we have

$$0 < X_{ij} \leq 1. \tag{9}$$

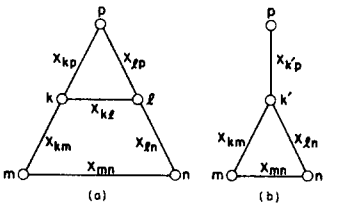
If J_{kl} vanishes, i.e., $X_{kl} = 1$, we erase the corresponding bond in the diagram. Another important operation is that of taking the limit $J_{kl} \to \infty$ or $X_{kl} \to 0$, which we call "combining" spins k and l. The effect of this operation on the partition function is easily verified: k and l may now be treated as a single spin, say k'. Further, the factors $X_{k'm}$ are simply given as products

$$X_{k'm} = X_{km}X_{lm} \tag{10}$$

(that is, $J_{k'm} = J_{km} + J_{lm}$). Note that if both X_{km} and X_{lm} satisfy (9), so does $X_{k'm}$. That is, the ferromagnetic nature of all bonds is preserved when two spins are combined. An example is shown in Fig. 2.

We use the same diagram to represent both the Hamiltonian (1) and the associated partition function (4). In connection with the latter it is convenient to introduce *restricted* partition functions in which instead of summing over all configurations, as in (4), one sums only over those in which certain spins have specified values. For example, $Z(k+)$, represented in Fig. 3(a), is a restricted partition function in which

FIG. 2. Illustration of the effect of combining spins k and l by letting X_{kl} go to zero. The result before combination is shown in (a) and the result after combination in (b). The factor $X_{k'p}$ is equal to $X_{kp}\cdot X_{lp}$.



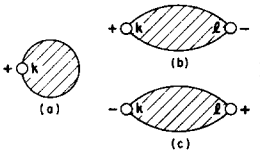


FIG. 3. Diagrams illustrating various restricted partition functions.

only configurations with $\sigma_k = +1$ are included in the sum (4). Figure 3(b) represents the restricted partition function $Z(k+\ l-)$ which, because (1) is invariant under time reversal (unaltered if each σ_i is replaced by $-\sigma_i$) is identical with $Z(k-\ l+)$ illustrated in Fig. 3(c). In these figures all other spins plus connecting bonds are, for brevity, represented by a cross-hatched region or "blob."

We add two terms to complete our notational and diagrammatic machinery. A diagram (Hamiltonian or partition function) is *complete* if every spin is joined to every other spin by a bond; that is, $J_{ij} \neq 0$ for any pair $i \neq j$. A diagram is *connected* if one can get from any spin to any other spin by passing along bonds from spin to spin.

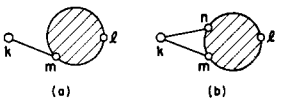
We now prove Theorem 1. In terms of restricted partition functions it suffices to show that

$$\begin{aligned}\tfrac{1}{2}Z\langle\sigma_k\sigma_l\rangle &= \tfrac{1}{2}[Z(k+\ l+) + Z(k-\ l-) - Z(k+\ l-) \\ &\quad -Z(k-\ l+)] = Q \\ &= [Z(k+\ l+) - Z(k-\ l+)] \geq 0, \end{aligned} \tag{11}$$

where we have used time reversal symmetry [e.g., $Z(k+\ l+) = Z(k-\ l-)$] to simplify the expression for Q.

For a system containing only two spins i and j, Q is simply $1 - X_{ij}$ and (11) is obviously true. Now let us proceed by induction. Suppose (11) holds for any system of N spins described by (1) and (2). Let us add one spin, k, to this system, initially connecting it by a single bond to a spin m as shown in Fig. 4(a). Q is a linear function of the factor X_{km}, so it suffices to check (11) at $X_{km} = 1$ and $X_{km} = 0$. In the former case k is disconnected from the diagram containing l, so $Z(k+\ l+) = Z(k-\ l+)$ and Q vanishes. In the

FIG. 4. Diagram illustrating proof of Theorem 1.



[1] For example, the very powerful results of T. D. Lee and C. N. Yang [Phys. Rev. 87, 410 (1952)] on the zeros of the Ising model partition function have provided information of great importance about the behavior of such models in a magnetic field, even though an exact solution to the statistical problem (in two and three dimensions) is still lacking.

latter case, $X_{km} = 0$ combines spins k and m which reduces our problem to N spins, for which Q is nonnegative by the induction hypothesis.

Suppose next that k is connected by two bonds to spins m and n as shown in Fig. 4(b). Q is linear in X_{kn}, and for $X_{kn} = 1$ the (kn) bond disappears and we have the problem considered in the preceding paragraph. But setting $X_{kn} = 0$ reduces the system to one of N spins, and thus the nonnegativity of Q is assured.

Clearly the same technique works as more and more bonds are added joining k to the original system of N spins. There is no difficulty if a bond is added directly connecting k and l. Thus the positivity of Q for all systems containing $N + 1$ spins is ensured, given its positivity for all systems of N spins, and our proof is complete.

III. PROOF OF THEOREMS 2 AND 3

Initially we assume that k, l, m, and n all denote different spins; the case where two or more are identical is considered later. We rewrite the requirement (6) in terms of restricted partition functions as follows

$$\begin{aligned} Z^2[&\langle\sigma_k\sigma_l\sigma_m\sigma_n\rangle - \langle\sigma_k\sigma_l\rangle\langle\sigma_m\sigma_n\rangle]/8 \\ &= \tfrac{1}{4}[(a+b+c+d-e-f-g-h) \\ &\quad\times(a+b+c+d+e+f+g+h) \\ &\quad-(a+b+e+f-c-d-g-h) \\ &\quad\times(a+b+g+h-c-d-e-f)] \\ &= F = (a+b)(c+d) - (e+f)(g+h) \geq 0, \end{aligned} \tag{12}$$

where

$$\begin{aligned} a &= Z(k+\,l+\,m+\,n+), \quad b = Z(k+\,l+\,m-\,n-), \\ c &= Z(k+\,l-\,m+\,n-), \quad d = Z(k+\,l-\,m-\,n+), \\ e &= Z(k+\,l+\,m+\,n-), \quad f = Z(k+\,l+\,m-\,n+), \\ g &= Z(k+\,l-\,m+\,n+), \quad h = Z(k+\,l-\,m-\,n-), \end{aligned} \tag{13}$$

and we have made free use of time-reversal invariance to replace, for example, $Z(k-\,l+\,m-\,n-)$ by g.

In order to gain insight into the algebraic structure of F, we consider first a simple example: the system illustrated in Fig. 1, a saturated diagram with four spins. Direct calculation yields

$$\begin{aligned} F = X_{kl}X_{mn}[&X_{kn}X_{lm}(1 - X_{km}^2)(1 - X_{ln}^2) \\ &+ X_{km}X_{ln}(1 - X_{kn}^2)(1 - X_{lm}^2)], \end{aligned} \tag{14}$$

a quantity obviously nonnegative for all X_{ij} between 0 and 1.

Note that F is the sum of terms with the structure gG where g, the "linear term,"[2] is a simple product of X's and linear in any particular X_{ij}. G, on the other hand, is a polynomial in which any X, if it occurs at all, appears as X^2. Such polynomials we call *quadratic terms*. A more precise definition of a *linear term* is the following: Let W be a set of distinct X's (note that X_{ij} and X_{ji} are considered equivalent) containing at least one member. The linear term g associated with W is simply the product of all X's appearing in W. No additional numerical factors are permitted. For example, $X_{12}X_{13}$ is by our definition a linear term, and $2X_{12}X_{13}$ is not. The latter is of the form gG, with $G = 2$ the "quadratic term."

Any restricted or unrestricted partition function is the sum of linear terms plus a constant (which may be zero). Thus F, the sum of products of pairs of restricted partition functions, may be written as a sum of terms each of which is either constant or the product of X's, some of which occur linearly and some quadratically. After classifying different terms according to the set W_p of linear factors, we may add up all terms with the same W_p and write the sum as g_pG_p, where g_p is the (unique) linear term associated with W_p, and G_p is a quadratic term. (Clearly it is not possible for a particular X to appear both in g_p and G_p.) Some quadratic terms occur without linear factors and we denote their sum by G_0. Thus F has the form

$$F = G_0 + \sum_{p=1} g_pG_p, \tag{15}$$

where $g_p \neq g_q$ for $p \neq q$. In general the G's will not have the simple form found in (14).

Provided the bonds kl and mn are present, F is always the product of $X_{kl}X_{mn}$ times a quantity not containing these factors, just as in the example (14). That this is true in general follows from (12) and the observation [see the definition (13)] that c, d, e, and f each contain X_{mn} to the first power while a, b, g, and h do not contain it at all. Similarly X_{kl} occurs to the first power in c, d, g, and h, and is absent from a, b, e, and f. The fact that X_{kl} and X_{mn} are simply multiplicative factors in F is the reason we do not need to require in Theorem 2 that J_{kl} and J_{mn} be nonnegative.

Consider next the example shown in Fig. 5 consisting of two disconnected diagrams A and B which,

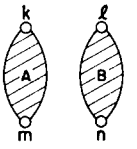


FIG. 5. Special case investigated in connection with Theorem 2.

[2] They could, more properly, be called multilinear.

apart from the fact that one contains the spins k and m and the other the spins l and n, are wholly arbitrary. Let Z_A and Z_B denote the [restricted] partition functions for these diagrams. Each term in (13) may be expressed as a suitable product; for example,

$$f = Z_A(k+ m-)Z_B(l+ n+). \tag{16}$$

Inserting these in (12) and making free use of the time reversal symmetry [$Z_A(k+ m+) = Z_A(k- m-)$, etc.] we have

$$F = [Z_A(k+ m+)^2 - Z_A(k+ m-)^2] \times [Z_B(l+ n+)^2 - Z_B(l+ n-)^2], \tag{17}$$

which is nonnegative by Theorem 1 [see (11)]. We later need the following result:

Lemma 1: If systems A and B in Fig. 5 are both connected systems and the corresponding F given by (17) is decomposed in the form (15), then G_o is nonnegative.

Let F^A and F^B denote the first and second factors on the right-hand side of (17). In analogy with (15) let us decompose F^A as

$$F^A = G_o^A + \sum_{p=1} g_p^A G_p^A \tag{18}$$

and F^B in similar fashion. Since none of the X's appearing in F^A appear in F^B and vice versa, G_o is simply the product $G_o^A G_o^B$. We need only prove that G_o^A is positive—the same proof suffices for G_o^B—in order to prove Lemma 1.

We may write (see Appendix A)

$$Z_A(k+ m+) = \sum_p r_p; \quad Z_B(k+ m-) = \sum_q t_q, \tag{19}$$

where each r_p is either 1 or a linear term and the same is true of the t's. By Lemma A1 of Appendix A, for $p \neq p'$, r_p does not contain the same factors as $r_{p'}$, so that $r_p r_{p'}$ will always contain a linear term and can make no contribution to G_o^A. The same holds for the t's. We conclude that

$$G_o^A = \sum_p r_p^2 - \sum_q t_q^2 = \tilde{Z}_A(k+ m+) - \tilde{Z}_A(k+ m-), \tag{20}$$

where by $\tilde{Z}_A$ we mean the (restricted) partition function for a system $\tilde{A}$ obtained from A by multiplying by 2 every J_{ij} which occurs in A. The result of this process is to replace every X_{ij} by X_{ij}^2. Since $\tilde{A}$ contains only ferromagnetic bonds, Theorem 1 applies and, by (11), the expression (20) must be nonnegative. This completes the proof of Lemma 1.

We now prove Theorem 2 for a complete system containing N spins, assuming that k, l, m, and n are distinct spins. The decomposition (15) for F lacks the term G_o, since, as noted above, F contains the linear factors $X_{kl}X_{mn}$. Choose a particular p, say $p = 2$, and set all the factors in W_2 (the set of X's in g_2) equal to 1 everywhere in the expression for F. The result, F', corresponds to a diagram in which every bond corresponding to some X in W_2 has been erased. This diagram, according to Lemma A2 of Appendix A, consists of two disconnected pieces, A and B, each of which is complete. Since X_{kl} belongs to W_2, it is evident that spins k and l cannot both belong to system A or both to system B, for then one or the other of these systems would be incomplete The same holds for spins m and n. Several possibilities remain; without loss of generality we may assume the one shown in Fig. 5.

Of course, F' may be decomposed in the form (15) as

$$F' = G_o' + \sum_p g_p' G_p'. \tag{21}$$

We now assert that G_o' and G_2 are identical. It is clear that setting all the X's in W_2 equal to one does not alter G_2, and thus G_2 is a quadratic term appearing in F' with no linear term as a factor. However, for $p \neq 2$, g_p' (that is, the term obtained from g_p by setting all X's in W_2 equal to one) contains at least one of the X_{ij}. This follows from part (ii) of Lemma A2 in Appendix A. Thus, in fact, G_2 is the *only* quadratic term appearing in F' without a linear term as a factor and must be identical with G_o'. But the latter is nonnegative by Lemma 1 above.

A similar argument works for any G_p in the decomposition (15) of F for a complete system. But if every G_p is nonnegative, so is F, which completes our proof. The same result holds for an incomplete system, since we need only take the limit of setting certain X's equal to 1, and F is a continuous function of the X's.

We next consider the case where spins k, l, m, and n are not all distinct. If k and l are the same, $\sigma_k\sigma_l$ becomes $\sigma_k^2 = 1$ and (6) simply vanishes. The case where m and n are identical is similarly uninteresting. The case where $l = m$ can be considered by taking the limit $X_{lm} \to 0$, that is, by combining the spins. In this case (6) becomes

$$\langle\sigma_k\sigma_n\rangle - \langle\sigma_k\sigma_l\rangle\langle\sigma_l\sigma_n\rangle \geq 0 \tag{22}$$

or, in other words, we have proved Theorem 3.

IV. APPLICATION: LONG-RANGE ORDER IN ISING FERROMAGNETS

The phase transition which occurs as the temperature is lowered in zero magnetic field for an Ising ferromagnet on a square lattice with nearest-neighbor interactions results in (among other things) the appearance of "long-range order"[3] which we define (in general) as

$$L = \liminf (r_{ij} \to \infty) \lim (N \to \infty) \langle \sigma_i \sigma_j \rangle_N, \quad (23)$$

where the $N \to \infty$ limit implies some "sensible" means of defining a correlation function as the number of spins N tends to infinity.[4] In the limit inferior as $r_{ij} \to \infty$ we allow the direction of the vector joining the two spins to vary, though Schultz, Mattis, and Lieb[3] have shown that the result is independent of direction for the Ising ferromagnet mentioned above.

An obvious application of Theorem 2 is the following: Given an Ising model A with purely ferromagnetic interactions, the long-range order L is never less for a model B obtained from A by adding ferromagnetic bonds. Further, the transition temperature (Curie point) of B, which we define as the highest temperature at which long-range order appears, is not less than that of A.

Thus suppose, for example, that we have a two-dimensional square Ising lattice with a ferromagnetic nearest-neighbor interaction, and also ferromagnetic interactions, of arbitrary magnitude, with second, third, and fourth nearest neighbors. This model must (to no one's great surprise!) exhibit long-range order at any temperature below the Curie temperature obtained by Onsager.[5] Or, as another example, consider the particular case of long-range interactions (decreasing exponentially in one of the lattice directions) for which Kac and Thompson[6] have recently shown that a two-dimensional Ising model exhibits long-range order at sufficiently low temperatures. Since the potential is obtained by adding ferromagnetic terms to a case with ferromagnetic interactions between nearest neighbors, the existence of long-range order at low enough temperatures follows at once from Theorem 2.

As another application, we note that the existence of long-range order for the Ising ferromagnet in a two-dimensional square lattice with nearest-neighbor interactions at sufficiently low temperatures implies the same for the corresponding three-dimensional simple cubic lattice. Suppose that spin i is located at $(0, 0, 0)$ and j at (n, m, p)—the three numbers giving x, y, and z coordinates. Let spin k be located at $(n, q, 0)$. By Theorem 3,

$$\langle \sigma_i \sigma_j \rangle \geq \langle \sigma_i \sigma_k \rangle \langle \sigma_k \sigma_j \rangle. \quad (24)$$

But at sufficiently low temperatures $\langle \sigma_i \sigma_k \rangle$ is bounded from below[7] since both spins lie in a plane perpendicular to the z axis, and similarly $\langle \sigma_k \sigma_j \rangle$, since both spins lie in a plane perpendicular to the x axis. We know that long-range order exists for such planar lattices, and the fact that they form portions of three dimensional lattices merely implies that the additional ferromagnetic interactions present serve to enhance (by Theorem 2) or, at the least, not decrease, the correlation functions calculated for planar lattices alone.

This last result is, once again, not unexpected, especially since the presence of spontaneous magnetization in the simple cubic lattice described can be proved by using a simple argument given by Peierls,[8] a rigorous version of which was developed by the author[9] and independently by Dobrushin.[10] The power of Theorem 2 is, we believe, illustrated in the fact that one can proceed immediately from the two- to the three-dimensional case with no need of invoking any new combinatorial argument. And, of course, the cubic lattice with ferromagnetic nearest-neighbor and next-nearest-neighbor interactions, or interactions decreasing as $1/r^4$, or a multitude of other cases, are known immediately to display long-range order at low enough temperatures.

APPENDIX. PARTITION FUNCTIONS FOR COMPLETE SYSTEMS

The partition function Z associated with any diagram [or Hamiltonian of the form (1)] is obtained as follows. A configuration γ denotes a division of indices labeling different spins into two disjoint complementary sets $U(\gamma)$ and $D(\gamma)$. For $j \in U(\gamma)$, $\sigma_j = +1$ ("up") and for $k \in D(\gamma)$, $\sigma_k = -1$ ("down"). Configurations γ and γ' are distinct if and only if $D(\gamma) \neq D(\gamma')$ [or, the equivalent, $U(\gamma) \neq U(\gamma')$]. We now define

$$Z = \sum_{\gamma} Z_\gamma, \quad (A1)$$

[3] T. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).

[4] See R. B. Griffiths, J. Math. Phys. **8**, 484 (1967) (following paper) for an approach which works for an Ising ferromagnet, and M. E. Fisher, J. Math. Phys. **6**, 1643 (1965) for a more general procedure.

[5] L. Onsager, Phys. Rev. **65**, 117 (1944).

[6] M. Kac and C. J. Thompson, Proc. Natl. Acad. Sci. U.S. **55**, 676 (1966). A recent note from these authors indicates that the proof as published is not correct and will require modification.

[7] In accordance with our definition (23) we must assume that spins i and k are sufficiently far apart, and similarly k and j. This may be accomplished by a proper choice of q.

[8] R. Peierls, Proc. Cambridge Phil. Soc. **32**, 477 (1936).

[9] R. B. Griffiths, Phys. Rev. **136**, A437 (1964).

[10] R. L. Dobrushin, Teoriya Veroyatnostei Primeneniya **10**, 209 (1965).

where

$$Z_\gamma = \prod_{i\in D(\gamma)} \prod_{j\in U(\gamma)} X_{ij} \qquad \text{(A2)}$$

and, in an unsaturated diagram, X_{ij} is set equal to 1 for absent bonds. Each Z_γ is either 1 or a linear term as defined in Sec. III.

Lemma A1: A restricted partition function Z' (that is, with the value of one or more of the σ's specified) for a connected diagram has the form

$$Z' = \sum_\gamma{}' g_\gamma \qquad \text{(A3)}$$

with g_γ (either 1 or a linear term) $\neq g_\eta$ for $\gamma \neq \eta$. The prime denotes a summation over all configurations satisfying the restriction.

The proof is almost obvious. We know that at least one σ has a specified value, say $\sigma_o = +1$. In a configuration γ we can determine the value (± 1) of any spin σ_i connected to σ_o by a bond by observing whether X_{oi} is present or absent in g_γ. The values of still other spins connected by bonds to these σ_i may be determined by repeating this process, and eventually the configuration γ is uniquely determined from a knowledge of g_γ, since the diagram is connected. [We remark that the lemma holds for the unrestricted partition function for a connected diagram if a factor of 2 is placed in front of the summation in (A3).]

Lemma A2: Let $Z^{(1)}$ and $Z^{(2)}$ be two restricted or unrestricted partition functions (they may be identical, or there may be different restrictions in the two cases) corresponding to the same complete diagram. Suppose the product is decomposed in the form (15):

$$Z^{(1)}Z^{(2)} = G_o + \sum_{p=1} g_p G_p . \qquad \text{(A4)}$$

The sets W_p corresponding to the linear terms g_p (see Sec. III)are, of course, distinct: $W_p \neq W_{p'}$ for $p \neq p'$.

(i) If all the X's in a *particular* W are set equal to 1 and the corresponbing bonds in the diagram erased, the resulting diagram consists of two disconnected pieces, each of which is complete.

(ii) For $p \neq p'$, W_p is not a subset of $W_{p'}$.

To prove part (i) we consider a particular term $Z_\gamma Z_{\gamma'}$ [see the definition (A2)] in the product $Z^{(1)}Z^{(2)}$, where γ and γ' are configurations permitted by the restrictions (if any) for Z_1 and Z_2, respectively. Now if $D(\gamma)$ is identical with either $D(\gamma')$ or $U(\gamma')$, then $Z_\gamma = Z_{\gamma'}$ and the product contains no linear term. When $D(\gamma)$ is not identical with $D(\gamma')$ or with $U(\gamma')$, we have a situation illustrated schematically in Fig. 6, where the horizontal line indicates the division of indices into $U(\gamma)$ and $D(\gamma)$ [U and D for short] and the vertical into $U(\gamma')$ and $D(\gamma')$ [U' and D' for short]. We now ask, which X's occur linearly in the product $Z_\gamma Z_{\gamma'}$? That is, which X's occur in one factor but not in the other? There are four possibilities: X_{ij} occurs linearly if (a) $i \in D \cap D'$, $j \in U \cap D'$; (b) $i \in U \cap D'$, $j \in U \cap U'$; (c) $i \in D \cap U'$, $j \in U \cap U'$; (d) $i \in D \cap D'$, $j \in D \cap U'$. These bonds, represented schematically in Fig. 6, constitute the set W for the term $Z_\gamma Z_{\gamma'}$.

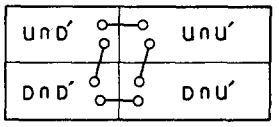


FIG. 6. Schematic diagram illustrating the division of all spin indices (represented by the complete rectangle) into sets according to two configurations γ and γ'. The small circles connected by straight lines indicate bonds whose factors enter linearly in the product $Z_\gamma Z_{\gamma'}$.

If we erase all bonds corresponding to X's in W, it is evident from Fig. 6 that the set of spins splits up into two disconnected sets,

$$\mathrm{A} = (D \cap D') \cup (U \cap U')$$

and

$$\mathrm{B} = (D \cap U') \cup (U \cap D').$$

That is, there are no bonds connecting the systems A and B. On the other hand, none of the bonds connecting two spins within A has been erased, nor any of the bonds connecting two spins within B. Therefore both A and B are complete. It is easily verified that if D is not identical to D' or to U', neither A nor B is a null set. This completes the proof of part (i).

To prove part (ii), assume that W_p is a proper subset of $W_{p'}$ (they cannot be identical for $p \neq p'$). In the initially complete diagram, erase all bonds corresponding to X's in W_p. The result, as we have just shown, is two disconnected systems A and B, each of which is complete. But if instead we were to erase all bonds corresponding to X's in the larger set $W_{p'}$, we would erase not only all the bonds connecting A and B, but additional bonds as well. That is, we would erase some of the bonds within A or within B. This would leave one or both systems incomplete in contradiction with part (i) of the lemma.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Correlations in Ising Ferromagnets. II. External Magnetic Fields*

ROBERT B. GRIFFITHS
Physics Department, Carnegie Institute of Technology, Pittsburgh, Pennsylvania

(Received 14 September 1966)

Results of a previous paper showing that $\langle\sigma_k\sigma_l\rangle$ and $[\langle\sigma_k\sigma_l\sigma_m\sigma_n\rangle - \langle\sigma_k\sigma_l\rangle\langle\sigma_m\sigma_n\rangle]$ are always positive for a system of Ising spins $\sigma_i = \pm 1$ coupled by a purely ferromagnetic interaction ($\langle\ \rangle$ denotes a thermal average) are extended to the cases where (i) certain spins are constrained to have the value $+1$ or (ii) the system is placed in an external ("parallel") magnetic field H. The theorems thus obtained provide a simple proof of the existence of "bulk" values for $\langle\sigma_k\sigma_l\rangle$ and for $\langle\sigma_k\rangle$; the latter is identical with the usual bulk magnetization per spin. The correlation functions $\langle\sigma_k\sigma_l\rangle$ are monotone nondecreasing in $|H|$ for fixed temperature T. Both $\langle\sigma_k\sigma_l\rangle$ and $\langle\sigma_k\rangle$ (and thus the bulk magnetization) are monotone nonincreasing in T for fixed $H \geq 0$.

I. INTRODUCTION

In a previous paper[1] (hereafter referred to as CIF I) we showed that for a system of N Ising spins $\sigma_i = \pm 1$ with Hamiltonian

$$\mathcal{H} = -\sum_{i<j} J_{ij}(\sigma_i\sigma_j - 1), \tag{1}$$

where for $i \neq j$,

$$0 \leq J_{ij} = J_{ji} < \infty, \tag{2}$$

the following statements are valid:

(A) For any k and l,

$$\langle\sigma_k\sigma_l\rangle \geq 0. \tag{3}$$

(B) For any spins k, l, m, and n (not necessarily distinct)

$$\beta^{-1}\partial\langle\sigma_k\sigma_l\rangle/\partial J_{mn} = \langle\sigma_k\sigma_l\sigma_m\sigma_n\rangle - \langle\sigma_k\sigma_l\rangle\langle\sigma_m\sigma_n\rangle \geq 0 \tag{4}$$

and J_{kl} or J_{mn} or both may be negative.

(C) For any spins k, l, and n

$$\langle\sigma_k\sigma_n\rangle \geq \langle\sigma_k\sigma_l\rangle\langle\sigma_l\sigma_n\rangle \tag{5}$$

and J_{kl} or J_{ln} or both may be negative.

The angular brackets denote a thermal average:

$$\langle\mathcal{O}\rangle = \mathrm{Tr}\,[\mathcal{O}e^{-\beta\mathcal{H}}]/Z, \tag{6}$$

where

$$Z = \mathrm{Tr}\,[e^{-\beta\mathcal{H}}] \tag{7}$$

is the partition function and Tr denotes the sum over all configurations (a configuration is a specific assignment of the value $+1$ or -1 to each σ_i). We always assume $\beta = (kT)^{-1}$ is nonnegative.

We extend these results as follows:

Theorem 4[2]*:* Statements A, B, and C are valid for a system described by (1) and (2) but having the restriction that certain spins belonging to a set U are all positive:

$$\text{for} \qquad i \in U, \quad \sigma_i = +1. \tag{8}$$

[Alternatively, one may require $\sigma_i = -1$ for all i in U.] By the restriction we mean that only configurations satisfying (8) appear in the traces (6) and (7).

Theorem 5: Statements A, B, and C are valid for a system with Hamiltonian [$\mathcal{H}$ is defined in (1) and (2)]:

$$\mathcal{H}_1 = \mathcal{H} - \sum_{i=1}^{N} H_i\sigma_i \tag{9}$$

and $H_i \geq 0$ for every i [the theorem also holds if $H_i \leq 0$ for every i] where $\mathcal{H}_1$ replaces $\mathcal{H}$ in calculating thermal averages, Eqs. (6) and (7). If $H_i = H$ for every i, $\mathcal{H}_1$ is the Hamiltonian of an Ising model in an "external parallel magnetic field."

Corollary 1: Under the conditions of either Theorem 4 or Theorem 5, for any k,

$$\langle\sigma_k\rangle \geq 0 \tag{10a}$$

and for any pair k, m,

$$\langle\sigma_k\sigma_m\rangle \geq \langle\sigma_k\rangle\langle\sigma_m\rangle. \tag{10b}$$

[In the case where the σ_i in U are -1, or where $H_i \leq 0$ for all i, the inequality (10a) is reversed.]

A brief summary of notation from CIF I is found in Sec. II followed by proofs of Theorems 4 and 5 in Secs. III and IV, respectively. In Sec. V we apply these results to prove the existence of binary correlation functions for a fairly general Ising model in the "bulk limit," that is, for a suitable infinite system. This existence proof is simple, rigorous, and quite different in its approach from other arguments for the existence of these functions of which we are aware. The same argument provides a bulk limit for

* Research supported in part by the National Science Foundation.
[1] R. B. Griffiths, J. Math. Phys. 8, 478 (1967).
[2] Theorems are numbered consecutively with those in Ref. 1.

$\langle\sigma_k\rangle$. Its equivalence with the ordinary "thermodynamic" magnetization per spin in a uniform, nonzero external magnetic field is shown in Sec. VI. In Sec. VII the temperature and field dependence of magnetization and binary correlation functions are discussed with the aid of Theorem 5.

II. NOTATION

We summarize the diagrammatic and notational conventions from CIF I. A diagram representing the Hamiltonian (1) or its associated partition function consists of small circles representing spins joined by straight lines or "bonds," one for every nonzero J_{ij} in (1). The partition function is a polynomial in the factors

$$X_{ij} = e^{-2\beta J_{ij}} \tag{11}$$

associated with the different bonds. In a *restricted* partition function the sum (7) is limited to configurations in which certain spins have specified values; thus $Z(p+\ k-)$ includes only configurations with $\sigma_p = +1$ and $\sigma_k = -1$. This may be indicated on the corresponding diagram by placing + or − beside the spins in question.

Provided no J_{ij} is 0, i.e., all spins are connected by bonds, the diagram is "complete." It is connected if one can move continuously from one spin to any other by means of bonds connecting spins. By letting $J_{kl} \to \infty$ or $X_{kl} \to 0$ we "combine" the spins k and l. That is, if a single spin k' in a new diagram replaces the two spins k and l in the previous diagram, and $X_{k'm} = X_{km}X_{lm}$ for all m, the new partition function is precisely that obtained by setting $X_{kl} = 0$ everywhere in the previous partition function.

Additional notation in connection with the Hamiltonian (9) is introduced in Sec. IV.

III. PROOF OF THEOREM 4

First consider the case where U contains the single spin h, and rewrite (3) as

$$Z(h+)\langle\sigma_k\sigma_l\rangle = [Z(h+\ k+\ l+) + Z(h+\ k-\ l-) - Z(h+\ k+\ l-) - Z(h+\ k-\ l+)] = Q \geq 0. \tag{12}$$

If h is the same as k, Q becomes $Z(h+\ l+) - Z(h+\ l-)$ which is nonnegative by Theorem 1 of CIF I.

When h, k, and l are distinct we argue by induction. Assume (12) is true for any system containing M spins (it is easily verified for $M = 3$) and consider a case of $M + 1$ spins with h connected by a single bond to another spin m [Fig. 1(a)]. As Q is linear in X_{hm}, it suffices to check (12) at $X_{hm} = 0$ and 1 [compare (2) and (11)]. But $X_{hm} = 0$ "combines"

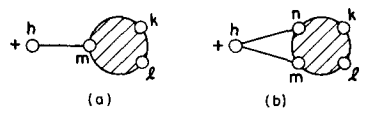


FIG. 1. Diagram illustrating the proof of statement A.

spins h and m, and (12) follows from the induction hypothesis, as we have but M spins. When $X_{hm} = 1$, h is disconnected from the remaining M spins and does not affect $\langle\sigma_k\sigma_l\rangle$, which is nonnegative by Theorem 1 of CIF I.

Next suppose [Fig. 1(b)] that h is connected to two spins m and n. Q is linear in X_{hn}. But $X_{hn} = 1$ reduces to the case just considered (the bond hn vanishes), and $X_{hn} = 0$ to a system of M spins. In analogous fashion we may add bonds from h to every other spin. No difficulty arises in adding the bond hk (or hl), since setting $X_{hk} = 0$ identifies h and k, a case considered earlier.

The inequality (4) presents a more difficult problem. We suppose h, k, l, m, and n are five distinct spins and rewrite (4) as

$$\tfrac{1}{4}Z(h+)^2\,[\langle\sigma_k\sigma_l\sigma_m\sigma_n\rangle - \langle\sigma_k\sigma_l\rangle\langle\sigma_m\sigma_n\rangle] = F = (\alpha + \alpha')(\beta + \beta') - (\gamma + \gamma')(\delta + \delta'), \tag{13}$$

where

$$\begin{aligned}
\alpha &= Z(h+\ k+\ l+\ m+\ n+) + Z(h+\ k+\ l+\ m-\ n-),\\
\beta &= Z(h+\ k+\ l-\ m+\ n-) + Z(h+\ k+\ l-\ m-\ n+),\\
\gamma &= Z(h+\ k+\ l+\ m+\ n-) + Z(h+\ k+\ l+\ m-\ n+),\\
\delta &= Z(h+\ k+\ l-\ m+\ n+) + Z(h+\ k+\ l-\ m-\ n-),
\end{aligned} \tag{14}$$

and α' is obtained from α by changing $k+$ to $k-$, $l+$ to $l-$, $m+$ to $m-$, $n+$ to $n-$, and vice versa, β' is similarly obtained from β, and so forth.

As shown in Sec. III of CIF I, F may be decomposed in the form

$$F = G_0 + \sum_{p=1} g_p G_p, \tag{15}$$

where the *linear terms* g_p are products of distinct X's in a set W_p, and the *quadratic terms* G_p are polynomials in which wherever any X_{ij} occurs (if it occurs at all) it appears as X_{ij}^2.

Our strategy closely follows that of CIF I, Sec. III. First consider the relatively simple case shown in Fig. 2, where the diagram consists of two disconnected pieces A and B, each of which is assumed to

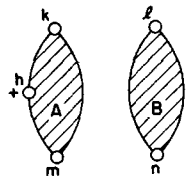


FIG. 2. Special case considered in the proof of statement B.

be complete, but is otherwise arbitrary. We show that G_0 in (15) is nonnegative for this example.

The partition functions (14) factor; for example:

$$Z(h+\,k+\,l-\,m-\,n+) = Z_{\mathrm{A}}(h+\,k+\,m-)Z_{\mathrm{B}}(l-\,n+) \tag{16}$$

with subscripts A and B referring to the separate pieces in Fig. 2. This factorization plus time-reversal invariance [for instance, $Z_{\mathrm{B}}(l-\,n+) = Z_{\mathrm{B}}(l+\,n-)$] allows us to write

$$F = F^{\mathrm{A}}F^{\mathrm{B}}, \tag{17}$$

$$F^{\mathrm{A}} = \{[Z_{\mathrm{A}}(h+\,k+\,m+) + Z_{\mathrm{A}}(h+\,k-\,m-)]^2 - [Z_{\mathrm{A}}(h+\,k+\,m-) + Z_{\mathrm{A}}(h+\,k-\,m+)]^2\}, \tag{18}$$

$$F^{\mathrm{B}} = [Z_{\mathrm{B}}(l+\,n+)]^2 - [Z_{\mathrm{B}}(l+\,n-)]^2. \tag{19}$$

The term G_0 in (15) is the product $G_0^{\mathrm{A}}G_0^{\mathrm{B}}$, where G_0^{A} is the quadratic term without a linear term as a factor in the decomposition of F^{A} in the form (15) and G_0^{B} the corresponding term for G^{B}. Lemma 1 of CIF I shows that $G_0^{\mathrm{B}} \geq 0$; we show the same is true of G_0^{A}. We may write

$$Z_{\mathrm{A}}(h+\,k+\,m+) + Z_{\mathrm{A}}(h+\,k-\,m-) = \sum_p r_p, \tag{20}$$

where each r_p is either 1 or a linear term, and for $p \neq p'$, $r_p \neq r_{p'}$, that is, the two contain different factors. Such a decomposition is possible for either one of the two terms on the left side of (20) according to Lemma A1 of CIF I, Appendix. But since the factor X_{hk} occurs in every linear term of $Z_{\mathrm{A}}(h+\,k-\,m-)$ and in none of the linear terms of $Z_{\mathrm{A}}(h+\,k+\,m+)$, the two classes of linear terms are disjoint. By a similar argument we may write

$$Z_{\mathrm{A}}(h+\,k+\,m-) + Z_{\mathrm{A}}(h+\,k-\,m+) = \sum_q t_q \tag{21}$$

with the t's linear terms and $t_q \neq t_{q'}$ for $q \neq q'$.

Following the argument used in Lemma 1, CIF I Sec. III, we see that

$$G_0^{\mathrm{A}} = \tilde{Z}_{\mathrm{A}}(h+\,k+\,m+) + \tilde{Z}_{\mathrm{A}}(h+\,k-\,m-) - \tilde{Z}_{\mathrm{A}}(h+\,k+\,m-) - \tilde{Z}_{\mathrm{A}}(h+\,k-\,m+), \tag{22}$$

where $\tilde{Z}_{\mathrm{A}}$ is the restricted partition function for a system $\tilde{\mathrm{A}}$ obtained from A by doubling every J_{ij} which appears in A. This does not affect the restriction (2), so we conclude from comparison with (12) that G_0^{A} is nonnegative. Thus G_0 also must be nonnegative.

To establish (4) for a general case, we consider a complete diagram containing N spins. From a comparison of (13) and (14) it is evident that F is equal to $X_{kl}X_{mn}$ times something which does not contain these factors at all; i.e., the term G_0 is absent in the decomposition (15). We next set equal to one everywhere in F every X appearing in a particular g_p, say g_2 (this leaves G_2 unaltered). The resulting F', as shown in Sec. III of CIF I, corresponds to a diagram of the form shown in Fig. 2, except that, for example, m and n could be interchanged, and h could appear in either system A or system B. In any case, the quadratic term in F' which appears without any linear term as a factor is simply G_2, and by the argument given above this is nonnegative.

We have thus established (4) for the case where U contains the single spin h, and h, k, l, m, and n are all denote distinct spins. The cases $k = l$ or $m = n$ are uninteresting [(4) vanishes]. The case $l = m$ may be obtained by letting $X_{lm} \to 0$, and the result is the inequality (5). In the case $h = k$, one finds that the terms α', β', γ', and δ' in (13) all vanish, leaving an expression identical with Eq. (12) of CIF I, which was there shown to be nonnegative.

The case where U contains more than one spin may be reduced to the case where U contains a single spin by a very simple argument. Suppose there are two spins h and j in U. Since $\sigma_h = \sigma_j = +1$ and X_{hj} appears in a partition function only for configurations in which $\sigma_h\sigma_j = -1$, it is clear that none of the restricted partition functions with which we deal in establishing (3), (4), and (5) contains X_{hj} as a factor. They are, therefore, unaltered if we set $X_{hj} = 0$, that is, if we combine spins h and j to form a new spin h'. This brings us back to the case where U contains only one spin. Clearly, the same argument works given any number of spins in the set U, and this completes our proof. Corollary 1 is proved by letting l and n belong to the set U. Then (3) becomes (10a) and (4) becomes (10b).

IV. PROOF OF THEOREM 5

Theorem 5 is readily seen as a corollary of Theorem 4 if we proceed as follows. Given a system of spins $\sigma_1, \sigma_2, \cdots, \sigma_N$, we introduce in addition a "ghost spin" σ_0 which is restricted to have the value $+1$. For $j = 1, 2, \cdots, N$ we set

$$J_{0j} = H_j \geq 0. \tag{23}$$

The situation is illustrated in Fig. 3 for the case

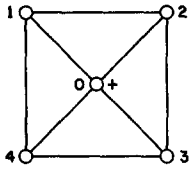


FIG. 3. Illustration of applying magnetic fields with the aid of a "ghost spin" (σ_0) restricted to the value +1.

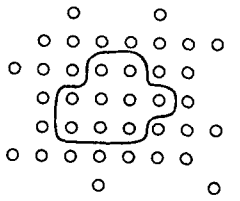


FIG. 4. Possible finite system Ω: all spins within the closed curve.

$N = 4$. Counting all $N + 1$ spins as one system, we have a Hamiltonian of the form (1):

$$\mathcal{H}_2 = -\sum_{i=1}^{N}\sum_{j=i+1}^{N} J_{ij}(\sigma_i\sigma_j - 1) - \sum_{i=1}^{N} H_i(\sigma_i\sigma_0 - 1), \quad (24)$$

which, because $\sigma_0 = +1$, differs from (9) only by an additive constant.[3] Thus statements A, B, and C applied to the original N spins are merely applications of Theorem 4. [In the event $H_i \leq 0$ for all i, we insert a minus sign in front of H_j in (23), require that σ_0 have the value -1 rather than $+1$, and proceed as before.] Corollary 1 is proved, as in Sec. IV, by identifying σ_l and σ_n with σ_0.

V. EXISTENCE OF BINARY CORRELATION FUNCTIONS

Theorem 5 may be used to prove the existence of a suitable "bulk limit" for binary correlation functions as the size of an Ising ferromagnet becomes infinite. The technique is best illustrated by a simple example: Ising spins on a two-dimensional square lattice with Hamiltonian

$$\mathcal{H} = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - H\sum_i \sigma_i, \quad (25)$$

where the J_{ij} satisfy (2) and possess the translational symmetry of the lattice; that is, they depend only on the vector $\mathbf{r}_{ij}$ joining spins i and j.

A finite system consists of a set Ω of spins which we may (but need not) assume are those found inside a simple closed curve, as illustrated in Fig. 4. The Hamiltonian $\mathcal{H}$ of this system is obtained by confining the sums in (25) to cases where i and j both lie in Ω. The diameter $D(\Omega)$ is the maximum distance between any pair of spins in Ω.

Consider a particular pair of spins k and l and a sequence of finite systems Ω_N, $N = 1, 2, \cdots$, with the properties

(i) Each Ω_N contains both the spins k and l.

(ii) If $d_{kl}(\Omega)$ is the minimum distance from either k or l to a spin outside Ω, then $d_{kl}(\Omega_N) \to \infty$ as $N \to \infty$.

(iii) $\Omega_N \subset \Omega_{N+1}$; that is, all spins in Ω_N are also in Ω_{N+1}.

[3] Which, needless to say, has no effect upon the correlation functions.

Let $\langle\sigma_k\sigma_l\rangle_N$ be the correlation function defined in the system Ω_N. Theorem 5 and condition (iii) imply that

$$\langle\sigma_k\sigma_l\rangle_{N+1} \geq \langle\sigma_k\sigma_l\rangle_N, \quad (26)$$

since Ω_{N+1} is obtained from Ω_N by adding ferromagnetic bonds to the latter (including bonds to the "ghost spin" of Sec. IV when $H \neq 0$). Since the $\langle\sigma_k\sigma_l\rangle_N$ form a monotone increasing sequence in N bounded from above by 1, they tend to a limit as $N \to \infty$.

The same limit is obtained using any other sequence of systems ω_N satisfying conditions (i) and (ii). For if N is large enough, we can always choose K and L, tending to infinity with N, such that

$$D(\Omega_K) < d(\omega_N) - 2a,$$
$$d(\Omega_L) > D(\omega_N) + 2a, \quad (27)$$

where a is the lattice constant and Ω_K and Ω_L belong to our standard sequence. We thus have

$$\Omega_K \subset \omega_N \subset \Omega_L \quad (28)$$

and $\langle\sigma_k\sigma_l\rangle$ for ω_N, bracketed between $\langle\sigma_k\sigma_l\rangle_K$ and $\langle\sigma_k\sigma_l\rangle_L$, tends as $N \to \infty$ to the limit previously obtained.

For spins k' and l' such that $\mathbf{r}_{k'l'} = \mathbf{r}_{kl}$, the bulk limit $\langle\sigma_{k'}\sigma_{l'}\rangle$ is the same as $\langle\sigma_k\sigma_l\rangle$. This is obvious because we can translate the systems Ω_N obtaining systems Ω'_N in which k' and l' have the same relative positions as k and l in Ω_N.

The above existence proof is quite simple and provides (at least for the systems considered) an approach quite different from others of which we are aware.[4] All that we require is a sequence of systems such that both k and l are eventually infinitely far from any walls. The sequence need *not* yield limiting values for bulk thermodynamic functions, and in particular there is no stability requirement for the energy (that is, it need not be bounded from below by $-CN$, where C is a constant and N the number of spins).[5] On the other hand we are unable to show that the same bulk limit would be obtained in, for example, a system with periodic boundary conditions.

The above argument applied to $\langle\sigma_0\sigma_k\rangle = \langle\sigma_k\rangle$,

[4] D. Ruelle, Ann. Phys. (N.Y.) **25**, 109 (1963); M. E. Fisher, J. Math. Phys. **6**, 1643 (1965); O. Penrose, *ibid.* **4**, 1312 (1963).

[5] We would guess that the bulk limit $\langle\sigma_k\sigma_l\rangle$ is 1 at any temperature if the stability condition is violated.

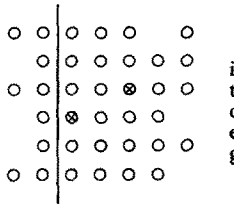


FIG. 5. The solid line represents an infinite wall; spins to the right are "inside" the system considered and those to the left outside. Spins k and l are marked by crosses; the correlation function $\langle\sigma_k\sigma_l\rangle$ will in general be different from the bulk value.

where σ_0 is the "ghost spin" of Sec. IV, proves the existence of a "bulk limit" for $\langle\sigma_k\rangle$. In Sec. VI below we show that this limit for $H \neq 0$ is identical with the average magnetization per spin obtained from the ordinary calculation of bulk thermodynamic quantities, in situations where the latter procedure is applicable. In particular, the limit as $H \to 0+$ of the bulk value of $\langle\sigma_k\rangle$ defines a spontaneous magnetization equivalent to definition A in Ref. 6.

By the above techniques it is also possible to define binary correlation functions $\langle\sigma_k\sigma_l\rangle$ (and likewise $\langle\sigma_k\rangle$) near a "wall" while the system is allowed to become infinite (illustrated in Fig. 5). We leave details of the argument to the reader, as also the extension, if not already obvious, of the above results to three (and, if preferred, higher)-dimensional systems.

VI. "THERMODYNAMIC" MAGNETIZATION AND $\langle\sigma_k\rangle$

For a finite system Ω_N in the shape of a square[7] containing $V_N = (2N)^2$ spins, we define the total magnetization operator

$$\mathcal{M} = \sum_{i\in\Omega_N} \sigma_i . \tag{29}$$

Provided suitable restrictions are placed on the J_{ij} appearing in (25),[8] one can show that for $H > 0$ the infinite volume or "thermodynamic" magnetization per spin

$$m(H) = \lim_{N\to\infty} V_N^{-1}\langle\mathcal{M}\rangle_N \tag{30}$$

exists and is an analytic function for $0 < H < \infty$.[9] By the arguments of Sec. V, any $\langle\sigma_i\rangle_N$ cannot exceed its bulk value, which we call σ for short, and thus

$$m(H) \leq \sigma. \tag{31}$$

We show that (31) is really an equality. Let σ_N (we assume H is fixed and positive) be the value of $\langle\sigma_k\rangle_N$ for the spin k lying nearest the center of the square Ω_N. From Theorem 5 it follows that for $M > N$, σ_{N-1} is a lower bound for $\langle\sigma_i\rangle_M$ for any spin i in Ω_M lying at least a distance Na (a is the lattice constant) from the nearest boundary. For spins closer to the boundary, $\langle\sigma_i\rangle_M$ is certainly nonnegative.

Thus we have a bound

$$V_M^{-1}\langle\mathcal{M}\rangle_M \geq V_M^{-1}(2M - 2N)^2\sigma_{N-1}. \tag{32}$$

The limit $M \to \infty$ followed by the limit $N \to \infty$ yields (31) with inequality sign reversed, and therefore

$$m(H) = \sigma. \tag{33}$$

Needless to say, similar arguments work for three (or one)-dimensional lattices.

VII. FIELD AND TEMPERATURE DEPENDENCE OF CORRELATION FUNCTIONS AND MAGNETIZATION

Consider the Hamiltonian (25) [satisfying condition (2)]. If H is greater than zero, an increase of H corresponds to increasing the ferromagnetic coupling between the "ghost spin" and the other spins, thus producing an increase, or at least not a decrease, in the binary correlation functions in accordance with (4). The same holds, of course, for their bulk limits. Likewise $\langle\sigma_i\rangle$ and its bulk limit σ increase with increasing field. The last result is not surprising, inasmuch as the increase of thermodynamic magnetization with increasing field reflects a general convexity property of the free energy for a spin system in which the Hamiltonian depends linearly on the magnetic field[8] (it is also true, for example, in an Ising antiferromagnet).

Next, suppose that H is fixed at some value ≥ 0. The X_{ij} depend on the temperature through (11). Increasing temperature or decreasing β has the effect of simultaneously decreasing all the J_{ij} which are greater than zero. Thus by Theorem 5, an increase of temperature at constant field leads to a decrease, or at least not an increase, of the binary correlation functions and the $\langle\sigma_i\rangle$ (for $H > 0$); the same holds true, of course, for their bulk limits. That the magnetization decreases with temperature in a fixed field is a nontrivial, though not unexpected, result.

The limit (at constant temperature) as $H \to 0+$ of the bulk magnetization is the spontaneous magnetization if one adopts definition A of Ref. 6. The above remarks imply that the spontaneous magnetization is a nonincreasing function of the temperature. Theorem 5 implies that the addition of ferromagnetic bonds to an Ising system with purely ferromagnetic interaction always increases (does not decrease) the spontaneous magnetization and therefore the Curie temperature (the lowest temperature at which the

[6] R. B. Griffiths, Phys. Rev. 152, 240 (1966).
[7] Other shapes are possible and we consider a square only for simplicity.
[8] R. B. Griffiths, J. Math. Phys. 5, 1215 (1964).
[9] C. N. Yang and T. D. Lee, Phys. Rev. 87, 404, 410 (1952).

spontaneous magnetization vanishes). Thus, for example, since it is known that spontaneous magnetization, in the above sense, occurs at sufficiently low temperatures for a variety of two-dimensional Ising ferromagnets, it follows immediately that the same is true of three-dimensional ferromagnets obtained by joining two-dimensional layers with ferromagnetic bonds.

ACKNOWLEDGMENTS

The author is indebted to Professor J. Lebowitz for some valuable suggestions and comments on the first draft of this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Embedding of a Simple Lie Group into a Simple Lie Group and Branching Rules

A. NAVON* AND J. PATERA†
Département de Physique, Université de Montréal, Montréal, Canada

(Received 27 June 1966)

A criterion established by Dynkin is used to specify the embedding of a connected simple Lie group $\mathfrak{G}'$ into a connected simple Lie group $\mathfrak{G}$, and to derive a standard procedure for evaluating branching rules. It is shown that the weight systems of the irreducible parts contained in the representation of $\mathfrak{G}'$ induced by a given finite dimensional representation φ of $\mathfrak{G}$ are obtained by projection of the weight system of φ. The projection mapping is determined directly from the specification of the embedding. The general procedure is supplemented with two constraint equations on the dimensions and indices of the irreducible representations.

I. INTRODUCTION

THE study of embeddings and branching rules for two simple Lie groups and their representations is motivated in elementary particle physics by the need one often has to relate a symmetry scheme to another possible symmetry scheme.[1,2] This study is also of interest in nuclear spectroscopy where nuclear states may be classified by using group chains.[3] In a recent paper, Whippman[2] summarized branching rules for various choices of two classical simple Lie groups. His choices are, however, particular from several points of view; for instance, only classical groups of certain ranks and types and which can be embedded one into the other in at most two distinct ways are considered.

Our purpose in this paper is to develop a standard procedure for obtaining branching rules for any choice of two connected simple Lie groups with no restriction concerning their ranks and types and the way they are embedded one into the other. Several of our arguments, and especially a criterion of equivalence of embeddings, follow the work of Dynkin.[4]

The main tools we use to study the embeddings and the branching rules for two simple connected Lie groups $\mathfrak{G}'$ and $\mathfrak{G}$, are a mapping f (embedding) between the two corresponding Lie algebras G' and G, and a related projection mapping f^* acting between the root spaces R and R' of these Lie algebras. We start, in Sec. II, by reviewing some useful properties of the root space and the weight system of a Lie algebra. The notion of the index of a representation is also recalled. In Sec. III the embedding f of a Lie algebra G' into a Lie algebra G is studied. When G is of the type[5] A_n, B_n, C_n, G_2, F_4, or E_6, f can be specified by that representation of G', embedded into the lowest dimensional representation of G. When G is of the type D_n, E_7, or E_8, f is specified by two representations of G' embedded respectively into the lowest dimensional and into some other representation of G. In Sec. IV it is shown how one can introduce from the embedding f, the mapping f^* which maps the root space of G onto the root space of G'. The most valuable property of f^* is that it maps the system of weights of any irreducible representation of G onto

* Holder of a Studentship from the National Research Council of Canada.

† Postdoctoral Fellow of N.R.C., on leave of absence from the Physics Institute of the Czechoslovak Academy of Science.

[1] For example, R. Gatto, in *Theoretical Physics* (International Atomic Energy Agency, Vienna, 1963), p. 197.

[2] M. L. Whippman, J. Math. Phys. 6, 1534 (1965), and references therein.

[3] M. Hamermesh, *Group Theory* (Addison-Wesley Publishing Company Inc., London 1964), Chap. 11.

[4] E. B. Dynkin, Matematiceskii Sbornik New Series 30, 349 (1952); also Am. Math. Soc., Transl., Ser. 2 6, 111 (1957).

[5] In Cartan's notation, we distinguish nine types of simple Lie groups; four classical A_n, B_n, C_n, D_n, and five exceptional G_2, F_4, E_6, E_7, E_8.

spontaneous magnetization vanishes). Thus, for example, since it is known that spontaneous magnetization, in the above sense, occurs at sufficiently low temperatures for a variety of two-dimensional Ising ferromagnets, it follows immediately that the same is true of three-dimensional ferromagnets obtained by joining two-dimensional layers with ferromagnetic bonds.

ACKNOWLEDGMENTS

The author is indebted to Professor J. Lebowitz for some valuable suggestions and comments on the first draft of this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Embedding of a Simple Lie Group into a Simple Lie Group and Branching Rules

A. NAVON* AND J. PATERA†
Département de Physique, Université de Montréal, Montréal, Canada

(Received 27 June 1966)

A criterion established by Dynkin is used to specify the embedding of a connected simple Lie group $\mathfrak{G}'$ into a connected simple Lie group $\mathfrak{G}$, and to derive a standard procedure for evaluating branching rules. It is shown that the weight systems of the irreducible parts contained in the representation of $\mathfrak{G}'$ induced by a given finite dimensional representation φ of $\mathfrak{G}$ are obtained by projection of the weight system of φ. The projection mapping is determined directly from the specification of the embedding. The general procedure is supplemented with two constraint equations on the dimensions and indices of the irreducible representations.

I. INTRODUCTION

THE study of embeddings and branching rules for two simple Lie groups and their representations is motivated in elementary particle physics by the need one often has to relate a symmetry scheme to another possible symmetry scheme.[1,2] This study is also of interest in nuclear spectroscopy where nuclear states may be classified by using group chains.[3] In a recent paper, Whippman[2] summarized branching rules for various choices of two classical simple Lie groups. His choices are, however, particular from several points of view; for instance, only classical groups of certain ranks and types and which can be embedded one into the other in at most two distinct ways are considered.

Our purpose in this paper is to develop a standard procedure for obtaining branching rules for any choice of two connected simple Lie groups with no restriction concerning their ranks and types and the way they are embedded one into the other. Several of our arguments, and especially a criterion of equivalence of embeddings, follow the work of Dynkin.[4]

The main tools we use to study the embeddings and the branching rules for two simple connected Lie groups $\mathfrak{G}'$ and $\mathfrak{G}$, are a mapping f (embedding) between the two corresponding Lie algebras G' and G, and a related projection mapping f^* acting between the root spaces R and R' of these Lie algebras. We start, in Sec. II, by reviewing some useful properties of the root space and the weight system of a Lie algebra. The notion of the index of a representation is also recalled. In Sec. III the embedding f of a Lie algebra G' into a Lie algebra G is studied. When G is of the type[5] A_n, B_n, C_n, G_2, F_4, or E_6, f can be specified by that representation of G', embedded into the lowest dimensional representation of G. When G is of the type D_n, E_7, or E_8, f is specified by two representations of G' embedded respectively into the lowest dimensional and into some other representation of G. In Sec. IV it is shown how one can introduce from the embedding f, the mapping f^* which maps the root space of G onto the root space of G'. The most valuable property of f^* is that it maps the system of weights of any irreducible representation of G onto

* Holder of a Studentship from the National Research Council of Canada.

† Postdoctoral Fellow of N.R.C., on leave of absence from the Physics Institute of the Czechoslovak Academy of Science.

[1] For example, R. Gatto, in *Theoretical Physics* (International Atomic Energy Agency, Vienna, 1963), p. 197.

[2] M. L. Whippman, J. Math. Phys. 6, 1534 (1965), and references therein.

[3] M. Hamermesh, *Group Theory* (Addison-Wesley Publishing Company Inc., London 1964), Chap. 11.

[4] E. B. Dynkin, Matematiceskii Sbornik New Series 30, 349 (1952); also Am. Math. Soc., Transl., Ser. 2 6, 111 (1957).

[5] In Cartan's notation, we distinguish nine types of simple Lie groups; four classical A_n, B_n, C_n, D_n, and five exceptional G_2, F_4, E_6, E_7, E_8.

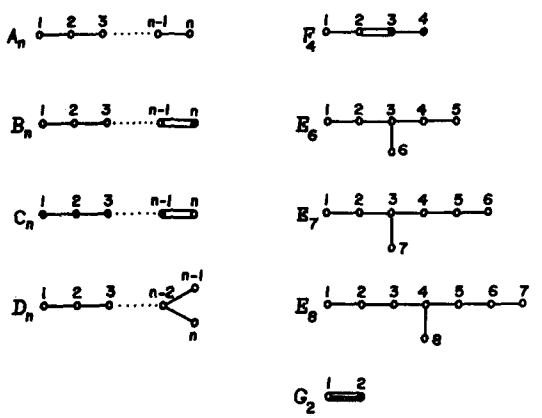


FIG. 1. Numbering of the simple roots.

the system of weights of the induced representation of G'. This property together with two constraint equations on the indices and dimensions of the irreducible representations, are used to present in Sec. V the standard procedure for obtaining the branching rules. An example is considered in detail. Tables I and II, contained in Sec. IV, give the mapping f^* for all types of simple Lie algebras G, once the embedding f is fixed.

II. MATHEMATICAL PRELIMINARIES[6] AND NOTATIONS

Because not all the ideas important for the present work are common in the physical literature, we first recall some of the definitions. Let us fix a system of simple roots $\{\alpha_i\}$, $i = 1, 2, \cdots, n$, of a simple Lie algebra G of rank n. Suppose the roots $\alpha_1, \alpha_2, \cdots, \alpha_n$ are numbered as in Fig. 1, and normalized by the condition

$$(\alpha_{\max}, \alpha_{\max}) = 2, \tag{1}$$

where $\alpha_{\max}$ is one of the longest roots of G, and $(,)$ denotes the scalar product in the root space R of G. The system of simple roots $\{\alpha_i\}$ forms a natural basis of R. However, for our purposes it is more convenient to introduce a conjugate basis $\{\nu_i\}$, $i = 1, 2, \cdots, n$, of R by the condition

$$(\alpha_i, \nu_j) = \tfrac{1}{2}(\alpha_i, \alpha_i)\,\delta_{ij}. \tag{2}$$

For an arbitrary vector

$$M = \sum_{i=1}^{n} m_i \nu_i$$

of R, we then have

$$m_i = [2/(\alpha_i, \alpha_i)](M, \alpha_i). \tag{3}$$

The introduction of the basis $\{\nu_i\}$ is justified by Cartan's theorem, which states that a vector M of R is the highest weight of an irreducible representation φ of G if and only if the coordinates (3) are nonnegative integers. In order to specify a particular representation φ we adopt the notation $\varphi = (m_1, m_2, \cdots, m_n)$, where the m_i are coordinates (3) of the highest weight M of φ.

An arbitrary weight L of the weight system $\Delta(\varphi)$ of a representation φ of G can be written in the form

$$L = M - \sum_{i=1}^{n} a^i \alpha_i,$$

where a^i are nonnegative integers. We say that a weight L belongs to the k layer of the system $\Delta(\varphi)$, if

$$\sum_{i=1}^{n} a^i = k.$$

Orderings in the root space R may be defined as usual.[6] Any ordering of R for which the fixed simple roots remain positive is called a simple-root ordering.

If a weight M is the only weight in an m layer and L is a weight from an l layer of a weight system $\Delta(\varphi)$ such that $m < l$ $(m > l)$, it follows from the above definitions and Dynkin's theorem[6] 0.15 that the relations $M > L$ $(M < L)$ hold in any simple-root ordering. The usefulness of this last remark is shown in Sec. IV.

For a given linear representation φ of G, the scalar product of the elements x and y of G being fixed by Eq. (1), we have[4,7]

$$\mathrm{Tr}\,[\varphi(x) \cdot \varphi(y)] = l(\varphi) \cdot (x, y). \tag{4}$$

The multiplicative factor $l(\varphi)$, called the index of the representation φ, does not depend on x and y and is given by

$$l(\varphi) = [d(\varphi)/d(G)] \cdot (\Lambda, \Lambda + g). \tag{5}$$

$d(\varphi)$ and $d(G)$ here denote the dimensions of the representation φ and of the algebra G, respectively, g is the sum of all the positive roots of G, and Λ is the highest weight of φ. From the definition, the following property of the index can be established:

$$l(\varphi_1 + \varphi_2 + \cdots + \varphi_s) = l(\varphi_1) + l(\varphi_2) + \cdots + l(\varphi_s). \tag{6}$$

III. THE EMBEDDING

Let us fix two simple Lie algebras G' and G corresponding respectively to two simple connected Lie groups $\mathfrak{G}'$ and $\mathfrak{G}$. An embedding f of G' into G is an isomorphic mapping of G' into its image $f(G')$ which is a simple subalgebra of G. Clearly f may also be called

[6] For more details, see, for instance, E. B. Dynkin, Am. Math. Soc., Transl., Ser. 2 6, 245 (1957), Supplement; or N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1962).

[7] J. Patera, Nuovo Cimento (to be published).

a representation of G' in G. In general, there are in G several subalgebras isomorphic with G', and thus there exist several possible embeddings of G' into G. We call two embeddings f_1 and f_2 equivalent, if for any finite dimensional linear representation φ of G, the representations φf_1 and φf_2 induced by φ in $f_1(G')$ and $f_2(G')$, respectively, are equivalent. Since by definition two equivalent embeddings induce the same branching rules, we characterize an embedding only up to equivalence.

A convenient criterion of equivalence of embeddings was established by Dynkin.[4]

Case I: Let ω denote the linear representations: $(1, 0, \cdots, 0)$ for $G = A_n$, B_n, C_n, and E_6, $(0, 1)$ for $G = G_2$, and $(0, 0, 0, 1)$ for $G = F_4$. Then two embeddings f_1 and f_2 of G' into G are equivalent, if and only if the representations ωf_1 and ωf_2 of G' are equivalent.

Case II: Let $\omega = (1, 0, 0, \cdots, 0)$ and $\bar{\omega} = (0, \cdots, 0, 1, 0)$ for $G = D_n$ and E_8, $\omega = (0, \cdots, 0, 1, 0)$ and $\bar{\omega} = (1, 0, \cdots, 0)$ for $G = E_7$. Then two embeddings f_1 and f_2 of G' into G are equivalent, if and only if ωf_1 is equivalent to ωf_2 and $\bar{\omega} f_1$ is equivalent to $\bar{\omega} f_2$.[8]

We use this criterion to specify an embedding by the highest weights of ωf in Case I, and by the highest weights of ωf and $\bar{\omega} f$ in Case II. Another characteristic of an embedding f, which we use in Sec. V, is Dynkin's index j_f, defined in analogy with the index of a representation [Eq. (4)] by

$$(f(x), f(y)) = j_f \cdot (x, y); \; x, y \in G'. \tag{7}$$

Here (x, y) and $(f(x), f(y))$ are scalar products in G' and G, respectively, fixed by the normalization (1). From Eqs. (4) and (7) one obtains

$$j_f = l(\varphi f)/l(\varphi) \tag{8}$$

for an arbitrary representation φ of G, and in particular

$$j_f = l(\omega f)/l(\omega) = l(\bar{\omega} f)/l(\bar{\omega}). \tag{9}$$

IV. THE MAPPING f^*

Gantmacher[9] proved that for a given Cartan subalgebra K' of G', it is always possible to choose a Cartan subalgebra K of G such that

$$f(K') \subset K. \tag{10}$$

Thus the restriction of the mapping f to K' determines a dual mapping

$$\hat{f}: R' \to \hat{f}(R') \subset R, \tag{11}$$

where R' and R are respectively the root spaces of G' and G. A mapping f^* projecting canonically R onto R' may be defined by

$$f^*(\hat{f}(L)) = L, \quad \text{for} \quad L \in R', \tag{12}$$

$$f^*(R_0) = 0, \quad \text{for} \quad R_0 = R - \hat{f}(R'). \tag{13}$$

The construction of f^* makes it to have the important property of mapping the weight system $\Delta(\varphi)$ of an arbitrary representation φ of G onto the weight system $\Delta(\varphi')$ of the induced representation $\varphi' = \varphi f$ of G'. For let the representation φ of G act in a vector space V, and let v be a vector of V corresponding to a weight Λ of $\Delta(\varphi)$; i.e.,

$$\varphi(K)v = \Lambda v. \tag{14}$$

Then v remains a weight vector for the induced representation φ' and from (10)–(14) one gets

$$\varphi f(K')v = f^*(\Lambda)v. \tag{15}$$

It follows that any weight of the weight system $\Delta(\varphi')$ must be of the form $f^*(\Lambda)$, $\Lambda \in \Delta(\omega)$, and we can write

$$\Delta(\varphi') = f^*(\Delta(\varphi)). \tag{16}$$

Consider now the weight systems $\Delta(\omega)$ and $\Delta(\omega f)$ of the representations ω and ωf defined in Sec. III. (We limit ourselves for simplicity to algebras G considered in Case I. In the Case II, one must take into account also the representations $\bar{\omega}$ and $\bar{\omega} f$, but the arguments are identical.) The weights of an irreducible weight system may be calculated by subtracting from the highest weight a certain number of times simple roots.[6] They can, at the same time, be ordered using a simple root ordering and partitioned in layers, as mentioned in Sec. II. Thus, we write

$$\Delta(\omega) = \{\Omega_1 > \Omega_2 \geqq \Omega_3 \geqq \cdots > \Omega_{d(\omega)}\}, \tag{17}$$

$$\Delta(\omega f) = \{M_1 \geqq M_2 \geqq \cdots \geqq M_{d(\omega)}\}, \tag{18}$$

where the weights of $\Delta(\omega)$ and $\Delta(\omega f)$ are ordered respectively following convenient simple-root-orderings O and O' of R and R', and $d(\omega)$ denotes the common dimension of the representations ω and ωf.

Now, although one has globally $\Delta(\omega f) = f^*(\Delta(\omega))$, the relation

$$f^*(\Omega_i) = M_i \tag{19}$$

is verified only if one of the following two conditions holds:

(A) The simple-root orderings in R' and R are consistent, that is, the relation $f^*(\Lambda_1) > f^*(\Lambda_2)$ in R' implies the relation $\Lambda_1 > \Lambda_2$ in R.

(B) The weight Ω_i is contained in a simple-weight layer of $\Delta(\omega)$; then one has $\Omega_i < \Omega_j$ ($\Omega_i > \Omega_j$) if

[8] One notes that for all G, ω is the lowest dimensional representation and for $G = D_n$, $\bar{\omega}$ is the spinor representation.

[9] F. R. Gantmacher, Mat. Sb. N.S. 5, 101 (1939), Theorem II.

TABLE I. Classical Lie algebras. The basis vectors $\{v_i\}$ of R are mapped by f^* into linear combinations of at most n weights of the representations ωf or $\bar{\omega} f$ of G'.

	i	$f^*(v_i)$
A_n, C_n	$1, 2, \cdots, n$	$\sum_{j=1}^{i} M_j$
B_n	$1, 2, \cdots, n-1$	$\sum_{j=1}^{i} M_j$
	n	$\frac{1}{2}\sum_{j=1}^{n} M_j$
D_n	$1, 2, \cdots, n-2$	$\sum_{j=1}^{i} M_j$
	$n-1$	$\overline{M}_1$
	n	$-\overline{M}_1 + M_{n-1} + M_{n-2} + 2\sum_{j=1}^{n-3} M_j = -\overline{M}_4$

$i > j (i < j)$ for any simple-root ordering of R, as follows from the remark of Sec. II.

Similar remarks can be made about weight systems $\Delta(\varphi)$ and $\Delta(\varphi f)$ corresponding to an arbitrary representation φ of G. The reason why the representation

TABLE II. Exceptional Lie algebras. The basis vectors $\{v_i\}$ of R are mapped by f^* into linear combinations of at most n weights of the representations ωf or $\bar{\omega} f$ of G'.

	i	$f^*(v_i)$
G_2	1	$M_1 + M_2$
	2	M_1
F_4	1	$M_3 + M_4$
	2	$M_1 + M_2 + M_3$
	3	$M_1 + M_2$
	4	M_1
E_6	1	M_1
	2	$M_1 + M_2$
	3	$M_1 + M_2 + M_3$
	4	$-M_{27} - M_{36}$
	5	$-M_{27}$
	6	$M_1 + M_2 + M_3 + M_4 + M_{36} + M_{27}$
E_7	1	$\overline{M}_1$
	2	$\overline{M}_1 + \overline{M}_2$
	3	$\overline{M}_1 + \overline{M}_2 + \overline{M}_3 = M_1 + M_2 + M_3 + M_4$
	4	$M_1 + M_2 + M_3$
	5	$M_1 + M_3$
	6	M_1
	7	$M_5 + \overline{M}_3 = M_1 + M_2 + M_3 + M_4 + M_5 - \overline{M}_1 - \overline{M}_2$
E_8	1	M_1
	2	$M_1 + M_2$
	3	$M_1 + M_2 + M_3$
	4	$M_1 + M_2 + M_3 + M_4$
	5	$M_1 + M_2 + M_3 + M_4 + M_5 = \overline{M}_1 + \overline{M}_2 + \overline{M}_3$
	6	$\overline{M}_1 + \overline{M}_2$
	7	$\overline{M}_1$
	8	$M_6 + \overline{M}_3 = M_1 + M_2 + M_3 + M_4 + M_5 + M_6 - \overline{M}_1 - \overline{M}_2$

ω (or ω and $\bar{\omega}$) is particularized in Dynkin's criterion is that $\Delta(\omega)$ [$\Delta(\omega)$ and $\Delta(\bar{\omega})$] contains n (n = rank G) linearly independent weights which satisfy condition (B). This can be verified readily by computing the weight system $\Delta(\omega)$ [$\Delta(\omega)$ and $\Delta(\bar{\omega})$] for each type of G. Consequently one has n linearly independent relations (19) connecting vectors of R and R'. In principle, one could use also the condition (A) for the specification of f^*, but the choice of consistent simple-root-orderings in R and R' is much less convenient to work with.

Tables I and II are intended to specify the mapping f^* on the basis vectors $\{v_i\}$ of R and $\{v_i'\}$ of R'. The quantities $f^*(v_i)$, $i = 1, 2, \cdots, n$, are given as linear combination of at most n weights $\{M_i\}$ or $\{\overline{M}_i\}$ of $\Delta(\omega f)$ or $\Delta(\bar{\omega} f)$. These weights are to be expressed in terms of the basis vectors $\{v_i\}$ once the embedding is fixed.

V. BRANCHING RULES

It is relatively simple now to formulate a standard procedure for obtaining branching rules. After a particular embedding of G' into G was specified, the mapping f^* is known explicitly and can be applied to the weight system $\Delta(\varphi)$ of any inducing representation φ of G. The general procedure can be, in principle, summarized in the following steps:

(1) Fix the embedding f, by expressing the weights $\{M_i\}$ of ωf in Table I or the weights $\{M_i\}$ of ωf and $\{\overline{M}_i\}$ of $\bar{\omega} f$ in Table II, in $\{v_i'\}$ notation.

(2) Obtain in $\{v_i\}$ notation the weight system

$$\Delta(\varphi) = \{\Lambda_1, \Lambda_2, \cdots, \Lambda_{d(\varphi)}\} \tag{20}$$

of the inducing representation φ of G.

(3) Apply the mapping f^* to the weights of $\Delta(\varphi)$ and, using Table I or II, find in $\{v_i'\}$ notation the weight system

$$\Delta(\varphi') = \{L_1, L_2, \cdots, L_{d(\varphi)}\} \tag{21}$$

of the induced representation φ' of G'.

(4) Separate in $\Delta(\varphi')$ all weight systems $\Delta(\varphi_I')$, $\Delta(\varphi_{II}')$, $\cdots$ which correspond to irreducible components.

These successive steps, in particular the last one, imply elaborate calculations, part of which is often superfluous. For instance, only the highest weights of irreducible components of φ' are subjects of our interest, but not the whole system $\Delta(\varphi')$. The general procedure is shortened by two equations of constraint.[7] The first one,

$$d(\varphi) = d(\varphi') = d(\varphi_I') + d(\varphi_{II}') + \cdots, \tag{22}$$

connects the dimensions of the unknown irreducible

TABLE III. Dimensions d and indices l of the representations of the algebras B_2 and A_5 which occur in the example of Sec. V.

B_2	(0, 0)	(0, 1)	(1, 0)	(0, 2)	(2, 0)	A_5	(1, 0, 0, 0, 0)	(1, 0, 0, 0, 1)
d	1	4	5	10	14	d	6	35
l	0	1	2	6	14	l	1	12

components in φ' to the known dimension of φ. The second equation,

$$l(\varphi') = j_f \cdot l(\varphi) = l(\varphi'_{\mathrm{I}}) + l(\varphi'_{\mathrm{II}}) + \cdots, \quad (23)$$

obtained from (6) and (8), relates the indices of the unknown nontrivial irreducible components of φ' to the known index of φ'.

Let us illustrate the whole method by an example. We consider an embedding of the algebra B_2 into A_5. According to Sec. III, the embedding can be specified as an isomorphism of B_2 into the representation $\omega = (1, 0, 0, 0, 0)$ of A_5. Thus one finds two possibilities,

$$B_2 \xrightarrow{f} \omega f = (1, 0) + (0, 0) \subset (1, 0, 0, 0, 0) \quad (24)$$

and

$$B_2 \xrightarrow{f} \omega f_1 = (0, 1) + 2(0, 0) \subset (1, 0, 0, 0, 0), \quad (25)$$

where ωf and ωf_1 are representations of B_2 which are matrix subalgebras of the representation ω of A_5. We have written explicitly the trivial representations (0, 0) necessary for the equality of dimensions of the embedded representations ωf, ωf_1, and ω.

Let us consider, for example, the embedding (24). The weight system $\Delta(\omega f)$ consists of

$$M_1 = \nu'_1; \quad M_2 = -\nu'_1 + 2\nu'_2; \quad M_3 = M_4 = 0;$$
$$M_5 = \nu'_1 - 2\nu'_2; \quad M_6 = -\nu'_1, \quad (26)$$

where ν'_1 and ν'_2 form the basis of the root space of B_2 defined by (2). As one can easily prove, the numbering (26) of the weights of $\Delta(\omega f)$ and the inequalities (18) imply a simple-root-ordering of the root space R' of B_2. Hence the mapping f^* is given explicitly by Table I and formulas (26).

Suppose, for instance, that the inducing representation φ of A_5 is the adjoint representation $\varphi = (1, 0, 0, 0, 1)$. Its weight system $\Delta(\varphi)$ consists of

$$\Lambda_1 = \nu_1 + \nu_5, \quad \Lambda_2 = -\nu_1 + \nu_2 + \nu_5, \cdots, \quad (27)$$

where ν_i, $i = 1, 2, \cdots, 5$, is the $\{\nu\}$-basis of the root space R of A_5. Using Table I and formulas (26) we apply the mapping f^* to the highest weight Λ_1 of φ:

$$f^*(\Lambda_1) = f^*(\nu_1) + f^*(\nu_5) = 2\nu'_1.$$

Here $2\nu'_1$ is a weight of the induced representation φ' of B_2. Since the highest weight Λ_1 of φ is always the only weight of the 0 layer of $\Delta(\varphi)$, it follows that Λ_1 is always mapped into a highest weight $f^*(\Lambda_1)$ of φ'. Consequently, one of the irreducible components of φ' is $\varphi'_{\mathrm{I}} = (2, 0)$. Similarly, the weight Λ_2 is mapped as follows

$$f^*(\Lambda_2) = -f^*(\nu_1) + f^*(\nu_2) + f^*(\nu_5) = 2\nu'_2.$$

One can check that the weight $2\nu'_2$ does not belong to the weight system $\Delta(\varphi'_{\mathrm{I}})$, hence it is the highest weight of another irreducible component $\varphi'_{\mathrm{II}} = (0, 2)$ of φ'.

In principle, the rest of the weight system $\Delta(\varphi')$ can be obtained by subsequent application of the mapping f^* to the rest of the system $\Delta(\varphi)$. However, the procedure is shortened by using the constraint Eqs. (22) and (23) as follows. From Eqs. (8) and (9) one has

$$l(\varphi') = [l(\omega f)/l(\omega)] \cdot l(\varphi) = 24,$$

where the values of indices and dimensions of the representations for the present example are summarized in Table III. The direct sum of the irreducible components of φ', which were not yet found, has the index $l(\varphi') - l(\varphi'_{\mathrm{I}}) - l(\varphi'_{\mathrm{II}}) = 4$ as it follows from (6), and the dimension $d(\varphi) - d(\varphi'_{\mathrm{I}}) - d(\varphi'_{\mathrm{II}}) = 11$. By inspection of Table III one sees that only the direct sum $(1, 0) + (1, 0) + (0, 0)$ has such index and dimension. Finally, the solution of our example is $\varphi' = (2, 0) + (0, 2) + 2(1, 0) + (0, 0)$.

ACKNOWLEDGMENTS

The authors are grateful to Dr. A. K. Bose and Dr. G. Paquette for several discussions and helpful comments. They acknowledge also the award of a Fellowship (J. P.) and a Studentship (A. N.) from the National Research Council of Canada, and the hospitality of the Université de Montréal.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Two-Body Scattering with Wave Packets

C. L. HAMMER AND T. A. WEBER
Institute for Atomic Research and Department of Physics, Iowa State University, Ames, Iowa

(Received 27 June 1966)

Two-body scattering is analyzed using a time-dependent formalism and wave packets of arbitrary shape as the initial wavefunctions. In particular, the case of single-channel resonant scattering is discussed in some detail. In addition to the usual amplitudes resulting from the potential term and the resonant term, a third amplitude is found which interferes with the other two. Another interesting result is that in this formalism the final state is not necessarily represented by spherical outgoing waves, but can be interpreted instead as two separate outgoing wave packets whose shape is different from the initial packets. The Wigner time delay is obtained as a natural consequence of the analysis, and the resonant cross sections are expressed as density matrices constructed from the wave packets.

1. INTRODUCTION

THERE are two limiting cases of resonant scattering which are of particular interest. One, which usually applies in atomic physics or low-energy nuclear physics, occurs when the width Γ of the resonant state is very small compared to the energy spread of the incident packets. Here the incident wave packets have a spatial spread much smaller than the resonant state, and Γ^{-1} is long compared to the interaction time. This limit can be thought of as an impulse approximation in the sense that a classical periodic system, set in motion by a force whose time duration is short compared to a period, will oscillate at the resonant frequency of the system. Thus one expects, in the quantum mechanical case, the final state wave packets to emerge from the scattering region with an average energy and momentum characteristic of the resonance.

The other limit, which applies primarily in high-energy physics, occurs when the incident packets have an energy spread much less than Γ. Here Γ^{-1} is not much larger than the interaction time, and the spatial extent of the incident packets is large compared to that of the resonant state. Consequently, the classical analog for this case is that of the driven periodic system which decays at its characteristic frequency as the driving force is removed. Therefore, one might expect interference effects which are not normally observed in low-energy phenomena.

In this paper the time-dependent formalism developed earlier[1,2] is used to investigate these two limiting cases for the particular physical circumstance when two initial wave packets interact in the center-of-momentum system via a time-independent short-range interaction $V(\mathbf{x}_1, \mathbf{x}_2)$.

A third scattering amplitude, different from the usual potential and resonant terms, is found which interferes destructively with the other terms. For small Γ this interference causes the familiar $[1 - \exp(-\Gamma t)]$ time dependence for the cross section of the scattered particles. For large Γ no interference occurs at large distances and the usual single-level Breit–Wigner cross section is obtained. However, in both cases, if a final state interaction occurs, that is, if a second scattering takes place in the neighborhood of the initial interaction, interference effects occur which tend to make the interaction look peripheral. As a consequence of the analysis, the Wigner "time delay" for the interaction is obtained and found to be in agreement with the recent work of Ohmura.[3]

Previous authors have also investigated some of the aspects of this problem. The excellent early work of Blatt and Biedenharn,[4] formalized using a time-independent phase-shift analysis, discussed the scattering of low-energy neutrons from a single resonance level of the compound nucleus. One of their results is that the scattering amplitude contains two terms, one due to potential (hard sphere) scattering and the other due to the resonance, which interfere with each other. Both of these terms are displayed in this paper in addition to the third term which, in the large Γ case, corresponds in the classical analog to the decay of the periodic system after the driving force diminishes.

Heitler[5] discusses the absorption and emission of light quanta for both limiting cases described above, but loses the transient effects by evaluating the transition probabilities only in the infinite time limit.

The work most closely resembling the presentation

[1] C. L. Hammer and T. A. Weber, J. Math. Phys. **6**, 1591 (1965).
[2] C. L. Hammer and T. A. Weber, Nuovo Cimento **37**, 88 (1965).
[3] T. Ohmura, Progr. Theoret. Phys. (Kyoto) Suppl. **29**, 108 (1964).
[4] J. M. Blatt and L. C. Biedenham, Rev. Mod. Phys. **24**, 258 (1952); see also J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley & Sons, Inc., New York, 1952), Chap. 8, Sec. 8, p. 398.
[5] W. Heitler, *Quantum Theory of Radiation* (Oxford University Press, London, 1954), 3rd ed., Sec. 20, p. 196.

here is the wave packet approach to scattering theory described in Goldberger and Watson's recent book.[6] Again, the important difference is primarily that they use an S-matrix formalism, which implies infinite time limits, rather than the U-matrix formalism, which implies large but finite times.

2. BASIC FORMALISM

Consider the elastic scattering of two particles of masses m_1 and m_2 and let the potential which describes the interaction be $V(\mathbf{x}_1, \mathbf{x}_2)$ where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the coordinates of particles 1 and 2. It is assumed that the wavefunction which describes the two particle system satisfies the differential equation

$$[H_0 + V(\mathbf{x}_1, \mathbf{x}_2)]\psi(t) = i(\partial/\partial t)\psi(t); \quad \hbar = c = 1. \tag{1}$$

If $V(\mathbf{x}_1, \mathbf{x}_2)$ is a short-range potential, there is a time t_0 before which the particles have not scattered. Thus for $t \leq t_0$, $\psi(t)$ satisfies the equation

$$H_0\psi(t) = i(\partial/\partial t)\psi(t); \quad t \leq t_0. \tag{2}$$

A formal solution to Eq. (1) for $t \geq t_0$ is

$$\psi(t) = \exp[-iH_0(t - t_0)]\psi(t_0) - i\exp(-iH_0t)\int_{t_0}^{t} d\xi \exp(iH_0\xi)V\psi(\xi). \tag{3}$$

Since H is presumed to be independent of time,

$$\psi(t) = \exp[-iH(t - t_0)]\psi(t_0), \tag{4}$$

so that Eq. (3) becomes

$$\psi(t) = \exp[-iH_0(t - t_0)]\psi(t_0) - i\exp(-iH_0t) \times \int_{t_0}^{t} d\xi\, |p', q'\rangle\langle p', q'| \exp(iH_0\xi)V \times \exp[-iH(\xi - t_0)]\, |\psi(t_0)\rangle, \tag{5}$$

where $|p', q'\rangle$ are the eigenstates of the free particle Hamiltonian H_0. Thus one has

$$H_0|p', q'\rangle = E'|p', q'\rangle, \tag{6}$$

$$E' = (p'^2 + m_1^2)^{\frac{1}{2}} + (q'^2 + m_2^2)^{\frac{1}{2}}, \tag{7}$$

$$|p', q'\rangle = (2\pi)^{-3}\exp(i\mathbf{p}' \cdot \mathbf{x}_1)\exp(i\mathbf{q}' \cdot \mathbf{x}_2)\chi_{s'}, \tag{8}$$

where $\chi_{s'}$ is a spinor which describes the spin state of the two particles. It should be noted at this point that the formalism to be developed can be extended to include all reactions of the type $A + B \to C + D$ if it is assumed that an equation of the form of Eq. (1) applies and that the free-particle wavefunction for any pairing of the four particles satisfies Eq. (6). In this event, $\chi_{s'}$ would be a spinor which depends not only on the spin but also upon other quantum numbers (isospin, strangeness, etc.) which are necessary to describe the pairing. Then all summations of the type $|p', q'\rangle\langle p', q'|$ should imply a summation over all pairings.

The time integration yields

$$\psi(t) = \exp[-H_0(t - t_0)]\psi(t_0) + |p', q'\rangle\langle p', q'|\, V \times \frac{\exp[-iH(t - t_0)] - \exp[-iE'(t - t_0)]}{H - E'}\, |\psi(t_0)\rangle. \tag{9}$$

For finite t and t_0, the integrand is analytic in the neighborhood of $H = E'$ so that $H - E'$ can be replaced by $H - E' - i\eta$ and the limit taken as $\eta \to 0$. Thus Eq. (9) can be written as

$$\psi(t) = \exp[-iH_0(t - t_0)]\psi(t_0) + \lim_{\eta\to 0} |p', q'\rangle\langle p', q'|\, V \frac{1}{H - E' - i\eta}\, |\psi(t)\rangle - \lim_{\eta\to 0} |p', q'\rangle\langle p', q'| \exp[-iE'(t - t_0)]V \times \frac{1}{H - E' - i\eta}\, |\psi(t_0)\rangle. \tag{10}$$

In terms of the stationary states of H,

$$H\varphi_{p,q} = E\varphi_{p,q}, \tag{11}$$

where

$$E = (p^2 + m_1^2)^{\frac{1}{2}} + (q^2 + m_2^2)^{\frac{1}{2}}, \tag{12}$$

$\psi(t)$ becomes

$$\psi(t) = \int d\mathbf{p}\, d\mathbf{q} A(\mathbf{p}, \mathbf{q})\varphi_{p,q} \exp(-iEt). \tag{13}$$

Here the labels p and q correspond to the momenta of the particles only in the asymptotic limit. Therefore, the initial conditions imply that for $t \leq t_0$,

$$\psi(t) = (2\pi)^{-3}\int d\mathbf{p}\, d\mathbf{q} A(\mathbf{p}, \mathbf{q}) \exp(i\mathbf{p} \cdot \mathbf{x}_1) \times \exp(i\mathbf{q} \cdot \mathbf{x}_2) \exp(-iEt)\chi_s. \tag{14}$$

Substitution into the matrix elements of Eq. (10) for $\psi(t)$ and $\psi(t_0)$ from Eqs. (13) and (14) gives

$$\psi(t) = \int d\mathbf{p}\, d\mathbf{q} A(\mathbf{p}, \mathbf{q})\varphi(t), \tag{15}$$

where

$$\varphi(t) = |p, q\rangle \exp(-iEt) + \lim_{\eta\to 0} \exp(-iEt)\, |p', q'\rangle \times \langle p', q'|\, V \frac{1}{E - E' - i\eta}\, |\varphi_{p,q}\rangle - \lim_{\eta\to 0} \exp(-iEt_0)\, |p', q'\rangle\langle p', q'| \times \exp[-iE'(t - t_0)]V \frac{1}{H - E' - i\eta}\, |p, q\rangle. \tag{16}$$

[6] M. L. Goldberger and K. M. Watson, *Collision Theory* (John Wiley & Sons, Inc., New York, 1964), p. 80. Time dependence is considered in Chap. 8 but not using wave packets.

The last term in Eq. (16) can be put in the more familiar T-matrix form by using the operator identity

$$\frac{1}{A} - \frac{1}{B} = \frac{1}{B}(B - A)\frac{1}{A},$$

with

$$A = H_0 - E' - i\eta \quad \text{and} \quad B = H - E' - i\eta$$

to express the function $(H - E' - i\eta)^{-1}$ as

$$\frac{1}{H - E' - i\eta} = \frac{1}{H_0 - E' - i\eta} + \frac{1}{E' - H + i\eta} \times V \frac{1}{H_0 - E' - i\eta}. \quad (17)$$

Substitution into Eq. (16) for this quantity gives

$$\varphi(t) = |p, q\rangle \exp(-iEt) + \lim_{\eta \to 0} \exp(-iEt)\, |p', q'\rangle \times \langle p', q'| V \frac{1}{E - E' - i\eta} |\varphi_{p,q}\rangle + \lim_{\eta \to 0} \exp(-iEt_0)\, |p', q'\rangle\langle p', q'| \times \frac{\exp[-iE'(t - t_0)]}{E' - E + i\eta} T(E')\, |p, q\rangle, \quad (18)$$

where, following Goldberger and Watson's[7] notation,

$$T(E') = V + V(E' - H + i\eta)^{-1} V. \quad (19)$$

An interesting point occurs here regarding the infinitesimal parameter η. Here, in contrast to the usual approach, the sign chosen for η is irrelevant since the integrand of Eq. (9) is analytic. With the positive choice of sign, as shown, the second term of Eq. (18) vanishes exponentially in the asymptotic limit of large $|\mathbf{x}_1 - \mathbf{x}_2|$. The third term gives the outgoing spherical waves for the pole at $E' = E - i|\eta|$ and the singularities of the T matrix. If the negative sign is chosen, as is normally done in the S-matrix formalism,[8] the contribution to the scattering cross section from the singularity at $E' = E + i|\eta|$ comes from the second term of Eq. (18). The singularities of the T matrix which appear in the third term give rise to contributions which damp to zero only in the extreme limit as $t_0 \to -\infty$ or $t \to \infty$. This establishes the connection between the S-matrix formalism and the one presented here.

Before obtaining the asymptotic limit of $\psi(t)$, it is convenient to change to variables which represent the center of energy of the system and the relative displacement of the particles. Consequently, consider the transformation of variables

$$\bar{\mathbf{x}} = \left(\frac{\bar{E}_1}{\bar{E}}\right)\mathbf{x}_1 + \left(\frac{\bar{E}_2}{\bar{E}}\right)\mathbf{x}_2, \quad (20)$$

$$\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2, \quad (21)$$

$$\mathbf{q}' = \mathbf{Q}'\left(\frac{\bar{E}_2}{\bar{E}}\right) - \mathbf{P}', \quad (22)$$

$$\mathbf{p}' = \mathbf{Q}'\left(\frac{\bar{E}_1}{\bar{E}}\right) + \mathbf{P}', \quad (23)$$

where

$$\bar{E}_1 = [\bar{P}^2 + m_1^2]^{\frac{1}{2}}, \quad (24)$$

$$\bar{E}_2 = [\bar{P}^2 + m_2^2]^{\frac{1}{2}}, \quad (25)$$

$$\bar{E} = \bar{E}_1 + \bar{E}_2. \quad (26)$$

The momentum $\bar{\mathbf{P}}$ is later chosen to be either the average momentum of the incident or scattered particles or the momentum corresponding to the center of a resonance, whichever is convenient for the approximation under study. With these substitutions, Eq. (18) becomes

$$\varphi(t) = |p, q\rangle \exp(-iEt) + \lim_{\eta \to 0} (2\pi)^{-3} \exp(-iEt) \times \sum_{s'} \int d\mathbf{P}'\, d\mathbf{Q}' \exp[i(\mathbf{P}' \cdot \mathbf{r} + \mathbf{Q}' \cdot \bar{\mathbf{x}})]\chi_{s'} \times \frac{\langle Q'(\bar{E}_1/\bar{E}) + P', Q'(\bar{E}_2/\bar{E}) - P'|\, V\, |\varphi_{p,q}\rangle}{E - E' - i\eta} + \lim_{\eta \to 0} (2\pi)^{-3} \exp(-iEt_0) \times \sum_{s'} \int d\mathbf{P}'\, d\mathbf{Q}' \exp[i(\mathbf{P}' \cdot \mathbf{r} + \mathbf{Q}' \cdot \bar{\mathbf{x}})]\chi_{s'} \times \frac{\langle Q'(\bar{E}_1/\bar{E}) + P', Q'(\bar{E}_2/\bar{E}) - P'| \times T(E') \exp[-iE'(t - t_0)]\, |p, q\rangle}{E' - E + i\eta}, \quad (27)$$

where now

$$E' = [(\mathbf{P}' + \mathbf{Q}'(\bar{E}_1/\bar{E})^2 + m_1^2]^{\frac{1}{2}} + [(\mathbf{P}' - \mathbf{Q}'(\bar{E}_2/\bar{E})^2 + m_2^2]^{\frac{1}{2}}, = E_1' + E_2'. \quad (28)$$

It is also convenient to factor the conservation of momentum delta functions from the matrix elements. Thus, let the reduced matrix element $\langle p', q'|\, (O)\, |p, q\rangle$ of an operator O be defined by

$$\langle p', q'|\, O\, |p, q\rangle = \delta(\mathbf{p}' + \mathbf{q}' - \mathbf{p} - \mathbf{q})\langle p', q'|\, (O)\, |p, q\rangle. \quad (29)$$

With this substitution the $\mathbf{Q}'$ integration in Eq. (27) can be done. The net effect is to replace $\mathbf{Q}'$ everywhere by $(\mathbf{p} + \mathbf{q})$.

[7] See Ref. 6, p. 215.
[8] See Ref. 6, p. 81, Eq. (91).

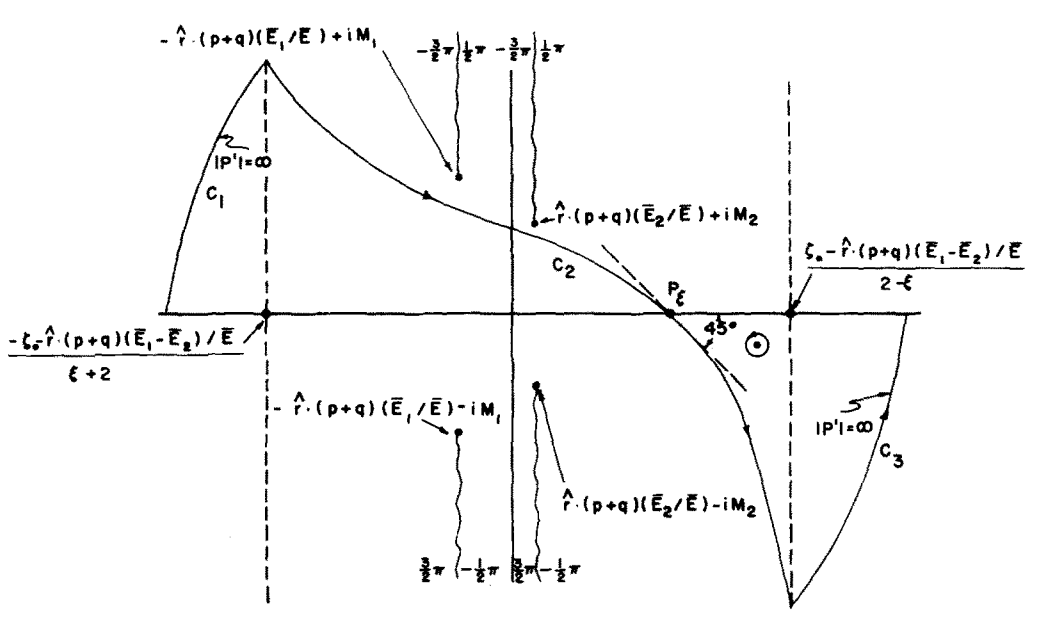


FIG. 1. The path of integration in the P' plane.

To obtain the asymptotic limit for large $|\mathbf{r}|$, the usual procedure is simplified if, wherever $f(E')\,|Q'(\bar{E}_1/\bar{E})+P', Q'(\bar{E}_2/\bar{E})-P'\rangle$ appears, it is replaced by $f(H_0)\,|Q'(\bar{E}_1/\bar{E})+P', Q'(\bar{E}_2/\bar{E})-P'\rangle$. Then if $V(\mathbf{x}_1, \mathbf{x}_2)$ falls off with sufficient rapidity so that negligible error is made by integrating over some finite region rather than over all of space, Eq. (27) becomes

$$\underset{r\to\infty}{\varphi(t)} \doteq |p, q\rangle \exp(-iEt) + \frac{i(2\pi)^{-2}}{r} \times \exp\{i[(\mathbf{p}+\mathbf{q})\cdot\bar{\mathbf{x}} - Et]\} \lim_{\eta\to 0} \sum_{s'} (I_{1s'} - I_{2s'}), \tag{30}$$

where

$$I_{1s'} = \int_{-\infty}^{\infty} P'\,dP' \exp(iP'r) \times \frac{\chi_{s'}\langle (p+q)(\bar{E}_1/\bar{E}) + P', (p+q)(\bar{E}_2/\bar{E}) - P'|(V)|\varphi_{p,q}\rangle}{E' - E + i\eta}, \tag{31}$$

$$I_{2s'} = \exp[iE(t-t_0)] \int_{-\infty}^{\infty} P'\,dP' \times \exp[i(P'\xi - E')(t-t_0)]\chi_{s'} \times \frac{\langle (p+q)(\bar{E}_1/\bar{E}) + P', (p+q)(\bar{E}_2/\bar{E}) - P'| \times (T(E'))\,|p, q\rangle}{E' - E + i\eta}, \tag{32}$$

$$\xi = r/(t-t_0), \tag{33}$$

$$\mathbf{P}' = P'(\mathbf{r}/r), \tag{34}$$

$$E' = \{[\mathbf{P}' + (\mathbf{p}+\mathbf{q})(\bar{E}_1/\bar{E})]^2 + m_1^2\}^{\frac{1}{2}} + \{[\mathbf{P}' - (\mathbf{p}+\mathbf{q})(\bar{E}_2/\bar{E})]^2 + m_2^2\}^{\frac{1}{2}}. \tag{35}$$

3. CONTOUR INTEGRATIONS

The integral $I_{1s'}$ can be done by using a contour which closes in the upper half of the P' plane, whereas the integral $I_{2s'}$ can be done by using a contour similar to that described by the authors in Refs. 1 and 2. For the purpose of clarity the procedure for obtaining this contour is briefly repeated here.

Because of the $E'(t-t_0)$ term in the exponent of the integrand of $I_{2s'}$, the contour cannot be closed in the upper P' plane alone for $(t-t_0) > 0$. Examination of the phases shows that it is possible to have an infinite contour C_1 and C_3 in the 2nd and 4th quadrants as shown in Fig. 1. The contour C_2, also shown in Fig. 1, which connects these two infinite contours, is uniquely determined by the requirement that the integral along this path be a Laplace transform so that it can be directly evaluated asymptotically without further distortion of the contour. This follows from considering the transformation suggested by the exponent in Eq. (32),

$$\zeta = P'\xi - E'. \tag{36}$$

The path in the ζ plane which corresponds to the path of integration of $I_{2s'}$ along the real P' axis is shown as the heavy line in Fig. 2 along with the appropriate phases. The branch points shown at $\pm\zeta_0$ and $\pm\zeta_1$ correspond to those values of P' which satisfy the equation $(d\zeta/dP') = 0$, that is, for

$$\xi = \frac{P' + (\bar{E}_1/\bar{E})\hat{\mathbf{r}}\cdot(\mathbf{p}+\mathbf{q})}{\{[\mathbf{P}' + (\mathbf{p}+\mathbf{q})(\bar{E}_1/\bar{E})]^2 + m_1^2\}^{\frac{1}{2}}} + \frac{P' - (\bar{E}_2/\bar{E})\hat{\mathbf{r}}\cdot(\mathbf{p}+\mathbf{q})}{\{[\mathbf{P}' - (\mathbf{p}+\mathbf{q})(\bar{E}_2/\bar{E})]^2 + m_2^2\}^{\frac{1}{2}}}. \tag{37}$$

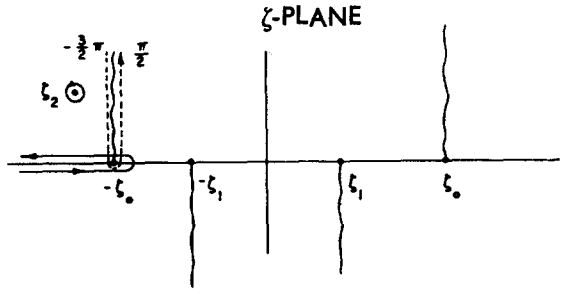


FIG. 2. The path of integration in the ζ plane.

In terms of the speeds of the two particles, it can be shown from Eqs. (36) and (37) that

$$\zeta_0 = M_1(1-\xi_1^2)^{-\frac{1}{2}}[1-\xi_1(\xi_1+\xi_2)(\bar{E}_2/\bar{E})] + M_2(1-\xi_2^2)^{-\frac{1}{2}}[1-\xi_2(\xi_1+\xi_2)(\bar{E}_1/\bar{E})], \quad (38)$$

$$\zeta_1 = M_1(1-\xi_1^2)^{-\frac{1}{2}}[1-\xi_1(\xi_1-\xi_2)(\bar{E}_2/\bar{E})] - M_2(1-\xi_2^2)^{-\frac{1}{2}}[1+\xi_2(\xi_1-\xi_2)(\bar{E}_1/\bar{E})], \quad (39)$$

where

$$\xi_1 = \frac{P' + (\bar{E}_1/\bar{E})\hat{\mathbf{r}}\cdot(\mathbf{p}+\mathbf{q})}{\{[\mathbf{P}' + (\mathbf{p}+\mathbf{q})(\bar{E}_1/\bar{E})]^2 + m_1^2\}^{\frac{1}{2}}},$$

$$\xi_2 = \frac{P' - (\bar{E}_2/\bar{E})\hat{\mathbf{r}}\cdot(\mathbf{p}+\mathbf{q})}{\{[\mathbf{P}' - (\mathbf{p}+\mathbf{q})(\bar{E}_2/\bar{E})]^2 + m_2^2\}^{\frac{1}{2}}},$$

$$M_1^2 = m_1^2 + (\bar{E}_1/\bar{E})^2\{(\mathbf{p}+\mathbf{q})^2 - [\hat{\mathbf{r}}\cdot(\mathbf{p}+\mathbf{q})]^2\},$$

$$M_2^2 = m_2^2 + (\bar{E}_2/\bar{E})^2\{(\mathbf{p}+\mathbf{q})^2 - [\hat{\mathbf{r}}\cdot(\mathbf{p}+\mathbf{q})]^2\}. \quad (40)$$

It should be noted, since the interaction takes place in the vicinity of the origin, that in the asymptotic limit in the center-of-energy system,

$$\xi_1 = \frac{|\mathbf{x}_1|}{t-t_0}, \quad \xi_2 = \frac{|\mathbf{x}_2|}{t-t_0}. \quad (41)$$

The path of integration in the ζ plane can be rotated about $-\zeta_0$ to the dotted line shown in Fig. 2, encircling poles and indenting around branch points where necessary, thereby changing the Fourier transform to a Laplace transform. The contour C_2 in the P' plane corresponds to the dotted path of integration in the ζ plane shown in Fig. 2. The point $-\zeta_0$ corresponds to the point in the P' plane where C_2 crosses the real axis. This value of P' shown as P_ξ in Fig. 1 satisfies Eq. (37). In the center-of-energy system this value of P' would be the momentum of m_1 or m_2 if the speed of separation of the particles after the collision is $r/(t-t_0)$.

The integration along C_2 can be evaluated asymptotically for large t using Watson's lemma described previously.[9] The branch points at ζ_0 and at $\pm\zeta_1$ can be shown to be sufficiently far removed from $-\zeta_0$ so that at most only the singularities of the T matrix shown symbolically at ζ_2 in Fig. 2 and the branch point at $-\zeta_0$ need be considered. In the following, the singularities of the T matrix are assumed removed far enough from $-\zeta_0$ so that in evaluating Eq. (32) along the contour C_2, expansions can be made about $-\zeta_0$. This restricts the discussion to a consideration of that region of time where the resonant states decay with an exponential type behavior as shown below. The contribution to Eq. (32) from the path along C_2 is found in a way parallel to Eq. (38) of Ref. 2. The result is to lowest order in $|t-t_0|$,

$$I_{2s'} \sim |t-t_0|^{-\frac{1}{2}}. \quad (42)$$

Since $|t-t_0|$ is related to r through Eq. (33), $(I_{2s'})_{C_2}$ is of order $r^{-\frac{1}{2}}$ and does not therefore contribute to the scattering cross section.

With this result, Eq. (30) can be reduced considerably. Firstly, it should be noted that the integrand of Eq. (30) as represented by the sum $(I_{1s'} - I_{2s'})$ cannot contain any singularities in the 2nd and 3rd quadrants of the P' plane. This follows because such singularities lead to incoming spherical waves which are a violation of the initial conditions. Thus, the simple pole which is in the second quadrant of the P' plane must cancel between the two terms $I_{1s'}$ and $I_{2s'}$ This cancellation can be seen explicitly if the substitution

$$\varphi_{p,q} = \left(1 + \frac{1}{E-H}V\right)|p,q\rangle \quad (43)$$

is made in Eq. (31). Since the contour integration for $I_{1s'}$ is in the upper-half P' plane, all other contributions from this term lead to functions which damp exponentially with r. Thus, $I_{1s'}$ does not contribute to the scattering cross section.

The time-independent [apart from a factor of $\exp(-iEt)$] part of the cross section comes from $I_{2s'}$ because of the pole at $E' - E + i\eta$ which lies in the 4th quadrant of the P' plane. As is shown in Fig. 1, this pole is included within the contour only if $P_\xi < P'$ or, equivalently, only if $(t-t_0) \geq (t-t')$, where t' is the time at which the particles elastically scatter. It therefore follows that $t' \geq t_0$ so that, as expected, causality is automatically taken into account.

The time-dependent parts of the scattering contributions can be separated from the time-independent

[9] T. A. Weber, D. M. Fradkin, and C. L. Hammer, Ann. Phys. (N.Y.) 27, 362 (1964); see also Ref. 2.

parts by using the residue theorem to write Eq. (30) as

$$\begin{aligned}\varphi(t)(\text{large } t) &= |p, q\rangle \exp(-iEt) \\ &\underset{r\to\infty}{} \\ &+ (2\pi)^{-3}r^{-1}\exp\{i[Pr - Et + (\mathbf{p}+\mathbf{q})\cdot\bar{\mathbf{x}}]\}\sum_{s'} f_{s'}\chi_{s'} \\ &- \lim_{\eta\to 0}(2\pi)^{-2}r^{-1}i\exp\{i[\mathbf{p}+\mathbf{q})\cdot\bar{\mathbf{x}} - Et_0]\} \\ &\times \sum_{s'}\int_{C_{E'\neq E}} P'\,dP'\exp[i(P'\xi - E')(t-t_0)]\chi_{s'} \\ &\langle P' + (p+q)(\bar{E}_1/\bar{E}), (p+q)(\bar{E}_2/\bar{E}) - P'| \\ &\times \frac{\times (T(E'))\,|p, q\rangle}{E' - E + i\eta},\end{aligned} \tag{44}$$

where the scattering amplitude is given by

$$\begin{aligned}f_{s'} = -(2\pi)^2\frac{E_1E_2}{\bar{E}}&\left[1 + \hat{\mathbf{r}}\cdot(\mathbf{p}+\mathbf{q})\frac{(\bar{E}_1E_2 - \bar{E}_2E_1)}{PE\bar{E}}\right]^{-1} \\ &\times \langle P + (p+q)(\bar{E}_1/\bar{E}), \\ &\times (p+q)(\bar{E}_2/\bar{E}) - P|\,(T(E))\,|p, q\rangle,\end{aligned} \tag{45}$$

$$\mathbf{P} = P\hat{\mathbf{r}}, \tag{46}$$

$$E_1 = \{[\mathbf{P} + (\mathbf{p}+\mathbf{q})(\bar{E}_1/\bar{E})]^2 + m_1^2\}^{\frac{1}{2}}, \tag{47}$$

$$E_2 = \{[\mathbf{P} - (\mathbf{p}+\mathbf{q})(\bar{E}_2/\bar{E})]^2 + m_2^2\}^{\frac{1}{2}}, \tag{48}$$

$$E = E_1 + E_2 = (p^2 + m_1^2)^{\frac{1}{2}} + (q^2 + m_2^2)^{\frac{1}{2}}, \tag{49}$$

and the subscript $E' \neq E$ means ignore the pole at $E' = E - i\eta$ in doing the indicated contour integration.

It is easy to show the resonant nature of Eq. (44) explicitly by writing $T(E')$ as

$$T(E') = V + VGV, \tag{50}$$

where G is the exact Green's function

$$G = [E' - H + i\eta]^{-1}. \tag{51}$$

Following Goldberger and Watson,[10] G can be written as the product of two operators

$$G = Fg, \tag{52}$$

where g is diagonal in the representation $|a\rangle$. As a particular example, the interaction V considered previously could be the sum of two terms

$$H - H_0 = V = U + \bar{V}, \tag{53}$$

so that

$$H - \bar{H}_0 = \bar{V}, \quad \bar{H}_0 - H_0 = U. \tag{54}$$

The states $|a\rangle$, which satisfy the equation

$$\bar{H}_0\,|a\rangle = E_a\,|a\rangle \tag{55}$$

could then be resonant states of the m_1, m_2 system that decay in the presence of the interaction $\bar{V}$. The state $|a\rangle$ can be written as

$$|a\rangle = (2\pi)^{-\frac{3}{2}}\exp(i\mathbf{P}_a\cdot\bar{\mathbf{x}})u_a(\mathbf{r}), \tag{56}$$

where P_a represents the momentum of the center of energy of the system and $u_a(\mathbf{r})$ describes the internal degrees of freedom. The energy E_a can be described by

$$E_a = [P_a^2 + m_a^2]^{\frac{1}{2}}, \tag{57}$$

where m_a is the mass of the resonant state. Since g is diagonal in this representation,

$$\langle a'|\,g\,|a\rangle = G_a\delta_{a'a}\delta(\mathbf{P}_{a'} - \mathbf{P}_a). \tag{58}$$

Further, let these diagonal elements be the diagonal elements of G so that

$$\langle a|\,(G)\,|a\rangle = G_a \quad \text{and} \quad \langle a|\,(F)\,|a\rangle = 1. \tag{59}$$

Substitution for G from Eq. (52) into Eq. (51) then gives G_a and, after some manipulation, F as

$$G_a = [E' - E_a - \langle a|\,(\bar{V}F)\,|a\rangle + i\eta]^{-1}, \tag{60}$$

$$F\,|a\rangle = |a\rangle + [E' - \bar{H}_0 + i\eta]^{-1}(1 - |a\rangle\langle a|)\bar{V}F\,|a\rangle. \tag{61}$$

The matrix element of VGV can be written as

$$\begin{aligned}&\langle p', q'|\,VGV\,|p, q\rangle \\ &\quad = \langle p', q'|\,VG\,|a\rangle\langle a|\,V\,|p, q\rangle \\ &\quad = G_a\langle p', q'|\,VF\,|a\rangle\langle a|\,V\,|p, q\rangle \\ &\quad = \frac{\delta(\mathbf{p}' + \mathbf{q}' - \mathbf{p} - \mathbf{q})\langle p', q'|\,(VF)\,|a\rangle\langle a|\,(V)\,|p, q\rangle}{E' - E_a - \langle a|\,(\bar{V}F)\,|a\rangle + i\eta},\end{aligned} \tag{62}$$

where, since the integral over $\mathbf{P}_a$ has been done,

$$\mathbf{P}_a = \mathbf{p} + \mathbf{q}, \quad E_a = [(\mathbf{p}+\mathbf{q})^2 + m_a^2]^{\frac{1}{2}}. \tag{63}$$

The reduced matrix element of $T(E')$ is

$$\begin{aligned}\langle p', q'|\,(\bar{T}(E'))\,|p, q\rangle &= \langle p', q'|\,(V)\,|p, q\rangle \\ &+ \frac{\langle p', q'|\,(VF)\,|a\rangle\langle a|\,(V)\,|p, q\rangle}{E' - E_a - \langle a|\,(\bar{V}F)\,|a\rangle + i\eta}.\end{aligned} \tag{64}$$

The real and imaginary parts of $\langle a|\,(\bar{V}F)\,|a\rangle$ can be defined (for real E') as

$$R_a(E') \equiv \langle a|\,(\bar{V}F)\,|a\rangle = D_a(E') - \tfrac{1}{2}i\Gamma_a(E'). \tag{65}$$

In general, the denominator in Eq. (64) is a very complicated function of E'. However, for the purposes of this paper, it is assumed that $|R_a| \ll E_a$ so that to lowest order

$$R_a(E') = D_a(E_a) - \tfrac{1}{2}i\Gamma_a(E_a). \tag{66}$$

[10] See Ref. 6, p. 425.

With these assumptions, Eq. (44) becomes

$$\begin{aligned}
&\underset{r\to\infty}{\varphi(t)}(\text{large } t) = |p, q\rangle \exp(-iEt) \\
&+ \frac{\exp\{i[Pr - Et + (\mathbf{p}+\mathbf{q})\cdot\bar{\mathbf{x}}]\}}{r}(2\pi)^{-3}\sum_{s'} f_{s'}\chi_{s'} \\
&- \lim_{\eta\to 0}\frac{i\exp\{i[(\mathbf{p}+\mathbf{q})\cdot\bar{\mathbf{x}} - Et_0]\}}{(2\pi)^2 r} \\
&\times \sum_{s'}\int_{C_{E'\neq E}} P'\,dP'\chi_{s'}\frac{\exp[i(P'\xi - E')(t-t_0)]}{E' - E + i\eta} \\
&\times \langle P' + (p+q)(\bar{E}_1/\bar{E}), (p+q)(\bar{E}_2/\bar{E}) - P'|(V)|p, q\rangle \\
&- \lim_{\eta\to 0}\frac{i\exp\{i[(\mathbf{p}+\mathbf{q})\cdot\bar{\mathbf{x}} - Et_0]\}}{(2\pi)^2 r} \\
&\times \sum_{s'}\int_{C_{E'\neq E}} P'\,dP'\chi_{s'}\frac{\exp[i(P'\xi - E')(t-t_0)]}{[E' - E + i\eta]} \\
&\quad \langle P' + (p+q)(\bar{E}_1/\bar{E}), (p+q)(\bar{E}_2/\bar{E}) - P'| \\
&\times \frac{\times (VF)|a\rangle\langle a|(V)|p, q\rangle}{[E' - E_a - D_a(E_a) + \frac{1}{2}i\Gamma_a(E_a) + i\eta]}.
\end{aligned} \tag{67}$$

However, the third term of Eq. (67) does not contribute to the scattering cross section since it contributes only to the incident flux. This follows directly from a consideration of the matrix element,

$$M = \langle P' + (p+q)(\bar{E}_1/\bar{E}), (p+q)(\bar{E}_2/\bar{E}) - P'|(V)|p, q\rangle.$$

Changing variables of integration according to Eqs. (20) and (21) gives, since V must be invariant to displacements,

$$M = \int d\mathbf{r}\exp[-i(\mathbf{P}' - \mathbf{P}_0)\cdot\mathbf{r}]f(\mathbf{r}), \tag{68}$$

where

$$\mathbf{P}_0 = \mathbf{p}(\bar{E}_2/\bar{E}) - \mathbf{q}(\bar{E}_1/\bar{E}).$$

A general interaction V must also be invariant to rotations. This determines that

$$f(r) = \sum_j A^j_{l,\dots m} r_l \cdots r_m g_j(r), \tag{69}$$

where $A^j_{l,\dots m}$ are tensors which contain the spin directions and, if V contains differential operators, the initial momenta. Thus,

$$\begin{aligned}
M = \sum_j A^j_{l,\dots m}(P'_l - P_{0l})\cdots(P'_m - P_{0m}) \\
\times g'(|\mathbf{P}' - \mathbf{P}_0|),
\end{aligned} \tag{70}$$

so that any singularities of M must be at the points

$$|\mathbf{P}' - \mathbf{P}_0|^2 + C^2 = 0, \tag{71}$$

where C^2 cannot be a negative real number since it is assumed that the integrand is analytic in the neighborhood of the real P' axis. The roots of Eq. (71) are

$$P' = \mathbf{P}_0\cdot\hat{\mathbf{r}} \pm i[C^2 + (\mathbf{P}_0\times\hat{\mathbf{r}})^2]^{\frac{1}{2}}. \tag{72}$$

If this singularity, with either sign, is included within the contour of Fig. 1, the result of the evaluation of the third term in Eq. (67) is proportional to

$$\begin{aligned}
r^{-1}\exp\{-|\mathrm{Im}E'|\,|t - t_0 - \mathrm{Re}\,[C^2 + (\mathbf{P}_0\times\hat{\mathbf{r}})^2]^{\frac{1}{2}} \\
\times |\mathrm{Im}\,E'|^{-1}r|\}\exp[i(\mathbf{p}\cdot\mathbf{x}_1 + \mathbf{q}\cdot\mathbf{x}_2)],
\end{aligned}$$

an exponentially damping contribution to the incident plane wave.

Similarly, since to lowest order in $\bar{V}$, $F|a\rangle \cong |a\rangle$, it follows that the singularities of the matrix element in the last term of Eq. (67), in this approximation, also contribute only to the incident plane wave. The residue theorem can then be used to write Eq. (67) as

$$\begin{aligned}
&\underset{r\to\infty}{\varphi(t)}(\text{large } t) = |p, q\rangle\exp(-iEt) \\
&+ \frac{(2\pi)^{-3}}{r}\exp\{i[Pr - Et + (\mathbf{p}+\mathbf{q})\cdot\bar{\mathbf{x}}]\} \\
&\times \sum_{s'} f_{s'}\chi_{s'} - \frac{(2\pi)^{-3}}{r}\exp\{i[(\mathbf{p}+\mathbf{q})\cdot\bar{\mathbf{x}} - Et_0]\} \\
&\times \sum_{s',a}\exp\{i[P_R r - E_R(t-t_0)]\} \\
&\times \exp[-\tfrac{1}{2}\Gamma_a(t' - t_0)]\chi_{s'}h_{s'a},
\end{aligned} \tag{73}$$

where

$$P_c = P_R + iP_I,$$

$$\begin{aligned}
\{[\mathbf{P}_c + (\mathbf{p}+\mathbf{q})(\bar{E}_1/\bar{E})]^2 + m_1^2\}^{\frac{1}{2}} \\
+ \{[\mathbf{P}_c - (\mathbf{p}+\mathbf{q})(\bar{E}_2/\bar{E})]^2 + m_2^2\}^{\frac{1}{2}} \\
= E_a + D_a(E_a) - \tfrac{1}{2}i\Gamma_a(E_a),
\end{aligned} \tag{74}$$

is the position of the pole in the P' plane,

$$E_R = E_a + D_a, \tag{75}$$

$$\begin{aligned}
h_{s',a} = -(2\pi)^2 P_c[\partial E'(P_c)/\partial P_c]^{-1} \\
\langle P_c + (p+q)(\bar{E}_1/\bar{E}), (p+q)(\bar{E}_2/\bar{E}) - P_c| \\
\times \frac{\times (VF)|a\rangle\langle a|(V)|p, q\rangle}{E - E_a - D_a + \frac{1}{2}i\Gamma_a},
\end{aligned} \tag{76}$$

$$t' = t + (2P_I/\Gamma_a)r. \tag{77}$$

The time t' is the retarded time at which the resonant state decays, since for small $\frac{1}{2}\Gamma_a$ and P_I, the Cauchy–Riemann condition gives

$$(\partial E_R/\partial P_R) = [\partial(-\tfrac{1}{2}\Gamma_a)/\partial P_I] \cong -\Gamma_a/2P_I,$$

so that

$$t' = t - [r/(\partial E_R/\partial P_R)] \tag{78}$$

or

$$t' - t_0 = [1 - \xi/(\partial E_R/\partial P_R)](t - t_0). \tag{79}$$

If $\Gamma_a > 0$, then the pole is included within the contour only for $\xi < (\partial E_R/\partial P_R)$. Equation (79) then shows that $(t' - t_0) > 0$ for $(t - t_0) > 0$. This constitutes a time delay and is therefore resonant scattering. If $\Gamma_a < 0$, then the pole is included within the contour only for $\xi > (\partial E_R/\partial P_R)$. Equation (79) then shows that $(t' - t_0) < 0$ for $(t - t_0) > 0$. This is a time advance and is therefore a case of antiresonance scattering. In both cases, the last term of Eq. (73) is exponentially damped at $(t - t_0) \to \infty$ because for $\xi \neq (\partial E_R/\partial P_R)$, $\Gamma_a(t' - t_0) \to \infty$. The remaining terms of Eq. (73) are just the usual S-matrix result. However, the last term can contribute to the scattering cross section since in a particular experiment $\frac{1}{2}\Gamma_a(t' - t_0)$ may not be large. This would be particularly apparent if Γ_a is very small or if a second scattering takes place close to the initial interaction region.

4. FINAL STATE WAVE PACKETS

Up to this point there has been no discussion of the nature of the wave packets. Since the incident particles are usually independently prepared; the wave packet amplitude $A(\mathbf{p}, \mathbf{q})$ [see Eq. (14)] is assumed to be of the form

$$A(\mathbf{p}, \mathbf{q}) = a(\mathbf{p} - \mathbf{p}_0)b(\mathbf{q} + \mathbf{p}_0), \tag{80}$$

so that initially

$$\psi(t) = (2\pi)^{-3}\int d\mathbf{p}\, d\mathbf{q} a(\mathbf{p} - \mathbf{p}_0)b(\mathbf{q} + \mathbf{p}_0) \times \exp\left[i(\mathbf{p}\cdot\mathbf{x}_1 + \mathbf{q}\cdot\mathbf{x}_2 - Et)\right]. \tag{81}$$

Also, as is usually the case physically, assume that the amplitudes $a(\mathbf{p} - \mathbf{p}_0)$ and $b(\mathbf{q} + \mathbf{p}_0)$ are sufficiently narrow that terms of order $[(\mathbf{p} - \mathbf{p}_0)^2/(p_0^2 + m_1^2)^{\frac{1}{2}}]$ and $[(\mathbf{q} + \mathbf{p}_0)^2/(p_0^2 + m_2^2)^{\frac{1}{2}}]$ can be ignored. In this approximation Eq. (81) becomes, after changing variables of integration according to

$$\boldsymbol{\tau} = \mathbf{p} - \mathbf{p}_0, \quad \boldsymbol{\sigma} = \mathbf{q} + \mathbf{p}_0,$$

$$\psi(t) \cong (2\pi)^{-3}\exp\left[i(\mathbf{p}_0\cdot\mathbf{r} - E_0 t)\right] \times \int d\boldsymbol{\tau}\, d\boldsymbol{\sigma} a(\boldsymbol{\tau})b(\boldsymbol{\sigma})\exp\left[i\boldsymbol{\tau}\cdot(\mathbf{x}_1 - \mathbf{v}_{01}t)\right] \times \exp\left[i\boldsymbol{\sigma}\cdot(\mathbf{x}_2 - \mathbf{v}_{02}t)\right], \tag{82}$$

where

$$E_0 = (p_0^2 + m_1^2)^{\frac{1}{2}} + (p_0^2 + m_2^2)^{\frac{1}{2}} = E_{01} + E_{02},$$

$$\mathbf{v}_{01} = \mathbf{p}_0/E_{01}, \quad \mathbf{v}_{02} = -\mathbf{p}_0/E_{02}. \tag{83}$$

The two independent packets therefore move toward one another with negligible change in shape, in a frame of reference where each has an average momentum p_0. The centers of the packets can be made to lie along a line perpendicular to $\mathbf{p}_0$ at $t = 0$, their distance of separation depending upon the choice of the phase of $a(\boldsymbol{\tau})b(\boldsymbol{\sigma})$. For wave packets with finite spatial extent the time t_0 can be chosen as the time that particular parts of the wave packets begin to interact. Consequently, t_0 would then be a function of the relative coordinates of the packets. For example, if the interaction distance is much smaller than the size of either wave packet, then for t in the neighborhood of zero,

$$t_0 \cong -\frac{(\mathbf{x}_1 - \mathbf{x}_2)\cdot\hat{\mathbf{v}}_{01}}{|\mathbf{v}_{01} - \mathbf{v}_{02}|} \tag{84}$$

Thus for $t_0 = 0$, the centers of the packets would begin to interact whereas the initial interaction would have taken place at $t_0 = -\delta\,|\mathbf{v}_{01} - \mathbf{v}_{02}|^{-1}$, where δ is the spatial extent of each packet. For $t > 0$, t_0 can be obtained from the definition $\xi = r(t - t_0)^{-1}$. Since t_0 is bounded in the vicinity of zero and since, as is shown below, r is bounded in the vicinity of $\langle r\rangle$, the distance between the centers of the final state packets, ξ approaches the constant value

$$\langle\xi\rangle = (\langle r\rangle/t) \tag{85}$$

in the asymptotic limit of large t. Therefore, for large t,

$$t_0 = t - (r/\langle\xi\rangle),$$

where r varies over the dimensions of the final states wave packets.

The wavefunction for the scattered particles as obtained from Eq. (73) is

$$\underset{r\to\infty}{\psi(t)}(\text{large } t) = (2\pi)^{-3}r^{-1}(\psi_1 - \psi_2), \tag{86a}$$

where

$$\psi_1 = \sum_{s'}\int d\mathbf{p}\, d\mathbf{q} a(\mathbf{p} - \mathbf{p}_0)b(\mathbf{q} + \mathbf{p}_0)f_{s'}\chi_{s'} \times \exp\left\{i[Pr - Et + (\mathbf{p} + \mathbf{q})\cdot\bar{\mathbf{x}}]\right\}, \tag{86b}$$

$$\psi_2 = \sum_{s',a}\int d\mathbf{p}\, d\mathbf{q} a(\mathbf{p} - \mathbf{p}_0)b(\mathbf{q} + \mathbf{p}_0)h_{s'a}\chi_{s'} \times \exp\left[-\tfrac{1}{2}\Gamma_a(t' - t_0)\right] \times \exp\left\{i[P_R r - E_R(t - t_0) - Et_0 + (\mathbf{p} + \mathbf{q})\cdot\bar{\mathbf{x}}]\right\}. \tag{86c}$$

To illustrate the procedure to be followed, consider first ψ_1 alone. By again changing to the $\boldsymbol{\tau}$, $\boldsymbol{\sigma}$ variables and retaining only terms linear in $\boldsymbol{\tau}$ and $\boldsymbol{\sigma}$, Eq. (86b) becomes

$$\psi_1 = \sum_{s'}\exp(-iE_0 t)\int d\boldsymbol{\tau}\, d\boldsymbol{\sigma} a(\boldsymbol{\tau})b(\boldsymbol{\sigma})f_{s'}\chi_{s'} \times \exp\left\{i[Pr + \boldsymbol{\tau}\cdot(\bar{\mathbf{x}} - \mathbf{v}_{01}t) + \boldsymbol{\sigma}\cdot(\bar{\mathbf{x}} - \mathbf{v}_{02}t)]\right\}, \tag{87}$$

where $f_{s'}\chi_{s'}$ are now functions of $\boldsymbol{\tau}$ and $\boldsymbol{\sigma}$, and from

Eqs. (47)–(49)

$$\{[\mathbf{P} + (\boldsymbol{\tau} + \boldsymbol{\sigma})(\bar{E}_1/\bar{E})]^2 + m_1^2\}^{\frac{1}{2}} + \{[\mathbf{P} - (\boldsymbol{\tau} + \boldsymbol{\sigma})(\bar{E}_2/\bar{E})]^2 + m_2^2\}^{\frac{1}{2}} = [(\boldsymbol{\tau} + \mathbf{p}_0)^2 + m_1^2]^{\frac{1}{2}} + [(\boldsymbol{\sigma} - \mathbf{p}_0)^2 + m_2^2]^{\frac{1}{2}}. \quad (88)$$

Because of the variation due to $\boldsymbol{\tau}$ and $\boldsymbol{\sigma}$, $\mathbf{P}$ will vary about some average value $\langle \mathbf{P} \rangle$. Therefore, if

$$\mathbf{P} = \langle \mathbf{P} \rangle + \Delta \mathbf{P}, \quad (89)$$

expansion of both sides of Eq. (88) gives the equation which defines $\langle \mathbf{P} \rangle$ as

$$\langle E \rangle = \langle E_1 \rangle + \langle E_2 \rangle = (\langle P \rangle^2 + m_1^2)^{\frac{1}{2}} + (\langle P \rangle^2 + m_2^2)^{\frac{1}{2}} = E_{01} + E_{02} \quad (90)$$

and

$$\Delta \mathbf{P} \cdot (\langle \mathbf{P} \rangle / |\langle \mathbf{P} \rangle|) = (\mathbf{v}_{01} \cdot \boldsymbol{\tau} + \mathbf{v}_{02} \cdot \boldsymbol{\sigma})\, |\langle \mathbf{v}_1 \rangle - \langle \mathbf{v}_2 \rangle|^{-1} - (\boldsymbol{\tau} + \boldsymbol{\sigma}) \cdot [\langle \mathbf{v}_1 \rangle (\bar{E}_1/\bar{E}) + \langle \mathbf{v}_2 \rangle (\bar{E}_2/\bar{E})]\, |\langle \mathbf{v}_1 \rangle - \langle \mathbf{v}_2 \rangle|^{-1}, \quad (91)$$

where

$$\langle \mathbf{v}_1 \rangle = \langle \mathbf{P} \rangle / \langle E_1 \rangle; \quad \langle \mathbf{v}_2 \rangle = -\langle \mathbf{P} \rangle / \langle E_2 \rangle. \quad (92)$$

The wavefunction ψ_1 then becomes

$$\psi_1 = \sum_{s'} \exp [i(\langle P \rangle r - \langle E \rangle t)] \int d\boldsymbol{\tau}\, d\boldsymbol{\sigma} a(\boldsymbol{\tau}) b(\boldsymbol{\sigma}) f_{s'} \chi_{s'} \times \exp (i\boldsymbol{\tau} \cdot \{\bar{\mathbf{x}} - [\langle \mathbf{v}_1 \rangle (\bar{E}_1/\bar{E}) + \langle \mathbf{v}_2 \rangle (\bar{E}_2/\bar{E})](t - t'_w) - \mathbf{v}_{01} t'_w\}) \times \exp (i\boldsymbol{\sigma} \cdot \{\bar{\mathbf{x}} - [\langle \mathbf{v}_1 \rangle (\bar{E}_1/\bar{E}) + \langle \mathbf{v}_2 \rangle (\bar{E}_2/\bar{E})](t_2 - t'_w) - \mathbf{v}_{02} t'_w\}), \quad (93)$$

where t'_w is the time of scattering defined by

$$t'_w = t - r\, |\langle \mathbf{v}_1 \rangle - \langle \mathbf{v}_2 \rangle|^{-1}. \quad (94)$$

Just as for t_0, t'_w varies as r varies over the dimensions of the final state packets. From the definition of $\bar{\mathbf{x}}$ from Eq. (20), it is clear that ψ_1 is a function only of $[\mathbf{x}_1 - \langle \mathbf{v}_1 \rangle (t - t'_w)]$ and $[\mathbf{x}_2 - \langle \mathbf{v}_2 \rangle (t - t'_w)]$. Therefore, ψ_1 resembles two separate wave packets, one centered about $\langle \mathbf{x}_1 \rangle = \langle \mathbf{v}_1 \rangle (t - \langle t'_w \rangle)$ and the other centered about $\langle \mathbf{x}_2 \rangle = \langle \mathbf{v}_2 \rangle (t - \langle t'_w \rangle)$, where $\langle t'_w \rangle$ is the value of t'_w when the centers of the packets are at the origin. If the choice $\langle \bar{\mathbf{P}} \rangle = \langle \mathbf{P} \rangle$ is made, then from Eq. (90), $|\mathbf{p}_0| = |\langle \mathbf{P} \rangle|$ and the center of energy is stationary, that is $\dot{\bar{\mathbf{x}}} = 0$ before $t = 0$ and after $t = \langle t'_w \rangle$.

The time $\langle t'_w \rangle$ can be obtained by using a method originated by Wigner.[11] This method in essence compares the phases of the final wave packets evaluated at their centers to the phases of the initial wave packets evaluated at their centers. If

$$f_{s'} = |f_{s'}| \exp (i\alpha), \quad (95)$$

then

$$\alpha(E) \cong \alpha(E_0) + [\partial \alpha(E_0)/\partial E_0](\boldsymbol{\tau} \cdot \mathbf{v}_{01} + \boldsymbol{\sigma} \cdot \mathbf{v}_{02}), \quad (96)$$

and the final state wave packets at $t = \langle t'_w \rangle$

$$\langle t'_w \rangle = \partial \alpha(E_0)/\partial E_0 \quad (97)$$

are at the same position as that of the initial wave packets at $t = 0$. This asymptotic time delay (or advance) is in agreement with the recent work of Ohmura.[3]

The expression for ψ_2 can be reduced in a manner similar to that used for ψ_1. The difference being that instead of Eq. (88), the conservation of energy equation as obtained from Eq. (74) is, in the $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ variables,

$$\{[\mathbf{P}_c + (\boldsymbol{\sigma} + \boldsymbol{\tau})(\bar{E}_1/\bar{E})]^2 + m_1^2\}^{\frac{1}{2}} + \{[\mathbf{P}_c - (\boldsymbol{\sigma} + \boldsymbol{\tau})(\bar{E}_2/\bar{E})]^2 + m_2^2\}^{\frac{1}{2}} = [(\boldsymbol{\sigma} + \boldsymbol{\tau})^2 + m_a^2]^{\frac{1}{2}} + D_a\{[(\boldsymbol{\sigma} + \boldsymbol{\tau})^2 + m_a^2]^{\frac{1}{2}}\} - \tfrac{1}{2} i \Gamma_a \{[(\boldsymbol{\sigma} + \boldsymbol{\tau})^2 + m_a^2]^{\frac{1}{2}}\}. \quad (98)$$

The zeroth- and first-order terms in $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ thus are

$$[\langle \mathbf{P}_c \rangle^2 + m_1^2]^{\frac{1}{2}} + [\langle \mathbf{P}_c \rangle^2 + m_2^2]^{\frac{1}{2}} = m_a + D_a(m_a) - \tfrac{1}{2} i \Gamma_a(m_a) \quad (99)$$

and

$$\Delta \mathbf{P}_c \cdot (\langle \mathbf{P}_c \rangle / |\langle \mathbf{P}_c \rangle|) = -(\boldsymbol{\sigma} + \boldsymbol{\tau}) \cdot [\langle \mathbf{v}_{1c} \rangle (\bar{E}_1/\bar{E}) + \langle \mathbf{v}_{2c} \rangle (\bar{E}_2/\bar{E})] \times |\langle \mathbf{v}_{1c} \rangle - \langle \mathbf{v}_{2c} \rangle|^{-1}, \quad (100)$$

where

$$\mathbf{P}_c = \langle \mathbf{P}_c \rangle + \Delta \mathbf{P}_c, \quad \langle \mathbf{v}_{1c} \rangle = \langle \mathbf{P}_c \rangle / [\langle \mathbf{P}_c \rangle^2 + m_1^2]^{\frac{1}{2}}, \quad \langle \mathbf{v}_{2c} \rangle = -\langle \mathbf{P}_c \rangle / [\langle \mathbf{P}_c \rangle^2 + m_2^2]^{\frac{1}{2}}. \quad (101)$$

The expression for ψ_2 is more complicated than that for ψ_1 since $\langle \mathbf{v}_{1c} \rangle$ and $\langle \mathbf{v}_{2c} \rangle$ are complex quantities. However, in the limit $[\Gamma_a(m_a)/E_R] \ll 1$, which covers most physical cases, products such as $|\boldsymbol{\sigma}|(\Gamma_a/E_R)$ or $|\boldsymbol{\tau}|(\Gamma_a/E_R)$ can be neglected. Thus, Eq. (100) becomes

$$\Delta \mathbf{P}_c \cdot (\langle \mathbf{P}_c \rangle / |\langle \mathbf{P}_c \rangle|) = -(\boldsymbol{\sigma} + \boldsymbol{\tau}) \cdot [\langle \mathbf{v}_{1c} \rangle_R (\bar{E}_1/\bar{E}) + \langle \mathbf{v}_{2c} \rangle_R (\bar{E}_2/\bar{E})] \times |\langle \mathbf{v}_{1c} \rangle_R - \langle \mathbf{v}_{2c} \rangle_R|^{-1}, \quad (102)$$

where use has been made of the expressions which are correct to order (Γ_a/E_R),

$$\langle \mathbf{v}_c \rangle = \langle \mathbf{v}_c \rangle_R + i \langle \mathbf{v}_c \rangle_I, \quad (103)$$

$$\langle \mathbf{v}_{1c} \rangle_R = (\langle \mathbf{P}_R \rangle / \langle E_{1R} \rangle), \quad \langle \mathbf{v}_{2c} \rangle_R = -(\langle \mathbf{P}_R \rangle / \langle E_{2R} \rangle), \quad (104)$$

$$\langle \mathbf{v}_{1c} \rangle_I = (\langle \mathbf{P}_I \rangle / \langle E_{1R} \rangle) + \tfrac{1}{2} [\Gamma_a(m_a)/\langle E_R \rangle] \times (\langle \mathbf{P}_R \rangle / \langle E_{1R} \rangle)(\langle E_{2R} \rangle / \langle E_{1R} \rangle), \quad (105)$$

[11] E. P. Wigner, Phys. Rev. 98, 145 (1955).

$$\langle \mathbf{v}_{2c}\rangle_I = -(\langle \mathbf{P}_I\rangle/\langle E_{2R}\rangle) - \tfrac{1}{2}[\Gamma_a(m_a)/\langle E_R\rangle] \times (\langle \mathbf{P}_R\rangle/\langle E_{2R}\rangle)(\langle E_{1R}\rangle/\langle E_{2R}\rangle), \quad (106)$$

$$\langle \mathbf{P}_R\rangle\langle \mathbf{P}_I\rangle = -\tfrac{1}{2}\Gamma_a(m_a)(\langle E_{1R}\rangle\langle E_{2R}\rangle/\langle E_R\rangle), \quad (107)$$

$$\langle E_R\rangle = \langle E_{1R}\rangle + \langle E_{2R}\rangle = [(\langle \mathbf{P}_R\rangle)^2 + m_1^2]^{\frac{1}{2}} + [(\langle \mathbf{P}_R\rangle)^2 + m_2^2]^{\frac{1}{2}} = m_a + D_a(m_a). \quad (108)$$

With these assumptions ψ_2 becomes

$$\psi_2 = \sum_{s',a} \exp\{i[\langle P_R\rangle r - \langle E_R\rangle(t - t_0) - E_0 t_0]\} \times \exp[-\tfrac{1}{2}\Gamma_a(m_a)(t' - t_0)] \times \int d\tau\, d\sigma a(\tau) b(\sigma) h_{s'a}\chi_{s'} \times \exp\big(i\tau \cdot \{\bar{\mathbf{x}} - [\langle \mathbf{v}_{1c}\rangle_R(\bar{E}_1/\bar{E}) + \langle \mathbf{v}_{2c}\rangle_R(\bar{E}_2/\bar{E})](t - t') - \mathbf{v}_{01}t_0\}\big) \times \exp\big(i\sigma \cdot \{\bar{\mathbf{x}} - [\langle \mathbf{v}_{1c}\rangle_R(\bar{E}_1/\bar{E}) + \langle \mathbf{v}_{2c}\rangle_R(\bar{E}_2/\bar{E})](t - t') - \mathbf{v}_{02}t_0\}\big), \quad (109)$$

where from Eq. (78) in this approximation

$$t' = t - r\,|\langle \mathbf{v}_{1c}\rangle_R - \langle \mathbf{v}_{2c}\rangle_R|^{-1}. \quad (110)$$

From the definition of $\bar{\mathbf{x}}$ from Eq. (20), it is clear from the exponents of Eq. (109) that ψ_2 is a function only of $[\mathbf{x}_1 - \langle \mathbf{v}_{1c}\rangle_R(t - t')]$ and $[\mathbf{x}_2 - \langle \mathbf{v}_{2c}\rangle_R(t - t')]$. Therefore, ψ_2 represents a state which decays at

$$\langle t'\rangle = t - \langle r\rangle\,|\langle \mathbf{v}_{1c}\rangle_R - \langle \mathbf{v}_{2c}\rangle_R|^{-1}, \quad (111)$$

into two separate wave packets, one centered about $\langle \mathbf{x}_2\rangle = \langle \mathbf{v}_{2c}\rangle_R(t - \langle t'\rangle)$ and the other centered about $\langle \mathbf{x}_1\rangle = \langle \mathbf{v}_{1c}\rangle_R(t - \langle t'\rangle)$. Again at $t = \langle t'\rangle$ the final state wave packets are at the same positions as those of the initial wave packets at $t = 0$.

Just as for $\langle t'_w\rangle$, the time $\langle t_0\rangle$,

$$\langle t_0\rangle = t - \langle r\rangle/\langle \xi\rangle, \quad (112)$$

can be obtained from

$$h_{s'a} = |h_{s'a}| \exp(i\beta), \quad (113)$$

$$\beta(E, E_a) \cong \beta(E_0, m_a) + [\partial\beta(E_0, m_a)/\partial E_0] \times (\mathbf{v}_{01} \cdot \tau + \mathbf{v}_{02} \cdot \sigma). \quad (114)$$

This gives $\langle t_0\rangle$ as

$$\langle t_0\rangle = \partial\beta(E_0, m_a)/\partial E_0\,. \quad (115)$$

It is clear from Eq. (84) that this expression for $\langle t_0\rangle$ is valid only if

$$\partial\beta/\partial E_0 \ll 2\delta/|\mathbf{v}_{01} - \mathbf{v}_{02}|. \quad (116)$$

The expressions for ψ_1 and ψ_2 can be further simplified if the spatial extent of the initial wave packets is much larger than the interaction region. Consequently, to lowest order in τ and σ, the matrix elements in $f_{s'}$ and $h_{s'a}$ can be factored from under the integral sign, giving

$$\psi_1 = \sum_{s'} \exp[i(p_0 r - E_0 t)] f_{s'B}\chi_{s'} \times \int d\tau\, d\sigma a(\tau) b(\sigma) \exp[i\tau \cdot (\bar{\mathbf{x}} - \mathbf{v}_{01}t'_w)] \times \exp[i\sigma \cdot (\bar{\mathbf{x}} - \mathbf{v}_{02}t'_w)] + \sum_{s'a} \exp[i(p_0 r - E_0 t)] M_{s'a}(\langle \hat{\mathbf{P}}\rangle, \hat{\mathbf{p}}_0)\chi_{s'} \times \int \frac{d\tau\, d\sigma a(\tau) b(\sigma) \exp[i\tau \cdot (\bar{\mathbf{x}} - \mathbf{v}_{01}t'_w)] \times \exp[i\sigma \cdot (\bar{\mathbf{x}} - \mathbf{v}_{02}t'_w)]}{E - m_a - D_a(E) + \tfrac{1}{2}i\Gamma_a(E)}, \quad (117a)$$

$$\psi_2 = \sum_{s'a} \exp\{i[\langle P_R\rangle r - \langle E_R\rangle(t - t_0) - E_0 t_0]\} \times \exp[-\tfrac{1}{2}\Gamma_a(m_a)(t' - t_0)] \times M_{s'a}(\langle \hat{\mathbf{P}}_c\rangle, \hat{\mathbf{p}}_0)\chi_{s'} \int d\tau\, d\sigma a(\tau) b(\sigma) \times \exp\big(i\tau \cdot \{\bar{\mathbf{x}} - [\langle \mathbf{v}_{1c}\rangle_R(E_{01}/E_0) + \langle \mathbf{v}_{2c}\rangle_R \times (E_{02}/E_0)](t - t') - \mathbf{v}_{01}t_0\}\big) \times \frac{\exp\big(i\sigma \cdot \{\bar{\mathbf{x}} - [\langle \mathbf{v}_{1c}\rangle_R(E_{01}/E_0) + \langle \mathbf{v}_{2c}\rangle_R(E_{02}/E_0)](t - t') - \mathbf{v}_{02}t_0\}\big)}{E - m_a - D_a(m_a) + \tfrac{1}{2}i\Gamma_a(m_a)}, \quad (117b)$$

where Eq. (19) has been used to write $f_{s'}$ as two terms with

$$f_B = -(2\pi)^2(E_{01}E_{02}/E_0) < \langle P\rangle, -\langle P\rangle|\,(V)\,|p_0, -p_0 > \quad (118)$$

as the Born approximation term, and with

$$M_{s'a}(\langle \hat{\mathbf{P}}\rangle, \hat{\mathbf{p}}_0) = -(2\pi)^2 \frac{\langle E_1\rangle\langle E_2\rangle}{\langle E\rangle} < \langle \mathbf{P}\rangle, -\langle \mathbf{P}\rangle|\,(VF)\,|a\rangle\langle a|\,(V)\,|p_0, -p_0 >; \quad (119)$$

as the matrix elements of the resonant terms. Also the choice $|\hat{\mathbf{P}}| = |\mathbf{p}_0| = |\langle \mathbf{P}\rangle|$ has been made. For initial energies near the resonance, the rate of change of the phases α and β is primarily due to the resonant denominators. Therefore, ignoring the change of phase due to the matrix elements $M_{s'a}$ and considering only the change of phase from the resonant denominators, Eqs. (97) and (115) become

$$\langle t'_w\rangle = \frac{\tfrac{1}{2}\Gamma_a(E_0)[1 - \partial D_a(E_0)/\partial E_0] - [E_0 - m_a - D_a(E_0)][\partial\Gamma_a(E_0)/2\partial E_0]}{[E_0 - m_a - D_a(E_0)]^2 + [\tfrac{1}{2}\Gamma_a(E_0)]^2} \quad (120)$$

and

$$\langle t_0\rangle = \tfrac{1}{2}\Gamma_a(m_a)\{[E_0 - m_a - D_a(m_a)]^2 + [\tfrac{1}{2}\Gamma_a(m_a)]^2\}^{-1}. \quad (121)$$

If, as is usually assumed in similar calculations but which may be unwarranted,

$$\Gamma_a(E_0) = \Gamma_a(m_a) = \text{const},$$
$$D_a(E_0) = D_a(m_a) = \text{const},$$

then $\langle t'_w \rangle = \langle t_0 \rangle$ and $\langle t'_w \rangle$ is the Wigner time delay.[3] This is analogous to the result obtained from the differential equation with constant coefficients for a damped driven resonant cavity. Here ψ_1 represents a pulse corresponding to the driving frequency which leaves the resonance region at $\langle t'_w \rangle$, and ψ_2 represents the subsequent decay (since $\langle t'_w \rangle = \langle t_0 \rangle$) of the resonance following the departure of the driving force.

If Γ_a is very small, $(\partial\alpha/\partial E_0)$ and $(\partial\beta/\partial E_0)$ become very large. Consequently, the approximate expansions for the phases α and β become invalid and $\langle t'_w \rangle$ and $\langle t_0 \rangle$ are no longer sharply defined. Ohmura[3] has shown that for such a case $\langle t'_w \rangle$ is of the order Γ_a^{-1} so that in contrast to the near equality of Eqs. (120) and (121)

$$(\langle t_0 \rangle / \langle t'_w \rangle) \sim (\delta\Gamma_a) \ll 1. \tag{122}$$

5. CROSS SECTIONS AND LIMITING CASES

The probability that one of the outgoing wave packets is in a particular counter volume V, very large compared to the volume of the wave packets, is [see Eq. (86a)]

$$P(t) = (2\pi)^{-6} \int_V d\mathbf{x}_1 \int d\mathbf{x}_2 r^{-2} \operatorname{Tr} \rho_f(\mathbf{x}_1; \mathbf{x}_2), \tag{123}$$

where the trace refers to the spinor components and $\rho_f(\mathbf{x}_1; \mathbf{x}_2)$ are the diagonal elements of the density matrix for the outgoing particle

$$\rho_f(\mathbf{x}_1; \mathbf{x}_2 : \mathbf{x}'_1; \mathbf{x}'_2) = \sum_s \int_{w_1}^{w_2} dE_0 W_s(E_0)(\psi_1 - \psi_2)(\psi_1 - \psi_2)^\dagger. \tag{124}$$

The density matrix for the initial particle in the spin state s is then

$$\rho_{si} = \int_{w_1}^{w_2} dE_0 W_s(E_0) \psi(t_0) \psi^\dagger(t_0), \tag{125}$$

where

$$\rho_i(\mathbf{x}_1; \mathbf{x}_2 : \mathbf{x}'_1; \mathbf{x}'_2) = \sum_s \rho_{si}. \tag{126}$$

If $\psi(t_0)$ is chosen to be an orthonormal set of spin states, then Eq. (125) implies

$$\int d\mathbf{x}_1 \int d\mathbf{x}_2 \operatorname{Tr} \rho_i(\mathbf{x}_1; \mathbf{x}_2) = \sum_s \int_{w_1}^{w_2} dE_0 W_s(E_0) = 1. \tag{127}$$

In general, in the calculation of $P(t)$ there is interference between the f_B term and the other terms of Eq. (117), as pointed out in Blatt and Weisskopf,[4] unless f_B is small in comparison to the other terms. For simplicity this assumption is made along with the assumption that there is only one state $|a\rangle$. The result is, after changing to the $\mathbf{r}$ and $\bar{\mathbf{x}}$ variables,

$$\begin{aligned} P(t) = {} & \sum_{ss'} d\Omega (2\pi)^{-2} \int d\boldsymbol{\tau} \int d\boldsymbol{\sigma} \int d\boldsymbol{\tau}'\, d\boldsymbol{\sigma}' a^\dagger(\boldsymbol{\tau}') a(\boldsymbol{\tau}) \\ & \times b^\dagger(\boldsymbol{\sigma}') b(\boldsymbol{\sigma}) \delta(\boldsymbol{\tau} + \boldsymbol{\sigma} - \boldsymbol{\tau}' - \boldsymbol{\sigma}') \delta[(\boldsymbol{\tau} - \boldsymbol{\tau}') \cdot \hat{\mathbf{v}}_{01}] \\ & \times \int_{w_1}^{w_2} dE_0 W_s(E_0) \Bigg(\frac{|M_{s'a}(\langle \hat{P} \rangle, \hat{p}_0)|^2}{[E - m_a - D_a(E)]^2 + [\frac{1}{2}\Gamma_a(E)]^2} \\ & \quad \exp[-\Gamma_a(\langle t' \rangle - \langle t_0 \rangle)] \, |M_{s'a}(\langle \hat{P}_c \rangle, \hat{p}_0)|^2 \\ & + \frac{\times \Theta(P_R - P_\xi + P_I)}{(E - \langle E_R \rangle)^2 + [\frac{1}{2}\Gamma_a(m_a)]^2} \\ & \times \frac{\langle \xi \rangle}{|\mathbf{v}_{01} - \mathbf{v}_{02}|} \Bigg) - 2 \operatorname{Re} \sum_{ss'} d\Omega (2\pi)^{-2} \int d\boldsymbol{\tau}\, d\boldsymbol{\sigma} \\ & \times \int d\boldsymbol{\tau}'\, d\boldsymbol{\sigma}' a^\dagger(\boldsymbol{\tau}') a(\boldsymbol{\tau}) b^\dagger(\boldsymbol{\sigma}') b(\boldsymbol{\sigma}) \delta(\boldsymbol{\tau} + \boldsymbol{\sigma} - \boldsymbol{\tau}' - \boldsymbol{\sigma}') \\ & \times \int_{w_1}^{w_2} dE_0 W_s(E_0) \exp[-\tfrac{1}{2}\Gamma_a(\langle t' \rangle - \langle t_0 \rangle)] \\ & \quad \times \frac{M^\dagger_{s'a}(\langle \hat{P}_c \rangle, \hat{p}_0) M_{s'a}(\langle \hat{P} \rangle, \hat{p}_0)}{[E(\boldsymbol{\tau}', \boldsymbol{\sigma}') - m_a - D_a(m_a) - \frac{1}{2}i\Gamma_a(m_a)]} \\ & \times [E(\boldsymbol{\tau}, \boldsymbol{\sigma}) - m_a - D_a(E) + \tfrac{1}{2}i\Gamma_a(E)]^{-1} \\ & \times \Theta(P_R - P_\xi + P_I) \delta(P - P_R + (\langle E_R \rangle - E)\langle \xi \rangle^{-1} \\ & + (\boldsymbol{\tau} - \boldsymbol{\tau}') \cdot (\mathbf{v}_{01} - \mathbf{v}_{02}) \langle \xi \rangle^{-1}) \exp\{ i[(P - P_R)\langle r \rangle \\ & + (\langle E_R \rangle - E)(t - \langle t_0 \rangle) + (\boldsymbol{\tau} - \boldsymbol{\tau}') \cdot (\mathbf{v}_{01} - \mathbf{v}_{02}) \langle t_0 \rangle] \}, \end{aligned} \tag{128}$$

where

$$E = E_0 + \boldsymbol{\tau} \cdot \mathbf{v}_{01} + \boldsymbol{\sigma} \cdot \mathbf{v}_{02}$$

and $d\Omega$ is the solid angle of the detector in the center-of-energy system. The step function $\Theta(P_R - P_\xi + P_I)$ is necessary because the terms involving ψ_2 are zero unless the contour shown in Fig. 1 includes the resonance singularity. The singularity is excluded when P_ξ exceeds the value where the singularity lies on the contour C_2. For small (Γ_a/E_R) the pole is near the real P' axis where the contour is essentially a 45° straight line. Thus for the singularity to be included within the contour,

$$P_\xi \leq P_R + P_I = P_R - \tfrac{1}{2}\Gamma_a \, |\langle \mathbf{v}_{1c} \rangle_R - \langle \mathbf{v}_{2c} \rangle_R|^{-1}. \tag{129}$$

In terms of the density matrix, the number of collisions/cm^2 per incident particle N is given by the overlap integral of $\rho_{si}(xyz; xyz_1)$ over a plane perpendicular to $\hat{\mathbf{v}}_{01}$ for all $\mathbf{z} = z\hat{\mathbf{v}}_{01}$ and $\mathbf{z}_1 = z_1\hat{\mathbf{v}}_{01}$. That is, N is given by

$$\begin{aligned} N &= \sum_s \int dz_1 \int d\mathbf{x} \operatorname{Tr} \rho_{si}(xyz; xyz_1) \\ &= (2\pi)^{-2} \int d\boldsymbol{\tau} \int d\boldsymbol{\sigma} \int d\boldsymbol{\tau}'\, d\boldsymbol{\sigma}' a^\dagger(\boldsymbol{\tau}') a(\boldsymbol{\tau}) b^\dagger(\boldsymbol{\sigma}') \\ & \quad \times b(\boldsymbol{\sigma}) \delta(\boldsymbol{\tau} + \boldsymbol{\sigma} - \boldsymbol{\tau}' - \boldsymbol{\sigma}') \delta[(\boldsymbol{\tau} - \boldsymbol{\tau}') \cdot \hat{\mathbf{v}}_{01}], \end{aligned} \tag{130}$$

where, for simplicity, it has been assumed that $a(\tau)$ and $b(\sigma)$ are independent of s. The cross section is then obtained from Eq. (128) as

$$d\sigma = P(t)N^{-1}. \tag{131}$$

The usual S-matrix result is obtained from Eq. (128) by allowing $t \to \infty$. This gives the Breit–Wigner single-level formula

$$\frac{d\sigma}{d\Omega} = \sum_{ss'} \int_{w_1}^{w_2} dE_0 \frac{W_s(E_0)\,|M_{s'a}(\langle \hat{P}\rangle, \hat{p}_0)|^2}{[E - m_a - D_a(E)]^2 + [\frac{1}{2}\Gamma_a(E)]^2} \tag{132}$$

as expected.

For small $\Gamma_a(\langle t'\rangle - \langle t_0\rangle)$ there are several limiting cases depending upon the momentum spread of the wave packets and the energy spread of $W_s(E_0)$ relative to Γ_a. If $\Gamma_a \ll (1/\delta)$, then from Eq. (122) it follows that $P_\xi < p_0$ so that the elastic pole is included within the original contour when the resonance pole is included. Therefore for this case, the step function in Eq. (128) can be taken as one. For energies near the resonance

$$P - P_R + (\langle E_R\rangle - E)\langle\xi\rangle^{-1} \sim \Gamma_a, \qquad \Gamma_a \ll (1/\delta); \tag{133a}$$

$$\sim \Gamma_a^2\langle\xi\rangle^{-1}\left[\frac{m_1^2}{(P^2 + m_1^2)^{\frac{3}{2}}} + \frac{m_2^2}{(P^2 + m_2^2)^{\frac{3}{2}}}\right], \qquad \Gamma_a \gg (1/\delta). \tag{133b}$$

In the limit of small Γ_a this term can be ignored when compared to $(\tau - \tau')\cdot(\mathbf{v}_{01} - \mathbf{v}_{02})\langle\xi\rangle^{-1}$. Also, $W_s(E_0)$ and the matrix elements in Eq. (128) can be considered as slowly varying in the vicinity of the resonance. For w_1 and w_2 bracketing the resonance energy, the remaining integrals are to lowest order

$$\int_{w_1}^{w_2} dE_0 \frac{1}{(E - \langle E_R\rangle)^2 + [\frac{1}{2}\Gamma_a(m_a)]^2} \cong 2\pi\Gamma_a^{-1}(m_a), \tag{134}$$

$$\int_{w_1}^{w_2} dE_0 \frac{\exp[(P - P_R)\langle r\rangle + (\langle E_R\rangle - E)(t - \langle t_0\rangle)]}{(E - \langle E_R\rangle)^2 + [\frac{1}{2}\Gamma_a(m_a)]^2} \cong 2\pi\Gamma_a^{-1}(m_a)\exp[-\tfrac{1}{2}\Gamma_a(m_a)(\langle t'\rangle - \langle t_0\rangle)]. \tag{135}$$

Subsitution for these integrals into Eq. (128) gives the familiar form for the cross section

$$\frac{d\sigma}{d\Omega} = \sum_{ss'} W_s(\langle E_R\rangle)2\pi\Gamma_a^{-1}(m_a)\,|M_{s'a}(\langle \hat{P}_R\rangle, \hat{p}_0)|^2 \times \{1 - \exp[-\Gamma_a(m_a)\langle t'\rangle]\}, \tag{136}$$

where to the lowest order in Γ_a,

$$\langle\xi\rangle \sim |\langle\mathbf{v}_{1c}\rangle_R - \langle\mathbf{v}_{2c}\rangle_R| \sim |\mathbf{v}_{01} - \mathbf{v}_{02}|,$$

and where $\langle t'\rangle$ is the retarded time or time at which the resonant state decays and $\langle t_0\rangle$ has been neglected since

$$\langle t'\rangle \gg (\delta/|\langle\mathbf{v}_{1c}\rangle_R - \langle\mathbf{v}_{2c}\rangle_R|).$$

Another limiting case occurs when $\Gamma_a \gg (1/\delta)$. Equations (120) and (121) now apply and in the asymptotic limit $P_\xi \to p_0$. Therefore from Eq. (129),

$$p_0 \leq P_R - \tfrac{1}{2}\Gamma_a\,|\langle\mathbf{v}_{1c}\rangle_R - \langle\mathbf{v}_{2c}\rangle_R|^{-1}, \tag{137}$$

or equivalently

$$E_0 \leq E_R - \tfrac{1}{2}\Gamma_a, \tag{138}$$

so that the terms of Eq. (128) that arise from ψ_2 are zero when p_0 exceeds the value where the resonance singularity lies on the contour C_2. Also in this limit Eq. (133b) applies, but $(\tau - \tau')\cdot(\mathbf{v}_{01} - \mathbf{v}_{02})\langle\xi\rangle^{-1}$ is now neglected when compared to $[P - P_R + (\langle E_R\rangle - E)\langle\xi\rangle^{-1}]$. The interference term therefore vanishes and

$$\frac{d\sigma}{d\Omega} = \sum_{ss'} \int_{w_1}^{w_2} dE W_{0s}(E_0) \{|M_{s'a}(\langle \hat{P}\rangle, \hat{p}_0)|^2 + |M_{s'a}(\langle \hat{P}_c\rangle, \hat{p}_0)|^2 \times \Theta(P_R - p_0 + P_I)\exp[-\Gamma_a(\langle t'\rangle - \langle t_0\rangle)]\} \times \frac{}{(E - \langle E_R\rangle)^2 + [\frac{1}{2}\Gamma_a(m_a)]^2}. \tag{139}$$

For large t Eq. (111) can be expressed as

$$\langle t'\rangle - \langle t_0\rangle \cong \left[1 - \frac{|\langle\mathbf{v}_1\rangle - \langle\mathbf{v}_2\rangle|}{|\langle\mathbf{v}_{1c}\rangle_R - \langle\mathbf{v}_{2c}\rangle_R|}\right]\frac{\langle r\rangle}{|\langle\mathbf{v}_1\rangle - \langle\mathbf{v}_2\rangle|}. \tag{140}$$

Since from Eq. (137) it is apparent that $|\langle\mathbf{v}_1\rangle - \langle\mathbf{v}_2\rangle| < |\langle\mathbf{v}_{1c}\rangle_R - \langle\mathbf{v}_{2c}\rangle_R|$, it is clear that ψ_2 or the last term of Eq. (139) can contribute only over a limited range of $\langle r\rangle$. This range can be estimated using the fact that the smallest value for $\langle t'\rangle - \langle t_0\rangle$ occurs when both the singularity at $E' = E$ and $E' = E_a + D_a - i\frac{1}{2}\Gamma_a$ lie on the contour C_2. Equation (137) then can be used to show that

$$|\langle\mathbf{v}_{1c}\rangle_R - \langle\mathbf{v}_{2c}\rangle_R| - |\langle\mathbf{v}_1\rangle - \langle\mathbf{v}_2\rangle| = \frac{\frac{1}{2}\Gamma_a}{|\langle\mathbf{v}_1\rangle - \langle\mathbf{v}_2\rangle|}\left[\frac{m_1^2}{(p_0^2 + m_1^2)^{\frac{3}{2}}} + \frac{m_2^2}{(p_0^2 + m_2^2)^{\frac{3}{2}}}\right]. \tag{141}$$

Consequently, $\Gamma_a(\langle t'\rangle - \langle t_0\rangle)$ can be expressed as

$$\Gamma_a(\langle t'\rangle - \langle t_0\rangle) = \frac{\Gamma_a^2(m_a)\langle r\rangle}{2\,|\langle\mathbf{v}_1\rangle - \langle\mathbf{v}_2\rangle|^3} \times \left[\frac{m_1^2}{(p_0^2 + m_1^2)^{\frac{3}{2}}} + \frac{m_2^2}{(p_0^2 + m_2^2)^{\frac{3}{2}}}\right]. \tag{142}$$

In order that Eq. (117) be consistent with the fact that various $(1/\langle r\rangle)$ contributions have been neglected in deriving the scattering solution, the exponential terms must be of order one where these contributions

TABLE I. Values for $\langle r\rangle$ in units of m_π^{-1} such that $\Gamma_a(\langle t'\rangle - t_0) = 1$.

Resonance	Mass MeV	Decay mode	$\langle r\rangle$ m_π^{-1}	Γ_a MeV
$N^*_{3/2}$	1236	π, P	13	120
Y^*	1405	π, Σ	41	35
ρ	765	π, π	171	124
ϕ	1019	K, K	68	3.3
f	1253	π, π	1092	118

are negligible. It is clear from Eq. (142) that this condition is always satisfied in the high-energy limit $p_0 \to \infty, p_0 < \langle P_R\rangle$.

In Table I, the values of $\langle r\rangle$ corresponding to $\Gamma_a(\langle t'\rangle - \langle t_0\rangle) = 1$ are given for various high-energy resonances. Whereas retaining the exponential terms as compared to the $(1/\langle r\rangle)$ terms for the Baryon resonances may be of borderline validity, for the π meson resonances it plays a dominant role if a second interaction takes place well outside the original interaction region but inside of the range

$$\langle r\rangle = 2\,|\langle \mathbf{v}_1\rangle - \langle \mathbf{v}_2\rangle|^3\,\Gamma_a^{-2}(m_a) \times [m_1^2(p_0^2 + m_1^2)^{-\frac{3}{2}} + m_2^2(p_0^2 + m_2^2)^{-\frac{3}{2}}]^{-1}. \quad (143)$$

The expressions $\psi_1 - \psi_2$ given in Eqs. (114) and (115) therefore represent a better approximation to the exact solution of the scattering problem than is obtained with the standard approach and may prove of value when taking into account final state interactions. It is interesting to note that any final state interaction that involves the resonant parts of $\psi_1 - \psi_2$ is peripheral because from Eq. (117)

$$\psi_1 - \psi_2 \sim 1 - Ae^{-br},$$

which tends to zero for small r.

6. DISCUSSION

In the interest of simplicity it was assumed above that there was only a single state $|a\rangle$. For a relativistic theory there would in general be at least the other state, say $|a_-\rangle$, the negative energy state of $|a_+\rangle$. The propagator for this state becomes

$$\sum_l \frac{|a_l\rangle\langle a_l|}{E - E_{a_l} - D_{a_l} + \frac{1}{2}i\Gamma_a} = \sum_l \frac{(\bar{H}_0 + E + D_{a_l} + \frac{1}{2}i\Gamma_a)\,|a_l\rangle\langle a_l|}{(E + \frac{1}{2}i\Gamma_a)^2 - (E_{a_+} + D_{a_+})^2}, \quad (144)$$

since iD_a and Γ_a are presumably invariant to charge conjugation. Near the resonance this can be written as

$$\sum_l \frac{|a_l\rangle\langle a_l|}{E - E_{a_l} - D_{a_l} + i(\Gamma_a/2)} \cong \frac{2m_a\,|a_+\rangle\langle a_+|}{(E + \frac{1}{2}i\Gamma_a)^2 - (E_{a_+} + D_{a_+})^2}, \quad (145)$$

where, as before, terms of order (Γ_a/E_R) are neglected. Consequently, the cross sections have a Lorentz shape rather than the shape of a Breit–Wigner resonance.

The formalism developed in this paper is easily extended to include closely coupled unstable states as well as isolated resonances by adopting the projection operator techniques developed by Feshbach[12] and recently extended by Mower.[13] The generalization is made simply by using the appropriate diagonal elements of the Green's function rather than those defined by Eq. (60).

ACKNOWLEDGMENTS

The authors wish to thank their colleagues at the Iowa State University for the many fruitful and stimulating discussions regarding this study.

This work was performed at the Ames Laboratory of the United States Atomic Energy Commission, Contribution No. 1892.

[12] H. Feshbach, Ann. Phys. (N.Y.) **19**, 287 (1962).
[13] L. Mower, Phys. Rev. **142**, 799 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

New Derivation of the Ising Model Cluster Expansion*

DAVID JASNOW† AND MICHAEL WORTIS
Department of Physics, University of Illinois, Urbana, Illinois

(Received 4 August 1966)

A new and simple derivation of the cluster expansion for the free energy of an Ising model with arbitrary range of interaction is presented. The proof explicitly isolates the dependence of the free energy on the strength of individual interaction bonds. Several points of principle are discussed. An illustrative appendix gives simple applications.

I. INTRODUCTION

THE term "cluster expansion" has several meanings in the literature. Roughly speaking, let us distinguish between (a) *perturbation* expansions in powers of the interaction strength, the terms of which may be represented graphically as linked clusters, and (b) *nonperturbative* expressions giving the physical properties of an ensemble in terms of those of its constituent subensembles. There are, of course, connections between these two types of expansions, the latter being in some sense partial summations of the former (see below). It is with cluster expansions of the type (b) that this paper is primarily concerned.

Past derivations of the cluster expansion for the Ising model have followed two separate lines. The first approach, due independently to Rushbrooke and Morgan[1] and to Elliott and Heap,[2] was motivated by the desire to treat the "randomly dilute" Ising model, in which a proportion, $p < 1$, of lattice sites are magnetically active and the remainder are inert. It is postulated that the active sites are distributed randomly, with no reference to energetic considerations. At sufficiently low concentrations p, the probable distribution of active sites looks like a collection of small magnetic clusters of various sizes and configurations, each isolated from its neighbors by nonmagnetic sites. Under these circumstances it is reasonable to expand the thermodynamics as a sum of contributions from isolated magnetic clusters, each multiplied by an "occurrence factor," involving p and the geometry of the perfect lattice. In this approach the occurrence factor for a given cluster is a cumbersome function of p, since one must require not only that all the cluster sites are active but also that all the neighboring sites, with which the cluster might interact, are inert. After some labor the expansion can be exhibited as a power series in p. The meaning of this expansion for large p, particularly in the limit $p = 1$, is not clear from the derivation. For large p the occurrence of *isolated* finite clusters becomes most improbable. In particular, for $p = 1$ the occurrence factor for every cluster other than the perfect lattice itself is rigorously zero.[3]

To clarify this situation, Brout and Klein[4,5] and later Rushbrooke[6] took an entirely different approach. High-temperature expansions[7,8] in powers of $\beta = 1/kT$ times the interaction potential had been in existence for some time. It proved possible to associate a concentration dependence with each of the terms in this expansion and then to sum all terms of given concentration dependence, thus rederiving the power series in p. This second derivation, while free of the conceptual difficulties of the first, is combinatorically quite complicated in a way that tends to obscure the appearance of the thermodynamic functions of the magnetic clusters.

The present derivation works for $p = 1$ but uses a magnetic field and an interaction strength varying from site to site and from bond to bond, respectively. The free energy is thus exhibited in its functional dependence on each site and bond. The derivation holds for arbitrary range of interaction. The result emerges directly in terms of the cluster free energies. It is easy to introduce powers of p appropriate to the random dilution problem into the $p = 1$ form of the free energy.

Section II presents the meat of the derivation. The result is discussed in Sec. III. Calculations of previous

* Research supported by National Science Foundation Grant No. NSF GP 4937.

† National Aeronautics and Space Administration Fellow.

[1] G. S. Rushbrooke and D. J. Morgan, Mol. Phys. 4, 1 (1961); see also D. J. Morgan and G. S. Rushbrooke, *ibid.* 4, 291 (1961).

[2] R. J. Elliott and B. R. Heap, Proc. Roy. Soc. (London) A265, 264 (1962); also, B. R. Heap, *ibid.* 82, 252 (1963).

[3] Each inert neighbor carries a factor $(1 - p)$, which vanishes at $p = 1$. Only the perfect lattice lacks inert sites.

[4] M. W. Klein and R. Brout, Phys. Rev. 132, 2412 (1963).

[5] R. Brout, Phys. Rev. 115, 824 (1959), provides background for Ref. 4.

[6] G. S. Rushbrooke, J. Math. Phys. 5, 1106 (1964).

[7] For a list of these references see G. F. Newell and E. W. Montroll, Rev. Mod. Phys. 25, 353 (1953).

[8] G. Horwitz and H. B. Callen, Phys. Rev. 124, 1757 (1961).

authors are resketched from our point of view in an Appendix.

We wish to emphasize that our result, Eq. (21), the cluster expansion in powers of p, is identical to that derived by previous authors.[1,2,4–6] Only the proof is different. Methods very similar to those we use have been employed in other contexts by Kubo, Strieb, Callen, and Horwitz.[9]

II. THE DERIVATION

Use numerical arguments $1, 2, \cdots$ to refer to the N sites of the Ising lattice. It is not necessary at this stage to specify the lattice geometry. Write the Ising Hamiltonian H as

$$-\beta H = \sum_1 b(1)\mu(1) + \sum_{\langle 12\rangle} v(12)\mu(1)\mu(2), \tag{1}$$

where $b(1)$ is the dimensionless external magnetic field at the site 1, $v(12)$ is the dimensionless interaction strength between sites 1 and 2, and

$$\sum_{\langle 12\rangle}$$

is to be read as the sum over all interacting pairs 1, 2. The dynamical variable μ at each site is restricted to the values $\mu(1) = \pm 1$. We emphasize that $b(1)$ and $v(12)$ are functions over sites and pairs of sites, respectively. In the particular case of uniform magnetic field and nearest-neighbor interactions

$$b(1) = b, \quad v(12) = \begin{cases} v \text{ for 1 and 2 nearest} \\ \quad \text{neighbors,} \\ 0 \text{ otherwise.} \end{cases} \tag{2}$$

The free energy F of the model is

$$-\beta F = W = \ln \operatorname{Tr} \exp(-\beta H), \tag{3}$$

where the trace is over the two values of each operator μ. Now, any operator A satisfying $A^2 = 1$ as a minimum equation obeys

$$\exp(\alpha A) = \cosh \alpha (1 + A \tanh \alpha) \tag{4}$$

for any number α for which the operator exponential is well defined. Use (4) to rewrite W:

$$W = W_0 + W_1 \tag{5}$$

with

$$W_0 = \sum_1 \ln 2 \cosh b(1) + \sum_{\langle 12\rangle} \ln \cosh v(12) \tag{6}$$

and

$$W_1[s, t] = \ln \left[2^{-N} \operatorname{Tr} \prod_1 \big(1 + \mu(1)s(1)\big) \times \prod_{\langle 23\rangle} \big(1 + \mu(2)\mu(3)t(23)\big) \right], \tag{7}$$

where

$$s(1) = \tanh b(1), \quad t(23) = \tanh v(23), \tag{8}$$

and the functional dependence of W_1 on the set of variables $s(1)$, $t(23)$ has been indicated. Observe that expansion of the multiple product in (7) produces a power series in the operators μ. The trace of any term which contains an *odd* power of one or more of the operators $\mu(1)$ is zero. The relation

$$W_1[s, 0] = 0 \tag{9}$$

is a direct consequence of this important property. What, now, is the structure of $W_1[s, t]$?

$W_1[s, t]$ may be regarded as the result of the operation,

$$W_1[s, t] = \exp\left[\sum_{\langle 12\rangle} t(12) \frac{\delta}{\delta \bar{t}(12)} \right] W_1[s, \bar{t}]\big|_{\bar{t}=0}. \tag{10}$$

This is just a formal way of writing the Taylor expansion[10] of W_1 in the t variables. It is convenient to regard (10) as a multidimensional translation operation:

$$W_1[s, t] = \prod_{\langle 12\rangle} T(12) W_1[s, \bar{t}], \tag{11}$$

where

$$T(12) = \exp[t(12)\, \delta/\delta \bar{t}(12)] \tag{12}$$

and it is left implicit that the barred variables are to be set to zero after differentiation. All the operators $T(12)$ commute. Note that every term in the expansion of $[T(12) - 1]W_1[s, \bar{t}]$ contains one or more powers of $t(12)$. Thus, the separation,

$$T(12) = 1 + [T(12) - 1]$$

isolates those terms which do not depend on $v(12)$ from those which do.

We now assert that the development of W_1 in $\Delta T = (T - 1)$ is the cluster expansion:

$$W_1[s, t] = \prod_{\langle 12\rangle} [1 + \Delta T(12)] W_1[s, \bar{t}], \tag{13}$$

$$\Delta T(12) = T(12) - 1. \tag{14}$$

The general term in the expansion of (13) is just a product of ΔT factors for a certain set, G, of interacting pairs:

$$\prod_{\alpha \in G} \Delta T(\alpha).$$

A simple graphical representation is achieved by drawing onto the labeled lattice of sites those bonds α contained in G. Every such graph appears once and only once in (13). See Fig. 1 for examples. Many of the terms in (13) correspond to graphs having two or more disconnected parts. We now prove the following proposition: *The contribution to* (13) *of every*

[9] R. Kubo, J. Phys. Soc. Japan **17**, 1100 (1962); B. Strieb, H. B. Callen, and G. Horwitz, Phys. Rev. **130**, 1798 (1963). We are grateful to the referee for bringing these references to our attention.

[10] Equations giving Ising thermodynamic functions as power series in $t = \tanh v$ have been given, for example, by T. Oguchi, J. Phys. Soc. Japan **6**, 31 (1951), and S. Katsura, Progr. Theoret. Phys. (Kyoto) **20**, 192 (1958). Such expansions are in a sense rearrangements of those given in Refs. 7 and 8. Equation (10) can be used as a basis for their derivation.

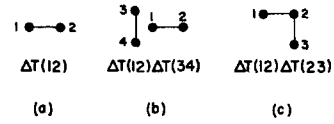


FIG. 1. Graphs in the expansion of Eq. (13). Graphs (a) and (c) are connected. Graph (b) is disconnected and does not contribute to the free energy.

disconnected graph is zero.[11] Consider a disconnected term,

$$\prod_{\alpha\in G_1}\Delta T(\alpha)\prod_{\beta\in G_2}\Delta T(\beta)W_1[s,t],$$

where no bond in G_1 shares a site with a bond in G_2. To make this more explicit, let F_1 and F_2 be the set of end points of the bonds in G_1 and G_2, respectively. F_1 and F_2 are disjoint. Notice that the t variables associated with neither G_1 nor G_2 may be set to zero immediately. Now evaluate $W_1[s,t]$ by (7). The trace factors. The logarithms of the factors add. The μ's associated with sites in neither F_1 nor F_2 cannot be paired. So, by a logic parallel to that leading to (9),

$$W_1[s,t]|_0 = W_1[s,t;G_1] + W_1[s,t;G_2], \quad (15)$$

where the bar on the left indicates that $t=0$ for bonds in neither G_1 nor G_2. The quantities on the right are defined by

$$W_1[s,t;G] = \ln\left[2^{-n}\,\mathrm{Tr}\prod_{1\in F}\big(1+\mu(1)s(1)\big)\times\prod_{\langle 23\rangle\in G}\big(1+\mu(2)\mu(3)t(23)\big)\right]. \quad (16)$$

The trace in (16) is restricted to the n sites contained in F. $W_1[s,t;G]$ is just the nontrivial part of the free energy of an Ising model containing only the sites in F and the bonds in G. Equation (15) expresses the additivity of the free energy of noninteracting subsystems. Since

$$\Delta T(\alpha\in G_1)W_1[s,t;G_2] = \Delta T(\beta\in G_2)W_1[s,t;G_1] = 0, \quad (17)$$

the proposition is proved. Thus,

$$W_1[s,t] = \sum_{\text{connected } G}\prod_{\alpha\in G}\Delta T(\alpha)W_1[s,t], \quad (18)$$

which shows the basic cluster property of the free energy. The term, $C[G]$, associated with each graph G contains the total contribution of the group of bonds in G acting together; i.e., the term is present for any dilute Ising model containing all the bonds in G but vanishes as soon as any one of the bonds in G is removed.

Equations (1), (5), and (18) constitute the cluster expansion for the undiluted Ising model with arbitrary strength and range of interaction. We now complete the circle[12] by pointing out that the contribution to W_1 of each graph G is expressible as a sum of free energies of "mutilated" Ising models, in which the only bonds are in subsets of G. Notice that

$$W_1[s,t;G] = \prod_{\alpha\in G}T(\alpha)W_1[s,t] \quad (19)$$

is a special case of (11). Thus, each term in (18) can be written as

$$C[G] = \prod_{\alpha\in G}\Delta T(\alpha)W_1[s,t] = \sum_{\gamma\subseteq G}(-1)^{n_\gamma}W_1[s,t;\gamma], \quad (20)$$

where the sum has one term for each subset, γ, of G and n_γ is the number of bonds in G but not in γ. For small clusters $W_1[s,t;\gamma]$ is easy to evaluate via (16) or other available methods. Then (20) expresses conveniently the contribution to the free energy of the graph G.

The extension to the randomly dilute system is now straightforward. We follow other authors[1,2,4-6] in arguing that for large N it is possible to realize the "randomly dilute" system described in Sec. I by assigning to each site *independently* a probability p of being active and correspondingly a probability $(1-p)$ of being inert. Any cluster of n labeled points then has a chance p^n of surviving intact a dilution of the lattice from $p=1$ to some concentration $p<1$. Thus the contribution of each graph G_n having n sites as bond endpoints must be decreased by a factor p^n. The cluster expansion then reads

$$-\beta F = W = p\sum_1\ln 2\cosh b(1) + p^2\sum_{\langle 12\rangle}\ln\cosh v(12) + \sum_{n=2}^{\infty}p^n\sum_{\substack{G_n\\ \text{connected}}}\prod_{\alpha\in G_n}\Delta T(\alpha)W_1[s,t], \quad (21)$$

which gives the free energy of the randomly dilute Ising model in an arbitrary external field and with arbitrary interaction range and strength.[13] The restriction to uniform magnetic field and, for example, nearest-neighbor interactions is direct. Some simple applications are presented in the Appendix.

III. DISCUSSION

(i) The skeptical reader will have observed a certain circularity in the argument for $p=1$. Equation (13)

[11] This proposition ensures that for an interaction of finite range the ratio F/N is finite as $N\to\infty$, i.e., that the free energy is extensive.

[12] See Sec. III (i).

[13] There is a hidden assumption here. Equation (21) is actually a rigorous evaluation of the average of W over an *ensemble* of dilute Ising models characterized by a site occupation *probability* p. The identification of (21) as the free energy of a *typical* Ising model of *concentration* p rests on two assertions: (i) that the peaking of the concentration distribution is such that (21) gives the *mean* free energy of Ising models with concentration p, and (ii) that the fluctuations of free energy among Ising models of concentration p are sufficiently small so that a typical one will have a free energy near the mean with high probability. We do not attempt to justify these assertions.

expresses T in terms of $(T-1)$, while (20) gives $(T-1)$ in terms of T. In effect, the substitution of (20) into (13) expresses $W_1[s,t]$ as $W_1[s,t]$ plus a set of terms summing to zero, involving the free energy of Ising models with one or more missing interaction bonds.[14] What is the significance of an "expansion" which is simply a rearrangement of a large number of extraneous terms adding to zero? We have indicated after (18) that the cluster expansion in our form does have the significance of isolating the contribution of each site and bond to the free energy. Suppose a given interaction bond $v(12)$ is deleted from the Hamiltonian (1). What is the decrease in free energy? In principle the answer is obtained by adding to the trivial term, ln cosh $v(12)$, the sum of the contributions to (18) of all graphs in which the bond $v(12)$ occurs. Of course, this computation is not easy in practice. The free energy associated with a given site can be calculated analogously. Part (i) of the Appendix should clarify these points.

(ii) A more subtle question concerns the convergence of (21), particularly at $p=1$. When the number of sites, N, is finite, there is no difficulty. It is the present authors' belief that no one has been able to discuss the $N\to\infty$ limit of the cluster expansion with any degree of mathematical rigor.[15]

(iii) Another mathematical question concerns the convergence of the translation operator $T(12)$. When $t=\tanh v$ approaches ± 1, one may well worry that the radius of convergence of such forms as ln $(1+t)$ is being approached. Again, there is no trouble for finite N: one may always work for very weak interaction, for which the Taylor expansion is certainly valid, and then invoke the analyticity of W in interaction strengths to extend the result. The $N\to\infty$ limit remains obscure.

(iv) An interesting conceptual point emerges in the comparison of our result with that of Ref. 1. Our analysis isolates individual bond contributions, while Rushbrooke's analysis is based on the contributions of isolated point clusters of magnetic sites. Let us restrict ourselves to a square plane lattice with nearest-neighbor interactions and develop an apparent paradox. Refer to Fig. 2. Both 2(a) and 2(b) appear as admissible graphs in our formulation. Only 2(b) is an admissible point cluster for Rushbrooke. Since sites 1 and 4 are nearest neighbors, they *must* interact in Rushbrooke's formulation. The paradox is that $W_1[s,t;2(a)]$ for the graph 2(a) seems to appear in our expansion but not in Rushbrooke's. The resolution is that both graphs contribute to order p^4 of our cluster expansion. When contributions are calculated according to (20), there is a cancellation between them in which $W_1[s,t;2(a)]$ disappears. The identity of our result with previous ones guarantees that this cancellation is a general feature. In our formulation there *will* be cancellations between different graphs G_n for the same value of n. This dead wood is the price we pay for the conceptual clarity of being able to exhibit explicitly individual bond contributions. Part (i) of the Appendix should make these remarks more concrete.

FIG. 2. Two p^4 graphs. Only (b) is a point cluster in the analysis of Ref. 1.

(v) At $b=0$ there is a simplification of the cluster expansion. Graphs with one or more free ends[16] do not contribute. This is easy to see: let the free site be 1 and its single connecting bond be $t(12)$. Then, since $s(1)=0$, the $\mu(1)$ in $\mu(1)\mu(2)t(12)$ can never be paired and

$$(T(12)-1)W_1[s,t]=0 \tag{22}$$

regardless of what other bonds the graph contains.

(vi) Calculations. Once the interaction $v(12)$ is made translationally invariant, all graphs of similar[17] geometry contribute identically. The contribution of a given graph type is, of course, the product of a single graph contribution and an occurrence factor. See Appendix (ii) and (iii) for some examples. In actual *computation* our form, (21), of the cluster expansion is not significantly less cumbersome than others.[18] We do feel that the *derivation* is both simple and transparent.

(vii) Note finally that Eq. (18) allows treatment of what we may call the "bond dilution" problem, in which bonds instead of sites are removed on a random basis. For a concentration p of bonds (all sites remaining present),

$$W_0=\sum_1 \ln 2\cosh b(1)+p\sum_{\langle 12\rangle}\ln\cosh v(12). \tag{23}$$

W_1 is given by the same graphical sum as in (21), only with n reinterpreted as the number of *bonds* in the graph G_n.

[14] It is just this feature which we noted in Rushbrooke's terminology in Sec. I and Ref. 3.

[15] R. J. Elliott, B. R. Heap, D. J. Morgan, and G. S. Rushbrooke [Phys. Rev. Letters 5, 366 (1960)] presented an argument showing that the critical concentration p_c below which long-range order is impossible no matter how low the temperature, is the same for the Heisenberg and Ising models, depending only on lattice structure. This proof was based on the susceptibility expansion, i.e., the second derivative of (21) with respect to the external field. In a critique of this proof Rushbrooke and Morgan [Mol. Phys. 6, 477 (1963)] suggest that the susceptibility expansion for the special case $b=0$, $T=0$ is *not* uniformly convergent.

[16] A free end is a site which serves as endpoint for one bond only.

[17] It is only neighbor relationships which count here, not bond angles. See Appendix, part (ii).

[18] By comparison with Ref. 1, for example, we have $(T-1)$ factors instead of $(1-p)$ factors.

ACKNOWLEDGMENTS

One of us (M. W.) is indebted to Professor R. J. Elliott for a very helpful discussion. It is a pleasure to acknowledge the hospitality of the Physics Division of the Aspen Institute for Humanistic Studies, where much of this paper was written.

APPENDIX. SAMPLE CALCULATIONS

(i) The Triangle Problem: Consider an Ising model consisting of three sites, 1, 2, and 3, and three bonds, α_1, α_2, and α_3. See Fig. 3(a). Since $N = 3$ is small here, the W given by Eq. (21) can *only* be interpreted as the ensemble average described in footnote 13. The calculation of W_1 involves seven graphs, Fig. 3(a) itself, three graphs with bond pairs, and three with single bonds. Label the graphs by *bonds* contained and use (20). Typical contributions are

$$\begin{aligned} C[123] &= W_1[123] - W_1[12] - W_1[13] - W_1[23] \\ &\quad + W_1[1] + W_1[2] + W_1[3], \\ C[12] &= W_1[12] - W_1[1] - W_1[2], \\ C_1[1] &= W_1[1], \end{aligned} \tag{A1}$$

where s and t dependences have been suppressed. From (16),

$$\begin{aligned} W_1[123] &= \ln [1 + s_1s_2t_3 + s_1t_2s_3 + t_1s_2s_3 \\ &\quad + t_1t_2s_1s_2 + t_2t_3s_2s_3 + t_1t_3s_1s_3 + t_1t_2t_3], \\ W_1[12] &= \ln [1 + s_1t_2s_3 + t_1s_2s_3 + t_1t_2s_1s_2], \\ W_1[1] &= \ln [1 + t_1s_2s_3], \end{aligned} \tag{A2}$$

where

$$s_i = \tanh b(i), \quad t_i = \tanh v(\alpha_i). \tag{A3}$$

If for simplicity we allow bond strengths and fields to become equal, Eq. (21) reads

$$\begin{aligned} W &= 3p \ln 2 \cosh b + 3p^2 \ln \cosh v + 3p^2 W_1[1] \\ &\quad + p^3(W_1[123] - 3W_1[1]). \end{aligned} \tag{A4}$$

Note that the non-point-cluster contribution, $W_1[12]$, has disappeared[19] from (A4), as discussed in Sec. III(iv).

Suppose now that bond α_3 were absent. See Fig. 3(b). The graphs, [123], [13], [23], and [3], would give zero. Instead of (A4),

$$\begin{aligned} W &= 3p \ln 2 \cosh b + 2p^2 \ln \cosh v + 2p^2 W_1[1] \\ &\quad + p^3(W_1[12] - 2W_1[1]). \end{aligned} \tag{A5}$$

Note that $W_1[12]$ does appear here. The difference, (A4)–(A5), is that part of the free energy due exclusively to the presence of bond α_3.

(ii) The Free Energy for Low Concentration: Let us calculate the cluster expansion (21) for plane square and simple cubic nearest neighbor Ising models through terms in p^3. Contributing graphs are shown in Fig. 4. Graphs 4(b) and 4(c) contribute identically.

[19] This cancellation holds even when bond strengths are unequal.

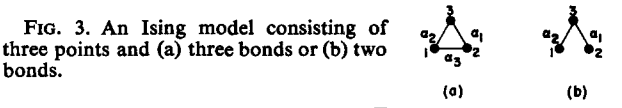


FIG. 3. An Ising model consisting of three points and (a) three bonds or (b) two bonds.

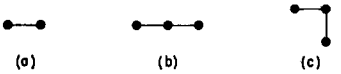


FIG. 4. Free energy graphs of orders p^2 [graph (a)] and p^3 [graphs (b) and (c)] for plane square and simple cubic nearest-neighbor lattices. Graphs (b) and (c) contribute identically.

$$\begin{aligned} C[4(b)] &= W_1[4(b)] - 2W_1[4(a)], \\ C[4(a)] &= W_1[4(a)]. \end{aligned} \tag{A6}$$

$$\begin{aligned} W_1[4(b)] &= \ln (1 + 2ts^2 + t^2s^2), \\ W_1[4(a)] &= \ln (1 + ts^2). \end{aligned} \tag{A7}$$

Occurrence factors for 4(a) and 4(b) are $\frac{1}{2}zN$ and $\frac{1}{2}z(z-1)N$, respectively, where z is the number of nearest neighbors. Thus, through order p^3

$$\begin{aligned} W/N &= p \ln 2 \cosh b + \tfrac{1}{2}p^2 z \\ &\quad \times \ln (\cosh v + \tanh^2 b \cdot \sinh v) + \tfrac{1}{2}p^3 z(z-1) \\ &\quad \times [\ln (1 + 2 \tanh^2 b \cdot \tanh v + \tanh^2 b \cdot \tanh^2 v) \\ &\quad - 2 \ln (1 + \tanh^2 b \cdot \tanh v)]. \end{aligned} \tag{A8}$$

This result agrees[20] with Ref. 1.

(iii) Ring Graphs: Consider an Ising model with $b = 0$ and arbitrary range of interaction. As shown in Sec. III(v), graphs with free ends do not contribute. Ring graphs are defined as those graphs consisting of a single closed loop of bonds, i.e., at each vertex of a ring graph two and only two bonds meet. Without discussing the relevance of this set of graphs,[21] we show, as a demonstration of technique, that it is possible to go some way towards evaluating its contribution to the free energy. Consider the set of ring graphs with n bonds and (thus) n vertices. Only the leading term of (20) fails to vanish. A typical graph, labeled now by its vertices taken in serial order around the perimeter, gives

$$C_n[1 \cdots n] = \ln \big(1 + t(12)t(23) \cdots t(n-1, n)\big). \tag{A9}$$

The total n-bond ring contribution to the free energy $W_{n\,\mathrm{ring}}$ is obtained by summing (A9). The sum on graphs is converted to a sum over vertex sites:

$$W_{n\,\mathrm{ring}} = \frac{p^n}{2n} \sum_{1, \cdots, n}{}' C_n[1 \cdots n], \tag{A10}$$

where each of the arguments is summed over all lattice sites. The prime on the summation restricts the vertices, $1, \cdots, n$, to *distinct* sites. The symmetry factor $1/2n$ compensates for multiple counting. The result (A10) has been derived elsewhere.[22]

[20] Actually, it is the susceptibility χ which is calculated in Ref. 1. The following equivalences facilitate comparison:

$$\begin{aligned} &\frac{d^2W}{db^2}\bigg|_{b=0} = \frac{4kT}{g^2\beta^2}\chi \qquad \text{(Ref. 1)}, \\ &v = K \text{ (Ref. 1)} = J/2kT \text{ (Ref. 1)}, \\ &b = g\beta H/2kT \qquad \text{(Ref. 1)}. \end{aligned}$$

[21] This is discussed in a particular context in Ref. 4.

[22] Reference 4, Eq. (10). The restriction on the summation is omitted there as part of the approximation scheme being used.

Padé Approximant and Partial-Wave Integral Equation*

D. MASSON
Department of Mathematics, University of Toronto, Toronto, Canada

(Received 19 July 1966)

The Padé approximant is applied to the partial-wave integral equation and is shown to yield approximate solutions which satisfy exact two-body unitarity and which converge to the N/D solution.

I. INTRODUCTION

THE Padé approximant[1] can be used to make approximate analytic continuations of the power series expansion of a function. For rational functions or functions whose power series expansion is a series of Stieltjes[1–3] the Padé approximant can be shown to "converge" to the function as the degree of the Padé numerator and denominator tend to infinity.[4] Fortunately, some functions of physical interest can be represented as series of Stieltjes, for example, the forward scattering amplitude[1] and partial-wave scattering amplitude for potential scattering[5] as a function of the potential strength. Here we demonstrate the convergence of the Padé for a relativistic model. We consider the s-wave amplitude for the scattering of two spinless particles as a function of the strength of the left-hand discontinuity. We show that it is related to a series of Stieltjes and hence the Padé can be used as an alternative to the N/D method[6] for solving the partial-wave integral equation. A natural approximate solution to the problem is thus given by the (N, N) Padé, which satisfies two-body unitarity and which would yield the exact amplitude if the left-hand cut consisted of N poles.

II. THE PADÉ APPROXIMANT

For the elastic scattering of two spinless particles, the s-wave amplitude

$$T = e^{i\delta(s)} \sin \delta(s)/\rho(s) \tag{1}$$

is a real analytic function of s (the square of the total center-of-mass energy) in the complex plane cut along the real axis, where ρ is a kinematical factor. We assume that for sufficiently small values of λ, T satisfies the dispersion relation

$$T(\lambda, s) = \lambda B(s) + \frac{1}{\pi}\int_{s_R}^{\infty} \frac{\operatorname{Im} T(\lambda, s')}{s' - s}\, ds', \tag{2}$$

where

$$B(s) = \frac{1}{\pi}\int_{-\infty}^{s_L} \frac{\sigma(s')}{s' - s}\, ds', \tag{3}$$

the left-hand cut contribution to T is given. In the two-body unitarity approximation, along the right-hand cut

$$\operatorname{Im} T(\lambda, s) = \rho(s)\,|T(\lambda, s)|^2. \tag{4}$$

Equation (2) can then be considered as a singular nonlinear integral equation for $T(\lambda, s)$. We propose to solve Eq. (2) by first iterating to get

$$T(\lambda, s) = \sum_{i=1}^{\infty} t_i(s)\lambda^i \tag{5}$$

and then forming the (N, M) Padé approximant to the power series (5)

$$[T(\lambda)]_{N,M} = P_{N,M}(\lambda)/Q_{N,M}(\lambda), \tag{6}$$

where $P_{N,M}$ and $Q_{N,M}$ are polynomials in λ of degree M and N, respectively, with coefficients which are functions of s and determined from the condition that (6) have the same power series expansion as (5) up to and including the term λ^{N+M+1}.

III. PROOF OF UNITARITY AND CONVERGENCE

We now prove the following:

(1) $[T(\lambda)]_{N,M}$ satisfies exact two-body unitarity for $N \geq M$.

* Supported in part by the National Research Council of Canada.

[1] G. A. Baker, Jr., in *Advances in Theoretical Physics*, K. A. Brueckner, Ed. (Academic Press Inc., New York, 1965), Vol. I, p. 1.

[2] H. S. Wall, *Analytic Theory of Continued Fractions* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1948), Chaps. XVII and XX.

[3] A function $f(z)$ is a series of Stieltjes (Ref. 1) if $f(z) = \int_0^{\infty} [d\phi(u)/(1 - uz)]$ has a power series expansion $f(z) = \sum f_n z^n$, where $\phi(u)$ is a bounded nondecreasing function taking on infinitely many values in the interval $0 \leq u < \infty$.

[4] The precise statements of convergence which we use are: (1) If $f(z)$ is a ratio of polynomials of degree m in the numerator and n in the denominator then the (N, M) Padé is equal to $f(z)$ if $N \geq n$ and $M \geq m$. (2) If $f(z)$ is a series of Stieltjes and has a finite radius of convergence, Baker (Ref. 1) proves that $\lim_{N\to\infty} (N, N + j)$ Padé converges to $f(z)$ when z is not a singular point and $j \geq -1$. By considering $f(z) = f(0)/[1 - zg(z)]$, one can show that $g(z)$ is a series of Stieltjes and thus extend the theorem to arbitrary j.

[5] S. Tani, Phys. Rev. **139**, B1011 (1965).

[6] G. F. Chew and S. Mandelstam, Phys. Rev. **119**, 467 (1960).

(2) In the limit as N and M become infinite $[T(\lambda)]_{N,M}$ is equal to $T(\lambda)$.

To prove (1) we need only know that the (N, M) Padé is unique.[2] Along the right-hand cut unitarity tells us that $\text{Im}\,[1/T(\lambda)] = -\rho$. The uniqueness of the Padé tells us that $\lambda/[T(\lambda)]_{N,M} = [\lambda/T(\lambda)]_{M-1,N}$, and for $M \leq N$, $\text{Im}\,\lambda/T(\lambda) = \text{Im}\,[\lambda/T(\lambda)]_{M-1,N}$. Combining these three statements we have, for $N \geq M$,

$$\text{Im}\,[T(\lambda)]_{N,M} = \rho(s)\,|[T(\lambda)]_{N,M}|^2. \tag{7}$$

To prove (2) we use a modified version of the N/D method of solution due to Ball[7] which reduces the solution of Eq. (2) to the solution of a Fredholm integral equation of the Hilbert–Schmidt type.

We write $T = N/D$ with

$$N(s) = \lambda B(s) + \frac{\lambda}{\pi}\int_{-\infty}^{s_L} ds' K(s, s', s_0) N(s') \tag{8}$$

and

$$D(s) = 1 - \frac{(s - s_0)}{\pi}\int_{s_R}^{\infty} \frac{N(s')\rho(s')\,ds'}{(s' - s)(s' - s_0)}, \tag{9}$$

where

$$K(s, s', s_0) = \frac{B(s')(s' - s_0) - (s - s_0)B(s)}{s' - s}\,\frac{\rho(s')}{(s' - s_0)} \tag{10}$$

and s_0 is an arbitrary point less than s_R.

Since the kernel for the integral Eq. (8) is of the polar type,[8] it can be transformed into one with a symmetric kernel,

$$k(s, s'; s_0) = K(s, s', s_0)\left(\frac{\rho(s)}{\rho(s')}\,\frac{(s' - s_0)}{(s - s_0)}\right)^{\frac{1}{2}}, \tag{11}$$

and if sufficiently well behaved has a denumerable set of real eigenvalues $\lambda_i(s_0)$ which we assume to be nondegenerate. N and D can then be expanded in terms of the eigenfunctions and eigenvalues of the kernel.[7–9] One gets

$$N(s) = \lambda B(s) + \lambda^2 \sum \frac{B_i(s_0)U_i(s, s_0)/\lambda_i}{1 - \lambda/\lambda_i}, \tag{12}$$

where

$$B_i(s_0) = \int_{s_R}^{\infty} ds' \frac{\rho(s')B(s')U_i(s', s_0)}{s' - s_0}, \tag{13}$$

$$U_i(s, s_0) = \lambda_i(s_0)\int_{s_R}^{\infty} K(s, s', s_0)U_i(s', s_0)\,ds'. \tag{14}$$

If one now considers $s = s_0$, then one has finally

$$T(\lambda, s) = \lambda B(s) + \lambda^2 f(\lambda, s), \tag{15}$$

where

$$f(\lambda, s) = \sum \frac{B_i^2(s)}{1 - \lambda/\lambda_i(s)}, \tag{16}$$

and $B_i(s)$ and $\lambda_i(s)$ are both real for $s_L < s < s_R$. There are three cases that one can consider.

(A) *The λ_i are finite in number:* A necessary and sufficient condition for this to occur is for the left-hand cut to consist of a finite number of n poles, in which case $[T(\lambda, s)]_{N,M} = T(\lambda, s)$ for N and $M \geq n$.

(B) *The λ_i are all of the same sign:* If $\sum B_i^2(s)\lambda_i(s)$ is uniformly convergent,[9] then $B(s) = \sum B_i^2(s)\lambda_i(s)$, and Eq. (15) becomes

$$T(\lambda, s) = \lambda \sum \frac{B_i^2(s)\lambda_i(s)}{1 - \lambda/\lambda_i(s)} \tag{17}$$

so that for $s_L < s < s_R$, $T(\lambda, s)/\lambda$ is a series of Stieltjes[3] in λ and therefore[4]

$$\lim_{N\to\infty} [T(\lambda, s)]_{N,N+j} = T(\lambda, s). \tag{18}$$

A necessary and sufficient condition for Case (B) to hold is for the kernel $k(s, s'; s_0)$ to be definite.[8] That is

$$\int_{s_R}^{\infty} k(s, s'; s_0) f(s) f(s')\,ds\,ds' \neq 0$$

for arbitrary real $f(s)$. From Eqs. (3), (10), and (11) one has the equivalent condition

$$\int_{-\infty}^{s_L} ds\sigma(s) f^2(s)(s - s_0) \neq 0. \tag{19}$$

Hence if $s_0 > s_L$, the kernel is definite if and only if $\sigma(s)$ is positive or negative semidefinite in the interval $(-\infty, s_L)$. This is indeed the situation in many bootstrap models.[7]

(C) *The λ_i differ in sign:* In this case one is dealing with the moment problem over the interval $(-\infty, \infty)$. From the form of $f(\lambda, s)$ in Eq. (16), one can prove[1,2] that for $j = \pm 1, \pm 3, \cdots$, and $s_L < s < s_R$,

$$\lim_{N\to\infty} [f(\lambda, s)]_{N,N+j} = f(\lambda, s). \tag{20}$$

Also, one can write $T(\lambda, s) = \lambda B(s)[1 - \lambda g(\lambda, s)]^{-1}$, where $g(\lambda, s)$ is of the same form as $f(\lambda, s)$, and hence the statement of convergence (20) applies also to $g(\lambda, s)$. Combining the statements of convergence for $f(\lambda, s)$ and $g(\lambda, s)$ with the uniqueness of the Padé, one obtains Eq. (18) for $j = \pm 1, \pm 3, \cdots$.

IV. DISCUSSION

We have shown that the Padé approximant can, with complete rigor, replace the N/D method of solution of the partial-wave integral equation. It provides also a natural method of obtaining approximate solutions which satisfy exact unitarity. These

[7] J. S. Ball, Phys. Rev. 137, B1573 (1965).

[8] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Pt. I, Chap. 8.

[9] W. V. Lovitt, *Linear Integral Equations* (Dover Publications Inc., New York, 1950).

solutions are roughly speaking equivalent to approximating the kernel by a finite sum of separable kernels, which in this case means replacing the left-hand cut by a finite number of poles. They have the advantage, however, that they are independent of the approximation. That is, the solution does not depend on the choice of these poles. For example, the (1, 1) Padé approximant is

$$[T(\lambda, s)]_{1,1} = \lambda B(s)\left(1 - \lambda B^{-1}(s)\int_{s_R}^{\infty} \frac{\rho(s')B^2(s')}{s' - s}\, ds'\right)^{-1}, \tag{21}$$

which would be exact if $B(s)$ consisted of one pole. It is interesting that the approximate solution (21) has already appeared in the literature and was considered as satisfactory for weak and moderately strong couplings.[10] The approximate solutions have the additional feature that they also supply bounds on the exact solution.[1]

The techniques illustrated here should be capable of generalization to include many channels and spin, and hopefully extended beyond two-body unitarity.[11] The Padé method may of course be applied to any integral equation. The proof of its convergence is at present, however, limited to solutions which are essentially series of Stieltjes.

[10] G. L. Shaw, Phys. Rev. Letters **12**, 345 (1964); J. Reinfelds and J. Smith, Phys. Rev. **146**, 1091 (1966).
[11] J. L. Gammel and F. A. McDonald, Phys. Rev. **142**, 1245 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 8 MARCH 1967

Upper and Lower Bounds on Knudsen Flow Rates

WILLIAM STRIEDER* AND STEPHEN PRAGER
University of Minnesota, Minneapolis, Minnesota

(Received 23 May 1966)

Reciprocal variational principles are used to formulate upper and lower bounds on the low-pressure (Knudsen) flow rate of a gas through a channel of arbitrary geometry. The upper bound is equivalent to one obtained earlier by DeMarcus, but we believe the lower bound to be new. Explicit calculations are given for a short parallel-plate channel. The variational principles discussed here may be applied to a wide range of problems involving linear inhomogeneous integral equations.

1. INTRODUCTION

IN the so-called Knudsen regime, a gas flowing through a channel is at sufficiently low pressures for molecule–molecule collisions to be negligible, and only molecule–wall collisions need be taken into account. In addition, one usually assumes that a molecule equilibrates with the wall at each collision, so that its paths before and after the collision are totally uncorrelated.

Now, suppose that we have a gas at equilibrium in front of the entrance of the channel (Fig. 1), at a density of n molecules per cc, that the exit of the channel opens into a vacuum, and that these conditions are maintained until a steady state has been reached. If the mean molecular speed at the prevailing temperature is $\bar{v}$, then $I_0 = \frac{1}{4}n\bar{v}A$ is the number of molecules entering the channel per unit time, A being the area of the entrance cross section. Of these, a number Q/sec will leave through the exit, never to return, while the remainder come back out through the entrance.

For all but the simplest channel geometries, the calculation of Q is an exceedingly difficult problem,[1] and must be treated by approximate methods. One of the most promising has been the variational procedure introduced by DeMarcus,[2] which leads to

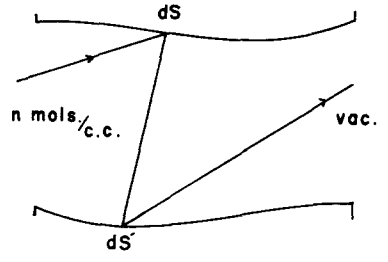


FIG. 1. Channel with sample molecular path.

* Present address: University of Notre Dame, South Bend, Indiana.

[1] E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938).
[2] W. C. DeMarcus, in *Advances in Applied Mechanics* Suppl. 1, *Rarefied Gas Dynamics* (Academic Press Inc., New York, 1961), p. 161.

solutions are roughly speaking equivalent to approximating the kernel by a finite sum of separable kernels, which in this case means replacing the left-hand cut by a finite number of poles. They have the advantage, however, that they are independent of the approximation. That is, the solution does not depend on the choice of these poles. For example, the (1, 1) Padé approximant is

$$[T(\lambda, s)]_{1,1} = \lambda B(s)\left(1 - \lambda B^{-1}(s)\int_{s_R}^{\infty} \frac{\rho(s')B^2(s')}{s' - s}\, ds'\right)^{-1}, \tag{21}$$

which would be exact if $B(s)$ consisted of one pole. It is interesting that the approximate solution (21) has already appeared in the literature and was considered as satisfactory for weak and moderately strong couplings.[10] The approximate solutions have the additional feature that they also supply bounds on the exact solution.[1]

The techniques illustrated here should be capable of generalization to include many channels and spin, and hopefully extended beyond two-body unitarity.[11] The Padé method may of course be applied to any integral equation. The proof of its convergence is at present, however, limited to solutions which are essentially series of Stieltjes.

[10] G. L. Shaw, Phys. Rev. Letters **12**, 345 (1964); J. Reinfelds and J. Smith, Phys. Rev. **146**, 1091 (1966).
[11] J. L. Gammel and F. A. McDonald, Phys. Rev. **142**, 1245 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 8 MARCH 1967

Upper and Lower Bounds on Knudsen Flow Rates

WILLIAM STRIEDER* AND STEPHEN PRAGER
University of Minnesota, Minneapolis, Minnesota

(Received 23 May 1966)

Reciprocal variational principles are used to formulate upper and lower bounds on the low-pressure (Knudsen) flow rate of a gas through a channel of arbitrary geometry. The upper bound is equivalent to one obtained earlier by DeMarcus, but we believe the lower bound to be new. Explicit calculations are given for a short parallel-plate channel. The variational principles discussed here may be applied to a wide range of problems involving linear inhomogeneous integral equations.

1. INTRODUCTION

IN the so-called Knudsen regime, a gas flowing through a channel is at sufficiently low pressures for molecule–molecule collisions to be negligible, and only molecule–wall collisions need be taken into account. In addition, one usually assumes that a molecule equilibrates with the wall at each collision, so that its paths before and after the collision are totally uncorrelated.

Now, suppose that we have a gas at equilibrium in front of the entrance of the channel (Fig. 1), at a density of n molecules per cc, that the exit of the channel opens into a vacuum, and that these conditions are maintained until a steady state has been reached. If the mean molecular speed at the prevailing temperature is $\bar{v}$, then $I_0 = \frac{1}{4}n\bar{v}A$ is the number of molecules entering the channel per unit time, A being the area of the entrance cross section. Of these, a number Q/sec will leave through the exit, never to return, while the remainder come back out through the entrance.

For all but the simplest channel geometries, the calculation of Q is an exceedingly difficult problem,[1] and must be treated by approximate methods. One of the most promising has been the variational procedure introduced by DeMarcus,[2] which leads to

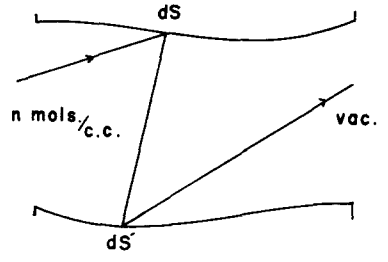


FIG. 1. Channel with sample molecular path.

* Present address: University of Notre Dame, South Bend, Indiana.

[1] E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938).
[2] W. C. DeMarcus, in *Advances in Applied Mechanics* Suppl. 1, *Rarefied Gas Dynamics* (Academic Press Inc., New York, 1961), p. 161.

rigorous upper bounds on the transmission probability Q/I_0. In this paper, we develop a companion procedure to bound Q/I_0 from below as well.[3]

2. MATHEMATICAL FORMULATION

Steady-state Knudsen flow through a channel of arbitrarily complex geometry is characterized by the number of molecules $\psi(S)$ colliding with unit area of the channel wall per unit time at each point S on the wall. The function satisfies the Clausing–DeMarcus integral equation,[2]

$$\psi(S) = I_0\phi(S) + \int K(S', S)\psi(S')\, dS', \qquad (1)$$

where $\phi(S)\, dS$ is the probability that a molecule entering the channel will make its *first* wall collision with the surface element dS located at S (Fig. 1), $K(S', S)\, dS$ is the probability that a molecule leaving the wall at S′ will make its next collision in dS at S, and the integral extends over the entire wall. Under the assumption that the molecule "forgets" its past history upon each collision, K and ϕ will depend only on the geometry of the system. Equation (1) is just a mass balance condition, and states that a molecule colliding with any portion of the channel wall must have come either from the entrance or from some other element of the wall. We find it more convenient in what follows to rewrite Eq. (1) in the form

$$I_0\phi(S) - \psi(S)M(S) - \int K(S', S) \times [\psi(S) - \psi(S')]\, dS' = 0, \qquad (2)$$

where

$$M(S) = 1 - \int K(S', S)\, dS'.$$

Once $\psi(S)$ is known, the ratio Q/I_0 is readily obtained from the relation[2]

$$\Omega/I_0 = 1 - A\int \psi(S)\phi(S)\, dS/I_0. \qquad (3)$$

The integral in Eq. (3) represents the rate at which molecules leave the channel through the entrance.

3. UPPER BOUND ON Q/I_0

In this section we rederive the variation principle of Davison[4] and DeMarcus[2] by a somewhat different route, so as to make clear its relationship to the reciprocal principle discussed in the next section. We show that solving Eq. (2) is equivalent to minimizing the functional

$$\Omega\{G, \psi\} \equiv \frac{1}{2}\iint K(S, S')G^2(S, S')\, dS\, dS' + \int \psi^2(S)M(S)\, dS \qquad (4)$$

with respect to G and ψ, subject to the subsidiary conditions that

$$Z \equiv \int \psi(S)\phi(S)\, dS \qquad (5)$$

has a specified value and that

$$G(S, S') = \psi(S) - \psi(S') \qquad (6)$$

for all S and S′. That a minimum value of Ω exists is evident from the fact that the coefficients of G^2 and ψ^2 are never negative, i.e.,

$$K(S, S') \geq 0, \quad M(S) \geq 0;$$

the latter inequality follows from the interpretation of $\int K(S, S')\, dS'$ as the probability that a molecule leaving the wall at S does not pass directly out of the channel.

The conditions (5) and (6) may be introduced through Lagrangian multipliers ω_1 and $\omega_2(S, S')$; instead of Ω, we then minimize the functional

$$\Omega'\{G, \psi\} \equiv \Omega\{G, \psi\} + \omega_1 Z + \iint \omega_2(S, S') \times [G(S, S') - \psi(S) + \psi(S')]. \qquad (7)$$

For the desired extremum the variation of Ω' must vanish,

$$\iint \delta G(S, S')[K(S, S')G(S, S') + \omega_2(S, S')]\, dS\, dS' + \int \delta\psi(S)\left[2\psi(S)\left(1 - \int K(S, S')\, dS'\right) + \omega_1\phi(S) - \int (\omega_2(S, S') - \omega_2(S', S))\, dS'\right] dS = 0. \qquad (8)$$

Setting the coefficients of δG and $\delta\psi$ independently equal to zero, we have the Euler–Lagrange equations

$$K(S, S')G(S, S') + \omega_2(S, S') = 0 \qquad (9)$$

and

$$2\psi(S)M(S) + \omega_1\phi(S) - \int [\omega_2(S, S') - \omega_2(S', S)]\, dS = 0, \qquad (10)$$

from which ω_2 and G can be eliminated by using the subsidiary condition (6). Because of the symmetry of the kernel $K(S, S')$ with respect to the interchange of S and S′, the result can be written as

$$\omega_1\phi(S) + 2\psi(S)M(S) + 2\int K(S', S) \times [\psi(S) - \psi(S')]\, dS' = 0. \qquad (11)$$

[3] For a general discussion of reciprocal variational principles see J. L. Synge, *The Hypercircle in Mathematical Physics* (Cambridge University Press, London, 1957).

[4] B. Davison, Phys. Rev. 71, 694 (1947).

Equation (11) is identical with (2), provided we make the identification

$$\omega_1 = -2I_0. \tag{12}$$

The minimum value of the functional Ω has a simple meaning, as can be seen by multiplying (2) on both sides by $\psi(S)$ and integrating over the wall surface. After some reduction one obtains

$$\Omega = I_0 \int \psi(S)\phi(S)\, dS = I_0 Z. \tag{13}$$

Of course, this relation is only valid if $\psi(S)$ actually satisfies (2): if instead we calculate Ω from trial functions $\psi^* \neq \psi$ and $G^* = \psi^*(S) - \psi^*(S') \neq G$, the result will necessarily be larger:

$$\Omega^* = \Omega\{G^*, \psi^*\} > I_0 Z. \tag{14}$$

It should be borne in mind that, in order for (14) to hold, ψ^* must satisfy condition (5). In other words, (14) is true only for trial functions which lead to the required value of Z. On the other hand, different choices for ψ^* lead to different values of the Lagrangian multiplier $\omega_1 = -2I_0$. Thus (14) gives an upper bound on I_0 for a fixed value of Z. Recalling that, according to Eq. (3),

$$Z = (I_0 - Q)/A, \tag{15}$$

we can equally well say that we have an upper bound on Q/Z for a given Z:

$$Q/Z \equiv AQ/[I_0 - Q] < (\Omega^*/Z^2) - A, \tag{16}$$

or, after rearrangement,

$$Q/I_0 < 1 - AZ^2/\Omega^*. \tag{17}$$

This is the bound used by DeMarcus in his treatment of Knudsen flow.

4. LOWER BOUND ON Q/I_0

To obtain a lower bound on the ratio Q/I_0, we reformulate the principle of the preceding section in terms of the net flux $J(S, S')$ between two points S and S' on the wall, and the rate $T(S)$ at which molecules leave the unit area of the wall surface at S to pass directly out through the channel entrance or exit without further collisions. We thus seek to minimize the functional

$$\Lambda\{J, T\} = \frac{1}{2}\iint [J^2(S, S')/K(S, S')]\, dS\, dS' + \int [T^2(S)/M(S)]\, dS \tag{18}$$

with respect to all J and T which are consistent with a given entrance rate I_0, i.e., which satisfy the mass balance condition

$$I_0\phi(S) = T(S) + \int J(S, S')\, dS' \tag{19}$$

at every S. In addition to (19), we require

$$J(S, S') + J(S', S) = 0 \tag{20}$$

for all S and S'.

To take these subsidiary conditions into account, we introduce Lagrangian multipliers $\lambda_1(S)$ and $\lambda_2(S, S')$ and minimize

$$\Lambda'\{J, T\} \equiv \Lambda\{J, T\} + \int \lambda_1(S) \times \left[T(S) + \int J(S, S')\, dS'\right] dS + \iint \lambda_2(S, S')[J(S, S') + J(S', S)]\, dS\, dS'. \tag{21}$$

Setting the variation of Λ',

$$\delta\Lambda' = \iint \delta J(S, S')[J(S, S')/K(S, S') + \lambda_1(S) + \lambda_2(S, S') + \lambda_2(S', S)]\, dS\, dS' + \int \delta T(S)[2T(S)/M(S) + \lambda_1(S)]\, dS, \tag{22}$$

equal to zero for arbitrary variations δJ and δT, we are led to the Euler–Lagrange equations

$$J(S, S') + K(S, S')[\lambda_1(S) + \lambda_2(S, S') + \lambda_2(S', S)] = 0, \tag{23}$$

$$2T(S) + \lambda_1(S)M(S) = 0. \tag{24}$$

When introduced into (16), the requirement (20) that J be antisymmetric with respect to the interchange of S and S' gives

$$\lambda_2(S, S') + \lambda_2(S', S) = -\tfrac{1}{2}[\lambda_1(S) + \lambda_1(S')]; \tag{25}$$

with the identification

$$\lambda_1(S) = -2\psi(S) \tag{26}$$

Eqs. (23) and (24) then become

$$J(S, S') = K(S, S')[\psi(S) - \psi(S')], \tag{27}$$

$$T(S) = M(S)\psi(S). \tag{28}$$

Together with (19), these equations are equivalent to the original Clausing–DeMarcus equation (2).

The minimum value of $\Lambda\{J, T\}$ is the same as the minimum of $\Omega\{G, \psi\}$, namely, $I_0 Z$, and is obtained if and only if T and J satisfy (19), (27), and (28). This time, however, the minimization has been carried out at fixed I_0 [Eq. (19)], rather than at fixed Z. Therefore, if in place of the minimizing J and T we use trial functions J^* and T^* satisfying the conditions (19) and (20), we obtain, *for any given value of* I_0, an upper bound on the product $I_0 Z$:

$$\Lambda^*\{J^*, T^*\} > I_0 Z = I_0(I_0 - Q)/A. \tag{29}$$

This inequality may be rewritten in the form of a lower bound on Q/I_0 to give the final result

$$\Omega/I_0 > 1 - A\Lambda^*/I_0^2. \tag{30}$$

5. SHORT CHANNELS BETWEEN PARALLEL PLATES

To illustrate the variational bounds (17) and (30), we consider the flow through a two-dimensional channel between two parallel lines separated by a distance A.[5] The channel has a length L in the direction of flow, which we take to be the x direction. The functions ϕ and K now depend only upon the x coordinates of the points S and S′: assuming that the gas molecules enter the channel and leave the wall with a two-dimensional cosine law distribution, simple geometrical considerations lead to[6]

$$\phi(x) = (1/2A)[1 - x/(x^2 + A^2)^{\frac{1}{2}}],$$
$$M(x) = \tfrac{1}{2}A[\phi(x) + \phi(L - x)], \quad (31)$$
$$K(|x - x'|) = \tfrac{1}{2}A^2[A^2 + (x - x')^2]^{-\frac{3}{2}}. \quad (32)$$

For a lower bound on Q/I_0, we take J^* identically equal to zero and calculate T^* from (19). It is clear that, in using these trial functions, we are in effect counting only those paths which involve no more than a single collision with the wall surface. The lower bound (30) thus becomes

$$Q/I_0 > 1 - 2\int_0^L \{\phi^2(x)/[\phi(x) + \phi(L - x)]\}\, dx. \quad (33)$$

For short channels, we may expand (33) in powers of $\rho = L/A$ to obtain

$$Q/I_0 > 1 - \rho/2 + \rho^2/4 - \rho^3/24 - \rho^4/12 + \cdots. \quad (34)$$

An upper bound on Q/I_0 for this system was derived by Berman[7] using a linear trial function $\psi = ax + b$; in expanded form, Berman's result is

$$Q/I_0 < 1 - \rho/2 + \rho^2/4 - \rho^3/24 - \rho^4/16 + \cdots, \quad (35)$$

which is identical with (34) through the term in ρ^3.

For very large values of ρ, however, the lower bound (33) goes to zero as ρ^{-1}, whereas Berman's upper bound for long channels is $\rho^{-1} \ln \rho$. While the one-collision trial function is thus quite effective for a short channel, improved trial functions are necessary for larger values of L. For example, one might choose

$$J^*(x, x') = aK(|x - x'|)(x - x'),$$

corresponding to the choice $\psi^*(x) = ax + b$.

6. SUMMARY AND CONCLUSION

To exhibit the bounds on the transmission probability in a more explicit manner, we can use the subsidiary conditions (5) and (6) to eliminate Z and G^* in (17), and the condition (19) to eliminate T^* in (30). In making the latter substitution, we may also write

$$J^*(\mathrm{S}, \mathrm{S}') = \mu j(\mathrm{S}, \mathrm{S}')$$

and, for any choice of the trial function $j(\mathrm{S}, \mathrm{S}')$ which is antisymmetric with respect to interchange of S and S′, maximize the lower bound on Q/I_0 given by (30) with respect to the parameter μ. The resulting inequalities are

$$1 - \frac{A\left[\int \phi(\mathrm{S})\psi^*(\mathrm{S})\, d\mathrm{S}\right]^2}{\frac{1}{2}\iint K(\mathrm{S}, \mathrm{S}')(\psi^*(\mathrm{S}) - \psi^*(\mathrm{S}')^2\, d\mathrm{S}\, d\mathrm{S}' + \int M(\mathrm{S})\psi^{*2}(\mathrm{S})\, d\mathrm{S}} \geq Q/I_0 \geq 1 - A\int \phi^2(\mathrm{S})M^{-1}(\mathrm{S})\, d\mathrm{S}$$
$$+ \frac{A\left(\iint j(\mathrm{S}, \mathrm{S}')\phi(\mathrm{S})M^{-1}(\mathrm{S})\, d\mathrm{S}\, d\mathrm{S}'\right)^2}{\frac{1}{2}\iint j^2(\mathrm{S}, \mathrm{S}')K^{-1}(\mathrm{S}, \mathrm{S}')\, d\mathrm{S}\, d\mathrm{S}' + \iiint j(\mathrm{S}, \mathrm{S}')j(\mathrm{S}, \mathrm{S}'')M^{-1}(\mathrm{S})\, d\mathrm{S}\, d\mathrm{S}'\, d\mathrm{S}''}$$
$$[j(\mathrm{S}, \mathrm{S}') = -j(\mathrm{S}', \mathrm{S})]. \quad (36)$$

Extension of these inequalities to diffuse-elastic scattering from the walls merely involves a modification of the kernel K, and DeMarcus has determined an upper bound on Q/I_0 for flow through cylinders under these conditions. Surface diffusion can also be included if the wall surface is sufficiently irregular for the effect to be appreciable.

Equations of the same form as (1) describe a wide range of phenomena, such as radiation transfer, neutron diffusion, etc. The reciprocal variation principles discussed in this paper apply in those cases where the kernel K is symmetric, and where both K and $[1 - \int K(\mathrm{S}, \mathrm{S}')\, d\mathrm{S}']$ are everywhere positive or zero.

ACKNOWLEDGMENTS

The authors are indebted to Dr. A. S. Berman for many helpful discussions.

This work was supported by a grant from the National Science Foundation.

[5] This is equivalent to three-dimensional flow between infinitely wide parallel plates.

[6] W. C. DeMarcus, United States Atomic Energy Commission Report K-1302 (1957).

[7] A. S. Berman, J. Appl. Phys. 36, 3356 (1965).

Weber's Mixed Boundary-Value Problem in Electrodynamics

O. LAPORTE
University of Michigan, Ann Arbor, Michigan

R. G. FOWLER
University of Oklahoma, Norman, Oklahoma

(Received 29 July 1966)

Electrodynamics problems with mixed boundary values promise to assume increasing practical importance in fields such as plasma physics. A new method of attacking such problems in three dimensions is presented and discussed.

1. STATEMENT OF THE PROBLEM

THE emergence of plasma physics as an important scientific discipline has disclosed many imperfections in the traditional approaches to certain aspects of mathematical physics. One of these is the mixed boundary-value problem in electrodynamics originating when a moderately conducting domain is incompletely bounded by perfect conductors. While it is evident that this problem can be solved in all two-dimensional symmetries, the simplest examples of complete three-dimensional character present great difficulties.

The classical example of such a problem is Nobili's[1] colored rings first considered by Riemann.[2] This is the problem of an infinite slab of conducting material on whose plane sides two circular metallic disks held at opposite potential are placed. The conductivity of the slab is finite, that of the disks infinite. The formulation of the problem is as follows:

A solution of Laplace's equation $\phi(\rho, z)$ is to be found such that for $z = \pm a$,

$$\phi(\rho, \pm a) = \pm\phi_0, \quad \rho < c,$$
$$(\partial\phi/\partial z)_{\pm a} = 0, \qquad \rho > c.$$

Here ρ is the two-dimensional distance $(x^2 + y^2)^{\frac{1}{2}}$, the thickness of the slab is $2a$, the radius of each electrode disk is c. A presumed solution of this cylindrically symmetric, mixed boundary-value problem was given by Weber,[3] but it was actually only an approximation for the case where the disk size is small compared with the slab thickness.

The method presented in the following pages reduces the problem to a standard integral equation of the Fredholm type (integral equation of the "second kind"). In another paper[4] the construction of solutions of this equation is presented in detail together with numerical results. Recently, the same problem was attacked by Tranter[5] using the method of integral transforms and an approximate solution was constructed. However, since the method given here is quite different and capable of considerable generalization (see Sec. 6 below), we think it merits a detailed presentation.

2. THE SINGLE CIRCULAR DISK

As a preparation, let the well-known solution for the metallic disk be rederived. We start with

$$\phi(\rho, z) = \frac{2}{\pi}\int_0^\infty d\lambda J_0(\rho\lambda)e^{-|z|\lambda}A(\lambda), \tag{1}$$

where $A(\lambda)$ is to be determined so that $\phi(\rho, 0)$ is equal to the constant potential ϕ_0 for $\rho < c$. Rather than trying to find $A(\lambda)$ directly, we make the following *Ansatz:*

$$A(\lambda) = \int_0^c d\xi \cos\lambda\xi f(\xi) \tag{2}$$

and attempt to find $f(\xi)$ from

$$\frac{2}{\pi}\int_0^\infty d\lambda J_0(\rho\lambda)\int_0^c d\xi \cos\lambda\xi f(\xi) = \phi_0. \tag{3}$$

Replacing ρ by η momentarily, we now operate on this with $(d/d\rho)\int_0^\rho \eta\, d\eta/(\rho^2 - \eta^2)^{\frac{1}{2}}$:

$$\frac{2}{\pi}\frac{d}{d\rho}\int_0^\rho \frac{\eta\, d\eta}{(\rho^2 - \eta^2)^{\frac{1}{2}}}\int_0^\infty d\lambda J_0(\eta\lambda)\int_0^c d\xi \cos\lambda\xi f(\xi) = \phi_0. \tag{4}$$

After drawing the λ integration to the left, the η integral can be performed:

$$\int_0^\rho \frac{\eta\, d\eta}{(\rho^2 - \eta^2)^{\frac{1}{2}}} J_0(\eta\lambda) = \frac{\sin\lambda\rho}{\lambda}$$

[1] L. Nobili, Poggendorf Ann. **9**, 183 (1827); **10**, 393, 410 (1827).
[2] B. Riemann, Poggendorf Ann. **95**, 130 (1855).
[3] H. Weber, Z. Angew. Math. **75**, 75 (1873).
[4] O. Laporte and R. G. Fowler, Phys. Rev. **148**, 170 (1966).
[5] C. J. Tranter, Quart. J. Math. **2**, 60 (1951).

and (3) becomes

$$\frac{2}{\pi}\frac{d}{d\rho}\int_0^\infty d\lambda \frac{\sin \lambda\rho}{\lambda}\int_0^c d\xi \cos \lambda\xi f(\xi) = \phi_0.$$

Again exchanging the order of integration gives

$$\frac{2}{\pi}\frac{d}{d\rho}\int_0^c d\xi f(\xi)\int_0^\infty \frac{d\lambda}{\lambda}\sin \lambda\rho \cos \lambda\xi = \phi_0, \quad (5)$$

in which the λ integral is recognizable as the "Dirichlet Discontinuous Factor." Therefore, we have

$$f(\rho) = \phi_0,$$

and $A(\lambda)$ of (2) can now be substituted into (1).

3. THE SLAB PROBLEM

Exactly the same method is now employed for the slab, with the modification that the vanishing of the normal derivative at the plates is brought about by assuming infinitely many equidistant image plates held at alternate potentials. (See Fig. 1.) Let a potential of the following form be assumed:

$$\phi = \frac{2}{\pi}\int_0^\infty d\lambda(\cdots + e^{-|z+3a|\lambda} - e^{-|z+a|\lambda} + e^{-|z-a|\lambda} - e^{-|z-3a|\lambda}\cdots)J_0(\rho\lambda)A(\lambda) \quad (6)$$

with $A(\lambda)$ given by (2). This can be summed to be

$$\phi = \frac{2}{\pi}\int_0^\infty d\lambda \frac{\sinh \lambda z}{\cosh \lambda a}J_0(\rho\lambda)A(\lambda), -a \le z \le +a. \quad (6a)$$

With solutions of this form we now seek to satisfy the boundary conditions on each of the plates. Let

$$\phi = +\phi_0 \quad \text{for} \quad z = (4m + 1)a,$$
$$\phi = -\phi_0 \quad \text{for} \quad z = (4m - 1)a.$$

Substituting these conditions into (6) leads to results such as the following: For

$$z = -3a: \quad +\phi_0 = \frac{2}{\pi}\int_0^\infty(\cdots + 1 - e^{-2a\lambda} + e^{+4a\lambda} - e^{-6a\lambda} + \cdots)J_0A\, d\lambda,$$

$$z = -a: \quad -\phi_0 = \frac{2}{\pi}\int_0^\infty(\cdots + e^{-2a\lambda} - 1 + e^{-2a\lambda} - e^{-4a\lambda}\cdots)J_0A\, d\lambda,$$

$$z = +a: \quad +\phi_0 = \frac{2}{\pi}\int_0^\infty(\cdots + e^{-4a\lambda} - e^{-2a\lambda} + 1 - e^{-2a\lambda}\cdots)J_0A\, d\lambda.$$

These and all other boundary conditions are seen to be identical, and give after summing

$$\frac{2}{\pi}\int_0^\infty d\lambda J_0(\lambda\rho)\tanh a\lambda\int_0^c d\xi \cos \lambda\xi f(\xi) = \phi_0. \quad (7)$$

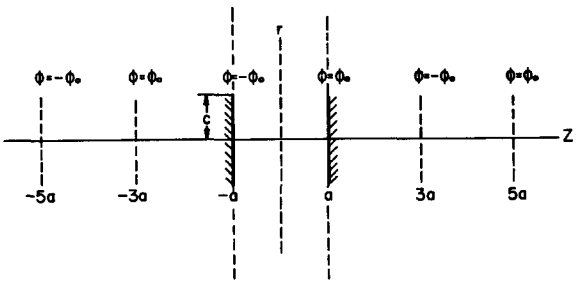


FIG. 1. Geometry of the image system for the two-disk problem.

This equation, which should be compared with (3) of the previous section is now subjected to the same transformations, which in the previous section led from (3) to (5). The result is

$$\frac{d}{d\rho}\int_0^c d\xi f(\xi)\int_0^\infty \frac{d\lambda}{\lambda}\tanh a\lambda \sin \lambda\rho \cos \lambda\xi = \phi_0. \quad (8)$$

In order to perform the differentiation with respect to ρ with complete safety, let the hyperbolic tangent be split by writing

$$\tanh a\lambda = 1 - 2/(e^{2a\lambda} + 1).$$

This results in the integral equation

$$f(\xi) - \frac{2}{\pi}\int_{-c}^{+c} d\xi_1 K(\xi\xi_1)f(\xi_1) = \phi_0, \quad (8a)$$

with the symmetric kernel

$$K(\xi\xi_1) = \int_0^\infty d\lambda \frac{\cos \lambda\xi \cos \lambda\xi_1}{e^{2a\lambda} + 1}. \quad (8b)$$

The range of integration with respect to ξ was extended to $-c$ by assuming $f(\xi)$ to be even.

4. EXPANSION IN LEGENDRE POLYNOMIALS

We introduce dimensionless variables into the integral equation by writing

$$\xi = cx, \quad \xi_1 = cx_1, \quad \lambda = \mu/2a, \quad f(\xi)/\phi_0 = F(x) \quad (9)$$

and later also into the equation for ϕ itself as

$$\rho/c = \sigma, \quad z/a = \zeta. \quad (10)$$

Equations (8a, b) are now

$$F(x) - \frac{2}{\pi}\epsilon\int_{-1}^{+1} dx_1 F(x_1)K(x, x_1) = 1, \quad (11a)$$

with

$$K(x, x_1) = \int_0^\infty d\mu \frac{\cos \epsilon\mu x \cos \epsilon\mu x_1}{e^\mu + 1}. \quad (11b)$$

The constant

$$\epsilon = c/2a$$

is the important ratio of the problem. For infinite plate distance (11a) reduces to $F(x) = 1$, and the

potential $\phi(z, \rho)$ becomes the original Weber expression

$$\phi(z, \rho) = \frac{2}{\pi}\phi_0 \int_0^\infty d\lambda \frac{\sinh z\lambda}{\cosh a\lambda} \frac{\sinh \lambda a}{\lambda} J_0(\rho\lambda).$$

Because of the range of the variables x and x_1, it was found most convenient to expand the unknown function $F(x)$ into a series of even Legendre polynomials

$$f(x) = \sum_0^\infty A_{2m} P_{2m}(x). \tag{12}$$

The integral equation (11a) now becomes an infinite system of linear equations for the expansion coefficients A_{2m}

$$(4n + 1)^{-1} A_{2n} - (\epsilon/\pi) \sum_m A_{2m} M_{2m,2n} = \delta_{0,2n},$$

with the matrix elements

$$M_{2m,2n} = \int_{-1}^{+1} dx \int_{-1}^{+1} dx_1 P_{2m}(x) P_{2n}(x_1) K(xx_1).$$

In the paper referred to in Ref. 4, we have reported the calculations necessary to obtain actual solutions and have shown that not only the matrix M decreases satisfactorily with increasing m and n, thereby making early truncation possible, but also the A_{2n} decrease rapidly.

5. THE FIELD FOR $\rho > c$ AS A RAPIDLY CONVERGING FOURIER SERIES

The expression for the complete potential $\phi(\rho, z)$ is using (6a), (2), (9), and (12)

$$\frac{\phi}{\phi_0} = \left(\frac{2\epsilon}{\pi}\right)^{\frac{1}{2}} \sum_m (-1)^m A_{2m} \times \int_0^\infty \frac{d\mu}{\mu^{\frac{1}{2}}} \frac{\sinh \frac{1}{2}\mu\zeta}{\cosh \frac{1}{2}\mu} J_0(\sigma\epsilon\mu) J_{2m+\frac{1}{2}}(\epsilon\mu).$$

The appearance of the half integer Bessel functions is explained below. The occurrence of the hyperbolic cosine in the denominator shows that in the complex μ plane there is a string of first-order poles along the imaginary axis located at

$$\mu = (2n + 1)\pi i.$$

This, therefore, invites one to decompose J_0 into two Hankel functions $H_0^{(1)}$ and $H_0^{(2)}$ and draw the integral containing the former into the upper, and the one containing the latter into the lower half-plane. (The integrals along the two large quarter circles in the first and fourth quadrants do not contribute, as can be shown readily.) Each integral of the m series thus becomes the sum of an integral from zero to $i\infty$ containing $H_0^{(1)}$ and a second one from zero to $-i\infty$ with $H_0^{(2)}$. Due to the circulation relation

$$H_0^{(1)}(iz) = -H_0^{(2)}(-iz),$$

we cancel these integrals, as soon as the paths become symmetrical with respect to the origin. As the "pictorial equation" Fig. 2 shows this can be achieved while, at the same time, the residues at the poles along the positive imaginary axis have to be taken into account. The result,

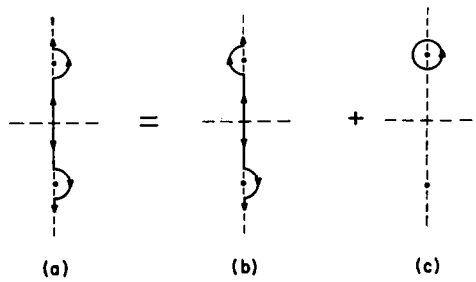


FIG. 2. Decomposition of the path of complex integration for the potential.

$$\frac{\phi}{\phi_0} = \left(\frac{2\epsilon}{\pi}\right)^{\frac{1}{2}} 4\pi \sum_m (-1)^m A_{2m} \sum_{n=0}^\infty (-1)^n H_0^{(1)}[i\sigma\epsilon(2n + 1)\pi] \times \frac{J_{2m+\frac{1}{2}}[i\epsilon(2n + 1)\pi]}{[i(2n + 1)\pi]^{\frac{1}{2}}} \sin \tfrac{1}{2}(2n + 1)\pi\zeta,$$

although a double series, should be very useful for the numerical calculation in the space outside the cylinder formed by the two plates $\sigma > 1$, where it converges rapidly. The appearance of the imaginary unit is only apparent. No corresponding expression for $\sigma < 1$, i.e., $\rho < a$ seems to exist.

6. THE NORMAL DERIVATIVE FOR $z = a$

A more detailed calculation of the normal derivative nets us the surface charge on the plates and also serves as a check on the fulfillment of the boundary condition for $\rho > c$. We have, from (2) and (6a) and using (9) and (10).

$$\left(\frac{\partial(\phi/\phi_0)}{\partial\zeta}\right)_{\zeta=1} = \frac{\epsilon}{\pi} \int_0^\infty \mu\, d\mu J_0(\epsilon\sigma\mu) \times \int_0^1 dx \cos \epsilon\mu x \sum A_{2m} P_{2m}(x).$$

The x integration can be performed and leads to spherical Bessel functions, so that we have

$$\left(\frac{\partial\phi/\phi_0}{\partial\zeta}\right)_{\zeta=1} = \frac{1}{(2\pi\epsilon)^{\frac{1}{2}}} \sum A_{2m} \times \int_0^\infty \mu_1^{\frac{1}{2}}\, d\mu_1 J_0(\sigma\mu_1) J_{2m+\frac{1}{2}}(\mu_1),$$

where, for the sake of simplicity, $\mu_1 = \epsilon\mu$ is introduced as variable of integration. The integrals which appear

here belong to the family of the Sonine–Schafheitlin discontinuous integrals.[6] They all represent different hypergeometric functions of σ, according as this variable is less or greater than one. What is of interest to us here is that for $\sigma > 1$ they all vanish, so that we have the result

$$[\partial(\phi/\phi_0)/\partial\zeta]_{\zeta=1} = 0, \quad \rho > c.$$

It is therefore seen that our form of solution does indeed satisfy the boundary condition outside the disks.

On the disks, i.e., for $\sigma < 1$ the reduction of these integrals to hypergeometric series would not constitute a particular advantage, were it not for the fact that these series can all be summed and reduced to Jacobi polynomials.[7] The result for the normal derivative is therefore, for $\sigma < 1$,

$$\left(\frac{\partial(\phi/\phi_0)}{\partial\zeta}\right)_{\zeta=1} = \frac{1}{2\pi\epsilon}\sum_{m=0}^{\infty} 2^{2m}\frac{(m-1)!}{(2m-1)!}A_{2m} \times \left(\frac{d}{d\tau}\right)^m [\tau^m(1-\tau)^{2m-1/2}], \quad (13)$$

where $\tau = \sigma^2$. For $m = 0$, unity should be substituted for the quotient of the two factorials. The ratio of Eq. (13) to Eq. (14) is plotted in Fig. 3.

To calculate the total charge, or for the current problem, the reciprocal resistance, the normal derivative has to be integrated over the disk surface. Because of the appearance of the m fold derivative with respect to $\tau = \sigma^2$, it is immediately seen that the contributions of all series terms with $m > 0$ vanish. The result is

$$\int_0^1 \sigma\, d\sigma\left(\frac{\partial(\phi/\phi_0)}{\partial\zeta}\right)_{\zeta=1} = \frac{1}{2\pi\epsilon}A_0. \quad (14)$$

7. THE PROBLEM OF IMPERFECT INFINITE ELECTRODES

Case I: A Single Plate

The family of inverse problems to those just solved presents interesting aspects. We consider first the case of a single infinite plate with a circular hole in it in a partially conducting medium, the hole being closed by a nonconductor. Here, the boundary conditions are

$$\phi = 0, \quad \rho > c, \quad z = 0,$$
$$\partial\phi/\partial z = 0, \quad \rho < c, \quad z = 0.$$

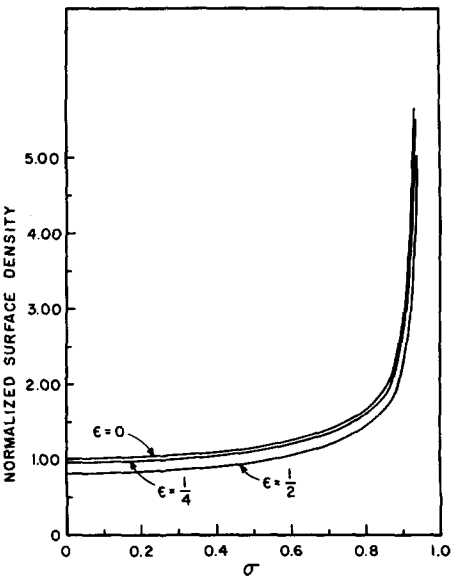


FIG. 3. The surface density of charge on either disk of the two-disk problem.

We introduce

$$\phi = -kz + \omega.$$

Then the new function ω fulfills the boundary conditions

$$\omega = 0, \quad \rho > c, \quad z = 0,$$
$$(\partial\omega/\partial z) = +k, \quad \rho < c, \quad z = 0.$$

This time, we propose to find the function $\partial\omega/\partial z$, and to make it satisfy the condition outside automatically. We assume that

$$\frac{\partial\omega}{\partial z} = \frac{2}{\pi}\int_0^\infty d\lambda e^{-\lambda z}J_0(\rho\lambda)A(\lambda)$$

with the new *Ansatz* that

$$A(\lambda) = \lambda\int_0^c \sin\lambda\xi f(\xi)\, d\xi.$$

Once again we rename ρ as η, and now operate on both sides with a new choice of operator, namely:

$$\frac{1}{\rho}\frac{d}{d\rho}\int_0^\rho \eta\, d\eta(\rho^2-\eta^2)^{\frac{1}{2}}.$$

Then

$$k\rho = \frac{2}{\pi\rho}\frac{d}{d\rho}\int_0^\rho \eta\, d\eta(\rho^2-\eta^2)^{\frac{1}{2}} \times \int_0^\infty \lambda\, d\lambda J_0(\lambda\eta)\int_0^c \sin\lambda\xi f(\xi)\, d\xi.$$

The integral over η can be performed by use of the discontinuous integrals of Weber and Schafheitlin. It yields

$$k\rho = \frac{2}{\pi\rho}\frac{d}{d\rho}\int_0^\infty \lambda\, d\lambda\left((\tfrac{1}{2}\pi)^{\frac{1}{2}}\frac{\rho^{\frac{3}{2}}}{\lambda^{\frac{3}{2}}}J_{\frac{3}{2}}(\rho\lambda)\right)\int_0^c \sin\lambda\xi f(\xi)\, d\xi.$$

[6] N. Nielsen, *Handbuch der Theorie der Cylinderfunktionen* (B. G. Teubner, Leipzig, 1905), formulas (4) and (11) of Sec. 74, p. 191 *et seq.* See also G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1958), Sec. 13.4, Eq. (2), p. 401.

[7] See R. Courant and D. Hilbert, *Methoden der Mathematischen Physik* (Julius Springer, Leipzig, 1924), Vol. I, p. 74.

The λ integral, after interchanging with ξ integration, can be reduced to

$$\frac{\rho^{\frac{3}{2}}}{\xi^{\frac{1}{2}}}\int_0^\infty d\lambda J_{\frac{3}{2}}(\rho\lambda)J_{\frac{1}{2}}(\xi\lambda),$$

and this is one of the generalized Dirichlet factors whose value is 1 or 0 according as $\xi < \rho$ or $\xi > \rho$. Therefore, the equation reduces to

$$k\rho = \frac{1}{\rho}\frac{d}{d\rho}\int_0^\rho \xi f(\xi)\,d\xi,$$

the solution of which is $f(\xi) = k\xi$. When this is returned to the equation for $A(\lambda)$, the ξ integration can be performed, yielding

$$\partial\omega/\partial z = (\tfrac{1}{2}\pi)^{\frac{1}{2}}c^{\frac{3}{2}}k\int_0^\infty e^{-\lambda z}J_0(\rho\lambda)J_{\frac{3}{2}}(\lambda c)\lambda^{\frac{1}{2}}\,d\lambda.$$

Hence

$$\omega = -(\tfrac{1}{2}\pi)^{\frac{1}{2}}c^{\frac{3}{2}}k\int_0^\infty e^{-\lambda z}J_0(\rho\lambda)J_{\frac{3}{2}}(\lambda c)\frac{d\lambda}{\lambda^{\frac{1}{2}}},$$

and it can be shown that this integral does indeed have the desired property that it is zero for $\rho > c$, $z = 0$.

Case II. Two Plates

The second case is that of two electrodes with opposite holes closed with a nonconductor. Here, the generalization employed in moving from the one-disk problem to the two-disk problem can be carried out again. We again imagine an infinite set of image plates so charged that our problem repeats between each pair of plates. As in the previous case, we note the need for a uniform field in the case of infinite plates, and let

$$\phi = -kz + \omega.$$

Then the boundary conditions are

$$z = a, \quad \rho < c, \quad (\partial\phi/\partial z) = 0, \quad (\partial\omega/\partial z) = k, \tag{15a}$$

$$z = -a, \quad \rho < c, \quad (\partial\phi/\partial z) = 0, \quad (\partial\omega/\partial z) = k, \tag{15b}$$

$$z = a, \quad \rho > c, \quad \phi = \phi_0, \quad \omega = \phi_0 + ka, \tag{15c}$$

$$z = -a, \quad \rho > c, \quad \phi = -\phi_0, \quad \omega = -\phi_0 - ka. \tag{15d}$$

If now we choose $\phi_0 = -k/a$, then $\omega = 0$ at $z = \pm a$.

An examination of the possibilities shows that the solution of the Laplace equation which fits the need to have the function be antisymmetrical, and zero at $z = \pm a$, while its derivative is symmetrical over the region $-a \le z \le a$ is

$$\frac{\partial\omega}{\partial z} = \frac{2}{\pi}\int_0^\infty d\lambda \frac{\cosh\lambda z}{\sinh\lambda a}J_0(\rho\lambda)A(\lambda) \tag{16}$$

and

$$\omega = \frac{2}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\frac{\sinh\lambda z}{\sinh\lambda a}J_0(\rho\lambda)A(\lambda).$$

Following the method of the single plate case, we search for an equation for $A(\lambda)$ which will satisfy (15a) and (15b). This leads to the integral equation

$$f(\rho) + \frac{4}{\pi}\int_0^a d\xi f(\xi)K_2(\rho\xi) = -k\rho,$$

where

$$K_2(\rho\xi) = \int_0^\infty d\lambda \frac{\sin\lambda\rho\sin\lambda\xi}{e^{2a\lambda} - 1}$$

is again a symmetric kernel with properties analogous to those of the two-disk kernel.

Once more, with the introduction of the dimensionless variables used before, the integral equation can be solved in terms of Legendre polynomials, this time of odd order. Letting

$$F(x) = \sum_0^\infty B_{2m+1}P_{2m+1}(x),$$

one finally obtains the formal solution

$$\omega(\rho, z) = \left(\frac{2c}{\pi}\right)^{\frac{1}{2}}k\sum_{m=0}^\infty (-1)^m B_{2m+1} \times \int_0^\infty \frac{d\lambda}{\lambda^{\frac{1}{2}}}\frac{\sinh\lambda z}{\sinh\lambda a}J_0(\lambda\rho)J_{2m+\frac{3}{2}}(\lambda c).$$

When $z = \pm a$, the integral over λ is a discontinuous function for all m, which has a zero value for $\rho > c$, so that the boundary conditions (15c) and (15d) are fulfilled identically.

8. CONCLUSION

The application of this method to hydrodynamical problems of interest such as counterflowing liquids that might be present in heat exchangers or reaction cells is evident. Our method has certain points in common with a paper by Sommerfeld[8] in which he employs a series of discontinuous integrals as the starting point for the solution of the problem of an oscillating disk.

[8] A. Sommerfeld, Ann. Physik **42**, 389 (1943).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Eigenvalue Problem for Lagrangian Systems*

E. M. BARSTON
Courant Institute of Mathematical Sciences, New York University, New York, New York

(Received 2 September 1966)

The problem of small oscillations about a state of steady motion of a Lagrangian system is considered. Upper and lower bounds for the growth rates of unstable systems are obtained; sufficient conditions for instability are given for finite dimensional systems; an existence theorem for stable modes for systems with an infinite number of degrees of freedom is presented (valid when the operators are completely continuous in Hilbert space); and finally the orthogonality and completeness properties of the modes of stable finite dimensional systems are discussed.

I. INTRODUCTION

LAGRANGIAN formulations of the equations of motion have been given for a wide class of hydrodynamic and plasma systems,[1-3] as well as for classical mechanical systems with a finite number of degrees of freedom.[4-7] The problem of the stability of stationary states of such systems often leads to an equation of the form[4-10]

$$H_1\ddot{\eta} + 2A_1\dot{\eta} + H_2\eta = 0.$$

Assuming a time dependence $\exp(i\omega t)$, one obtains the following eigenvalue problem[4-10]:

$$\omega^2 H_1\eta - 2\omega i A_1\eta - H_2\eta = 0, \qquad (1)$$

where H_1 and H_2 are linear Hermitian operators (usually real) defined on some complex linear vector space E, A_1 is a linear anti-Hermitian operator (usually real or pure imaginary) on E, H_1 is positive definite, and the eigenvector η corresponding to the eigenvalue ω is a nonzero member of E.

The most comprehensive discussions of the solutions of Eq. (1) (for nontrivial A_1) appear to be due to Routh[4-6] and Whittaker.[7] Both authors were concerned only with finite dimensional systems and relied heavily on the theory of determinants; Routh derived a number of eigenvalue properties for positive or negative definite operators (in particular, both he and Whittaker demonstrated stability for H_1 and H_2 positive definite), and gave a necessary and sufficient condition for stability (the Routh–Hurwitz criterion) valid for real H_1, A_1, and H_2; Whittaker obtained the general solution for H_1 and H_2 positive definite by directly integrating the equations of motion in Hamiltonian form. The problem seems to have received little further attention, outside of discussions of marginal stability[8,9] and forced oscillations.[11] In this paper we present a number of new results, many of which are valid when E is a Hilbert space, and extend some of Routh's results for finite dimensional space to Hilbert space. In particular, in Sec. II we obtain upper and lower bounds for complex eigenvalues; sufficient conditions for instability are given in Sec. III, valid for finite dimensional systems; Sec. IV consists of an existence theorem for real eigenvalues and corresponding eigenfunctions for completely continuous operators in Hilbert space; and Sec. V concerns itself with orthogonality relations and completeness properties of the eigenvectors for H_2 positive definite.

II. BOUNDS FOR COMPLEX EIGENVALUES

In this section we derive upper and lower bounds for the real and imaginary parts of an eigenvalue ω, assuming that $\operatorname{Im}\omega \neq 0$. It proves to be convenient to assume that $H_1 = I$. Indeed, if H_1 is a bounded positive definite linear Hermitian operator with a bounded inverse H_1^{-1}, then Eq. (1) can be cast into the form

$$(\omega^2 I - 2i\omega A - H)\xi = 0, \qquad (2)$$

where $A \equiv SA_1S$, $H \equiv SH_2S$, $\xi \equiv S^{-1}\eta$, and $S \equiv (H_1^{-1})^{\frac{1}{2}}$, $S^{-1} = (H_1)^{\frac{1}{2}}$. The operator S is Hermitian, so that iA and H are Hermitian if iA_1 and H_2 are

* The work presented here was supported by the U.S. Atomic Energy Commission under Contract AT(30-1)1480.

[1] F. E. Low, Proc. Roy. Soc. (London) **A248**, 283 (1958).
[2] P. A. Sturrock, Ann. Phys. (N.Y.) **4**, 306 (1958).
[3] A. H. Taub, Proc. Symp. Appl. Math. **1**, 148 (1949).
[4] E. T. Routh, *The Advanced Part of a Treatise on the Dynamics of a System of Rigid Bodies* (The Macmillan Company, New York, 1905), 6th ed., Chaps. 3, 6, 7.
[5] E. T. Routh, *Essay on the Stability of Motion* (1877).
[6] J. W. Strutt (Lord Rayleigh), *The Theory of Sound* (Dover Publications, Inc., New York, 1945), Chap. 5, 2nd ed.
[7] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge University Press, New York, 1937), 4th ed., Chap. 7.
[8] F. E. Low, Phys. Fluids **4**, 842 (1961).
[9] G. Laval, R. Pellat, M. Cotsaftis, and M. Trocheris, Nucl. Fusion **4**, 25 (1964).
[10] E. M. Barston, Phys. Rev. **139**, A394 (1965).
[11] R. E. D. Bishop, G. M. L. Gladwell, and S. Michaelson, *The Matrix Analysis of Vibration* (Cambridge University Press, New York, 1965), Chap. 5.

Hermitian. We restrict our attention to Eq. (2), and assume throughout that H and iA are linear Hermitian operators defined on a linear vector space E with a complex inner product (f, g) defined for all f and g in E. Further assumptions are introduced as required.

A sufficient (but not necessary) condition for all the eigenvalues ω of Eq. (2) to be real is that H be nonnegative definite.[4–10] Indeed, suppose that Eq. (2) holds. Forming the inner product of (2) with ξ and solving for ω we obtain

$$\omega = \frac{(\xi, iA\xi)}{(\xi, \xi)} \pm \left\{\left[\frac{(\xi, iA\xi)}{(\xi, \xi)}\right]^2 + \frac{(\xi, H\xi)}{(\xi, \xi)}\right\}^{\frac{1}{2}}. \quad (3)$$

The operators are Hermitian, so that the inner products are real. Defining $\omega_1 \equiv \operatorname{Re}\omega$, $\omega_2 \equiv \operatorname{Im}\omega$, Eq. (3) and $\omega_2 \neq 0$ then imply

$$\omega_1 = (\xi, iA\xi)/(\xi, \xi), \quad (4)$$

$$|\omega|^2 = \omega_1^2 + \omega_2^2 = -(\xi, H\xi)/(\xi, \xi). \quad (5)$$

Therefore, for $\omega_2 \neq 0$

$$|\omega|^2 \leq -\lambda_0, \quad \lambda_0 \equiv \inf_E \frac{(\zeta, H\zeta)}{(\zeta, \zeta)}. \quad (6)$$

The complex eigenvalues of Eq. (2) all lie in a circle of radius $|\lambda_0|^{\frac{1}{2}}$ centered at the origin of the complex ω plane. Assuming that λ_0 is an eigenvalue of Eq. (2) for $A = 0$ (this will be the case, for example, if E is a Hilbert space and H is completely continuous), we see that the introduction of a nonzero A into Eq. (2) does not increase the maximum growth rate of the system. In fact, we show that ω_2 tends to zero as

$$\inf_E \frac{(A\zeta, A\zeta)}{(\zeta, \zeta)}$$

approaches infinity. Suppose that Eq. (2) holds, so that $H_\omega \xi = 0$, where $H_\omega \equiv \omega^2 I - 2\omega iA - H$. Then

$$\{H_\omega^2 + 2[2\omega iA + H]H_\omega\}\xi = 0 \quad (7)$$

or

$$\{\omega^4 I + 4\omega^2 A^2 - 2\omega[HiA + iAH] - H^2\}\xi = 0. \quad (8)$$

Taking the inner product with ξ, and equating the real and imaginary parts of the resulting expression equal to zero, we obtain

$$\{[\omega_1^2 - \omega_2^2]^2 - 4\omega_1^2\omega_2^2\}(\xi, \xi) + 4(\omega_1^2 - \omega_2^2)(\xi, A^2\xi) - 2\omega_1(\xi, [HiA + iAH]\xi) - (\xi, H^2\xi) = 0, \quad (9)$$

$$\omega_2\{2\omega_1(\omega_1^2 - \omega_2^2)(\xi, \xi) + 4\omega_1(\xi, A^2\xi) - (\xi, [HiA + iAH]\xi)\} = 0. \quad (10)$$

Assuming $\omega_2 \neq 0$, and eliminating $(\xi, [HiA + iAH]\xi)$ from Eq. (9) by means of (10), we find

$$|\omega|^2\{\omega_2^2 - 3\omega_1^2 + 4(A\xi, A\xi)/(\xi, \xi)\} - (\xi, H^2\xi)/(\xi, \xi) = 0. \quad (11)$$

Equation (11) is quadratic in ω_2^2, and the only nonnegative root of (11) gives

$$\omega_2^2 = [y^2 + (\xi, H^2\xi)/(\xi, \xi)]^{\frac{1}{2}} - y - \omega_1^2, \quad (12)$$

where

$$y \equiv 2\{(A\xi, A\xi)/(\xi, \xi) - \omega_1^2\}. \quad (13)$$

Equation (4) and the Cauchy–Schwarz inequality imply that $y \geq 0$. The following results can now be easily established:

Theorem I: Assume that Eq. (2) holds and that $\omega_2 \neq 0$. Then the following inequalities hold:

$$\omega_1^2 \leq \|A\|^2, \quad (14)$$

$$[4\|A\|^4 + \mu_H]^{\frac{1}{2}} - 2\|A\|^2 \leq |\omega|^2 \leq \|H\|^2/\mu_A, \quad (15)$$

$$-\lambda_0^{-1}\mu_H - 4\|A\|^2 \leq \omega_2^2 \leq [x^2 + \|H\|^2]^{\frac{1}{2}} - x, \quad (16)$$

where

$$x \equiv 2(\mu_A + \lambda_0), \quad \mu_A \equiv \inf_E \frac{(A\zeta, A\zeta)}{(\zeta, \zeta)},$$

$$\mu_H \equiv \inf_E \frac{(H\zeta, H\zeta)}{(\zeta, \zeta)}.$$

Proof: Equation (14) is obtained by applying the Cauchy–Schwarz inequality to Eq. (4). Equation (12) gives

$$|\omega|^2 = [y^2 + (\xi, H^2\xi)/(\xi, \xi)]^{\frac{1}{2}} - y \geq [y^2 + \mu_H]^{\frac{1}{2}} - y. \quad (17)$$

For μ_H fixed, the last term of (17) is a monotone decreasing function of y; therefore since $y \leq 2\|A\|^2$ by Eq. (13), we obtain the left-hand side of Eq. (15). Equation (4) and the Cauchy–Schwarz inequality imply $\omega_1^2 \leq (A\xi, A\xi)/(\xi, \xi)$. Equation (11) then gives $\mu_A |\omega|^2 \leq \|H\|^2$, which is the right-hand side of Eq. (15).

From Eqs. (11) and (6) we obtain

$$\omega_2^2 = [(\xi, H^2\xi)/(\xi, \xi)]\,|\omega|^{-2} + 3\omega_1^2 - 4[(A\xi, A\xi)/(\xi, \xi)] \geq -\lambda_0^{-1}\mu_H - 4\|A\|^2. \quad (18)$$

Equation (12) leads to

$$\omega_2^2 \leq [y^2 + \|H\|^2]^{\frac{1}{2}} - y. \quad (19)$$

The right-hand side is a monotone decreasing function of y, so that, since Eq. (6) and (13) imply $x \leq y$, we obtain

$$\omega_2^2 \leq [x^2 + \|H\|^2]^{\frac{1}{2}} - x. \quad (20)$$

This completes the proof of Theorem I.

Corollary: Suppose that

$$\delta \equiv \inf_E \frac{(\zeta, -H\zeta)}{(\zeta, \zeta)} > 0,$$

i.e., H is negative definite. Then if $\delta\mu_A > \|H\|^2$, all the eigenvalues ω are real.

Proof: Let $\delta\mu_A > \|H\|^2$, and suppose that there exists an eigenvalue $\omega = \omega_1 + i\omega_2$ with $\omega_2 \neq 0$.

Equation (5) implies $|\omega|^2 \geq \delta$, and Eq. (15) then yields $\|H\|^2 \geq \delta\mu_A$, which is a contradiction.

Equation (16) can be used to infer the existence of a complex eigenvalue $\omega(\omega_2 \neq 0)$ if $\lambda_0 < 0$, λ_0 is an eigenvalue of H, and if ω depends continuously on the operator A, for fixed H. Indeed, if we replace A by ϵA in Eq. (2), and suppose that $\omega(\epsilon)$ is a continuous curve in the complex ω plane for $0 \leq \epsilon \leq 1$ (this certainly is the case if E is a finite dimensional Euclidean space), then it follows at once from Eq. (16) that if $\|A\|^2 < -\mu_H/4\lambda_0$, we have $\omega_2 \neq 0$, $0 \leq \epsilon \leq 1$. Under these circumstances, a somewhat sharper condition can be given. Suppose that $\omega(\epsilon) = \omega_1(\epsilon) + i\omega_2(\epsilon)$ became real for some ϵ in [0, 1], and let ϵ_0 be the smallest such ϵ. Then $\omega_2^2(0) = -\lambda_0$, $\epsilon_0 > 0$, $\omega(\epsilon_0) = \omega_1(\epsilon_0)$, $\omega_2(\epsilon_0) = 0$, and $\omega_2(\epsilon) \neq 0$ for $0 \leq \epsilon < \epsilon_0$. Due to the assumed continuity of $\omega(\epsilon)$, Eq. (11) must hold with A replaced by $\tilde{A} \equiv \epsilon_0 A$, $\omega_1 = \omega_1(\epsilon_0)$, and $\omega_2 = 0$, so that

$$\omega_1^2 = \frac{2}{3}\frac{(\tilde{A}\xi, \tilde{A}\xi)}{(\xi, \xi)} \pm \left\{\left[\frac{2(\tilde{A}\xi, \tilde{A}\xi)}{3(\xi, \xi)}\right]^2 - \frac{(\xi, H^2\xi)}{3(\xi, \xi)}\right\}^{\frac{1}{2}}. \tag{21}$$

A real solution for ω_1^2 is possible only if

$$[\tfrac{2}{3}\|\tilde{A}\|^2]^2 \geq \tfrac{1}{3}\mu_H.$$

Hence

$$\|A\|^4 < \tfrac{3}{4}\mu_H \tag{22}$$

implies $\omega_2 \neq 0$.

III. SUFFICIENT CONDITIONS FOR INSTABILITY FOR FINITE DIMENSIONAL SYSTEMS

We restrict our attention in this section to the case where E is an n-dimensional unitary space. We begin with the following result:

Theorem II: Let P, iA, and H be $n \times n$ Hermitian matrices. Then the coefficients of the polynomial

$$P_{2n}(x) \equiv \det(x^2P - 2xiA - H) \tag{23}$$

are all real and the eigenvalues ω of the system

$$(\omega^2P - 2\omega iA - H)\xi = 0 \tag{24}$$

occur in complex conjugate pairs.

Proof: ω is an eigenvalue of Eq. (24) if and only if ω is a root of the $2n$th-degree polynomial $P_{2n}(x)$ defined by Eq. (23). Using a bar to denote complex conjugation, and a superscript T to denote the transpose, we have

$$\begin{aligned}\overline{P_{2n}(x)} &= \det(\bar{x}^2\bar{P} + 2\bar{x}i\bar{A} - \bar{H}) \\ &= \det(\bar{x}^2P^T + 2\bar{x}iA^T - H^T) \\ &= \det(\bar{x}^2P - 2\bar{x}iA - H) = P_{2n}(\bar{x})\end{aligned} \tag{25}$$

since P, iA, and H are all Hermitian. Thus $P_{2n}(\omega) = 0$ implies $P_{2n}(\bar{\omega}) = 0$.

Theorem III: Let iA and H be $n \times n$ Hermitian matrices, and suppose that $\lambda_0 < 0$ and $\|A\|^4 < \frac{3}{4}\mu_H$. Then the system described by Eq. (2) is unstable.

Proof: Theorem II implies that the eigenvalues ω of Eq. (2) occur in complex conjugate pairs, and must be roots of the polynomial

$$P_{2n}(x) \equiv \det(x^2I - 2xiA - H).$$

But the roots of $P_{2n}(x)$ are continuous functions of the elements of A; the theorem then follows at once from the last paragraph of Sec. I.

It proves to be convenient for the following theorem to assume that the matrix H is diagonal. This may be done without loss of generality; indeed, referring to Eq. (1), it is well known that for H_1 and H_2 Hermitian and H_1 positive definite, there exists a nonsingular transformation S such that $S^*H_1S = I$ and $S^*H_2S = H$, where H is diagonal (S^* is the adjoint of S). Furthermore, S is real if H_1 and H_2 are real. Equation (1) can then be cast into the form of Eq. (2), where $A = S^*A_1S$, $H = S^*H_2S$, and $\xi = S^{-1}\eta$. Clearly iA_1 Hermitian implies that iA is also Hermitian.

Theorem IV: Let A be a real antisymmetric $n \times n$ matrix and H be diagonal with (real) eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. Suppose that $\prod_{k=1}^{n} \lambda_k < 0$ (i.e., H has an odd number of negative eigenvalues, and no eigenvalue is zero). Then the system described by Eq. (2) is unstable, possessing (at least) the complex eigenvalues $\pm i\omega_2$ where ω_2 satisfies the inequality

$$[x^2 + \|H\|^2]^{\frac{1}{2}} - x \geq \omega_2^2 \geq [4\|A\|^4 + \mu_H]^{\frac{1}{2}} - 2\|A\|_0^2 \tag{26}$$

where all quantities are as defined in Theorem I.

Proof: The polynomial $P_{2n}(\omega)$ defined by Eq. (23) takes the following form, for $P = I$ and H diagonal:

$$P_{2n}(\omega) = \begin{bmatrix} \omega^2 - \lambda_1 & -2i\omega a_{12} & -2i\omega a_{13} & \cdots & -2i\omega a_{1n} \\ 2i\omega a_{12} & \omega^2 - \lambda_2 & -2i\omega a_{23} & \cdots & -2i\omega a_{2n} \\ 2i\omega a_{13} & 2i\omega a_{23} & \omega^2 - \lambda_3 & \cdots & -2i\omega a_{3n} \\ \cdot & & & & \\ \cdot & & & & \\ 2i\omega a_{1n} & 2i\omega a_{2n} & \cdots & \cdots & \omega^2 - \lambda_n \end{bmatrix}, \tag{27}$$

where the a_{ij} are the elements of the real antisymmetric matrix A, $a_{ij} = -a_{ji}$, and λ_k ($k = 1, 2, \cdots, n$) are the eigenvalues of H. Clearly $P_{2n}(\omega) = P_{2n}(-\omega)$, since the value of the determinant is unchanged if rows and columns are interchanged, so that $P_{2n}(\omega)$ is a

polynomial of degree n in ω^2. Therefore

$$P_{2n}(\omega) = \prod_{k=1}^{n} (\omega^2 - \Omega_k),$$

and since

$$P_{2n}(0) = \prod_{k=1}^{n} (-\lambda_k),$$

we have

$$\prod_{k=1}^{n} \Omega_k = \prod_{k=1}^{n} \lambda_k < 0. \tag{28}$$

Let C denote the set of all integers k such that $\operatorname{Im} \Omega_k \neq 0$ $(1 \leq k \leq n)$. From Eq. (25), $\overline{P_{2n}(\omega)} = P_{2n}(\bar{\omega})$, so that C can be decomposed into disjoint sets of pairs of integers $\{p, q\}$ such that $\Omega_p = \bar{\Omega}_q$ and $C = \bigcup \{p, q\}$. Therefore $\prod_{k \text{ in } C} \Omega_k > 0$, if C is nonempty. Equation (28) then implies the existence of a real $\Omega_k < 0$. Therefore $P_{2n}(\omega)$ has roots $\pm i(-\Omega_k)^{\frac{1}{2}}$. Equation (26) follows immediately from Eqs. (15) and (16).

This result shows that if we confine our attention to real operators, there exist systems unstable for $A = 0$ which cannot be stabilized by the introduction of a nonzero A, no matter how large. However, by Theorem I, the growth rate may be made arbitrarily small by taking μ_A sufficiently large. Note that there is no conflict between Theorem IV and the Corollary to Theorem I, as $\mu_A = 0$ for a real antisymmetric $n \times n$ matrix for odd n.

IV. AN EXISTENCE THEOREM

We now turn our attention to the case where iA and H are completely continuous Hermitian operators defined on a Hilbert space E, and present an existence theorem for stable solutions to Eq. (2), valid whenever H admits of one or more positive eigenvalues.

We have seen in Sec. II that Eq. (2) can possess an unstable mode ($\operatorname{Im} \omega < 0$) only if the system described by Eq. (2) with $A \equiv 0$ is itself unstable; furthermore, the maximum growth rate of unstable modes of Eq. (2) for $A \neq 0$ never exceeds that for the system with $A \equiv 0$, and can be made arbitrarily small (or zero) provided A is sufficiently "large" [i.e.,

$$\inf_{E} \frac{(A\zeta, A\zeta)}{(\zeta, \zeta)}$$

is sufficiently large]. The operator A can therefore be thought of as a stabilizing influence. We now demonstrate that A also preserves stable modes with real eigenfrequencies. [For E finite dimensional, we can state that Eq. (2) with $A \neq 0$ has at least the same number of stable modes as does Eq. (2) with $A \equiv 0$.] For every positive eigenvalue λ of H, the system $(\omega^2 I - H)\zeta = 0$ admits of two modes, with eigenfrequencies $\pm\lambda^{\frac{1}{2}}$; for each such positive λ (counted as often as its degeneracy) we can guarantee the existence of two stable modes of Eq. (2) for $A \neq 0$—one with $\omega > 0$, the other with $\omega < 0$.

Theorem V: Let H and iA be completely continuous Hermitian operators on the Hilbert space E. Let H have N positive eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N > 0$$

with associated eigenvectors η_j $(j = 1, 2, \cdots, N)$, $(\eta_j, \eta_k) = \delta_{jk}$. Then Eq. (2) admits of N real positive eigenvalues ω_j^+ $(j = 1, 2, \cdots, N)$ (with corresponding eigenvectors ξ_j^+) and N real negative eigenvalues ω_j^- (with corresponding eigenvectors ξ_j^-).[12] If $\omega_j^{\pm} = \omega_k^{\pm}$ for $j \neq k$, then $(\xi_j^{\pm}, \xi_k^{\pm}) = 0$.

Proof: Since iA and H are completely continuous Hermitian operators, $K_\omega \equiv H + 2\omega iA$ is a completely continuous Hermitian operator on E for each ω on the real line. For each such ω, $\omega \neq 0$, we define

$$F_1(\omega) \equiv \frac{1}{\omega^2} \sup_{E} \frac{(\zeta, K_\omega \zeta)}{(\zeta, \zeta)}. \tag{29}$$

Then we have

$$|F_1(\omega)| \leq \frac{1}{\omega^2} \|K_\omega\| \leq \frac{\|H\|}{\omega^2} + 2\frac{\|A\|}{|\omega|} \tag{30}$$

so that $F_1(\omega) \to 0$ as $\omega \to \pm\infty$.

Now $H\eta_1 = \lambda_1\eta_1$, $\lambda_1 > 0$, so that

$$(\eta_1, K_\omega\eta_1)/(\eta_1, \eta_1) = \lambda_1 + 2\omega(\eta_1, iA\eta_1)/(\eta_1, \eta_1) \geq \lambda_1 - |\omega|\, 2\, \|A\|. \tag{31}$$

Thus

$$F_1(\omega) \geq (1/\omega^2)(\tfrac{1}{2}\lambda_1) \quad \text{for} \quad |\omega| < \lambda_1/(4\, \|A\|). \tag{32}$$

We now show that $F_1(\omega)$ is continuous on $(-\infty, 0)$ and $(0, \infty)$. For future reference we define

$$G_n(\omega) \equiv \inf_{\substack{\phi_j \in E \\ j=1,2,\cdots,n-1}} \left\{ \sup_{\substack{(\zeta,\phi_j)=0 \\ j=1,2,\cdots,n-1}} \frac{(\zeta, K_\omega\zeta)}{(\zeta, \zeta)} \right\}, \qquad -\infty < \omega < \infty \quad n = 1, 2, 3, \cdots \tag{33}$$

and we have $|G_n(\omega)| \leq \|K_\omega\|$. In particular, $F_1(\omega) = \omega^{-2}G_1(\omega)$. Since $K_{\omega+\Delta} = K_\omega + 2\Delta iA$, Eq. (33) yields

$$G_n(\omega \pm \Delta) = \inf_{\substack{\phi_j \in E \\ j=1,2,\cdots,n-1}} \left\{ \sup_{\substack{(\zeta,\phi_j)=0 \\ j=1,2,\cdots,n-1}} \left[\frac{(\zeta, K_\omega\zeta)}{(\zeta, \zeta)} \pm 2\Delta \frac{(\zeta, iA\zeta)}{(\zeta, \zeta)} \right] \right\}$$

$$\leq \inf_{\substack{\phi_j \in E \\ j=1,2,\cdots,n-1}} \left\{ \sup_{\substack{(\zeta,\phi_j)=0 \\ j=1,2,\cdots,n-1}} \frac{(\zeta, K_\omega\zeta)}{(\zeta, \zeta)} \right\} + 2\,|\Delta|\, \|A\|$$

so that

$$G_n(\omega \pm \Delta) \leq G_n(\omega) + 2\,|\Delta|\, \|A\|. \tag{34}$$

[12] Routh has proved a similar theorem (using determinants) for E finite dimensional. See Ref. 4.

Replacing ω by $\omega \mp \Delta$ in Eq. (34), there results

$$G_n(\omega) \leq G_n(\omega \mp \Delta) + 2|\Delta| \|A\|. \tag{35}$$

Equations (34) and (35) imply

$$|G_n(\omega + \Delta) - G_n(\omega)| \leq 2|\Delta| \|A\| \tag{36}$$

so that $G_n(\omega)$ is a continuous function of ω on $(-\infty, \infty)$ for all n. Therefore $F_1(\omega)$ is continuous on $(-\infty, 0)$ and $(0, \infty)$, so that by Eqs. (30) and (32), there exist $a_1 > 0$ and $b_1 < 0$ such that $F_1(a_1) = F_1(b_1) = \frac{1}{2}$, and $F_1(\omega) \geq \frac{1}{2}$ for

$$\omega \in S_1 \equiv (0, a_1) \cup (b_1, 0).$$

Since K_ω is completely continuous and Hermitian, there exists $\xi_1(\omega) \in E$ such that $\|\xi_1(\omega)\| = 1$ and

$$\omega^2 F_1(\omega)\xi_1(\omega) = K_\omega \xi_1(\omega) \tag{37}$$

for all $\omega \in S_1$. By Eq. (32), there exists $\omega_1^+ \in (0, a_1)$ and $\omega_1^- \in (b_1, 0)$ such that $F_1(\omega_1^\pm) = 1$. Hence

$$[(\omega_1^\pm)^2 I - 2\omega_1^\pm iA - H]\xi_1^\pm = 0, \tag{38}$$

where $\xi_1^\pm \equiv \xi_1(\omega_1^\pm)$.

Let

$$F_2(\omega) \equiv \frac{1}{\omega^2} \sup_{(\zeta, \xi_1)=0} \frac{(\zeta, K_\omega \zeta)}{(\zeta, \zeta)}, \quad \omega \in S_1. \tag{39}$$

Clearly

$$F_2(\omega) \leq F_1(\omega), \quad \omega \in S_1 \tag{40}$$

and in particular,

$$F_2(\omega_1^\pm) \leq F_1(\omega_1^\pm) = 1. \tag{41}$$

Now $H\eta_j = \lambda_j \eta_j$, $j = 1, 2$; $\lambda_1 \geq \lambda_2 > 0$; $(\eta_j, \eta_k) = \delta_{jk}$. Then for any vector x, $\|x\| > 0$, in the linear manifold M_2 spanned by η_1 and η_2, we have

$$(x, Hx)/(x, x) \geq \lambda_2 > 0.$$

Given $\xi_1(\omega)$ $\omega \in S_1$, there exists $x_1(\omega) \in M_2$ such that $\|x_1\| = 1$ and $(x_1, \xi_1) = 0$. Therefore

$$(x_1, K_\omega x_1) \geq \lambda_2 - 2|\omega| \|A\| \geq \lambda_2/2 > 0 \tag{42}$$

for $|\omega| \leq \lambda_2/4 \|A\|$, so that

$$F_2(\omega) \geq \lambda_2/2\omega^2 \quad \text{for} \quad |\omega| \leq \frac{\lambda_2}{4\|A\|}. \tag{43}$$

We shortly show that $\omega^2 F_2(\omega) = G_2(\omega)$ on S_1, so that $F_2(\omega)$ is continuous on S_1. Thus there exist real numbers a_2 and b_2 such that $0 < a_2 \leq a_1$, $0 > b_2 \geq b_1$, $F_2(a_2) = F_2(b_2) = \frac{1}{2}$, $F_2(\omega) \geq \frac{1}{2}$ for $\omega \in S_2 \equiv (b_2, 0) \cup (0, a_2)$. Note that $S_2 \subset S_1$. For every $\omega \in S_2$, since K_ω is a completely continuous Hermitian operator, there exists $\xi_2(\omega) \in E$, $\|\xi_2\| = 1$, satisfying

$$\omega^2 F_2(\omega)\xi_2 = K_\omega \xi_2, \quad (\xi_2, \xi_1) = 0. \tag{44}$$

By the continuity of $F_2(\omega)$ on S_2, and Eqs. (40) and (43), there exists $\omega_2^+ \in (0, a_2)$ and $\omega_2^- \in (b_2, 0)$ such that $\omega_1^+ \geq \omega_2^+ > 0$, $\omega_1^- \leq \omega_2^- < 0$, and $F_2(\omega_2^\pm) = 1$, so that

$$[(\omega_2^\pm)^2 I - 2\omega_2^\pm iA - H]\xi_2^\pm = 0, \tag{45}$$

where $\xi_2^\pm \equiv \xi_2(\omega_2^\pm)$. Note that by construction, $\omega_2^\pm = \omega_1^\pm$ implies that $(\xi_2^\pm, \xi_1^\pm) = 0$.

Suppose now that we have constructed $2p$ $(p < N)$ eigenvalues $\omega_1^+ \geq \omega_2^+ \geq \cdots \geq \omega_p^+ > 0$, $\omega_1^- \leq \omega_2^- \leq \cdots \leq \omega_p^- < 0$ and eigenvectors $\xi_k^\pm$, $k = 1, 2, \cdots, p$, so that we have for $n = 1, 2, \cdots, p$:

$$F_n(\omega) = \frac{1}{\omega^2} \sup_{\substack{(\zeta, \xi_k)=0 \\ k=1,2,\cdots,n-1}} \frac{(\zeta, K_\omega \zeta)}{(\zeta, \zeta)}, \quad \omega \in S_{n-1}, \tag{46}$$

$$\omega^2 F_n(\omega)\xi_n(\omega) = K_\omega \xi_n, \quad \omega \in S_n, \tag{47}$$

$S_n \equiv (0, a_n) \cup (b_n, 0)$; $0 < a_p \leq a_{p-1} \leq \cdots \leq a_1 < a_0 = \infty$; $-\infty = b_0 < b_1 \leq \cdots \leq b_{p-1} \leq b_p < 0$; $F_n(a_n) = F_n(b_n) = \frac{1}{2}$; $F_n(\omega) \geq \frac{1}{2}$ for $\omega \in S_n$; $F_n(\omega) \leq F_{n-1}(\omega)$ on S_{n-1};

$$\lim_{\omega \to 0} F_n(\omega) = \infty;$$

$F_n(\omega)$ continuous on S_{n-1}; $(\xi_l, \xi_m) = \delta_{lm}$; $F_n(\omega_n^\pm) = 1$; and $\xi_n^\pm = \xi_n(\omega_n^\pm)$. Then we define

$$F_{p+1}(\omega) \equiv \frac{1}{\omega^2} \sup_{\substack{(\zeta, \xi_k)=0 \\ k=1,2,\cdots,p}} \frac{(\zeta, K_\omega \zeta)}{(\zeta, \zeta)}, \quad \omega \in S_p. \tag{48}$$

Therefore

$$F_{p+1}(\omega) \leq F_p(\omega), \quad \omega \in S_p. \tag{49}$$

Let M_{p+1} be the linear manifold spanned by $\{\eta_1, \eta_2, \cdots, \eta_{p+1}\}$. Then for any $x \in M_{p+1}$, $\|x\| > 0$, we have $(x, Hx)/(x, x) \geq \lambda_{p+1} > 0$. Given the p orthonormal vectors $\xi_1(\omega)$, $\xi_2(\omega)$, $\cdots$, $\xi_p(\omega)$, there exists $x_p(\omega) \in M_{p+1}$ such that $\|x_p\| = 1$ and $(x_p, \xi_k) = 0$, $k = 1, 2, \cdots, p$. Therefore

$$F_{p+1}(\omega) \geq (1/\omega^2)\{(x_p, Hx_p) - 2|\omega| \|A\|\}$$
$$\geq \lambda_{p+1}/2\omega^2 > 0 \tag{50}$$

for $|\omega| \leq \lambda_{p+1}/4\|A\|$. Assume for the moment that $\omega^2 F_{p+1}(\omega) = G_{p+1}(\omega)$ on S_p, so that $F_{p+1}(\omega)$ is continuous on S_p. Then there exist real numbers a_{p+1} and b_{p+1} such that $0 < a_{p+1} \leq a_p$, $b_p \leq b_{p+1} < 0$, $F_{p+1}(a_{p+1}) = F_{p+1}(b_{p+1}) = \frac{1}{2}$, and $F_{p+1}(\omega) \geq \frac{1}{2}$ for $\omega \in S_{p+1} \equiv (b_{p+1}, 0) \cup (0, a_{p+1})$. For each $\omega \in S_{p+1}$, K_ω is a completely continuous Hermitian operator, so that there exists $\xi_{p+1}(\omega) \in E$, $\|\xi_{p+1}\| = 1$, satisfying

$$\omega^2 F_{p+1}(\omega)\xi_{p+1} = K_\omega \xi_{p+1}, \ (\xi_{p+1}, \xi_m) = 0,$$
$$m = 1, 2, \cdots, p. \tag{51}$$

Continuity of $F_{p+1}(\omega)$ on S_p and Eqs. (49) and (50) imply that there exist $\omega_{p+1}^+ \in (0, a_{p+1})$ and $\omega_{p+1}^- \in (b_{p+1}, 0)$ such that $\omega_p^+ \geq \omega_{p+1}^+ > 0$, $\omega_p^- \leq \omega_{p+1}^- < 0$, and $F_{p+1}(\omega_{p+1}^\pm) = 1$, so that

$$[(\omega_{p+1}^\pm)^2 I - 2\omega_{p+1}^\pm iA - H]\xi_{p+1}^\pm = 0, \tag{52}$$

where $\xi_{p+1}^\pm \equiv \xi_{p+1}(\omega_{p+1}^\pm)$.

It remains to show that $\omega^2 F_n(\omega) = G_n(\omega)$ on S_{n-1}. Clearly Eqs. (33) and (46) imply that $G_n(\omega) \leq \omega^2 F_n(\omega)$ on S_{n-1}, so that it suffices to show that $G_n(\omega) \geq \omega^2 F_n(\omega)$ for $\omega \in S_{n-1}$. Now for any $n-1$ vectors $\phi_1, \phi_2, \cdots, \phi_{n-1}$ in E,

$$\sup_{\substack{(\zeta,\phi_j)=0 \\ j=1,2,\cdots,n-1}} \frac{(\zeta, K_\omega \zeta)}{(\zeta,\zeta)} \geq \sup_{\substack{(\zeta,\xi_j)=0 \\ j=1,2,\cdots,n-1}} \frac{(\zeta, K_\omega \zeta)}{(\zeta,\zeta)} = \omega^2 F_n(\omega), \quad \omega \in S_{n-1}. \tag{53}$$

Indeed, let $\zeta \in E$ such that $\|\zeta\| = 1$ and $(\zeta, \xi_k) = 0$, $k = 1, 2, \cdots, n-1$. By construction, the ξ_k ($k = 1, 2, \cdots, n-1$) are orthonormal eigenvectors of K_ω (for each fixed ω) and $(\zeta, K_\omega \zeta) \leq (\xi_k, K_\omega \xi_k)$, $k = 1, 2, \cdots, n-1$, since $(\zeta, \xi_k) = 0$, $k = 1, 2, \cdots, n-1$. The linear manifold spanned by the n orthonormal vectors $\xi_0 \equiv \zeta, \xi_1, \cdots, \xi_{n-1}$ is n dimensional, so that there exist constants α_k ($k = 0, 1, \cdots, n-1$) such that $y \equiv \sum_{k=0}^{n-1} \alpha_k \xi_k$ satisfies $\|y\|^2 = \sum_{k=0}^{n-1} |\alpha_k|^2 = 1$ and $(y, \phi_k) = 0$, $k = 1, 2, \cdots, n-1$. Then

$$(y, K_\omega y) = \sum_{k,l=0}^{n-1} \bar{\alpha}_k \alpha_l (\xi_k, K_\omega \xi_l) = \sum_{k=0}^{n-1} (\xi_k, K_\omega \xi_k) |\alpha_k|^2$$
$$\geq (\zeta, K_\omega \zeta) \sum_{k=0}^{n-1} |\alpha_k|^2 = (\zeta, K_\omega \zeta).$$

Hence Eq. (53) holds, and $G_n(\omega) \geq \omega^2 F_n(\omega)$, $\omega \in S_{n-1}$, follows at once. This completes the proof of the theorem.

V. ORTHOGONALITY AND COMPLETENESS RELATIONS FOR STABLE FINITE-DIMENSIONAL SYSTEMS

We now consider the orthogonality and completeness properties of the eigenvectors of Eq. (2), restricting our attention primarily to the case where H is positive definite and E is a finite-dimensional unitary space.

The time-dependent counterpart of Eq. (2) is

$$\ddot{\eta} + 2A\dot{\eta} + H\eta = 0. \tag{54}$$

This second-order differential equation allows the initial displacement η_0 and velocity $\dot{\eta}_0$ to be arbitrarily prescribed vectors in E. Thus, if the set of eigenvectors $\{\xi_n\}$ of Eq. (2) is sufficiently large, the most general solution of Eq. (54) would be given by (assuming that $A\eta = 0$ and $H\eta = 0$ implies $\eta = 0$)

$$\eta(t) = \sum_n \alpha_n \xi_n e^{i\omega_n t} \tag{55}$$

with

$$\eta_0 = \sum_n \alpha_n \xi_n, \tag{56}$$

$$-i\dot{\eta}_0 = \sum_n \alpha_n \omega_n \xi_n, \tag{57}$$

where the α_n are constants, determined by η_0 and $\dot{\eta}_0$. Equations (56) and (57) lead us to the following definition: We say that a set of eigenvectors $\{\xi_n\}$ of Eq. (2) is complete if for any two vectors η_0 and $-i\dot{\eta}_0$ in E, there exist constants α_n such that Eqs. (56) and (57) hold. In the circumstance that H is positive definite and E is an N-dimensional unitary space E_N, we show that

(1) there exists a complete set of $2N$ eigenvectors $\{\xi_n\}_{n=1}^{2N}$;

(2) the coefficients α_n are given by

$$\alpha_n = \frac{\omega_n(\xi_n, -i\dot{\eta}_0) + (\xi_n, H\eta_0)}{\omega_n^2(\xi_n, \xi_n) + (\xi_n, H\xi_n)}; \tag{58}$$

(3) $(\dot{\eta}_0, \dot{\eta}_0) + (\eta_0, H\eta_0)$

$$= \sum_{n=1}^{2N} |\alpha_n|^2 \{\omega_n^2(\xi_n, \xi_n) + (\xi_n, H\xi_n)\}; \tag{59}$$

(4) $$\{\xi_n\}_{n=1}^{2N} = \{\xi_n^+\}_{n=1}^{N} \cup \{\xi_n^-\}_{n=1}^{N},$$

where ξ_n^+ is an eigenvector with positive eigenvalue ω_n^+, ξ_n^- is an eigenvector with negative eigenvalue ω_n^-, and each of the sets $\{\xi_n^\pm\}_{n=1}^N$ spans E_N;

(5) for each of the sets $\{\xi_n^\pm\}_{n=1}^N$, there exists a linear operator $Q^\pm$ and an inner product $(\ ,\ )^\pm$ defined on E_N with the following properties: Q^+ is a positive-definite Hermitian operator with respect to $(\ ,\ )^+$; Q^- is a negative definite Hermitian operator with respect to $(\ ,\ )^-$; $Q^+\xi_n^+ = \omega_n^+\xi_n^+$, $Q^-\xi_n^- = \omega_n^-\xi_n^-$ ($n = 1, 2, \cdots, N$); and $(\xi_l^+, \xi_m^+)^+ = 0 = (\xi_l^-, \xi_m^-)^-$ for $m \neq l$.

In establishing the results to follow, we assume throughout that H and iA are linear Hermitian operators defined on a linear vector space E with a complex inner product (f, g) defined for all f and g in E.

Lemma I: Let ω_k and ω_l be eigenvalues of Eq. (2) with corresponding eigenvectors ξ_k and ξ_l. Then $\bar{\omega}_k \neq \omega_l$ implies

$$\bar{\omega}_k \omega_l (\xi_k, \xi_l) + (\xi_k, H\xi_l) = 0, \tag{60}$$

$$(\bar{\omega}_k + \omega_l)(\xi_k, \xi_l) - 2(\xi_k, iA\xi_l) = 0. \tag{61}$$

Proof: We have

$$\omega_l^2(\xi_k, \xi_l) - 2\omega_l(\xi_k, iA\xi_l) - (\xi_k, H\xi_l) = 0, \tag{62}$$

$$\omega_k^2(\xi_l, \xi_k) - 2\omega_k(\xi_l, iA\xi_k) - (\xi_l, H\xi_k) = 0. \tag{63}$$

Subtracting the complex conjugate of Eq. (63) from Eq. (62), and using the fact that H and iA are Hermitian, we obtain

$$(\omega_l - \bar{\omega}_k)[(\bar{\omega}_k + \omega_l)(\xi_k, \xi_l) - 2(\xi_k, iA\xi_l)] = 0. \tag{64}$$

$\bar{\omega}_k \neq \omega_l$ therefore implies Eq. (61). Equation (60) now follows by substituting $(\xi_k, iA\xi_l) = \frac{1}{2}(\bar{\omega}_k + \omega_l) \times (\xi_k, \xi_l)$ into the complex conjugate of Eq. (63).

When $H \geq 0$, the eigenvalues of Eq. (2) are all real, so that Eqs. (60) and (61) hold (with $\bar{\omega}_k = \omega_k$) whenever $\omega_k \neq \omega_l$. If we are given any set of eigenvectors $\{\xi_k\}$, Eq. (60) holds for $k \neq l$ provided $\omega_k \neq \omega_l$. If the $\{\xi_k\}$ contains a subset of p linearly independent eigenvectors all having eigenvalue ω_k, one can always take linear combinations of these so as to form a new set of p linearly independent eigenvectors (with eigenvalue ω_k) which themselves satisfy Eq. (60) for $k \neq l$ (see Lemma II). Thus for $H \geq 0$, given any set of eigenvectors $\{\xi_k'\}$, we can construct from them (if necessary) a new set of eigenvectors $\{\xi_k\}$ such that Eq. (60) holds for $k \neq l$ and which possesses for each ω_k the same number of linearly independent eigenvectors.

Lemma II: Let E be a Hilbert space and $M(\omega)$ be the linear subspace spanned by p linearly independent eigenvectors of Eq. (2) having the eigenvalue ω. Then there exists p orthonormal vectors ξ_k $(k = 1, 2, \cdots, p)$ in $M(\omega)$ such that

$$|\omega|^2(\xi_l, \xi_m) + (\xi_l, H\xi_m) = 0, \quad l \neq m. \tag{65}$$

Each ξ_k is an eigenvector of Eq. (2) with eigenvalue ω.

Proof: By hypothesis, $M(\omega)$ is a p-dimensional unitary space, and every nonzero vector in $M(\omega)$ is an eigenvector of Eq. (2) with eigenvalue ω. Let P be the projection onto $M(\omega)$, i.e., $P\xi$ is the projection of ξ onto $M(\omega)$ for all ξ in E. Then $\tilde{H} \equiv PHP$ is Hermitian (since P is Hermitian) and maps $M(\omega)$ into itself. Therefore there exist p vectors ξ_k $(k = 1, 2, \cdots, p)$ in $M(\omega)$ satisfying

$$\tilde{H}\xi_k = \lambda_k \xi_k, \quad k = 1, 2, \cdots, p, \tag{66}$$

$$(\xi_l, \xi_m) = \delta_{lm}, \quad l, m = 1, 2, \cdots, p, \tag{67}$$

where λ_k is a real constant and δ_{lm} is the Kronecker delta. Therefore $l \neq m$ implies

$$\begin{aligned} |\omega|^2(\xi_l, \xi_m) + (\xi_l, H\xi_m) &= (P\xi_l, HP\xi_m) = (\xi_l, \tilde{H}\xi_m) \\ &= \lambda_m(\xi_l, \xi_m) = 0. \end{aligned} \tag{68}$$

Suppose that H is positive definite on E, so that all the eigenvalues ω_k of Eq. (2) are real and nonzero. We then make the following definitions.

Definition 1: A set of eigenvectors $\{\xi_k\}_{k=1}^n$ of Eq. (2) is said to be canonical if, for all $k, l = 1, 2, \cdots, n$

$$\omega_k\omega_l(\xi_k, \xi_l) + (\xi_k, H\xi_l) = \epsilon_k \delta_{kl}, \tag{69}$$

where

$$\epsilon_k \equiv \omega_k^2(\xi_k, \xi_k) + (\xi_k, H\xi_k) > 0, \quad k = 1, 2, \cdots, n. \tag{70}$$

Definition 2: Let $\{\xi_k\}_{k=1}^n$ be a canonical set of eigenvectors of Eq. (2), and let x and $\dot{x}$ be any two vectors in E. The generalized Fourier coefficient $\alpha_k[\dot{x}, x]$ is defined by

$$\alpha_k[\dot{x}, x] \equiv \{\omega_k(\xi_k, \dot{x}) + (\xi_k, Hx)\}\epsilon_k^{-1}. \tag{71}$$

The next theorem shows that our generalized Fourier coefficients play much the same role in the representation of the two vectors x and $\dot{x}$ by sums of canonical eigenvectors of the form of Eqs. (56) and (57) as do the ordinary Fourier coefficients in the approximation of a vector by a sum of orthonormal vectors.

Theorem VI: Let H be positive definite on E, $\{\xi_k\}_{k=1}^n$ be a canonical set of eigenvectors of Eq. (2), and $\dot{x}$ and x be any two vectors in E. Let $\{\beta_k\}_{k=1}^n$ be any n complex numbers. Then

$$\begin{aligned} &\left\| \dot{x} - \sum_1^n \beta_k\omega_k\xi_k \right\|^2 + \left\| x - \sum_1^n \beta_k\xi_k \right\|_H^2 \\ &\quad = \left\| \dot{x} - \sum_1^n \alpha_k\omega_k\xi_k \right\|^2 + \left\| x - \sum_1^n \alpha_k\xi_k \right\|_H^2 \\ &\qquad + \left\| \sum_1^n (\alpha_k - \beta_k)\omega_k\xi_k \right\|^2 + \left\| \sum_1^n (\alpha_k - \beta_k)\xi_k \right\|_H^2 \\ &\quad \geq \left\| \dot{x} - \sum_1^n \alpha_k\omega_k\xi_k \right\|^2 + \left\| x - \sum_1^n \alpha_k\xi_k \right\|_H^2 \\ &\quad = (\dot{x}, \dot{x}) + (x, Hx) - \sum_1^n |\alpha_k|^2\epsilon_k, \end{aligned} \tag{72}$$

where the α_k are defined by Eq. (71) and $\|x\|_H^2 \equiv (x, Hx)$.

Proof: Let

$$f \equiv \dot{x} - \sum_1^n \alpha_k\omega_k\xi_k, \quad f \equiv x - \sum_1^n \alpha_k\xi_k,$$

$$\dot{g} \equiv \sum_1^n (\alpha_k - \beta_k)\omega_k\xi_k, \quad g \equiv \sum_1^n (\alpha_k - \beta_k)\xi_k,$$

then

$$\dot{x} - \sum_1^n \beta_k\omega_k\xi_k = f + \dot{g}, \quad x - \sum_1^n \beta_k\xi_k = f + g,$$

so that

$$\begin{aligned} &\left\| \dot{x} - \sum_1^n \beta_k\omega_k\xi_k \right\|^2 + \left\| x - \sum_1^n \beta_k\xi_k \right\|_H^2 \\ &\qquad = (f, f) + (f, Hf) + (\dot{g}, \dot{g}) \\ &\qquad\quad + (g, Hg) + 2\,\mathrm{Re}\,\{(f, \dot{g}) + (f, Hg)\}. \end{aligned}$$

But

$$\begin{aligned} &(f, \dot{g}) + (f, Hg) \\ &\quad = \left(\dot{x} - \sum_1^n \alpha_k\omega_k\xi_k, \sum_1^n (\alpha_l - \beta_l)\omega_l\xi_l\right) \\ &\qquad + \left(x - \sum_1^n \alpha_k\xi_k, H\left[\sum_1^n (\alpha_l - \beta_l)\xi_l\right]\right) \\ &\quad = \sum_1^n (\alpha_l - \beta_l)\{\omega_l(\dot{x}, \xi_l) + (x, H\xi_l)\} \\ &\qquad - \sum_{k,l=1}^n \bar{\alpha}_k(\alpha_l - \beta_l)\{\omega_k\omega_l(\xi_k, \xi_l) + (\xi_k, H\xi_l)\} \\ &\quad = \sum_1^n (\alpha_l - \beta_l)\bar{\alpha}_l\epsilon_l - \sum_1^n \bar{\alpha}_l(\alpha_l - \beta_l)\epsilon_l = 0, \end{aligned}$$

where we have used Eqs. (69)–(71). Now

$$\left\| \dot{x} - \sum_1^n \alpha_k \omega_k \xi_k \right\|^2 + \left\| x - \sum_1^n \alpha_k \xi_k \right\|_H^2$$
$$= (\dot{x}, \dot{x}) + (x, Hx) - \sum_1^n \alpha_k \{\omega_k (\dot{x}, \xi_k) + (x, H\xi_k)\}$$
$$- \sum_1^n \bar{\alpha}_k \{\omega_k (\xi_k, \dot{x}) + (\xi_k, Hx)\}$$
$$+ \sum_{k,l=1}^n \bar{\alpha}_k \alpha_l \{\omega_k \omega_l (\xi_k, \xi_l) + (\xi_k, H\xi_l)\}$$
$$= (\dot{x}, \dot{x}) + (x, Hx) - 2 \sum_1^n |\alpha_k|^2 \epsilon_k + \sum_1^n |\alpha_k|^2 \epsilon_k$$
$$= (\dot{x}, \dot{x}) + (x, Hx) - \sum_1^n |\alpha_k|^2 \epsilon_k .$$

This completes the proof of Theorem VI.

Note that the last line of Eq. (72) gives a Bessel inequality for the α_k, viz.,

$$(\dot{x}, \dot{x}) + (x, Hx) \geq \sum_k |\alpha_k|^2 \epsilon_k . \tag{73}$$

Furthermore, we see that a denumerable canonical set of eigenvectors $\{\xi_k\}$ is complete if and only if the equality sign holds in Eq. (73) for arbitrary x and $\dot{x}$ in the Hilbert space E, for H bounded and positive definite, with a bounded inverse H^{-1}.

Lemma III: Let ω_l $(l = 1, 2, \cdots, n)$ be n eigenvalues of Eq. (2) with corresponding eigenvectors ξ_l $(l = 1, 2, \cdots, n)$, where the n vectors ξ_l are linearly independent. Suppose that Ω is an eigenvalue of Eq. (2) with eigenvector ζ, where $\zeta = \sum_{l=1}^n a_l \xi_l \neq 0$. Then Ω and ω_l $(l = 1, 2, \cdots, n)$ are all roots of the real polynomial $P_{2n}(x)$ of degree $2n$ defined by

$$P_{2n}(x) \equiv \det F(x), \tag{74}$$

where $F(x)$ is the $n \times n$ matrix given by

$$F(x) \equiv x^2 G - 2xiA' - H' \tag{75}$$

with $G_{kl} \equiv (\xi_k, \xi_l)$, $A'_{kl} \equiv (\xi_k, A\xi_l)$, and $H'_{kl} \equiv (\xi_k, H\xi_l)$. The matrices G, iA', and H' are all Hermitian, and G is positive definite. Furthermore, if H is positive definite, then $P_{2n}(x)$ has precisely n positive and n negative roots.

Proof: The matrices G, iA', and H' are clearly Hermitian since iA and H are Hermitian. For any column vector β with components b_i $(i = 1, 2, \cdots, n)$ we have $(\eta, \eta) = \langle \beta, G\beta \rangle$, where $\eta \equiv \sum_{k=1}^n b_k \xi_k$ and $\langle \alpha, \beta \rangle \equiv \sum_{k=1}^n \bar{a}_k b_k$ (the usual inner product in E_n). The ξ_k $(k = 1, 2, \cdots, n)$ are linearly independent, so that $\eta = 0$ if and only if $\beta = 0$. Thus G is positive definite. The polynomial $P_{2n}(x)$ is clearly of degree $2n$ in x, and is real by Theorem II. Now

$$F_{kl}(\omega_l) = (\xi_k, \{\omega_l^2 I - 2\omega_l iA - H\}\xi_l) = 0, \quad k = 1, 2, \cdots, n, \tag{76}$$

so that

$$P_{2n}(\omega_l) = 0, \quad l = 1, 2, \cdots, n. \tag{77}$$

Suppose that Ω is an eigenvalue of Eq. (2) with eigenvector $\zeta = \sum_{l=1}^n a_l \xi_l \neq 0$. Then the column vector α with components a_k $(k = 1, 2, \cdots, n)$ is nonzero (the ξ_l are linearly independent) and

$$(\xi_k, \{\Omega^2 I - 2\Omega iA - H\}\zeta)$$
$$= \sum_{l=1}^n a_l \{\Omega^2 (\xi_k, \xi_l) - 2\Omega (\xi_k, iA\xi_l) - (\xi_k, H\xi_l)\} = 0 \tag{78}$$

for $k = 1, 2, \cdots, n$, i.e., $F(\Omega)\alpha = 0$. Thus $\alpha \neq 0$ implies $P_{2n}(\Omega) = 0$.

Finally, suppose $H > 0$, so that $H' > 0$. We define

$$P_{2n}(\epsilon; x) \equiv \det F_\epsilon(x), \quad 0 \leq \epsilon \leq 1, \tag{79}$$

where

$$F_\epsilon(x) \equiv x^2 G - 2\epsilon xiA' - H', \tag{80}$$

so that $P_{2n}(1; x) = P_{2n}(x)$ and $F_1(x) = F(x)$. Now each root of $P_{2n}(\epsilon; x)$ is an eigenvalue of the system

$$F_\epsilon(x)\eta = 0 \tag{81}$$

with a nonzero eigenvector η. Therefore $H' > 0$ and Eqs. (3)–(6) imply that all the roots of $P_{2n}(\epsilon; x)$ are real and nonzero for all real ϵ. Since the roots of $P_{2n}(\epsilon; x)$ are continuous functions of ϵ, $0 \leq \epsilon \leq 1$, the number of positive and negative roots of $P_{2n}(\epsilon; x)$ must be independent of ϵ, $0 \leq \epsilon \leq 1$. But $P_{2n}(0; x)$ has precisely n positive and n negative roots, and the proof is complete.

Lemma IV: Suppose H is positive definite. Then any set of eigenvectors $\{\xi_k\}_{k=1}^n$ of Eq. (2) with distinct positive (negative) eigenvalues ω_k (i.e., $\omega_k \neq \omega_l$ for $l \neq k$) is linearly independent.

Proof: We begin by demonstrating that if the $\{\xi_k\}_{k=1}^q$ are q linearly independent eigenvectors of Eq. (2) having distinct positive (negative) eigenvalues ω_k, then the set $\{\xi_k\}_{k=1}^{q+1}$ is also linearly independent, where ξ_{q+1} is an eigenvector of Eq. (2) with positive (negative) eigenvalue ω_{q+1}, and $\omega_{q+1} \neq \omega_k$, $k = 1, 2, \cdots, q$. Indeed, if the set $\{\xi_k\}_{k=1}^{q+1}$ were linearly dependent, we would have $P_{2q}(\omega_{q+1}) = 0$ [see Eq. (74)] by Lemma III, so that $P_{2q}(x)$ would have at least $q + 1$ distinct positive (negative) roots. But this is impossible, as H is positive definite (see Lemma III). Thus $\{\xi_k\}_{k=1}^q$ linearly independent implies that $\{\xi_k\}_{k=1}^{q+1}$ is also linearly independent. Since one eigenvector constitutes a

linearly independent set, Lemma IV follows by induction.

Theorem VII: Let H be positive definite, and $\{\xi_k\}_{k=1}^n$ be a canonical set of eigenvectors of Eq. (2) with positive (negative) eigenvalues. Then the set $\{\xi_k\}_{k=1}^n$ is linearly independent.

Proof: We denote the *distinct* elements of the set of eigenvalues corresponding to the eigenvectors ξ_k $(k = 1, 2, \cdots, n)$ by $\omega_1, \omega_2, \cdots, \omega_L$, where $L \leq n$ $(\omega_k \neq \omega_l$ for $k \neq l)$. Let $S(\omega_l)$ be the subset of $\{\xi_k\}_{k=1}^n$ containing precisely those eigenvectors possessing the eigenvalue ω_l. We relabel the ξ_k as $\xi_j^{(l)}$ where $\xi_j^{(l)} \in (S\omega_l)$ and $j = 1, 2, \cdots, m_l \leq n$. We have $\{\xi_k\}_{k=1}^n = \bigcup_{l=1}^L S(\omega_l)$, Lemma IV implies $S(\omega_l) \cap S(\omega_k) = \phi$ for $k \neq l$, and $n = \sum_{l=1}^L m_l$. Now suppose $\{\xi_k\}_{k=1}^n$ is linearly dependent, so that

$$\sum_{l=1}^{L} \sum_{j=1}^{m_l} a_j^{(l)} \xi_j^{(l)} = 0 \tag{82}$$

for some $a_q^{(p)} \neq 0$. Let $\phi_l \equiv \sum_{j=1}^{m_l} a_j^{(l)} \xi_j^{(l)}$, $l = 1, 2, \cdots, L$, so that ϕ_l is an eigenvector of Eq. (2) with eigenvalue ω_l if $\phi_l \neq 0$. Since the set $\{\xi_k\}_{k=1}^n$ is canonical, $a_q^{(p)} \neq 0$ implies $\phi_p \neq 0$. [Note that Eq. (69) states that if ξ_k and ξ_l both have the eigenvalue ω, then $(\xi_k, [\omega^2 I + H]\xi_l) = 0$ for $k \neq l$.] Equation (82), which takes the form

$$\sum_{l=1}^{L} \phi_l = 0$$

contradicts Lemma IV. Thus $\{\xi_k\}_{k=1}^n$ is linearly independent.

Theorem VIII: Let H be positive definite and $E = E_N$ (an N-dimensional unitary space). Then there exists a canonical set C of $2N$ eigenvectors of Eq. (2); $C = \{\xi_k^+\}_{k=1}^N \cup \{\xi_k^-\}_{k=1}^N$, where ξ_k^+ is an eigenvector with positive eigenvalue, ξ_k^- has a negative eigenvalue, and each of the sets $\{\xi_k^\pm\}_{k=1}^N$ spans E_N.

Proof: The existence of C is an immediate consequence of Theorem V and Lemma II. Theorem VII implies that the sets $\{\xi_k^\pm\}_{k=1}^N$ are each linearly independent and therefore span E_N.

Theorem IX: Let H be positive definite on $E = E_N$, and $\{\xi_k^+\}_{k=1}^N$ and $\{\xi_k^-\}_{k=1}^N$ be the sets of eigenvectors of Eq. (2) (with corresponding eigenvalues ω_k^+, ω_k^-) introduced in Theorem VIII. Then there exist linear operators Q^+ and Q^- defined on E_N and inner products $(\ ,\ )^+$ and $(\ ,\ )^-$ with the following properties:

(1) Q^+ is Hermitian positive definite with respect to the inner product $(\ ,\ )^+$; Q^- is Hermitian negative definite with respect to $(\ ,\ )^-$.

(2) $Q^\pm \xi_k^\pm = \omega_k^\pm \xi_k^\pm \quad k = 1, 2, \cdots, N.$

(3) $(\xi_k^+, \xi_l^+)^+ = 0 = (\xi_k^-, \xi_l^-)^- \quad \text{for} \quad k \neq l.$

Proof: Since $\{\xi_k^\pm\}_{k=1}^N$ is linearly independent and spans E_N, the following expressions uniquely define the linear operators Q^+ and Q^- on E_N:

$$Q^+\left\{\sum_1^N a_k \xi_k^+\right\} \equiv \sum_1^N a_k \omega_k^+ \xi_k^+,$$
$$Q^-\left\{\sum_1^N a_k \xi_k^-\right\} \equiv \sum_1^N a_k \omega_k^- \xi_k^-. \tag{83}$$

In particular, we have $Q^\pm \xi_k^\pm = \omega_k^\pm \xi_k^\pm$. We define, for all $\zeta, \eta \in E_N$,

$$(\zeta, \eta)^+ \equiv (Q^+\zeta, Q^+\eta) + (\zeta, H\eta),$$
$$(\zeta, \eta)^- \equiv (Q^-\zeta, Q^-\eta) + (\zeta, H\eta), \tag{84}$$

and it is clear that $(\ ,\ )^\pm$ possesses all the requirements of an inner product. Now

$$(\xi_k^\pm, \xi_l^\pm)^\pm = (Q^\pm \xi_k^\pm, Q^\pm \xi_l^\pm) + (\xi_k^\pm, H\xi_l^\pm)$$
$$= \omega_k^\pm \omega_l^\pm (\xi_k^\pm, \xi_l^\pm) + (\xi_k^\pm, H\xi_l^\pm) = \epsilon_k^\pm \delta_{kl}, \tag{85}$$

[ϵ_k is defined by Eq. (70)] since the sets $\{\xi_k^\pm\}_{k=1}^N$ are canonical. Finally, for any two vectors

$$\zeta = \sum_{k=1}^N a_k^\pm \xi_k^\pm, \quad \eta = \sum_{l=1}^N b_l^\pm \xi_l^\pm$$

in E_N, we have

$$(\zeta, Q^\pm\eta)^\pm = \sum_{k,l=1}^N \bar{a}_k^\pm b_l^\pm \omega_l^\pm (\xi_k^\pm, \xi_l^\pm)^\pm = \sum_{k=1}^N \bar{a}_k^\pm b_k^\pm \omega_k^\pm \epsilon_k^\pm$$
$$= \sum_{k,l=1}^N \omega_k^\pm \bar{a}_k^\pm b_l^\pm (\xi_k^\pm, \xi_l^\pm)^\pm = (Q^\pm\zeta, \eta)^\pm, \tag{86}$$
$$(\zeta, Q^\pm\zeta)^\pm = \sum_{k=1}^N \epsilon_k^\pm \omega_k^\pm |a_k^\pm|^2, \tag{87}$$

which proves statement (1).

Theorem X: Let H be positive definite on $E = E_N$, and let C be the canonical set of eigenvectors of Eq. (2) introduced in Theorem VIII. Then C is complete, and we have

$$\dot{x} = \sum_{k=1}^N \{\alpha_k^+ \omega_k^+ \xi_k^+ + \alpha_k^- \omega_k^- \xi_k^-\}, \quad x = \sum_{k=1}^N \{\alpha_k^+ \xi_k^+ + \alpha_k^- \xi_k^-\}, \tag{88}$$

$$(\dot{x}, \dot{x}) + (x, Hx) = \sum_1^N \{|\alpha_k^+|^2 \epsilon_k^+ + |\alpha_k^-|^2 \epsilon_k^-\}, \tag{89}$$

for all x, $\dot{x}$ in E_N, where

$$\alpha_k^\pm = \{\omega_k^\pm (\xi_k^\pm, \dot{x}) + (\xi_k^\pm, Hx)\}[\epsilon_k^\pm]^{-1},$$
$$\epsilon_k^\pm = [\omega_k^\pm]^2 (\xi_k^\pm, \xi_k^\pm) + (\xi_k^\pm, H\xi_k^\pm). \tag{90}$$

Proof: We begin by noting that Eqs. (88) and the fact that C is canonical imply that the $\alpha_k^\pm$ are necessarily given by Eqs. (90), i.e., the $\alpha_k^\pm$ are uniquely determined. Indeed, if we insert Eqs. (88) into the

quantity $(\omega_k^{\pm}\xi_k^{\pm}, \dot{x}) + (\xi_k^{\pm}, Hx)$, Eqs. (69) and (70) lead immediately to Eqs. (90). It therefore suffices (by Theorem IX) to demonstrate the existence of the vectors $y^+(=\sum_1^N \alpha_k^+\xi_k^+)$ and $y^-(=\sum_1^N \alpha_k^-\xi_k^-)$ satisfying the simultaneous equations

$$y^+ + y^- = x, \quad Q^+y^+ + Q^-y^- = \dot{x}. \tag{91}$$

Equations (91) possess the unique solution

$$y^+ = K^{-1}[\dot{x} - Q^-x], \quad y^- = K^{-1}[Q^+x - \dot{x}] \tag{92}$$

provided K^{-1} exists, where $K \equiv Q^+ - Q^-$. We complete the proof by showing that $Ky = 0$ implies $y = 0$. Suppose $Ky = 0$. There exist $2N$ constants $\gamma_k^{\pm}$ such that $y = \sum_1^N \gamma_k^+\xi_k^+ = -\sum_1^N \gamma_k^-\xi_k^-$ by Theorem VIII, so that $Ky = 0$ leads to

$$0 = \sum_1^N \{\gamma_k^+\omega_k^+\xi_k^+ + \gamma_k^-\omega_k^-\xi_k^-\},$$
$$0 = \sum_1^N \{\gamma_k^+\xi_k^+ + \gamma_k^-\xi_k^-\}. \tag{93}$$

Equations (93) are of the form of Eqs. (88) with $\dot{x} = 0$, $x = 0$, $\alpha_k^{\pm} = \gamma_k^{\pm}$, so that the $\gamma_k^{\pm}$ are uniquely given by Eqs. (90), i.e., $\gamma_k^{\pm} = 0$ for all k, and $y = 0$.

Note added in proof: The extension of many of the results obtained in this section to positive completely continuous Hermitian operators H in Hilbert space has been obtained by reducing Eq. (2) to a linear eigenvalue problem in the Hilbert space $E \times E$. This analysis will be presented in a forthcoming paper.

ACKNOWLEDGMENTS

The author would like to thank C. Goldstein, Professor H. Weitzner, and Professor J. Berkowitz for reading the manuscript. The work presented here was supported by the Magneto-Fluid Dynamics Division, Courant Institute of Mathematical Sciences, New York University, under contract AT(30-1)1480 with the U.S. Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Lowest-Order Correction to Many-Fermion Vertex Function

ARTHUR LAYZER
Stevens Institute of Technology, Hoboken, New Jersey

(Received 10 June 1966)

The lowest-order vertex correction, the exchange counterpart of the polarization correction, is evaluated for all values of the energy and momentum transfer variables and for an arbitrary rotationally invariant two-body interaction. The method employs a dispersion theory technique and an angular expansion of the interaction in spherical harmonics. For any partial wave, the result is expressed as a one-dimensional integral over a closed interval that can easily be evaluated numerically once the form of the interaction is specified. Explicit expressions can be obtained for the long wavelength limit, and for an expansion in powers of the momentum transfer variable. In this case the result depends on the value of the interaction and its momentum derivatives when one of the momentum variables lies on the Fermi surface.

1. INTRODUCTION

THE vertex function plays an important role in the investigation of many-fermion systems. It determines, for example, the weak response of the system to an external agent.[1] Furthermore, the Landau theory of the normal system can be re-expressed in terms of the vertex function.[2]

The complete vertex function, Γ, can be factored into the inverse dielectric constant, ϵ^{-1}, and the proper vertex function, $\tilde{\Gamma}$. The former takes account of the polarization screening of the interaction. The latter incorporates exchange effects.

Although the perturbation value of Γ is not adequate for realistic calculations, it does provide a very rough estimate of the result. Moreover, it gives a clue to the mathematical properties of more accurate nonperturbative approximations.

The perturbation expression for ϵ^{-1}, in other words the evaluation of the familiar "bubble" diagram is well known and has been given in explicit form.[3,4] Somewhat surprisingly, the perturbation expression for $\tilde{\Gamma}$, corresponding to the triangle diagram of Fig. 1, has not been evaluated to the same degree of completeness, although partial evaluations have been given.

This has been partly the result of a lack of interest

[1] A. Layzer, Ann. Phys. (N.Y.) **35**, 67 (1965).
[2] P. Nozieres, *Proprietes generales des gaz de fermions* (Dunod Cie., Paris, 1963).
[3] J. Lindhard, Kgl. Danske Videnskab. Selskab. Mat-Fys. Medd. **28**, 8 (1954).
[4] D. V. Dubois, Ann. Phys. (N.Y.) **7**, 174 (1959); **8**, 24 (1959).

quantity $(\omega_k^{\pm}\xi_k^{\pm}, \dot{x}) + (\xi_k^{\pm}, Hx)$, Eqs. (69) and (70) lead immediately to Eqs. (90). It therefore suffices (by Theorem IX) to demonstrate the existence of the vectors $y^+(=\sum_1^N \alpha_k^+\xi_k^+)$ and $y^-(=\sum_1^N \alpha_k^-\xi_k^-)$ satisfying the simultaneous equations

$$y^+ + y^- = x, \quad Q^+y^+ + Q^-y^- = \dot{x}. \tag{91}$$

Equations (91) possess the unique solution

$$y^+ = K^{-1}[\dot{x} - Q^-x], \quad y^- = K^{-1}[Q^+x - \dot{x}] \tag{92}$$

provided K^{-1} exists, where $K \equiv Q^+ - Q^-$. We complete the proof by showing that $Ky = 0$ implies $y = 0$. Suppose $Ky = 0$. There exist $2N$ constants $\gamma_k^{\pm}$ such that $y = \sum_1^N \gamma_k^+\xi_k^+ = -\sum_1^N \gamma_k^-\xi_k^-$ by Theorem VIII, so that $Ky = 0$ leads to

$$0 = \sum_1^N \{\gamma_k^+\omega_k^+\xi_k^+ + \gamma_k^-\omega_k^-\xi_k^-\},$$
$$0 = \sum_1^N \{\gamma_k^+\xi_k^+ + \gamma_k^-\xi_k^-\}. \tag{93}$$

Equations (93) are of the form of Eqs. (88) with $\dot{x} = 0$, $x = 0$, $\alpha_k^{\pm} = \gamma_k^{\pm}$, so that the $\gamma_k^{\pm}$ are uniquely given by Eqs. (90), i.e., $\gamma_k^{\pm} = 0$ for all k, and $y = 0$.

Note added in proof: The extension of many of the results obtained in this section to positive completely continuous Hermitian operators H in Hilbert space has been obtained by reducing Eq. (2) to a linear eigenvalue problem in the Hilbert space $E \times E$. This analysis will be presented in a forthcoming paper.

ACKNOWLEDGMENTS

The author would like to thank C. Goldstein, Professor H. Weitzner, and Professor J. Berkowitz for reading the manuscript. The work presented here was supported by the Magneto-Fluid Dynamics Division, Courant Institute of Mathematical Sciences, New York University, under contract AT(30-1)1480 with the U.S. Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Lowest-Order Correction to Many-Fermion Vertex Function

ARTHUR LAYZER
Stevens Institute of Technology, Hoboken, New Jersey

(Received 10 June 1966)

The lowest-order vertex correction, the exchange counterpart of the polarization correction, is evaluated for all values of the energy and momentum transfer variables and for an arbitrary rotationally invariant two-body interaction. The method employs a dispersion theory technique and an angular expansion of the interaction in spherical harmonics. For any partial wave, the result is expressed as a one-dimensional integral over a closed interval that can easily be evaluated numerically once the form of the interaction is specified. Explicit expressions can be obtained for the long wavelength limit, and for an expansion in powers of the momentum transfer variable. In this case the result depends on the value of the interaction and its momentum derivatives when one of the momentum variables lies on the Fermi surface.

1. INTRODUCTION

THE vertex function plays an important role in the investigation of many-fermion systems. It determines, for example, the weak response of the system to an external agent.[1] Furthermore, the Landau theory of the normal system can be re-expressed in terms of the vertex function.[2]

The complete vertex function, Γ, can be factored into the inverse dielectric constant, ϵ^{-1}, and the proper vertex function, $\tilde{\Gamma}$. The former takes account of the polarization screening of the interaction. The latter incorporates exchange effects.

Although the perturbation value of Γ is not adequate for realistic calculations, it does provide a very rough estimate of the result. Moreover, it gives a clue to the mathematical properties of more accurate nonperturbative approximations.

The perturbation expression for ϵ^{-1}, in other words the evaluation of the familiar "bubble" diagram is well known and has been given in explicit form.[3,4] Somewhat surprisingly, the perturbation expression for $\tilde{\Gamma}$, corresponding to the triangle diagram of Fig. 1, has not been evaluated to the same degree of completeness, although partial evaluations have been given.

This has been partly the result of a lack of interest

[1] A. Layzer, Ann. Phys. (N.Y.) **35**, 67 (1965).
[2] P. Nozieres, *Proprietes generales des gaz de fermions* (Dunod Cie., Paris, 1963).
[3] J. Lindhard, Kgl. Danske Videnskab. Selskab. Mat-Fys. Medd. **28**, 8 (1954).
[4] D. V. Dubois, Ann. Phys. (N.Y.) **7**, 174 (1959); **8**, 24 (1959).

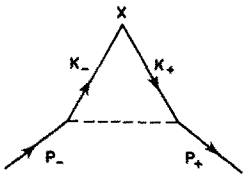


FIG. 1. Triangle diagram for lowest-order vertex correction. The cross denotes an external potential or interaction.

arising from the common but overly optimistic belief that exchange corrections are not really important. Partly, it is due to the somewhat increased mathematical complexity of the evaluation. In particular, unlike the bubble diagram, the integration now depends explicitly on the particular form of the interaction. The increased complexity, however, turns out to be fairly minor when one makes use of dispersion relation techniques and a partial wave analysis of the interaction.

In the work below, with the aid of these techniques, the evaluation of the triangle diagram of Fig. 1 is carried through to the point of one-dimensional integrations over a closed domain of integration, for all momentum and energy transfers.

To increase the region of validity of the result somewhat beyond that of ordinary perturbation theory, we have allowed the two-body interaction v to be an arbitrary *nonlocal*, but still instantaneous, potential. As in the perturbation case, we assume that v is real and symmetric in momentum space, though the symmetry property is not actually used.

In the static long wavelength limit, the results for the vertex correction agree with those obtained earlier from the use of a Ward's identity.[5]

2. EVALUATION OF $\tilde{\Gamma}$ IN LOWEST ORDER

The diagram for the lowest-order vertex correction, γ, is shown in Fig. 1. According to the Feynman rules, we can write γ in the form

$$\gamma(p, q) \equiv \tilde{\Gamma} - 1 = ci\int d^4k G_0(k_-)G_0(k_+)v(\mathbf{p}, \mathbf{k}), \quad (1)$$

$$c = (2\pi)^{-4},$$

where

$$k_\pm \equiv k \pm \tfrac{1}{2}q \quad (2)$$

and G_0 is the free propagator,

$$G_0 = [k_0 - k^2/(2m) + i\epsilon \operatorname{sgn}(|\mathbf{k}| - k_F)]^{-1}. \quad (3)$$

The invariance of γ under time reversal is expressed by the symmetry property, evident from (1)

$$\gamma(p, -q) = \gamma(p, +q). \quad (4)$$

Provided only that v is instantaneous γ is independent of $E \equiv p_0$. For rotationally invariant v, γ then depends on only four variables which we take to be the quantities $|\mathbf{q}|$, $|\mathbf{p}|$, $x_{pq} \equiv \cos(\mathbf{p}, \mathbf{q})$, and $w \equiv q_0$. Instead of w it is more convenient to deal with the ratio $u \equiv w/|\mathbf{q}|$ since the vertex function is not uniquely defined in the $\mathbf{q} \to 0$, $w \to 0$ limit unless the ratio u is also specified.

In the following, we obtain explicit results, with no remaining nontrivial integrations, for several cases: the $q = 0$ limit for all u and for arbitrary interaction v; corrections to this of order q^2 and, for the special case of the delta function interaction, results for all q and u. In the case of the general interaction away from $q = 0$, the reduction is carried to the point of one-dimensional integrals over a closed interval. These integrals can easily be evaluated numerically once the form of the interaction is specified.

It turns out that the calculation is simplified by employing a formal dispersion relation technique with respect to the energy-transfer variable w or u, considered as a complex variable. Thus, we first evaluate the simpler quantity $\operatorname{Im}\gamma(u)$ and then express $\operatorname{Re}\gamma(u)$ in terms of this result by a Cauchy principal-value integral. As might be expected we need to know $\operatorname{Im}\gamma$ only "on the energy shell," when w is equal to the energy of intermediate particle-hole pairs with the momenta $(\mathbf{k}_-, \mathbf{k}_+)$.

Our reduction also utilizes a "partial wave" expansion of $v(\mathbf{p}, \mathbf{k})$ in Legendre polynomials in the variable $x_{pk} \equiv \cos(\mathbf{p}, \mathbf{k})$. In the static $q \to 0$ limit only the s-wave term contributes to the result.

We begin the evaluation of (1) by performing the k_0 integration, closing the contour of integration in the upper half of the complex k_0 plane. We then obtain two terms, a time-reversed pair, one proportional to $\theta(k_F - k_-)\theta(k_+ - k_F)$ due to a pole at $(k_-)_0 = T(k_-) + i\epsilon$, the other proportional to $\theta(k_F - k_+)\theta(k_- - k_F)$ due to a pole at $(k_+)_0 = T(k_+) + i\epsilon$. Here $T(k) = k^2/2m$. The result is

$$\gamma = \gamma^+ + \gamma^-, \quad (5a)$$

$$\gamma^-(p, q) = \gamma^+(p, -q), \quad (5b)$$

where explicitly

$$\gamma^+ = 2\pi i^2 c\int d^3k v(\mathbf{p}, \mathbf{k}) \frac{\theta(1 - k_-)\theta(k_+ - 1)}{w - \mathbf{q}\cdot\mathbf{k} + 2i\epsilon} \quad (6)$$

with the units

$$\hbar = m = k_F = 1. \quad (7)$$

The imaginary part of this expression is somewhat easier to evaluate than the real part. We wish to take advantage of this situation by using a Kramers–Kronig type of dispersion relation to express the real part in terms of the imaginary part. Let us observe, however, that γ^+ and γ^- have different analytic properties as a function of w or u regarded as a complex variable: the former is analytic in the upper

[5] A. Layzer, Phys. Letters 13, 121 (1964). See also Ref. 1.

half-plane, the latter in the lower half-plane. Thus, we must treat γ^+ and γ^- separately as far as the dispersion relations are concerned.

From the preceding discussion, we can write for real u

$$\operatorname{Im}\gamma^+(u) = f^{\pm}(u)\theta(\pm u), \tag{8}$$

$$\operatorname{Re}\gamma^+(u) = +\frac{1}{\pi}\int_0^\infty \frac{du' f^+(u')}{u'-u}, \tag{9a}$$

$$\operatorname{Re}\gamma^-(u) = -\frac{1}{\pi}\int_{-\infty}^0 \frac{du' f^-(u')}{u'-u}. \tag{9b}$$

Equation (8) expresses the fact that γ^+ and γ^- are also the positive and negative frequency parts of γ. In the integrals in (9), principal values are understood.

From (6), Im γ^+ has the value

$$\operatorname{Im}\gamma^+ = 2\pi^2 c\int d^3k v(\mathbf{p},\mathbf{k})\theta(1-k_-) \times \theta(k_+ - 1)\delta(w - \mathbf{q}\cdot\mathbf{k}). \tag{10}$$

Note that, in agreement with (8), this is nonvanishing only for

$$w = \mathbf{q}\cdot\mathbf{k} = T(k_+) - T(k_-) \geq 0. \tag{11}$$

In order to facilitate the evaluation of γ, it is convenient to expand it in Legendre polynominals of the cosine of the angle between $\mathbf{p}$ and $\mathbf{q}$,

$$\gamma = \sum_{l=0}^{\infty}\gamma_l P_l(x_{pq}), \tag{12}$$

and similarly for f.

For this purpose we introduce the corresponding expansion coefficients v_l according to

$$v(\mathbf{p},\mathbf{k}) = \sum_{l=0}^{\infty} v_l(p,k)P_l(x_{pk}). \tag{13}$$

As far as the angular k integration is concerned we can now make the replacement

$$P_l(x_{pk}) \to P_l(x_{pq})P_l(x_{qk}). \tag{14}$$

This yields the desired expansion (12). The angular k-integration is now trivial due to the delta function in (10). We obtain for f_l the expression

$$f_l^+ = 2\pi^3 c\frac{1}{q}\int_{1-w-q^2/4}^{1+w-q^2/4} d(k^2)v_l(p,k)P_l(u/k)\theta(k^2-u^2), \tag{15}$$

and from (5b), for negative u

$$f_l^-(q,u) = (-1)^l f_l^+(q,-u), \tag{16}$$

where now $q \equiv |\mathbf{q}|$.

This last relation allows us to express γ_l in terms of f_l^+ alone through the dispersion relations (9). These take different forms according to whether l is even or odd,

for even l:

$$\operatorname{Im}\gamma_l(u) = f_l^+(|u|), \tag{17a}$$

$$\operatorname{Re}\gamma_l(u) = \frac{2}{\pi}\int\!\!\!\!\!\!\diagup_0^\infty f_l^+(u')\frac{u'\,du'}{u'^2-u^2}; \tag{18a}$$

for odd l:

$$\operatorname{Im}\gamma_l(u) = f_l^+(|u|)\epsilon(u), \tag{17b}$$

$$\operatorname{Re}\gamma_l(u) = \frac{2u}{\pi}\int\!\!\!\!\!\!\diagup_0^\infty f_l^+(u')\frac{du'}{u'^2-u^2}, \tag{18b}$$

where $\epsilon(u) = \theta(u) - \theta(-u)$. The slash through the integral sign indicates that the principal value must be taken.

3. THE LIMIT $\mathbf{q} \to 0$

It is useful at this point to consider special cases in which exact final results can be obtained analytically without too much effort. One of these is the important limiting case $\mathbf{q} \to 0$ with u held finite, the long wavelength or semi-classical limit.

In connection with the $\mathbf{q} \to 0$ limit, we first note that f_l^+, given in (15), is finite in this limit: as $\mathbf{q} \to 0$, the upper and lower integration limits coincide (since $w = qu$) preventing a divergence due to the factor q^{-1}. To show this more explicitly it is convenient to eliminate the k integration in favor of a parametric z integration through the substitution

$$k^2 = 1 + quz - \tfrac{1}{4}q^2. \tag{19}$$

We then obtain, from (15)

$$f_l^+ = 2\pi^3 cu\int_{-1}^{1} dz v_l(p,k)P_l(u/k)\theta(k^2-u^2). \tag{20}$$

In the limit $\mathbf{q} \to 0$, the function k may be replaced by unity:

$$\lim_{q\to 0} f_l^+(\mathbf{q},u) \equiv f_l^+(0,u) = 4\pi^3 cuv_l(p,1)P_l(u)\theta(1-u^2). \tag{21}$$

We see that Im γ vanishes in the static limit, $u \to 0$, and also if the absolute value of u is greater than unity.

From (21) we can obtain Re γ via the dispersion relations (18). The u' integration is trivial for any fixed value of l. Let us concentrate attention on the $l = 0$ term.

We obtain then in the $\mathbf{q} \to 0$ limit

$$\operatorname{Re}\gamma_0(0,u;p) = 4\pi^3 cv_0(p,1)I(u), \tag{22}$$

where

$$I(u) = \frac{2}{\pi}\int\!\!\!\!\!\!\diagup_0^1 \frac{u'^2\,du'}{u'^2-u^2} = \frac{2}{\pi}\left[1 - \tfrac{1}{2}u\ln\left|\frac{1+u}{1-u}\right|\right]. \tag{23}$$

Let us note that I has an integrable, logarithmic singularity at $u = 1$.

We see that Re γ vanishes for $u \to \infty$ as it should but not in the static limit, $u \to 0$. In the static limit, (22) yields

$$\gamma_0(0,0,p) = \frac{1}{2\pi^2}v_0(p,1). \tag{24}$$

One easily verifies that, due to the orthogonality properties of the Legendre polynomials, only the $l = 0$

term contributes to $\mathrm{Re}\,\gamma$ in the $q, u \to 0$ limit. Equation (24) agrees with the value of $\gamma(0, 0; p)$ obtained earlier on the basis of gauge invariance.[5]

4. DELTA-FUNCTION INTERACTION

There is at least one form of interaction v simple enough to allow us to evaluate the vertex correction γ exactly for all q and u. This is the case of a delta-function interaction with strength a.

$$v(\mathbf{p}, \mathbf{k}) = a \tag{25}$$

or

$$v(\mathbf{x}, \mathbf{x}') = a\delta^3(\mathbf{x} - \mathbf{x}'). \tag{26}$$

Clearly $v(\mathbf{p}, \mathbf{k})$ has only an $l = 0$ component, $v = v_0$.

When we substitute (25) for v into (20) we obtain

$$\mathrm{Im}\,\gamma_\delta = 2\pi^3 cau \int_{-1}^{1} dz\theta(1 + quz - \tfrac{1}{4}q^2 - u^2), \tag{27}$$

where γ_δ is the value of γ for the delta-function interaction (25).

The z integration is trivial and yields the result

$$\mathrm{Im}\,\gamma_\delta = \begin{cases} (4\pi)^{-1}a\,|u|, & |u| + \tfrac{1}{2}q < 1, \quad (28a) \\ (8\pi)^{-1}a\,\dfrac{1}{q}\,[(1 - (|u| - \tfrac{1}{2}q)^2], & \\ \qquad\qquad ||u| - \tfrac{1}{2}q| < 1 < |u| + \tfrac{1}{2}q, & (28b) \\ 0, \qquad\qquad ||u| - \tfrac{1}{2}q| > 1. & (28c) \end{cases}$$

To calculate $\mathrm{Re}\,\gamma$ we employ the dispersion relation (18a). After carrying out the elementary u' integration we obtain

$$\mathrm{Re}\,\gamma_\delta = (4\pi^2)^{-1}a\Big\{1 + \frac{1}{2q}\,[1 - (u + \tfrac{1}{2}q)^2] \times \ln\left|\frac{1 + \tfrac{1}{2}q + u}{1 - \tfrac{1}{2}q - u}\right| + \frac{1}{2q}\,[1 - (u - \tfrac{1}{2}q)^2] \times \ln\left|\frac{1 + \tfrac{1}{2}q - u}{1 - \tfrac{1}{2}q + u}\right|\Big\}. \tag{29}$$

The result (28) and (29) for γ_δ has a familiar mathematical appearance. Aside from a numerical factor, it agrees precisely in form with the well-known value of the polarization loop. This is hardly surprising since for a delta-function interaction the dotted line of the vertex diagram is effectively collapsed to a point (in position space). Thus, it is clear that for a delta-function interaction the lowest-order polarization correction, $a\pi$, should have the same value as the proper vertex correction except for a factor of -2 arising from the exchange nature of the vertex correction:

$$a\pi = -2\gamma_\delta, \tag{30}$$

which is indeed true.

5. GENERAL FORM OF INTERACTION

Equation (18) cannot be explicitly evaluated for all q and u for the general interaction, v. One can, however, further reduce the expression for $\mathrm{Re}\,\gamma$ by eliminating one of the two remaining integrations.

According to (18) and (20) we can write $\mathrm{Re}\,\gamma$ in the form

$$\mathrm{Re}\,\gamma_l = 2\pi^2 c \int_0^\infty du' \left[\frac{1}{u' - u} + (-1)^l \frac{1}{u' + u}\right] \times \int_{-1}^{+1} u'\,dz\{v_l(p, \bar{k})P_l(u'/\bar{k})\theta(\bar{k}^2 - u'^2)\}, \tag{31}$$

where

$$\bar{k}^2 \equiv 1 + qu'z - \tfrac{1}{4}q^2 \geq 0. \tag{32}$$

We make the substitution of variables $z \to y$, where

$$y \equiv u'z - \tfrac{1}{2}q. \tag{33}$$

Then, we have

$$\bar{k}^2 = 1 + qy + \tfrac{1}{4}q^2, \tag{34}$$

where $\bar{k}$ is now independent of the variable u'. Interchanging the order of integration in (31) we obtain

$$\mathrm{Re}\,\gamma_l = 2\pi^2 c \int_{-1}^{+1} v_l(p, \bar{k})F_l(y)dy, \tag{35}$$

where

$$F_l(y) = \int_{|y + \frac{1}{2}q|}^{k} du' P_l(u'/\bar{k})\left[\frac{1}{u' - u} + (-1)^l \frac{1}{u' + u}\right]. \tag{36}$$

From (34) we see that

$$\bar{k}(y = \pm 1) = |\pm 1 + \tfrac{1}{2}q| \tag{37}$$

and therefore

$$F_l(\pm 1) = 0. \tag{38}$$

The integration for $F_l(y)$ is trivial for any value of l. Therefore only one nontrivial integration remains, as in the case of $\mathrm{Im}\,\gamma$.

For $l = 0$, F_l has the value

$$F_0(y) = 2 \ln\left|\frac{\bar{k}^2 - u^2}{(y + \tfrac{1}{2}q)^2 - u^2}\right|. \tag{39}$$

We can easily verify that the previous results for the $\mathbf{q} \to 0$ limit and the delta-function interaction follow from (35) and (39). In the static limit, one finds for the coefficient of the q^2 term the following result, which agrees with (29) for the case of the delta-function interaction

$$\gamma_0(q, 0, p) - \gamma(0, 0, p) = \frac{1}{2\pi^2}\,a(p)q^2 + 0(>q^2), \tag{40}$$

$$2a(p) = -\tfrac{1}{3}v_0(p, 1) - \tfrac{2}{9}v_0^{(1)}(p, 1) + \tfrac{1}{18}v_0^{(2)}(p, 1), \tag{41}$$

where

$$v^{(n)}(p, 1) \equiv \lim_{k \to 1} (d^n/dk^n)v(p, k). \tag{42}$$

In general, the expansion of γ in powers of q involves higher-order derivatives of the various $v_l(p, k)$ at the Fermi surface, $k = 1$. In the static limit, $u = 0$, one sees from (18), (19), and (20) that only even powers of q and even values of l can occur.

Relationship of the Internal and External Multiplicity Structure of Compact Simple Lie Groups*

A. J. MACFARLANE, L. O'RAIFEARTAIGH,† AND P. S. RAO
Department of Physics, Syracuse University, Syracuse, New York

(Received 11 April 1966)

As a preliminary step in a program aimed at developing the Racah algebra of an arbitrary compact simple Lie group L, this paper gives a unified review with various extensions of the work of Biedenharn, Speiser, and others on the relationship between the internal and external multiplicity structures of L, the former being that of the weights of the representations of L, the latter being that of the terms of the Clebsch–Gordan series of L.

1. INTRODUCTION

IN view of the present preoccupation of particle and other physicists with symmetry groups of various types, it is undoubtedly desirable that the Racah algebra of such groups be developed as extensively as is the Racah algebra of SU_2, i.e., the familiar quantum theory of angular momentum. If L denotes a compact simple Lie group, then, as stressed by Wigner, Racah and Biedenharn,[1] a variety of problems have to be solved before the Racah algebra of L can be systematically investigated. The first problem, which is closely related to the labeling of the irreducible representations (IR's) of L, is the construction of the invariants or Casimir operators of L. This problem has already been solved.[2] The second problem concerns the determination of operators whose eigenvalues yield a complete characterization of the states of the IR's of L. A given IR of L is specified by its eigenvalues of the invariants of L, or else equivalently and more usually, by the components of its highest weight. From this highest weight all the weights of the IR can be directly deduced.[3] However, the weights other than the highest are not in general simple but rather of multiplicity greater than one. In this context, we speak of the internal multiplicity structure (of the IR's) of L. The various states belonging to the multiple weights are to be distinguished by the operators mentioned above and we refer to the problem of determining these operators as the internal labeling problem. The third problem is that of the Clebsch–Gordan series and coefficients of L. Its solution involves the explicit reduction of direct products of IR's of L. Most of the difficulties here stem from the fact that such products are not in general simply reducible; i.e., the repeated occurrence of IR's in a direct product is possible. We refer to the multiplicity structure of the reductions of direct products of IR's of L as the external multiplicity structure of L. Associated with it there is an external labeling problem, that of determining operators whose eigenvalues can distinguish the multiple occurrences of IR's of L in reductions of direct products.

At present the internal multiplicity structure of any compact simple Lie group is either known[4] or at least accessible from Kostant's formula,[5] but the internal labeling problem has been solved only for unitary groups[6] and orthogonal groups.[7] An implicit determination of the external multiplicity structure is contained in the formulas of Steinberg and Straumann,[8] while more explicit knowledge can be obtained by a rather wide variety of methods, for example, by using tensorial methods[9] or by using Speiser's

* Research supported in part by U.S. Atomic Energy Commission.
† On leave of absence from Dublin Institute for Advanced Studies.
[1] See L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963).
[2] G. Racah, Rend. Lincei **8**, 108 (1950); G. Racah, *Lecture Notes on Group Theory and Spectroscopy* (Institute for Advanced Study, Princeton, New Jersey, 1951), reprinted as CERN report 61-8, CERN, Geneva (1961), and published in *Ergebnisse der Exakten Naturwissenschaften* (Springer-Verlag, Berlin, 1965) Vol. 37; L. C. Biedenharn, Ref. 1; B. Gruber and L. S. O'Raifeartaigh, J. Math. Phys. **5**, 1796 (1964); M. Umezawa, Nucl. Phys. **48**, 111 (1963); **53**, 54 (1964); **57**, 65 (1964); Strasbourg Preprint; T. S. Santhanam, ICTP reprint 65-86, Trieste (1965).
[3] G. Racah, in *Lectures on Lie Groups in Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gursey, Ed. (Gordon and Breach Science Publishers, Inc., New York, 1964); A. Salam, *Formalism of Lie Groups* in *Proceedings of the 1962 Trieste Seminar in Theoretical Physics* (IAEA, Vienna, 1963); J. P. Antoine and D. R. Speiser, J. Math. Phys. **5**, 1226, 1560 (1964); D. R. Speiser, Helv. Phys. Acta **38**, 73 (1965); R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[4] See the papers cited in Ref. 3, especially that of Racah.
[5] See N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1962), p. 261. In connection with Kostant's formula, see J. Tarski, J. Math. Phys. **4**, 569 (1963).
[6] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1499 (1963). See also H. Weyl, *Theory of Groups and Quantum Mechanics* (Methuen and Company, Ltd., London, 1931). S. Gasiorowicz, Argonne National Laboratory Report No. 6729 (1963) (unpublished).
[7] G. Racah has obtained but not published a solution of internal labeling problem for orthogonal groups (private communication).
[8] Steinberg's formula is discussed in Jacobson, Ref. 5, p. 262. N. Straumann [Helv. Phys. Acta **38**, 56 (1965)] has discussed its use in practical situations and obtained an alternative formula, CERN preprints 65 320 5 Th. 527, (1965).
[9] See R. E. Brehends *et al.* or N. Mukunda and L. K. Pandit, Progr. Theoret. Phys. (Kyoto) **34**, 46 (1965).

method,[10] or else, in the case of unitary groups, by using Young diagrams.[11] So far no published work of a general nature on the external labeling problem is available, although it has been solved in the case of SU_3 by Moshinsky.[12] Finally, as far as Clebsch–Gordan coefficients are concerned, progress of a general nature has been made by Derome and Sharpe,[13] who recognize the existence of the external multiplicity problem but perform their investigation without considering any special solution of it. Explicit work on coefficients for situations in which the external multiplicity question enters nontrivially has already been performed in special cases, especially the case of SU_3.[14]

The present paper is designed to serve as an introduction to a program aimed at solving both the internal and external labeling problems for arbitrary compact simple Lie groups. The idea of this approach is first to solve the external labeling problem explicitly and then to proceed to a solution of the internal labeling problem by exploiting the intimate relationship between the internal and external multiplicity structures. We have constructed the operators—polarized Casimir operators—which solve the external multiplicity problem for the classical groups.[15] The present paper is a unified review which contains, however, some new results and new proofs and which places emphasis suitably for later parts of our program on the work of Biedenharn, Speiser, and others (whose contributions are cited below) on the relationship of the internal and external multiplicity structures of an arbitrary compact simple Lie group L. In Sec. 2, we state and prove the following lemma which we call Biedenharn's lemma. The lemma was first given by Kostant[16] and was later rediscovered by Biedenharn,[16] who was the first to realize its value for physical applications.

Lemma: Let R and R' be two IR's of a compact simple Lie group L with highest weights Λ and Λ', respectively. Let m with multiplicity γ_m denote the weights of R. If Λ' is so much higher than Λ that $\Lambda' + m$ is a dominant weight of L for each $m \in R$, then in the reduction of $R \times R'$ the IR of L of highest weight $\Lambda' + m$ occurs exactly γ_m times for each $m \in R$.

Despite the restriction of the type of direct product $R \times R'$ to which this lemma applies, this is the fundamental result for our purposes, as is explained below. We have proved it very simply using only general properties of an arbitrary compact simple Lie group. In Sec. 3, we obtain necessary and sufficient conditions on Λ' for fixed Λ, for each compact simple Lie group, which ensure that $\Lambda' + m$ is dominant for each $m \in R$. The content of Secs. 2 and 3 has been discussed in detail by Mukunda and Pandit,[9] and by Preziosi, Simone, and Vitale[17] in the case of SU_3. These discussions given by these authors use specific details of SU_3 representation theory and these results are obtained much less directly than ours are here. Later Vitale[18] extended the discussion of Preziosi, Simone, and Vitale to the other rank two Lie groups. Also, Nussinov[19] has obtained results agreeing with those of Sec. 3 for SU_n, and very recently Zaccaria[20] has done likewise for the classical groups. Our results were obtained independently of those of Zaccaria; since our approach is purely algebraic and in the spirit of the rest of our paper, whereas Zaccaria has used intuitive geometrical notions, we feel justified in presenting our method in detail. In Sec. 4 we state and prove the generalization of Biedenharn's lemma to the case of a general direct product $R \times R'$. In

[10] D. R. Speiser, in *Lectures on Theory of Compact Lie Groups in Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gursey, Ed. (Gordon and Breach Science Publishers, Inc., New York, 1964), and Speiser, Ref. 3.

[11] D. E. Littlewood, *Theory of Group Characters* (Oxford University Press, New York, 1950), p. 94. See also A. R. Edmonds, Proc. Roy. Soc. (London) **A268**, 567 (1962); and C. Itzykson and M. Nauenberg, Rev. Mod. Phys. **38**, 95 (1966).

[12] (a) M. Moshinsky, J. Math. Phys. **4**, 1128 (1963) and (b) private communication to A. J. Macfarlane.

[13] J. R. Derome and W. T. Sharpe, J. Math. Phys. **6**, 1584 (1965); see also A. J. Macfarlane, N. Mukunda, and E. C. G. Sudarshan, J. Math. Phys. **5**, 576 (1964).

[14] K. T. Hecht, Nucl. Phys. **62**, 1 (1965); J. G. Kuriyan, D. Lurie, and A. J. Macfarlane, J. Math. Phys. **6**, 722 (1965); T. A. Brody, M. Moshinsky, and I. Renero, J. Math. Phys. **6**, 1540 (1965); G. E. Baird and L. C. Biedenharn, Duke University preprint (1965); M. Resnikoff, Ph.D. thesis, Michigan (1965), and J. Math. Phys. **8**, 63, 79 (1967); L. Banyai, N. Marinesen, I. Raszillier, and V. Rittenberg, Phys. Letters **14**, 156 (1965), and Bucharest preprint (1965); Hou Tei-yu, Sci. Sinica **14**, 367 (1965); G. Ponzano, Torino preprint (1965). See also M. Moshinsky, Rev. Mod. Phys. **34**, 813 (1962), and Ref. 12(a); J. J. deSwart, Rev. Mod. Phys. **35**, 916 (1963); Refs. 14–18 of the paper by J. G. Kuriyan *et al.* noted above; I. S. Gerstein and K. T. Mahanthappa, Nuovo Cimento **32**, 239 (1964); L. K. Pandit and N. Mukunda, J. Math. Phys. **6**, 1574 (1965). For C–G coefficients of groups other than SU_2 and SU_3, see J. C. Carter, J. J. Coyne, and S. Meshkov, Phys. Rev. Letters **14**, 523, 1850(E) (1965); C. L. Cook and G. Murtaza, Nuovo Cimento **39**, 532 (1965) for SU_6, and K. T. Hecht, Nucl. Phys. **63**, 177 (1965), and J. N. Ginocchio, Rochester preprint UR875-75 (1965) for R_5.

[15] A. J. Macfarlane, L. O'Raifeartaigh, and P. S. Rao (to be published). Preliminary accounts of this work have been given by L. O'Raifeartaigh, Bull. Am. Phys. Soc. **10**, 483 (1965).

[16] B. Kostant, Trans. Am. Math. Soc. **93**, 53 (1959); L. C. Biedenharn, Phys. Letters **3**, 254 (1963); G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5**, 1730 (1964). For the classical groups the content of the lemma is implicitly contained in H. Weyl, *Classical Groups* (Princeton University Press, Princeton, New Jersey, 1946), p. 231, Theorem (7.10A). We are indebted to Professor A. J. Coleman for informing us that the general idea of the lemma had already occurred to him in 1957 and apparently was known to R. Brauer and H. Weyl as early as 1930.

[17] B. Preziosi, A. Simone, and B. Vitale, Nuovo Cimento **34**, 1101 (1964). See also A. Simone and B. Vitale, *ibid.* **38**, 1199 (1964).

[18] B. Vitale, University of Wisconsin preprint (1965).

[19] S. Nussinov, University of Washington preprint (1965).

[20] A. Zaccaria, Napoli preprint (1965).

view of the facts (a) that the form of our result has been given already by Racah,[3] and (b) that it is essentially an algebraic statement of the result underlying Speiser's geometrical method[10,3] for reducing direct products, we refer to the generalization of Biedenharn's lemma as the Racah–Speiser lemma. The first available discussions of Speiser's method to be found in the literature were those of Antoine[21] and de-Swart[14] both of which deal with the case of SU_3. Some details regarding the application of Speiser's geometrical method to other cases are given by Speiser himself. In view of the difficulty of applying geometrical techniques in spaces of more than two (even three!) dimensions, the desirability of the algebraic statement should be apparent. Finally, it should be mentioned that the work of Sec. IV-D of the paper by Behrends, Dreitlein, Fronsdal, and Lee[22] is essentially equivalent to Speiser's method.

Sections 2–4 contain the basic results of the relationship between the external and internal multiplicity structures of an arbitrary compact simple Lie group. We conclude this Introduction by explaining our view that it is Biedenharn's lemma rather than the Racah–Speiser lemma which has the more fundamental significance. The view stems from the fact that we hope to proceed to the development of the Racah algebra of an arbitrary compact simple Lie group L and in particular to define a complete set of Clebsch–Gordan coefficients of L as matrix elements of a complete set of irreducible tensor operators. We thus regard such operators as fundamental entities in the Racah algebra of L and explain how Biedenharn's lemma rather than the Racah–Speiser lemma is exploited in classifying them. The basic irreducible tensor operators of L possess, in addition to the representation labels and internal labels which specify their transformation properties under L, external labels which specify the changes they induce on the representation labels of states of IR's of L. Let $|R'\,i'\,m'\rangle$ be a state of the irreducible representation R' of L with highest weight Λ', with weights m' and eigenvalues i' of the operators which solve the internal multiplicity problem for L. Let T^R_{jq} be an irreducible tensor operator which transforms like $|R\,j\,q\rangle$ under L. If Λ' is high enough with respect to Λ then T^R_{jq} can, according to Biedenharn's lemma, induce on $|R'\,i'\,m'\rangle$ any or all of the changes

$$R' \to R'' = R' + m,$$

where m is a weight of R, on the representation labels R' of the state $|R'\,i'\,m'\rangle$. Let us then define for each weight m of R, tensor operators $T^R_{m,jq}$ which induce only the change $R' \to R'' + m$, i.e., such that the reduced matrix elements

$$\langle R''\| T^R_m \|R'\rangle$$

vanish unless $R'' = R' + m$. By the lemma again there are γ_m independent such operators for each $m \in R$ since $R' + m$ occurs γ_m times in $R \otimes R'$, so that in all there are dim R independent components $T^R_{m,jq}$ for fixed R, j, q. Thus Biedenharn's lemma—giving us the m label—is of vital significance in the classification of tensor operators of L. Of course, our discussion has related only to Λ' "high enough" with respect to Λ; i.e., it applies only when $T^R_{m,jq}$ acts on states $|\Lambda'\,i'\,m'\rangle$ of high enough Λ'. If Λ' is not high enough, the situation of Sec. 4 obtains and for any given allowed set of values R', i', m', R, j, q, i'' there are less than dim R nonvanishing matrix elements,

$$\langle R' + m \quad i'' \quad m'' = m' + q|\, T^R_{m,jq}\, |R'\,i'\,m'.\rangle$$

Information as to how many and which ones these are is contained implicitly and somewhat awkwardly in the Racah–Speiser lemma. This, however, does not concern us here. The point is that it is not that any of the dim R independent tensor components $T^R_{m,jq}$ with fixed R, j, q are identically zero, but simply that some of them annihilate certain states $|R'\,i'\,m'\rangle$ whose highest weights are not high enough.[23] In other words, Biedenharn's lemma features significantly in the classification of the dim R independent tensor operators $T^R_{m,jq}$ with fixed R, j, q, and the Racah–Speiser lemma simply describes edge effects of minor importance regarding the vanishing of certain matrix elements of such operators. It is for this reason that we have not only stated and proved Biedenharn's lemma, which is after all no more than a special case of the Racah–Speiser lemma, but also stressed it over and above the general discussion. The simplicity and vital significance of the lemma could easily be lost within the complication of the general treatment. It should also be mentioned that the philosophy expounded here is in obvious agreement with that of Biedenharn. Further, it might be convenient to illustrate the argument by means of the case of SU_2. In familiar notation, if $j' \geq j$, then the tensor operator t^j_m has nonvanishing matrix elements

$$\langle j''\ m' + m|\, t^j_m\, |j'm'\rangle$$

for $j'' = j' + \Delta$, $\Delta = -j, -j+1, \cdots, j$, since the Clebsch–Gordan series of SU_2 can be written as

$$D^{j'} \times D^j = D^{j'+j} + D^{j'+j-1} + \cdots + D^{j'-j} = \sum_{\Delta=-j}^{j} D^{j+\Delta},$$

[21] J. P. Antoine, Ann, Soc. Sci. Bruxelles 77, 150 (1963).

[22] R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. 34, 1 (1962). See also C. Fronsdal, *1962 Brandeis Lectures* (W. A. Benjamin, New York, 1963), Vol. I.

[23] This point should be clarified by the example discussed below.

which is an explicit statement of Biedenharn's lemma. Accordingly, we may define, for fixed jm, dim $D^j = (2j+1)$ tensor components $t^j_{\Delta m}$ such that

$$\langle j''m' + m | t^j_{\Delta m} | j'm' \rangle$$

vanishes unless $j'' = j' + \Delta$. The set $t^j_{\Delta m}$, $j = 0, \frac{1}{2}, 1 \cdots$, $-j \leq \Delta, m \leq j$, is a complete set of tensor operators, whose matrix elements yield all Clebsch–Gordan coefficients of SU_2. If $j' < j$, i.e., if $j' = j - p$, $0 < p \leq j$, then

$$D^{j'} \times D^j = D^{j+j'} + \cdots + D^{j-j'} = \sum_{\Delta=-j+2p}^{j} D^{j'+\Delta},$$

with

$$-j < -j + 2p \leq j,$$

and

$$\langle j''m' + m | t^j_{\Delta m} | j'm' \rangle$$

vanishes for $j'' = j' + \Delta$ with $-j \leq \Delta < -j + 2p$. Clearly this does not imply the identical vanishing of any component $t^j_{\Delta m}$ and does not carry any relevance to the tensor operator classification question. Tensors of the type $t^j_{\Delta m}$ occur in Schwinger's theory of angular momentum[24] and will be discussed in a forthcoming paper by one of the authors.[25]

2. BIEDENHARN'S LEMMA FOR AN ARBITRARY COMPACT SIMPLE LIE GROUP

In this section we establish Biedenharn's lemma for an arbitrary compact simple Lie group L and discuss its geometrical interpretation.

We precede a statement of Biedenharn's lemma with a convenient definition. Let R and R' be the IR's of L with highest weights Λ and Λ'. We say that R' dominates R if $(\Lambda' + m)$ is a dominant weight of L for each weight m of L.

Biedenharn's lemma: Let R and R' denote two IR's of a compact simple Lie group L. Let Λ and Λ' be the highest weights of R and R', and let m of multiplicity γ_m denote the various weights of R. If R' dominates R, then in the reduction of the direct product representation $R \times R'$ of L:

(a) only those IR's of L with highest weight $\Lambda' + m$ occur,

(b) the IR of highest weight $\Lambda' + m$ occurs γ_m times.

Proof: We have two distinct but equivalent formulas for the character $\chi^R(\phi)$ of any IR of L. One of these is

$$\chi^R(\phi) = \sum_m \gamma_m \exp[i(m, \phi)], \tag{2.1}$$

which follows directly from the definition of the character.

The other is the Weyl character formula[26]

$$\chi^R(\phi) = \chi^\Lambda(\phi)/x^0(\phi),$$
$$\chi^\Lambda(\phi) = \sum_S \delta_S \exp\{i[S(\Lambda + \delta), \phi]\}. \tag{2.2}$$

Here the superscript zero refers to the identity IR of L whose highest (and only) weight is zero. Also S denotes an element of the Weyl group of L, and $\delta_S = \pm 1$ according to whether an even or odd number of Weyl reflections are needed to obtain S, and δ is half the sum of the positive roots of L. The crucial point of our proof of Biedenharn's lemma is to use (2.1) for R and (2.2) for R' in forming the character of the direct product $R \times R'$. More precisely, from the general theory of characters, we have

$$\begin{aligned}\chi^{R\times R'}(\phi) &= \chi^R(\phi)\chi^{R'}(\phi) \\ &= \sum_m \gamma_m \exp[i(m, \phi)] \\ &\quad \times \sum_S \delta_S \exp\{i[S(\Lambda' + \delta), \phi]\}/x^0(\phi) \\ &= \sum_S \delta_S \sum_m \gamma_m \\ &\quad \times \exp\{i[S(\Lambda' + \delta) + m, \phi]\}/x^0(\phi).\end{aligned} \tag{2.3}$$

But, since the weight diagram, including multiplicities, of any IR of L is invariant under any element S of the Weyl group of L and since the sum $\sum_m$ runs over all distinct weights m of R, we have

$$\sum_m \gamma_m \exp[i(S(\Lambda' + \delta) + m, \phi)] = \sum_m \gamma_m \exp\{i[S(\Lambda' + \delta) + Sm, \phi]\} \tag{2.4}$$

for each S separately. Inserting this into (2.3) and using

$$Sa + Sb = S(a + b) \tag{2.5}$$

we obtain

$$\begin{aligned}&\chi^{R\times R'}(\phi) \\ &= \sum_m \delta_S \sum_S \gamma_m \exp\{i[S(\Lambda' + \delta + m), \phi]\}/x^0(\phi) \\ &= \sum_m \gamma_m \{\sum_S \delta_S \exp\{i[S(\Lambda' + \delta + m), \phi]\}/x^0(\phi)\}.\end{aligned} \tag{2.6}$$

It is at this point that the condition that R' dominates R enters. When the condition is satisfied, we can immediately write

$$\sum_S \delta_S \exp\{i[S(\Lambda' + \delta + m), \phi]\}/x^0(\phi) = \chi^{R(\Lambda'+m)}(\phi), \tag{2.7}$$

[24] J. Schwinger, "On Angular Momentum," NYO-3071 (1951).
[25] A. J. Macfarlane (to be published).
[26] H. Weyl. Z. Math. **24**, 328 (1924), reprinted in H. Weyl, *Selecta* (Birkhauser, Basel, 1956). See also G. Racah, Lecture notes (1951)[2]; B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill Book Company, Inc., New York, 1963), p. 131.

where $R(\Lambda' + m)$ is the IR of L with the highest weight $\Lambda' + m$, for all m. Then (2.6) and (2.7) yield

$$\chi^{R\times R'}(\phi) = \sum_m \gamma_m \chi^{R(\Lambda'+m)}(\phi). \qquad (2.8($$

Hence by the general theory of characters, the reduction of $R \times R'$ contains only those IR's of L with characters $\chi^{R(\Lambda'+m)}$ and these with multiplicities γ_m. This establishes the lemma. Note that the condition that R' dominates R is a necessary as well as sufficient condition for the validity of the lemma, since, if $\Lambda' + m$ is not a dominant weight of L, it cannot be the highest weight of any IR of L and $\chi^{R(\Lambda'+m)}$ is a meaningless expression.

The geometrical interpretation of Biedenharn's lemma in weight space is fairly obvious. Let the point of weight space which corresponds to the highest weight of any irreducible representation R of L be called the site of R. To find which IR's of L are contained in $R \times R'$ and with what multiplicity, we begin by drawing the weight diagram of R, i.e., we plot the weights of R together with an indication of their multiplicities in weight space. We refer to the origin of weight space which may or may not correspond to a weight of R as the center of the weight diagram of R. Now, if we translate the weight diagram of R rigidly and without rotation until its center lies at the site of some representation R' which dominates R, then the weights of the translated weight diagram lie at the sites of the IR's of L contained in $R \times R'$, each IR occurring a number of times equal to the multiplicity of the weight of L which it has at its site after translation. It is to be stressed that this geometrical picture is not identical to that obtained by direct specialization to the case when R' dominates R of the geometrical method of Speiser for reducing $R \times R'$. In fact the picture just described is simpler. It defines the site of any IR of L in weight space to be the point which corresponds to its highest weight, whereas in Speiser's method the site of R is defined to be the point $\Lambda + \delta$ in weight space. Apart from this simplification, which clearly does not affect the result, our picture agrees with that obtained from Speiser's method. Of course, Speiser's definition must be adopted in the general case of $R \times R'$ and our simplification applies only when R' dominates R. However, it is in keeping with our general philosophy (which regards the case in which R' dominates R as being of paramount importance) to find it worthwhile to use the simplified definition of site in this case. Further use of the general definition of site would be one aspect, somewhat innocent, of how the complicated detail of the general case can obscure the simplicity of the important case in which R' dominates R. All this should in no way be construed as criticism of the excellent work of Speiser. It rather reflects our different motivation: Speiser's being to reduce general direct products, ours to proceed toward the Racah algebra of L.

We conclude this section with some simple illustrations. For SU_2, if we refer to IR's by means of their j value, so that the IR j has weights $m = -j, -j+1, \cdots, j$, then Biedenharn's lemma applies to $j \times j'$ with $j' \geq j$, in which case the Clebsch–Gordan series

$$j \times j' = (j' + j) + (j' + j - 1) + \cdots + (j' - j)$$

can be written as

$$j \times j' = \sum_{\Delta=-j}^{j} (j' + \Delta),$$

explicitly illustrating the lemma. For SU_3, reduction of the general product[27] $(\lambda, \mu) \times (\lambda', \mu')$ has been studied by many authors.[28] The conditions that (λ', μ') dominate (λ, μ) are satisfied if and only if $\lambda', \mu' \geq \lambda + \mu$.[29] Since the highest weight of any (α, β) is $[\frac{1}{2}(\alpha + \beta), \frac{1}{3}(\alpha - \beta)]$, the product $(\lambda, \mu) \times (\lambda', \mu')$ with $\lambda', \mu' \geq \lambda + \mu$ can be seen from Biedenharn's lemma to contain (λ'', μ'') with

$$\lambda'' = \lambda' + (M + \tfrac{3}{2}Y),$$
$$\mu'' = \mu' + (M - \tfrac{3}{2}Y), \qquad (2.9)$$

a number of times equal to the multiplicity of the weight (M, Y) of (λ, μ) for each weight of (λ, μ). Mukunda and Pandit have obtained this result by applying tensor methods to the direct product in question, while Preziosi *et al.* have obtained an equivalent one by Young diagram manipulation. While no very convenient formula exists for the multiplicity of (M, Y) in (λ, μ), in any special case this can be readily inferred as follows. The IR (λ, μ) contains those pairs of (I, Y) eigenvalues given by[30]

$$I = \tfrac{1}{2}(f - g),$$
$$Y = f + g - \tfrac{2}{3}(\lambda + 2\mu),$$

for f, g integers with ranges $\lambda + \mu \geq f \geq \mu \geq g \geq 0$, and each pair of eigenvalues (I, Y) is associated with $(2I + 1)$ states $|\lambda\, \mu\, I\, M\, Y)$, with $-I \leq M \leq Y$. For example, the quark IR $(1, 0)$ contains the simple weights $(\frac{1}{2}, \frac{1}{3})$, $(-\frac{1}{2}, \frac{1}{3})$, and $(0, -\frac{2}{3})$, and (2.9) tells us that, as long as $\lambda, \mu \geq 1$, we have $(1, 0) \times (\lambda, \mu) = (\lambda + 1, \mu) + (\lambda, \mu - 1) + (\lambda - 1, \mu + 1)$, a well-known

[27] Notation (λ, μ) for SU_3 IR's is explained, for example, by R. E. Behrends *et al.* in Ref. 22.

[28] See H. Goldberg, Nuovo Cimento **27**, 532 (1963); V. B. Mandelsveig, Zh. Eksperim. i Teor. Fiz. **47**, 1836 (1964) [English transl. Soviet Phys.—JETP **20**, 1237 (1965)]; S. Coleman, J. Math. Phys. **5**, 1343 (1964), as well as the papers of Refs. 9, 17, and 18.

[29] This result has been given by Mukunda and Pandit, Ref. 9, as well as by the authors of the papers cited in Refs. 17–20. See also Sec. 3 of this paper.

[30] A simple proof of this result is given by C. R. Hagen and A. J. Macfarlane, J. Math. Phys. **5**, 1335 (1964).

result.[31] In the case of the octet (1, 1) the simple weights are $(\pm\frac{1}{2}, 1)$, $(\pm\frac{1}{2}, -1)$, and $(\pm 1, 0)$, and there is one double weight (0, 0). From this (2.9) tells us that, as long as $\lambda, \mu \geq 2$, we have[32]

$$\begin{aligned}(1,1) \times (\lambda,\mu) = {} & (\lambda+2,\mu-1)+(\lambda+1,\mu-2) \\ & +(\lambda-1,\mu+2)+(\lambda-2,\mu+1) \\ & +(\lambda+1,\mu+1)+(\lambda-1,\mu-1) \\ & +2(\lambda,\mu).\end{aligned}$$

Kuriyan *et al.*[14] have not only used this result in their tabulation of SU_3 Clebsch–Gordan coefficients for $(\lambda, \mu) \times (1, 1)$, but also have actually sharpened it by placing the two orthogonal sets of coefficients connecting $(\lambda, \mu) \times (1, 1)$ to (λ, μ) in 1:1 correspondence with the $I = 1$ and $I = 0$ states of weight (0, 0) in the IR (1, 1). While the actual correspondence is made in a somewhat *ad hoc* manner, such a correspondence is expected to emerge as a general feature in the study of CG coefficients of SU_3 or indeed *mutatis mutandis* of any other group.

3. EXPLICIT CONDITIONS FOR THE VALIDITY OF BIEDENHARN'S LEMMA

In the preceding section we derived Biedenharn's lemma which relates to a direct product representation $R \times R'$ of a compact simple Lie group L in which R' dominates R. More precisely, the necessary and sufficient condition for Biedenharn's lemma to be valid is that $\Lambda' + m$ must be a dominant weight of L for all weights m of R, where Λ' is the highest weight of R'. It is clearly of interest to express this condition in more explicit form.

We first recall[33] that if L is rank l, any IR of L can be denoted by

$$\{\lambda_1, \cdots, \lambda_l\}$$

and realized as the leading[34] IR in the reduction of the direct product

$$\underbrace{D_1 \times D_1 \times \cdots D_1}_{|\leftarrow \lambda_1 \text{ factors} \rightarrow|} \times \underbrace{D_2 \times D_2 \times \cdots D_2}_{|\leftarrow \lambda_2 \text{ factors} \rightarrow|} \times \cdots \underbrace{D_l \times D_l \times \cdots D_l}_{|\leftarrow \lambda_l \text{ factors} \rightarrow|},$$

where D_k $(k = 1, 2 \cdots l)$ are the fundamental IR's of L. Thus the irreducible representation $R \equiv \{\lambda_1, \cdots, \lambda_l\}$ of L has highest weight

$$\Lambda \equiv \Lambda(\lambda_1, \cdots, \lambda_l) = \sum_{k=1}^{l} \lambda_k \Lambda^{(k)},$$

where $\Lambda^{(k)}$ is the highest weight of D_k, and the highest weight Λ' of $R' \equiv \{\lambda_1', \cdots, \lambda_l'\}$

$$\Lambda' = \sum_{k=1}^{l} \lambda_k' \Lambda^{(k)}.$$

It is our purpose in this section to express the condition that $\Lambda' + m$ be a dominant weight of L for each weight m of R as an explicit statement in terms of λ_k, λ_k', $(k = 1, \cdots, l)$ for all of the classical groups and G_2. Our results are collected into the accompanying table.

As a preliminary we state and prove Lemma A.

Lemma A: Any weight m of the IR of L with highest weight Λ can be written in the form

$$m = \Lambda - \sum_{\alpha} c_\alpha r(\alpha), \tag{3.1}$$

where the $r(\alpha)$ are the positive roots of L and the c_α are nonnegative integers.

Proof: In any IR, the highest weight state $|\Lambda\rangle$ is the only state such that[35]

$$E_\alpha |\Lambda\rangle = 0, \quad \text{for all positive } \alpha.$$

Hence given any state $|m\rangle$, either $m = \Lambda$ or else there is at least one E_α with positive α such that

$$|m + r(\alpha)\rangle = E_\alpha |m\rangle.$$

Similarly either $m + r(\beta) = \Lambda$, or else there is at least one E_β with positive β such that

$$|m + r(\alpha) + r(\beta)\rangle = E_\beta |m + r(\alpha)\rangle = E_\beta E_\alpha |m\rangle.$$

If we proceed in this way, the fact that all IR's are of finite dimension implies that we eventually reach

$$|m + r(\alpha) + r(\beta) + \cdots r(\gamma)\rangle = E_\gamma \cdots E_\beta E_\alpha |m\rangle$$

such that

$$E_\delta |m + r(\alpha) + r(\beta) + \cdots + r(\gamma)\rangle = 0$$

for all E_δ with positive δ. In this case, we have

$$\Lambda = m + r(\alpha) + r(\beta) + \cdots + r(\gamma) \tag{3.2}$$

and since any given $r(\tau)$ can occur on the right c_τ times, $c_\tau = 0, 1, 2, \cdots$, we see that (3.2) is equivalent to the statement (3.1) of the lemma.

We now study the four families of classical groups one at a time.[36]

SU_n $(n = l + 1)$: For this group, the positive roots are the n-component vectors $r(ij)$, $n \geq i > j \geq 1$, with ath components

$$r_a(ij) = \delta_{ia} - \delta_{ja} \tag{3.3}$$

[31] H. A. Jahn and H. Van Wieringer, Proc. Roy. Soc. (London) **A209**, 502 (1951).

[32] D. Lurie and A. J. Macfarlane, J. Math. Phys. **5**, 565 (1964).

[33] See G. Racah's lecture notes, Ref. 2, and R. E. Behrends *et al.*, Ref. 22.

[34] The one which has the highest weight.

[35] Notation E_α etc., explained in G. Racah's lecture notes, Ref. 2. In this proof, kets denote unnormalized states with the indicated weight.

[36] See G. Racah's lecture notes Ref. 2 for most of the background information regarding the roots and weights of compact simple Lie groups used in this section.

and similarly $r(ij)$ with $n \geq j > i \geq 1$ gives the negative roots (i.e., pairs of number $i, j, n \geq i > j \geq 1$ play the role of the single label α of Lemma A). The Weyl group of SU_n consists of all permutations of the components W_i of any n-component vector W.

From the definition of the Weyl group, it follows that $\Lambda' + m$, for any given $m \in R$, is a dominant weight of SU_n if and only if

$$(\Lambda' + m)_1 \geq (\Lambda' + m)_2 \geq \cdots \geq (\Lambda' + m)_n$$

or, equivalently, if and only if

$$\begin{aligned} \Lambda'_1 - \Lambda'_2 &\geq m_2 - m_1, \\ &\cdot \\ &\cdot \\ &\cdot \\ \Lambda'_{n-1} - \Lambda'_n &\geq m_n - m_{n-1}. \end{aligned} \tag{3.4}$$

Hence the conditions that $\Lambda' + m$ be a dominant weight for all $m \in R$ are that $\Lambda'_1 - \Lambda'_2$ exceed the maximum value of $m_2 - m_1$ as m ranges throughout R, and that $\Lambda'_2 - \Lambda'_3$ exceed $\cdots$ etc. But the fact that the Weyl group permutes the components m_i of m implies that $\max (m_i - m_j)$, $i > j$, is independent of i and j, so that (3.4) may be replaced by

$$\Lambda'_1 - \Lambda'_2, \cdots, \Lambda'_{n-1} - \Lambda'_n \geq \max (m_1 - m_n). \tag{3.5}$$

But now the definition (3.3) of the positive roots yields

$$r_1(ij) \geq 0 \geq r_n(ij),$$

i.e.,

$$r_1(ij) - r_n(ij) \geq 0$$

and Lemma A yields

$$\begin{aligned} \Lambda_1 - \Lambda_n &= m_1 - m_n + \sum_\alpha c_\alpha [r_1(\alpha) - r_n(\alpha)] \\ &\geq m_1 - m_n \end{aligned}$$

so that the maximum value of $m_1 - m_n$ is attained for $m = \Lambda$. Hence (3.5) can be given as

$$\Lambda'_1 - \Lambda'_2, \cdots, \Lambda'_{n-1} - \Lambda'_n \geq \Lambda_1 - \Lambda_n. \tag{3.6}$$

These are the required conditions in terms of the highest weights Λ and Λ' of R and R'. To write them in terms of the numbers λ_k, λ'_k $(k = 1, \cdots, l)$, we use the familiar relations[37]

$$\Lambda_1 - \Lambda_2 = \lambda_1, \quad \Lambda_2 - \Lambda_3 = \lambda_2, \cdots, \Lambda_{n-1} - \Lambda_n = \lambda_l$$

and the results displayed in the table emerge directly from (3.6).

O_{2l+1}: For this group, the positive roots are the l-component vectors $r(i)$, $1 \leq i \leq l$, and $r(ij)$, $r'(ij)$, $l \geq i > j \geq 1$, with ath components given by

$$\begin{aligned} r_a(i) &= \delta_{ai}, \\ r_a(ij) &= \delta_{ai} + \delta_{aj}, \\ r'_a(ij) &= \delta_{ai} - \delta_{aj}, \end{aligned} \tag{3.7}$$

and the Weyl groups consists of all permutations of and changes of sign of the components W_i of any l-component vector W. It now follows that $\Lambda' + m$, for any given $m \in R$, is a dominant weight of O_{2l+1} if and only if

$$(\Lambda' + m)_1 \geq (\Lambda' + m)_2 \geq \cdots \geq (\Lambda' + m)_l \geq 0$$

or, equivalently, if and only if

$$\begin{aligned} \Lambda'_1 - \Lambda'_2 &\geq m_2 - m_1, \\ &\cdot \\ &\cdot \\ &\cdot \\ \Lambda'_{l-1} - \Lambda'_l &\geq m_l - m_{l-1}, \\ \Lambda'_l &\geq -m_l. \end{aligned} \tag{3.8}$$

Validity of (3.8) for all $m \in R$ implies its validity when the quantities on the right sides of these inequalities attain their maximum values in R. From the definition of the Weyl group, it is clear that

$$\begin{aligned} \max (m_1 - m_2) &= \max (m_2 - m_3) \\ &= \cdots \max (m_l - m_{l-1}) \\ &= \max (m_i + m_j) = \max (m_1 + m_2) \end{aligned}$$

and

$$\max (-m_l) = \max m_i = \max m_1.$$

Hence, the conditions for validity of (3.8) for all $m \in R$ become

$$\begin{aligned} \Lambda'_1 - \Lambda'_2, \cdots \Lambda'_{l-1} - \Lambda'_l &\geq \max (m_1 + m_2), \\ \Lambda'_l &\geq \max m_1. \end{aligned} \tag{3.9}$$

Now the definition (3.7) of positive roots $r(\alpha)$ yields

$$r_1(\alpha) \geq 0 \quad \text{and} \quad r_1(\alpha) + r_2(\alpha) \geq 0$$

so that, from Lemma A, we get

$$\begin{aligned} \Lambda_1 &= m_1 + \sum_\alpha c_\alpha r_1(\alpha) \geqslant m, \\ \Lambda_1 + \Lambda_2 &= m_1 + m_2 + \sum_\alpha c_\alpha [r_1(\alpha) + r_2(\alpha)] \\ &\geq m_1 + m_2. \end{aligned}$$

Hence (3.9) becomes

$$\Lambda'_1 - \Lambda'_2, \cdots \Lambda'_{l-1} - \Lambda'_l \geq \Lambda_1 + \Lambda_2, \quad \Lambda'_l \geq \Lambda_1. \tag{3.10}$$

To rewrite (3.10) in terms of λ_k, λ'_k $(k = 1, \cdots, l)$ we use the standard relation[37,38]

$$\Lambda_1 - \Lambda_2 = \lambda_1, \cdots, \quad \Lambda_{l-1} - \Lambda_l = \lambda_{l-1}, \quad \Lambda_l = \tfrac{1}{2}\lambda_l \tag{3.11}$$

The results then appear as in the table. The factor $\frac{1}{2}$ in the last part of (3.11) reflects the fact that lth fundamental IR of O_{2l+1} is a spinor representation with highest weight $(\frac{1}{2}, \frac{1}{2} \cdots \frac{1}{2})$.

[37] In this and the corresponding equation below our notation is essentially that of the paper by Dynkin [Am. Math. Soc. Transl. 17, (1950), Table 24]. See also Eq. (94) of Racah's lecture notes, Ref. 2.

[38] This is essentially Eq. (96) of Racah's lecture notes, Ref. 2. We have taken the spinor IR as D_l instead of Racah's D_1. See also H. Boerner, *Representations of Groups* (North-Holland Publishing Company, Amsterdam, 1963), Chap, VII, Sec. 14, and Chap. VIII, 3-5.

Sp_{2l}: In this case the positive roots are the same as for O_{2l+1} except that $n_a(i) = 2\delta_{ai}$. Furthermore the Weyl group of Sp_{2l} is the same as for O_{2l+1}. It therefore follows exactly as for O_{2l+1} that (3.10) also gives the condition that $\Lambda' + m$ is a dominant weight of Sp_{2l} for all $m \in R$. However, the standard relations[37,39]

$$\Lambda_1 - \Lambda_2 = \lambda_1, \cdots, \quad \Lambda_{l-1} - \Lambda_l = \lambda_{l-1}, \quad \Lambda_l = \lambda_l \tag{3.12}$$

for Sp_{2l} differ from (3.11) in the absence of the factor $\frac{1}{2}$ in the last equality, so that the final results for Sp_{2l}, as displayed in the table, differ from those for O_{2l+1}. The reason for the difference between (3.11) and (3.12), of course, stems from the different nature of the lth fundamental IR's in the two cases, that for Sp_{2l} being a tensor representation of highest weight $(1, 1 \cdots 1)$

O_{2l}: For this group, the positive roots are the l-component vectors

$$r(ij), \quad r'(ij), \quad l \geq i > j \geq 1,$$

with ath components

$$\begin{aligned} r_a(ij) &= \delta_{ia} + \delta_{ja}, \\ r_a(ij) &= \delta_{ia} - \delta_{ja}. \end{aligned} \tag{3.13}$$

The Weyl group consists of all permutations and all changes of sign in pairs of the components W_i of an l-component vector W. It follows that $\Lambda' + m$, for any given $m \in R$, is a dominant weight of O_{2l} if and only if

$$\begin{aligned} (\Lambda' + m)_1 &\geq (\Lambda' + m)_2 \\ &\geq \cdots (\Lambda' + m)_{l-1} \geq |(\Lambda' + m)_l|, \end{aligned}$$

or, equivalently, if and only if

$$\begin{aligned} \Lambda'_1 - \Lambda'_2 &\geq m_2 - m_1, \\ &\cdots \\ \Lambda'_{l-1} - \Lambda'_l &\geq m_l - m_{l-1}, \\ \Lambda'_{l-1} + \Lambda'_l &\geq -m_l - m_{l-1}. \end{aligned} \tag{3.14}$$

The departure from pattern in the last inequality is to be noted. As before, from the properties of the Weyl group we deduce that, for (3.14) to hold for all $m \in R$, we must have

$$\begin{aligned} \Lambda'_1 - \Lambda'_2, \cdots \Lambda'_{l-1} - \Lambda'_l &\geq \max (m_1 + m_2), \\ \Lambda'_{l-1} + \Lambda'_l &\geq \max (m_1 + m_2). \end{aligned} \tag{3.15}$$

Again as before, we can use (3.13) and Lemma A to show that $\max (m_1 + m_2)$ is attained for $m = M$.

[39] L. O'Raifeartaigh, "Lectures on Local Lie Groups and Their Representations," Matscience Report 25 (1964).

TABLE I. Conditions for compact simple Lie groups L of rank l that the IR $(\lambda'_1, \lambda'_2, \cdots, \lambda'_l)$ dominate the IR $(\lambda_1, \lambda_2, \cdots, \lambda_l)$.

L	Conditions
$SU(n)$ $(n = l + 1)$	$\lambda'_1, \lambda'_2, \cdots \lambda'_l \geq (\lambda_1 + \lambda_2 + \cdots + \lambda_l)$
$O(2l + 1)$	$\lambda'_1, \lambda'_2, \lambda'_{l-1} \geq \lambda_1 + 2(\lambda_2 + \cdots + \lambda_{l-1}) + \lambda_l$, $\lambda'_l \geq 2(\lambda_1 + \cdots + \lambda_{l-1}) + \lambda_l$
$Sp(2l)$	$\lambda'_1, \lambda'_2, \lambda'_{l-1} \geq \lambda_1 + 2(\lambda_2 + \cdots + \lambda_l)$, $\lambda'_l \geq \lambda_1 + \lambda_2 + \cdots + \lambda_l$
$O(2l)$	$\lambda'_1, \lambda'_2, \cdots \lambda'_l$ $\geq \lambda_1 + 2(\lambda_2 + \cdots \lambda_{l-2}) + \lambda_{l-1} + \lambda_l$
G_2	$\lambda'_1 \geq 2\lambda_1 + 3\lambda_2$ $\lambda'_2 \geq \lambda_1 + 2\lambda_2$

Hence (3.15) reads as

$$\begin{aligned} \Lambda'_1 - \Lambda'_2, \cdots \Lambda'_{l-1} - \Lambda'_l &\geq \Lambda_1 + \Lambda_2, \\ \Lambda'_{l-1} + \Lambda'_l &\geq \Lambda_1 + \Lambda_2. \end{aligned} \tag{3.16}$$

These inequalities are translated into terms of λ_k, λ'_k $(k = 1, \cdots, l)$ by means of[37,40]

$$\begin{aligned} \Lambda_1 - \Lambda_2 = \lambda_1, \cdots, \Lambda_{l-2} - \Lambda_{l-1} = \lambda_{l-2}, \\ \Lambda_{l-1} - \Lambda_l = \lambda_l, \quad \Lambda_{l-1} + \Lambda_l = \lambda_{l-1}, \end{aligned} \tag{3.17}$$

the final results being given in Table I. The inverted order of λ_{l-1} and λ_l in the second line of (3.17) stems from the fact that the fundamental IR's D^{l-1} and D^l of O_{2l} are spinor representations, respectively, with highest weights $(\frac{1}{2}, \frac{1}{2} \cdots \frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{2} \cdots \frac{1}{2}, -\frac{1}{2})$.

G_2: G_2 is a rank two group with positive roots[41]

$$\left(\frac{1}{2\sqrt{3}}, 0\right), \quad \left(\frac{\sqrt{3}}{4}, \pm\frac{1}{4}\right), \quad \left(\frac{1}{4\sqrt{3}}, \pm\frac{1}{4}\right), \quad \left(0, \frac{1}{2}\right). \tag{3.18}$$

The Weyl group is conveniently specified by saying that, if $W = (W_1, W_2)$ is any two component vector, the vectors equivalent to it are

$$\begin{aligned} &(\pm W_1, \pm W_2), \\ &[\pm\tfrac{1}{2}(W_1 + \sqrt{3}W_2), \pm\tfrac{1}{2}(\sqrt{3}W_1 - W_2)], \\ &[\pm\tfrac{1}{2}(W_1 - \sqrt{3}W_2), \pm\tfrac{1}{2}(\sqrt{3}W_1 + W_2)], \end{aligned} \tag{3.19}$$

where all possible combinations of sign are to be taken. The vector (W_1, W_2) can be seen from (3.19) to be dominant if and only if

$$W_1 \geq \sqrt{3}W_2 \geq 0$$

so that $(\Lambda' + m)$, for fixed $m \in R$, is a dominant weight of G_2 if and only if

$$\Lambda'_1 - \sqrt{3}\,\Lambda'_2 \geq \sqrt{3}\,m_2 - m_1, \quad \Lambda'_2 \geq -m_2. \tag{3.20}$$

[40] See H. Boerner, Ref. 38.

[41] For information regarding G_2, see R. E. Behrends *et al.*, Ref. 22.

As before, we find that, as m ranges through R

$$\max(\sqrt{3}m_2 - m_1) = \max m_1,$$
$$\max(-m_2) = \max \tfrac{1}{2}(\sqrt{3}m_1 + m_2), \quad (3.21)$$

and use (3.18) and Lemma A to prove the maxima on the right are attained for $m = \Lambda$. Hence (3.20) becomes

$$\Lambda_1' - \sqrt{3}\,\Lambda_2' \geq \Lambda_1,$$
$$\Lambda_2' \geq \tfrac{1}{2}(\sqrt{3}\,\Lambda_1 + \Lambda_2). \quad (3.22)$$

This translates into the results displayed in the table when one uses

$$\Lambda_1 = (1/2\sqrt{3})(\lambda_1 + \tfrac{3}{2}\lambda_2),$$
$$\Lambda_2 = \tfrac{1}{4}\lambda_2, \quad (3.23)$$

which follow from the fact the seven-dimensional fundamental IR D_1 of G_2 has highest weight $(1/2\sqrt{3})(1, 0)$, while the fourteen-dimensional fundamental IR D_2 has highest weight $\frac{1}{4}(3, 1)$.

4. THE RACAH–SPEISER LEMMA

In Sec. 2 we stated and proved Biedenharn's lemma relating to a direct product representation $R \times R'$ of a compact simple Lie group L when R' dominates R. In this section we turn to the case of the general product $R \times R'$, and prove that this case can be described by the following lemma, which, for reasons given in the Introduction, we call the Racah–Speiser lemma.

Racah–Speiser lemma: Let L, R, R', Λ, Λ', m and γ_m be defined as in our statement of Biedenharn's lemma. Let $\Lambda' + m_d$ denote those weights of the set of weights $\Lambda' + m$ which are dominant weights of L. In the reduction of the general product representation $R \times R'$ of L

(a) only those IR's of L of highest weight $\Lambda' + m_d$ can occur, and

(b) each of these occurs with multiplicity

$$\Gamma(m_d) = \sum \gamma_m(-1)^{n_m}, \quad (4.1)$$

where the summation $\sum$ extends over all those weights m (including m_d itself) for which a succession of Weyl reflections of $\Lambda' + m + \delta$ yields $\Lambda' + m_d + \delta$, and where n_m is the number of reflexions required. $\Gamma(m_d)$ may turn out to be zero, but is never negative.

We precede the proof with two lemmas, which we call Lemmas B and C.

Lemma B: Let m be a weight of R and m' a weight of R'. Let S_α be the Weyl reflection in the hyperplane perpendicular to the root $r(\alpha)$ of L. If S_α is such that

$$S_\alpha m' \leq m',$$
$$S_\alpha(m' + m) > m' + m, \quad (4.2)$$

then there exists in R a weight $m(\alpha)$ such that

$$m(\alpha) \geq m,$$
$$S_\alpha(m' + m) = m' + r(\alpha) + m(\alpha). \quad (4.3)$$

Proof of Lemma B: From (4.2), there exists in R' and R, respectively, the strings of weights[42]

$$m', m' - r(\alpha), \cdots m' - qr(\alpha) = S_\alpha m',$$
$$m, m + r(\alpha), \cdots m + pr(\alpha) = S_\alpha m, \quad (4.4)$$

with

$$p > q \geq 0. \quad (4.5)$$

Hence, we have

$$S_\alpha(m' + m) = S_\alpha m' + S_\alpha m$$
$$= m' + r(\alpha) + m + (p + q - 1)r(\alpha),$$

and from (4.5) it follows that $m(\alpha)$ given by

$$m(\alpha) = m + (p - q - 1)r(\alpha) \quad (4.6)$$

is a weight of R which satisfies (4.3).

In Lemma C, we need to introduce the primitive roots[43] of L. If L is of rank l, its primitive roots are a set of l positive roots $r(k)$ $(k = 1, 2, \cdots, l)$ such that each positive root $r(\alpha)$ can be written in the form

$$r(\alpha) = \sum_{k=1}^{l} c_k(\alpha) r(k), \quad (4.7)$$

where the $c_k(\alpha)$ are nonnegative integers. We refer to the Weyl reflection S_k in the hyperplane orthogonal to r_k as a primitive Weyl reflection. We now state and prove Lemma C.

Lemma C: If any weight vector W of L satisfies

$$S_k W \leq W \quad (4.8)$$

for all $k = 1, 2 \cdots l$, then W is a dominant weight of L.

Proof of Lemma C: We have

$$S_k W = W - \frac{2W \cdot r(k)}{r(k) \cdot r(k)} r(k), \quad k = 1, 2, \cdots, l, \quad (4.9)$$

whence, using (4.8) and the positive nature of the $r(k)$, we get

$$W \cdot r(k) \geq 0, \quad k = 1, 2, \cdots, l.$$

But then from (4.7) it follows that $W \cdot r(\alpha) \geq 0$ for all α, so that $S_\alpha W \leq W$ for all α; i.e., W is a dominant weight.

Proof of the Racah–Speiser lemma: We recall that in proving Biedenharn's lemma we reached (2.6) without using the assumption that R' dominates R. Hence Eq. (2.6) is valid even when the assumption is dropped; it is used as a starting point in the present discussion.

[42] See G. Racah's lecture notes, Ref. 2, Lecture 2, Sec. 2, and Lectures 3 and 4, Sec. 1.

[43] For properties of primitive roots used here, see B. Gruber and L. O'Raifeartaigh, Ref. 2.

Suppose that for some $m \in R$, $\Lambda' + m$ is not a dominant weight. Then by Lemma C, there exists at least one primitive Weyl reflection S_k such that

$$S_k(\Lambda' + m) > (\Lambda' + m). \tag{4.10}$$

On the other hand, Λ' is a dominant weight so that

$$S_k\Lambda' \leq \Lambda. \tag{4.11}$$

However, (4.10) and (4.11) are just the conditions for the validity of Lemma B, and it follows there exists in R a weight $m(k)$ such that

$$m(k) \geq m;\ S_k(\Lambda' + m) = \Lambda' + m(k) + r(k). \tag{4.12}$$

We now use the fact that S_k is a primitive reflection to write[43]

$$S_k\delta = \delta - r(k), \tag{4.13}$$

where δ as before is half the sum of the positive roots of L.

Now (4.12) and (4.13) may be combined to give for the quantity $(\Lambda' + m + \delta)$ occurring in (2.6) the result

$$S_k(\Lambda' + m + \delta) = \Lambda' + m(k) + \delta \geq \Lambda' + m + \delta. \tag{4.14}$$

In other words S_k reflects $\Lambda' + m + \delta$ into a vector of the form $\Lambda' + m' + \delta$ at least as positive as $\Lambda' + m + \delta$.

Let us rename k as k_1. If $\Lambda' + m(k_1)$ is neither equal to $\Lambda' + m$ nor dominant, we can repeat the process and reflect $\Lambda' + m(k_1) + \delta$ onto $\Lambda' + m(k_2) + \delta$, where $m(k_2) \geq m(k_1)$. It is easy to see that by repeating the process a sufficient number of times, we eventually reach a situation wherein one of two possibilities obtains

(a) $S_{k_p}(\Lambda' + m(k_{p-1}) + \delta) = \Lambda' + m(k_{p-1}) + \delta$,

(b) $\Lambda' + m(k_p)$ is a dominant weight of Λ.

We examine cases (a) and (b) separately.

Case (a): In this case $\Lambda' + m(k_{p-1}) + \delta$ lies in the Weyl hyperplane orthogonal to $r(k_p)$. On the other hand, it is obtained from $\Lambda' + m(k_{p-2}) + \delta$ by means of the Weyl reflection in the hyperplane orthogonal to $r(k_{p-1})$. Hence $\Lambda + m(k_{p-2}) + \delta$ lies in the hyperplane obtained from the Weyl hyperplane orthogonal to $r(k_p)$ by reflection in the Weyl hyperplane orthogonal to $r(k_{p-1})$. But the Weyl reflection of a Weyl hyperplane is itself a Weyl hyperplane. Hence $\Lambda' + m(k_{p-2}) + \delta$ lies in a Weyl hyperplane, and we prove by repeating the argument that the original $\Lambda' + m + \delta$ does also.

Suppose $\Lambda' + m + \delta$ belongs to the Weyl hyperplane orthogonal to $r(j)$, say. Then, using the group property of Weyl reflections we get

$$\begin{aligned}\sum_S \delta_S \exp\{i[S(\Lambda' + m + \delta), \phi]\} &= \sum_S \delta_{SS_j} \exp\{i[SS_j(\Lambda' + m + \delta), \phi]\} \\ &= \sum_S (-\delta_S) \exp\{i[S(\Lambda' + m + \delta), \phi]\} \\ &= 0.\end{aligned}$$

Hence any term $m \in R$ such that $\Lambda' + m + \delta$ lies in a Weyl hyperplane makes zero contribution to, and can hence be omitted from the summation over m in (2.6).

Case (b): In this case, we have

$$\begin{aligned}\sum_S \delta_S \exp\{i[S(\Lambda' + m + \delta), \phi]\} &= \sum_S \delta_{SS_{k_1}\cdots S_{k_p}} \exp\{i[SS_{k_1}\cdots S_{k_p} \\ &\qquad \times (\Lambda' + m + \delta), \phi]\} \\ &= (-)^p \sum_S \delta_S \exp\{i[S(\Lambda' + m(k_n) + \delta), \phi]\}.\end{aligned} \tag{4.15}$$

Now, since $\Lambda' + m(k_p)$ is a dominant weight, we can, by (2.2) write the contribution to $\chi^{R\times R'}$ from the weight m of R from which we set out as

$$(-)^p \chi^{[R\Lambda' + m(k_p)]}(\phi). \tag{4.16}$$

Substitution of (4.16) into (2.6) gives rise now to the Racah–Speiser lemma. Note, however, that each $m(k_p)$ is to be identified with some one of the m_d of the above statement of the lemma.

Also the numbers of primitive Weyl reflections is equal modulo two to the number of Weyl reflections of any kind.

Some comments regarding the nature and application of the Racah–Speiser lemma are now given. First, we should emphasize that while IR's that can occur in $R \times R'$ are determined by the dominance or non-dominance of the $\Lambda' + m$, the multiplicity of their occurrence, which can be zero, is determined by the Weyl reflection properties $\Lambda' + m + \delta$. Second, we note that in forming the sequence of weights $\Lambda' + m + \delta$, $\Lambda' + m(k_i) + \delta$, $i = 1 \cdots p$, it is not necessary to check at each step whether $\Lambda' + m(k_i)$ is dominant, but only whether $\Lambda' + m(k_i) + \delta$ is dominant. This follows because, after the m's of case (a) have been dropped, dominance of $\Lambda' + m(k_i)$ implies and is implied by dominance of $\Lambda' + m(k_i) + \delta$. This observation affords considerable simplification of practical calculation of $\Gamma(m_d)$. Finally, we turn to the question of geometrical significance.

The Racah–Speiser lemma in fact affords an explicit proof of the geometrical rules given by

Speiser[3,10] for the reduction of $R \times R'$. A very clear statement of these rules can be obtained by paralleling the discussion given by deSwart[14] of the application of Speiser's method to SU_3. For SU_3 deSwart starts with a suitable preparation of the weight space of SU_3 for the performance of the geometrical operations associated with Speiser's method. We describe this preparation as follows.

The highest weight points of the IR's of the rank two group SU_3 lie on the boundaries of or within that region of SU_3 weight space, which is bounded by the two lines (Weyl hyperplanes) perpendicular to the primitive roots of SU_3 and which contains the vector δ which is half the sum of the positive roots of SU_3. Call this region the fundamental domain D_0 of SU_3 weight space. Define the site[44] of the irreducible representation R of SU_3 of highest weight Λ to be the point $\Lambda + \delta$ of weight space. The sites of all IR's of SU_3 lie strictly inside D_0. We prepare D_0 by attaching to the site of each R the label $+R$. Any other domain D of SU_3 weight space is bounded by lines (Weyl hyperplanes) perpendicular to some pair of positive roots of SU_3. Its points are equivalent under some number k of Weyl reflections to the points of D_0. We prepare the interior of D by attaching the label $(-)^k R$ to the points of D equivalent to the site of R in D_0. Preparation of weight space is completed by attaching the label 0 to any allowed weight point of SU_3 lying on a Weyl reflection axis. Now to reduce $R \times R'$, we

(a) construct the weight diagram of R', which involves not only specification of the m but also the γ_m, and

(b) translate it rigidly without rotation so that its center moves from the origin of weight space to the site of R' in D_0. Then, if m is a weight of R of multiplicity γ_m, m now lies at a point of weight space labeled by $(-)^k R''$ or 0, and corresponding to this m there is a contribution $(-)^k \gamma_m R''$ or 0 to the reduction of $R \times R'$. Adding the contribution from each m of R leads to the reduction of $R \times R'$. This statement has been written so that its generalization to arbitrary compact simple L of rank l is immediate and, this being done, it is surely clear that it corresponds exactly to the Racah–Speiser lemma. We refrain from exhibiting examples due to their availability in the papers of deSwart[14] and Speiser.[3,10] Finally, we are indebted to Dr. C. Anderson for pointing out that an equivalent, but in practice somewhat simpler, method of carrying out the construction of this paragraph is to identify the site R with Λ (instead of $\Lambda + \delta$) and then to carry out the reflections in a set of Weyl planes intersecting at $-\delta$ (instead of 0).

We conclude with a final reminder of our idea of the relative importance of the Racah–Speiser lemma and Biedenharn's lemma. We regard the former not so much as a proof of Speiser's rules, but rather as a means of making these rules understandable as edge effects when the simple picture provided by the latter breaks down. Indeed, one hardly needs to know the rules explicitly for the development of the Racah algebra of L. One needs Biedenharn's lemma as a vital structural ingredient of the Racah algebra and one uses it as one computes the CG coefficients of L for $R \times R'$ with R' tacitly assumed to dominate R. All that happens if R' does not dominate R is that certain sets of the CG coefficients so constructed automatically drop out—one does not need to get rid of them in advance. Thus, the role of the Racah–Speiser lemma is that it gives a simple explanation of *why* these sets of CG coefficients drop out in the actual calculation.[45]

ACKNOWLEDGMENTS

The authors wish to thank Dr. M. Moshinsky, Dr. B. Vitale, Dr. J. Nagel, and Dr. T. A. Brody for many stimulating discussions.

[44] Note this is not the same definition of the site of an IR of L, which we use in Sec. 2.

[45] The reasonability of this remark can be appreciated by referring to the paper of Kuriyan *et al.*,[14] in particular to Eq. (1.2) and Table I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Chapman–Enskog–Hilbert Expansion for a Class of Solutions of the Telegraph Equation

H. P. McKean, Jr.
Rockefeller University, New York, New York

(Received 2 September 1966)

The Chapman–Enskog–Hilbert expansion is a method for describing a gas in the "hydrodynamical stage," beginning from the Boltzmann equation. The present paper is devoted to the analog of this expansion for the problem $\partial p/\partial t + e\partial p/\partial x = p(-e) - p(+e)$, where $e = \pm 1$ and $x \in R^1$. Though the situation is vastly simpler than in the Boltzmann case, new and amusing mathematical phenomena are encountered. One studies solutions of $\partial p/\partial t + e\partial p/\partial x = \epsilon^{-1}[p(-e) - p(+e)]$ which are (formal) power series in ϵ (Hilbert solutions): such a solution solves $\partial p/\partial t = \epsilon^{-1}[(1 + \epsilon^2\partial^2/\partial x^2)^{\frac{1}{2}} - 1]p$ (hydrodynamical equation) and is completely determined by the initial value of $p(-e) + p(+e)$ (Hilbert paradox). Also, every solution of the original problem comes very rapidly close to a Hilbert solution which is actually convergent (hydrodynamical stage).

1. INTRODUCTION

CONSIDER a dilute gas of molecules of mass 1, filling the whole of 3-dimensional space R^3, subject to an external field $\mathfrak{f}$. The corresponding molecular distribution function $p = p(t, x, v)$ $[t \geq 0, x \in R^3, v \in R^3]$ is a solution of Boltzmann's problem:

$$\frac{\partial p}{\partial t} + v\frac{\partial p}{\partial x} + \mathfrak{f}\frac{\partial p}{\partial v} = B[p \otimes p], \qquad (1.1)$$

in which B stands for special quadratic functional of *p as a function of $v \in R^3$ only*.[1]

Chapman–Enskog–Hilbert's development of p describes the so-called *hydrodynamical stage*. Hilbert's recipe[2] is, *first*, to expand p as a *formal power series* $p_0 + \epsilon p_1 + \epsilon^2 p_2 + \cdots$, which is required to solve

$$\frac{\partial p}{\partial t} + v\frac{\partial p}{\partial x} + \mathfrak{f}\frac{\partial p}{\partial v} = \frac{1}{\epsilon}B[p \otimes p], \qquad (1.2)$$

and, *second*, to put $\epsilon = 1$, hoping for convergence to an actual solution of (1.1). This is clearly a very radical thing to do. Equation (1.2) means that

$$\left[\frac{\partial}{\partial t} + v\frac{\partial}{\partial x} + \mathfrak{f}\frac{\partial}{\partial v}\right]p_{n-1} = \sum_{i+j=n} B[p_i \otimes p_j],$$
$$n \geq 0, \quad p_{-1} \equiv 0, \qquad (1.3a)$$

especially, for $n = 0$, $B[p_0 \otimes p_0] = 0$, and this turns out to be the same as to say that p_0 is a (local) Maxwellian function $[c_1 \exp(-c_2|v - u|^2)]$ depending upon 5 unknown functions of $(t, x) \in [0, \infty) \times R^3$: the 5 hydrodynamical moments

$$\int v^n p_0\, dv \quad (n = 0, 1, 2).$$

At the stage $n \geq 1$, it is required to solve

$$\left[\frac{\partial}{\partial t} + v\frac{\partial}{\partial x} + \mathfrak{f}\frac{\partial}{\partial v}\right]p_{n-1} - \sum_{1 \leq i < n} B[p_i \otimes p_{n-i}] = C[p_n] \qquad (1.3b)$$

for p_n with $C[f] \equiv B[f \otimes p_0] + B[p_0 \otimes f]$. C turns out to be a nice integral operator with null space comprising the 5 functions v^n $(n = 0, 1, 2)$, and to continue the recipe, it is necessary to meet the terms of the Fredholm alternative by making the left side of (1.3b) perpendicular to this null space. Doing this for $n = 1$ gives the Eulerian hydrodynamical equations for the 5 hydrodynamical moments of p_0; for $n = 2$, it gives the Navier–Stokes equations for the hydrodynamical moments of $p_0 + p_1$; etc. The curious thing about this expansion is that *the formal power series for p is completely determined by the initial values of its 5 hydrodynamical moments* $\int v^n p\, dv$ $(n = 0, 1, 2)$.[3] Ford and Uhlenbeck[4] call this *the Hilbert paradox*. This seemingly accidental feature of the recipe is highly satisfactory, as it substantiates, in part, the following possibly over-optimistic diagram (Fig. 1) of what is going on. For *any* solution p of Boltzmann's problem with initial data f, there is (or should be) a Hilbert solution p^0 with initial data f^0 which is *closest* to p, i.e., for which $p - p^0$ becomes small most rapidly. f^0 should be a projection of f, and this projection should commute with the Boltzmann streaming so that p^0 is the *same* projection of p.

[1] G. Ford and G. E. Uhlenbeck, *Lectures on Statistical Mechanics* (American Mathematical Society, Providence, Rhode Island, 1963), p. 77.

[2] D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen* (B. G. Teubner, Leipzig, 1912), p. 270. See also Ref. 1, p. 108.

[3] D. Hilbert, Ref. 2, p. 280.

[4] Reference 1, p. 110.

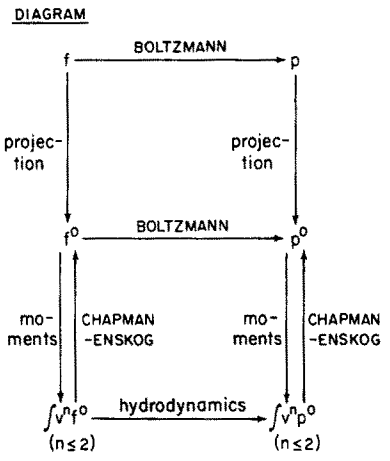


FIG. 1. A schematic view of the connection between hydrodynamics and Boltzmann's problem.

Hilbert's paradox implies the existence of a (very complicated) functional expressing f^0 by means of its 5 hydrodynamical moments [Chapman–Enskog development]. This functional should also commute with the Boltzmann streaming so that p^0 stands in the *same* relationship to *its* hydrodynamical moments. Hilbert's paradox also implies the existence of a self-contained parabolic problem governing the hydrodynamical state $\int v^n p^0$ $(n \leq 2)$.

Except for the existence of formal power series solutions of (1.2) exhibiting the Hilbert paradox and the (formal) computation of the Chapman–Enskog development, we think it is fair to say that all this is up in the air and very difficult to verify.

The purpose of this paper is to verify this picture in a very simple case, replacing (1.1) by

$$\frac{\partial p}{\partial t} + e\frac{\partial p}{\partial x} = D[p] \equiv p(-e) - p(+e), \quad (1.4a)$$

$$\lim_{t \downarrow 0} p = f, \quad (1.4b)$$

with $t \geq 0$, $x \in R^1$, $e = \pm 1$, and nice data f; this is the same as the telegraph equation

$$\frac{\partial^2}{\partial t^2} p + 2\frac{\partial}{\partial t} p = \frac{\partial^2}{\partial x^2} p \quad (1.5a)$$

with initial data (1.4b) and

$$\lim_{t \downarrow} \frac{\partial p}{\partial t} = f' \equiv -ef' + D[f]. \quad (1.5b)$$

Kaç[5] found a nice probabilistic model for (1.4): if $\#$ is a Poisson process with rate 1 and jumps $+1$ and if $z(0) = [x, e]$ is distributed according to $f = f(x, e)$ with

$$f \geq 0, \quad \int_{R^1} \sum_{e=\pm 1} f = 1,$$

then $z(t) = [x + e\int_0^t (-1)^{\#}\, ds, e(-1)^{\#}]$ is distributed according to p, especially, $p \geq 0$ if $f \geq 0$ [see (2.2) in Sec. 2].

Hilbert's recipe for (1.4) would be to find a formal power series solution $p = p_0 + \epsilon p_1 + \epsilon^2 p_2 + \cdots$ of

$$\frac{\partial p}{\partial t} + e\frac{\partial p}{\partial x} = \frac{1}{\epsilon} D[p], \quad (1.6a)$$

$$\lim_{t \downarrow 0} p = \text{a formal power series } f, \quad (1.6b)$$

i.e.,

$$\left[\frac{\partial}{\partial t} + e\frac{\partial}{\partial x}\right] p_{n-1} = D[p_n], \quad n \geq 0, \quad p_{-1} \equiv 0, \quad (1.7a)$$

$$\lim_{t \downarrow 0} p_n = f_n, \quad n \geq 0. \quad (1.7b)$$

For $n = 0$, (1.7a) states that $D[p_0] = 0$. This means that p_0 is an *even* function of e, and the Fredholm condition for solving $p_0^{\cdot} + ep_0' = D[p_1]$ is that the left side should be an *odd* function of e. Thus, $p_0^{\cdot} = 0$ [Eulerian equation], and $p_0 \equiv f_0$. $ep_0' = D[p_1]$ is now solved for the odd part of $p_1[(p_1)_{\text{odd}} = -\frac{1}{2}ef_0']$, and the Fredholm condition for solving $p_1^{\cdot} + ep_1' = D[p_2]$, i.e., $p_1^{\cdot} + ep_1'$ odd, permits us to compute the even part of $p_1[(p_1)_{\text{even}} = (f_1)_{\text{even}} + \frac{1}{2}tf_0'']$, etc.

The actual facts can be expressed much more compactly and elegantly: *a formal power series* $f = \sum f_n \epsilon^n$ *with coefficients from*[6] $C^\infty(R^1 \times Z)$ *gives rise to a formal power series solution* p *of (1.6) if and only if*[7]

$$-ef_{\text{odd}} = \frac{(1 + \epsilon^2\partial^2)^{\frac{1}{2}} - 1}{\epsilon\partial} f_{\text{even}} \equiv \sum_{m=1}^{\infty} \binom{\frac{1}{2}}{m} (\epsilon\partial)^{2m-1} f_{\text{even}}, \quad (1.8)$$

in which case[6]

$$\frac{\partial p}{\partial t} = \epsilon^{-1}[(1 + \epsilon^2\partial^2)^{\frac{1}{2}} - 1]p \equiv \sum_{m=1}^{\infty} \binom{\frac{1}{2}}{m} \epsilon^{2m-1}\partial^{2m} p \quad (1.9a)$$

and

$$-ep_{\text{odd}} = \frac{(1 + \epsilon^2\partial^2)^{\frac{1}{2}} - 1}{\epsilon\partial} p_{\text{even}} \quad t \geq 0, \quad (1.9b)$$

especially, the odd and even parts of p *propagate separately, and* p *is completely specified by the data* f_{even} *(Hilbert paradox).* Equation (1.9a) plays the role of the hydrodynamical equations and (1.9b) the role of the Chapman–Enskog development.

Now, suppose $f_{\text{even}} = f_0$.[8] *Then the formal power series* $-ef_{\text{odd}} = (1.8)$ *converges in* $C^\infty(R^1 \times Z)$ *for* $|\epsilon| < 1$ *if and only if* f_0 *is an integral function of*

[5] M. Kaç, *Some Stochastic Problems in Physics and Mathematics* (Magnolia Petroleum Company, Dallas, Texas, 1956).

[6] Z stands for the 2-point space $e = \pm 1$.

[7] f_{odd} [f_{even}] always denotes the odd [even] part of f *as a function of* e. ∂ stands for $\partial/\partial x$.

[8] By Eq. (1.8), f_0 is automatically even.

exponential type ≤ 1, *in which case* p *converges to an integral function of the same type for any* $t \geq 0$. *Under the additional condition* $\|f_{\text{even}}\|_1 < \infty$, p *is continuous for* $|\epsilon| \leq 1$ *and is a bona fide solution of* (1.4) *at* $\epsilon = 1$. *Any solution* p *of* (1.4) *with data*[9] $f \in C^\infty_\downarrow(R^1 \times Z)$ *rapidly comes close to a Hilbert solution* p^0. *The correct recipe for the data of* p^0 *is*

$$f^0 = \int_{-1}^{+1} e^{i\gamma x} \frac{1}{2}\left[1 + \frac{1 + D - i\gamma e}{(1-\gamma^2)^{\frac{1}{2}}}\right] \hat{f}\, d\gamma \quad (1.10a)$$

with

$$\hat{f} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\gamma x} f\, dx. \quad (1.10b)$$

This map is a projection onto the class of functions satisfying (1.8) *for* $\epsilon = 1$, *and the corresponding Hilbert solution*

$$p^0 = \int_{-1}^{+1} \exp(i\gamma x) \exp t[(1-\gamma^2)^{\frac{1}{2}} - 1] \hat{f}^0\, d\gamma \quad (1.11)$$

differs from p *by terms of magnitude* e^{-t}, *roughly. Unfortunately, the projection* $f \to f^0$ *does not preserve positivity, though nonnegative summable Hilbert solutions do exist.*

Carleman[10] proposed a less artificial caricature of (1.1)[11]:

$$\frac{\partial p}{\partial t} + e\frac{\partial p}{\partial x} = D[p^2]. \quad (1.12)$$

A nice description of its formal power series solutions would bring us quite close to a satisfactory picture of the Chapman–Enskog–Hilbert development for the actual Boltzmann problem, but we were unable to form any simple picture of them.

Grad[12] studied the Chapman–Enskog–Hilbert development for the linearized Boltzmann problem:

$$\frac{\partial p}{\partial t} + v\frac{\partial p}{\partial x} + \mathfrak{f}\frac{\partial p}{\partial v} = C[p] \quad (1.13)$$

and also for a problem isomorphic to (1.4), verifying the presence of a Hilbert paradox. Grad's statement is not very explicit, especially the analog of (1.9) is not proved, but he did find that actual solutions must be functions of exponential type.

2. CLASSICAL SOLUTIONS

Because the solution of $p^\cdot + ep' = \epsilon^{-1}D[p]$ $(\epsilon > 0)$ coincides with the solution of $p^\cdot + ep' = D[p]$ at $(t/\epsilon, x/\epsilon, e)$, it suffices to discuss the case $\epsilon = 1$:

$$p^\cdot(+1) + p'(+1) = p(-1) - p(+1), \quad (2.1a)$$

$$p^\cdot(-1) - p'(-1) = p(+1) - p(-1). \quad (2.1b)$$

This problem has just 1 solution $p \in C^n([0, \infty) \times R^1 \times Z)$ with data $f \in C^n(R^1 \times Z)$ for any $1 \leq n \leq \infty$.[13] Kaç[5] noticed that if $\#$ is a standard Poisson process with rate 1, jumps $+1$, and expectation E, then

$$p(t, x, e) = E\left\{f\left[x - e\int_0^t (-1)^{\#(t)-\#(s)} ds, e(-1)^{\#(t)}\right]\right\}, \quad (2.2)$$

especially, $p \geq 0$ if $f \geq 0$. The fact that (2.2) solves (2.1) is easily proved by an explicit computation of $\partial p/\partial t$. A more concrete formula for p can be obtained from the transform[14]:

$$I = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{\sinh(1-\gamma^2)^{\frac{1}{2}}t}{(1-\gamma^2)^{\frac{1}{2}}} \cos\gamma x\, d\gamma = \begin{cases} I_0(t^2 - x^2)^{\frac{1}{2}} & t \geq |x|, \\ 0 & t < |x|. \end{cases} \quad (2.3)$$

Define $f^\cdot = -ef' + D[f]$ as before. Then[15]

$$p = e^{-t}I * \tfrac{1}{2}[f^\cdot + f] + e^{-t}\frac{\partial}{\partial t} I * f/2. \quad (2.4)$$

3. FORMAL POWER SERIES SOLUTIONS

Consider a formal power series $f = \sum f_n \epsilon^n$ with coefficients from $C^\infty(R^1 \times Z)$ and let us ask if there exists a formal power series $p = \sum p_n \epsilon^n$ with coefficients from $C^\infty([0, \infty) \times R^1 \times Z)$ that solves

$$\frac{\partial p}{\partial t} + e\frac{\partial p}{\partial x} = \frac{1}{\epsilon} D[p], \quad (3.1a)$$

$$\lim_{t \downarrow 0} p = f, \quad (3.1b)$$

or, what is the same,

$$p^\cdot_{n-1} + ep'_{n-1} = D[p_n], \quad n \geq 0, \quad p_{-1} \equiv 0, \quad (3.2a)$$

$$\lim_{t \downarrow 0} p_n = f_n, \quad n \geq 0. \quad (3.2b)$$

Define an operator Q on such formal power series f by the rule:

$$Q = \frac{1}{\epsilon}[(1 + \epsilon^2\partial^2)^{\frac{1}{2}} - 1] = \sum_{m=1}^{\infty} \binom{\frac{1}{2}}{m} \epsilon^{2m-1}\partial^{2m}$$

and let us verify the following facts. f *gives rise to a*

[9] $C^\infty_\downarrow(R^1 \times Z)$ is the class of rapidly decreasing functions from $C^\infty(R^1 \times Z)$.

[10] T. Carleman, *Problèmes mathematiques dans la théorie cinetique des gaz* (Almqvist Wiksells, Uppsala, Sweden, 1957).

[11] I. I. Kolodner has proved that Eq. (1.12) is well-posed. See I. I. Kolodner, Ann. Mat. 73, 11 (1963).

[12] H. Grad, Phys. Fluids 6, 147 (1963).

[13] The standard proof is given by I. G. Petrovskii, *Lectures on Partial Differential Equations* (Interscience Publishers Inc., New York, 1954).

[14] I_0 below is the usual modified Bessel function; see H. Bateman, *Tables of Integral Transforms*, A. Erdélyi, Ed. (McGraw-Hill Book Company, Inc., New York, 1954), Vol. 1.

[15] The sign $*$ below is the customary convolution on R^1.

formal power series solution p of (3.1) = (3.2) *if and only if*

$$f^{\cdot} \equiv -ef' + \frac{1}{\epsilon} D[f] = Q[f], \tag{3.3a}$$

or, what is the same,

$$-ef_{\text{odd}} = Q\partial^{-1}f_{\text{even}} = \sum_{m=1}^{\infty} \binom{\frac{1}{2}}{m} (\epsilon\partial)^{2m-1} f_{\text{even}}. \tag{3.3b}$$

In this case, the coefficients of p are polynomials in t with coefficients from $C^{\infty}(R^1 \times Z)$, and

$$\frac{\partial p}{\partial t} = Q[p], \quad -ep_{\text{odd}} = Q\partial^{-1}p_{\text{even}} \tag{3.4a, b}$$

for any $t \geq 0$, especially, p is completely specified by the knowledge of f_{even} (Hilbert paradox; see Sec. 1 for comment).

A number of simple examples are tabulated below; the proofs occupy the rest of this section.

TABLE I. A brief list of special solutions of (3.1).

f_{even}	$-ef_{\text{odd}}$	p
1	0	1
x	$\epsilon/2$	$x - \epsilon e/2$
x^2	ϵx	$x + \epsilon t$ $-\epsilon e x$

Given a formal power series solution p, it follows from (3.1) and the rules

$$D^2 = -2D, \quad eD + De = -2e \tag{3.5a,b}$$

that

$$\left[\frac{\partial^2}{\partial t^2} + \frac{2}{\epsilon}\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}\right] p = 0, \tag{3.6a}$$

which can be re-expressed as

$$\{\partial/\partial t + \epsilon^{-1}[1 + (1 + \epsilon^2\partial^2)^{\frac{1}{2}}]\} \times \{\partial/\partial t + \epsilon^{-1}[1 - (1 + \epsilon^2\partial^2)^{\frac{1}{2}}]\} p = 0 \tag{3.6b}$$

with $(1 + \epsilon^2\partial^2)^{\frac{1}{2}}$ expanded according to the binomial series as before. Drop the first operator. What is left is the formal power series $p^{\cdot} - Q[p]$, so

$$(\partial/\partial t + \epsilon^{-1}[1 + \sqrt{1 + \epsilon^2\partial^2}])q = 0 \tag{3.7}$$

has the formal power series solution $q = p^{\cdot} - Q[p]$. Equation (3.4a) follows from the fact that the only formal power series solution of (3.7) is $q = 0$.

Proof: A formal power series solution q of (3.7) satisfies $0 = -q_1^{\cdot} = 2q_0$, so q/ϵ is likewise a formal power series solution of (3.7), $q_1 = 0$ by the same argument, etc.

Comparison of (3.1) and (3.4a) at $t = 0$ gives (3.3a), and (3.4b) follows as soon as it is proved that (3.3b) is the same as (3.3a). The proof is as follows: To derive (3.3b) from (3.3a), take even and odd parts in (3.3a) to obtain

$$-ef'_{\text{odd}} = Qf_{\text{even}}, \quad -ef'_{\text{even}} - \frac{2}{\epsilon} f_{\text{odd}} = Qf_{\text{odd}}, \tag{3.8a, b}$$

and then substitute f'_{odd} from (3.8a) into Qf_{odd} in (3.8b) to obtain (3.3b):

$$\begin{aligned} -ef_{\text{odd}} &= \frac{e\epsilon}{2}[ef'_{\text{even}} + Qf_{\text{odd}}] \\ &= \frac{e\epsilon}{2}[ef'_{\text{even}} - eQ^2\partial^{-1}f_{\text{even}}] \\ &= \frac{\epsilon}{2}\{f'_{\text{even}} - \epsilon^{-2}[1 + \epsilon^2\partial^2 \\ &\quad - 2(1 + \epsilon^2\partial^2)^{\frac{1}{2}} + 1]\partial^{-1}f_{\text{even}}\} \\ &= Q\partial^{-1}f_{\text{even}}. \end{aligned}$$

Now begin with (3.3b). Equation (3.8a) is immediate, so to prove (3.3a), it suffices to derive (3.8b) as follows:

$$\begin{aligned} Qf_{\text{odd}} &= -eQ^2\partial^{-1}f_{\text{even}} \\ &= \frac{-e}{\epsilon^2}[1 + \epsilon^2\partial^2 - 2(1 + \epsilon^2\partial^2)^{\frac{1}{2}} + 1]\partial^{-1}f_{\text{even}} \\ &= -ef'_{\text{even}} + \frac{2}{\epsilon} eQ\partial^{-1}f_{\text{even}} \\ &= -ef'_{\text{even}} - \frac{2}{\epsilon} f_{\text{odd}}. \end{aligned}$$

The problem is now to prove that *any formal power series f subject to* (3.3) *gives rise to just* 1 *formal power series solution p of* (3.1). *The coefficients of p should be polynomials in t with coefficients from $C^{\infty}(R^1 \times Z)$.* Put $n = 0$ in (3.2a). $p_{-1} \equiv 0$, so $D[p_0] = 0$, i.e., p_0 is even. Now put $n = 1$, take even parts, and conclude that $p_0^{\cdot} = 0$, i.e., $p_0 \equiv f_0$. Beginning with $p_0 = f_0$, compute the rest of p from (3.4a), expressed in the form

$$p_n^{\cdot} = \sum_{\substack{2m-1+\ell=n \\ m\geq 1 \\ 0\leq\ell<n}} \binom{\frac{1}{2}}{m} \partial^{2m} p_\ell, \quad n \geq 1. \tag{3.9}$$

This gives a formal power series solution of (3.4a) which turns out to be a solution of (3.1) also. The fact that this is the only possible formal power series solution with data f is trivial. Also, it is clear from (3.9) that p_n is a polynomial in t (of degree n) with coefficients from $C^{\infty}(R^1 \times Z)$.

To prove that p solves (3.1), consider the expression $q = -ep' + \epsilon^{-1}D[p] - Q[p]$. Because p_0 is even, this is a formal power series. Also, it solves (3.4a), it vanishes at $t = 0$ since f satisfies (3.3a), and since

from (3.9):

$$p_1 = p_0''/2, \quad \text{i.e.,} \quad p_1 = f_1 + tf_0''/2,$$

and from (3.3b):

$$-e(f_1)_{\text{odd}} = f_0'/2,$$

it develops that

$$q_0 = -ep_0' + D[p_1] = -ef_0' + D[f_1] = -ef_0' - 2(f_1)_{\text{odd}} = 0.$$

But now (3.9) implies that $q \equiv 0$, so $\partial p/\partial t = Q[p] = -ep' + \epsilon^{-1}D[p]$, as stated.

4. REGULAR SOLUTIONS

Consider a formal power series solution p of (3.1) with data

$$f_{\text{even}} = f_0, \tag{4.1a}$$

$$-ef_{\text{odd}} = Q\partial^{-1}f_{\text{even}} = \sum_{m=1}^{\infty} \binom{\frac{1}{2}}{m} (\epsilon\partial)^{2m-1} f_{\text{even}}, \tag{4.1b}$$

and let us verify that f_{odd} *converges in*[16] $C^\infty(R^1 \times Z)$ *for* $|\epsilon| < 1$ *if and only if* $f_{\text{even}} = f_0$ *is an integral function of exponential type* ≤ 1, *in which case* p *also converges in* $C^\infty([0, \infty) \times R^1 \times Z)$ *for* $|\epsilon| < 1$ *to an integral function of the same exponential type.*

The formal power series (4.1b) converges in $C^\infty(R^1 \times Z)$ for $|\epsilon| < 1$ if and only if

$$\limsup_{m \uparrow \infty} m^{-1} \lg \|\partial^m f_0\|_\infty \leq 0$$

on compact figures of $R^1 \times Z$, so the first statement is plain. The statement about p is easily proved by expressing the integral function f_{even} in Polya's fashion[17]:

$$f_{\text{even}} = \oint_{|z|=R} e^{xz} \hat{f}_{\text{even}}\, dz, \quad R > 1, \tag{4.2a}$$

with $\hat{f}_{\text{even}}$ regular outside $|z| = 1$. Then

$$f_{\text{odd}} = \oint -\exp \frac{(1 + \epsilon^2 z^2)^{\frac{1}{2}} - 1}{\epsilon z}\, e^{xz} \hat{f}_{\text{even}}\, dz, \quad R\,|\epsilon| < 1, \tag{4.2b}$$

and putting $\hat{f} = (1 - (e/\epsilon z)[(1 + \epsilon^2 z^2)^{\frac{1}{2}} - 1])\hat{f}_{\text{even}}$, the statement about p can be read off from the formula

$$p = \oint \exp\{t\epsilon^{-1}[(1 + \epsilon^2 z^2)^{\frac{1}{2}} - 1]\}\, e^{xz} \hat{f}\, dz, \quad R\,|\epsilon| < 1. \tag{4.2c}$$

5. APPROXIMATION BY HILBERT SOLUTIONS

Now add

$$\|f_{\text{even}}\|_1 < \infty \tag{5.1}$$

to the conditions (4.1). Then the Polya integrals in (4.2) can be replaced by Fourier integrals for $|\epsilon| < 1$:

$$f_{\text{even}} = \int_{-1}^{+1} e^{i\gamma x} \hat{f}_{\text{even}}\, d\gamma, \tag{5.2a}$$

$$f_{\text{odd}} = \int_{-1}^{+1} -e\, \frac{(1 - \epsilon^2\gamma^2)^{\frac{1}{2}} - 1}{i\epsilon\gamma}\, e^{i\gamma x} \hat{f}_{\text{even}}\, d\gamma, \tag{5.2b}$$

$$p = \int_{-1}^{+1} \exp\{t\epsilon^{-1}[(1 - \epsilon^2\gamma^2)^{\frac{1}{2}} - 1]\}\, e^{i\gamma x} \hat{f}\, d\gamma, \tag{5.2c}$$

with

$$\hat{f} = \{1 - e(i\epsilon\gamma)^{-1}[(1 - \epsilon^2\gamma^2)^{\frac{1}{2}} - 1]\}\hat{f}_{\text{even}}, \tag{5.3a}$$

$$\hat{f}_{\text{even}} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\gamma x} f_{\text{even}}\, dx. \tag{5.3b}$$

Because $\hat{f}_{\text{even}} \in C[-1, +1]$, p is not only convergent on the open disk $|\epsilon| < 1$, but is also continuous on the closed disk $|\epsilon| \leq 1$, and p *at* $\epsilon = 1$ is a bona fide solution of

$$p^{\cdot} + ep' = D[p]. \tag{5.4}$$

Now, consider any solution p of (5.4) with data $f \in C_{\downarrow}^{\infty}(R^1 \times Z)$. By (2.3) and (2.4),

$$p = e^{-t} \int_{-\infty}^{+\infty} e^{i\gamma x} \frac{\sinh(1 - \gamma^2)^{\frac{1}{2}}t}{(1 - \gamma^2)^{\frac{1}{2}}} [\hat{f}^{\cdot} + \hat{f})\, d\gamma + e^{-t} \int_{-\infty}^{+\infty} e^{i\gamma x} \cosh(1 - \gamma^2)^{\frac{1}{2}}t f\, d\gamma \tag{5.5}$$

with $f^{\cdot} = -ef' + D[f]$. Define

$$\hat{f}^0 = \frac{1}{2}\left[1 + \frac{1 + D - i\gamma e}{(1 - \gamma^2)^{\frac{1}{2}}}\right]\hat{f}, \quad |\gamma| \leq 1, \qquad = 0, \quad |\gamma| > 1. \tag{5.6}$$

Then, for $|\gamma| \leq 1$,

$$2(1 - \gamma^2)^{\frac{1}{2}} \hat{f}^0_{\text{even}} = [(1 - \gamma^2)^{\frac{1}{2}} + 1]\hat{f}_{\text{even}} - i\gamma e \hat{f}_{\text{odd}}, \tag{5.7a}$$

$$2(1 - \gamma^2)^{\frac{1}{2}} \hat{f}^0_{\text{odd}} = [(1 - \gamma^2)^{\frac{1}{2}} - 1]\hat{f}_{\text{odd}} - i\gamma e \hat{f}_{\text{even}}, \tag{5.7b}$$

and the map $f \to f^0$ is a *projection* onto the class of functions satisfying (3.3a) = (3.3b) for $\epsilon = 1$, as a little algebra will verify. The corresponding Hilbert solution

$$p^0 = \int_{-1}^{+1} e^{i\gamma x} \exp\{t[(1 - \gamma^2)^{\frac{1}{2}} - 1]\}\, \hat{f}^0\, d\gamma \tag{5.8}$$

differs from $p =$ (5.5) by

$$e^{-t} \int_{|\gamma|>1} e^{i\gamma x} \frac{\sinh(1 - \gamma^2)^{\frac{1}{2}}t}{(1 - \gamma^2)^{\frac{1}{2}}} [\hat{f}^{\cdot} + \hat{f}] + e^{-t} \int_{|\gamma|>1} e^{i\gamma x} \cosh(1 - \gamma^2)^{\frac{1}{2}} t \hat{f} + e^{-t} \int_{-1}^{+1} e^{i\gamma x} e^{-t(1-\gamma^2)^{\frac{1}{2}}} \frac{1}{2}\left[1 - \frac{1 + D - i\gamma e}{(1 - \gamma^2)^{\frac{1}{2}}}\right]\hat{f}, \tag{5.9}$$

[16] f is convergent in $C^\infty(R^1 \times Z)$ if each of the formal power series $\partial^m f$ ($m \geq 0$) converges uniformly on compact figures of $R^1 \times Z$.

[17] R. Boas, *Entire Functions* (Academic Press Inc., New York, 1954).

and all this is of magnitude e^{-t}, roughly. This is the best fit that could be expected. Because a positive-definite function has to be continuous, it is obvious from (5.6) that the projection $f \to f^0$ does not preserve positivity.

6. POSITIVE HILBERT SOLUTIONS

An example will settle the existence of nonnegative summable Hilbert solutions.

Example 1: Define

$$\hat{f}_{\text{even}} = [(1-\gamma^2)^{\frac{1}{2}} + 1](1-|\gamma|)^2, \quad |\gamma| \le 1,$$
$$= 0, \quad |\gamma| > 1.$$

Then $f_{\text{even}} = (5.2a)$ is an integral function of exponential type ≤ 1, and, with f_{odd} defined by (5.2b), $f \ge 0$ (see proof below), $p \ge 0$ by (2.2), and since

$$\hat{p} = \exp\{t[(1-\gamma^2)^{\frac{1}{2}} - 1]\}\left[1 - e\frac{(1-\gamma^2)^{\frac{1}{2}} - 1}{\gamma i}\right] \times [(1-\gamma^2)^{\frac{1}{2}} + 1](1-|\gamma|)_2$$

is finite at $\gamma = 0$, $\int p < \infty$.

Proof that $f \ge 0$:

$$\hat{f} = \left[1 - e\frac{(1-\gamma^2)^{\frac{1}{2}} - 1}{i\gamma}\right][(1-\gamma^2)^{\frac{1}{2}} + 1](1-|\gamma|)^2$$
$$= [(1-\gamma^2)^{\frac{1}{2}} + 1 + ei\gamma](1-|\gamma|)^2, \quad |\gamma| \le 1,$$

is split into 2 pieces: $A = (1-\gamma^2)^{\frac{1}{2}}(1-|\gamma|)^2$ and $B = (1+ei\gamma)(1-|\gamma|)^2$. A is positive-definite since

$$A'' = \left(\frac{1-\gamma}{1+\gamma}\right)^{\frac{1}{2}}[2 + 6\gamma - (1+\gamma)^{-1}], \quad 0 < \gamma < 1,$$

so that

$$\int_{-1}^{+1} e^{i\gamma x} A = 2\int_0^1 \cos \gamma x A$$
$$= 2x^{-2}\int_0^1 [1 - \cos \gamma x]A'' \ge 0.$$

B is also positive-definite, but the proof is not so cheap.

$$\int_{-1}^{+1} e^{i\gamma x} B = (1 + e\gamma)x^{-3}(x - \sin x)$$
$$= x^{-3}[(1 - 3e/x)(x - \sin x) + e(1 - \cos x)],$$

and by a trivial manipulation, it is enough to verify

$$(1 \mp 3/x)(x - \sin x) \pm (1 - \cos x) \ge 0, \quad x > 0.$$

But for $0 < x \le 3$,

$$\left(1 - \frac{3}{x}\right)(x - \sin x) + 1 - \cos x$$
$$\ge \left(1 - \frac{3}{x}\right)\frac{x^3}{6} + \frac{x^2}{2} - \frac{x^4}{4!}$$
$$= \frac{x^3}{6}\left(1 - \frac{x}{4}\right) \ge \frac{x^3}{6}(1 - \tfrac{3}{4}) > 0$$

and

$$\left(1 + \frac{3}{x}\right)(x - \sin x) - 1 + \cos x$$
$$\ge \left(1 + \frac{3}{x}\right)\left(\frac{x^3}{6} - \frac{x^5}{5!}\right) - \frac{x^2}{2}$$
$$= x^3\left(\frac{1}{6} - \frac{x^2}{5!} - \frac{x}{40}\right) \ge x^3(\tfrac{1}{6} - \tfrac{3}{20}) > 0,$$

while the same bounds are obvious for $x > 3$. This completes the proof.

Example 2: A simpler example is available if R^1 is replaced by the circle $0 \le x < 2\pi$. Now, f_{even} is *both* periodic and of exponential type ≤ 1, so it has to be of the form $a + b\cos x + c \sin x$, and

$$f = \{1 - e[(1+\partial^2)^{\frac{1}{2}} - 1]\partial^{-1}\}f_{\text{even}}$$
$$= a + (b + ec)\cos x + (c - eb)\sin x$$

is nonnegative if and only if $a \ge [2(b^2 + c^2)]^{\frac{1}{2}}$.

ACKNOWLEDGMENTS

This work was completed while the author was at the Massachusetts Institute of Technology. The partial support of the National Science Foundation (NSF GP 4364) is gratefully acknowledged.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Derivations of Hierarchies for N-Particle Systems and Vlasov Systems by Means of the Functional Calculus

TOSHIO NAKAYAMA AND JOHN M. DAWSON*
Institute of Plasma Physics, Nagoya University, Nagoya, Japan

(Received 8 March 1966)

The methods of the functional calculus are applied to the Klimontovich equation and to the Vlasov equations. The equation for the generating functional is derived for both cases. By taking moments of these equations, the BBGKY hierarchy is obtained for the case of the Klimontovich equation, and a similar hierarchy is obtained for the case of the Vlasov equation. The two hierarchies are identical, except for some terms involving individual particle interactions. Methods of solving the Vlasov hierarchy are discussed and it is shown how quasi-linear theory can be obtained.

I. INTRODUCTION

FOR studies of nonequilibrium systems a basic starting point has been the Liouville equation. Because of its mathematical complexity and because of one's interest in functions of only a few dynamical variables, the main problem in nonequilibrium statistical mechanics is that of finding equations for reduced distribution functions. In recent years much progress has been made in this field. A diagrammatic expansion method originated by Prigogine and his school[1] has been successfully employed in deriving equations for the reduced distribution functions. However, this method has the disadvantages that it is complex, it is difficult to prove that all the important diagrams have been included, and it is difficult to verify the convergency of the series involved.

A second approach has been to employ the so-called BBGKY hierarchy[2] for the reduced distribution functions. This hierarchy is open and must be truncated by making some assumptions and approximations. One such truncation scheme is to assume that three-particle correlations are negligibly small. This approach has been successful in deriving the Boltzmann equation for low-density gases of neutral particles[3] and for the derivation of the Fokker–Planck equation for plasmas.[4]

In 1957, Klimontovich[5] proposed to study the N-body problem through the use of an equation which is identical in form with the Vlasov[6] equation, but in which he took the distribution function to be that due to a discrete set of particles (therefore, a set of delta functions whose arguments are specified by the classical laws of mechanics). Starting from the Klimontovich equation, one can derive the BBGKY hierarchy by averaging the solutions over a distribution of initial positions and velocities.[7] This method has been successfully applied to a number of problems.

Another area of nonequilibrium phenomenon which has been receiving considerable attention recently is the field of turbulence. This is particularly true in the field of plasma physics. Here, also, it appears that a statistical approach is called for and some success has been achieved in developing such theories.[8,9]

It is the purpose of the present paper to first show how the functional calculus can be used to derive the BBGKY hierarchy of statistical mechanics and second to derive a similar hierarchy[10,11] which is applicable to turbulences of a Vlasov fluid (a plasma, for example).

Our starting points for the derivation of these hierarchies are the Klimontovich equation and Vlasov

* Permanent address: Plasma Physics Laboratory, Princeton University, Princeton, New Jersey.

[1] For example, see I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1963); R. Balescu, *Statistical Mechanics of Charged Particles* (Interscience Publishers, Inc., New York, 1963).

[2] Among many works, see, for example, N. N. Bogoliubov, in *Studies in Statistical Mechanics*, translated by E. K. Gora, J. de Boer and G. E. Uhlenbeck, Eds. (North-Holland Publishing Company, Amsterdam, 1962).

[3] J. G. Kirkwood, J. Chem. Phys. **15**, 72 (1947).

[4] N. Rostoker and M. Rosenbluth, Phys. Fluids **3**, 1 (1960); N. Rostoker, *ibid.* **3**, 922 (1960); A. Lenard, Ann. Phys. (N.Y.) **10**, 390 (1960).

[5] Iu. L. Klimontovich, Zh. Eksperim. i Teor. Fiz. **33**, 982 (1957) [English transl.: Soviet Phys.—JETP **6**, 753 (1958)].

[6] A. A. Vlasov, *Many-Particle Theory and Its Application to Plasma* (Gordon and Breach Science Publishers, Inc., New York, 1961).

[7] E. P. Gross, J. Nucl. Energy **C2**, 173 (1961). Also, see reports by W. E. Brittin *et al.*, University of Colorado; the most complete self-contained work with many references is found in W. R. Chappel, Ph.D. thesis, University of Colorado (1965).

[8] W. Drummond and D. Pines, Ann. Phys. (N. Y.) **28**, 478 (1964); A. A. Vedenov, J. Nucl. Energy **C5**, 169 (1963).

[9] B. Kadomtsev, *Plasma Turbulence*, translated by L. C. Ramson, M. G. Rusbridge, trans. Ed. (Academic Press Inc., New York, 1965).

[10] The authors have recently learned, from private communication, that Von P. Graff of the Institute für Plasmaphysik, Kernforschungsanlage Julich, has also obtained this hierarchy (unpublished work).

[11] C. Oberman, Bull. Am. Phys. Soc. **6**, 185 (1961).

equations. Because these equations are formally identical, the same techniques work for both. The derivation of the BBGKY hierarchy simply gives another method of obtaining this hierarchy. However, the observation that a similar hierarchy applies to the Vlasov equation gives us another method for attacking problems of plasma turbulence. In fact, by making suitable approximations one can obtain the "quasi-linear theory" of plasma turbulence. By making different approximations, one could obtain different theories of plasma turbulence. Finally, we might point out that this approach may shed some light on the question which has often been raised, "is it sufficient to consider only the Vlasov equation when dealing with unstable plasmas or do we need to take into account particle correlations through the BBGKY hierarchy." It has been argued that since particle correlations become very large in such plasmas, we should include their effects. However, since a very similar hierarchy holds for the Vlasov equation, we may argue that the correlations due to instabilities are to a large degree already contained in the Vlasov equation and that it is sufficient to consider only this equation (perhaps including some appropriate initial disturbances).

II. FUNCTIONALS AND FUNCTIONAL EQUATIONS

A. Review of the Functional Calculus

Our approach in many ways follows the treatment by Hopf[12] of hydrodynamic turbulence. The details of the formalism can be found in his work. However, we include a brief account of the derivation of the functional equation for the sake of clarity and completeness.

Consider a function of $\mathbf{x}$ and t, $f(\mathbf{x}, t)$, which satisfies a deterministic equation,

$$\partial f(\mathbf{x}, t)/\partial t = L(f). \tag{1}$$

Here L is a polynomial function of f only and is assumed not to depend explicitly on time t. This equation may describe the time development of some system, as for example, it might be the Vlasov equation for a plasma. In place of (1) we may define a time-development operator T^t such that T^t acting on $f(\mathbf{x}, 0)$ gives $f(\mathbf{x}, t)$,

$$T^t f(\mathbf{x}, 0) = f(\mathbf{x}, t). \tag{2}$$

This is simply another way of writing (1).

Now denote the space spanned by the solutions of Eq. (1) by Ω and introduce the probability $P(\mathrm{A})$ of finding f in the region A, $\mathrm{A} \subset \Omega$.

Since P is a probability we must have

$$P(\mathrm{A}) \geq 0, \quad P(\Omega) = 1. \tag{3}$$

If the probability is conserved following the motion (no systems are added to or removed from those under consideration), then

$$P^t(\mathrm{A}) = P^0(T^{-t}\mathrm{A}), \tag{4}$$

where P^t denotes the probability at time t, and $T^{-t}\mathrm{A}$ denotes the region in which f must be at $t = 0$ to be in A at time t.

Define an average of any function of f, $F(f)$, by

$$\langle F(f)\rangle_{\mathrm{av}} = \int_\Omega F(f) P^t(df). \tag{5}$$

By making use of Eq. (4) we have

$$\langle F\rangle_{\mathrm{av}} = \int_\Omega F(f) P^t(df) = \int_\Omega F[T^t f(\mathbf{x}, 0)] P^0[df(\mathbf{x}, 0)]. \tag{6}$$

With this preparation we now introduce the following generating functional:

$$\begin{aligned}\Phi(y, t) &= \int_\Omega \exp i(y, f) P^t(df) \\ &= \int_\Omega \exp\{i[y, T^t f(\mathbf{x}, 0)]\} P^0[df(\mathbf{x}, 0)],\end{aligned} \tag{7}$$

where f satisfies Eq. (1) and we have made use of Eq. (6). Here, $y(\mathbf{x})$ is an arbitrary continuous function of $\mathbf{x}$ and does not depend on time; the bracket quantity (y, f) stands for the scalar product of y and f,

$$(y, f) = \int y(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}.$$

Differentiating both sides of (7) with respect to t and making use of Eq. (1) gives

$$\begin{aligned}\frac{\partial \Phi(y, t)}{\partial t} &= \int i\left[\int d\mathbf{x} y(\mathbf{x}) \frac{\partial T^t}{\partial t} f(\mathbf{x}, 0)\right] \\ &\qquad \exp\{i[y, T^t f(\mathbf{x}, 0)]\} P^0[df(\mathbf{x}, 0)] \\ &= \int i\left[\int d\mathbf{x} y(\mathbf{x}) L[f(\mathbf{x}, t)]\right] \\ &\qquad \exp\{i[y, f(\mathbf{x}, t)]\} P^0[df(\mathbf{x}, 0)] \\ &= i \int y(\mathbf{x}) L\left(\frac{\delta}{i\delta y}\right) \Phi(y, t)\, dx.\end{aligned} \tag{8}$$

Here the notation $\delta/i\delta y$ denotes functional differentiation with respect to $y(\mathbf{x})$ and is adopted from Bogoliubov. A rigorous proof of Eq. (8) is given by Hopf.[12]

[12] E. Hopf, J. Ratl. Mech. Analysis 1, 87 (1952); E. Hopf and E. W. Titt, *ibid.* 2, 587 (1953).

B. Application of the Functional Calculus to the Klimontovich and Vlasov Equations

An exact equation for the dynamics of a single system of N identical particles is given by the Klimontovich equation.[5] This description sometimes leads to a better understanding of the physics of many-body problems.[5,7,13]

Klimontovich considers the distribution function for a single system of N identical particles,

$$f(\mathbf{x}, t) = \sum_{i=1}^{N} \delta[\mathbf{x} - x_i(t)]. \tag{9}$$

Here $\mathbf{x}$ stands for a set of coordinates and associated canonical momenta of a particle $(\mathbf{q}, \mathbf{p})$ and the delta function is a six-dimensional one. The $\mathbf{x}_i(t)$'s are the classical orbits of the particles and are determined by solving Hamilton's equations of motion. The Hamiltonian for the system under consideration is

$$H = \int \frac{p^2}{2m} f(\mathbf{x}, t)\, d\mathbf{x} + \tfrac{1}{2}\int d\mathbf{x}\, d\mathbf{x}' \phi(|\mathbf{q} - \mathbf{q}'|) \times [f(\mathbf{x}, t) f(\mathbf{x}', t) - \delta(\mathbf{x} - \mathbf{x}') f(\mathbf{x}, t)], \tag{10}$$

where ϕ is the two-body potential (the only one assumed present). The term $\delta(\mathbf{x} - \mathbf{x}')f(\mathbf{x}, t)$ is subtracted in (10) because it is assumed that a particle does not interact with itself.

The equation which f satisfies is the Klimontovich equation,

$$\frac{\partial f(\mathbf{x}, t)}{\partial t} = -\frac{\mathbf{p}}{m} \cdot \frac{\partial f(\mathbf{x}, t)}{\partial \mathbf{q}} + \int \frac{\partial \phi(|\mathbf{q} - \mathbf{q}'|)}{\partial \mathbf{q}} \cdot \frac{\partial f(\mathbf{x}, t)}{\partial \mathbf{p}} \times [f(\mathbf{x}', t) - \delta(\mathbf{x} - \mathbf{x}')]\, d\mathbf{x}'. \tag{11}$$

This is most simply verified by direct substitution.

Under some circumstances (particularly when dealing with plasmas), we may treat the system as a continuous Vlasov fluid rather than a set of discrete particles. In this case we treat f as a continuous function and we drop the $\delta(\mathbf{x} - \mathbf{x}')$ terms from Eqs. (10) and (11). Thus in this case the Hamiltonian is given by

$$H = \int \frac{p^2}{2m} f(\mathbf{x}, t)\, d\mathbf{x} + \tfrac{1}{2}\int d\mathbf{x}\, d\mathbf{x}' \phi(|\mathbf{q} - \mathbf{q}'|) \times f(\mathbf{x}, t) f(\mathbf{x}', t), \tag{12}$$

while the Vlasov equation is

$$\frac{\partial f(\mathbf{x}, t)}{\partial t} = -\frac{\mathbf{p}}{m} \cdot \frac{\partial f(\mathbf{x}, t)}{\partial \mathbf{q}} + \int \frac{\partial \phi(|\mathbf{q} - \mathbf{q}'|)}{\partial \mathbf{q}} \cdot \frac{\partial f(\mathbf{x}, t)}{\partial \mathbf{p}} f(\mathbf{x}', t)\, d\mathbf{x}'. \tag{13}$$

We now identify the f appearing in the Klimontovich and Vlasov equations with the f appearing in Eq. (1) and the right-hand sides of Eq. (11) or (13), depending on which system is under consideration, with $L(f)$ appearing in Eq. (1).

1. *Application to the Klimontovich Equations*

For the case of the Klimontovich equation, we must, however, note that the space Ω is not the whole space spanned by the solutions of Eq. (11), but is the subspace spanned only by f's of the form of Eq. (9). Because of the special nature of this Ω space, we first consider a few examples to obtain a better understanding of the following sections.

The average of $f(\mathbf{x}_1)$ is simply

$$\langle f(\mathbf{x}_1)\rangle = \frac{\delta\Phi}{i\delta y(\mathbf{x}_1)}\bigg|_{y=0} = \int_\Omega f(\mathbf{x}_1) \exp[i(y,f)] P^t(df)\bigg|_{y=0} = N f_1(\mathbf{x}_1, t). \tag{14}$$

In the above and hereafter the subscript on f indicates the reduced function in the usual sense (the one-particle distribution function in the above). The last step in the above equation follows by considering the meaning of $P(df)$. This is the differential increment of the probability due to changing the function f. In this case the change is a change in the delta functions due to changes in initial positions and momenta. The average in Eq. (14), thus, is nothing, but an average over initial values of the coordinates and momenta of the particles. Thus it is equivalent to the average Klimontovich and others take.

In a similar manner, one can compute the second moment of f by operating on Φ with the second-order functional derivatives with respect to $y(\mathbf{x}_1)$ and $y(\mathbf{x}_2)$,

$$\langle f(\mathbf{x}_1) f(\mathbf{x}_2)\rangle_{\mathrm{av}} = \frac{\delta^2\Phi}{i\delta y(\mathbf{x}_1)\delta y(\mathbf{x}_2)}\bigg|_{y=0} = \int_\Omega f(\mathbf{x}_1) f(\mathbf{x}_2) \exp[i(y,f)] P^t(df)\bigg|_{y=0}. \tag{15}$$

The average $\langle f(\mathbf{x}_1) f(\mathbf{x}_2)\rangle_{\mathrm{av}}$, however, differs from the usual two-particle distribution function by the amount $\delta(\mathbf{x}_1 - \mathbf{x}_2) f(\mathbf{x}_1)$ since we do not wish to correlate a particle with itself. Thus for $f_2(\mathbf{x}_1, \mathbf{x}_2, t)$, we write

$$N(N-1) f_2(\mathbf{x}_1, \mathbf{x}_2, t) = \langle f(\mathbf{x}_1) f(\mathbf{x}_2)\rangle_{\mathrm{av}} - N\delta(\mathbf{x}_1 - \mathbf{x}_2) f_1(\mathbf{x}_1, t). \tag{16}$$

In general, to extract the correct s-particle distribution function from the generating functional, we operate with

$$\frac{1}{N(N-1)\cdots(N-s+1)} \frac{\delta}{i\delta y(\mathbf{x}_1)} \times \left(\frac{\delta}{i\delta y(\mathbf{x}_2)} - \delta(\mathbf{x}_2 - \mathbf{x}_1)\right) \cdots \times \left(\frac{\delta}{i\delta y(\mathbf{x}_s)} - \sum_{i<s} \delta(\mathbf{x}_s - \mathbf{x}_i)\right), \tag{17}$$

and then set all y's equal to zero.

[13] J. M. Dawson and T. Nakayama, Phys. Fluids 9, 252 (1966).

The equation for the characteristic functional for the case of the Klimontovich equation is obtained as shown in Subsection A and is given by Eq. (8) making use of the identification of L from Eq. (11). This equation is

$$\frac{\partial\Phi(y,t)}{\partial t} = -i\int y(\mathbf{x})\frac{\mathbf{p}}{m}\cdot\frac{\partial}{\partial\mathbf{q}}\frac{\delta}{i\delta y(\mathbf{x})}\Phi(y,t)\,d\mathbf{x} + i\iint y(\mathbf{x})\frac{\partial\phi(|\mathbf{q}-\mathbf{q}'|)}{\partial\mathbf{q}}\left(\frac{\delta}{i\delta y(\mathbf{x}')}-\delta(\mathbf{x}-\mathbf{x}')\right) \times(\partial/\partial\mathbf{p})[\delta/i\delta y(\mathbf{x})]\Phi(y,t)\,d\mathbf{x}\,d\mathbf{x}'. \quad (18)$$

It should be noted that the functional equation derived here is somewhat different from that derived by Bogoliubov[2] since the functional (7) contains an extra functional,

$$\Phi(y,t) = 1+\sum_{s\geq 1}\frac{(i)^s}{s!}\int\cdots\int N(N-1)\cdots(N-s+1) \times f_s(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_s,t)y(\mathbf{x}_1)y(\mathbf{x}_2)\cdots \times y(\mathbf{x}_s)\,d\mathbf{x}_1\,d\mathbf{x}_2\cdots d\mathbf{x}_s+\theta(y,t) = \mathfrak{L}(Nig,t)+\theta(y,t)\ (\text{in the limit of } N\to\infty). \quad (19)$$

Here $\mathfrak{L}$ is the one used by Bogoliubov[2] and $\theta(y,t)$ is

$$\theta(y,t) = \sum_{s=2}\frac{(i)^s}{s!}\int\cdots\int\left[\prod_{i=1}^{s}f(\mathbf{x}_i)-N^s \times f_s(\mathbf{x}_1,\cdots,\mathbf{x}_s,t)\right]P^t(df)\prod_{i=1}^{s}y(\mathbf{x}_i)\,d\mathbf{x}_i. \quad (20)$$

2. Application to the Vlasov Equation

In like manner, we obtain the equation for the characteristic functional for the case of the Vlasov equation by identifying L as the right-hand side of Eq. (13). The functional equation in this case is

$$\frac{\partial\Phi}{\partial t}(y,t) = -i\int y(\mathbf{x})\frac{\mathbf{p}}{m}\cdot\frac{\partial}{\partial\mathbf{q}}\frac{\delta\Phi}{i\delta y(\mathbf{x})}(y,t)\,d\mathbf{x} + i\iint y(\mathbf{x})\frac{\partial\phi(|\mathbf{q}-\mathbf{q}'|)}{\partial\mathbf{q}}\frac{\delta}{i\delta y(\mathbf{x}')} \cdot\frac{\partial}{\partial\mathbf{p}}\frac{\delta\Phi}{i\delta y(\mathbf{x})}(y,t)\,d\mathbf{x}\,d\mathbf{x}'. \quad (21)$$

III. DERIVATION OF THE HIERARCHIES

A. BBGKY Hierarchy

Since Eq. (18) is different from Bogoliubov's, the derivation of the hierarchy in this formalism is given here. To obtain the first member, we operate on Eq. (18) with the functional derivative with respect to $iy(\mathbf{x}_1)$ and set all y's equal to zero. The left-hand side then becomes

$$\frac{\delta}{i\delta y(\mathbf{x}_1)}\frac{\partial}{\partial t}\Phi\Big|_{y=0} = \frac{\partial}{\partial t}\int\frac{\delta}{i\delta y(\mathbf{x}_1)}\exp[i(y,f)]P^t(df)\Big|_{y=0} = \frac{\partial\langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}}{\partial t}.$$

The first term on the right-hand side of Eq. (18) becomes

$$-\frac{i\delta}{i\delta y(\mathbf{x}_1)}\int y(\mathbf{x})\frac{\mathbf{p}}{m}\frac{\partial}{\partial\mathbf{q}}\cdot\frac{\delta}{i\delta y(\mathbf{x})}\Phi(y,t)\,d\mathbf{x}\Big|_{y=0} = -\frac{\mathbf{p}_1}{m}\cdot\frac{\partial}{\partial\mathbf{q}_1}\int f(\mathbf{x}_1)\exp[i(y,f)]P^t(df)\Big|_{y=0} - i\int y(\mathbf{x})\frac{\mathbf{p}}{m}\cdot\frac{\partial}{\partial\mathbf{q}}\frac{\delta}{i\delta y(\mathbf{x}_1)}\frac{\delta}{i\delta y(\mathbf{x})}\Phi\,d\mathbf{x}\Big|_{y=0} = -\frac{\mathbf{p}_1}{m}\cdot\frac{\partial\langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}}{\partial\mathbf{q}_1}.$$

In a similar way, the second term on the right-hand side of (18) becomes

$$\frac{i\delta}{i\delta y(\mathbf{x}_1)}\iint y(\mathbf{x})\left[\frac{\partial\phi(|\mathbf{q}-\mathbf{q}'|)}{\partial\mathbf{q}}\left(\frac{\delta}{i\delta y(\mathbf{x}')}-\delta(\mathbf{x}'-\mathbf{x})\right)\right] \cdot\frac{\partial}{\partial\mathbf{p}}\frac{\delta\Phi}{i\delta y(\mathbf{x})}\,d\mathbf{x}\,d\mathbf{x}'\Big|_{y=0} = \int\frac{\partial\phi(|\mathbf{q}_1-\mathbf{q}'|)}{\partial\mathbf{q}_1}[f(\mathbf{x}')-\delta(\mathbf{x}'-\mathbf{x}_1)] \times\frac{\partial}{\partial\mathbf{p}_1}f(\mathbf{x}_1)\exp[i(y,f)]P^t(df)\,d\mathbf{x}'\Big|_{y=0} + \iint y(\mathbf{x})\frac{\partial\phi(|\mathbf{q}-\mathbf{q}'|)}{\partial\mathbf{q}}\frac{\delta}{i\delta y(\mathbf{x}_1)} \times\left(\frac{\delta}{i\delta y(\mathbf{x}')}-\delta(\mathbf{x}-\mathbf{x}')\right)\cdot\frac{\partial}{\partial\mathbf{p}}\frac{\delta\Phi}{i\delta y(\mathbf{x})}\,d\mathbf{x}\,d\mathbf{x}'\Big|_{y=0} = \int\frac{\partial\phi(|\mathbf{q}_1-\mathbf{q}'|)}{\partial\mathbf{q}}\cdot\frac{\partial}{\partial\mathbf{p}_1} \times\langle[f(\mathbf{x}')-\delta(\mathbf{x}-\mathbf{x}_1)]f(\mathbf{x}_1)\rangle_{\mathrm{av}}\,d\mathbf{x}',$$

where the following relation has been used:

$$\int\sum_i\psi(\mathbf{q},\mathbf{q}')\delta(\mathbf{x}'-\mathbf{x})\frac{\partial}{\partial\mathbf{p}}\delta(\mathbf{x}-\mathbf{x}_i)\,d\mathbf{x}' = \int\sum_i\psi(\mathbf{q},\mathbf{q}')\delta(\mathbf{x}'-\mathbf{x}_i)\frac{\partial}{\partial\mathbf{p}}\delta(\mathbf{x}-\mathbf{x}')\,d\mathbf{x}' = \int\sum_i\psi(\mathbf{q},\mathbf{q}')\frac{\partial\delta(\mathbf{x}'-\mathbf{x}_i)}{\partial\mathbf{p}'}\delta(\mathbf{x}-\mathbf{x}')\,d\mathbf{x}' = \int\psi(\mathbf{q},\mathbf{q}')\delta(\mathbf{x}-\mathbf{x}')\frac{\partial}{\partial\mathbf{p}}\sum_i\delta(\mathbf{x}-\mathbf{x}_i)\,d\mathbf{x}.$$

Thus we obtain the first member of the hierarchy, which can be found in many texts,

$$\frac{\partial f_1(\mathbf{x}_1)}{\partial t} + \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} f_1(\mathbf{x}_1) - N \int \frac{\partial \phi(|\mathbf{q}_1 - \mathbf{q}'|)}{\partial \mathbf{q}_1} \cdot \frac{\partial f_2(\mathbf{x}_1, \mathbf{x}')}{\partial \mathbf{p}_1} d\mathbf{x}' = 0. \quad (22)$$

In a similar way, one can obtain the sth member of the hierarchy by operation of the functional derivative (17),

$$\frac{\partial f_s}{\partial t} + \sum_{i \atop i \subset s} \frac{\mathbf{p}_i}{m} \cdot \frac{\partial}{\partial \mathbf{q}_i} f_s - \sum_{\substack{i \neq j \\ i \subset s \\ j \subset s}} \frac{\partial(|\mathbf{q}_i - \mathbf{q}_j|)}{\partial \mathbf{q}_i} \cdot \frac{\partial f_s}{\partial \mathbf{p}_i} = N \sum_{i \atop i \subset s} \int \frac{\partial \phi(|\mathbf{q}_i - \mathbf{q}_{s+1}|)}{\partial \mathbf{p}_i} \cdot \frac{\partial f_{s+1}}{\partial \mathbf{p}_i} d\mathbf{x}_{s+}, \quad (23)$$

where the third term on the left-hand side as well as the terms on the right-hand side arise from the delta-function term in Eq. (18).

B. Hierarchy for the Vlasov Equation

We may employ exactly the same procedure to the Vlasov equation using Eq. (21) for the characteristic functional in place of Eq. (18). Everything goes through in exactly the same manner except that the terms resulting from the delta functions are now absent. We thus obtain a hierarchy very similar to the BBGKY hierarchy. The first member of this hierarchy is obtained by operating on Eq. (21) with $\delta/i\delta y(\mathbf{x}_1)$ and is given by

$$\frac{\partial \langle f(\mathbf{x}_1, t)\rangle_{\mathrm{av}}}{\partial t} + \frac{\mathbf{p}_1}{m} \cdot \frac{\partial \langle f(\mathbf{x}_1, t)\rangle_{\mathrm{av}}}{\partial \mathbf{q}_1} \times \int \frac{\partial \phi(|\mathbf{q}_1 - \mathbf{q}_2|)}{\partial \mathbf{q}_1} \cdot \frac{\partial \langle f(\mathbf{x}_1 t) f(\mathbf{x}_2, t)\rangle_{\mathrm{av}}}{\partial \mathbf{p}_1} d\mathbf{x}_2 = 0. \quad (24)$$

The sth member is given by operating on Eq. (21) with the functional derivative,

$$\frac{(N-s)!}{N!} \frac{\delta}{i\delta y(\mathbf{x}_1)} \cdot \frac{\delta}{i\delta y(\mathbf{x}_2)} \cdot \cdots \cdot \frac{\delta}{i\delta y(\mathbf{x}_s)},$$

and is

$$\frac{\partial \langle f(\mathbf{x}_1) \cdots f(\mathbf{x}_s)\rangle_{\mathrm{av}}}{\partial t} + \sum_{i=1}^{s} \frac{\mathbf{p}_i}{m} \cdot \frac{\partial \langle f(\mathbf{x}_1) \cdots f(\mathbf{x}_s)\rangle_{\mathrm{av}}}{\partial \mathbf{q}_i} - \int \sum_i \frac{\partial \phi(|\mathbf{q}_i - \mathbf{q}_{s+1}|)}{\partial \mathbf{q}_i} \cdot \frac{\partial \langle f(\mathbf{x}_1) \cdots f(\mathbf{x}_s) f(\mathbf{x}_{s+1})\rangle_{\mathrm{av}}}{\partial \mathbf{p}_i} d\mathbf{x}_{s+1}. \quad (25)$$

Here, $\langle f(\mathbf{x}_1) \cdots f(\mathbf{x}_s)\rangle_{\mathrm{av}}$ is equivalent to the s-particle correlation function in the BBGKY hierarchy. It is, in fact, the correlation function for f at the s phase points, $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_s$, which amounts to correlations of s particles with positions and momenta given by the x's. The difference between f_s and $\langle f(\mathbf{x}_1) \cdots f(\mathbf{x}_s)\rangle_{\mathrm{av}}$ arises only from the interpretation of f, that is, only from interpreting f as either the singular functions associated with systems of discrete particles or as the continuous functions associated with a Vlasov fluid. The interpretation of f [hence of Eq. (13) as the Klimontovich equation or the Vlasov equation] specifies the functional space Ω spanned by f.

Equations (25) should apply when collective or many-particle interactions dominate. It can, in fact, be obtained from the BBGKY hierarchy by neglecting the interactions between the individual particles belonging to the group s. This result can be obtained by subdividing the particle into smaller and smaller units.[14]

It thus appears that for unstable plasma where collective effects certainly play a large role, the s-particle correlations should be obtainable from Eq. (25) when some suitable initial conditions are employed. Since Eq. (25) has as its base the Vlasov equation, this equation should be sufficient for studying unstable situations even though particle correlations become large. The only thing missing is the excitation of disturbances by the particles. However, once the instabilities get going, further excitation of them should make little difference. Also, if turbulence develops and is self-maintaining, this weak tickling should not play much of a role. There will, of course, be some unstable situations where individual particle interactions are important, for example, situations where collisions play an important role in the growth of an instability.

IV. METHODS FOR SOLVING THE VLASOV HIERARCHY

A. Exact Solution

We may first note that the hierarchy allows a product solution[14] of the form

$$\langle f(\mathbf{x}_1) \cdots f(\mathbf{x}_s)\rangle_{\mathrm{av}} = \prod_{i=1}^{s} \bar{f}(\mathbf{x}_i), \quad (26)$$

where $\bar{f}(\mathbf{x}_i)$ satisfies the nonlinear Vlasov equation,

$$\frac{\partial \bar{f}(\mathbf{x})}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial \bar{f}(\mathbf{x})}{\partial \mathbf{q}} - \int \frac{\partial \phi(|\mathbf{q} - \mathbf{q}'|)}{\partial \mathbf{q}} \cdot \frac{\partial \bar{f}(\mathbf{x})}{\partial \mathbf{p}} \bar{f}(\mathbf{x}')\, d\mathbf{x}' = 0. \quad (27)$$

Direct substitution of Eq. (26) into Eq. (25) verifies this. This form is a natural consequence of the form

[14] N. Rostoker and M. Rosenbluth, Phys. Fluids 3, 1 (1960).

of Eq. (25). Physically we see that such a solution should result if $P(df)$ is a delta function, i.e., zero everywhere except at $\langle f(\mathbf{x})\rangle_{\text{av}}$, and the integral of $P(df)$ over a small volume of function space containing $\langle f(\mathbf{x})\rangle_{\text{av}}$ is 1.

While (26) is a solution, it is not the most general one. However, the general solution can be built up from (26). This must be so since all the averages can be found in terms of the general solution to the Vlasov equation. To show this, consider the average value of $f(\mathbf{x}_1)\cdots f(\mathbf{x}_s)$,

$$\left\langle \prod_{i=1}^{s} f(\mathbf{x}_i)\right\rangle_{\text{av}} = \int f(\mathbf{x}_1)\cdots f(\mathbf{x}_s)P^t(df) = \int \prod_{i=1}^{s} T^t f(\mathbf{x}_i, 0)P^0(df). \tag{28}$$

The integral is, in fact, simply a sum of product solutions of the form of (26) with appropriate weighting for their occurrence. Now compute the time derivative of $\left\langle \prod_{i=1}^{s} f(\mathbf{x}_i)\right\rangle_{\text{av}}$,

$$\frac{\partial}{\partial t}\left\langle \prod_{i=1}^{s} f(\mathbf{x}_1)\right\rangle_{\text{av}} = \int \sum_j \frac{\partial f(\mathbf{x}_j, t)}{\partial t} \prod_{\substack{i=1\\ i\neq j}}^{s} T^t f(\mathbf{x}_i, 0)P^0(df). \tag{29}$$

Substituting for the time derivative of $f(\mathbf{x}_j, t)$ from the Vlasov equation gives

$$\begin{aligned}\frac{\partial}{\partial t}\left\langle \prod_{i=1}^{s} f(\mathbf{x}_i)\right\rangle_{\text{av}} &= \int -\sum_j \frac{\mathbf{p}_j}{m}\cdot\frac{\partial}{\partial \mathbf{q}_j}\prod_{i=1}^{s} T^t f(\mathbf{x}_i, 0)P^0(df)\\ &\quad + \iint \sum_j \frac{\partial\phi(|\mathbf{q}_j - \mathbf{q}_{s+1}|)}{\partial \mathbf{q}_j}\cdot\frac{\partial}{\partial \mathbf{p}_j}\\ &\quad \times \prod_{i=1}^{s} T^t f(\mathbf{x}_i, 0)\, d\mathbf{x}_{s+1}P^0(df)\\ &= -\sum_j \frac{\mathbf{p}_j}{m}\cdot\frac{\partial}{\partial \mathbf{q}_j}\left\langle \prod_{i=1}^{s} f(\mathbf{x}_i)\right\rangle_{\text{av}}\\ &\quad + \sum_j \int \frac{\partial\phi(|\mathbf{q}_j - \mathbf{q}_{s+1}|)}{\partial \mathbf{q}_j}\\ &\quad \cdot\frac{\partial}{\partial \mathbf{p}_j}\left\langle \prod_{i=1}^{s+1} f(\mathbf{x}_i)\right\rangle_{\text{av}} d\mathbf{x}_{s+1}.\end{aligned} \tag{30}$$

Equation (30) is just Eq. (25), and this, in fact, gives another derivation of the hierarchy.

Although all solutions can be built up from those of the form of (26), to proceed in this way would gain us nothing. We should have to solve the full nonlinear Vlasov equation, and if we could do that we should know everything there is to know about the problem. The advantage to be gained from a statistical approach—that it gives us only the information we are interested in and ignores the great mass of details which are present—would then be lost. Our primary aim, thus, is to find approximate solutions to (25). There are many approximations which one can make, and to a large extent one's success depends on the accuracy and simplicity of the approximation. The most obvious things to try are the approximations which have been made in treating the BBGKY hierarchy.

B. No Correlations

The simplest thing one could do is to neglect all correlations and assume that all disturbances are small. One then simply obtains the linearized Vlasov equation for f.

C. Weak Correlations Quasi-Linear Theory

The next simplest thing to do is to include two-point correlations but assume they are small and neglect three-point correlations. We thus write for $\langle f(\mathbf{x}_1)f(\mathbf{x}_2)\rangle_{\text{av}}$ and $\langle f(\mathbf{x}_1)f(\mathbf{x}_2)f(\mathbf{x}_3)\rangle_{\text{av}}$ as,

$$\langle f(\mathbf{x}_1)f(\mathbf{x}_2)\rangle_{\text{av}} = \langle f(\mathbf{x}_1)\rangle_{\text{av}}\langle f(\mathbf{x}_2)\rangle_{\text{av}} + G(\mathbf{x}_1, \mathbf{x}_2), \tag{31}$$

$$\begin{aligned}&\langle f(\mathbf{x}_1)f(\mathbf{x}_2)f(\mathbf{x}_3)\rangle_{\text{av}}\\ &\quad = \langle f(\mathbf{x}_1)f(\mathbf{x}_2)f(\mathbf{x}_3)\rangle_{\text{av}} + G(\mathbf{x}_1, \mathbf{x}_2)\langle f(\mathbf{x}_3)\rangle_{\text{av}}\\ &\qquad + G(\mathbf{x}_2, \mathbf{x}_3)\langle f(\mathbf{x}_1)\rangle_{\text{av}} + G(\mathbf{x}_1, \mathbf{x}_3)\langle f(\mathbf{x}_2)\rangle_{\text{av}}.\end{aligned} \tag{32}$$

Substituting these expressions into Eq. (25) gives the following equation for $\langle f\rangle$ and G:

$$\begin{aligned}&\frac{\partial\langle f(\mathbf{x}_1)\rangle_{\text{av}}}{\partial t} + \frac{\mathbf{p}_1}{m}\frac{\partial\langle f(\mathbf{x}_1)\rangle_{\text{av}}}{\partial \mathbf{q}_1} + F(\mathbf{q}_1)\cdot\frac{\partial\langle f(\mathbf{x}_1)\rangle_{\text{av}}}{\partial \mathbf{p}_1}\\ &\qquad - \int \frac{\partial\phi(|\mathbf{q}_1 - \mathbf{q}_2|)}{\partial \mathbf{q}_1}\cdot\frac{\partial G(\mathbf{x}_1, \mathbf{x}_2)}{\partial p_1}\, d\mathbf{x}_2 = 0,\end{aligned} \tag{33}$$

$$\begin{aligned}&\frac{\partial G(\mathbf{x}_1, \mathbf{x}_2)}{\partial t} + \left(\frac{\mathbf{p}_1}{m}\cdot\frac{\partial}{\partial \mathbf{q}_1} + \frac{\mathbf{p}_2}{m}\cdot\frac{\partial}{\partial \mathbf{q}_2}\right)G(\mathbf{x}_1, \mathbf{x}_2)\\ &\quad - \left(\mathbf{F}(\mathbf{q}_1)\cdot\frac{\partial}{\partial \mathbf{p}_1} + \mathbf{F}(\mathbf{q}_2)\cdot\frac{\partial}{\partial \mathbf{p}_2}\right)G(\mathbf{x}_1, \mathbf{x}_2)\\ &\quad - \frac{\partial\langle f(\mathbf{x}_1)\rangle_{\text{av}}}{\partial \mathbf{p}_1}\cdot\int \frac{\partial\phi(|\mathbf{q}_1 - \mathbf{q}_3|)}{\partial \mathbf{q}_1} G(\mathbf{x}_2, \mathbf{x}_3)\, d\mathbf{x}_3\\ &\quad - \frac{\partial\langle f(\mathbf{x}_2)\rangle_{\text{av}}}{\partial \mathbf{p}_2}\cdot\int \frac{\partial\phi(|\mathbf{q}_2 - \mathbf{q}_3|)}{\partial \mathbf{q}_2} G(\mathbf{x}_1, \mathbf{x}_3)\, d\mathbf{x}_3 = 0,\end{aligned} \tag{34}$$

with

$$\mathbf{F}(\mathbf{q}) = \int \frac{\partial\phi(|\mathbf{q} - \mathbf{q}'|)}{\partial \mathbf{q}}\langle f(\mathbf{x}')\rangle_{\text{av}}\, d\mathbf{x}'. \tag{35}$$

Equation (33) is quite similar to the usual Fokker–Planck equation for a plasma. Equation (35) is similar to the usual two-particle correlation function, except that the driving terms, due to discrete particles, are absent.

Equation (34) may be solved by the method of separation of variables. First we note that

$$G(\mathbf{x}_2, \mathbf{x}_1) = G(\mathbf{x}_1, \mathbf{x}_2) \tag{36}$$

because

$$\langle f(\mathbf{x}_1) f(\mathbf{x}_2)\rangle_{\text{av}} = \langle f(\mathbf{x}_2) f(\mathbf{x}_1)\rangle_{\text{av}}.$$

Thus we could write Eq. (34) in the form

$$\begin{aligned}
&\frac{\partial G(\mathbf{x}_1, \mathbf{x}_2)}{\partial t} + \left(\frac{\mathbf{p}_1}{m}\cdot\frac{\partial}{\partial \mathbf{q}_1} + \frac{\mathbf{p}_2}{m}\cdot\frac{\partial}{\partial \mathbf{q}_2}\right) G(\mathbf{x}_1, \mathbf{x}_2) \\
&- \left(\mathbf{F}(\mathbf{q}_1)\cdot\frac{\partial}{\partial \mathbf{p}_1} + \mathbf{F}(\mathbf{q}_2)\cdot\frac{\partial}{\partial \mathbf{p}_2}\right) G(\mathbf{x}_1, \mathbf{x}_2) \\
&- \frac{\partial\langle f(\mathbf{x}_1)\rangle_{\text{av}}}{\partial \mathbf{p}_1}\cdot\int \frac{\partial\phi(|\mathbf{q}_1 - \mathbf{q}_3|)}{\partial \mathbf{q}_1} G(\mathbf{x}_3, \mathbf{x}_2)\, d\mathbf{x}_3 \\
&- \frac{\partial\langle f(\mathbf{x}_2)\rangle_{\text{av}}}{\partial \mathbf{p}_2}\cdot\int \frac{\partial\phi(|\mathbf{q}_2 - \mathbf{q}_3|)}{\partial \mathbf{q}_2} G(\mathbf{x}_1, \mathbf{x}_3)\, d\mathbf{x}_3 = 0.
\end{aligned} \tag{37}$$

Now assume (37) has a solution of the form

$$G(\mathbf{x}_1, \mathbf{x}_2) = X_1(\mathbf{x}_1) X_2(\mathbf{x}_2). \tag{38}$$

Substitution of (38) into (37) gives

$$\begin{aligned}
&\frac{1}{X_1(\mathbf{x}_1)}\left(\frac{\partial X_1(\mathbf{x}_1)}{\partial t} + \frac{\mathbf{p}_1}{m}\cdot\frac{\partial}{\partial \mathbf{q}_1} X_1(\mathbf{x}_1) - \mathbf{F}(\mathbf{q}_1)\cdot\frac{\partial X_1(\mathbf{x}_1)}{\partial \mathbf{p}_1}\right. \\
&\qquad \left. - \frac{\partial\langle f(\mathbf{x}_1)\rangle_{\text{av}}}{\partial \mathbf{p}_1}\cdot\int \frac{\partial\phi(|\mathbf{q}_1 - \mathbf{q}_3|)}{\partial \mathbf{q}_1} X_1(\mathbf{x}_3)\, d\mathbf{x}_3\right) \\
&+ \frac{1}{X_2(\mathbf{x}_2)}\left(\frac{\partial X_2(\mathbf{x}_2)}{\partial t} + \frac{\mathbf{p}_2}{m}\cdot\frac{\partial}{\partial \mathbf{q}_2} X_2(\mathbf{x}_2) - \mathbf{F}(\mathbf{q}_2)\right. \\
&\qquad \cdot\frac{\partial X_2(\mathbf{x}_2)}{\partial \mathbf{p}_2} - \frac{\partial\langle f(\mathbf{x}_2)\rangle_{\text{av}}}{\partial \mathbf{p}_2}\cdot\int \frac{\partial\phi(|\mathbf{q}_2 - \mathbf{q}_3|)}{\partial \mathbf{q}_2} \\
&\qquad \left.\times X_2(\mathbf{x}_3)\, d\mathbf{x}_3\right) = 0.
\end{aligned} \tag{39}$$

Equation (39) is a sum of two factors, one depending on $\mathbf{x}_1$ and the other on $\mathbf{x}_2$. Thus, one of these factors must be equal to the constant γ and the other to $-\gamma$.

We may write (39) in the form

$$\begin{aligned}
\frac{d}{dt} X_1(\mathbf{x}_1)\Big|_{\text{evaluated along 0}} &= \frac{\partial X_1(\mathbf{x}_1)}{\partial t} + \frac{\mathbf{p}_1}{m}\cdot\frac{\partial}{\partial \mathbf{q}_1} X_1(\mathbf{x}_1) \\
&\quad - \mathbf{F}(\mathbf{q}_1)\cdot\frac{\partial X_1(\mathbf{x}_1)}{\partial \mathbf{p}_1} \\
&= \frac{\partial\langle f(\mathbf{x}_1)\rangle_{\text{av}}}{\partial \mathbf{p}_1}\cdot\int \frac{\partial\phi(|\mathbf{q}_1 - \mathbf{q}_3|)}{\partial \mathbf{q}_1} \\
&\quad \times X_1(\mathbf{x}_3)\, d\mathbf{x}_3 - \gamma X_1(\mathbf{x}_1), \\
\frac{dX_2(\mathbf{x}_2)}{dt}\Big|_{\text{evaluated along 0}} &= \frac{\partial\langle f(\mathbf{x}_2)\rangle_{\text{av}}}{\partial \mathbf{p}_2}\cdot\int \frac{\partial\phi(|\mathbf{q}_2 - \mathbf{q}_3|)}{\partial \mathbf{q}_2} \\
&\quad \times X_2(\mathbf{x}_3)\, d\mathbf{x}_3 + \gamma X_2(\mathbf{x}_2).
\end{aligned} \tag{40}$$

Here 0 represents the trajectories which give $\langle f(\mathbf{x})\rangle_{\text{av}}$. Integrating gives

$$\begin{aligned}
X_1(\mathbf{x}_1, t) &= X_1(\mathbf{x}_1, 0)\, e^{-\gamma t} + \int_0^t \exp[-\gamma(t - t')] \\
&\qquad \text{integrating along trajectories} \\
&\quad \times \left(\frac{\partial\langle f(\mathbf{x}_1)\rangle_{\text{av}}}{\partial \mathbf{p}_1}\cdot\int \frac{\partial\phi(|\mathbf{q}_1 - \mathbf{q}_3|)}{\partial \mathbf{q}_1} X_1(\mathbf{x}_3)\, d\mathbf{x}_3\right) dt', \\
X_2(\mathbf{x}_2, t) &= X_2(\mathbf{x}_2, 0) e^{\gamma t} + \int_0^t \exp[\gamma(t - t')] \\
&\qquad \text{integrating along trajectories} \\
&\quad \times \left(\frac{\partial\langle f(\mathbf{x}_2)\rangle_{\text{av}}}{\partial \mathbf{p}_2}\cdot\int \frac{\partial\phi(|\mathbf{q}_2 - \mathbf{q}_3|)}{\partial \mathbf{q}_2} X_2(\mathbf{x}_3)\, d\mathbf{x}_3\right) dt'.
\end{aligned} \tag{41}$$

Since we are only interested in the product $X_1(\mathbf{x}_1) X_2(\mathbf{x}_2)$, we see we may choose γ to be zero.[15] Equations (40) and (41) then just give the solutions to the linearized Vlasov equation.

It should, perhaps, be noted here that the product $X_1(\mathbf{x}_1) X_2(\mathbf{x}_2)$ is not symmetric in $\mathbf{x}_1$ and $\mathbf{x}_2$. However, it is easy to construct symmetric solutions from these products, i.e., $[X_1(\mathbf{x}_1) X_2(\mathbf{x}_2) + X_1(\mathbf{x}_2) X_2(\mathbf{x}_1)]$. We must allow only symmetric combinations in the final solution. If we start with symmetric solutions, they remain symmetric.

We may build up the solution to the general initial-value problem for Eq. (37) from these product solutions. First we write the initial value of G as a sum of products of delta functions,

$$\begin{aligned}
G(\mathbf{x}_1, \mathbf{x}_2, t = 0) = \iint &G(\mathbf{x}', \mathbf{x}'', t = 0) \\
&\times \delta(\mathbf{x}_1 - \mathbf{x}')\delta(\mathbf{x}_2 - \mathbf{x}'')\, d\mathbf{x}'\, d\mathbf{x}''.
\end{aligned} \tag{42}$$

Next let $\chi(\mathbf{x}_1, t; \mathbf{x}')$ and $\chi(\mathbf{x}_2, t; \mathbf{x}'')$ be the solutions given by (41) for which $X_1(\mathbf{x}_1, 0)$ and $X_2(\mathbf{x}_2, 0)$ are $\delta(\mathbf{x}_1 - \mathbf{x}')$ and $\delta(\mathbf{x}_2 - \mathbf{x}'')$, respectively. Then the general solution to Eq. (37) is given by

$$\begin{aligned}
G(\mathbf{x}_1, \mathbf{x}_2, t) = \iint &G(\mathbf{x}', \mathbf{x}'', t = 0) \\
&\times \chi(\mathbf{x}_1, t; \mathbf{x}')\chi(\mathbf{x}_2, t; \mathbf{x}'')\, d\mathbf{x}'\, d\mathbf{x}''.
\end{aligned} \tag{43}$$

If $G(\mathbf{x}_1, \mathbf{x}_2, t = 0)$ is symmetric in $\mathbf{x}_1$, $\mathbf{x}_2$, then (43) will be too.

We may note that the above solution holds even if $\langle f(\mathbf{x})\rangle_{\text{av}}$ is a function of t. If $\langle f(\mathbf{x}_1)\rangle_{\text{av}}$ is time independent or slowly varying with t the above method is

[15] *Note added in proof:* This is perhaps most easily seen by noting that the solutions can be written as $X_1(\mathbf{x}_1, t) = e^{-\gamma t}\xi_1(\mathbf{x}_1, t)$, $X_2(\mathbf{x}_2, t) = e^{\gamma t}\xi_2(\mathbf{x}_2, t)$, where ξ_1 and ξ_2 are the solutions for $\gamma = 0$.

equivalent to Dupree's[16] operational method for solving the second member of the BBGKY hierarchy. In fact the same method of separation of variables may be employed to obtain his result. The only difference is that there are source terms due to the discrete particles. These are, however, readily included once we know the solutions of the homogeneous equation.

As given above, Eqs. (33) and (34) would be difficult to solve, for they apply to the general case where $\langle f(\mathbf{x})\rangle_{\mathrm{av}}$ may be spatially nonuniform and time varying. However, they can be applied to many types of systems; uniform and nonuniform plasmas, and clusters of stars. One case which is very often considered is that in which $\langle f(\mathbf{x})\rangle_{\mathrm{av}}$ is spatially uniform and is either stationary in time or varies only slowly in time. In this case G may be solved for adiabatically in terms of the variations of $\langle f(\mathbf{x})\rangle_{\mathrm{av}}$, and one obtains quasi-linear theory without mode coupling.

D. Other Approximations

There are many other methods of terminating the hierarchy which one can think of. One such method is to employ the superposition approximation which Kirkwood originally proposed for use in statistical mechanics. Here one approximates the three-point distribution function in terms of the two by

$$\langle f(\mathbf{x}_1)f(\mathbf{x}_2)f(\mathbf{x}_3)\rangle_{\mathrm{av}} = \frac{\langle f(\mathbf{x}_1)f(\mathbf{x}_2)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_1)f(\mathbf{x}_3)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_2)f(\mathbf{x}_3)\rangle_{\mathrm{av}}}{\langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_2)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_3)\rangle_{\mathrm{av}}} \tag{44}$$

We may write

$$\langle f(\mathbf{x}_1)f(\mathbf{x}_2)\rangle_{\mathrm{av}} = \langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_2)\rangle_{\mathrm{av}}[1 + G(\mathbf{x}_1, \mathbf{x}_2)]. \tag{45}$$

Equation (44) then becomes

$$\begin{aligned}\langle f(\mathbf{x}_1)f(\mathbf{x}_2)f(\mathbf{x}_3)\rangle_{\mathrm{av}} &= \langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_2)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_3)\rangle_{\mathrm{av}}[1 + G(\mathbf{x}_1, \mathbf{x}_2) \\ &\quad + G(\mathbf{x}_1, \mathbf{x}_3) + G(\mathbf{x}_2, \mathbf{x}_3) + G(\mathbf{x}_1, \mathbf{x}_2)G(\mathbf{x}_1, \mathbf{x}_3) \\ &\quad + G(\mathbf{x}_1, \mathbf{x}_2)G(\mathbf{x}_2, \mathbf{x}_3) + G(\mathbf{x}_1, \mathbf{x}_3)G(\mathbf{x}_2, \mathbf{x}_3) \\ &\quad + G(\mathbf{x}_1, \mathbf{x}_2)G(\mathbf{x}_1\mathbf{x}_3)G(\mathbf{x}_2\mathbf{x}_3)].\end{aligned} \tag{46}$$

One further approximation we could make is to assume that G is relatively small and neglect terms containing three G's. If we make this approximation and further assume that $\langle f(\mathbf{x})\rangle_{\mathrm{av}}$ is spatially uniform, the first two members of the hierarchy become

$$\begin{aligned}&\frac{\partial\langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}}{\partial t} \\ &- \int \frac{\partial\phi(|\mathbf{q}_1 - \mathbf{q}_2|)}{\partial\mathbf{q}_1}\cdot\frac{\partial}{\partial\mathbf{p}_1} f(\mathbf{x}_1)f(\mathbf{x}_2)G(\mathbf{x}_1, \mathbf{x}_2)\, d\mathbf{x}_2 \\ &\times \frac{\partial}{\partial t}\langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_2)\rangle_{\mathrm{av}}G(\mathbf{x}_1, \mathbf{x}_2) \\ &+ \left(\frac{\mathbf{p}_1}{m}\cdot\frac{\partial}{\partial\mathbf{p}_1} + \frac{\mathbf{p}_2}{m}\cdot\frac{\partial}{\partial\mathbf{q}_2}\right)\langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_2)\rangle_{\mathrm{av}}G(\mathbf{x}_1, \mathbf{x}_2) \\ &- \int \frac{\partial\phi(|\mathbf{q}_2 - \mathbf{q}_3|)}{\partial\mathbf{q}_2}\cdot\frac{\partial}{\partial\mathbf{p}_2} \\ &\times \langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_2)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_3)\rangle_{\mathrm{av}}G(\mathbf{x}_1, \mathbf{x}_3)\, d\mathbf{x}_3 \\ &- \int \frac{\partial\phi(|\mathbf{q}_1 - \mathbf{q}_3|)}{\partial\mathbf{q}_1}\langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_2)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_3)\rangle_{\mathrm{av}} \\ &\times G(\mathbf{x}_2, \mathbf{x}_3)\, d\mathbf{x}_3 \\ &- \int \left(\frac{\partial\phi(|\mathbf{q}_1 - \mathbf{q}_3|)}{\partial\mathbf{q}_1}\cdot\frac{\partial}{\partial\mathbf{p}_1} + \frac{\partial\phi(|\mathbf{q}_2 - \mathbf{q}_3|)}{\partial\mathbf{q}_2}\cdot\frac{\partial}{\partial\mathbf{p}_2}\right) \\ &\times \langle f(\mathbf{x}_1)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_2)\rangle_{\mathrm{av}}\langle f(\mathbf{x}_3)\rangle_{\mathrm{av}},\end{aligned} \tag{47}$$

$$\sum_{\substack{ijk\\ i\neq j\\ i\neq k\\ j\neq k}} G(\mathbf{x}_i, \mathbf{x}_j)G(\mathbf{x}_j, \mathbf{x}_k)\, d\mathbf{x}_3. \tag{48}$$

If terms quadratic in G are neglected, then this reduces to the previous case. When these terms are kept, they give mode coupling terms. If it is further assumed that $\langle f(\mathbf{x})\rangle_{\mathrm{av}}$ varies slowly with respect to G, then this equation should be similar to quasi-linear theory with mode coupling (the equivalence has not been shown, however).

ACKNOWLEDGMENTS

The authors would like to express sincere thanks to Dr. T. Tatsumi and Dr. T. Taniuti for their helpful discussions and suggestions. One of the authors (J. M. D.) wishes to acknowledge indebtedness to the Institute of Plasma Physics at Nagoya University for the warm, friendly, and stimulating atmosphere it provided during his stay there when much of this work was carried out. He is also indebted to Dr. C. Oberman for many stimulating discussions.

One of the authors (J. M. D.) wishes to acknowledge partial support from a Fulbright grant and partial support from the United States Air Force, Contract AF49(638)-1555.

[16] T. H. Dupree, Phys. Fluids **4**, 696 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Phase Problem in Coherence Theory

H. M. NUSSENZVEIG*
Department of Physics and Astronomy, University of Rochester, Rochester, New York

(Received 6 May 1966)

For a class of quasi-monochromatic spectra, including Lorentzian and Gaussian line shapes, it is shown that knowledge of the modulus of the complex degree of temporal coherence $\gamma(\tau)$ does not suffice to reconstruct the spectrum. This is due to the existence of zeros of $\gamma(\tau)$ in the complex τ plane, giving rise to a significant contribution to the phase of $\gamma(\tau)$. The position of the zeros and their physical interpretation are investigated. The case of band-limited spectra is also treated, and some general properties of the distribution of zeros in this case are given.

I. INTRODUCTION

LET a light beam be split into two parts, which are later reunited, after a time delay τ has been introduced into one of them. A sufficiently accurate measurement of the resultant intensity as a function of τ would enable one to determine the energy spectrum of the light beam. This is the basis of Michelson's method of interference spectroscopy.[1]

In fact, we have[1–4]

$$\gamma(\tau) = \int_0^\infty g(\omega) e^{-i\omega\tau}\, d\omega, \tag{1.1}$$

where $\gamma(\tau)$ is the *complex degree of temporal coherence* of the light beam and $g(\omega)$ is the *spectral density*, which is real and nonnegative:

$$g(\omega) \geq 0. \tag{1.2}$$

The experiment described above would lead to a measurement of $\mathrm{Re}\,\gamma(\tau)$. According to (1.1), $g(\omega)$ might then be obtained by taking the inverse cosine transform. This corresponds to the Wiener–Khintchine theorem for stochastic processes: the autocorrelation function and the power spectrum are Fourier transforms of each other.

The function $\gamma(\tau)$ is usually normalized so that

$$\gamma(0) = 1. \tag{1.3}$$

The integral in (1.1) extends only over positive frequencies. In the classical theory of coherence, this arises from the fact that the field is real, so that, in its Fourier representation, the negative-frequency components are the complex conjugates of the positive-frequency ones. This allows one to eliminate negative frequencies by working with analytic signals.[1] In the quantum theory of coherence,[2,4] the same restriction arises from the preferential role played by the annihilation operators for the field in the photoelectric detection process.

For a quasi-monochromatic beam, $\mathrm{Re}\,\gamma(\tau)$ is a rapidly oscillating function, with mean period given by the average beam frequency. Thus, direct measurements become quite difficult at high frequencies, although they have been performed in the far infrared.[5,6]

It is much easier to measure the envelope of the rapid oscillations, which is a slowly varying function (it is proportional to the *visibility* of the interference fringes[1]). If we write

$$\gamma(\tau) = |\gamma(\tau)| \exp\,[i\varphi(\tau)], \tag{1.4}$$

it is readily seen that $|\gamma(\tau)|$ can be taken as defining the envelope, while $\varphi(\tau)$ gives the phase of the oscillations.

Even for very narrow spectra, it is often possible to employ correlation techniques, such as the Hanbury Brown–Twiss effect,[7,8] to measure $|\gamma|$, but it would be very difficult to measure φ.

The question then arises whether knowledge of $|\gamma|$ alone is sufficient to reconstruct the spectral density $g(\omega)$. This is analogous to the well-known phase problem in x-ray diffraction,[9] and similar problems arise in the theory of image reconstruction[10] and in the Fock–Krylov formulation of the quantum theory of decay.[11]

When the spectrum is known to be symmetric

* On leave of absence from Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil.

[1] M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, London, 1959), Secs. 7.5.8 and 10.4.1.

[2] L. Mandel and E. Wolf, Rev. Mod. Phys. **37**, 231 (1965).

[3] E. Wolf, Japan. J. Appl. Phys. **4**, Suppl. I, 1 (1965).

[4] R. J. Glauber, Phys. Rev. **131**, 2766 (1963), Eq. (10.16).

[5] J. Strong and G. A. Vanasse, J. Opt. Soc. Am. **49**, 844 (1959).

[6] P. Jacquinot, Rept. Progr. Phys. **23**, 267 (1960).

[7] L. Mandel, in *Proceedings of the Symposium on Electromagnetic Theory and Antennas* (Pergamon Press, London, 1963), p. 811.

[8] M. L. Goldberger, H. W. Lewis, and K. M. Watson, Phys. Rev. **142**, 25 (1966).

[9] M. L. Goldberger, H. W. Lewis, and K. M. Watson, Phys. Rev. **132**, 2764 (1963).

[10] A. Walther, Opt. Acta **10**, 41 (1963).

[11] L. A. Khalfin, Zh. Eksperim. i Teor. Fiz. **33**, 1371 (1957) [English transl.: Soviet Phys.—JETP **6**, 1053 (1958)].

about its mean frequency $\bar{\omega}$, it is possible to reconstruct it if $|\gamma|$ and $\bar{\omega}$ are known, because the phase is then given essentially by a trivial factor $\exp(-i\bar{\omega}\tau)$ (cf. Sec. III).

It has been suggested by Wolf[12] that it may be possible to solve the phase problem also in more general cases, by taking into account the analytic properties of $\gamma(\tau)$ that follow from (1.1) and the nonnegative definiteness condition (1.2).

Since the integral in (1.1) contains only positive frequencies, $\gamma(\tau)$ can be analytically continued as a regular function of τ in the lower half of the complex τ plane $I_-(\tau)$. The fact that $g(\omega)$ is real implies

$$\gamma(-\tau) = \gamma^*(\tau^*). \tag{1.5}$$

If we assume that the Paley–Wiener condition

$$\int_{-\infty}^{\infty} \frac{|\ln|\gamma(\tau)||}{1+\tau^2}\, d\tau < \infty \tag{1.6}$$

is satisfied, it is possible to express the phase $\varphi(\tau)$ in terms of $\ln|\gamma(\tau)|$ by a dispersion relation[12]

$$\varphi(\tau) = \varphi_M(\tau) + \varphi_B(\tau), \tag{1.7}$$

where

$$\varphi_M(\tau) = \frac{2}{\pi}\tau P\int_0^{\infty} \frac{\ln|\gamma(\tau')|}{\tau'^2-\tau^2}\, d\tau', \tag{1.8}$$

the symbol P denoting the Cauchy principal value, and

$$\varphi_B(\tau) = \sum_n \arg\left(\frac{\tau-\tau_n}{\tau-\tau_n^*}\right), \tag{1.9}$$

where the sum is extended over all the zeros τ_n of $\gamma(\tau)$ in $I_-(\tau)$. This corresponds to the representation

$$\gamma(\tau) = |\gamma(\tau)|\exp[i\varphi_M(\tau)]\prod_n\left(\frac{\tau-\tau_n}{\tau-\tau_n^*}\right). \tag{1.10}$$

The last factor, which represents the contribution from the zeros, is known as a *Blaschke product*, and we call $\varphi_B(\tau)$ the *Blaschke phase*.

Since $\varphi_B(\tau) \geq 0$ with a proper definition of the phase (cf. Sec. II), the function $\varphi_M(\tau)$ given by (1.8) represents the *minimal phase*. Furthermore,

$$\frac{d\varphi_B}{d\tau} = 2\sum_n \frac{\operatorname{Im}\tau_n}{(\tau-\operatorname{Re}\tau_n)^2+(\operatorname{Im}\tau_n)^2} \leq 0, \tag{1.11}$$

so that

$$d\varphi/d\tau \leq d\varphi_M/d\tau. \tag{1.12}$$

The phase problem is therefore reduced to the determination of the zeros τ_n. According to (1.5), if τ_n is a zero, so is $-\tau_n^*$, so that the distribution of zeros is symmetrical with respect to the imaginary axis. It follows from (1.2) that there cannot be any zeros on the imaginary axis,[13] so that all the zeros must occur in pairs.

It has been conjectured by Wolf[12] that, for light emitted by ordinary sources, $\gamma(\tau)$ may have no zeros at all in I_-, or at least in that part of I_- which would significantly affect the reconstruction of the spectrum. On the other hand, if zeros are present, they should have some physical significance, the understanding of which would be of considerable importance for interference spectroscopy. In the absence of zeros, the solution to the phase problem would be given by the minimal phase.

This conjecture was supported by an explicit calculation of $\gamma(\tau)$ for blackbody radiation.[14] It was shown that there are no zeros in I_- in this case, so that knowledge of $|\gamma(\tau)|$ suffices to reconstruct the spectrum, in spite of its not being symmetric.

It has also been shown[15] that, in the case of complete coherence, i.e., if $|\gamma(\tau)| = 1$ for all τ, condition (1.2) leads to a unique solution, namely a perfectly monochromatic beam: $\gamma(\tau) = \exp(-i\omega_0\tau)$.

These examples correspond to extreme cases: a very broad spectrum (blackbody radiation) and an infinitely narrow one (monochromatic radiation). The case of greatest practical interest is that of quasi-monochromatic radiation, such as that from an optical maser.

The following questions arise in the quasi-monochromatic case: (i) Are there any zeros? If so: (ii) What is their location? (iii) What is their physical interpretation? (iv) Do they give a significant contribution to the phase or is the minimal phase a good approximation for reconstructing the spectrum?

The present paper is concerned with the investigation of these questions. There are two quite different approaches that may be taken in such an investigation. The first one is to deal with specific classes or examples of quasi-monochromatic spectra given *a priori* and assumed to be physically realizable. The second and far more ambitious approach would be trying to derive the relevant features of the spectrum from a physical model of the source, taking into account the statistical features implied in the definition of the spectral density. As a further refinement, the effect of the measurement process might also be considered.

Only the first approach is taken here. The classes of spectra that are treated include the Lorentzian and Gaussian line shapes as special cases, and particular

[12] E. Wolf, Proc. Phys. Soc. (London) **80**, 1269 (1962).

[13] P. Roman and A. S. Marathay, Nuovo Cimento **30**, 1452 (1963).

[14] Y. Kano and E. Wolf, Proc. Phys. Soc. (London) **80**, 1273 (1962).

[15] C. L. Mehta, E. Wolf, and A. P. Balachandran, J. Math. Phys. **7**, 133 (1966).

attention is devoted to them. In all the cases treated, it is shown that there exists a large number of zeros. Their location is determined and their physical interpretation discussed. It turns out, in all these cases, that most of the phase information is contained in the zeros and very little in the minimal phase, so that knowledge of $|\gamma(\tau)|$ alone is not sufficient to reconstruct the spectrum.

The classes of spectra to be considered certainly do not encompass all possible quasi-monochromatic spectra, so that the results need not apply to all cases. However, they indicate that, in general, one cannot expect the minimal phase to be a good approximation to the solution of the phase problem. To determine to what extent these results can be applied to actual light beams, and to make further progress in the solution, a deeper investigation, based on the second approach, would seem necessary.

The distribution of zeros depends not only on the behavior of the spectrum near the main peaks, but also far away from them, and in particular on its behavior near the end points of the region where it is nonvanishing. In Sec. II, we consider quasi-monochromatic spectra extending over all frequencies, from zero to infinity. In Sec. III, we consider the case of band-limited spectra, which is of particular importance in connection with the phase problem in x-ray diffraction. Some general theorems about the distribution of zeros are given. The main conclusions are summarized in Sec. IV.

II. QUASI-MONOCHROMATIC SPECTRA

A spectrum is called quasi-monochromatic if the spectral density $g(\omega)$ takes on appreciable values only for

$$|\omega - \omega_0| \leqslant \delta, \tag{2.1}$$

where ω_0 is the midfrequency, and

$$\epsilon = \delta/\omega_0 \ll 1. \tag{2.2}$$

In typical cases, $\epsilon \leqslant 10^{-6}$ for thermal light; for laser light, it can be several orders of magnitude smaller.

We consider in this section several examples of quasi-monochromatic spectra for which $\gamma(\tau)$ has a large number of zeros in I_-. Before discussing them in detail, we shall give a simple argument that relates the existence of these zeros to some common features of the spectra.

The spectra considered have a (very small) tail extending down to $\omega = 0$. In practice, the low-frequency behavior of the spectrum will be affected by the frequency response of the measuring device, which acts as a bandpass filter. It is largely a matter of taste whether one considers the present model or that of a band-limited spectrum, treated in Sec. III, as a more faithful representation of an actual situation. We shall see that the basic results are very similar in both cases.

To determine the number of zeros of $\gamma(\tau)$ in I_-, we apply the well-known theorem[16] according to which the number of zeros of the regular analytic function $\gamma(\tau)$ within a contour C is

$$N = \Delta_C \arg \gamma(\tau)/2\pi, \tag{2.3}$$

where $\Delta_C \arg \gamma$ denotes the variation of $\arg \gamma$ round the contour C. We take as contour C the segment $-T \leq \tau \leq T$ of the real axis, closed by a half-circle of radius T in $I_-(\tau)$, where $T \to \infty$.

Let us rewrite (1.1) as

$$\gamma(\tau) = \exp(-i\omega_0\tau) \int_{-\omega_0}^{\infty} g(\omega_0 + \eta) \exp(-i\tau\eta)\, d\eta. \tag{2.4}$$

According to (2.1) and (2.2), during the time interval for which $\gamma(\tau)$ is appreciable, i.e., for $|\tau| \leqslant \tau_0$, with

$$\tau_0 \gtrsim \tau_c = \delta^{-1}, \tag{2.5}$$

where τ_c is the coherence time, the integral in (2.4) is a slowly varying function as compared with the exponential factor, so that most of the phase variation is given by this factor:

$$\Delta \arg \gamma(\tau) \approx -\omega_0 \Delta\tau \quad (-\tau_0 \leqslant \tau \leqslant \tau_0). \tag{2.6}$$

The asymptotic behavior of $\gamma(\tau)$ for very large times depends on the behavior of $g(\omega)$ for $\omega \to 0$. We assume for the sake of simplicity that

$$g(\omega) \approx A\omega^{\alpha} \quad \text{for} \quad \omega \to 0, \tag{2.7}$$

where $\alpha \geq 0$ and A are constants. It then follows from the Abelian theorem on the asymptotic behavior of Laplace (Fourier) transforms[17] that

$$\gamma(\tau) \sim A\Gamma(\alpha + 1)/(i\tau)^{\alpha+1} \quad \text{for} \quad |\tau| \to \infty \text{ in } I_-. \tag{2.8}$$

Thus, for sufficiently large times, we no longer have oscillatory behavior, but only a slowly varying (algebraic) decay. It follows from (2.8) that the variation of $\arg \gamma$ around the half-circle of radius T, for $T \to \infty$, is

$$\Delta \arg \gamma(\tau) = -(\alpha + 1)\Delta \arg \tau = (\alpha + 1)\pi. \tag{2.9}$$

If we now assume that the transition from oscillatory behavior to the asymptotic behavior (2.8) takes place smoothly, for $\tau \gtrsim \tau_0$, without introducing large additional phase variations, it follows

[16] E. C. Titchmarsh, *The Theory of Functions* (Oxford University Press, London, 1939), 2nd ed., p. 116.

[17] G. Doetsch, *Handbuch der Laplace-Transformation* (Verlag Birkhäuser, Basel, 1950), Vol. I, p. 473.

from (2.3), (2.6), and (2.9) that

$$N \sim (\omega_0\tau_0/\pi) - \tfrac{1}{2}(\alpha + 1). \tag{2.10}$$

According to (2.5) and (2.2),

$$\omega_0\tau_0 \gtrsim \omega_0\tau_c \sim \omega_0/\delta = \epsilon^{-1} \tag{2.11}$$

and we have, typically, $\epsilon^{-1} \gtrsim 10^6$. Thus, unless α is also of this order of magnitude in (2.7), the number of zeros N is very large. We assume that α is much smaller; in fact, $\alpha = 0$ in the examples to be considered, but nothing would be essentially changed if we took $\alpha \gg 1$, provided that $\alpha \ll \epsilon^{-1}$.

It should be emphasized that the above result gives only an order-of-magnitude estimate of the number of zeros, based on the assumption that there are no compensating phase changes for $|\tau| \gtrsim \tau_0$. This assumption is verified in the examples discussed below. The large number of zeros then arises from the large phase change that takes place within a few coherence times. It is by no means implied, however, that the same result is valid for any arbitrary quasi-monochromatic spectrum.

Within this approximation, one can also say something about the distribution of zeros: most of them must be located within a half-circle with radius of the order of τ_0, centered at the origin. To see this, it suffices to take $T \sim \tau_0$ in the contour C considered above.

To find out more about the location of the zeros and their effect on the phase of $\gamma(\tau)$, we now turn to specific examples.

Example 1: Lorentzian Peak. This corresponds to taking

$$g(\omega) = B\delta/[(\omega - \omega_0)^2 + \delta^2] \tag{2.12}$$

in (1.1). This spectrum is usually associated with the emission of a single line. The constant B is determined by the normalization condition (1.3), which gives $B \approx 1/\pi$ for $\epsilon = \delta/\omega_0 \ll 1$; we take $B = 1/\pi$ for simplicity.

Let us investigate the behavior of $\gamma(\tau)$ in $I_-(\tau)$. According to the symmetry relation (1.5), it suffices to consider the behavior in the fourth quadrant. Substituting (2.12) in (2.4), and introducing the dimensionless variables

$$\omega_0\tau = z = x + iy, \quad \epsilon = \delta/\omega_0, \tag{2.13}$$

we find

$$\gamma(z) = \frac{\epsilon}{\pi} e^{-iz} \int_{-1}^{\infty} \frac{e^{-iuz}}{u^2 + \epsilon^2}\, du. \tag{2.14}$$

The evaluation of this integral is undertaken in Appendix A.

The behavior of $\gamma(z)$ for $|z| \ll 1$ follows from (A3)–(A6):

$$\gamma(z) = 1 + (i\epsilon/\pi)z \ln z - i\{1 + (\epsilon/\pi)[1 - C - \tfrac{1}{2}i\pi - \tfrac{1}{2}\ln(1 + \epsilon^2)]\}z + O(z^2 \ln z) \quad (|z| \ll 1). \tag{2.15}$$

For $|z| \gg 1$, employing (A7), we find

$$\gamma(z) = e^{-iz-\epsilon z} - [i\epsilon/\pi(1 + \epsilon^2)z] \times \{1 - [2i/(1 + \epsilon^2)z] + O(z^{-2})\} \quad \text{for} \quad z \in 1, |z| \gg 1. \tag{2.16}$$

For $z \in 2$, the first term on the right-hand side is to be omitted (regions 1 and 2 are defined in Appendix A).

In particular, along the positive real axis ($z = x > 0$), (2.16) gives the behavior of $\gamma(\tau)$ for times much larger than the mean period ($\tau \gg \omega_0^{-1}$). The first term has the familiar exponentially decaying behavior associated with the Lorentzian spectrum, with lifetime $\tau_c = \delta^{-1}$ (coherence time). The remaining terms represent corrections to the exponential decay law, arising from the restriction of (1.1) to positive frequencies only. The correction terms decay much more slowly than the exponential, following an inverse power law, so that they must ultimately become dominant.

According to (2.16), the transition from one decay law to the other takes place for $\exp(-\epsilon x_0) \approx (\pi x_0)^{-1}\epsilon$ or, since $\epsilon \ll 1$, for

$$x = x_0 \approx \epsilon^{-1} \ln(\pi\epsilon^{-2}). \tag{2.17}$$

Thus, the transition time is given by $\tau_0 \sim \ln(\pi\epsilon^{-2})\tau_c$, which agrees with (2.11). For $\epsilon \lesssim 10^{-6}$, we find $\tau_0 \gtrsim 30\tau_c$, so that the degree of coherence would be unobservable, in practice, for $\tau \gtrsim \tau_0$.

The nonexponential nature of the decay law for very large times as a consequence of the one-sidedness of the spectrum is a well-known effect in the decay of unstable particles.[11,18–20]

Let us now determine the zeros of $\gamma(z)$ in the fourth quadrant. It is readily seen that there are no zeros with $|z| \lesssim 1$ and no zeros in region 2, so that it suffices to consider $|z| \gg 1$ in the region 1. For $\epsilon \ll 1$, according to (2.16), the zeros are approximately given by the roots

$$z_n = x_n + iy_n = r_n \exp(i\theta_n) \quad (-\tfrac{1}{2}\pi < \theta_n \leq 0)$$

of

$$\exp(-iz - \epsilon z) = i\epsilon/(\pi z). \tag{2.18}$$

[18] M. Lévy, Nuovo Cimento **14**, 612 (1959).
[19] G. Beck and H. M. Nussenzveig, Nuovo Cimento **16**, 416 (1960).
[20] J. Schwinger, Ann. Phys. (N.Y.) **9**, 169 (1960).

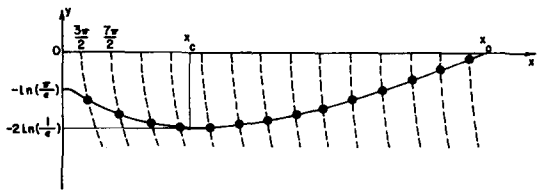


FIG. 1. The zeros of the complex degree of coherence for a Lorentzian peak. •—zeros.

Equating modulus and phase of both sides, we get

$$\epsilon x - y = \ln(\pi/\epsilon) + \tfrac{1}{2}\ln(x^2 + y^2), \quad (2.19)$$

$$x_n + \epsilon y_n = 2n\pi - \tfrac{1}{2}\pi + \theta_n \quad (n = 1, 2, 3, \cdots). \quad (2.20)$$

The zeros[21] are located at the intersections of the curve (2.19), drawn in full line in Fig. 1, with the family of curves (2.20), drawn in broken line in Fig. 1. They are approximately equally spaced, with spacing $|\Delta z| \sim 2\pi$. The total number of zeros N is given by

$$N \sim x_0/\pi \approx -2(\pi\epsilon)^{-1}\ln\epsilon \gg x_c = \epsilon^{-1}, \quad (2.21)$$

in agreement with (2.10).

For $x \ll x_c$, the zeros are relatively far from the real axis, but a large number of them falls within this interval, so that they can have a large effect on the phase of the degree of coherence. This indeed happens, as will now be seen.

The evaluation of the minimal phase $\varphi_M(x)$ is carried out in Appendix B. The results are contained in (B5)–(B7), and they are represented by the curve in broken line in Fig. 2.

The actual phase $\varphi(x)$ is given by (B8)–(B10) and it is represented by the curve in full line in Fig. 2. The only points it has in common with $\varphi_M(x)$ are the vertical tangent at $x = 0$ and the asymptotic limit $-\frac{1}{2}\pi$ as $x \to \infty$ (however, φ approaches this limit much faster than φ_M). The nearly linear behavior of $\varphi(x)$ up to $x \sim x_0$ is due to the near symmetry of the spectrum. Note that $\varphi_M \leq \varphi$ and $d\varphi_M/dx \geq d\varphi/dx$, in agreement with (1.12).

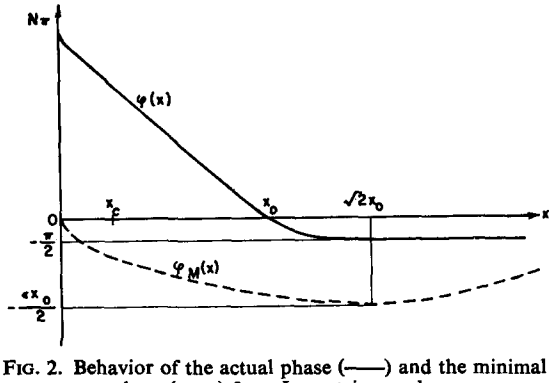


FIG. 2. Behavior of the actual phase (——) and the minimal phase (– – –) for a Lorentzian peak.

Since φ and φ_M have very little in common, most of the phase information must be contained in the Blaschke phase φ_B. How does this affect the reconstruction of the spectrum?

It follows from (1.1) and (1.5) that

$$g(\omega) = \frac{1}{\pi}\,\mathrm{Re}\int_0^\infty |\gamma(\tau)|\exp[i\varphi(\tau) + i\omega\tau]\,d\tau. \quad (2.22)$$

Let $g_M(\omega)$ denote the function obtained when $\varphi(\tau)$ is replaced by $\varphi_M(\tau)$ in the above integral. According to (2.16), the main contribution to the integral arises from $x \ll x_0$, so that, by (B5),

$$\varphi_M(x) \approx (2\epsilon/\pi)[x\ln x - (1 + \ln x_0)x + O(x^3/x_0^3)]. \quad (2.23)$$

The coefficient of the linear term determines the center frequency of the peak. By comparison with (B9), we see that there is a large frequency shift. However, we can always assume that the center frequency is independently measured and adjusted. We are interested mainly in the shape of the peak, which depends on the curvature of the phase, rather than its slope.

The shape will be distorted mainly by the first term of (2.23). This term has a large curvature and gives rise to a large phase variation in the interval $(0, x_c)$, so that it should produce appreciable distortion.

This is confirmed by a numerical evaluation, as shown in Fig. 3, where $g(\omega)$ (curve in full line) and $g_M(\omega)$ (curve in broken line) are plotted[22] as a function of $(\omega - \omega_0)/\delta$. The results were adjusted so that the maxima of the two peaks would coincide.

Although the height and half-width of the peak in $g_M(\omega)$ do not differ too much from the correct values, there is considerable distortion. It would be a far better approximation in this case to omit the minimal phase altogether, employing only the phase factor $\exp(-i\omega_0 t)$ corresponding to the center frequency.

It should be stressed that the term in the minimal phase that is mainly responsible for the distortion [first term of (2.23)] arises precisely from the range of values of x in the dispersion integral [first integral of (B1)] where the exponential decay law is approximately valid, i.e., where $|\gamma(x)|$ would be measurable

[21] L. A. Khalfin, Dokl. Akad. Nauk SSSR **141**, 599 (1961) [English transl.: Soviet Phys.—Doklady **6**, 1,010 (1962)], asserts that a function essentially identical to (2.16) has no zeros at all in the lower half-plane. His result seems to arise from an incorrect analytic continuation.

[22] Since both $|\gamma(x)|$ and the first term of (2.34) depend only on ϵx in the relevant part of the domain of integration, one obtains a universal curve (independent of ϵ) as a function of this variable.

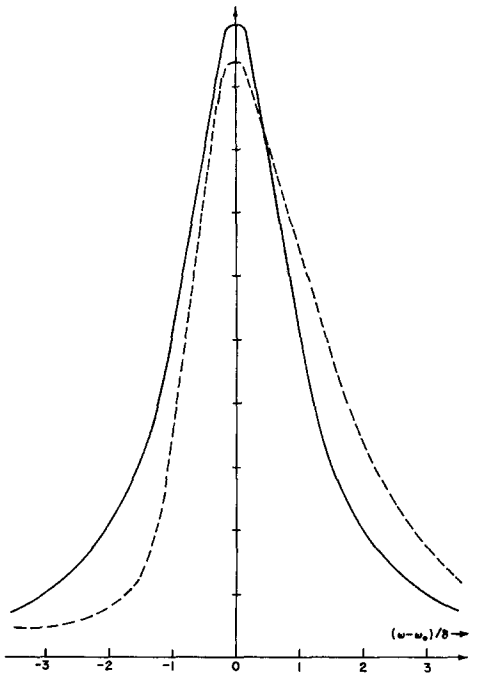


FIG. 3. The functions $g(\omega)$ (——) and $g_M(\omega)$ (– – –) as a function of $(\omega - \omega_0)/\delta$.

in practice. This result therefore depends only on the Lorentzian shape of the spectrum in the neighborhood of the peak, and is insensitive to the low-frequency behavior.

We conclude that the reconstruction procedure by means of the minimal phase would lead to incorrect results in the present example.

Example 2: Gaussian Peak. Let us now take

$$g(\omega) = B \exp\{-[(\omega - \omega_0)/2\delta]^2\}, \qquad (2.24)$$

which might be regarded as a representation of a Doppler-broadened spectral line. Again with $\epsilon = \delta/\omega_0 \ll 1$, the normalization condition (1.3) gives $B \approx (2\pi^{\frac{1}{2}}\delta)^{-1}$.

The corresponding function $\gamma(z)$ is given by (C1) (cf. Appendix C). Taking into account the asymptotic expansion of the error function,[19] we find

$$\gamma(z) \approx \exp(-iz - \epsilon^2 z^2) + \frac{\exp[-(4\epsilon^2)^{-1}]}{2i(\pi^{\frac{1}{2}})(\epsilon z + i/2\epsilon)} \times \left[1 + \frac{1}{2(\epsilon z + i/2\epsilon)^2} + \cdots\right] \quad \text{for} \quad \operatorname{Im} z > -(2\epsilon^2)^{-1}. \qquad (2.25)$$

This approximation is valid, in particular, along the real axis. Thus, $\gamma(x)$ decays according to a Gaussian law which, for very large times, is replaced by an inverse power law, in agreement with (2.8). The transition between the two types of behavior takes place for $x \sim x_0$, where

$$x_0 = (2\epsilon^2)^{-1} \gg x_c = \epsilon^{-1}. \qquad (2.26)$$

According to (C1)–(C4), the zeros of $\gamma(z)$ in the fourth quadrant of the z plane are approximately given by

$$z_n \approx -\frac{i}{2\epsilon^2} + \frac{e^{i\pi/4}}{\epsilon}[(2n - \tfrac{1}{4})\pi]^{\frac{1}{2}} \times \exp\left\{-i\frac{\ln[2\pi(2n - \tfrac{1}{4})^{\frac{1}{2}}]}{2\pi(2n - \tfrac{1}{4})}\right\}. \qquad (2.27)$$

The distribution of zeros is shown in Fig. 4. The distance between two successive zeros decreases as $n^{-\frac{1}{2}}$, in contrast with the uniform spacing found in Fig. 1. The total number of zeros is again given by (2.10).

For $x \ll x_c$, the zeros are far from the real axis and far apart. This suggests that the minimal phase may be a better approximation here than in the previous example. This is confirmed by the expressions for $\varphi_M(x)$ and $\varphi(x)$, given in (C6)–(C9).

If we replace φ by φ_M in (2.22), the reconstructed spectrum will be approximately Gaussian, although the center frequency still suffers a large shift. Thus, assuming the center frequency to be independently adjusted, we get a reasonably good approximation to the spectrum. However, we would again get a better approximation by omitting the minimal phase altogether, since the correction to the linear term in (C6) is entirely different from that in (C8).

Example 3: A Pair of Lorentzian Peaks. The examples hitherto considered correspond to nearly symmetric spectra. Since the phase problem becomes trivial for an exactly symmetric spectrum (cf. Sec. III), it is of interest to consider also larger departures

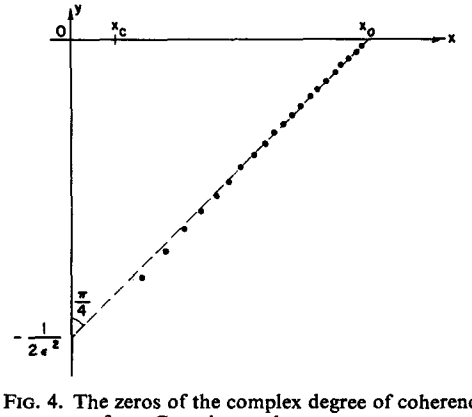


FIG. 4. The zeros of the complex degree of coherence for a Gaussian peak. •—zeros.

from symmetry. For this purpose, let us take

$$g(\omega) = \frac{B\delta}{2\pi}\left[\frac{1+\alpha}{(\omega-\omega_1)^2+\delta^2} + \frac{1-\alpha}{(\omega-\omega_2)^2+\delta^2}\right]. \tag{2.28}$$

This spectrum corresponds to a pair of Lorentzian peaks, centered at ω_1 and ω_2, and taken, for simplicity, to have the same width δ. Their intensity ratio is $(1+\alpha)/(1-\alpha)$. For definiteness, we take $\omega_1 < \omega_2$ and $0 \leq \alpha < 1$. The asymmetry is measured by α. Let us also introduce the dimensionless parameters

$$z = \tfrac{1}{2}(\omega_1+\omega_2)\tau, \quad \epsilon = \frac{2\delta}{\omega_1+\omega_2} \ll 1,$$
$$\beta = \frac{\omega_2-\omega_1}{\omega_2+\omega_1} \ll 1. \tag{2.29}$$

Then, it follows immediately from (2.16) that

$$\gamma(z) \approx B\left\{\tfrac{1}{2}e^{-iz-\epsilon z}[(1+\alpha)e^{i\beta z} + (1-\alpha)e^{-i\beta z}] - \frac{i\epsilon}{\pi z}\right\}$$
$$\text{for} \quad z \in 1, |z| \gg 1. \tag{2.30}$$

The zeros can be obtained by equating modulus and phase of both sides, as in Example 1. We find a family of curves (shown in broken line in Fig. 5) similar to those in (2.20), but the curve corresponding to (2.19) (shown in solid line in Fig. 5) has a different behavior. It has been drawn in Fig. 5 for the case of two well-separated peaks, $\beta \gg \epsilon$, and we have also assumed that $\beta \ln(\epsilon^{-1}) \gg 1$. The curve is tangent to that of Fig. 1 (shown in dash-and-dotted line in Fig. 5) at the points $x_j = j\pi/\beta$ $(j = 0, 1, 2, \cdots)$.

The oscillations of this curve, which are reflected in the positions of the zeros, correspond to beats between the center frequencies of the two peaks. The beat frequency is measured by the parameter β. In the limiting case of two identical, infinitely narrow lines, all the zeros would be located on the real axis, at the points $x_j' = (j+\frac{1}{2})\pi/\beta$ $(j = 0, 1, 2, \cdots)$.

The amplitude of the oscillations is determined by the asymmetry parameter α. It is largest for $\alpha = 0$ (two identical peaks) and it vanishes, as it must, for $\alpha = 1$ (single peak).

The explicit evaluation of the minimal phase is more difficult in the present example. However, it should be clear, by comparing Fig. 5 with Fig. 1, that the Blaschke phase must be even more significant here, since the zeros are located closer to the real axis. The shape and asymmetry of the spectrum are directly reflected in the distribution of zeros, so that most of the information must be contained in the Blaschke phase, and not in the minimal phase.

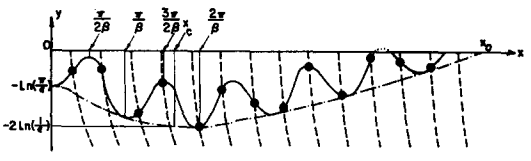


FIG. 5. The zeros of the complex degree of coherence for a pair of Lorentzian peaks. •—zeros.

III. BAND-LIMITED SPECTRA

If we assume that the bandpass filtering effect of the measuring device can be adequately represented by a sharp cutoff, the resulting spectrum is band-limited:

$$g(\omega) = 0 \quad \text{for} \quad |\omega-\omega_0| > a, \tag{3.1}$$

where $2a$ is the bandwidth and ω_0 is the central frequency of the band. This case is also of considerable importance in connection with the phase problem in x-ray scattering,[9] where the band limitation is due to the finite size of the crystal.

According to (1.1) and (3.1),

$$\gamma(\tau) = \int_{\omega_0-a}^{\omega_0+a} g(\omega)\exp(-i\omega\tau)\,d\omega$$
$$= \exp(-i\omega_0\tau)G(\omega), \tag{3.2}$$

where

$$G(\tau) = \int_{-a}^{a} g(\omega_0+u)e^{-iu\tau}\,du. \tag{3.3}$$

We assume that $(-a, a)$ are the effective cutoff points in (3.3), so that the domain of integration cannot be reduced without altering the value of the integral.

It then follows from (3.3) and the Paley–Wiener theorem[23] that $G(\tau)$ is an entire function of exponential order 1 and type a, i.e.,

$$|G(\tau)| \leq A\exp(a|\tau|) \tag{3.4}$$

as $|\tau| \to \infty$ in any direction.

If we make the additional assumption that the spectrum is *symmetric*, i.e., $g(\omega_0-u) = g(\omega_0+u)$, Eq. (3.3) becomes

$$G(\tau) = 2\int_0^a g(\omega_0+u)\cos(u\tau)\,du$$
$$= G(-\tau) = \pm|\gamma(\tau)|. \tag{3.5}$$

Thus, apart from a sign factor, the phase of $\gamma(\tau)$ is given just by the factor $\exp(-i\omega_0\tau)$ in (3.2). The sign may be determined by arguments of physical plausibility, so that knowledge of $|\gamma|$ and ω_0 would enable us to reconstruct the spectrum in this case, by taking the inverse cosine transform of (3.5).

This does not imply, however, that $\gamma(\tau)$ has no

[23] R. P. Boas, *Entire Functions* (Academic Press Inc., New York, 1954), p. 103.

zeros for symmetric spectra; on the contrary, it follows from (3.5) that it has an infinite number of them. In fact, since $G(-\tau) = G(\tau)$, we see that $G(\tau)$ is an entire function of τ^2 of order $\frac{1}{2}$ [according to (3.4)], and an entire function of nonintegral order has an infinite set of zeros.[24] According to Theorem 1 below, this result is valid for any band-limited spectrum, symmetric or not. For a symmetric spectrum, according to (3.5), the distribution of zeros is symmetric not only with respect to the imaginary axis, but also with respect to the real axis: if τ_j is a zero, so are $-\tau_j^*$, $-\tau_j$ and τ_j^*.

The zeros of entire functions of the form (3.2) have been investigated by several authors.[25-27] We quote below some of the most significant results that have been found, adapted to the present case.

The following theorems have been proved by Titchmarsh[26]:

Theorem 1: The function $\gamma(\tau)$ given by (3.2) has an infinite set of zeros.

Let $\tau_1 = t_1 \exp(i\theta_1)$, $\tau_2 = t_2 \exp(i\theta_2)$, $\cdots$ be the zeros, *arranged in order of increasing modulus*: $0 < t_1 \leq t_2 \leq \cdots$ (according to (1.5), this implies that the zeros τ_n and $-\tau_n^*$ will always be paired together).

Theorem 2: The series $\sum_{n=1}^{\infty} t_n^{-1}$ is divergent, but $\sum_{n=1}^{\infty} t_n^{-1} \sin\theta_n$ is absolutely convergent.

This implies that the zeros tend to be located near the real axis for large n. Note that, since τ_n and $-\tau_n^*$ are paired together,

$$\sum_n t_n^{-1} \sin\theta_n = i \sum_n \tau_n^{-1}. \tag{3.6}$$

Theorem 3: If $n(t)$ denotes the number of zeros having $|\tau| \leq t$, we have

$$n(t) \approx 2at/\pi \quad \text{as} \quad t \to \infty. \tag{3.7}$$

Thus, the spacing between two adjacent zeros must approach π/a for large n.

Theorem 4: We have

$$\gamma(\tau) = \exp(-i\omega_0\tau) \prod_{n=1}^{\infty} \left(1 - \frac{\tau}{\tau_n}\right), \tag{3.8}$$

where the product is extended over all the zeros. Furthermore, we have the sum rule

$$\sum_{n=1}^{\infty} \tau_n^{-1} = -i\omega_0 - \gamma'(0) = -i(\omega_0 + \bar{\omega}), \tag{3.9}$$

where

$$\bar{\omega} = \int_{\omega_0 - a}^{\omega_0 + a} \omega g(\omega)\, d\omega \tag{3.10}$$

in the mean frequency. This gives the sum of the series (3.6).

The product expansion (3.8) is simpler than the Hadamard canonical product representation employed in Ref. 9. This expansion brings out the artificial character of the separation (1.7) into minimal phase and Blaschke phase in the present case.

In fact, since $\gamma(\tau)$ is now an entire function, there is no reason to assign a preferential role either to the lower or to the upper half-plane. If we arbitrarily decide to split the phase into a Blaschke phase associated with the zeros in $I_-(\tau)$ and a minimal phase, as in (1.10), it is clear, by comparison with (3.8), that the minimal phase corresponds to just another Blaschke product, involving also the zeros in the upper half-plane $I_+(\tau)$.

Thus, all the zeros, both in I_+ and in I_-, are relevant to the phase problem. It has been shown by Walther[10] (cf. also Ref. 9) that, given $|\gamma(\tau)|$, the positions of all the zeros τ_n are determined up to a reflection on the real axis. All possible solutions differ only by "zero flips" of the type $\tau_n \to \tau_n^*$ (accompanied by $-\tau_n^* \to -\tau_n$ for the other member of the zero pair, in order to preserve the symmetry relation (1.5)). This result is essentially due to the fact that a Blaschke factor $(\tau - \tau_n)/(\tau - \tau_n^*)$ introduces a pole in the opposite half-plane, unless τ_n^* also is a zero; for a representation restricted to a half-plane, this would not matter, but here $\gamma(\tau)$ is an entire function.

So far no use has been made of the nonnegative definiteness condition (1.2). This condition entails new restrictions on the distribution of zeros, as illustrated by the following theorems[25]:

Theorem 5: If $g(\omega) \geq 0$ is a nonincreasing (nondecreasing) function, all the zeros of $\gamma(\tau)$ have positive (negative) imaginary part, except possibly when $g(\omega)$ is piecewise constant.[28]

An example is

$$g(\omega) = A \exp(b\omega) \tag{3.11}$$

in (3.2). The corresponding zeros are

$$\tau_n = (n\pi/a) - ib. \tag{3.12}$$

This can also be taken as an illustration of the "zero

[24] Cf. Ref. 23, p. 24.
[25] G. Pólya, Math. Z. **2**, 352 (1918).
[26] E. C. Titchmarsh, Proc. London Math. Soc. (2) **25**, 283 (1926).
[27] M. L. Cartwright, Quart. J. Math., Oxford Ser. (1) **1**, 38 (1930) and **2**, 113 (1931).

[28] C. L. Mehta and E. Wolf (private communication) have derived the same result independently, by a simple graphical argument based on interpreting the integral (3.2) as the limit of a sum of complex vectors. The exceptional case is that of piecewise constant $g(\omega)$ with a finite number of jump discontinuities at the points $\omega_j = \omega_0 - a + p_j a/q$ where p_j and q are integers. In this case, there are infinitely many zeros on the real axis.

flip" ambiguity: $A\exp(b\omega)$ and $A\exp(-b\omega)$ give rise to complex conjugate zeros, but to the same value of $|\gamma(\tau)|$. More generally, $|\gamma(\tau)|$ does not allow us to distinguish between $g(\omega)$ and the "inverted" spectrum $g(2\omega_0 - \omega)$.

Theorem 6: If $g(\omega) \geq 0$ is continuous and (except perhaps at a finite number of points) differentiable, and if

$$\alpha \leq -g'(\omega)/g(\omega) \leq \beta \quad (\omega_0 - a < \omega < \omega_0 + a), \tag{3.13}$$

then, all the zeros of $\gamma(\tau)$ lie in the open strip

$$\alpha < \operatorname{Im}\tau < \beta, \tag{3.14}$$

the only exception being the example (3.11) (for which the strip reduces to the line $\operatorname{Im}\tau = -b$).

The result contained in Theorem 3 can be made more precise by specifying the behavior of $g(\omega)$ near the endpoints of the interval $(\omega_0 - a, \omega_0 + a)$, which determines the asymptotic behavior of the Fourier integral[29] (3.2), and consequently also the asymptotic distribution of zeros. As an example, we have[27]

Theorem 7: If $g(\omega)$ is continuous within the interval $(\omega_0 - a, \omega_0 + a)$ and $g(\omega_0 \pm a) \neq 0$, and if $g'(\omega)$ is integrable, the zeros of $\gamma(\tau)$ are asymptotically given by

$$\tau_n \approx \frac{n\pi}{a} + \frac{i}{2a}\ln\left[\frac{g(\omega_0 - a)}{g(\omega_0 + a)}\right] + \epsilon_n, \tag{3.15}$$

where $\epsilon_n \to 0$ as $n \to \infty$.

It is clear from Theorems 3 and 7 that the asymptotic distribution of zeros is determined by the properties of the cutoff, and therefore contains no information about the shape of $g(\omega)$ within the interval $(\omega_0 - a, \omega_0 + a)$. This information must be contained in the zeros that are located closer to the origin.[30]

As an illustration of these results, we close this section by reconsidering the example of a Gaussian peak, but now with band limitation:

Example 4: Band-Limited Gaussian Peak. Let

$$g(\omega) = \frac{B}{2(\pi^{\frac{1}{2}})\delta}\exp\left[-\left(\frac{\omega - \omega_0}{2\delta}\right)^2\right] \quad (|\omega - \omega_0| < a),$$
$$= 0 \quad (|\omega - \omega_0| > a). \tag{3.16}$$

Since this is a symmetric spectrum, the corresponding phase problem is trivial, but we are interested only in the distribution of zeros. We assume that the bandwidth is much larger than the width of the peak, so that

$$\chi = \delta/a \ll 1. \tag{3.17}$$

Introducing the dimensionless variable

$$a\tau = \zeta = \xi + i\eta, \tag{3.18}$$

we find, similarly to (2.25),

$$\gamma(\tau) \approx B\exp(-i\omega_0\tau)\left\{\exp(-\chi^2\zeta^2) + \frac{\exp(-\frac{1}{2}\xi_0)}{2i(\pi^{\frac{1}{2}})} \times \left[\frac{\exp(i\zeta)}{\chi\zeta + (i/2\chi)} - \frac{\exp(-i\zeta)}{\chi\zeta - (i/2\chi)}\right]\right\} \quad \text{for } \eta > -\xi_0, \tag{3.19}$$

where

$$\xi_0 = (2\chi^2)^{-1}. \tag{3.20}$$

For $\eta < -\xi_0$, the first term within the curly brackets is to be omitted, so that there are no zeros in this region. All the zeros are contained within the strip $|\eta| < \xi_0$, in agreement with Theorem 6. The distribution of zeros is symmetric with respect to both the real and the imaginary axis, so that it suffices to consider ζ in the fourth quadrant.

Within the strip, for $|\zeta| \ll \xi_0$, we find, just as in (2.27), that the zeros are located very close to

$$\zeta_n = -\frac{i}{2\chi^2} + \frac{e^{i\pi/4}}{\chi}[(2n - \tfrac{1}{4})\pi]^{\frac{1}{2}} \times \exp\left\{-i\frac{\ln[2\pi(2n - \frac{1}{4})^{\frac{1}{2}}]}{2\pi(2n - \frac{1}{4})}\right\} \quad (|\zeta_n| \ll \xi_0). \tag{3.21}$$

Their distribution is very similar to that shown in Fig. 4, with the variables appropriately relabeled.

On the other hand, for $|\zeta| \gg \xi_0$, the zeros are located on the real axis, very close to the points

$$\zeta_n = n\pi, \tag{3.22}$$

in agreement with Theorem 7.

These results also illustrate the remarks about the role of "distant" and "nearby" zeros. It is clear that information about the shape of the spectrum is contained primarily in the zeros (3.21), and not in the distant zeros (3.22).

Finally, expressing the zeros in terms of the variable $z = \omega_0\tau = \omega_0\zeta/a$, we see that (3.21) differs from (2.27) only by the replacement: $(2\epsilon^2)^{-1} \to (2\epsilon\chi)^{-1}$. Thus, if the admitted band extends down to frequencies well below ω_0 ($\chi \sim \epsilon$), the distribution of nearby zeros is almost unaffected by the cutoff, in agreement with the remarks made at the beginning of Sec. II.

IV. CONCLUSION

In all the examples of quasi-monochromatic spectra that have been considered here, it has been found that $\gamma(\tau)$ has a large number of zeros. Their location has

[29] A. I. Erdélyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956), p. 46.

[30] A related conclusion was reached in Ref. 9, by a different argument.

been determined and it has been shown that they play an important role in the phase problem. The spectral reconstruction procedure based on the minimal phase does not lead to satisfactory results in these cases.

What is the physical interpretation of the zeros that have been found? For spectra with a single peak, the behavior of the degree of coherence from $\tau = 0$ up to several coherence times is determined mainly by the behavior of the spectrum in the neighborhood of the peak. This is the region ordinarily accessible to experiment; it corresponds to the domain of validity of the exponential or Gaussian decay law in Examples 1 and 2, respectively. However, this domain is limited by the restriction of the spectrum to positive frequencies. The asymptotic behavior of $\gamma(\tau)$ for $\tau \gg \tau_c$ is determined by the low-frequency behavior of the spectrum.

The zeros that have been found in this case arise from the interference between the initial and the asymptotic decay laws, so that they depend both on the shape of the peak and on the low-frequency behavior. The latter affects most strongly the distribution of distant zeros.

In the case of multiple peaks, as in Example 3, the beats arising from the combination frequencies give rise to zeros close to the real axis; in the limiting case of infinitely narrow lines, these zeros must tend to the real axis.

For band-limited spectra, there is an infinite number of zeros in the τ plane, but only the nearby ones contain information about the shape of the spectrum. The distant zeros arise from the band limitation and depend mainly on the behavior of the spectrum near the end points of the band. In this case, there is no reason to give a preferred role to either the lower or the upper half-plane, so that the splitting into minimal phase and Blaschke phase seems quite artificial.

As was emphasized in the Introduction, the above results need not apply to arbitrary quasi-monochromatic spectra. It is even possible to construct examples of such spectra[31] for which $\gamma(\tau)$ does not have any zeros in $I_-(\tau)$. The assumptions that led to the existence of a large number of zeros in the examples considered here have already been discussed in Sec. II.

In particular, changes in the low-frequency behavior of the spectrum can modify considerably the distribution of zeros. This is related to the well-known fact that small changes in the value of a function on the real axis can produce a large effect on its analytic continuation into the complex plane.

On the other hand, the minimal phase depends only on the behavior of $|\gamma(\tau)|$ along the real axis. Changes in the low-frequency behavior of $g(\omega)$ affect mainly the asymptotic behavior of $\gamma(\tau)$ for very large τ, where it would not be measurable in practice. We have seen in Example 1 that it is precisely the domain where $\gamma(\tau)$ is measurable that gives rise to distortions in the spectral reconstruction by means of the minimal phase; this result is independent of the low-frequency behavior.

One can make the dispersion relation less sensitive to the high-τ behavior by making additional subtractions at $\tau = 0$. However, this would require knowledge of higher-order moments of the spectrum (e.g., the mean frequency, for one subtraction). While this should improve the accuracy of the reconstruction, there is no reason to believe that the subtracted minimal phase would be a useful approximation to the remainder.

As was mentioned in the Introduction, a more ambitious approach would be to attempt a theoretical derivation of the spectrum associated with a given quasi-monochromatic source. However, even for a single emission process, the asymptotic decay law is not uniquely defined; it is known not to be exponential, but its exact form depends on the details of the measurement process.[18–20] The effect of taking an ensemble average to obtain the degree of coherence is unknown. One would also have to take into account the effect of the measuring apparatus, e.g., its frequency response and the limitations due to noise.

In practice, if the spectrum is known to be nearly symmetric, one may obtain a good first approximation by assuming complete symmetry (in which case the phase problem can be trivially solved), and one may then investigate the effects of a small departure from symmetry. If the spectrum is strongly asymmetric, it may still be possible to represent it by a superposition of symmetric spectra, e.g., when there are several peaks. The number and relative position of the peaks may be inferred by analyzing the resulting beat pattern, which also appears in $|\gamma|$. This seems to have been the procedure originally applied by Michelson.[32]

In conclusion, we see that the determination of $|\gamma(\tau)|$ alone does not seem to suffice, in general, for the unambiguous reconstruction of quasi-monochromatic spectra, even when the requirements of analyticity and nonnegative definiteness are taken into account. This is perhaps not very surprising, since the

[31] D. Dialetis, J. Math. Phys. (to be published).

[32] A. A. Michelson, Phil. Mag. **31**, 338 (1891); **34**, 280 (1892).

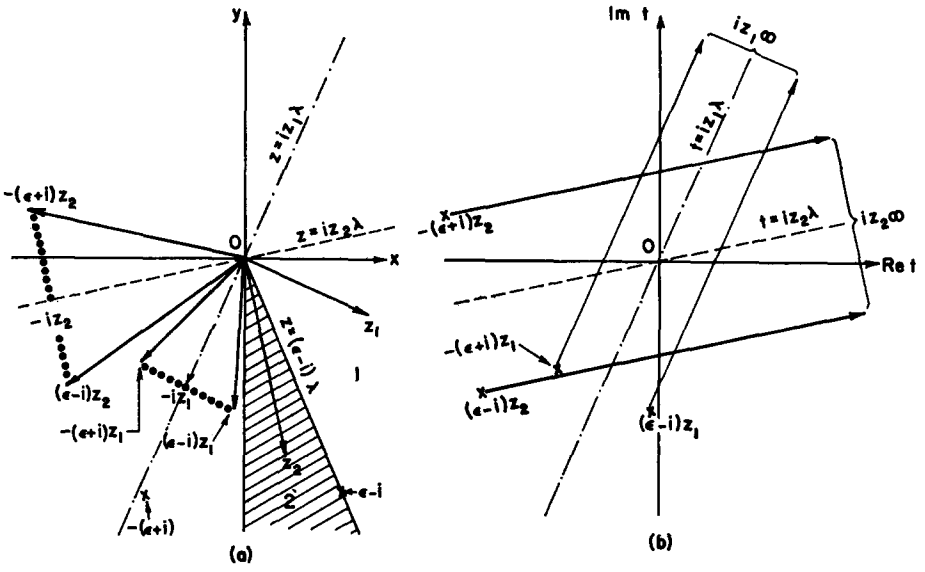


FIG. 6. (a) Division into regions of the fourth quadrant of the z plane. (b) Corresponding paths of integration in the t plane.

analytic properties of $\gamma(\tau)$ do not originate from any far-reaching physical principle, such as causality (cf. Sec. I).

Additional information, either in the form of theoretical restrictions derived from a physical model of the source or by further measurements,[33,34] seems necessary to solve the phase problem. It is not clear at the present time which procedure would be most suitable for this purpose.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor E. Wolf for suggesting this problem. The author is also indebted to him and to Professor L. Mandel for interesting discussions.

This work was supported by the Air Force Cambridge Research Laboratories.

APPENDIX A. EVALUATION OF $\gamma(z)$ FOR A LORENTZIAN PEAK

It follows from (2.14) that

$$\gamma(z)=\frac{e^{-iz}}{2i\pi}\left[e^{\epsilon z}\int_{(\epsilon-i)z}^{iz\infty}e^{-t}\frac{dt}{t}-e^{-\epsilon z}\int_{-(\epsilon+i)z}^{iz\infty}e^{-t}\frac{dt}{t}\right], \tag{A1}$$

where the paths of integration are straight lines parallel to $t=iz\lambda$. The integrals can be reduced to the exponential integral, defined by[35]

$$E_1(u)=\int_u^\infty e^{-t}\frac{dt}{t}\quad(|\arg u|<\pi), \tag{A2}$$

where it is assumed that the path of integration avoids the origin and does not cross the negative real axis.

To perform the reduction, we must subdivide the fourth quadrant of the z plane into the regions 1 and 2 shown in Fig. 6(a). In region 1,

$$-\pi<\arg(-\epsilon z-iz)<-\tfrac{1}{2}\pi;$$

in region 2, $\frac{1}{2}\pi<\arg(-\epsilon z-iz)<\pi$. For a point $z_1\in 1$, the path of integration in the second integral of (A1) crosses the negative real axis [cf. Fig. 6(b)], so that it has to be taken across the pole at the origin to reduce it to (A2). This does not happen for a point $z_2\in 2$, nor does it happen in the first integral, so that we finally get

$$\gamma(z)=\theta(1,2)e^{-iz-\epsilon z}+(e^{-iz}/2\pi i)\times[e^{\epsilon z}E_1(\epsilon z-iz)-e^{-\epsilon z}E_1(-\epsilon z-iz)], \tag{A3}$$

where

$$\theta(1,2)=1\quad\text{if}\quad z\in 1,\quad\theta(1,2)=0\quad\text{if}\quad z\in 2. \tag{A4}$$

We have[35]

$$E_1(u)=Ein(u)-\ln u-C, \tag{A5}$$

where C is Euler's constant and

$$Ein(u)=\int_0^u(1-e^{-t})\frac{dt}{t}=-\sum_{n=1}^{\infty}\frac{(-1)^n u^n}{nn!} \tag{A6}$$

is an entire function. The discontinuity across the cut of the logarithmic term in $E_1(-\epsilon z-iz)$ compensates the discontinuity (A4), so that (A3) represents a regular, single-valued function in the fourth quadrant of the z plane. Note, however, that it has a logarithmic branch point at the origin.

To find the behavior of $\gamma(z)$ for $|z|\gg 1$, one can

[33] H. Gamo, J. Appl. Phys. 34, 875 (1963).
[34] C. L. Mehta, Nuovo Cimento 36, 202 (1965).
[35] *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964), p. 228.

employ the asymptotic expansion of $E_1(u)$:

$$E_1(u) = \frac{e^{-u}}{u}\left[1 - \frac{1}{u} + \cdots + \frac{(-1)^n n!}{u^n} + O(u^{-n-1})\right], \tag{A7}$$

where the remainder is smaller than the first neglected term.

APPENDIX B. EVALUATION OF THE PHASE FOR A LORENTZIAN PEAK

It follows from (1.8), (2.15), and (2.16) that the minimal phase, for $x \gg 1$, is approximately given by

$$\begin{aligned}\varphi_M(x) &\approx -\frac{2x}{\pi}\left[\epsilon P\int_0^{x_0}\frac{x'\,dx'}{x'^2 - x^2}\right. \\ &\quad \left. + \ln\left(\frac{\pi}{\epsilon}\right)P\int_{x_0}^{\infty}\frac{dx'}{x'^2 - x^2} + P\int_{x_0}^{\infty}\frac{\ln x'\,dx'}{x'^2 - x^2}\right] \\ &= -\frac{\epsilon x}{\pi}\ln\left|1 - \frac{x_0^2}{x^2}\right| - \frac{1}{\pi}\ln\left(\frac{\pi x_0}{\epsilon}\right)\ln\left|\frac{x_0 + x}{x_0 - x}\right| \\ &\quad - \frac{1}{\pi}\left[f\left(\frac{x}{x_0}\right) - f\left(-\frac{x}{x_0}\right)\right],\end{aligned} \tag{B1}$$

where

$$f(x) = -\int_0^x y^{-1}\ln|1 - y|\,dy \tag{B2}$$

is the dilogarithm function.[36] We have

$$f(x) = \sum_{n=1}^{\infty} n^{-2}x^n \quad (|x| \leq 1) \tag{B3}$$

and[37]

$$f(x) - f(-x) + f(x^{-1}) - f(-x^{-1}) = \tfrac{1}{2}\pi^2 \quad (x > 0). \tag{B4}$$

These relations allow us to obtain rapidly convergent expansions of (B1) both for $x \ll x_0$ and for $x \gg x_0$:

$$\varphi_M(x) = -\frac{\epsilon x}{\pi}\ln\left|\frac{x_0^2}{x^2} - 1\right| - \frac{2}{\pi}(1 + \epsilon x_0)\frac{x}{x_0} + O\left(\frac{x^3}{x_0^3}\right) \quad \text{for } 1 \ll x \ll x_0, \tag{B5}$$

$$\varphi_M(x) = -\frac{\pi}{2} - \frac{1}{\pi}(\epsilon x_0 - 2)\frac{x_0}{x} + O\left(\frac{x_0^3}{x^3}\right) \quad \text{for } x \gg x_0. \tag{B6}$$

Finally, for $x \to 0$, we find

$$\varphi_M(x) = (\epsilon/\pi)x\ln x + O(x) \to 0 \quad \text{as} \quad x \to 0. \tag{B7}$$

These results are to be compared with the actual phase $\varphi(x)$. It follows from (1.5) and (2.9) (with $\alpha = 0$) that we must take $\varphi(\infty) = -\varphi(-\infty) = -\frac{1}{2}\pi$, and (2.3) then implies that $\varphi(0+) = -\varphi(0-) = N\pi$, where N is the total number of zeros. Taking into account (2.15) and (2.16), we get

$$\varphi(x) \approx N\pi + (\epsilon/\pi)x\ln x + O(x) \quad (0 < x \leq 1), \tag{B8}$$

$$\varphi(x) \approx N\pi - x - (x_0/x)\cos x\exp\left[-\epsilon(x_0 - x)\right] \quad (1 \ll x \ll x_0), \tag{B9}$$

$$\varphi(x) \approx -\tfrac{1}{2}\pi + (x/x_0)\cos x\exp\left[-\epsilon(x - x_0)\right] \quad (x \gg x_0). \tag{B10}$$

APPENDIX C. EVALUATION OF $\gamma(z)$ FOR A GAUSSIAN PEAK

Substituting (2.24) in (1.1), we find, in terms of the dimensionless variables (2.13),

$$\gamma(z) = \tfrac{1}{2}\exp(-iz - \epsilon^2 z^2)\operatorname{erfc}(w), \tag{C1}$$

where

$$w = i\epsilon z - 1/2\epsilon \tag{C2}$$

and

$$\operatorname{erfc}(w) = \frac{2}{\pi^{\frac{1}{2}}}\int_w^{\infty}\exp(-\zeta^2)\,d\zeta \tag{C3}$$

is the error function, which is an entire function of w. Thus, $\gamma(z)$ in this case is an entire function of z.

The zeros of erfc (w) are symmetrically distributed with respect to the real axis and they are contained in the half-plane $\operatorname{Re} w < 0$. Those located in the second quadrant are approximately given by[38]

$$w_n \approx e^{3i\pi/4}[(2n - \tfrac{1}{4})\pi]^{\frac{1}{2}}\exp\left\{-i\frac{\ln[2\pi(2n - \frac{1}{4})^{\frac{1}{2}}]}{2\pi(2n - \frac{1}{4})}\right\}. \tag{C4}$$

According to (2.25), the minimal phase is given by

$$\begin{aligned}\varphi_M(x) &\approx -\frac{2x}{\pi}\left\{\epsilon^2 P\int_0^{x_0}\frac{x'^2\,dx'}{x'^2 - x^2} + \left[\frac{1}{4\epsilon^2} + \ln(2(\pi^{\frac{1}{2}})\epsilon)\right]\right. \\ &\quad \left. \times P\int_{x_0}^{\infty}\frac{dx'}{x'^2 - x^2} + P\int_{x_0}^{\infty}\frac{\ln x'\,dx'}{x'^2 - x^2}\right\} \\ &= -\frac{2x}{\pi}\left\{\frac{1}{2} - \frac{\epsilon^2}{2x}(x^2 - x_0^2)\ln\left|\frac{x + x_0}{x - x_0}\right|\right. \\ &\quad \left. + \frac{1}{2x}\left[f\left(\frac{x}{x_0}\right) - f\left(-\frac{x}{x_0}\right)\right]\right\},\end{aligned} \tag{C5}$$

where $f(x)$ has been defined in (B2). It follows that

$$\varphi_M(x) \approx -\frac{2x}{\pi} + \frac{2x^3}{3\pi x_0^2} + \cdots \quad (x \ll x_0), \tag{C6}$$

$$\varphi_M(x) \approx -\frac{\pi}{2} - \frac{2x_0^2}{3\pi x} + \cdots \quad (x \gg x_0), \tag{C7}$$

whereas, according to (2.25), the actual phase is given by

$$\varphi(x) \approx N\pi - x - (\epsilon/\pi^{\frac{1}{2}})\sin x\exp\left[-(x_0^2 - x^2)/2x_0\right] \quad (0 < x \ll x_0), \tag{C8}$$

$$\varphi(x) \approx -\tfrac{1}{2}\pi - \tan^{-1}(x_0/x) + \cdots \quad (x \gg x_0). \tag{C9}$$

[36] K. Mitchell, Phil. Mag. 40, 351 (1949).

[37] This result follows from term-by-term subtraction of equations (4.2) (with x replaced by $-x$) and (4.3) in Ref. 36.

[38] This follows from the asymptotic expansion of erfc (w) given in Ref. 19, p. 33. Similar results for erf $(w) = 1 - \operatorname{erfc}(w)$ are given in Ref. 35, p. 329.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Equations of Motion in Classical Nonlinear Field Theories

GERALD ROSEN
Drexel Institute of Technology, Philadelphia, Pennsylvania

(Received 16 July 1966)

A Rayleigh–Ritz procedure is outlined for deriving approximate equations of "particle" motion (the over-all motion of a singularity-free particlelike solution) in a superimposed small-amplitude external field. The method is illustrated here for a classical nonlinear model scalar field theory. A relativistic generalization of the approximation procedure is described.

I. INTRODUCTION

RIGOROUS particlelike solutions to Lorentz-covariant nonlinear model field theories, solutions which are spatially localized, time-independent, singularity-free, and of finite energy, have been the subject of recent papers.[1,2] Our purpose in the present paper is to outline a systematic Rayleigh–Ritz approximation procedure for deriving the motion of such a "particle" in a superimposed external field, more precisely, the over-all motion of a particlelike solution induced by the nonlinearity of the field equations with the addition of a small-amplitude wave solution. This Rayleigh–Ritz approximation method appears to be ideal for obtaining unambiguous equations of "particle" motion from any nonlinear field theory associated with an action principle that admits a singularity-free particlelike solution of finite energy, the method applying irrespective of whether the field theory features Einsteinian general covariance and concomitant Bianchi identities.[3] We also sketch an analogous approximation theory for the quantum motion of the particlelike solution, application of a standard canonical quantization to the dynamical theory based on the "reduced Lagrangian" of the classical Rayleigh–Ritz procedure. The approximation theory is illustrated here for the rigorous particlelike solutions to a nonlinear model scalar field theory, and a relativistic generalization is described.

II. CLASSICAL AND QUANTUM NONRELATIVISTIC EQUATIONS OF MOTION

Let us consider a Lorentz-covariant nonlinear field theory with the invariant Lagrangian density,

$$\mathfrak{L} = \mathfrak{L}(\phi, \dot{\phi}, \nabla\phi), \tag{1}$$

and the action principle,

$$\delta \int L\,dt = 0, \quad L \equiv \int \mathfrak{L}\, d^3\mathbf{x}, \tag{2}$$

where $\phi = \phi(\mathbf{x}, t)$ is a generic (multicomponent) real field. Suppose that the field equations derived from (2) admit a spatially localized time-independent singularity-free particlelike solution,

$$\phi = \phi_0(\mathbf{x}), \tag{3}$$

and also a (nonlocalized) small-amplitude singularity-free wave solution,

$$\phi = \phi_{ex}(\mathbf{x}, t). \tag{4}$$

With additive superposition of the right members of (3) and (4), we seek an approximate solution of the form

$$\phi = \phi_0(\mathbf{x} - \boldsymbol{\xi}) + \phi_{ex}(\mathbf{x}, t), \tag{5}$$

where the coordinates $\boldsymbol{\xi} = \boldsymbol{\xi}(t)$ locate the center of the "particle." Keeping terms at most quadratic in ϕ_{ex}, the Lagrangian density (1) is evaluated with (5) and the resulting function of $\boldsymbol{\xi}$, $\dot{\boldsymbol{\xi}}$, $\mathbf{x}$, and t is integrated over $\mathbf{x}$. Thus, the Lagrangian in the definition part of (2) is expressed explicitly as

$$L = L(\boldsymbol{\xi}, \dot{\boldsymbol{\xi}}, t) \tag{6}$$

by dropping an additive function of t alone. Note that the Lagrangian (6) is suitable for describing the

[1] G. Rosen, J. Math. Phys. **6**, 1269 (1965).
[2] G. Rosen, J. Math. Phys. **7**, 2066 (1966).
[3] For obtaining the equations of motion in a general relativistic field theory which does not admit a *singularity-free* particlelike solution, one must use either the Einstein–Infeld–Hoffman method [e.g., L. Infeld, Rev. Mod. Phys. **29**, 398 (1957)] or the Fock method [e.g., V. A. Fock, *Theory of Space, Time, and Gravitation* (Pergamon Press, Inc., New York, 1962)]. Each of the latter methods, depending in an essential way on general covariance and the concomitant Bianchi identities, involves a series expansion and iterative solutional procedure which is certainly not justifiable in a rigorous sense, for close to the mass point (or field singularity) the field equations are essentially nonlinear. Furthermore, unambiguous equations of motion are obtainable with an Einstein–Infeld–Hoffman or Fock method only if supplemented with certain suitable coordinate conditions {e.g., H. Wojewoda, Zh. Eksperim. i Teor. Fiz. **45**, 2051 (1963) [English transl.: Soviet Phys.—JETP **18**, 1408 (1964)]} and certain additional "nonradiative" conditions on the mass point itself [e.g., A. Peres, Phys. Rev. **137**, B1126 (1965)]. It is for these reasons that the Rayleigh–Ritz approximation procedure discussed here would be preferable mathematically for a *singularity-free* particlelike solution in the context of an Einsteinian general relativistic theory, notwithstanding the venerability of the Einstein–Infeld–Hoffman and Fock methods.

dynamics of a particle with coordinates ξ. Indeed, the action principle (2) supplemented with a Rayleigh–Ritz argument guarantees the approximate validity of the Euler–Lagrange equations derived from (6). Hence, the Euler–Lagrange equations derived from (6) are approximate classical equations of motion for a particle with the coordinates ξ in a superimposed small-amplitude external field. It also follows that the "reduced Lagrangian" (6), representing the particle system accurately to within the approximation of three principal degrees of freedom, can be used to formulate a corresponding Schrödinger equation for the approximate quantum motion of the particle.

In order to illustrate this Rayleigh–Ritz approximation method for the motion of a particle in an external field, consider the solvable nonlinear model field theory[1] based on the Lagrangian density,

$$\mathcal{L} = (\dot{\theta})^2 - (\nabla\theta)^2 + g\theta^6, \tag{7}$$

with $\theta = \theta(\mathbf{x}, t)$ a real scalar field and g a positive physical constant. The field equation associated with (7) admits rigorous singularity-free spherically symmetric static solutions of the form

$$\theta = \theta_0(\mathbf{x}) \equiv Z(Z^4 g + |\mathbf{x}|^2)^{-\frac{1}{2}} \tag{8}$$

in which the "size parameter" Z is a free nonzero constant of integration. For a solution of the form (8) the total field energy or "particle rest mass" is

$$\int [(\nabla\theta_0)^2 - g\theta_0^6]\, d^3\mathbf{x} = \frac{\pi^2}{2g^{\frac{1}{2}}} \equiv m_0, \tag{9}$$

a quantity independent of Z. The particlelike solutions (8) are dynamically unstable, but the characteristic time for dissolution of such a solution is of the order $Z^2 g^{\frac{1}{2}}$ and is therefore arbitrarily large for $|Z|$ sufficiently large.

Now let $\theta_{ex} = \theta_{ex}(\mathbf{x}, t)$ denote a small-amplitude singularity-free wave field which satisfies the linearized field equation, $\ddot{\theta}_{ex} - \nabla^2\theta_{ex} = 0$. With θ_{ex} superimposed additively on the rigorous particlelike solution (8), we seek an approximate solution in which (8) is generalized dynamically,

$$\theta = \theta_0(\mathbf{x} - \xi) + \theta_{ex}(\mathbf{x}, t), \quad \xi = \xi(t). \tag{10}$$

Evaluated with (10) the Lagrangian density (7) works out to give

$$\begin{aligned}\mathcal{L} \cong\ & (\dot{\xi}\cdot\nabla\theta_0)^2 - 2(\dot{\xi}\cdot\nabla\theta_0)\dot{\theta}_{ex} + (\dot{\theta}_{ex})^2 - (\nabla\theta_0)^2 \\ & - 2\nabla\theta_0\cdot\nabla\theta_{ex} - (\nabla\theta_{ex})^2 \\ & + g(\theta_0^6 + 6\theta_0^5\theta_{ex} + 15\theta_0^4\theta_{ex}^2)\end{aligned} \tag{11}$$

up to terms of quadratic in θ_{ex} and where the argument of θ_0 is understood to be $(\mathbf{x} - \xi)$. We simplify the integration of (11) over all $\mathbf{x}$ by making use of the formulas

$$\begin{aligned}&\nabla^2\theta_0 + 3g\theta_0^5 = 0, \\ &\int (\nabla\theta_0)^2\, d^3\mathbf{x} = \int 3g\theta_0^6\, d^3\mathbf{x} = \tfrac{3}{2}m_0, \\ &\int \nabla_i\theta_0\nabla_j\theta_0\, d^3\mathbf{x} = \tfrac{1}{2}m_0\delta_{ij}.\end{aligned} \tag{12}$$

Discarding the term $\int[(\dot{\theta}_{ex})^2 - (\nabla\theta_{ex})^2]\, d^3\mathbf{x}$, an additive function of t alone, we thus obtain

$$\begin{aligned}L \cong\ & \frac{1}{2}m_0(\dot{\xi})^2 - 2\dot{\xi}\cdot\int(\nabla\theta_0)\dot{\theta}_{ex}\, d^3\mathbf{x} - m_0 \\ & - 2\int\nabla\cdot(\theta_{ex}\nabla\theta_0)\, d^3\mathbf{x} + 15g\int\theta_0^4\theta_{ex}^2\, d^3\mathbf{x}.\end{aligned} \tag{13}$$

The three integral terms in (13) can be evaluated easily with explicit integration provided that $|\nabla\dot{\theta}_{ex}/\dot{\theta}_{ex}|$ and $|\nabla\theta_{ex}/\theta_{ex}|$ are small compared to $(Z^2 g^{\frac{1}{2}})^{-1}$ in the neighborhood of the particle. With the latter conditions satisfied by θ_{ex}, we have

$$\begin{aligned}&\int(\nabla\theta_0)\dot{\theta}_{ex}\, d^3\mathbf{x} \cong 0, \\ &\int\nabla\cdot(\theta_{ex}\nabla\theta_0)\, d^3\mathbf{x} \cong -4\pi Z\theta_{ex}(\xi, t), \\ &\int\theta_0^4\theta_{ex}^2\, d^3\mathbf{x} = \pi^2 Z^2 g^{-\frac{1}{2}}\theta_{ex}(\xi, t)^2,\end{aligned} \tag{14}$$

and hence (13) becomes

$$L \cong \tfrac{1}{2}m_0\dot{\xi}^2 - m_0 - V(\xi, t), \tag{15}$$

in which

$$V(\xi, t) \equiv -8\pi Z\theta_{ex}(\xi, t) - 15\pi^2 Z^2 g^{\frac{1}{2}}\theta_{ex}(\xi, t)^2. \tag{16}$$

Thus the external field acts on the particle through the "effective potential" (16). In view of the particlelike solution (8), Z is analogous to an electric charge with θ analogous to an electrostatic potential, and so the linear (dominant) term in (16) is of an "anti-Coulombic" character, similar to an electrostatic potential energy except for the minus sign. Finally we note that the Euler–Lagrange equations derived from (15)

$$m_0\ddot{\xi} = -\nabla V(\xi, t) \tag{17}$$

give the approximate classical motion of the particle. The reduced Lagrangian (15) also leads to the Schrödinger equation,

$$\frac{i\hbar\partial\psi}{\partial t} = \left[-\frac{\hbar^2}{2m_0}\left(\frac{\partial}{\partial\xi}\right)^2 + m_0 + V(\xi, t)\right]\psi \tag{18}$$

for the corresponding quantum motion.

In the case of two particles separated by a distance

large compared to their radii, the field equation admits an approximate solution of the form

$$\theta = \theta_0(1) + \theta_0(2), \quad \theta_0(i) \equiv Z_{(i)}(Z_{(i)}^4 g + |\mathbf{x} - \boldsymbol{\xi}_{(i)}|^2)^{-\frac{1}{2}},$$

in place of (10). Straightforward integration over all $\mathbf{x}$ of cross terms in the Lagrangian density (7) produces the potential energy of interaction

$$\begin{aligned} V = &-8\pi Z_{(1)}Z_{(2)}\,|\boldsymbol{\xi}_{(1)} - \boldsymbol{\xi}_{(2)}|^{-1} \\ &- 30\pi^2 Z_{(1)}^2 Z_{(2)}^2 g^{\frac{1}{2}}\,|\boldsymbol{\xi}_{(1)} - \boldsymbol{\xi}_{(2)}|^{-2} \\ &+ (\text{higher-order terms in the reciprocal} \\ &\quad \text{of the separation distance } |\boldsymbol{\xi}_{(1)} - \boldsymbol{\xi}_{(2)}|). \end{aligned}$$

The leading potential energy term is of an anti-Coulombic character, two particles with $Z_{(1)}Z_{(2)} > 0$ attracting each other with a force proportional to the product of their size parameters divided by their separation distance squared, the particles repelling each other if their size parameters have opposite signs; on the other hand, the potential energy term proportional to the reciprocal of the separation distance squared is of a manifestly attractive short-range character, two particles attracting each other with a force proportional to the product of their radii ("particle radius" $\cong Z^2 g^{\frac{1}{2}}$) divided by their separation distance cubed. Owing to the large numerical prefactor $30\pi^2$, the latter short-range attractive force is one or two orders of magnitude greater than a value suggested by naïve dimensional considerations. Note that the leading anti-Coulombic term in the potential energy for two interacting particles follows from the linear term in expression (16) by evoking the elementary algorithm of mechanics: "To obtain the potential energy of interaction for two particles, evaluate the potential field due to the first particle at the location of the second particle and multiply the result by an appropriate physical constant (a 'charge' or mass) associated with the second particle." However, the elementary algorithm of mechanics applied to (16) with $Z\theta_{\text{ex}}(\boldsymbol{\xi}, t) \to Z_{(1)}Z_{(2)}\,|\boldsymbol{\xi}_{(1)} - \boldsymbol{\xi}_{(2)}|^{-1}$ does not produce a numerically correct expression for the potential energy beyond the leading anti-Coulombic term. For example, the term proportional to the reciprocal of the separation distance squared contains an extra factor of 2, reflecting the symmetrical contributions to the integral over all $\mathbf{x}$ from the neighborhoods of both particles. We suspect that the elementary algorithm of mechanics is in general only valid for long-range $|\boldsymbol{\xi}_{(1)} - \boldsymbol{\xi}_{(2)}|^{-1}$ type potential fields.

III. RELATIVISTIC GENERALIZATION

The preceding Rayleigh–Ritz approximation method yields nonrelativistic equations of particle motion, valid for $\dot{\boldsymbol{\xi}}^2$ small compared to unity. It is easy to generalize the method to obtain relativistic equations of particle motion, valid for larger values of $\dot{\boldsymbol{\xi}}^2 < 1$. In place of (5) we seek a more general approximate solution of the form

$$\phi = S\phi_0(\hat{\mathbf{x}}) + \phi_{\text{ex}}(\mathbf{x}, t), \tag{19}$$

where $\hat{\mathbf{x}}$ is related to $\mathbf{x}$ at any instant of time by an inhomogeneous Lorentz transformation,

$$\begin{aligned} &\hat{\mathbf{x}} \equiv \boldsymbol{\Lambda}(\mathbf{x} - \boldsymbol{\xi}), \\ &(\boldsymbol{\Lambda})_{ij} = \delta_{ij} + [1 - \dot{\boldsymbol{\xi}}^2 + (1 - \dot{\boldsymbol{\xi}}^2)^{\frac{1}{2}}]^{-1}\dot{\boldsymbol{\xi}}_i\dot{\boldsymbol{\xi}}_j, \end{aligned} \tag{20}$$

and S in (19) is a generic matrix function of $\boldsymbol{\Lambda}$ that mixes the components of ϕ according to their Lorentz transformation character. The eigenvalues of the 3×3 time-dependent matrix $\boldsymbol{\Lambda}$ are 1, 1, and $(1 - \dot{\boldsymbol{\xi}}^2)^{-\frac{1}{2}}$, $\boldsymbol{\Lambda}$ being the spatial part of a 4×4 homogeneous Lorentz transformation; thus we have $d^3\mathbf{x} = (\det \boldsymbol{\Lambda})^{-1}\, d^3\hat{\mathbf{x}} = (1 - \dot{\boldsymbol{\xi}}^2)^{\frac{1}{2}}\, d^3\hat{\mathbf{x}}$. Since $\boldsymbol{\Lambda}$ changes slowly with time if ϕ_{ex} is relatively small (for then $\ddot{\boldsymbol{\xi}}$ is relatively small), it follows that

$$\int \mathfrak{L}(\hat{\phi}_0, \dot{\hat{\phi}}_0, \nabla\hat{\phi}_0)\, d^3\mathbf{x} \cong -m_0(1 - \dot{\boldsymbol{\xi}}^2)^{\frac{1}{2}}, \tag{21}$$

where $\hat{\phi}_0 \equiv S\phi_0(\hat{\mathbf{x}})$ and use is made of the Lorentz invariant character of the Lagrangian density and the expression for the "particle rest mass,"

$$m_0 = -\int \mathfrak{L}(\phi_0, 0, \nabla\phi_0) d^3\mathbf{x}, \tag{22}$$

with $\phi_0 = \phi_0(\mathbf{x})$. Hence, by substituting (19) into the Lagrangian density, expanding the result up to terms quadratic in $\phi_{\text{ex}}(\mathbf{x}, t)$, and discarding an additive function of t alone, we obtain the reduced Lagrangian in relativistic form

$$L = -m_0(1 - \dot{\boldsymbol{\xi}}^2)^{\frac{1}{2}} - V(\boldsymbol{\xi}, t)(1 - \dot{\boldsymbol{\xi}}^2)^{\frac{1}{2}}, \tag{23}$$

where the generic "effective potential," $V(\boldsymbol{\xi}, t)$, is in general composed of terms linear and quadratic in $\phi_{\text{ex}}(\boldsymbol{\xi}, t)$ and Lorentz transforms as an invariant. For the nonlinear model scalar field theory, the reduced Lagrangian (23) is the relativistic generalization of (15) with $V(\boldsymbol{\xi}, t)$ in (23) working out to give (16) unaltered.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Random Walk with an Excluded Origin

ROBERT J. RUBIN
National Bureau of Standards, Washington, D.C.

(Received 29 July 1966)

The mean square end-to-end distance R_N^2 is calculated for the subset of all random walk configurations on a D-dimensional simple cubic lattice which do not return to the starting point. Explicit results are obtained in the limit $N \gg 1$ for the one-, two-, and three-dimensional lattices. The values of the first two terms in the asymptotic series for R_N^2 are, respectively, $N + N$, $N + N/\log N$, and $N + 0.435/N^{\frac{1}{2}}$. An unexpected relation is obtained between R_N^2 and S_N, the average number of different lattice sites visited in an N-step random walk on a perfect lattice. It is $R_N^2 = S_N(S_{N+1} - S_N)^{-1}$.

1. INTRODUCTION

IN this paper we calculate R_N^2, the mean square end-to-end distance, for a particular subset of all N-step random walk configurations on the one-, two-, and three-dimensional simple cubic lattices. The subset consists of those random walk configurations which satisfy the $N - 1$ restrictions that the 2nd, 3rd, $\cdots$, and Nth steps cannot overlap the first step. This restricted random walk, or excluded origin problem is a special case of the problem in which no step can overlap any other step. This latter problem arises in lattice models of long polymer chain molecules which account for the self-excluded volume of the polymer chain.[1] Some time ago,[2] we determined the increase in the mean square end-to-end distance for a D-dimensional random walk which results from the introduction of the single restriction that the jth and kth steps cannot overlap. The method employed there is not useful in the present problem where the number of overlap restrictions is large and the restrictions are "nested" in the sense that, for a given pair of interacting steps, intermediate steps are involved in other overlap restrictions. The method of solution which we use is based on a one-to-one correspondence between the excluded origin random walk configurations and a set of random walk configurations on a lattice containing an absorbing point at the origin.

The problem of calculating the probability distribution of the end-to-end distance is formulated in Sec. 2 and solved formally in Sec. 3. The absorbing point is treated as a defect in an otherwise perfect lattice. The method used is similar to that used by Rubin[3-5] in treating random walk models of polymer chain adsorption on plane solution surfaces and thin rods, and is equivalent to the method presented by Montroll[6,7] and Montroll and Weiss[8] for discussing random walk problems on slightly defective lattices. Lifshitz[9] appears to have made the earliest use of these methods in lattice vibration problems.

The expression for R_N^2 is evaluated in Sec. 3 in the limit $N \gg 1$. The formal expression which is obtained for R_N^2 is unexpectedly given in terms of S_N, the average number of different lattice points visited in a random walk of N steps on a perfect lattice. The relation is

$$R_N^2 = S_N/(S_{N+1} - S_N).$$

In obtaining the asymptotic value of R_N^2 in the one- and three-dimensional lattices, we have used the asymptotic series for S_N given by Montroll and Weiss.[8] The necessary analysis is also carried out for the two-dimensional lattice.

Some aspects of the results are discussed in Sec. 4.

2. RECURRENCE EQUATIONS

We consider restricted random walks on a D-dimensional simple cubic lattice such that the walker steps between nearest-neighbor lattice points only but cannot return to the starting point. Let $C(D; N)$ denote the set of all N-step random walk paths on a D-dimensional lattice which originate at the origin; and let $C_0(D; N)$ denote the subset of $C_0(D; N)$ which do not return to the origin. We wish to calculate the mean square displacement at the Nth step for the subset $C_0(D; N)$ in the limit $N \gg 1$ for the one-, two-, and three-dimensional lattices. A simple one-to-one correspondence can be established between the random walk paths in $C_0(D; N)$ and random walk paths

[1] There is an extensive and growing literature on this problem. See, e.g., M. E. Fisher and M. F. Sykes, Phys. Rev. **114**, 45 (1959); C. Domb, J. Chem. Phys. **38**, 2957 (1963); S. F. Edwards, Proc. Phys. Soc. (London) **85**, 613 (1965); J. Mazur, J. Chem. Phys. **43**, 4354 (1965).

[2] R. J. Rubin, J. Chem. Phys. **20**, 1940 (1952).

[3] R. J. Rubin, J. Chem. Phys. **43**, 2392 (1965).

[4] R. J. Rubin, J. Res. Natl. Bur. Std. (U.S.) **69B**, 301 (1965).

[5] R. J. Rubin, J. Chem. Phys. **44**, 2130 (1966).

[6] E. W. Montroll, *Applied Combinatorial Mathematics* (John Wiley & Sons, Inc., New York, 1964), Chap. 4.

[7] E. W. Montroll, Proc. Symp. Appl. Math. Am. Math. Soc. **16**, 193 (1964).

[8] E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).

[9] Lifshitz's work appeared in the Russian literature in the 1940's. A summary appears in I. M. Lifshitz, Nuovo Cimento Suppl. **3**, 716 (1956).

on the same lattice in case the origin is an absorbing point. It is clear that if the origin is an absorbing point, all $(N-1)$-step random walk paths which start from the $2D$ nearest-neighbor lattice points to the origin and which are not located at the origin at the $(N-1)$th step are in one-to-one correspondence with the paths in $C_0(D;N)$. Consequently, we now consider the random walk problem when the origin is an absorbing point.

For simplicity of exposition, we treat the two-dimensional lattice explicitly. The result for the D-dimensional lattice is similar and is obtained in an identical manner. If $P(m_1, m_2; N)$ denotes the probability that the random walker is located at lattice site (m_1, m_2) at the Nth step, then $P(m_1, m_2; N+1)$ is related to the probabilities one step earlier by the expression

$$P(m_1, m_2; N+1) = \tfrac{1}{4}\{P(m_1-1, m_2; N) + P(m_1+1, m_2; N) + P(m_1, m_2-1; N) + P(m_1, m_2+1; N)\} \quad (1)$$

provided that (m_1, m_2) is not a nearest-neighbor to the origin, i.e., $(m_1, m_2) \neq (\pm 1, 0)$ and $(m_1, m_2) \neq (0, \pm 1)$. The probabilities $P(\pm 1, 0; N+1)$ and $P(0, \pm 1; N+1)$ are related to the probabilities one step earlier by expressions analogous to (1)

$$\begin{aligned} P(\pm 1, 0; N+1) &= \tfrac{1}{4}\{P(\pm 2, 0; N) + P(\pm 1, 1; N) + P(\pm 1, -1; N)\}, \\ P(0, \pm 1; N+1) &= \tfrac{1}{4}\{P(1, \pm 1; N) + P(-1, \pm 1; N) + P(0, \pm 2; N)\}. \end{aligned} \quad (2)$$

For convenience, we assume that the probability of being at the origin at the $(N+1)$st step is the sum of the probabilities of arrival from neighboring sites plus the probability of having been at the origin at the preceding step,

$$P(0, 0; N+1) = \tfrac{1}{4}\{P(1, 0; N) + P(-1, 0; N) + P(0, 1; N) + P(0, -1; N)\} + P(0, 0; N). \quad (3)$$

The recurrence equations (1)–(3) have the property that the total probability is conserved, i.e.,

$$\sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} P(m_1, m_2; N+1) = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} P(m_1, m_2; N).$$

In the next section we solve the recurrence equations (1)–(3) for the starting conditions

$$P(1, 0; 0) = P(-1, 0; 0) = P(0, 1; 0) = P(0, -1; 0) = \tfrac{1}{4}, \quad (4)$$

with all other $P(m_1, m_2; 0)$'s equal to zero. Once the solution $P(m_1, m_2; N)$ has been obtained, the mean square displacement at the Nth step is determined from the expression

$$R_N^2 = \frac{\sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty}{}' (m_1^2 + m_2^2) P(m_1, m_2; N)}{\sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty}{}' P(m_1, m_2; N)}, \quad (5)$$

where the primes on the double sums indicate that the $P(0, 0; N)$ term is omitted.

3. SOLUTION OF RECURRENCE EQUATIONS

The recurrence equations (1)–(3) can be solved for the initial condition (4) by introducing a generating function. Multiply the equation for $P(m_1, m_2; N+1)$ by $(2\pi)^{-1} \exp(i\theta_1 m_1 + i\theta_2 m_2)$ and sum over all values of m_1 and m_2. The result is

$$G(\theta_1, \theta_2; N+1) = \tfrac{1}{2}(\cos\theta_1 + \cos\theta_2) G(\theta_1, \theta_2; N) + (1/2\pi)[1 - \tfrac{1}{2}(\cos\theta_1 + \cos\theta_2)]\, P(0, 0; N), \quad (6)$$

where

$$G(\theta_1, \theta_2; N) = \frac{1}{2\pi} \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} P(m_1, m_2; N) \times \exp[i\theta_1 m_1 + i\theta_2 m_2].$$

Next, multiply the equation for $G(\theta_1, \theta_2, N+1)$ by y^{N+1} and sum from $N=0$ to $N=\infty$ to obtain

$$\begin{aligned} -G(\theta_1, \theta_2; 0) + \Gamma(\theta_1, \theta_2; y) &= \tfrac{1}{2} y(\cos\theta_1 + \cos\theta_2)\Gamma(\theta_1, \theta_2; y) \\ &\quad + (1/2\pi) y[1 - \tfrac{1}{2}(\cos\theta_1 + \cos\theta_2)] h(0, 0; y), \end{aligned} \quad (7)$$

where

$$\Gamma(\theta_1, \theta_2; y) = \sum_{N=0}^{\infty} y^N G(\theta_1, \theta_2; y), \quad (8)$$

$$h(m_1, m_2; y) = \sum_{N=0}^{\infty} y^N P(m_1, m_2; N), \quad (9)$$

$$\Gamma(\theta_1, \theta_2; y) = \frac{1}{2\pi} \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} h(m_1, m_2; y) \times \exp(im_1\theta_1 + im_2\theta_2). \quad (10)$$

When the starting values (4) are substituted in $G(\theta_1, \theta_2; 0)$ in Eq. (7), one obtains, after some rearranging of terms,

$$\Gamma(\theta_1, \theta_2; y) = \frac{\frac{1}{2}(\cos\theta_1 + \cos\theta_2) + (y-1)\,h(0,0;y)}{2\pi[1 - \frac{1}{2}y(\cos\theta_1 + \cos\theta_2)]} + \frac{1}{2\pi} y h(0,0;y). \tag{11}$$

Equation (11) is an implicit equation containing $h(0, 0; y)$ on both the right- and left-hand sides. An equation for determining $h(0, 0; y)$ can be obtained by multiplying Eq. (11) by $(2\pi)^{-1}$ and integrating with respect to θ_1 and θ_2 from $-\pi$ to $+\pi$. The result is

$$h(0,0;y) = \tfrac{1}{4}(I_{1,0} + I_{-1,0} + I_{0,1} + I_{0,-1}) + yh(0,0;y)[I_{0,0} - \tfrac{1}{4}(I_{1,0} + I_{-1,0} + I_{0,1} + I_{0,-1})], \tag{12}$$

where

$$I_{m,n} = \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} d\theta_1 \int_{-\pi}^{\pi} d\theta_2 \frac{\exp(im\theta_1 + in\theta_2)}{1 - \frac{1}{2}y(\cos\theta_1 + \cos\theta_2)}. \tag{13}$$

The integral $I_{0,0}$ in (12) is evaluated as

$$I_{0,0} = {}_2F_1[\tfrac{1}{2}, \tfrac{1}{2}; 1; y^2], \tag{14}$$

where ${}_2F_1[a, b; c; x]$ denotes the hypergeometric function; and the integrals $I_{\pm1,0}$ and $I_{0,\pm1}$ are equal and expressible in terms of $I_{0,0}$ as[5]

$$I_{1,0} = y^{-1}(I_{0,0} - 1). \tag{15}$$

As a consequence, the explicit expression for $h(0, 0; y)$ can be written as

$$h(0,0;y) = y^{-1}(1-y)^{-1}(1 - I_{0,0}^{-1}), \tag{16}$$

and that for $\Gamma(\theta_1, \theta_2; y)$ as

$$\Gamma(\theta_1, \theta_2; y) = \frac{1}{2\pi}\frac{1}{1-y} \times \left\{1 - \frac{1 - \frac{1}{2}(\cos\theta_1 + \cos\theta_2)}{1 - \frac{1}{2}y(\cos\theta_1 + \cos\theta_2)} \cdot \frac{1}{I_{0,0}}\right\}. \tag{17}$$

The generating function $\Gamma(\theta_1, \theta_2; y)$ in (17) is a weighted sum of all the $P(m_1, m_2; N)$'s. In order to obtain an explicit expression for R_N^2, the mean square displacement defined in terms of the $P(m_1, m_2; N)$'s in Eq. (5), we first subtract $(1/2\pi)h(0, 0; y)$ from $\Gamma(\theta_1, \theta_2; y)$ and then select the coefficient of y^N

$$\frac{1}{2\pi i}\int_{C_0} \frac{dy}{y^{N+1}}\left[\Gamma(\theta_1, \theta_2; y) - \frac{1}{2\pi} h(0,0;y)\right] = \frac{1}{2\pi} \sum_{m_1=-\infty}^{\infty} {\sum_{m_2=-\infty}^{\infty}}' P(m_1, m_2; N) \exp(i\theta_1 m_1 + i\theta_2 m_2), \tag{18}$$

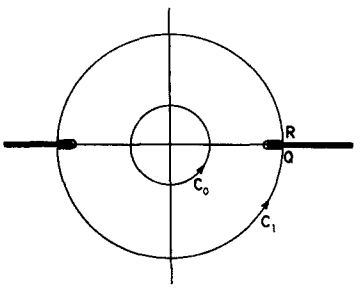


FIG. 1. Cuts and integration contours C_0 and C_1 in the complex y plane.

where C_0 is a counterclockwise contour around the origin in the complex y plane (see Fig. 1). The explicit expression for R_N^2 in the case of the 2-dimensional lattice is

$$R_N^2 = \frac{\frac{1}{2\pi i}\int_{C_0}\frac{dy}{y^{N+1}}\left(-\frac{d^2}{d\theta_1^2} - \frac{d^2}{d\theta_2^2}\right)\Gamma(\theta_1, \theta_2; y)\Big|_{\theta_1=\theta_2=0}}{\frac{1}{2\pi i}\int_{C_0}\frac{dy}{y^{N+1}}\left[\Gamma(0,0;y) - \frac{1}{2\pi}h(0,0;y)\right]} = \frac{\frac{1}{2\pi i}\int_{C_0}\frac{dy}{y^{N+1}}\frac{1}{(1-y)^2 I_{0,0}}}{\frac{1}{2\pi i}\int_{C_0}\frac{dy}{y^{N+2}}\frac{1}{(1-y)I_{0,0}}}, \tag{19}$$

where $I_{0,0}$ is the hypergeometric function (14). The expression for R_N^2 in a D-dimensional lattice is identical with (19) except that $I_{0,0}$ is replaced by

$$I_{0,\cdots,0} = \left(\frac{1}{2\pi}\right)^D \int_{-\pi}^{\pi} d\theta_1 \cdots \times \int_{-\pi}^{\pi} d\theta_D [1 - D^{-1}y(\cos\theta_1 + \cdots + \cos\theta_D)]^{-1}. \tag{20}$$

The solution of our random walk problem on a slightly defective lattice has led to a result involving $I_{0,\cdots,0}(y)$, the generating function for a random walk on a perfect lattice, a well-known connection.[3-9] However, the particular form of the result is somewhat unexpected. It is the ratio of two quantities appearing in the paper of Montroll and Weiss[8]

$$R_N^2 = S_N/(S_{N+1} - S_N) = S_N/\Delta_{N+1}, \tag{21}$$

where, in the notation of Montroll and Weiss, S_N is the average number of different lattice points visited in an N-step walk; and Δ_N, an auxilliary function, is the average number of new lattice points visited when the $(N + 1)$th step is taken. Montroll and Weiss have studied the function S_N in considerable detail, and we can utilize their results in one and three dimensions

to calculate R_N^2. In the case of a one-dimensional lattice, where the asymptotic series for S_N is

$$S_N \sim (8N/\pi)^{\frac{1}{2}}\{1 + (1/4N) - \cdots\},$$

we obtain the well-known result

$$R_N^2 \sim \frac{N^{\frac{1}{2}}\{1 + (1/4N)\}}{(N+1)^{\frac{1}{2}}\{1 + [1/4(N+1)]\} - N^{\frac{1}{2}}\{1 + (1/4N)\}} \sim 2N, \quad D = 1. \tag{22}$$

In the case of the three-dimensional simple cubic lattice, where the asymptotic series for S_N is

$$S_N \sim (N/u_0) + (2u_1/u_0^2)(N/\pi)^{\frac{1}{2}} + \cdots \tag{23}$$

and $u_0 = 1.51639$ and $u_1 = 1.16955$, we obtain the result

$$R_N^2 \sim \frac{\frac{N}{u_0} + \frac{2u_1}{u_0^2}\left(\frac{N}{\pi}\right)^{\frac{1}{2}}}{\frac{1}{u_0} + \frac{2u_1}{u_0^2}\pi^{-\frac{1}{2}}[(N+1)^{\frac{1}{2}} - N^{\frac{1}{2}}]} \sim N[1 + (0.4351/N^{\frac{1}{2}})], \quad D = 3. \tag{24}$$

Montroll and Weiss have only obtained the leading term in the asymptotic series for S_N in two dimensions using a Tauberian theorem. We have carried out the required analysis for S_N in this case in the Appendix; and the value of R_N^2 is

$$R_N^2 \sim N[1 + (1/\ln N)], \quad D = 2. \tag{25}$$

4. REMARKS

Although the calculations in this paper have been carried out explicitly for simple cubic lattices, it appears from an analysis along the lines used by Montroll[6,7] that Eq. (21) for R_N^2 is valid on face-centered and body-centered cubic lattices as well. Montroll and Weiss[8] have obtained the asymptotic series for S_N in these two lattices; and they only differ from the simple cubic lattice result (23) in the numerical values of u_0 and u_1.

The denominator Δ_{N+1} in Eq. (21) for R_N^2 is

$$\Delta_{N+1} = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty}{}' P(m_1, m_2; N),$$

the probability of not visiting the origin up to the Nth step. The asymptotic values which we have obtained for this quantity are

$$\Delta_{N+1} \sim \begin{cases} (2/\pi N)^{\frac{1}{2}}, & D = 1, \\ \pi/\ln N, & D = 2, \\ u_0^{-1} + u_1 u_0^{-2}\pi^{-\frac{1}{2}}N^{-1}, & D = 3. \end{cases} \tag{26}$$

These values are consistent with the results of Polya[10] that the probability of eventual return to the origin on a D-dimensional cubic lattice is unity in the infinite one- and two-dimensional lattices and $1 - (1/u_0)$ for the three-dimensional lattice. The results in Eq. (26) show the rate of approach to the limiting values.

It is of interest to compare the contribution to the mean square end-to-end distance of the $N-1$ nested overlap restrictions considered in this paper with the contributions of $N-1$ independent or separate overlap restrictions. It was shown in Ref. 2 that δ_s, the increase in R_N^2 resulting from the single restriction that the kth and the $(k+s)$th steps cannot overlap, is proportional to $s^{\frac{1}{2}}$, s^0, and $s^{-\frac{1}{2}}$ in the case of the one-, two-, and three-dimensional random walks, respectively. Thus, in the case of the 2-dimensional random walk, the quantity

$$\sum_{s=1}^{N-1} \delta_s$$

is proportional to N, whereas the actual contribution of the set of nested overlap restrictions to R_N^2 is $N/\ln N$.

APPENDIX. ASYMPTOTIC EXPRESSIONS FOR $R_N^2 = S_N/\Delta_{N+1}$ FOR $N \gg 1$ IN THE 2-DIMENSIONAL LATTICE

Montroll and Weiss[8] and Dvoretsky and Erdös[11] have given an asymptotic formula for S_N in the case of the 2-dimensional lattice. We use the contour integral expressions for S_N and Δ_N and obtain the first two terms in the asymptotic series for S_N/Δ_{N+1}. First consider the integral for Δ_N

$$\Delta_N = \frac{1}{2\pi i}\int_{C_0} \frac{dy}{y^{N+1}} \frac{1}{(1-y)_2F_1[\frac{1}{2}, \frac{1}{2}; 1; y^2]}. \tag{A1}$$

The integrand contains logarithmic branch points at $y = \pm 1$. Introduce cuts in the complex y plane which start at $y = \pm 1$ and extend out to $\pm\infty$ respectively as shown in Fig. 1. The contour C_0 can be deformed into the contour C_1 shown in Fig. 1. The dominant asymptotic contribution of the line integral around C_1 to Δ_N for $N \gg 1$ comes from the portion of the contour between Q and R in the immediate vicinity of $y = 1$, where there is a pole superimposed on the logarithmic branch point (see Fig. 1)

$$\Delta_N \sim \frac{1}{2\pi i}\int_Q^R \frac{dy}{y^{N+1}} \frac{1}{(1-y)_2F_1[\frac{1}{2}, \frac{1}{2}; 1; y^2]}. \tag{A2}$$

For our purposes it is sufficient to set the distance of Q and R from $y = 1$ equal to $N^{-\frac{1}{2}}$ and assume that

[10] G. Polya, Math. Ann. **84**, 149 (1921).

[11] A. Dvoretzky and P. Erdös, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, California, 1951), p. 353.

$\ln N \gg 1$. It is then possible to replace the hypergeometric function ${}_2F_1[\frac{1}{2}, \frac{1}{2}; 1; y^2]$ by its analytic continuation[12] where, in the interval of integration, the only significant term is

$$ {}_2F_1[\tfrac{1}{2}, \tfrac{1}{2}; 1; y^2] \sim \pi^{-1} \ln [1/(1-y)]. \tag{A3}$$

The phase of the logarithm is chosen so that the logarithm is real on the real axis between $y = 0$ and $y = 1$. Thus, we have for the value of ${}_2F_1[\frac{1}{2}, \frac{1}{2}; 1; y^2]$ on the lower and upper sides of the cut

$$ {}_2F_1[\tfrac{1}{2}, \tfrac{1}{2}; 1; y^2] \sim \begin{cases} \pi^{-1}\{\ln [1/(y-1)] - i\pi\}, & \text{lower}, \\ \pi^{-1}\{\ln [1/(y-1)] + i\pi\}, & \text{upper}. \end{cases} \tag{A4}$$

Substituting (A4) in (A2), one obtains the following expression for Δ_N

$$\begin{aligned} \Delta_N &\sim \frac{1}{2\pi i}\left\{\int_{1+N^{-\frac{1}{2}}}^{1} \frac{dy}{y^{N+1}} \frac{1}{1-y} \frac{\pi}{\ln [1/(y-1)] - i\pi} \right. \\ &\quad \left. + \int_{1}^{1+N^{-\frac{1}{2}}} \frac{dy}{y^{N+1}} \frac{1}{1-y} \frac{\pi}{\ln [1/(y-1)] + i\pi}\right\} \\ &\sim \pi \int_{1}^{1+N^{-\frac{1}{2}}} \frac{dy}{y^{N+1}} \frac{1}{y-1} \frac{1}{\ln^2 [1/(y-1)] + \pi^2}. \end{aligned} \tag{A5}$$

Replacing y by $1 + x/N$ in (A5), neglecting π^2, and integrating by parts, one obtains

$$\Delta_N \sim \pi \int_0^{N^{\frac{1}{2}}} \frac{dx}{[1 + (x/N)]^{N+2}} \frac{1}{\ln N - \ln x}. \tag{A6}$$

Now, consider the corresponding integral for S_N

$$S_N \sim \frac{1}{2\pi i} \int_Q^R \frac{dy}{y^{N+1}} \frac{1}{(1-y)^2} \frac{\pi}{\ln [1/(1-y)]}. \tag{A7}$$

Integrate (A7) by parts using the fact that

$$-\frac{d}{dy} E_1\left[-\ln\left(\frac{1}{1-y}\right)\right] = \frac{1}{(1-y)^2} \frac{1}{\ln [1/(1-y)]},$$

where

$$E_1(z) = \int_z^\infty e^{-t} t^{-1}\, dt$$

and obtain

$$\begin{aligned} S_N \sim \frac{N}{2i}\left\{ -\int_{1+N^{-\frac{1}{2}}}^{1} \frac{dy}{y^{N+2}} E_1\left[-\ln\left(\frac{1}{y-1}\right) + i\pi\right] \right. \\ \left. - \int_{1+N^{-\frac{1}{2}}}^{1} \frac{dy}{y^{N+2}} E_1\left[-\ln\left(\frac{1}{y-1}\right) - i\pi\right]\right\}. \end{aligned} \tag{A8}$$

In the interval of integration in (A8), one can replace the exponential integrals[13]

$$\begin{aligned} E_1\left[-\ln\left(\frac{1}{y-1}\right) \pm i\pi\right] &\sim E_1\left[-\ln\left(\frac{1}{y-1}\right)\right] \\ &\quad + \frac{1}{y-1}\left\{\frac{2}{\ln [1/(y-1)]} + \frac{2 \pm i\pi}{\ln^2 [1/(y-1)]} \right. \\ &\quad \left. + \frac{4 - \pi^2 \pm 2i\pi}{\ln^3 [1/(y-1)]} + \cdots\right\}. \end{aligned}$$

The result is

$$\begin{aligned} S_N &\sim N\pi \int_1^{1+N^{-\frac{1}{2}}} \frac{dy}{y^{N+2}} \frac{1}{y-1} \\ &\quad \times \left\{\frac{1}{\ln^2 [1/(y-1)]} + \frac{2}{\ln^3 [1/(y-1)]} + \cdots\right\}. \end{aligned} \tag{A9}$$

Finally, as in Eq. (A5), replace y by $1 + x/N$, and integrate by parts

$$\begin{aligned} S_N &\sim N\pi \int_0^{N^{\frac{1}{2}}} \frac{dx}{[1 + (x/N)]^{N+3}} \\ &\quad \times \left\{\frac{1}{\ln N - \ln x} + \left(\frac{1}{\ln N - \ln x}\right)^2 + \cdots\right\}. \end{aligned} \tag{A10}$$

The expression for R_N^2, from (A6) and (A10) is

$$\begin{aligned} R_N^2 &= S_N/\Delta_{N+1} \\ &\sim N + N \frac{\displaystyle\int_0^{N^{\frac{1}{2}}} \frac{dx}{[1 + (x/N)]^{N+3}} \left(\frac{1}{\ln N - \ln x}\right)^2}{\displaystyle\int_0^{N^{\frac{1}{2}}} \frac{dx}{[1 + (x/N)]^{N+3}} \left(\frac{1}{\ln N - \ln x}\right)}. \end{aligned} \tag{A11}$$

In arriving at Eq. (A11), we have consistently neglected terms of higher order in $(\ln N)^{-1}$. We must now estimate the value of the two integrals in (A11). This can be done by splitting the interval of integration into three parts: from 0 to $(\ln N)^{-1}$, from $(\ln N)^{-1}$ to $\ln N$, and from $\ln N$ to $N^{\frac{1}{2}}$. It is a simple matter to show that for the denominator

$$\begin{aligned} &\int_0^{(\ln N)^{-1}} \frac{dx}{[1 + (x/N)]^{N+3}} \frac{1}{\ln N - \ln x} < (\ln N)^{-2}, \\ &[\ln N + \ln (\ln N)]^{-1} \\ &\qquad < \int_{(\ln N)^{-1}}^{\ln N} \frac{dx}{[1 + (x/N)]^{N+3}} \frac{1}{\ln N - \ln x}, \\ &[\ln N - \ln (\ln N)]^{-1}, \\ &\qquad > \int_{(\ln N)^{-1}}^{\ln N} \frac{dx}{[1 + (x/N)]^{N+3}} \frac{1}{\ln N - \ln x}, \end{aligned} \tag{A12}$$

[12] A. Erdelyi, W. Magnus, and F. Oberhettinger, *Higher Transcendental Functions* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 1, p. 110.

[13] A. Erdelyi, W. Magnus, and F. Oberhettinger, *Higher Transcendental Functions* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 2, p. 145.

and

$$\int_{\ln N}^{N^{\frac{1}{2}}} \frac{dx}{[1+(x/N)]^{N+3}} \frac{1}{\ln N - \ln x} < 2(\ln N)^{-1}(1+N^{-1}\ln N)^{-(N+2)}.$$

Therefore, the dominant contribution of the denominator is $(\ln N)^{-1}$. The same procedure can be used for the numerator to give

$$\int_0^{N^{\frac{1}{2}}} \frac{dx}{[1+(x/N)]^{N+3}} \left(\frac{1}{\ln N - \ln x}\right)^2 \sim (\ln N)^{-2}.$$

Thus, we finally arrive at the following estimate

$$R_N^2 \sim N[1+(\ln N)^{-1}]. \quad \text{(A13)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Short Simple Evaluation of Expressions of the Debye–Waller Form Valid for Anharmonic Modes not in Thermal Equilibrium

T. D. SCHULTZ
IBM Watson Research Center, Yorktown Heights, New York

(Received 13 August 1966)

A one-sentence nonalgebraic evaluation of averages, such as occur in the theory of the Debye–Waller factor, is given, which is valid for certain types of anharmonic modes and also for ensembles more general than the canonical ensemble.

IN a recent paper, Mermin[1] reconsidered the evaluation of the average

$$\langle e^S \rangle \equiv \text{tr}\, \rho e^S / \text{tr}\, \rho,$$

where

$$S = \sum_i (c_i a_i + d_i a_i^\dagger), \quad \rho = \exp(-\beta \sum_i \omega_i a_i^\dagger a_i), \quad \beta = 1/k_B T, \quad (1)$$

and the a_i and $a_i^\dagger$ are boson annihilation and creation operators. He has given an algebraic derivation of the well-known result

$$\langle e^S \rangle = \exp[\tfrac{1}{2} \sum_i c_i d_i \coth \tfrac{1}{2}\beta\omega_i] \quad (2)$$

in one sentence, albeit a rather long one, using the identity $e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]}$ ($[A, B]$ a c-number).

Here, we offer another (and shorter) one-sentence derivation which, though restricted to a macroscopic lattice (i.e., very many nonvanishing c_i's and/or d_i's), is *more general* than the usual *algebraic* proofs in that we assume only that the density operator ρ has the form

$$\rho = \prod_i p_i(n_i), \quad n_i = a_i^\dagger a_i \quad (3)$$

with the p_i any reasonably general probability density functions [i.e., $p_i(n_i)$ should be a positive semi-definite operator with unit trace].

DERIVATION

If we calculate $\langle e^S \rangle$ in the representation in which all the operators $X_i \equiv c_i a_i + d_i a_i^\dagger$ are diagonal, the diagonal elements of p_i define a normalized set of classical probabilities and the Central Limit Theorem of probability theory,[2] which is then applicable, states that the sum

$$S \equiv \sum X_i$$

has a Gaussian distribution (in this case with mean zero, because $\langle X_i \rangle \equiv 0$), from which we deduce

$$\begin{aligned} \langle e^S \rangle &= \exp \tfrac{1}{2}\langle S^2 \rangle = \exp \tfrac{1}{2} \sum_i \langle X_i^2 \rangle \\ &= \exp[\sum_i c_i d_i (\langle n_i \rangle + \tfrac{1}{2})], \quad \text{for general } p_i(n_i), \\ &= \exp[\tfrac{1}{2} \sum c_i d_i \coth \tfrac{1}{2}\beta\omega_i], \quad \text{if } p_i(n_i) \propto e^{-\beta\omega_i n_i}. \end{aligned} \quad (4)$$

REMARKS

It should be emphasized that all other proofs, including Mermin's, implicitly or explicitly make use of the fact that *each* normal mode amplitude has a Gaussian distribution at any temperature (the utilization of this fact is the aim of all the various algebraic maneuvers), a fact which is unnecessary in treating a macroscopic system provided only that no small set of modes contributes to $\langle e^S \rangle$ out of proportion to its number. The derivation is thus valid for independent but anharmonic modes, and for independent modes not at thermal equilibrium, but not, for example for the Mössbauer effect whenever the emitting (or absorbing) nucleus produces a localized mode in the lattice.

[1] N. D. Mermin, J. Math. Phys. 7, 1038 (1966).

[2] See any book on probability theory, such as R. v. Mises, *Mathematical Theory of Probability and Statistics* (Academic Press Inc., New York, 1964), p. 294; or W. Feller, *Probability Theory and Its Applications* (John Wiley & Sons, Inc., New York, 1950), Vol. I, p. 201.

and

$$\int_{\ln N}^{N^{\frac{1}{2}}} \frac{dx}{[1+(x/N)]^{N+3}} \frac{1}{\ln N - \ln x} < 2(\ln N)^{-1}(1+N^{-1}\ln N)^{-(N+2)}.$$

Therefore, the dominant contribution of the denominator is $(\ln N)^{-1}$. The same procedure can be used for the numerator to give

$$\int_0^{N^{\frac{1}{2}}} \frac{dx}{[1+(x/N)]^{N+3}} \left(\frac{1}{\ln N - \ln x}\right)^2 \sim (\ln N)^{-2}.$$

Thus, we finally arrive at the following estimate

$$R_N^2 \sim N[1+(\ln N)^{-1}]. \quad \text{(A13)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Short Simple Evaluation of Expressions of the Debye–Waller Form Valid for Anharmonic Modes not in Thermal Equilibrium

T. D. SCHULTZ
IBM Watson Research Center, Yorktown Heights, New York

(Received 13 August 1966)

A one-sentence nonalgebraic evaluation of averages, such as occur in the theory of the Debye–Waller factor, is given, which is valid for certain types of anharmonic modes and also for ensembles more general than the canonical ensemble.

IN a recent paper, Mermin[1] reconsidered the evaluation of the average

$$\langle e^S \rangle \equiv \text{tr}\, \rho e^S / \text{tr}\, \rho,$$

where

$$S = \sum_i (c_i a_i + d_i a_i^\dagger), \quad \rho = \exp(-\beta \sum_i \omega_i a_i^\dagger a_i), \quad \beta = 1/k_B T, \quad (1)$$

and the a_i and $a_i^\dagger$ are boson annihilation and creation operators. He has given an algebraic derivation of the well-known result

$$\langle e^S \rangle = \exp[\tfrac{1}{2} \sum_i c_i d_i \coth \tfrac{1}{2}\beta\omega_i] \quad (2)$$

in one sentence, albeit a rather long one, using the identity $e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]}$ ($[A, B]$ a c-number).

Here, we offer another (and shorter) one-sentence derivation which, though restricted to a macroscopic lattice (i.e., very many nonvanishing c_i's and/or d_i's), is *more general* than the usual *algebraic* proofs in that we assume only that the density operator ρ has the form

$$\rho = \prod_i p_i(n_i), \quad n_i = a_i^\dagger a_i \quad (3)$$

with the p_i any reasonably general probability density functions [i.e., $p_i(n_i)$ should be a positive semi-definite operator with unit trace].

DERIVATION

If we calculate $\langle e^S \rangle$ in the representation in which all the operators $X_i \equiv c_i a_i + d_i a_i^\dagger$ are diagonal, the diagonal elements of p_i define a normalized set of classical probabilities and the Central Limit Theorem of probability theory,[2] which is then applicable, states that the sum

$$S \equiv \sum X_i$$

has a Gaussian distribution (in this case with mean zero, because $\langle X_i \rangle \equiv 0$), from which we deduce

$$\begin{aligned} \langle e^S \rangle &= \exp \tfrac{1}{2}\langle S^2 \rangle = \exp \tfrac{1}{2} \sum_i \langle X_i^2 \rangle \\ &= \exp[\sum_i c_i d_i (\langle n_i \rangle + \tfrac{1}{2})], \quad \text{for general } p_i(n_i), \\ &= \exp[\tfrac{1}{2} \sum c_i d_i \coth \tfrac{1}{2}\beta\omega_i], \quad \text{if } p_i(n_i) \propto e^{-\beta\omega_i n_i}. \end{aligned} \quad (4)$$

REMARKS

It should be emphasized that all other proofs, including Mermin's, implicitly or explicitly make use of the fact that *each* normal mode amplitude has a Gaussian distribution at any temperature (the utilization of this fact is the aim of all the various algebraic maneuvers), a fact which is unnecessary in treating a macroscopic system provided only that no small set of modes contributes to $\langle e^S \rangle$ out of proportion to its number. The derivation is thus valid for independent but anharmonic modes, and for independent modes not at thermal equilibrium, but not, for example for the Mössbauer effect whenever the emitting (or absorbing) nucleus produces a localized mode in the lattice.

[1] N. D. Mermin, J. Math. Phys. 7, 1038 (1966).

[2] See any book on probability theory, such as R. v. Mises, *Mathematical Theory of Probability and Statistics* (Academic Press Inc., New York, 1964), p. 294; or W. Feller, *Probability Theory and Its Applications* (John Wiley & Sons, Inc., New York, 1950), Vol. I, p. 201.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Quaternionic Representations of Compact Metric Groups

S. NATARAJAN AND K. VISWANATH
Indian Statistical Institute, Calcutta, India

(Received 7 April 1966)

Representations of compact metric groups in Hilbert spaces over the quaternions are studied. A generalization of the Peter–Weyl theorem is formulated and proved. The problem of finding all the irreducible quaternionic representations of an arbitrary compact metric group is solved, and a rule is given for computing the "Q-characters" of all the irreducible quaternionic representations once the characters of all the irreducible complex representations are known. For the Abelian case, it is shown that every irreducible quaternionic representation is equivalent to a complex representation and hence one dimensional. An example is given of a non-Abelian group whose irreducible quaternionic representations are all one dimensional.

I. INTRODUCTION

IT is well known (see, e.g., Birkhoff and von Neumann,[1] Yang,[2] Mackey,[3] Michel[4]) that the lattice of closed linear manifolds of a quaternionic Hilbert space is a possible candidate for the logic of propositions (see Varadarajan[5]) of a quantum mechanical system, and that there is nothing canonical about the (classical) choice of the complex number system for the development of quantum mechanics. But, in spite of the wide-spread knowledge of this fact, very little work has been done toward setting up a theory of quaternionic quantum mechanics apart from the fundamental work[6–9] of Finkelstein, Jauch, Speiser, and Schiminovitch. We hope that our present work is of some help in this context, as the theory of group representations is indispensable for the exposition of quantum mechanics and compact metric groups are an important special case.

II. PRELIMINARY IDEAS

We present this section in some detail as our orientation differs from that of Finkelstein *et al.*

Let Q denote the division ring of real quaternions. We denote an arbitrary element q of Q by $q = q_0 + q_1 i + q_2 j + q_3 k$, where q_0, q_1, q_2, q_3 are real. We identify the reals with the set of all quaternions q with $q_1 = q_2 = q_3 = 0$ and the complex numbers with the set of all quaternions q with $q_2 = q_3 = 0$. Every $q \in Q$ may be written in the form $\alpha + \beta j$, where α and β are complex. We denote by q^* the conjugate of the quaternion q.

1. Vector Spaces

By a vector space over Q (to be called a Q-space) we always mean a left-vector space over Q. A Q-Banach space is a complete normed Q-space. If X is a topological space, we denote by $\mathsf{C}_Q(X)$ the Q-Banach space of all bounded quaternion-valued continuous functions on X with the supremum norm.

An inner product on a Q-space V is a quaternion-valued function on $\mathsf{V} \times \mathsf{V}$, denoted by $(\ldots)$, with the properties:

(i) $(x, y) = (y, x)^*$,

(ii) $(px + p'x', y) = p(x, y) + p'(x', y)$,

(iii) $(x, x) \geq 0, = 0$ if and only if $x = 0$,

where $x, x', y \in \mathsf{V}$, and $p, p' \in Q$. From (i) and (ii) we have

$$(x, py + p'y') = (x, y)p^* + (x, y')p'^*.$$

It is easy to prove that, on an inner product Q-space, $\|x\| = (x, x)^{\frac{1}{2}}$ defines a norm.[7] A Q-space V is called a Q-Hilbert space if there exists an inner product on V such that the induced norm makes V a complete normed Q-space. The concepts of orthogonality, basis, etc., for Q-Hilbert spaces are defined in the usual way. In what follows H denotes a Q-Hilbert space.

An operator on H is a bounded linear transformation of H into itself. An automorphism of H is a bijective operator on H. For every automorphism A, there exists an unique automorphism A^{-1} such that $AA^{-1} = A^{-1}A = I$. The set of all automorphisms is a group in a natural way.

[1] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936).
[2] C. N. Yang, in *Proceedings of the Seventh Annual Rochester Conference* (Interscience Publishers, Inc., New York, 1957), p. IX-26.
[3] G. W. Mackey, *The Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, Inc., New York, 1963), p. 73.
[4] L. Michel, *Invariance in Quantum Mechanics and Group Extension, Group-Theoretical Concepts and Methods in Elementary Particles* (Gordon and Breach Science Publishers, Inc, New York, 1964), p. 148.
[5] V. S. Varadarajan, Indian Statistical Institute preprint (1965), p. 207.
[6] D. Finkelstein, J. M. Jauch, and D. Speiser, "Notes on Quaternion Quantum Mechanics I, II, and III", CERN (1959).
[7] D. Finkelstein, J. M. Jauch, S. Schiminovitch, and D. Speiser, J. Math. Phys. **3**, 207 (1962).
[8] D. Finkelstein, J. M. Jauch, S. Schiminovitch, and D. Speiser, J. Math. Phys. **4**, 788 (1963).
[9] D. Finkelstein, J. M. Jauch, and D. Speiser, J. Math. Phys. **4**, 136 (1963).

The elementary theory of Q-Hilbert spaces can now be developed as in the complex case. We note in particular that, for every operator A on H, there exists a unique operator A^* on H such that $(Ax, y) = (x, A^*y)$ for all $x, y \in \mathsf{H}$. A^* is called the adjoint of A. An operator A on H is called Hermitian if $A = A^*$ and unitary if $AA^* = A^*A = I$.

The spectral theory of Hermitian operators in Q-Hilbert spaces parallels the theory in the complex case.[6,7]

Let now V be a finite-dimensional Q-space. (Note that V may be endowed with a Q-Hilbert space structure.) Given a basis $(e_1, \cdots, e_n)$ of V, every linear transformation A on V has a matrix representation (a_{rs}), defined by

$$Ae_s = \sum_r a_{rs}e_r .$$

If A and B are two linear transformations with matrices (a_{rs}) and (b_{rs}), respectively, then the matrix of AB is given by (c_{rs}), where

$$c_{rs} = \sum_t b_{ts}a_{rt} .$$

Observe that our rule for matrix multiplication differs from the usual rule for matrices over a field.

If A has the matrix (a_{rs}) with respect to an orthonormal basis (e_r), then $a_{rs} = (Ae_s, e_r)$. The matrix of A^* with respect to the same basis is then (b_{rs}), where $b_{rs} = (A^*e_s, e_r) = a_{sr}^*$. If A is Hermitian, then $A = A^*$ and hence $a_{rs} = a_{sr}^*$. If A is unitary, $A^*A = AA^* = I$ and hence

$$\sum_t a_{st}^* a_{rt} = \delta_{rs} = \sum_t a_{ts}a_{tr}^* .$$

We note here that, if A has the matrix (a_{rs}) with respect to a basis (e_r), then

$$\mathrm{Re}\,(\mathrm{tr}\,A) = \mathrm{Re}\,(\sum_r a_{rr})$$

is defined independently of the basis (e_r).

2. The Symplectic Picture

It is convenient for our purposes to restate the usual definition[10] in geometric language.

If V is a Q-space, then the additive group of V can be considered as a C-space (i.e., a vector space over the complex numbers). This we denote by V^C and call the symplectic picture of V. If $(e_1, \cdots, e_n)$ is a basis for V, then $(e_1, \cdots, e_n, je_1, \cdots, je_n)$ is a basis for V^C. Hence V^C is of dimension $2n$. A linear transformation A on V is also a linear transformation on V^C. This we denote by A^C. If the matrix of A with respect to the basis $(e_1, \cdots, e_n)$ is $A_1 + A_2 j$, where A_1 and A_2 are complex matrices, then the matrix of A^C with respect to the basis $(e_1, \cdots, e_n, je_1, \cdots, je_n)$ is

$$\begin{vmatrix} A_1 & A_2 \\ -\bar{A}_2 & \bar{A}_1 \end{vmatrix},$$

where $\bar{\alpha}$ denotes the complex conjugate of the complex number α and $\bar{B} = (\bar{b}_{rs})$ if B is the complex matrix (b_{rs}).

3. Integration Theory

Let (X, Σ, μ) be a measure space. We always identify functions which differ only on μ-null sets. A quaternion-valued measurable function

$$f(x) = f_0(x) + f_1(x)i + f_2(x)j + f_3(x)k$$

on X, where f_r [$r = 0, 1, 2, 3$, are real-valued (measurable) functions on X] is said to be integrable with respect to μ if and only if f_0, f_1, f_2, f_3 are integrable with respect to μ. If f is integrable, the integral of f with respect to μ is defined as

$$\int f\,d\mu = \int f_0\,d\mu + \left(\int f_1\,d\mu\right)i + \left(\int f_2\,d\mu\right)j + \left(\int f_3\,d\mu\right)k.$$

The following properties of the integral are easily verified ($q \in Q$ is arbitrary):

(i) $\int (f + g)\,d\mu = \int f\,d\mu + \int g\,d\mu,$

(ii) $\int (pfq)\,d\mu = p\left(\int f\,d\mu\right)q,$

(iii) $\left(\int f\,d\mu\right)^* = \int f^*\,d\mu,$

(iv) $\left|\int f\,d\mu\right| \leq \int |f|\,d\mu.$

The only nontrivial relation is (iv). This may be proved by a slight modification of Cramér's proof[11] for the complex case.

We define $\mathsf{L}_Q^2(X)$ as the set of all quaternion-valued measurable functions f such that $|f|^2$ is integrable with respect to μ. It follows that $f \in \mathsf{L}_Q^2(X)$ implies that $f^* \in \mathsf{L}_Q^2(X)$. If we define for f and g in $\mathsf{L}_Q^2(X)$ $(f, g) = \int fg^*\,d\mu$ then $\mathsf{L}_Q^2(X)$ becomes a Q-Hilbert space with $(.\,,.)$ as inner product.

If $f, g \in \mathsf{L}_Q^2(X)$ and $\int f^*g\,d\mu = 0$, we say that f and g are left orthogonal. If f and g are also orthogonal, we say that f and g are bothways orthogonal.

[10] C. Chevalley, *Theory of Lie Groups, I* (Princeton University Press, Princeton, New Jersey, 1946), p. 18.

[11] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946), p. 65.

We note that if $f \in L^2_Q(X)$ and $p \in Q$, then $fp \in L^2_Q(X)$. If f and g are left orthogonal then fp and gq are left orthogonal for any $p, q \in Q$.

III. Q-REPRESENTATIONS

In what follows, we denote by G a compact metric group and by μ the unique normalized Haar measure on Σ, the class of Borel sets of G.

Let H be a separable Q-Hilbert space and $\mathsf{A}(\mathsf{H})$ the group of automorphisms of H. By a Q-representation[9] A of G in H we mean a homomorphism $g \to A_g$ from G to $\mathsf{A}(\mathsf{H})$ such that $g \to A_g x$ from G to H is continuous for every fixed $x \in \mathsf{H}$. The Q-representation A is called unitary if A_g is unitary for every $g \in G$. An example of a Q-representation of G in $L^2_Q(G)$ is the right regular representation. This is, in fact, unitary.

When H is finite-dimensional, we may, on occasion, regard the A_g as matrices with respect to some fixed basis of H.

The notions of equivalence, irreducibility, etc., of Q-representations are defined in the usual way.[12]

We now state some basic theorems. The departure from the complex case is only slight and so we omit the proofs.

Theorem 1: Any Q-representation A of G in H is equivalent to a unitary Q-representation.

Theorem 2: Every unitary Q-representation of G is a direct sum of irreducible unitary Q-representations of G. Every irreducible Q-representation of G is finite-dimensional.

The irreducible unitary Q-representations of G split up into equivalence classes in a natural way. We shall index these equivalence classes by α. (It follows from our analysis that the set of all α's is countable.) Let n_α be the dimension of any irreducible Q-representation of type α.

Consider now a unitary Q-representation of G in H. Let

$$\mathsf{H} = \bigoplus_t S_t$$

be a direct sum decomposition of H into irreducible subspaces and let the irreducible subspaces S_t of type α be indexed by a set of cardinality c_α. We call c_α the multiplicity of type α in the decomposition

$$\mathsf{H} = \bigoplus_t S_t .$$

Theorem 3: In any decomposition of H into irreducible subspaces the same types occur with the same multiplicities.

Schur's Lemma[9]*:* Let H_1 and H_2 be two finite-dimensional Q-spaces. Let (A_α) and (B_β) be irreducible collections of linear transformations on H_1 and H_2, respectively. If M is any linear transformation from H_1 to H_2 such that $(B_\beta M) = (M A_\alpha)$, then M is either 0 or an isomorphism.

Corollary 1: If U and V be two inequivalent irreducible unitary Q-representations of G in Q-Hilbert spaces H_1 and H_2, respectively, then

$$\int (V_g M U_g^{-1} x, y)\, dg = 0, \quad x \in \mathsf{H}_1, \quad y \in \mathsf{H}_2$$

for any linear transformation M from H_1 to H_2.

Corollary 2: Let U be an irreducible unitary Q-representation of G in a Q-Hilbert space H of dimension n. Then for any Hermitian operator M of H into itself

$$\int (U_g M U_g^{-1} x, y)\, dg = \frac{\mathrm{Re}\,(\mathrm{tr}\, M)}{n} (x, y).$$

Remark: Note that with our geometric approach Corollary 2 may be proved directly without invoking the *ersatz* determinant used by Finkelstein *et al.*[9]

IV. ORTHOGONALITY RELATIONS AND THE PETER–WEYL THEOREM

We now begin an analysis of the irreducible (and hence finite-dimensional) Q-representations of a compact metric group G.

Let A be an irreducible Q-representation of G in H of dimension n and let $[a_{rs}(g)]$ be the matrix of A_g with respect to an orthonormal basis (e_r). The function $a_{rs}(g) = (A_g e_r, e_s)$ is a continuous function on G for every r, s; i.e., the matrix entries $[a_{rs}(.)]$ of A with respect to an orthonormal basis are continuous. It follows that the matrix entries of A with respect to any basis of H are continuous, i.e., are elements of $C_Q(G)$ and hence of $L^2_Q(G)$.

We know that (see Theorem 24 in Ref. 13) in the complex case the matrix entries of two inequivalent irreducible unitary representations are orthogonal. A similar result holds in the quaternionic case. To see this, let U and V be inequivalent irreducible unitary Q-representations acting on Q-Hilbert spaces H and K, respectively, and let $u_{rs}(g)$ [respectively $v_{rs}(g)$] be the matrix entries of $U_g(V_g)$ with respect to the orthonormal basis $(e_s)[(f_s)]$. If $M: \mathsf{H} \to \mathsf{K}$ is the linear

[12] G. W. Mackey, "Theory of Group Representations," Lecture Notes, The University of Chicago (1955), p. 3.

[13] L. Pontrjagin, *Topological Groups* (Princeton University Press, Princeton, New Jersey, 1958).

transformation defined by $Me_t = f_w$, $Me_s = 0$ if $s \neq t$, then by Corollary 1 to Schur's Lemma

$$0 = \int (V_g M U_g^{-1} e_r, f_s)\, dg$$
$$= \int u_{rt}^*(g) v_{sw}(g)\, dg,$$

and also by the invariance of the integral,

$$= \int (V_g^{-1} M U_g e_r, f_s)\, dg$$
$$= \int u_{tr}(g) v_{ws}^*(g)\, dg.$$

In words, every matrix entry of U is bothways orthogonal to every matrix entry of V.

To study the orthogonal relations between the matrix entries of a single representation U, let $M: \mathsf{H} \to \mathsf{H}$ be the linear transformation defined by $Me_t = e_w$ and $Me_s = 0$ if $s \neq t$. Then we have, as above,

$$\int (U_g M U_g^{-1} e_r, e_s)\, dg = \int u_{rt}^*(g) u_{sw}(g)\, dg$$
$$= \int u_{tr}(g) u_{ws}^*(g)\, dg. \quad \text{(A)}$$

In case $r = s$ and $t = w$, M is Hermitian with $\mathrm{Re}\,(\mathrm{tr}\, M) = 1$ and so, from Corollary 2 to Schur's Lemma, it follows that

$$\int |u_{ws}(g)|^2\, dg = \frac{1}{n} \quad \text{for all } w, s.$$

Further, Eq. (A) shows that the n^2 matrix entries are mutually orthogonal if and only if they are mutually left orthogonal. However, in contrast to the complex case, it is not necessary that they be orthogonal, as the following example shows.

Example 1: Let G be the symmetric group of degree 3. The elements of G are

$$g_0 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix},\ g_1 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix},\ g_2 = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \end{pmatrix},$$
$$g_3 = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \end{pmatrix},\ g_4 = \begin{pmatrix} 0 & 1 & 2 \\ 2 & 0 & 1 \end{pmatrix},\ g_5 = \begin{pmatrix} 0 & 1 & 2 \\ 2 & 1 & 0 \end{pmatrix}.$$

Let $I = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$, $J = \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}$ and define

$$U_0 = I, \quad U_1 = [(i+j)/\sqrt{2}]J,$$
$$U_2 = [(\sqrt{3}-1)i/2\sqrt{2} - (\sqrt{3}+1)j/2\sqrt{2}]J,$$
$$U_3 = [(-1+\sqrt{3}k)/2]I, \quad U_4 = [(-1-\sqrt{3}k)/2]I,$$
$$U_5 = [-(\sqrt{3}+1)i/2\sqrt{2} + (\sqrt{3}-1)j/2\sqrt{2}]J.$$

Then the representation $g_i \to U_i$ is unitary and, moreover, irreducible, because the only vector sent into a multiple of itself by all the U_i is the null vector. Since in each matrix the two elements in the principal diagonal are equal, two of the matrix entries are identical.

Let $A_g = [a_{rs}(g)]$ be any irreducible Q-representation of G of type α. Define

$$\mathsf{F}_\alpha = \text{Span } [a_{rs}(g)q: 1 \leq r, s \leq n_\alpha, q \in Q].$$

It is easy to check that F_α depends only on the type α of the representation and not on the particular representation chosen. We call F_α the space of matrix entries of type α. Since every element of the generating set of F_α is a (real) linear combination of the $4n_\alpha^2$ elements of the type

$$a_{rs}(g),\ a_{rs}(g)i,\ a_{rs}(g)j,\ a_{rs}(g)k, \quad 1 \leq r, s \leq n_\alpha,$$

F_α is a closed linear manifold of $\mathsf{L}_Q^2(G)$ of dimension at most $4n_\alpha^2$ (see also Theorem 11, this paper).

The following theorem generalizes the Peter–Weyl theorem to the quaternionic case.

Theorem 4: The subspaces F_α and F_β are bothways orthogonal if $\alpha \neq \beta$. If $\sum_\alpha \mathsf{F}_\alpha$ denotes the set of finite sums of elements of $\bigcup_\alpha \mathsf{F}_\alpha$, where α ranges over all types and $\overline{\sum_\alpha \mathsf{F}_\alpha}$ the uniform closure of $\sum_\alpha \mathsf{F}_\alpha$, then

$$\overline{\sum_\alpha} \mathsf{F}_\alpha = \mathsf{C}_Q(G) \quad \text{and} \quad \bigoplus_\alpha \mathsf{F}_\alpha = \mathsf{L}_Q^2(G).$$

Proof: Let $[u_{rs}(g)]$, $[v_{rs}(g)]$ be unitary representations of types α and β respectively. For any $p, q \in Q$

$$\int [u_{rs}(g)p][v_{tw}(g)q]^*\, dg$$
$$= pq^* \int [(pq^*)^{-1} u_{rs}(g) pq^*] v_{tw}^*(g)\, dg = 0,$$

by the orthogonality relations proved earlier, since, for any quaternion q, the representation $[q^{-1}u_{rs}(g)q]$ is equivalent to $[u_{rs}(g)]$. Since the elements of F_α and F_β are linear combinations of elements of the form $[u_{rs}(g)p]$ and $[v_{tw}(g)q]$, respectively, we have shown that F_α and F_β are orthogonal. To prove that F_α and F_β are left orthogonal, it is enough to show that, for $p, q \in Q$, $pu_{rs}(g)$ and $qv_{tw}(g)$ are left orthogonal. But

$$\int [pu_{rs}(g)]^* [qv_{tw}(g)]\, dg$$
$$= \int u_{rs}^*(g) [p^* q v_{tw}(g)(p^*q)^{-1}]\, dg(p^*q) = 0.$$

For the second part, let us denote by Δ the set of all real functions arising from all possible real representations of G. Then (Ref. 13, p. 119) the finite real

linear combinations of elements of Δ are dense in $C_R(G)$, the Banach space of real-valued continuous functions on G. It follows that finite quaternion linear combinations of elements of Δ are dense in $C_Q(G)$.

Therefore, to prove that

$$\overline{\sum_\alpha F_\alpha} = C_Q(G),$$

it is enough to show that every function in Δ is a linear combination (and hence a finite sum) of functions in $U_\alpha F_\alpha$. But since every real representation A is equivalent to a direct sum of irreducible Q-representations and since the matrix entries of irreducible Q-representations belong to $U_\alpha F_\alpha$, it follows that every matrix entry of A and hence every element of Δ is a linear combination of elements of $U_\alpha F_\alpha$.

Since $C_Q(G)$ is dense in $L^2_Q(G)$ and uniform convergence implies L^2-convergence and since the F_α are mutually orthogonal subspaces of $L^2_Q(G)$, we have

$$L^2_Q(G) = \bigoplus_\alpha F_\alpha.$$

Corollary: There exists at most a countable number of inequivalent irreducible Q-representations of G.

Proof: $L^2_Q(G)$ is separable.

The following theorem (cf. Ref. 13, p. 120) may now be proved exactly as in the complex case.

Theorem 5: We select one representative from each equivalence class of irreducible Q-representations of G and denote them by

$$U^{(1)}, \cdots, U^{(n)}, \cdots.$$

Then for every element $g \in G$ distinct from the identity, there exists an n such that $U^{(n)}_g$ is not the identity transformation.

V. Q-CHARACTERS

Let $A_g = [a_{rs}(g)]$ be a Q-representation of G of degree n. Define

$$X(A_g) = \mathrm{Re}\,[\sum_r a_{rr}(g)].$$

Then it is easy to see that if A and B are equivalent Q-representations, then $X(A_g) = X(B_g)$. In this way we may associate with every equivalence class of Q-representations a real-valued function $X(g)$ which we call (see also Finkelstein, Jauch, and Speiser[9]) its Q-character (to distinguish it from the usual definition of the character of a complex representation which we call the C-character). We denote by $X_\alpha(g)$ the Q-character of any irreducible Q-representation of type α. Note that if A_g is of type α, then

$$X_\alpha(g) = \tfrac{1}{4}\sum_r [a_{rr}(g) + ia_{rr}(g)i^* + ja_{rr}(g)j^* + ka_{rr}(g)k^*] \in F_\alpha.$$

Thus we have the following theorem.

Theorem 6: Two irreducible Q-representations are equivalent if and only if they have the same Q-character. Moreover, Q-characters of inequivalent irreducible Q-representations are orthogonal.

VI. CLASSIFICATION OF IRREDUCIBLE Q-REPRESENTATIONS

We now proceed to study the inter-relations between the irreducible Q-representations and the irreducible C-representations of G. Let B be an irreducible C-representation of G and $\bar{B}$ its contragredient.[14] Recall that (if χ denotes the complex character) $\chi(\bar{B}_g) = \overline{\chi(B_g)}$. B satisfies exactly one of the following three conditions[15,16]:

(a) B is not equivalent to $\bar{B}$.

(b) There exists a matrix M such that $M = M^T$ (the transpose of M) and $MB_gM^{-1} = \bar{B}_g$ for all $g \in G$.

(c) There exists a matrix M such that $M = -M^T$ and $MB_gM^{-1} = \bar{B}_g$ for all $g \in G$. We say (cf. Ref. 16) that B is nonreal, potentially real or pseudoreal according as it satisfies (a), (b), or (c).

Note that every C-matrix representation B may be considered to be a Q-matrix representation since we have identified the complex field with a fixed subfield of the quaternions. However, even if B is irreducible as a C-representation, it need not be irreducible as a Q-representation. The following theorem[9] gives a necessary and sufficient condition.

Theorem 7: An irreducible C-representation B is an irreducible Q-representation if and only if B is not pseudoreal. If B is pseudoreal, then B decomposes over Q into the direct sum of two equivalent irreducible Q-representations.

Consider now an irreducible Q-representation A of G. We say that A is (i) of class **R** if it is equivalent to a real representation, (ii) of class **C** if it is equivalent to a C-representation but not equivalent

[14] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover Publications, Inc., New York, 1931), p. 123.

[15] G. Frobenius and I. Schur, Sitzber. Akad. Wiss. Berlin Kl. Phys. Math. 186 (1906).

[16] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959), p. 285 *et seq.*

to any real representation, and (iii) of class **Q** if it is neither of class **R** nor of class **C**. The following three theorems establish correspondences between the various classes of irreducible Q-representations and C-representations.

Theorem 8: A Q-representation is of class **R** if and only if it is equivalent to a potentially real representation. Two potentially real representations are Q-inequivalent if and only if they are C-inequivalent.

Proof: Since a C-representation is potentially real if and only if it is equivalent to a real representation,[16] the first part follows. For the second part, we have only to note that the C-character of a potentially real representation is real and hence equal to its Q-character.

Theorem 9: A Q-representation is of class **C** if and only if it is equivalent to a nonreal representation. Two nonreal representations B and C are Q-inequivalent if and only if B is C-inequivalent to both C and $\bar{C}$.

Proof: If A be a Q-representation of class **C**, Q-equivalent to a C-representation B, then it is clear that B cannot be potentially real. Also, since B is Q-irreducible, B cannot be pseudoreal by Theorem 7. Hence B must be nonreal. To prove the converse, we have only to show that a nonreal representation B cannot be Q-equivalent to a potentially real representation D. But this is evident, since $X(B_g) = \frac{1}{2}[\chi(B_g) + \chi(\bar{B}_g)]$ is orthogonal to $X(D_g) = \chi(D_g)$, using the classical orthogonality relations.

If B and C are Q-inequivalent, then $X(B_g)$ is not equal to $X(C_g)$ and hence $\chi(B_g)$ is not equal to either $\chi(C_g)$ or $\chi(\bar{C}_g)$, i.e., B is C-inequivalent to both C and $\bar{C}$. Conversely, if B is C-inequivalent to both C and $\bar{C}$, then $X(B_g) = \frac{1}{2}[\chi(B_g) + \chi(\bar{B}_g)]$ is orthogonal to $X(C_g) = \frac{1}{2}[\chi(C_g) + \chi(\bar{C}_g)]$ and hence B and C are Q-inequivalent.

We now turn our attention to pseudoreal representations. If B is one such, then by Theorem 7, $B = B^1 \oplus B^2$ where B^1 and B^2 are equivalent irreducible Q-representations. Since $\chi(B)$ is real, $X(B^1_g) = \frac{1}{2}\chi(B_g)$ and hence the equivalence class of B^1 is uniquely determined by B. We call any member of this equivalence class a Q-representation induced by B.

Theorem 10: A Q-representation A is of class **Q** if and only if it is induced by a pseudoreal representation. Two pseudoreal representations are C-inequivalent if and only if their induced Q-representations are Q-inequivalent.

Proof: Let the Q-representation of class **Q** of dimension n act on the Q-space V. We may assume that A is unitary. Then $g \to A^C_g$ is a unitary C-representation of G in V^C.

We first prove that $g \to A^C_g$ is irreducible. If it is not, let $(e_1, \cdots, e_r)$ be a basis in V^C of some invariant subspace S for A^C. Since A^C is unitary, by replacing S by $S^\perp$ if necessary, we may assume that $r \leq n$. The Q-subspace spanned by $(e_1, \cdots, e_r)$ in V is then invariant under A. Since A is irreducible, we can conclude that $r = n$. But then the matrix of A_g with respect to $(e_1, \cdots, e_n)$ is the matrix of A^C_g restricted to S with respect to $(e_1, \cdots, e_n)$ which is complex—a contradiction since A is of class **Q**. Hence A^C is irreducible.

We show next that A^C is pseudoreal. If A_g has the matrix $A^1_g + A^2_g j$ (where A^1_g and A^2_g are complex) with respect to some basis in V, then with respect to the corresponding basis in V^C, A^C_g has the matrix

$$\begin{vmatrix} A^1_g & A^2_g \\ -\bar{A}^2_g & \bar{A}^1_g \end{vmatrix}.$$

Since A^C_g is unitary, $\bar{A}^C_g$ has the matrix

$$\begin{vmatrix} \bar{A}^1_g & \bar{A}^2_g \\ -A^2_g & A^1_g \end{vmatrix}.$$

The matrix

$$M = \begin{vmatrix} 0 & -I \\ I & 0 \end{vmatrix}$$

has the properties $M = -M^T$ and $MA^C_g M^{-1} = \bar{A}^C_g$, i.e., A^C is pseudoreal.

Since the equality $X(A_g) = \frac{1}{2}\chi(A^C_g)$ is evident by looking at the matrices of A_g and A^C_g, we conclude that A is induced by A^C.

Conversely, if B is a pseudoreal representation inducing the Q-representation A, then A has to be of class **Q**. For, if not, we may assume, by what has been proved so far, that A is either a potentially real or a nonreal representation. In either case $\chi(A)$ is orthogonal to $\chi(B) = 2X(A) = 2\,\mathrm{Re}\,[\chi(A)]$—a contradiction.

The second part is proved by a comparison of characters.

To sum up, the situation is as follows: There is a one-to-one correspondence between the equivalence classes of potentially real (respectively pseudoreal) representations and the equivalence classes of Q-representations of class **R** (class **Q**). There is a one-to-one correspondence between pairs of equivalence classes of nonreal representations, each pair consisting

of the equivalence classes of a representation and its contragredient, and the equivalence classes of Q-representations of class **C**.

This leads us to the following rule for the computation of irreducible Q-characters. Recall that an irreducible C-representation with character χ is nonreal, potentially real or pseudoreal according as

$$\int \chi(g^2)\, dg = 0 \tag{1}$$

$$= +1 \tag{2}$$

or

$$= -1. \tag{3}$$

Rule: Every real irreducible C-character $\chi(g)$ determines an irreducible Q-character $X(g) = \chi(g)$ or $\frac{1}{2}\chi(g)$ according as χ satisfies (2) or (3). Every nonreal irreducible C-character $\chi(g)$ determines an irreducible Q-character $X(g) = \mathrm{Re}\,[\chi(g)]$. All the irreducible Q-characters are obtained in this way.

In the complex case, a C-character χ is irreducible if and only if its (L^2-) norm is unity. For the quaternionic case, we may show that the square of the norm of an irreducible Q-character is 1, $\frac{1}{2}$, or $\frac{1}{4}$ according as the corresponding representation is of class **R**, **C**, or **Q**. This does not in general give us a criterion for deciding the irreducibility of an arbitrary finite-dimensional Q-representation, but if the square of the norm of its Q-character is $\frac{1}{4}$, we can conclude that the representation is irreducible and is of class **Q**.

Every Q-character $X(g)$ is an invariant function, i.e., $X(g) = X(hgh^{-1})$ for all $h \in G$. In contrast to the complex case, it is not in general true that the irreducible Q-characters form a basis for the subspace of invariant functions I in $L^2_Q(G)$. In Example 2 of Sec. VII, for instance, there are only five irreducible Q-characters, whereas $L^2_Q(G)$ is of dimension 8. However, since, as is easily checked, the irreducible C-characters form a basis for I, we may conclude from our analysis that the irreducible Q-characters form a basis for I if and only if every irreducible C-character is real. This happens, for instance, when $G = SO(3)$.

In passing we note that $SO(3)$ does not admit of any irreducible Q-representation of class **Q**, since it does not admit of any irreducible C-representation of even degree.

We conclude this section with the following result.

Theorem 11: If A is an irreducible Q-representation of type α and degree n, then the subspace F_α has dimension n^2, $2n^2$, or $4n^2$ according as A is of class **R**, **C**, or **Q**.

Proof: If A is of class **R**, we may assume that A is real and orthogonal. Since the (real-valued) matrix entries of A are then orthogonal and the reals commute with all the quaternions, F_α is of dimension n^2.

If A is of class **C**, then again we may take A to be complex and unitary. If A_g has the matrix $[a_{rs}(g)]$, its contragredient has the matrix $[\bar{a}_{rs}(g)]$. By definition, every element of F_α is a linear combination of elements of the form $a_{rs}(g)(\beta + \gamma j) = \beta a_{rs}(g) + \gamma j \bar{a}_{rs}(g)$, where β and γ are complex. Again using the classical orthogonal relations, we may conclude that F_α is of dimension $2n^2$.

Now, let A be in class **Q**. Consider A^C. By Theorems 7 and 10, there exists a matrix M such that

$$MA^C_g M^{-1} = \begin{vmatrix} B_g & 0 \\ 0 & C_g \end{vmatrix},$$

where B and C are equivalent to A. Therefore, F_α is spanned by the right Q-multiples of the matrix entries of MA^CM^{-1} and hence of A^C. But the set of matrix entries of A^C is closed (except possibly for sign) with respect to complex conjugation and by the same method used earlier in the proof, we can conclude that F_α is spanned by the matrix entries of A^C. But, by Theorem 10 again, A^C is an irreducible C-representation. Invoking the classical orthogonal relations once more, we conclude that F_α is of dimension $4n^2$.

VII. ABELIAN GROUPS

Let now G denote a compact metric Abelian group. Since every irreducible C-representation of G is one dimensional, it follows from Theorem 10 that G does not admit of any irreducible Q-representations of class **Q**, i.e., every irreducible Q-representation of G is equivalent to a C-representation. It follows immediately that every irreducible Q-representation of G is one dimensional. However, in contrast to the complex case, it is not true that if every irreducible Q-representation of a compact metric group G is one dimensional, then G is Abelian, as the following example shows. We denote by G^0 the group opposite to G (i.e., the elements of G^0 are those of G and the group operation in G^0 is given by $g \cdot h = hg$).

Example 2: Let G be the quaternion group, i.e., $G = [\pm 1, \pm i, \pm j, \pm k]$. Consider G^0. We show that every irreducible Q-representation of G^0 is one dimensional.

If $q \in Q$, let R_q denote the linear transformation of the Q-space Q, given by $R_q(p) = pq$ for all p in Q.

Consider the representations:

$$(1)\ g \to R_g;$$
$$(2)\ g \to A_g = R_1 \quad \text{for all} \quad g \in G^0;$$
$$(3)\ g \to A_g = R_1 \quad \text{if} \quad g = \pm 1, \pm i,$$
$$= R_{-1} \quad \text{otherwise};$$
$$(4)\ g \to A_g = R_1 \quad \text{if} \quad g = \pm 1, \pm j,$$
$$= R_{-1} \quad \text{otherwise};$$
$$(5)\ g \to A_g = R_1 \quad \text{if} \quad g = \pm 1, \pm k,$$
$$= R_{-1} \quad \text{otherwise}.$$

It is easy to verify that the above five (one-dimensional and hence irreducible) Q-representations are mutually inequivalent. If F_r is the subspace in $\mathsf{L}^2_Q(G^0) = Q^{(8)}$ associated with the rth-representation above, then F_1 has dimension four and each of the remaining F_r has dimension one. It follows that G^0 cannot have any irreducible Q-representation inequivalent to all the five above and in particular that G^0 does not have any Q-representation of degree greater than one.

ACKNOWLEDGMENTS

We are indebted to Dr. V. S. Varadarajan for encouraging us to consider these problems and for valuable suggestions. Our thanks are also due the referee for many critical comments.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Multiple Scattering of Electromagnetic Waves by Arbitrary Configurations

VICTOR TWERSKY*
Sylvania Electronic Defense Laboratories, Mountain View, California

(Received 29 April 1966)

This paper extends to three-dimensional vector electromagnetic scattering problems our previous development of the scalar problems. We introduce a vector–dyadic formalism that facilitates exploiting the previous results, and derive analogous integral equations which specify the multiple-scattering amplitudes for many objects in terms of the corresponding functions for isolated scatterers. One representation is in terms of the dyadic analog of Beltrami's operator. For arbitrary configurations, the multi-scattered amplitudes are developed as series in inverse powers of the separations of scatterers (with coefficients in terms of isolated scatterer amplitudes and their derivatives); for two scatterers, we derive a corresponding closed form in terms of a differential operator. Another representation is a system of algebraic equations for the many-body multipole coefficients in terms of the isolated scatterer values. Explicit closed forms are derived for two arbitrarily spaced elementary scatterers (electric dipoles, magnetic dipoles, etc.) both by separations of variables, and by working with elementary dyadic fields.

1. INTRODUCTION

IN previous papers[1–3] we considered the two- and three-dimensional scalar problems of multiple scattering of waves by arbitrary configurations of arbitrary scatterers. In the present paper, the results are extended to the three-dimensional electromagnetic case. We parallel our previous analysis of the three-dimensional scalar case,[3] and exploit as much of that development as feasible; similarly, because recent surveys of the literature of scattering by more than one object are available[4,5] we restrict citations to explicitly related work. Conventional integral and series scalar–vector representations which are adequate (although not particularly convenient) for isolated-scatterer problems, are too cumbersome for multiple scattering problems. We therefore work with vector–dyadic representations essentially as in Morse and Feshbach,[6] and in Saxon[7,8]; we supplement these with

* Present address: Department of Mathematics, University of Illinois at Chicago Circle, Chicago, Illinois.

[1] V. Twersky, in *Electromagnetic Waves*, R. E. Langer, Ed. (University of Wisconsin Press, Madison, Wisc., 1962), pp. 361–389.

[2] J. E. Burke, D. Censor, and V. Twersky, J. Acoust. Soc. Am. 37, 5 (1965).

[3] V. Twersky, J. Math. Phys. 3, 83 (1962).

[4] J. E. Burke and V. Twersky, Radio Sci. 68D, 500 (1964).

[5] V. Twersky, J. Res. Natl. Bur. Std. 64D, 715 (1960).

[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), particularly p. 1897 and Chap. 13.

[7] D. S. Saxon, "Scattering of Light," Scientific Report No. 9, Department of Meteorology, UCLA (1955).

[8] D. S. Saxon, Phys. Rev. 100, 1771 (1955).

Consider the representations:

$$(1)\ g \to R_g;$$
$$(2)\ g \to A_g = R_1 \quad \text{for all} \quad g \in G^0;$$
$$(3)\ g \to A_g = R_1 \quad \text{if} \quad g = \pm 1, \pm i,$$
$$= R_{-1} \quad \text{otherwise};$$
$$(4)\ g \to A_g = R_1 \quad \text{if} \quad g = \pm 1, \pm j,$$
$$= R_{-1} \quad \text{otherwise};$$
$$(5)\ g \to A_g = R_1 \quad \text{if} \quad g = \pm 1, \pm k,$$
$$= R_{-1} \quad \text{otherwise}.$$

It is easy to verify that the above five (one-dimensional and hence irreducible) Q-representations are mutually inequivalent. If F_r is the subspace in $\mathsf{L}^2_Q(G^0) = Q^{(8)}$ associated with the rth-representation above, then F_1 has dimension four and each of the remaining F_r has dimension one. It follows that G^0 cannot have any irreducible Q-representation inequivalent to all the five above and in particular that G^0 does not have any Q-representation of degree greater than one.

ACKNOWLEDGMENTS

We are indebted to Dr. V. S. Varadarajan for encouraging us to consider these problems and for valuable suggestions. Our thanks are also due the referee for many critical comments.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Multiple Scattering of Electromagnetic Waves by Arbitrary Configurations

VICTOR TWERSKY*
Sylvania Electronic Defense Laboratories, Mountain View, California

(Received 29 April 1966)

This paper extends to three-dimensional vector electromagnetic scattering problems our previous development of the scalar problems. We introduce a vector–dyadic formalism that facilitates exploiting the previous results, and derive analogous integral equations which specify the multiple-scattering amplitudes for many objects in terms of the corresponding functions for isolated scatterers. One representation is in terms of the dyadic analog of Beltrami's operator. For arbitrary configurations, the multi-scattered amplitudes are developed as series in inverse powers of the separations of scatterers (with coefficients in terms of isolated scatterer amplitudes and their derivatives); for two scatterers, we derive a corresponding closed form in terms of a differential operator. Another representation is a system of algebraic equations for the many-body multipole coefficients in terms of the isolated scatterer values. Explicit closed forms are derived for two arbitrarily spaced elementary scatterers (electric dipoles, magnetic dipoles, etc.) both by separations of variables, and by working with elementary dyadic fields.

1. INTRODUCTION

IN previous papers[1–3] we considered the two- and three-dimensional scalar problems of multiple scattering of waves by arbitrary configurations of arbitrary scatterers. In the present paper, the results are extended to the three-dimensional electromagnetic case. We parallel our previous analysis of the three-dimensional scalar case,[3] and exploit as much of that development as feasible; similarly, because recent surveys of the literature of scattering by more than one object are available[4,5] we restrict citations to explicitly related work. Conventional integral and series scalar–vector representations which are adequate (although not particularly convenient) for isolated-scatterer problems, are too cumbersome for multiple scattering problems. We therefore work with vector–dyadic representations essentially as in Morse and Feshbach,[6] and in Saxon[7,8]; we supplement these with

* Present address: Department of Mathematics, University of Illinois at Chicago Circle, Chicago, Illinois.

[1] V. Twersky, in *Electromagnetic Waves*, R. E. Langer, Ed. (University of Wisconsin Press, Madison, Wisc., 1962), pp. 361–389.

[2] J. E. Burke, D. Censor, and V. Twersky, J. Acoust. Soc. Am. 37, 5 (1965).

[3] V. Twersky, J. Math. Phys. 3, 83 (1962).

[4] J. E. Burke and V. Twersky, Radio Sci. 68D, 500 (1964).

[5] V. Twersky, J. Res. Natl. Bur. Std. 64D, 715 (1960).

[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), particularly p. 1897 and Chap. 13.

[7] D. S. Saxon, "Scattering of Light," Scientific Report No. 9, Department of Meteorology, UCLA (1955).

[8] D. S. Saxon, Phys. Rev. 100, 1771 (1955).

results of Hansen,[9] Stratton,[10] Silver,[11] and Wilcox,[12] as well as with additional representations and theorems derived in the course of the present development (e.g., by separating variables in the vector wave equation). We use dyadic surface integral forms, complex integral dyadic plane-wave representations, inverse distance series involving the vector scattering amplitudes acted on by the dyadic analog of Beltrami's operator, series for dyadic fields in terms of dyads of vector harmonics, etc. To facilitate discussion we start with a relatively conventional vector formalism, and then switch to dyadic representations.

In the following we always indicate dyadics by using a tilde—$\tilde{g}$, $\tilde{u}$, $\tilde{\varphi}$, etc., and write vectors as $\mathbf{g}$, $\mathbf{u}$, $\boldsymbol{\varphi}$, etc.; a caret always indicates a unit vector—$\hat{g}$, $\hat{x}$, $\hat{\theta}$, etc., but we also define some special symbols ($\mathbf{o}$, $\mathbf{i}$, $\boldsymbol{\epsilon}$, $\mathbf{n}$, etc.) to represent unit vectors. For brevity, we regard the numbered equations and figures of Ref. 3 as part of the present text, and cite them as Eq. (3:8), Fig. 3:1, etc.

2. ONE SCATTERER

2.1. Vector Fields

The three-dimensional scattering of a plane electromagnetic wave (with $e^{-i\omega t}$ suppressed) is specified in the external region by a solution of

$$\nabla \times \nabla \times \psi - k^2\psi = 0, \quad \nabla \cdot \psi = 0, \qquad k = |k| = 2\pi/\lambda, \tag{1}$$

subject to prescribed conditions on the scatterer's surface, and subject to the condition that ψ consist of a plane wave $\boldsymbol{\varphi}$ plus a radiated wave $\mathbf{u}$. With increasing distance from the scatterer ($r \to \infty$) the function ψ (which represents either the $\mathbf{E}$ or $\mathbf{H}$ field) reduces to a plane wave

$$\boldsymbol{\varphi}(\mathbf{i}:\boldsymbol{\epsilon}) = \boldsymbol{\epsilon} e^{i\mathbf{k}\cdot\mathbf{r}}, \quad \mathbf{k} = k\mathbf{i}, \quad \mathbf{r} = r\mathbf{o}, \tag{2}$$

where $\boldsymbol{\epsilon}$, $\mathbf{i}$, and $\mathbf{o}$ are unit vectors. Because of the divergence condition $\nabla \cdot \boldsymbol{\varphi} = 0$, the "polarization vector" $\boldsymbol{\epsilon}$ is perpendicular to the direction of incidence, $\boldsymbol{\epsilon} \cdot \mathbf{i} = 0$; to make this explicit, we write

$$\boldsymbol{\varphi}(\mathbf{i}:\boldsymbol{\epsilon}) = \boldsymbol{\epsilon} \cdot (\tilde{I} - \mathbf{ii})e^{i\mathbf{k}\cdot\mathbf{r}} = \boldsymbol{\epsilon} \cdot \tilde{\varphi}(\mathbf{i}), \qquad \tilde{\varphi}(\mathbf{i}) \equiv (\tilde{I} - \mathbf{ii})e^{ik\mathbf{i}\cdot\mathbf{r}}, \tag{3}$$

where $\tilde{I}$ is the unit dyadic, and $\tilde{\varphi}$ is a dyadic plane wave. The difference $\psi - \boldsymbol{\varphi} = \mathbf{u}$, the scattered wave, may be specified by the Sommerfeld–Silver radiation condition[11,12]

$$\lim r[\mathbf{o} \times (\nabla \times \mathbf{u}) + ik\mathbf{u}] = 0, \quad \text{as } r \to \infty. \tag{4}$$

For concreteness, we may take the origin of coordinates of $\mathbf{r}$ as the center of the smallest sphere which completely encloses the scatterer; we use the same geometry as in Fig. 3:1.

From (4) and Green's theorem it follows[12] that for $r \sim \infty$,

$$\mathbf{u} \sim \mathbf{g}(\mathbf{o}, \mathbf{i}: \boldsymbol{\epsilon})h(kr), \quad h(r) \equiv h_0^{(1)}(r) = e^{ir}/ir, \tag{5}$$

where the normalized "scattering amplitude" $\mathbf{g}(\mathbf{o}, \mathbf{i}: \boldsymbol{\epsilon})$ specifies the "far-field" response in the direction of observation $\mathbf{o}$ to plane-wave excitation of direction of incidence $\mathbf{i}$ and polarization $\boldsymbol{\epsilon}$. Since $\nabla \cdot \mathbf{u} = 0$, we have $\mathbf{o} \cdot \hat{g} = 0$, and we may write

$$\mathbf{g}(\mathbf{o}, \mathbf{i}: \boldsymbol{\epsilon}) = (\tilde{I} - \mathbf{oo}) \cdot \mathbf{g}(\mathbf{o}, \mathbf{i}: \boldsymbol{\epsilon}).$$

In general, we take $\psi = \mathbf{E}$, and $\nabla \times \psi = \mathbf{H}i\omega\mu_0 = \mathbf{H}i\omega$. At the surface of a perfect conductor,

$$\mathbf{n} \times \psi = \mathbf{n} \times (\boldsymbol{\varphi} + \mathbf{u}) = 0, \tag{6}$$

where $\mathbf{n}$ is the surface normal. For a scatterer specified by relative electrical constants ϵ and μ we introduce the internal field ψ' such that

$$\nabla \times \nabla \times \psi' - k'^2\psi' = 0, \quad \nabla \cdot \psi' = 0, \qquad k' = k(\epsilon\mu)^{\frac{1}{2}}, \tag{7}$$

and use the surface conditions

$$\mathbf{n} \times \psi = \mathbf{n} \times \psi', \quad \mathbf{n} \times (\nabla \times \psi) = \mathbf{n} \times (\nabla \times \psi'/\mu). \tag{8}$$

Surface integral representation: Introducing the free-space dyadic Green's function[13,7,6]

$$\tilde{\Gamma}(\mathbf{r}, \mathbf{r}') = \left(\tilde{I} + \frac{\nabla\nabla}{k^2}\right)\frac{kh(k\,|\mathbf{r} - \mathbf{r}'|)}{4\pi i}, \qquad \nabla \times \nabla \times \tilde{\Gamma} - k^2\tilde{\Gamma} = -\tilde{I}\delta(\mathbf{r} - \mathbf{r}'), \tag{9}$$

we apply Gauss' theorem for dyadics to construct

$$\begin{aligned}\int [(\nabla \times \nabla \times \mathbf{u}) \cdot \tilde{\Gamma} - \mathbf{u} \cdot (\nabla \times \nabla \times \tilde{\Gamma})]\, dV \\ = \int \mathbf{n} \cdot [(\nabla \times \mathbf{u}) \times \tilde{\Gamma} + \mathbf{u} \times (\nabla \times \tilde{\Gamma})]\, dS \\ = -\int [(\nabla \times \mathbf{u}) \cdot (\mathbf{n} \times \tilde{\Gamma}) - (\mathbf{n} \times \mathbf{u}) \cdot (\nabla \times \tilde{\Gamma})]\, dS.\end{aligned} \tag{10}$$

In the region external to the scatterer, we use (1), (9),

[9] W. W. Hansen, Phys. Rev. **47**, 139 (1935); see also Physics **7**, 460 (1936); J. Appl. Phys. **8**, 282 (1937); W. W. Hansen and J. G. Beckerly, Physics **7**, 220 (1936); Proc. IRE **24**, 1594 (1936).

[10] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941).

[11] S. Silver, *Microwave Antenna Theory and Design* (McGraw-Hill Book Company, Inc., New York, 1949).

[12] C. H. Wilcox, Commun. Pure Appl. Math. **9**, 115 (1956).

[13] H. Levine and J. Schwinger, Commun. Pure Appl. Math. **3**, 355 (1950).

and (10) to reduce the dyadic form of Green's theorem (10) to

$$\mathbf{u}(\mathbf{r}) = \int [(\nabla \times \mathbf{u}) \cdot (\mathbf{n} \times \tilde{\Gamma}) - (\mathbf{n} \times \mathbf{u}) \cdot (\nabla \times \tilde{\Gamma})]\, dS, \tag{11}$$

where now $\mathbf{n}$ points away from the scatterer, and where the integral is over any surface enclosing the scatterer and excluding $\mathbf{r}$. We rewrite (11) as

$$\mathbf{u}(\mathbf{r}; \mathbf{i}: \boldsymbol{\epsilon}) = \frac{k}{4\pi i} \int [(\nabla \times \mathbf{u}) \cdot (\mathbf{n} \times \tilde{h}) - (\mathbf{n} \times \mathbf{u}) \cdot (\nabla \times \tilde{h})]\, dS$$

$$\equiv \{\tilde{h}(k\,|\mathbf{r} - \mathbf{r}'|), \mathbf{u}(\mathbf{r}'; \mathbf{i}: \boldsymbol{\epsilon})\};$$

$$\tilde{h}(k\,|\mathbf{r} - \mathbf{r}'|) = (\tilde{I} + \nabla\nabla/k^2) h(k\,|\mathbf{r} - \mathbf{r}'|) = \tilde{\Gamma} 4\pi i/k,$$

$$h(r) = e^{ir}/ir. \tag{12}$$

If we replace $\tilde{\Gamma}$ by $\tilde{\Gamma} \cdot \mathbf{e}$ in the above, where $\mathbf{e}$ is an arbitrary constant vector, then (10) reduces to the usual vector form of Green's theorem [say (10) $\cdot$ e] and the left-hand sides of (11) and (12) reduce to $\mathbf{u} \cdot \mathbf{e}$.

Since $\{\tilde{h}(k\,|\mathbf{r} - \mathbf{r}'|), \boldsymbol{\varphi}(\mathbf{r}')\} = 0$ for $\mathbf{r}$ outside S, we may also write

$$\mathbf{u}(\mathbf{r}) = \{\tilde{h}(k\,|\mathbf{r} - \mathbf{r}'|), \boldsymbol{\psi}(\mathbf{r}')\}. \tag{13}$$

From (13), (9), and (8), we obtain

$$\boldsymbol{\psi} = \boldsymbol{\varphi} - \int \left[\left(\frac{K^2}{\mu} - k^2 \right) \boldsymbol{\psi}' \cdot \tilde{\Gamma} + \left(1 - \frac{1}{\mu} \right) (\nabla \times \boldsymbol{\psi}') \cdot (\nabla \times \tilde{\Gamma}) \right] dV, \tag{14}$$

which also holds for an interior point, in which case $\boldsymbol{\psi} = \boldsymbol{\psi}'$ is supplied by the internal (instead of the external) singularity of $\tilde{\Gamma}$. The case $\mu = 1$ is discussed in detail by Saxon,[7] and a generalization of (14) is considered in Ref. 14.

If $k\,|\mathbf{r} - \mathbf{r}'| \gg 1$ and $r \gg r'$, then

$$\tilde{h}(k\,|\mathbf{r} - \mathbf{r}'|) \sim (\tilde{I} - \mathbf{oo}) e^{-ik\mathbf{o}\cdot\mathbf{r}'} h(kr) = \tilde{\varphi}(-\mathbf{o}) h(kr), \tag{15}$$

and (12) reduces to the far-field form (5) with

$$\mathbf{g}(\mathbf{o}, \mathbf{i}: \boldsymbol{\epsilon}) = \{(\tilde{I} - \mathbf{oo}) e^{-ik\mathbf{o}\cdot\mathbf{r}'}, \mathbf{u}(\mathbf{r}'; \mathbf{i}: \boldsymbol{\epsilon})\} = \{\tilde{\varphi}(-\mathbf{o}), \mathbf{u}\}. \tag{16}$$

For any unit vector $\boldsymbol{\gamma}$ perpendicular to $\mathbf{o}$ we have

$$\boldsymbol{\gamma} \cdot \mathbf{g}(\mathbf{o}, \mathbf{i}: \boldsymbol{\epsilon}) = \{\boldsymbol{\varphi}(-\mathbf{o}: \boldsymbol{\gamma}), \mathbf{u}(\mathbf{i}: \boldsymbol{\epsilon})\}, \tag{17}$$

where $\mathbf{r}'$ has been suppressed. If $\boldsymbol{\gamma} = \hat{g} = \mathbf{g}/g$, then the left side of (17) reduces to $g(\mathbf{o}, \mathbf{i}: \boldsymbol{\epsilon})$.

Scattering theorems: To facilitate subsequent applications we use the present formalism to derive certain theorems which $\mathbf{g}$ fulfills. See Saxon[8] for derivation based on a tensor scattering matrix.

[14] V. Twersky, J. Math. Phys. 3, 716 (1962).

Consider two solutions of a scattering problem for two different incident waves, say $\boldsymbol{\psi}_1 = \boldsymbol{\varphi}_1 + \mathbf{u}_1$ and $\boldsymbol{\psi}_2 = \boldsymbol{\varphi}_2 + \mathbf{u}_2$, such that $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}(\mathbf{i}_1: \boldsymbol{\epsilon}_1)$, etc. Since $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ satisfy the same conditions at the scatterer [i.e., (6), or (7) plus (8)], we have $\{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2\}_S = 0$ on its surface S, and since $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ fulfill (1) in the external region, it follows from (10) $\cdot$ e that

$$\{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2\} = \{(\boldsymbol{\varphi}_1 + \mathbf{u}_1), (\boldsymbol{\varphi}_2 + \mathbf{u}_2)\} = 0 \tag{18}$$

for any surface (including the surface at infinity S_∞) surrounding the scatterer. Since

$$\{\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2\} = \{\mathbf{u}_1, \mathbf{u}_2\}_{S\infty} = 0,$$

(18) reduces to

$$\{\boldsymbol{\varphi}_1, \mathbf{u}_2\} = -\{\mathbf{u}_1, \boldsymbol{\varphi}_2\} = \{\boldsymbol{\varphi}_2, \mathbf{u}_1\}, \tag{19}$$

where the last equality follows from the explicit form of the operator in (12). Thus since $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}(\mathbf{i}_1: \boldsymbol{\epsilon}_1) = \boldsymbol{\epsilon}_1 \cdot \tilde{\varphi}(\mathbf{i}_1)$, we use (17) in (19) to obtain the reciprocity relation

$$\boldsymbol{\epsilon}_1 \cdot \mathbf{g}(-\mathbf{i}_1, \mathbf{i}_2: \boldsymbol{\epsilon}_2) = \boldsymbol{\epsilon}_2 \cdot \mathbf{g}(-\mathbf{i}_2, \mathbf{i}_1: \boldsymbol{\epsilon}_1). \tag{20}$$

This holds for the relatively weak surface condition $\{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2\}_S = 0$, which includes (6), etc.

If $\boldsymbol{\psi}_1$ is replaced by its complex conjugate $\boldsymbol{\psi}_1^*$, then for lossless scatterers

$$\{\boldsymbol{\psi}_1^*, \boldsymbol{\psi}_2\} = \{(\boldsymbol{\varphi}_1^* + \mathbf{u}_1^*), (\boldsymbol{\varphi}_2 + \mathbf{u}_2)\} = 0. \tag{21}$$

We have $\{\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2\} = 0$, and

$$\{\mathbf{u}_1^*, \mathbf{u}_2\} = \frac{k}{4\pi i} 2ik \int (\mathbf{o} \times \mathbf{g}_2 h) \cdot (\mathbf{o} \times \mathbf{g}_1 h)^*\, dS_\infty = \frac{1}{2\pi} \int \mathbf{g}(\mathbf{o}; \mathbf{i}_2: \boldsymbol{\epsilon}_2) \cdot \mathbf{g}^*(\mathbf{o}, \mathbf{i}_1: \boldsymbol{\epsilon}_1)\, d\Omega_o, \tag{22}$$

where $d\Omega_o$ is the differential solid angle around $\mathbf{o}$, and the integration is over all angles of observation. Since $\boldsymbol{\varphi}^* = \boldsymbol{\epsilon}^* e^{-ik\mathbf{i}^*\cdot\mathbf{r}} = \boldsymbol{\varphi}(-\mathbf{i}^*: \boldsymbol{\epsilon}^*)$, we reduce (21) to

$$\boldsymbol{\epsilon}_1^* \cdot \mathbf{g}(\mathbf{i}_1^*, \mathbf{i}_2: \boldsymbol{\epsilon}_2) + \boldsymbol{\epsilon}_2^* \cdot \mathbf{g}^*(\mathbf{i}_2^*, \mathbf{i}_1: \boldsymbol{\epsilon}_1) = \frac{-1}{2\pi} \int \mathbf{g}(\mathbf{o}, \mathbf{i}_2: \boldsymbol{\epsilon}_2) \cdot \mathbf{g}^*(\mathbf{o}, \mathbf{i}_1: \boldsymbol{\epsilon}_1)\, d\Omega_o. \tag{23}$$

In particular, in the forward scattered direction, such that all $\mathbf{i}$'s reduce to $\mathbf{i}$, and all $\boldsymbol{\epsilon}$'s to $\boldsymbol{\epsilon}$, we obtain the "energy theorem"

$$-\mathrm{Re}\, \boldsymbol{\epsilon} \cdot \mathbf{g}(\mathbf{i}, \mathbf{i}: \boldsymbol{\epsilon}) = \frac{1}{4\pi} \int |\mathbf{g}(\mathbf{o}, \mathbf{i}: \boldsymbol{\epsilon})|^2\, d\Omega_o = \frac{k^2}{4\pi} Q(\mathbf{i}: \boldsymbol{\epsilon}), \tag{24}$$

where Q is the total scattering cross section. If the scatterer is not lossless, then $4\pi/k^2$ times the left-most form of (24) equals the sum of scattering plus absorption cross sections.

Plane-wave representation: If $\mathbf{u}$ is known, then (16) gives $\mathbf{g}$ by integration. An inverse relation follows from (14) by using[15]

$$h(k\,|\mathbf{r}-\mathbf{r}'|) = \frac{1}{2\pi}\int e^{ik\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')}\,d\Omega_p, \tag{25}$$

where the limits of the complex paths of the angles associated with the unit vector $\mathbf{p}(\tau,\beta)$ (each path analogous to one in Sommerfeld's integral for $H_0^{(1)}$) are chosen to ensure $\operatorname{Im}\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')>0$. See additional discussion in Noether[15] and after (3:8).

Substituting the corresponding dyadic

$$\tilde{h}(k\,|\mathbf{r}-\mathbf{r}'|) = \left(\tilde{I}+\frac{\nabla\nabla}{k^2}\right)h(k\,|\mathbf{r}-\mathbf{r}'|) = \frac{1}{2\pi}\int(\tilde{I}-\mathbf{pp})e^{ik\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')}\,d\Omega_p \tag{26}$$

into $\mathbf{u}$ of (12), and using definition (16), we obtain the vector analog of (3:9):

$$\mathbf{u}(\mathbf{r};\mathbf{i}) = \frac{1}{2\pi}\int e^{ik\mathbf{p}\cdot\mathbf{r}}\{(\tilde{I}-\mathbf{pp})e^{-ik\mathbf{p}\cdot\mathbf{r}'},\,\mathbf{u}(\mathbf{r}';\mathbf{i}:\epsilon)\}\,d\Omega_p = \frac{1}{2\pi}\int e^{ik\mathbf{p}\cdot\mathbf{r}}\mathbf{g}(\mathbf{p},\mathbf{i}:\epsilon)\,d\Omega_p, \tag{27}$$

which holds at least for $r > r'_{\max} \equiv a$. (See Ref. 3 for weaker condition.)

Cartesian representation in inverse powers of r: The asymptotic form given in (5) is the leading term of a series expansion of $\mathbf{u}$ in inverse powers of r which converges for $r > r'_{\max} = a$; see Wilcox[12] for a detailed discussion. This series, with coefficients expressed in different forms, may be obtained from (27) by various procedures, e.g., by means of

$$\frac{1}{2\pi}\int e^{ir\mathbf{p}\cdot\mathbf{o}}F(\mathbf{p})\,d\Omega_p = h(r)\mathfrak{D}(r;D)F(\mathbf{o}),$$

$$\mathfrak{D}(r;D) = 1+\frac{i}{2r}D+\left(\frac{i}{2r}\right)^2\frac{D(D-1\cdot 2)}{2!}+\cdots + \left(\frac{i}{2r}\right)^n\frac{D(D-1\cdot 2)(D-2\cdot 3)\cdots(D-[n-1]n)}{n!},$$

$$D = \frac{-1}{\sin^2\theta}[\partial_\varphi^2+\sin\theta\partial_\theta(\sin\theta\partial_\theta)], \tag{28}$$

where r is a parameter, $F(\mathbf{o})$ is representable as series of surface harmonics, and D is Beltrami's operator; see (3:10) to (3:16) for details. Using (28) for the Cartesian components of (27), we obtain

$$\mathbf{u} = h(kr)\mathfrak{D}(kr;D)\mathbf{g}(\mathbf{o}) = h(kr)\left[\mathbf{g}+\frac{i}{2kr}D\mathbf{g}+\left(\frac{i}{2kr}\right)^2\frac{D(D-2)}{2}\mathbf{g}+\cdots\right], \tag{29}$$

subject to $\nabla\cdot\mathbf{u}=0$ and $\nabla\times\nabla\times\mathbf{u}-k^2\mathbf{u}=0$.

Special function representations: In the following, except for normalization factors, we work with the transverse vector spherical functions introduced by Hansen,[9] and discussed by Stratton,[10] Morse and Feshbach,[6] Saxon,[7] Stein,[16] and others. We use

$$\begin{aligned}\mathbf{M}_{nm}(\mathbf{r}) &= h_n(kr)\mathbf{C}_n^m(\mathbf{o}),\\ \mathbf{C}_n^m(\mathbf{o}) &= -\mathbf{r}\times\nabla Y_n^m(\mathbf{o}) = \left(\hat{\theta}\frac{\partial_\varphi}{\sin\theta}-\hat{\varphi}\partial_\theta\right)Y_n^m(\mathbf{o})\\ &= -\mathbf{L}(\mathbf{o})Y_n^m(\mathbf{o}),\\ Y_n^m(\mathbf{o}) &= P_n^m(\cos\theta)e^{im\varphi},\\ Y_n^{-m}(\mathbf{o}) &= (-1)^m[(n-m)!/(n+m)!]P_n^m(\cos\theta)e^{-im\varphi}.\end{aligned} \tag{30}$$

Here $h_n = h_n^{(1)}$ is a radiating spherical Hankel function, and P_n^m is an associated Legendre function. Similarly

$$\begin{aligned}\mathbf{N}_{nm}(\mathbf{r}) &= [n(n+1)h_n(kr)\mathbf{P}_n^m(\mathbf{o})\\ &\quad+\partial_{kr}[krh_n(kr)]\mathbf{B}_n^m(\mathbf{o})]/kr,\\ \mathbf{P}_n^m(\mathbf{o}) &= \mathbf{o}Y_n^m(\mathbf{o}),\quad \mathbf{B}_n^m(\mathbf{o}) = r\nabla Y_n^m(\mathbf{o}) = \mathbf{o}\times\mathbf{C}_n^m(\mathbf{o}).\end{aligned} \tag{31}$$

The two sets are related through $k\mathbf{N}=\nabla\times\mathbf{M}$ and $k\mathbf{M}=\nabla\times\mathbf{N}$. For real directions, the corresponding even and odd vector harmonics $\mathbf{P}_{nm}^{e,o}$ and $\mathbf{C}_{nm}^{e,o}$ of Morse and Feshbach[6] (pp. 1865, 1898, 1899) are the real and imaginary parts, respectively, of the present $\mathbf{P}_n^m$ and $\mathbf{C}_n^m/[n(n+1)]^{\frac{1}{2}}$. We have

$$\mathbf{C}_n^{-m}(\mathbf{o}) = (-1)^m[(n-m)!/(n+m)!][\mathbf{C}_n^m(\mathbf{o}^*)]^*.$$

We also work with $\mathbf{N}$ in an alternative form, essentially as in Morse and Feshbach[6] (p. 1866):

$$i^{n-1}\mathbf{N}_n(r) = \frac{n(n+1)}{2n+1}[h_{n-1}i^{n-1}\mathbf{E}_{n-1}-h_{n+1}i^{n+1}\mathbf{H}_{n+1}],$$

$$\mathbf{E}_{n-1}\equiv\mathbf{P}_n+\frac{\mathbf{B}_n}{n},\quad \mathbf{H}_{n+1}\equiv\mathbf{P}_n-\frac{\mathbf{B}_n}{n+1}, \tag{32}$$

where we have dropped arguments and the index m for brevity. Henceforth, also for brevity, all four-digit page numbers we cite are to be found in Morse and Feshbach.[6]

[15] F. Noether, in *Theory of Functions*, R. Rothe, F. Ollendorf, and K. Pohlhausen, Eds. (Technology Press, Cambridge, Mass., 1948), p. 167, Eq. (7).

[16] S. Stein, Quart. Appl. Math. **19**, 15 (1961).

The angular functions satisfy the following orthogonality relations:

$$\int \mathbf{C}_n^{-m} \cdot \mathbf{B}_\nu^\mu \, d\Omega = \int \mathbf{C}_n^{-m} \cdot \mathbf{P}_\nu^\mu \, d\Omega = \int \mathbf{P}_n^{-m} \cdot B_\nu^\mu \, d\Omega = 0,$$

$$\int \mathbf{C}_n^{-m} \cdot \mathbf{C}_\nu^\mu \, d\Omega = \int \mathbf{B}_n^{-m} \cdot \mathbf{B}_\nu^\mu \, d\Omega$$

$$= n(n+1) \int \mathbf{P}_n^{-m} \cdot \mathbf{P}_\nu^\mu \, d\Omega$$

$$= (-1)^m 4\pi \delta_{n\nu} \delta_{m\mu} n(n+1)/(2n+1),$$

$$\int \mathbf{E}_{n-1}^{-m} \cdot \mathbf{H}_{n+1}^\mu \, d\Omega = \int \mathbf{E}_{n-1}^{-m} \cdot \mathbf{C}_\nu^\mu \, d\Omega$$

$$= \int \mathbf{H}_{n+1}^{-m} \cdot \mathbf{C}_\nu^\mu \, d\Omega = 0,$$

$$n \int \mathbf{E}_{n-1}^{-m} \cdot \mathbf{E}_{\nu-1}^\mu \, d\Omega = (n+1) \int \mathbf{H}_{n+1}^{-m} \cdot \mathbf{H}_{\nu+1}^\mu \, d\Omega$$

$$= (-1)^m 4\pi \delta_{n\nu} \delta_{m\mu},$$

$$\mathbf{P}_n^m \cdot \mathbf{B}_n^m = \mathbf{P}_n^m \cdot \mathbf{C}_n^m = \mathbf{B}_n^m \cdot \mathbf{C}_n^m = 0,$$

$$\mathbf{C}_n^m \cdot \mathbf{C}_\nu^\mu = \mathbf{B}_n^m \cdot \mathbf{B}_\nu^\mu. \tag{33}$$

The asymptotic forms of Hansen's functions are

$$i^n \mathbf{M}_n(kr) \sim h(kr) \mathbf{C}_n(\mathbf{o}), \quad i^{n-1} \mathbf{N}_n(kr) \sim h(kr) \mathbf{B}_n(\mathbf{o}), \tag{34}$$

where $h = h_0^{(1)}$ as in (5).

From pp. 1782 and 1875 we may write the normalized dyadic Green's function for $r > r'$ as

$$\tilde{h}(k|\mathbf{r} - \mathbf{r}'|) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} [\mathbf{M}_{nm}(\mathbf{r}) \mathbf{M}_{n,-m}^1(\mathbf{r}') + \mathbf{N}_{nm}(\mathbf{r}) \mathbf{N}_{n,-m}^1(\mathbf{r}')](-1)^m \, d_n,$$

$$d_n \equiv \frac{2n+1}{n(n+1)}, \tag{35}$$

where the functions with superscript 1 are the nonsingular nonradiating functions (j type), and those without superscripts are the radiating functions ($h^{(1)}$ type). If we substitute (35) into (12), we obtain

$$\mathbf{u}(\mathbf{r}; \mathbf{i}: \epsilon) = \sum [\mathbf{M}_{nm}(\mathbf{r}) c_{nm}(\mathbf{i}: \epsilon) - i \mathbf{N}_{nm}(\mathbf{r}) b_{nm}(\mathbf{i}: \epsilon)] i^n,$$

$$c_{nm}(\mathbf{i}: \epsilon) = i^{-n}(-1)^m \, d_n \{\mathbf{M}_{n,-m}^1(\mathbf{r}'), \mathbf{u}(\mathbf{r}'; \mathbf{i}: \epsilon)\},$$

$$b_{nm} = i^{-n+1}(-1)^m \, d_n \{\mathbf{N}_{n,-m}^1, \mathbf{u}\}. \tag{36}$$

The scattering coefficients (or multipole coefficients) c and b are of the magnetic-type and electric-type, respectively. If we introduce (34) into (36), and compare with (5) we have

$$\mathbf{g}(\mathbf{o}, \mathbf{i}: \epsilon) = \sum [\mathbf{C}_n^m(\mathbf{o}) c_{nm}(\mathbf{i}: \epsilon) + \mathbf{B}_n^m(\mathbf{o}) b_{nm}(\mathbf{i}: \epsilon)], \tag{37}$$

which may also be derived directly from (16) by substituting [from (35) with $r \sim \infty$, or from p. 1866].

$$\tilde{\varphi}(\mathbf{r}'; -\mathbf{o}) = (\tilde{I} - \mathbf{oo}) e^{-ik\mathbf{o}\cdot\mathbf{r}'}$$

$$= \sum [\mathbf{C}_n^m(\mathbf{o}) \mathbf{M}_{n,-m}^1(\mathbf{r}') + i \mathbf{B}_n^m(\mathbf{o}) \mathbf{N}_{n,-m}^1(\mathbf{r}')] i^{-n}(-1)^m \, d_n. \tag{38}$$

We could also have obtained (36) for $\mathbf{u}$ by substituting (37) into (27). Thus

$$\mathbf{u} = \frac{1}{2\pi} \int e^{ik\mathbf{p}\cdot\mathbf{r}} \sum [\mathbf{C}_n^m(\mathbf{p}) c_{nm} + \mathbf{B}_n^m(\mathbf{p}) b_{nm}] \, d\Omega_p,$$

$$c_{nm} = \frac{(-1)^m \, d_n}{4\pi} \int \mathbf{C}_n^{-m}(\mathbf{o}) \cdot \mathbf{g}(\mathbf{o}) \, d\Omega_o,$$

$$b_{nm} = \frac{(-1)^m \, d_n}{4\pi} \int \mathbf{B}_n^{-m}(\mathbf{o}) \cdot \mathbf{g}(\mathbf{o}) \, d\Omega_o, \tag{39}$$

which reduces to (36) on using

$$i^n \mathbf{M}_{nm}(\mathbf{r}) = \frac{1}{2\pi} \int e^{ik\mathbf{p}\cdot\mathbf{r}} \mathbf{C}_n^m(\mathbf{p}) \, d\Omega_p,$$

$$i^{n-1} \mathbf{N}_{nm}(\mathbf{r}) = \frac{1}{2\pi} \int e^{ik\mathbf{p}\cdot\mathbf{r}} \mathbf{B}_n^m(\mathbf{p}) \, d\Omega_p \tag{40}$$

(the radiating function analogs of the forms on pp. 1865–1866).

General representation in inverse powers of r: The present series leads to an inverse-distance expansion fully analogous to (3:16). For the scalar case,[3] we substituted Hankel's polynomial form

$$h_n(r) i^n = h(r) \Big[1 + n(n+1)(i/2r) + n(n+1)[n(n+1) - 1 \cdot 2] (1/2!)(i/2r)^2 + \cdots\Big]$$

$$\equiv h(r) \mathfrak{D}(r; n[n+1]), \tag{41}$$

into $u = \sum a_{nm}(i) h_n(kr) i^n Y_n^m(\mathbf{o})$, and then used Legendre's equation

$$n(n+1) Y_n(\mathbf{o}) = D Y_n(\mathbf{o}), \tag{42}$$

and the scalar amplitude $g = \sum a_{nm}(\mathbf{i}) Y_n^m(\mathbf{o})$ to obtain the form $u = h(kr) \mathfrak{D}(kr; D) g(\mathbf{o})$ [implicit in the Cartesian representation (29)].

We obtain the analog of (42) for the vector spherical harmonics by separating variables in the vector wave equation (1); we write $\boldsymbol{\psi}(\mathbf{r})$ as a series of functions $R_n(r) \mathbf{F}_n(\mathbf{0})$, and obtain

$$\mathbf{F}_n(\mathbf{o})[r^2 k^2 + (1/R_n) \partial_r (r^2 \partial_r R_n)] = r^2 [\nabla \times (\nabla \times \mathbf{F}_n) - \nabla\nabla \cdot \mathbf{F}_n] \equiv \tilde{D} \cdot \mathbf{F}_n(\mathbf{o}), \tag{43}$$

where $\tilde{D}$ reduces to $D\tilde{I}$ in Cartesian coordinates. In polar coordinates, with $\mathbf{F} = F_r \hat{r} + F_\theta \hat{\theta} + F_\varphi \hat{\varphi}$, $\mathbf{o} = \hat{r}$,

we have

$$\tilde{D}\cdot\mathbf{F} = \hat{r}\left\{DF_r + 2F_r + \frac{2}{\sin\theta}[\partial_\theta(\sin\theta F_\theta) + \partial_\varphi F_\varphi]\right\}$$
$$+ \hat{\theta}\left\{DF_\theta + \frac{1}{\sin^2\theta}[F_\theta + 2\cos\theta\partial_\varphi F_\varphi] - 2\partial_\theta F_r\right\}$$
$$+ \hat{\varphi}\left\{DF_\varphi - \frac{1}{\sin^2\theta}[-F_\varphi + 2\cos\theta\partial_\varphi F_\theta] - \frac{2}{\sin\theta}\partial_\varphi F_r\right\}. \tag{44}$$

If we specialize (43) to $R\mathbf{F} = \mathbf{M}$ of (30), and use Bessel's equation

$$[k^2r^2 + \partial_r(r^2\partial_r h_n)/h_n] = n(n+1), \tag{45}$$

we obtain the vector analog of (42):

$$[n(n+1) - \tilde{D}\cdot]\mathbf{C}_n(\mathbf{o}) = 0. \tag{46}$$

Similarly, if we specialize (43) to $R\mathbf{F} = \mathbf{N}$ of (32), apply (45) for h_{n-1} and h_{n+1} and use the orthogonality properties of $\mathbf{E}_{n-1}$, $\mathbf{H}_{n+1}$ as in (33), we obtain

$$[(n-1)n - \tilde{D}\cdot]\mathbf{E}_{n-1} = 0, \quad \mathbf{E}_{n-1} = \mathbf{P}_n + \mathbf{B}_n/n, \tag{47}$$

$$[(n+1)(n+2) - \tilde{D}\cdot]\mathbf{H}_{n+1} = 0,$$
$$\mathbf{H}_{n+1} = \mathbf{P}_n - \mathbf{B}_n/(n+1). \tag{48}$$

The above provides a different procedure than the usual one of synthesizing solutions of the vector wave equation from known solutions of the scalar equation: We separate variables in the vector equation to obtain (43) and work with solutions of the form $h_\nu(r)\mathbf{F}_\nu(\mathbf{o})$, where $\mathbf{F}$ represents the three sets of eigenvectors $\mathbf{E}_{n-1}$, $\mathbf{C}_n$, and $\mathbf{H}_{n+1}$ of the linear operator $\tilde{D}$.

Using (41) we rewrite $\mathbf{M}$ of (30) and $\mathbf{N}$ of (32) as

$$i_n\mathbf{M}_n = h\mathfrak{D}(n[n+1])\mathbf{C}_n,$$
$$i^{n-1}\mathbf{N}_n = h\frac{n(n+1)}{2n+1}$$
$$\times\left[\mathfrak{D}([n-1]n)\mathbf{E}_{n-1} - \mathfrak{D}([n+1][n+2])\mathbf{H}_{n+1}\right], \tag{49}$$

where the three $\mathfrak{D}$'s are polynomials in $(n[n+1])$, $([n-1]n)$, and $([n+1][n+2])$, respectively. From (46)–(48), we have

$$\mathfrak{D}(\nu[\nu+1])\mathbf{F}_\nu = \mathfrak{D}(\tilde{D}\cdot)\mathbf{F}_\nu \equiv \tilde{\mathfrak{D}}\cdot\mathbf{F}_\nu,$$
$$\tilde{\mathfrak{D}}(kr;\tilde{D}) = \tilde{I} + (i/2kr)\tilde{D}$$
$$+ (i/2kr)^2\tilde{D}\cdot(\tilde{D} - 1\cdot 2\tilde{I}) + \cdots. \tag{50}$$

Using (50) in (49), we obtain

$$i^n\mathbf{M}_n(\mathbf{r}) = h(kr)\tilde{\mathfrak{D}}(kr;\tilde{D})\cdot\mathbf{C}_n(\mathbf{o}), \tag{51}$$

$$i^{n-1}\mathbf{N}_n(\mathbf{r})$$
$$= h(kr)\tilde{\mathfrak{D}}(kr;\tilde{D})\cdot[n(n+1)/(2n+1)](\mathbf{E}_{n-1} - \mathbf{H}_{n+1}). \tag{52}$$

From the definitions in (32),

$$\mathbf{E}_{n-1} - \mathbf{H}_{n+1} = \frac{2n+1}{n(n+1)}\mathbf{B}_n,$$

(52) reduces to

$$i^{n-1}\mathbf{N}_n(\mathbf{r}) = h(kr)\tilde{\mathfrak{D}}(kr;\tilde{D})\cdot\mathbf{B}_n(\mathbf{o}). \tag{53}$$

We may now construct the full vector analog of the scalar solution (3:16). Substituting (51) and (53) into (36), reduces the solution to

$$\mathbf{u} = h(kr)\tilde{\mathfrak{D}}(kr;\tilde{D})\cdot\sum[\mathbf{C}_n^m(\mathbf{o})c_{nm} + \mathbf{B}_n^m(\mathbf{o})b_{nm}]$$
$$= h(kr)\tilde{\mathfrak{D}}(kr;\tilde{D})\cdot\mathbf{g}(\mathbf{o},\mathbf{i}:\epsilon), \tag{54}$$

where the differentiations are with respect to the angles of $\mathbf{o}$.

The longitudinal (with respect to $\mathbf{o}$) $\mathbf{P}$ terms do not appear explicitly in (54) [or in (53)]; however, except for the leading term (the far-field form $h\mathbf{g}$), components along $\mathbf{o}$ are generated by the $\tilde{D}\cdot$ operation. The polar representation of the above series form obtained by using (44) for $\tilde{D}$, with polar components of subsequent terms expressed recursively in terms of the first $(\mathbf{g})$, was derived originally by Wilcox,[12] who also showed that the series in $r^{-\nu}$ converged absolutely and uniformly in r, θ, and φ in any region $r > r'_{\max} = a$.

Since our series for the scattering amplitude $\mathbf{g}(\mathbf{o})$ of (37) is a general transverse form, we see from (27) and (54) that

$$\frac{1}{2\pi}\int e^{ik\mathbf{p}\cdot\mathbf{r}}\mathbf{F}(\mathbf{p})\,d\Omega_p = h(r)\tilde{\mathfrak{D}}(r;\tilde{D})\cdot\mathbf{F}(\mathbf{o}), \tag{55}$$

where r is a parameter, and where $\mathbf{F}(\mathbf{o})$ is representable as a series of transverse vector surface harmonics. To cover vector problems for which $\nabla\cdot\psi \neq 0$, we generalize (55) to include nontransverse components $(\mathbf{P}_n^m)$. This corresponds to fields which involve the longitudinal functions

$$\mathbf{L}_{nm}(\mathbf{r}) = \partial_{kr}[h_n(kr)]\mathbf{P}_n^m(\mathbf{o}) + (h_n/kr)\mathbf{B}_n^m(\mathbf{o})$$
$$= \frac{i^{-n+1}}{2n+1}[nh_{n-1}i^{n-1}\mathbf{E}_{n-1} + (n+1)h_{n+1}i^{n+1}\mathbf{H}_{n+1}]$$
$$= \frac{i^{-n+1}}{2\pi}\int e^{ik\mathbf{p}\cdot\mathbf{r}}\mathbf{P}_n^m(\mathbf{p})\,d\Omega_p, \tag{56}$$

essentially as on p. 1865, and in terms of $\mathbf{E}$ and $\mathbf{H}$ of (32). Thus if

$$\mathbf{F}(\mathbf{o}) = \sum[\mathbf{C}_n^m(\mathbf{o})c_{nm} + \mathbf{B}_n^m(\mathbf{o})b_{nm} + \mathbf{P}_n^m(\mathbf{o})p_{nm}], \tag{57}$$

then substituting (57) for $\mathbf{F}(\mathbf{p})$ in (55) we obtain

$$\frac{1}{2\pi}\int e^{ik\mathbf{p}\cdot\mathbf{r}}\mathbf{F}(\mathbf{p})\,d\Omega_p$$
$$= \sum[\mathbf{M}_{nm}(\mathbf{r})c_{nm} - i\mathbf{N}_{nm}b_{nm} - i\mathbf{L}_{nm}p_{nm}]i^n \equiv \mathbf{V}. \tag{58}$$

From the second equality of (56), and from (41), (47), and (48) we obtain

$$\begin{aligned}\mathbf{L}_{nm}i^{n-1} &= \frac{h}{2n+1}\Big[n\mathfrak{D}([n-1]n)\mathbf{E}_{n-1} \\ &\quad + (n+1)\mathfrak{D}([n+1][n+2])\mathbf{H}_{n+1}\Big] \\ &= h\tilde{\mathfrak{D}}(\tilde{D})\cdot\frac{1}{2n+1}\left[n\mathbf{E}_{n-1}+(n+1)\mathbf{H}_{n+1}\right] \\ &= h\tilde{\mathfrak{D}}(\tilde{D})\cdot\mathbf{P}_n^m(\mathbf{o}), \qquad (59)\end{aligned}$$

where the final form followed from

$$n\mathbf{E}_{n-1}+(n+1)\mathbf{H}_{n+1}=(2n+1)\mathbf{P}_n.$$

Substituting (51), (53), and (59) into the series $\mathbf{V}$ of (58) gives

$$\begin{aligned}\mathbf{V} &= h\tilde{\mathfrak{D}}(\tilde{D})\cdot\sum\left[\mathbf{C}_n^m(\mathbf{o})c_{nm}+\mathbf{B}_n^m(\mathbf{o})b_{nm}+\mathbf{P}_n^m(\mathbf{o})p_{nm}\right] \\ &= h(kr)\tilde{\mathfrak{D}}(kr;\tilde{D})\cdot\mathbf{F}(\mathbf{o}). \qquad (60)\end{aligned}$$

Thus (55) holds for any $\mathbf{F}(\mathbf{o})$ representable in terms of any series of vector surface harmonics.

2.2. Dyadic Fields

We may parallel the above development of the scattering problem of the vector plane wave $\boldsymbol{\varphi}(\mathbf{i}:\boldsymbol{\epsilon})$ of (2) by the analogous development for the dyadic plane wave introduced in (3):

$$\begin{aligned}\tilde{\varphi}(\mathbf{i}) &= (\tilde{I}-\mathbf{ii})e^{ik\mathbf{i}\cdot\mathbf{r}} = (\tilde{I}+\nabla\nabla/k^2)e^{ik\mathbf{i}\cdot\mathbf{r}} \\ &= \nabla\times\nabla\times\tilde{I}e^{ik\mathbf{i}\cdot\mathbf{r}}/k^2. \qquad (61)\end{aligned}$$

The dyadic scattering problem, because of its higher symmetry, is often the easier one: for the vector form (2) we must specify both a direction of incidence $\mathbf{i}$ plus a direction of polarization $\boldsymbol{\epsilon}$ in the plane perpendicular to $\mathbf{i}$, but in (61) we specify only the direction of incidence $\mathbf{i}=\mathbf{k}/k$. The vector plane wave follows from $\boldsymbol{\varphi}(\mathbf{i}:\boldsymbol{\epsilon})=\tilde{\varphi}(\mathbf{i})\cdot\boldsymbol{\epsilon}$, and we may introduce a dyadic scattering amplitude[6–8] $\tilde{g}(\mathbf{o},\mathbf{i})$, such that the vector amplitude follows from

$$\mathbf{g}(\mathbf{o},\mathbf{i}:\boldsymbol{\epsilon})=\tilde{g}(\mathbf{o},\mathbf{i})\cdot\boldsymbol{\epsilon}. \qquad (62)$$

We may rewrite (61) as

$$\tilde{\varphi}(\mathbf{i})=(\boldsymbol{\epsilon}\boldsymbol{\epsilon}+\boldsymbol{\delta}\boldsymbol{\delta})e^{ik\mathbf{i}\cdot\mathbf{r}}=\boldsymbol{\varphi}(\mathbf{i}:\boldsymbol{\epsilon})\boldsymbol{\epsilon}+\boldsymbol{\varphi}(i:\boldsymbol{\delta})\boldsymbol{\delta}, \qquad (63)$$

where $\mathbf{i}$, $\boldsymbol{\epsilon}$, $\boldsymbol{\delta}$ form an orthogonal set of unit vectors. From the superposition principle, the corresponding dyadic scattered wave is thus

$$\tilde{u}(\mathbf{r};\mathbf{i})=\mathbf{u}(\mathbf{i}:\boldsymbol{\epsilon})\boldsymbol{\epsilon}+\mathbf{u}(\mathbf{i}:\boldsymbol{\delta})\boldsymbol{\delta}. \qquad (64)$$

Asymptotically, we have

$$\tilde{u}(\mathbf{r};\mathbf{i})\sim h(kr)[\mathbf{g}(\mathbf{o},\mathbf{i}:\boldsymbol{\epsilon})\boldsymbol{\epsilon}+\mathbf{g}(\mathbf{o},\mathbf{i}:\boldsymbol{\delta})\boldsymbol{\delta}]=h(kr)\tilde{g}(\mathbf{o},\mathbf{i}) \qquad (65)$$

with

$$\tilde{g}(\mathbf{o},\mathbf{i})=\mathbf{g}(\mathbf{o},\mathbf{i}:\boldsymbol{\epsilon})\boldsymbol{\epsilon}+\mathbf{g}(\mathbf{o},\mathbf{i}:\boldsymbol{\delta})\boldsymbol{\delta} \qquad (66)$$

in accord with (62) and with Saxon's definition.[7,8]

Although we could construct the dyadic functions from the vector ones by using (64) and (66), it is somewhat simpler to consider the dyadic scattering problem systematically. In the following, (1d) means Eq. (1) in terms of $\tilde{\psi}$, etc.

Surface integrals: If we transpose the dyadic $\tilde{\Gamma}$ terms in (10), we obtain

$$\begin{aligned}&\int[\tilde{\Gamma}^{\mathrm{T}}\cdot(\nabla\times\nabla\times\mathbf{u})-(\nabla\times\nabla\times\tilde{\Gamma})^{\mathrm{T}}\cdot\mathbf{u}]\,dV \\ &\quad=-\int[(\mathbf{n}\times\tilde{\Gamma})^{\mathrm{T}}\cdot(\nabla\times\mathbf{u})-(\nabla\times\tilde{\Gamma})^{\mathrm{T}}\cdot(n\times\mathbf{u})]\,dS, \qquad (67)\end{aligned}$$

where the superscript T indicates the transposed (Gibbs' conjugate) dyadic. For any dyadic solution of (1d),

$$\begin{aligned}&(\nabla\times\nabla\times\tilde{F})^{\mathrm{T}}=k^2\tilde{F}^{\mathrm{T}}=\nabla\times\nabla\times\tilde{F}^{\mathrm{T}}, \\ &(\mathbf{n}\times\tilde{F})^{\mathrm{T}}=-\tilde{F}^{\mathrm{T}}\times\mathbf{n}, \quad (\nabla\times\tilde{F})^{\mathrm{T}}=-\tilde{F}^{\mathrm{T}}(\times\nabla), \qquad (68)\end{aligned}$$

where $(\times\nabla)$ operating to the left on $\tilde{F}$ in the last equality means differentiate to the left on $\tilde{F}$ but leave the vector part of $\times\nabla$ on the right of $\tilde{F}$. In particular, for $\tilde{F}=\tilde{\Gamma}$ of (9), we have

$$\begin{aligned}&\tilde{\Gamma}^{\mathrm{T}}=\tilde{\Gamma}, \quad (n\times\tilde{\Gamma})^{\mathrm{T}}=-\tilde{\Gamma}\times\mathbf{n}, \\ &(\nabla\times\tilde{\Gamma})^{\mathrm{T}}=-\nabla\times\tilde{\Gamma}=h_1(k\,|\mathbf{r}-\mathbf{r}'|)k\mathbf{o}\times\tilde{I}. \qquad (69)\end{aligned}$$

From the steps leading to (10) and (67), we obtain

$$\begin{aligned}&\int[\tilde{F}^{\mathrm{T}}\cdot(\nabla\times\nabla\times\tilde{u})-(\nabla\times\nabla\times\tilde{F})^{\mathrm{T}}\cdot\tilde{u}]\,dV \\ &\quad=-\int[(\mathbf{n}\times\tilde{F})^{\mathrm{T}}\cdot(\nabla\times\tilde{u})-(\nabla\times\tilde{F})^{\mathrm{T}}\cdot(\mathbf{n}\times\tilde{u})]\,dS. \qquad (70)\end{aligned}$$

In the region external to the scatterer, we use (70) for $\tilde{F}=\tilde{\Gamma}=\tilde{h}k/4\pi i$ to obtain

$$\begin{aligned}\tilde{u}(\mathbf{r};\mathbf{i}) &= \frac{k}{4\pi i}\int[(n\times\tilde{h})^{\mathrm{T}}\cdot(\nabla\times\tilde{u}) \\ &\quad-(\nabla\times\tilde{h})^{\mathrm{T}}\cdot(\mathbf{n}\times\tilde{u})]\,dS\equiv\{\tilde{h},\tilde{u}\}, \qquad (71)\end{aligned}$$

where $\mathbf{n}$ points away from the scatterer. It is this definition of the brace operation for dyadics, equivalent to (12) for $\tilde{u}$ replaced by a vector $\mathbf{u}$, that we use henceforth. Since $(\tilde{A}^{\mathrm{T}}\cdot\tilde{B})^{\mathrm{T}}=\tilde{B}^{\mathrm{T}}\cdot\tilde{A}$, we have $\tilde{u}^{\mathrm{T}}=\{\tilde{h},\tilde{u}^{\mathrm{T}}\}$, and also

$$\begin{aligned}\tilde{u}^{\mathrm{T}}=\{\tilde{h},\tilde{u}\}^{\mathrm{T}} &= \frac{k}{4\pi i}\int[(\nabla\times\tilde{u})^{\mathrm{T}}\cdot(\mathbf{n}\times\tilde{h}) \\ &\quad-(\mathbf{n}\times\tilde{u})^{\mathrm{T}}\cdot(\nabla\times\tilde{h})]\,dS=-\{\tilde{u},\tilde{h}\}. \qquad (72)\end{aligned}$$

Similarly

$$\tilde{g}(\mathbf{o}, \mathbf{i}) = \{\tilde{\varphi}(-\mathbf{o}), \tilde{u}(\mathbf{r}'; \mathbf{i})\}. \tag{73}$$

From (18d), i.e., $\{\tilde{\varphi}_1, \tilde{\varphi}_2\} = 0$, we proceed as for (19) to obtain

$$\{\tilde{\varphi}_1, \tilde{u}_2\} = -\{\tilde{u}_1, \tilde{\varphi}_2\} = \{\tilde{\varphi}_2, \tilde{u}_1\}^{\mathrm{T}}, \tag{74}$$

where the last equality follows from (72). Thus using (73) in (74), we obtain Saxon's result[8]

$$\tilde{g}(-\mathbf{i}_1, \mathbf{i}_2) = \tilde{g}^{\mathrm{T}}(-\mathbf{i}_2, \mathbf{i}_1), \tag{75}$$

which also follows from (20) and (62):

$$\begin{aligned}\boldsymbol{\epsilon}_1 \cdot [\tilde{g}(-\mathbf{i}_1, \mathbf{i}_2) \cdot \boldsymbol{\epsilon}_2] &= \boldsymbol{\epsilon}_2 \cdot [\tilde{g}(-\mathbf{i}_2, \mathbf{i}_1) \cdot \boldsymbol{\epsilon}_1] \\ &= \boldsymbol{\epsilon}_1 \cdot \tilde{g}^{\mathrm{T}}(-\mathbf{i}_2, \mathbf{i}_1) \cdot \boldsymbol{\epsilon}_2.\end{aligned}$$

From (74), we see that $\mathbf{i}_1 \cdot \tilde{g}(\mathbf{i}_1, \mathbf{i}_2) = \tilde{g}(\mathbf{i}_1, \mathbf{i}_2) \cdot \mathbf{i}_2 = 0$, i.e., $\tilde{g}$ is transverse both fore and aft [cf. (66)]. From (66) and (74), we have

$$\begin{aligned}\tilde{g}(\mathbf{i}_1, \mathbf{i}_2) = \tilde{g}^{\mathrm{T}}(-\mathbf{i}_2, -\mathbf{i}_1) &= \boldsymbol{\epsilon}_1 \mathbf{g}(-\mathbf{i}_2, -\mathbf{i}_1 : \boldsymbol{\epsilon}_1) \\ &\quad + \boldsymbol{\delta}_1 \mathbf{g}(-\mathbf{i}_2, -\mathbf{i}_1 : \boldsymbol{\delta}_1),\end{aligned} \tag{76}$$

which supplements (66) in providing a vector representation for $\tilde{g}(\mathbf{i}_1, \mathbf{i}_2)$ in terms of observed instead of incident polarizations.

Similarly from (21d), i.e., $\{\tilde{\psi}_1^*, \tilde{\psi}_2\} = 0$, we proceed as for (21) to obtain

$$\{\tilde{\varphi}_1^*, \tilde{u}_2\} + \{\tilde{u}_1^*, \tilde{\varphi}_2\} + \{\tilde{u}_1^*, \tilde{u}_2\} = 0. \tag{77}$$

The first term equals $\tilde{g}(\mathbf{i}_1^*, \mathbf{i}_2)$, the second reduces to $+\{\tilde{\varphi}_2^*, \tilde{u}_1\}^{\mathrm{T}*} = \tilde{g}^{\mathrm{T}*}(\mathbf{i}_2^*, \mathbf{i}_1)$, and the last equals

$$\begin{aligned}\{\tilde{u}_1^*, \tilde{u}_2\} &= \frac{k}{4\pi i} 2ik \int (\mathbf{o} \times \tilde{g}_1 h)^{*\mathrm{T}} \cdot (\mathbf{o} \times \tilde{g}_2 h)\, dS \\ &= \frac{1}{2\pi} \int \tilde{g}^{\mathrm{T}*}(\mathbf{o}, \mathbf{i}_1) \cdot \tilde{g}(\mathbf{o}, \mathbf{i}_2)\, d\Omega.\end{aligned} \tag{78}$$

Thus the dyadic analog of (23) is

$$\tilde{g}(\mathbf{i}_1^*, \mathbf{i}_2) + \tilde{g}^{\dagger}(\mathbf{i}_2^*, \mathbf{i}_1) = -\frac{1}{2\pi} \int \tilde{g}^{\dagger}(\mathbf{o}, \mathbf{i}_1) \cdot \tilde{g}(\mathbf{o}, \mathbf{i}_2)\, d\Omega, \qquad \tilde{g}^{\dagger} \equiv \tilde{g}^{\mathrm{T}*}, \tag{79}$$

as obtained originally by Saxon[8] by a briefer, more abstract procedure. The symbol $\tilde{g}^{\dagger}$ represents the Hermitian adjoint of $\tilde{g}$. In the forward direction $\mathbf{i}_1^* = \mathbf{i}_2^* = \mathbf{i}_1 = \mathbf{i}_2$ we may reduce (79) to (24):

$$\begin{aligned}-\boldsymbol{\epsilon} \cdot [\tilde{g}(\mathbf{i}, \mathbf{i}) + \tilde{g}^{\dagger}(\mathbf{i}, \mathbf{i})] \cdot \boldsymbol{\epsilon} &= -2 \operatorname{Re} [\boldsymbol{\epsilon} \cdot \tilde{g}(\mathbf{i}, \mathbf{i}) \cdot \boldsymbol{\epsilon}] \\ &= \frac{1}{2\pi} \int |\tilde{g}(\mathbf{o}, \mathbf{i}) \cdot \boldsymbol{\epsilon}|^2\, d\Omega \\ &= \frac{k^2}{4\pi} Q(\mathbf{i} : \boldsymbol{\epsilon}).\end{aligned} \tag{80}$$

Plane wave form: To construct the dyadic analog of (27), we use $(\tilde{I} - \mathbf{pp}) \cdot (\tilde{I} - \mathbf{pp}) = \tilde{I} - \mathbf{pp}$, and rewrite (26) in terms of the form (61) as

$$\tilde{h}(k\,|\mathbf{r} - \mathbf{r}'|) = \frac{1}{2\pi} \int \tilde{\varphi}(\mathbf{r}; \mathbf{p}) \cdot \tilde{\varphi}(\mathbf{r}'; -\mathbf{p})\, d\Omega_p. \tag{81}$$

Substituting in (71) and using (73), we obtain

$$\begin{aligned}\tilde{u}(\mathbf{r}; \mathbf{i}) &= \frac{1}{2\pi} \int \tilde{\varphi}(\mathbf{r}; \mathbf{p}) \cdot \tilde{g}(\mathbf{p}, \mathbf{i})\, d\Omega_p \\ &= \frac{1}{2\pi} \int e^{ik\mathbf{p}\cdot\mathbf{r}} \tilde{g}(\mathbf{p}, \mathbf{i})\, d\Omega_p.\end{aligned} \tag{82}$$

Similarly for (55d), etc.

If the scatterer is not at the origin $r = 0$, but at $\mathbf{r} = \mathbf{b}$ then we may work with

$$\begin{aligned}\tilde{u} \cdot \tilde{\varphi}(\mathbf{b}; \mathbf{i}) \sim h(kr) \tilde{\varphi}(\mathbf{b}; -\mathbf{o}) \cdot \tilde{g}(\mathbf{o}, \mathbf{i}) \cdot \tilde{\varphi}(\mathbf{b}; \mathbf{i}) \\ = h e^{ik\mathbf{b}\cdot(\mathbf{i}-\mathbf{o})} \tilde{g}(\mathbf{o}, \mathbf{i}).\end{aligned} \tag{83}$$

Special function series: Corresponding to

$$\begin{aligned}\tilde{\varphi}(\mathbf{i}) = \tilde{\varphi}^{\mathrm{T}}(\mathbf{i}) &= (\tilde{I} - \mathbf{ii}) e^{ik\mathbf{i}\cdot\mathbf{r}} \\ &= \sum [\mathbf{M}_{nm}^1(\mathbf{r}) \mathbf{C}_n^{-m}(\mathbf{i}) - i\mathbf{N}_{nm}^1 \mathbf{B}_n^{-m}] i^n (-1)^m\, d_n,\end{aligned}$$

$$d_n = \frac{2n+1}{n(n+1)}, \tag{84}$$

we have

$$\tilde{u}(\mathbf{r}; \mathbf{i}) = \sum_{m,n} [\mathbf{M}_{nm}(\mathbf{r}) \mathbf{c}_{nm}(\mathbf{i}) - i\mathbf{N}_{nm} \mathbf{b}_{nm}(\mathbf{i})], \tag{85}$$

$$\tilde{g}(\mathbf{o}, \mathbf{i}) = \sum [\mathbf{C}_n^m(\mathbf{o}) \mathbf{c}_{nm}(\mathbf{i}) + \mathbf{B}_n^m \mathbf{b}_{nm}(\mathbf{i})], \tag{86}$$

where

$$\begin{aligned}\mathbf{c}_{nm}(\mathbf{i}) &= \sum_{\nu,\mu} [\alpha_{nm\nu\mu} \mathbf{C}_\nu^{-\mu}(\mathbf{i}) + \beta_{mn\nu\mu} \mathbf{B}_\nu^{-\mu}(\mathbf{i})], \\ \mathbf{b}_{nm}(\mathbf{i}) &= \sum [\gamma_{mn\nu\mu} \mathbf{C}_\nu^{-\mu}(\mathbf{i}) + \delta_{nm\nu\mu} \mathbf{B}_\nu^{-\mu}(\mathbf{i})].\end{aligned} \tag{87}$$

The reciprocity relation (75) gives

$$\alpha_{\nu,-\mu,n,-m} = (-1)^{n+\nu} \alpha_{nm\nu\mu},$$

and similarly for δ; for β and γ we obtain this form with $(-1)^{n+\nu+1}$.

For a spherically symmetric scatterer,

$$\tilde{u}(\mathbf{r}; \mathbf{i}) = \sum [\mathbf{M}_{nm}(\mathbf{r}) \mathbf{C}_n^{-m}(\mathbf{i}) c_n - i\mathbf{N}_{nm} \mathbf{B}_n^{-m} b_n] i^n (-1)^m, \tag{88}$$

$$\tilde{g}(\mathbf{o}, \mathbf{i}) = \sum [\mathbf{C}_n^m(\mathbf{o}) \mathbf{C}_n^{-m}(\mathbf{i}) c_n + \mathbf{B}_n^m(\mathbf{o}) \mathbf{B}_n^{-m}(\mathbf{i}) b_n] (-1)^n, \qquad \sum = \sum_{n=1}^{\infty} \sum_{m=-n}^{n}, \tag{89}$$

where b and c are independent of directions. We may rewrite (89) as

$$\tilde{g}(\mathbf{o}, \mathbf{i}) = \sum_{n=1}^{\infty} [\tilde{C}_n(\mathbf{o}, \mathbf{i}) c_n + \tilde{B}_n(\mathbf{o}, \mathbf{i}) b_n],$$

$$\begin{aligned}\tilde{C}_n(\mathbf{o}, \mathbf{i}) &= \sum_{m=-n}^{n} \mathbf{C}_n^m(\mathbf{o}) \mathbf{C}_n^{-m}(\mathbf{i}) (-1)^m = \mathbf{L}(\mathbf{o}) L(\mathbf{i}) P_n(\mathbf{o} \cdot \mathbf{i}), \\ \tilde{B}_n(\mathbf{o}, \mathbf{i}) &= [\mathbf{o} \times \mathbf{L}(\mathbf{o})][\mathbf{i} \times \mathbf{L}(\mathbf{i})] P_n(\mathbf{o} \cdot \mathbf{i}),\end{aligned} \tag{90}$$

where $P_n(\mathbf{o}\cdot\mathbf{i}) = \sum Y_n^m(\mathbf{o})\, Y_n^{-m}(\mathbf{i})(-1)^m$ is the Legendre polynomial, and $\mathbf{L}$ is defined in (30).

The form (90) has essentially the same symmetry as for the scalar problem: The reciprocity relation (75) reduces to

$$\tilde{g}(\mathbf{o}, \mathbf{i}) = \tilde{g}(-\mathbf{o}, -\mathbf{i}) = \tilde{g}^{\mathrm{T}}(\mathbf{i}, \mathbf{o}). \tag{91}$$

Substituting (91) into (79) gives the simpler form

$$-\operatorname{Re} \tilde{g}(\mathbf{i}_1, \mathbf{i}_2) = \frac{1}{4\pi}\int \tilde{g}^*(\mathbf{i}_1, \mathbf{o}) \cdot \tilde{g}(\mathbf{o}, \mathbf{i}_2)\, d\Omega, \tag{92}$$

and using (90), and

$$\int \tilde{C}_n(\mathbf{i}_1, \mathbf{o}) \cdot \tilde{B}_\nu(\mathbf{o}, \mathbf{i}_2)\, d\Omega = 0,$$

$$\int \tilde{D}_n(\mathbf{i}_1, \mathbf{o}) \cdot \tilde{D}_\nu(\mathbf{o}, \mathbf{i}_2)\, d\Omega = 4\pi \tilde{D}_n(\mathbf{i}_1, \mathbf{i}_2)\delta_{n\nu}/d_n,$$

with $\tilde{D} = \tilde{C}$ or $\tilde{B}$, we obtain

$$-d_n \operatorname{Re} c_n = |c_n|^2, \quad -d_n \operatorname{Re} b_n = |b_n|^2. \tag{93}$$

In the forward direction, we have

$$\tilde{g}(\mathbf{i}, \mathbf{i}) = (\tilde{I} - \mathbf{ii}) \sum [\tfrac{1}{2}n(n+1)(b_n + c_n)], \tag{94}$$

and the total cross section equals $-4\pi/k^2$ times $\operatorname{Re} \sum [\ ]$.

If only the dipole terms are significant, then

$$\begin{aligned}
&\tilde{g}(\mathbf{o}, \mathbf{i}) = \tilde{C}_1 c_1 + \tilde{B}_1 b_1, \quad \tilde{C}_1 \equiv \tilde{C}_1^0 + \tilde{C}_1^1, \quad \tilde{B}_1 \equiv \tilde{B}_1^0 + \tilde{B}_1^1,\\
&\tilde{C}_1^0 = \mathbf{C}_1^0(\mathbf{o})\mathbf{C}_1^0(\mathbf{i}) = \hat{\varphi}\hat{\varphi}_i \sin\theta \sin\theta_i,\\
&\tilde{C}_1^1 = \operatorname{Re} \mathbf{C}_1^1(\mathbf{o})\mathbf{C}_1^{1*}(\mathbf{i})\\
&\quad = (\hat{\theta}\hat{\theta}_i + \hat{\varphi}\hat{\varphi}_i \cos\theta \cos\theta_i)\cos(\varphi - \varphi_i)\\
&\qquad + (\hat{\theta}\hat{\varphi}_i \cos\theta_i - \hat{\varphi}\hat{\theta}_i \cos\theta)\sin(\varphi - \varphi_i),\\
&\tilde{B}_1^0 = \mathbf{B}_1^0(\mathbf{o})\mathbf{B}_1^0(\mathbf{i}) = \hat{\theta}\hat{\theta}_i \sin\theta \sin\theta_i,\\
&\tilde{B}_1^1 = \operatorname{Re} \mathbf{B}_1^1(\mathbf{o})\mathbf{B}_1^{1*}(\mathbf{i})\\
&\quad = (\hat{\theta}\hat{\theta}_i \cos\theta \cos\theta_i + \hat{\varphi}\hat{\varphi}_i)\cos(\varphi - \varphi_i)\\
&\qquad + (\hat{\theta}\hat{\varphi}_i \cos\theta - \hat{\varphi}\hat{\theta}_i \cos\theta_i)\sin(\varphi - \varphi_i).
\end{aligned} \tag{95}$$

For a homogeneous sphere of radius a, for the surface conditions (6d),

$$c_n = \frac{-j_n(x)}{h_n(x)}\, d_n, \quad b_n = \frac{-\partial_x[x j_n(x)]}{\partial_x[x h_n(x)]}\, d_n, \quad x = ka. \tag{96}$$

For conditions (7d) plus (8d), we supplement (84) and (88) with the internal field

$$\tilde{\psi}' = \sum [\mathbf{M}_{nm}^1(k'\mathbf{r})\mathbf{C}_n^{-m}(\mathbf{i})c_n' - i\mathbf{N}_{nm}^1\mathbf{B}_n^{-m}b_n'] i^n(-1)^m, \tag{97}$$

and obtain

$$\begin{aligned}
c_n &= -\frac{j_n(X)\partial_x[x j_n(x)] - j_n(x)\partial_x[X j_n(X)]/\mu}{j_n(X)\partial_x[x h_n(x)] - h_n(x)\partial_x[X j_n(X)]/\mu}\, d_n\\
&= c_n(\mu), \quad b_n = c_n(\epsilon), \quad X = k'a.
\end{aligned} \tag{98}$$

See Morse and Feshbach[6] (pp. 1882ff), Stratton[10] (pp. 563ff), and Van de Hulst[17] (pp. 113ff).

Small scatterer of arbitrary shape: For an arbitrarily shaped scatterer with all dimensions very small compared to wavelength, in terms of dyadic electric ($\tilde{p}$) and magnetic ($\tilde{m}$) dipole moments (pp. 1886ff), we have

$$\begin{aligned}
\tilde{u}\cdot\tilde{\varphi} &= \tilde{h}(kr)\cdot\tilde{p}\cdot\tilde{\varphi} + (\nabla\times\tilde{h})\cdot\tilde{m}\cdot(\nabla\times\tilde{\varphi}/k^2)\\
&= [\tilde{h}\cdot\tilde{p} + (\nabla\times\tilde{h})\cdot\tilde{m}\cdot(\mathbf{i}\times\tilde{I})i/k]\cdot\tilde{\varphi},
\end{aligned} \tag{99}$$

where $\tilde{p}$ arises from the E field $\tilde{\varphi}$, and $\tilde{m}$ from the associated H field proportional to $\nabla\times\tilde{\varphi} = \mathbf{i}\times\tilde{\varphi}ik = \mathbf{i}\times\tilde{I}\cdot\tilde{\varphi}ik$; both $\tilde{p}$ and $\tilde{m}$ are independent of $\mathbf{i}$ and $\mathbf{o}$. From the definition of $\tilde{h}$ in (12), we obtain

$$\begin{aligned}
&\tilde{h} = (\tilde{I} - \mathbf{oo})\mathcal{H} + \mathbf{oo}H = \tilde{h}^{\mathrm{T}},\\
&\mathcal{H}(x) = \frac{\partial_x[x h_1(x)]}{x}, \quad H(x) = \frac{2h_1}{x},\\
&\nabla\times\tilde{h} = -kh_1\mathbf{o}\times\tilde{I} = -kh_1\tilde{I}\times\mathbf{o} = -(\nabla\times h)^{\mathrm{T}}.
\end{aligned} \tag{100}$$

Using $\tilde{h} \sim (\tilde{I} - \mathbf{oo})h$, $\nabla\times\tilde{h} \sim i\mathbf{o}\times\tilde{I}h$ in (99) to obtain $\tilde{u} \sim \tilde{g}h$, we write

$$\begin{aligned}
\tilde{g}(\mathbf{o}, \mathbf{i}) &= (\tilde{I} - \mathbf{oo})\cdot\tilde{p}\cdot(\tilde{I} - \mathbf{ii})\\
&\quad - (\mathbf{o}\times\tilde{I})\cdot\tilde{m}\cdot(\tilde{I}\times\mathbf{i}) \equiv \tilde{g}_e + \tilde{g}_m.
\end{aligned} \tag{101}$$

Here

$$\begin{aligned}
\tilde{I} - \mathbf{ii} &= \boldsymbol{\epsilon}\boldsymbol{\epsilon} + \boldsymbol{\delta}\boldsymbol{\delta} = -(\tilde{I}\times\mathbf{i})\cdot(\tilde{I}\times\mathbf{i}),\\
\tilde{I}\times\mathbf{i} &= \mathbf{i}\times\tilde{I} = \boldsymbol{\delta}\boldsymbol{\epsilon} - \boldsymbol{\epsilon}\boldsymbol{\delta}
\end{aligned}$$

are both planar dyadics; the first is symmetrical, and the second is antisymmetrical. Both annihilate components of vectors parallel to $\mathbf{i}$; the second ($\tilde{I}\times\mathbf{i}$) turns perpendicular components through 90° around $\mathbf{i}$ as an axis, and the first [$\tilde{I} - \mathbf{ii} = -(\tilde{I}\times\mathbf{i})^2$] is the negative of a turn through 180°; see Gibbs[18] for detailed discussion of $(\tilde{I}\times\mathbf{i})^n$.

From theorem (75) applied to (101), we obtain

$$\tilde{p} = \tilde{p}^{\mathrm{T}}, \quad \tilde{m} = \tilde{m}^{\mathrm{T}}; \tag{102}$$

thus each is symmetrical and may be put in the form $\tilde{p} = p_\xi\hat{\xi}\hat{\xi} + p_\eta\hat{\eta}\hat{\eta} + p_\zeta\hat{\zeta}\hat{\zeta}$, where the vectors correspond to the principal axis. From theorem (79), we obtain

$$\begin{aligned}
-\operatorname{Re}\tilde{p} &= \frac{1}{4\pi}\int \tilde{p}\cdot(\tilde{I} - \mathbf{oo})\cdot\tilde{p}^*\, d\Omega = \tfrac{2}{3}\tilde{p}\cdot\tilde{p}^*,\\
-\operatorname{Re}\tilde{m} &= -\frac{1}{4\pi}\int \tilde{m}\cdot(\tilde{I}\times\mathbf{o})\cdot(\tilde{I}\times\mathbf{o})\cdot\tilde{m}^*\, d\Omega\\
&= \tfrac{2}{3}\tilde{m}\cdot\tilde{m}^*.
\end{aligned} \tag{103}$$

[17] H. C. van de Hulst, *Light Scattering by Small Particles* (John Wiley & Sons, Inc., New York, 1953), Chap. 9.

[18] See J. Willard Gibbs, *Vector Analysis*, Vol. II *Collected Works, Vol. II* (Yale University Press, New Haven, Conn., 1948), pp. 61ff, for $(I\times\mathbf{i})^n$; and also E. B. Wilson, *Gibb's Vector Analysis* (Yale University Press, New Haven, Conn., 1943), pp. 299ff. More generally, the dyadic operations of this paper are based on their development, and also on C. E. Weatherburn, *Advanced Vector Analysis* (Bell and Sons, London, 1949), and on Ref. 6.

With s equal to either ξ, η, or ζ, we have $-\mathrm{Re}\, p_s = \frac{2}{3}|p_s|^2$; similarly, with s or t or r equal to either x, y, or z, we have $-\mathrm{Re}\, p_{st} = -\mathrm{Re}\, p_{ts} = \frac{2}{3}\sum p_{sr}p^*_{rt}$.

The special case (95) corresponds to $\tilde{p} = b_1\tilde{I}$, $\tilde{m} = c_1\tilde{I}$:

$$\begin{aligned}\tilde{g}(\mathbf{o},\mathbf{i}) &= b_1(\tilde{I}-\mathbf{oo})\cdot(\tilde{I}-\mathbf{ii}) - c_1(\mathbf{o}\times\tilde{I})\cdot(\tilde{I}\times\mathbf{i})\\ &= b_1(\hat{\theta}\hat{\theta}+\hat{\varphi}\hat{\varphi})\cdot(\hat{\theta}_i\hat{\theta}_i+\hat{\varphi}_i\hat{\varphi}_i)\\ &\quad + c_1(\hat{\varphi}\hat{\theta}-\hat{\theta}\hat{\varphi})\cdot(\hat{\varphi}_i\hat{\theta}_i-\hat{\theta}_i\hat{\varphi}_i). \quad (104)\end{aligned}$$

For later use, we make the relations between (95) and (104) explicit by rewriting $\tilde{h}$ in terms of Hansen's functions. From (35) for $r'\to 0$ we see that all terms vanish except

$$\begin{aligned}\mathbf{N}_{10} &\to \tfrac{2}{3}(\mathbf{P}^0_1+\mathbf{B}^0_1) = \tfrac{2}{3}(\hat{r}\cos\theta - \hat{\theta}\sin\theta) = \tfrac{2}{3}\hat{z},\\ \mathbf{N}_{11} &\to \tfrac{2}{3}(\mathbf{P}^1_1+\mathbf{B}^1_1) = \tfrac{2}{3}[\hat{r}e^{i\varphi}\sin\theta + e^{i\varphi}(\cos\theta\hat{\theta}+i\hat{\varphi})]\\ &= \tfrac{2}{3}(\hat{x}+i\hat{y}),\\ \mathbf{N}_{1-1} &= -\tfrac{1}{2}(\mathbf{N}^*_{11}) \to -\tfrac{1}{2}\tfrac{2}{3}(\hat{x}-i\hat{y}),\end{aligned} \quad (105)$$

where $\hat{z}$, $(\hat{x}+i\hat{y})/\sqrt{2}$ and $(\hat{x}-i\hat{y})/\sqrt{2}$ form a set of orthonormal vectors. Consequently,

$$\begin{aligned}\tilde{h}(kr) = h^{\mathrm{T}} &= \hat{z}\mathbf{N}_{10}(kr) + \tfrac{1}{2}(\hat{x}-i\hat{y})\mathbf{N}_{11} - (\hat{x}+i\hat{y})\mathbf{N}_{1-1}\\ &\equiv \hat{z}\mathbf{N}_{10} + \hat{x}\mathbf{N}_{11e} + \hat{y}\mathbf{N}_{11o}, \quad (106)\end{aligned}$$

Using the asymptotic forms of the left- and right-hand sides, we also have

$$\begin{aligned}\tilde{I}-\mathbf{oo} &= \hat{z}\mathbf{B}^0_1(\mathbf{o}) + \tfrac{1}{2}(\hat{x}-i\hat{y})\mathbf{B}^1_1 - (\hat{x}+i\hat{y})\mathbf{B}^{-1}_1\\ &= \hat{z}\mathbf{B}^0_1(\mathbf{o}) + \hat{x}\,\mathrm{Re}\,\mathbf{B}^1_1 + \hat{y}\,\mathrm{Im}\,\mathbf{B}^1_1\\ &\equiv \hat{z}B_z(\mathbf{o}) + \hat{x}B_x(\mathbf{o}) + \hat{y}B_y(\mathbf{o}), \quad (107)\end{aligned}$$

where since, $(\tilde{I}-\mathbf{oo}) = (\tilde{I}-\mathbf{oo})^{\mathrm{T}}$, we may transpose the left and right members of each term. Similarly, since $\nabla\times\mathbf{N}\hat{z} = (\nabla\times\mathbf{N})\hat{z} = k\mathbf{M}\hat{z}$ etc.,

$$\begin{aligned}\nabla\times\tilde{h}/k &= -(\nabla\times\tilde{h})^T/k\\ &= \mathbf{M}_{10}\hat{z} + \tfrac{1}{2}\mathbf{M}_{11}(\hat{x}-i\hat{y}) - \mathbf{M}_{1-1}(\hat{x}+i\hat{y})\\ &= -\hat{z}\mathbf{M}_{10} - \hat{x}\mathbf{M}_{11e} - \hat{y}\mathbf{M}_{11o}, \quad (108)\end{aligned}$$

$$\begin{aligned}\mathbf{o}\times\tilde{I} &= \tilde{I}\times\mathbf{o}\\ &= -(\mathbf{o}\times\tilde{I})^T = \hat{z}\mathbf{C}^0_1 + \hat{x}\,\mathrm{Re}\,\mathbf{C}^1_1 + \hat{y}\,\mathrm{Im}\,\mathbf{C}^1_1\\ &\equiv \hat{z}\mathbf{C}_z + \hat{x}\mathbf{C}_x + \hat{y}\mathbf{C}_y. \quad (109)\end{aligned}$$

Since $\tilde{I}\times\mathbf{o} = (\tilde{I}-\mathbf{oo})\cdot(\tilde{I}\times\mathbf{o})$, we also have $\tilde{I}\times\mathbf{o} = \mathbf{B}_z(\mathbf{o})\mathbf{C}_z(\mathbf{o}) + \mathbf{B}_x\mathbf{C}_x + \mathbf{B}_y\mathbf{C}_y$.

Substituting (107) into the electric term of (101), and letting $\hat{s}$ and $\hat{t}$ range over $\hat{x}$, $\hat{y}$, $\hat{z}$ we may write

$$\tilde{g}_e(\mathbf{o},\mathbf{i}) = \sum_{s,t}\mathbf{B}_s(\mathbf{o})p_{st}\mathbf{B}_t(\mathbf{i}), \quad p_{st} = \hat{s}\cdot\tilde{p}\cdot\hat{t}; \quad \hat{s},\hat{t} = \hat{x},\hat{y},\hat{z}. \quad (110)$$

We may also work with the first form of (107) to obtain

$$\tilde{g}_e(\mathbf{o},\mathbf{i}) = \sum_{m\mu}\mathbf{B}^m_1(\mathbf{o})\mathbf{B}^{-\mu}_1(\mathbf{i})p_{m\mu}; \quad m,\mu = -1,0,+1, \quad (111)$$

where, e.g.,

$$p_{01} = \hat{z}\cdot\tilde{p}\cdot(\hat{x}+i\hat{y}) = (\mathbf{P}^0_1+\mathbf{B}^0_1)\cdot\tilde{p}\cdot(\mathbf{P}^1_1+\mathbf{B}^1_1).$$

If the principal axes of $\tilde{p}$ coincide with $\hat{x}$, $\hat{y}$, $\hat{z}$ then (110) reduces to

$$\tilde{g}_e(\mathbf{o},\mathbf{i}) = \mathbf{B}_x(\mathbf{o})\mathbf{B}_x(\mathbf{i})p_x + \mathbf{B}_y(\mathbf{o})\mathbf{B}_y(\mathbf{i})p_y + \mathbf{B}_z(\mathbf{o})\mathbf{B}_z(\mathbf{i})p_z. \quad (112)$$

The analogous discussion goes through for $\mathbf{g}_m$, i.e.,

$$\begin{aligned}\mathbf{g}_m &= -(\tilde{I}\times\mathbf{o})\cdot\tilde{m}\cdot(\tilde{I}\times\mathbf{i}) = (\tilde{I}\times\mathbf{o})^T\cdot\tilde{m}\cdot(\tilde{I}\times\mathbf{i})\\ &= \sum_{st}\mathbf{C}_s(\mathbf{o})m_{st}\mathbf{C}_t(\mathbf{i}), \quad \text{etc.} \quad (113)\end{aligned}$$

See Morse and Feshbach[6] (pp. 1886ff) for an alternative development and for illustrations of $\tilde{p}$ and $\tilde{m}$. Electric dipole dyadics are also considered by Yvon,[19] Mazur,[20] Fixman,[21] Brown,[22] and others.

3. MANY SCATTERERS

For many scatterers in the geometry of Fig. 3:1, we write the vector field as

$$\Psi = \varphi(\mathbf{i}:\epsilon) + \mathfrak{U}, \quad \mathfrak{U} \sim h(kr)\mathfrak{S}(\mathbf{o},\mathbf{i}:\epsilon), \quad (114)$$

where $\mathfrak{U}$ and $\mathfrak{S}$ have the forms (12) and (16) with $\mathbf{u}$ replaced by $\mathfrak{U}$. The "compound amplitude" $\mathfrak{S}$ fulfills the same theorems as $\mathbf{g}$.

Proceeding as in Refs. 1 and 3, we express the total scattered field of a configuration of scatterers (whose "centers" are at $\mathbf{b}_s$) as

$$\mathfrak{U} = \sum\mathbf{U}_s(\mathbf{r}-\mathbf{b}_s)e^{i\mathbf{k}\cdot\mathbf{b}_s}, \quad \mathbf{U}_s = \{\tilde{h}(k|\mathbf{r}_s-\mathbf{r}'_s|), \mathbf{U}_s(\mathbf{r}'_s)\}, \quad (115)$$

where $\mathbf{r}_s = \mathbf{r}-\mathbf{b}_s$ and $\mathbf{r}'_s$ are an observation point and surface point respectively in the local coordinates of scatterer s. For $kr_s\sim\infty$,

$$\mathbf{U}_s \sim h(kr_s)\{\tilde{\varphi}(\mathbf{r}'_s;-\mathbf{o}), \mathbf{U}_s(\mathbf{r}'_s)\} \equiv h(kr_s)\mathbf{G}_s(\mathbf{o}), \quad (116)$$

where $\mathbf{G}_s$, the "multiple-scattered amplitude" of scatterer s, reduces to the single-scattered function $\mathbf{g}_s$ as the others recede to infinity. In terms of $\mathbf{G}_s$ the compound amplitude equals

$$\mathfrak{S}(\mathbf{o},\mathbf{i}:\epsilon) = \sum_s e^{ik(\mathbf{i}-\mathbf{o})\cdot\mathbf{b}_s}\mathbf{G}_s(\mathbf{o},[\mathbf{i}:\epsilon]), \quad (117)$$

where the brackets are to indicate that $\mathbf{i}:\epsilon$ plays a less complete role in $\mathbf{G}$ than in $\mathfrak{S}$ or $\mathbf{g}$.

[19] J. Yvon, *Actualités scientifiques et industrielles* (Hermann Cie., Paris, 1937), Nos. 542 and 543.
[20] P. Mazur and M. Mandel, Physica **22**, 289 (1956).
[21] M. Fixman, J. Chem. Phys. **23**, 2074 (1955).
[22] W. F. Brown, Jr., in *Handbuch der Physik* (Springer-Verlag, Berlin, 1956), Vol. 17.

Integral equations: Substituting $\tilde{h}$ of (26) into $\mathbf{U}_s$ of (115), and rewriting in terms of $\mathbf{G}_s$ of (116), we obtain

$$\mathbf{U}_s(\mathbf{r}_s) = \frac{1}{2\pi}\int e^{i\mathbf{k}\mathbf{p}\cdot\mathbf{r}_s}\mathbf{G}_s(\mathbf{p})\,d\Omega_p, \tag{118}$$

$$\mathfrak{U} = \sum e^{i\mathbf{k}\cdot\mathbf{b}_s}\int e^{i\mathbf{k}\mathbf{p}\cdot(\mathbf{r}-\mathbf{b}_s)}\mathbf{G}_s(\mathbf{p})\,d\Omega_p/2\pi. \tag{119}$$

Proceeding as in Ref. 1, we use $\mathbf{r}_s = \mathbf{r}_t + \mathbf{b}_t - \mathbf{b}_s \equiv \mathbf{r}_t + \mathbf{b}_{ts}$ to express $\boldsymbol{\varphi}$ and $\mathfrak{U}$ in the local coordinates of scatterer t, and write the total field referred to t as a set of plane waves plus one outgoing wave $\mathbf{U}_t$:

$$\boldsymbol{\Psi}(\mathbf{b}_t + \mathbf{r}_t) = e^{i\mathbf{k}\cdot\mathbf{b}_t}\left[e^{i\mathbf{k}\mathbf{i}\cdot\mathbf{r}_t}\boldsymbol{\epsilon} + \sum{}'\int e^{i\mathbf{k}(\mathbf{p}-i)\cdot\mathbf{b}_{ts}}e^{i\mathbf{k}\mathbf{p}\cdot\mathbf{r}_t}\mathbf{G}_s(\mathbf{p})\,d\Omega/2\pi + \mathbf{U}_t\right] = e^{i\mathbf{k}\cdot\mathbf{b}_t}[\boldsymbol{\Phi}_t + \mathbf{U}_t], \tag{120}$$

where $\sum'$ means sum over $s \neq t$ and where $\boldsymbol{\Phi}_t$ is the total excitation at t. Then knowing the response $(\mathbf{u})$ of the scatterer to one plane wave, we use the superposition principle to write

$$\mathbf{U}_t = \mathbf{u}_t(\mathbf{i}:\boldsymbol{\epsilon}) + \sum{}'\int e^{i\mathbf{k}(\mathbf{p}-\mathbf{i})\cdot\mathbf{b}_{ts}}\mathbf{u}_t(\mathbf{p}:\boldsymbol{\gamma}_s)\mathbf{G}_s(\mathbf{p})\,d\Omega/2\pi, \qquad \boldsymbol{\gamma}_s \equiv \mathbf{G}_s/G_s, \tag{121}$$

where $\boldsymbol{\gamma}_s$ is the polarization of $\mathbf{G}_s$.

The asymptotic form of (121) for $\mathbf{r}_t \to \infty$ gives a "self-consistent" system of integral equations for the multiple-scattering amplitude:

$$\mathbf{G}_t(\mathbf{o}) = \mathbf{g}_t(\mathbf{o}, \mathbf{i}:\boldsymbol{\epsilon}) + \sum{}'\int e^{i\mathbf{k}(\mathbf{p}-\mathbf{i})\cdot\mathbf{b}_t}\mathbf{g}_t(\mathbf{o}, \mathbf{p}:\boldsymbol{\gamma}_s)G_s(\mathbf{p})\,d\Omega_p/2\pi, \tag{122}$$

where in general $\mathbf{g}(\mathbf{o}, \mathbf{i}:\boldsymbol{\epsilon})$ and $\mathbf{g}(\mathbf{o}, \mathbf{p}:\boldsymbol{\gamma}_s)$ are not parallel. Forming $\mathbf{e}\cdot\mathbf{G}_t$, and using the reciprocity relation (20) to replace $\mathbf{e}\cdot\mathbf{g}_t(\mathbf{o}, \mathbf{p}:\boldsymbol{\gamma}_s)$ by $\boldsymbol{\gamma}_s\cdot\mathbf{g}_t(-\mathbf{p},-\mathbf{o}:\mathbf{e})$ we see from the definition (116) for $\mathbf{G}_s$ that the integral converges if $\operatorname{Im}\mathbf{p}\cdot(\mathbf{b}_{ts} + \mathbf{r}'_t - \mathbf{r}'_s) > 0$. In terms of $\mathbf{b}_{ts} = b_{ts}\hat{b}_{ts}$, we require $b_{ts} > [(\mathbf{r}'_t + \mathbf{r}'_s)\cdot\hat{b}_{ts}]_{\max}$, i.e., that the sum of the scatterer's projections on $\mathbf{b}_{ts}$ do not overlap.

The integral equation (122) is essentially a "reciprocity relation" between G and g. This follows on applying Green's theorem (10) to ψ_1 and Ψ_2, with ψ_1 as the solution for $\boldsymbol{\varphi}_1$ incident on an isolated scatterer t, and Ψ_2 as the solution for $\boldsymbol{\varphi}_2$ incident on a collection of scatterers which includes t; ψ_1 and Ψ_2 satisfy the same conditions at t's surface and the same wave equation in its interior. Consequently, essentially as for (18), we obtain

$$0 = \{\psi_1, \Psi_2\}_t = \{(\boldsymbol{\varphi}_1 + \mathbf{u}_{t1}), (\boldsymbol{\varphi}_2 + \sum{}'\mathbf{U}_{s2}e^{i\mathbf{k}\cdot\mathbf{b}_{ts}} + \mathbf{U}_{t2})\}_t, \tag{123}$$

where the subscript t indicates integration is over a surface that isolates scatterer t from the others. We have $\{\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2\} = 0$ as previously; similarly $\{\boldsymbol{\varphi}_1, \mathbf{U}_{s2}\}_t = 0$ since $\mathbf{U}_{s2}$ has no singularities inside the surface that isolates t; finally $\{\mathbf{u}_{t1}, \mathbf{U}_{t2}\}_t = \{\mathbf{u}_{t1}, \mathbf{U}_{t2}\}_\infty = 0$ follows from the asymptotic forms (5) and (116). Consequently (122) reduces to

$$\{\boldsymbol{\varphi}_1, \mathbf{U}_{t2}\} = -\{\mathbf{u}_{t1}, \boldsymbol{\varphi}_2\} - \{\mathbf{u}_{t1}, \sum{}'\mathbf{U}_{s2}e^{-i\mathbf{k}\cdot\mathbf{b}_{ts}}\}. \tag{124}$$

Using the definitions of $\mathbf{G}$ and $\mathbf{g}$ as in (116) and (16), and proceeding as for (20), we reduce (124) to

$$\boldsymbol{\epsilon}_1\cdot\mathbf{G}_t(-\mathbf{i}_1, \mathbf{i}_2:\boldsymbol{\epsilon}_2) = \boldsymbol{\epsilon}_2\cdot\mathbf{g}_t(-\mathbf{i}_2, \mathbf{i}_1:\boldsymbol{\epsilon}_1) + \{\sum{}'\mathbf{U}_{s2}e^{-i\mathbf{k}\cdot\mathbf{b}_{ts}}, \mathbf{u}_{t1}\}. \tag{125}$$

Introducing the plane wave representation (118) for $\mathbf{U}_{s2}$ and the definition of $\mathbf{g}$ in the kernel, gives

$$\boldsymbol{\epsilon}_1\cdot\mathbf{G}_t(-\mathbf{i}_1) = \boldsymbol{\epsilon}_2\cdot\mathbf{g}_t(-\mathbf{i}_2, \mathbf{i}_1:\boldsymbol{\epsilon}_1) + \sum{}'\int e^{i\mathbf{k}(\mathbf{p}-\mathbf{r})\cdot\mathbf{b}_{ts}}\mathbf{g}_t(-\mathbf{p}, \mathbf{i}_1:\boldsymbol{\epsilon}_2)\cdot\mathbf{G}_s(\mathbf{p})\,d\Omega_p/2\pi. \tag{126}$$

Applying (20) to $\mathbf{g}_t$, and replacing $-\mathbf{i}_1$ by $\mathbf{o}$ we reduce (126) to $\boldsymbol{\epsilon}_1\cdot(122)$.

Equation (122) is a mixed vector–scalar form. The analogous mixed vector–dyadic form is obtained by introducing the dyadic isolated-scattering amplitude $\tilde{g}$ of (66). Thus since $\mathbf{g}(\mathbf{o}, \mathbf{i}:\boldsymbol{\epsilon}) = \tilde{g}(\mathbf{o}, \mathbf{i})\cdot\boldsymbol{\epsilon}$, and $\mathbf{g}(\mathbf{o}, \mathbf{p}, \boldsymbol{\gamma}_s) = \tilde{g}(\mathbf{o}, \mathbf{p})\cdot\boldsymbol{\gamma}_s$, we may rewrite (122) as

$$\mathbf{G}_t(\mathbf{o}) = \tilde{g}_t(\mathbf{o}, \mathbf{i})\cdot\boldsymbol{\epsilon} + \sum{}'\int e^{i\mathbf{k}(\mathbf{p}-\mathbf{i})\cdot\mathbf{b}_{ts}}\tilde{g}_t(\mathbf{o}, \mathbf{p})\cdot\mathbf{G}_s(\mathbf{p})\,d\Omega_p/2\pi. \tag{127}$$

Similarly, we obtain a complete dyadic representation by introducing a multiple-scattered dyadic amplitude $\tilde{G}$, such that

$$\mathbf{G}(\mathbf{o}) = \tilde{G}(\mathbf{o})\cdot\boldsymbol{\epsilon}, \tag{128}$$

and dropping $\boldsymbol{\epsilon}$:

$$\tilde{G}(\mathbf{o}) = \tilde{g}_t(\mathbf{o}, \mathbf{i}) + \sum{}'\int e^{i\mathbf{k}(\mathbf{p}-\mathbf{i})\cdot\mathbf{b}_{ts}}\tilde{g}_t(\mathbf{o}, \mathbf{p})\cdot\tilde{G}_s(\mathbf{p})\,d\Omega/2\pi, \tag{129}$$

which is the complete analog of (3:34). Alternatively from $\{\tilde{\psi}_1, \Psi_2\}_t = 0$ we obtain (124d), i.e.,

$$\{\tilde{\varphi}_1, \tilde{U}_{t2}\} = -\{\tilde{u}_{t1}, \tilde{\varphi}_2\} - \{\tilde{u}_{t1}, \sum{}'\tilde{U}_{s2}e^{-i\mathbf{k}\cdot\mathbf{b}_{ts}}\} = -\{\tilde{u}_{t1}, \tilde{\varphi}_2\} - \sum{}'\int e^{i\mathbf{k}(\mathbf{p}-\mathbf{i})\cdot\mathbf{b}_{ts}}\{\tilde{u}_{t1}, \tilde{\varphi}(\mathbf{r}_t;\mathbf{p})\}\cdot\tilde{G}_s(\mathbf{p})\,d\Omega/2\pi, \tag{130}$$

where the last form followed from (118d). From (116d) and (73) and (74), we reduce (130) to

$$\tilde{G}(-\mathbf{i}_1) = \tilde{g}^{\mathrm{T}}(-\mathbf{i}_2, \mathbf{i}_1) + \sum{}'\int e^{i\mathbf{k}(\mathbf{p}-\mathbf{i})\cdot\mathbf{b}_{ts}}\tilde{g}^{\mathrm{T}}(-\mathbf{p}, \mathbf{i}_1)\cdot\tilde{G}_s(\mathbf{p})\,d\Omega/2\pi, \tag{131}$$

from which we obtain (129) by using the reciprocity relation (75) to convert $\tilde{g}^{\mathrm{T}}$ to $\tilde{g}(-\mathbf{i}_1, \mathbf{i}_2)$ and $\tilde{g}(-\mathbf{i}_1, p)$, and then replacing $-\mathbf{i}_1$ by $\mathbf{o}$. See Appendix A for additional relations and discussion of reciprocity.

Large spacings: We obtain forms of (129) convenient for large k $|\mathbf{b}_t - \mathbf{b}_s| = kb_{ts}$ by applying (55):

$$\tilde{G}_t(\mathbf{o}, \mathbf{i}) = \tilde{g}_t(\mathbf{o}, \mathbf{i}) + \sum_s{}' \tilde{\mathcal{F}}_{ts} \cdot \tilde{g}_t(\mathbf{o}, \hat{b}_{ts}) \cdot \tilde{G}_s(\hat{b}_{ts}, \mathbf{i}),$$
$$\tilde{\mathcal{F}}_{ts} = h(kb_{ts})e^{-i\mathbf{k}\cdot\mathbf{b}_{ts}}\tilde{\mathfrak{D}}_{ts} \equiv \tilde{\mathcal{H}}_{ts}(b^{-1}) + \tilde{\mathcal{M}}_{ts}(b^{-2}) + \tilde{\mathcal{N}}_{ts}(b^{-3}) + \cdots, \tag{132}$$

where $\tilde{\mathfrak{D}}$ is given terms of $\tilde{D}$ in (50), and the present subscripts indicate that the differentiations of (44) in $\tilde{D}$ are to be performed with respect to the angles associated with the unit vector $\hat{b}_{ts}$. We introduced the additional factor $\mathbf{i}$ in the argument of $\tilde{G}$ to facilitate iteration. If we keep only the leading term of $\tilde{\mathfrak{D}}$ (i.e., $\tilde{I}$), then

$$\tilde{G}_t(\mathbf{o}, \mathbf{i}) \sim \tilde{g}_t(\mathbf{o}, \mathbf{i}) + \sum_s{}' \frac{e^{ikb_{ts} - i\mathbf{k}\cdot\mathbf{b}_{ts}}}{ikb_{ts}} \tilde{g}_t(\mathbf{o}, \hat{b}_{ts}) \cdot \tilde{G}_s(\hat{b}_{ts}, \mathbf{i}); \tag{133}$$

if we dot-multiply from the right by $\boldsymbol{\epsilon}$ we have the system of equations discussed by Saxon[7] (pp. 92–99). (The analogous equations for the scalar problem, and the iterated orders-of-scattering form are discussed by Karp, and by Twersky in the papers cited in the survey, Ref. 5.)

The leading term of (132) is the single-scattered value, or equivalently the "first-order" of scattering $\tilde{g}_t(\mathbf{o}, \mathbf{i})$. Iterating (132) starting with $\tilde{g}_t(\mathbf{o}, \mathbf{i})$ yields a series in inverse powers of kb_{ts} which involves $\tilde{g}$ and its derivatives. Thus the $(kb)^{-1}$ term [either of (132) or (133)] is the far-field multiple scattering form of the second order of scattering:

$$\sum_s{}' \tilde{\mathcal{H}}_{ts} \cdot \tilde{g}_t(\mathbf{o}, \hat{b}_{ts}) \cdot \tilde{g}_s(\hat{b}_{ts}, \mathbf{i}), \quad \tilde{\mathcal{H}}_{ts} \equiv h(kb_{ts})e^{-i\mathbf{k}\cdot\mathbf{b}_{ts}}\tilde{I},$$

i.e., the dyadic analog of (3:37). Terms to order $(kb)^{-2}$ are given by

$$\sum_s{}' \tilde{\mathcal{H}}_{ts} \cdot \tilde{g}_t(\mathbf{o}, \hat{b}_{ts}) \cdot \sum_p{}' \tilde{\mathcal{H}}_{sp} \cdot \tilde{g}_s(\hat{b}_{ts}, \hat{b}_{sp}) \cdot \tilde{g}_p(\hat{b}_{sp}, \mathbf{i}) + \sum_s{}' \tilde{\mathcal{M}}_{ts} \cdot \tilde{g}_t(\mathbf{o}, \hat{b}_{ts}) \cdot \tilde{g}_s(\hat{b}_{ts}, \mathbf{i}),$$
$$\tilde{\mathcal{M}}_{ts} \equiv (i/2kb_{ts})\tilde{\mathcal{H}}_{ts} \cdot \tilde{D}_{ts},$$

where the double sum corresponds to the third far-field order, and the single sum is the first "mid-field" correction to the second far-field order; this is the analog of (3:38). The next terms in the expansion of $\tilde{G}$, the terms or order $(kb)^{-3}$ are given by (3:39d), obtained from (39) of Ref. 3 by replacing g by $\tilde{g}$, the previous scalar operators $\mathcal{H}$ and $\mathcal{M}$ by the present dyadics, and the previous $\mathcal{N}$ by

$$\tilde{\mathcal{N}} = [(i/2kb_{ts})^2/2]\tilde{\mathcal{H}}_{ts} \cdot \tilde{D}_{ts} \cdot (\tilde{D}_{ts} - 2\tilde{I}).$$

Algebraic equations: If we substitute spherical harmonic representations for $\tilde{g}_t$ and $\tilde{G}_t$ in (129), i.e.,

$$\tilde{g}_t(\mathbf{o}, \mathbf{i}) = \sum [\mathbf{C}_n^m(\mathbf{o})\mathbf{c}_{nm}^t(\mathbf{i}) + \mathbf{B}_n^m(\mathbf{o})\mathbf{b}_{nm}^t(\mathbf{i})], \tag{134}$$
$$\tilde{G}_t(\mathbf{o}) = \sum [\mathbf{C}_n^m(\mathbf{o})\,\mathbf{C}_{nm}^t + \mathbf{B}_n^m(\mathbf{o})\mathfrak{B}_{nm}^t], \tag{135}$$

and use the orthogonality of the $\mathbf{C}$'s and $\mathbf{B}$'s we obtain

$$\mathbf{C}_{nm}^t = \mathbf{c}_{nm}^t(\mathbf{i}) + \sum_{}{}' \sum \int e^{ik(\mathbf{p}-\mathbf{i})\cdot\mathbf{b}_{ts}}\, \mathbf{c}_{nm}^t(\mathbf{p}) \cdot [\mathbf{C}_r^q(\mathbf{p})\mathbf{C}_{rq}^s + \mathbf{B}_r^q(\mathbf{p})\mathfrak{B}_{rq}^s]\, d\Omega_p/2\pi,$$
$$\mathfrak{B}_{nm}^t = \mathbf{b}_{nm}^t(\mathbf{i}) + \sum_{}{}' \sum \int e^{ik(\mathbf{p}-\mathbf{i})\cdot\mathbf{b}_{ts}}\mathbf{b}_{nm}^t(\mathbf{p}) \cdot [\mathbf{C}_r^q(\mathbf{p})\mathbf{C}_{rq}^s + \mathbf{B}_r^q(\mathbf{p})\mathfrak{B}_{rq}^s]\, d\Omega_p/2\pi. \tag{136}$$

If we expand the isolated scattering coefficients as series of spherical harmonics as in (87), then we may write

$$\mathbf{C}_{nm}^t = \mathbf{c}_{nm}^t(\mathbf{i}) + \sum_{}{}' \sum [\alpha_{nm\nu\mu}^t \mathbf{C}_{rq}^s \mathcal{E}(st; \nu\mu, rq) + \beta\mathcal{C}\mathcal{E}' - \alpha\mathcal{B}\mathcal{E}' + \beta\mathcal{B}\mathcal{E}],$$
$$\mathfrak{B}_{nm}^t = \mathbf{b}_{nm}^t(\mathbf{i}) + \sum_{}{}' \sum [\gamma\mathcal{C}\mathcal{E} + \delta\mathcal{C}\mathcal{E}' - \gamma\mathcal{B}\mathcal{E}' + \delta\mathcal{B}\mathcal{E}], \tag{137}$$

where the scheme for the indices is shown only once, and where

$$\mathcal{E}(st; \nu\mu, rq) = \frac{e^{-i\mathbf{k}\cdot\mathbf{b}_{ts}}}{2\pi}\int e^{ik\mathbf{p}\cdot\mathbf{b}_{ts}}\mathbf{C}_\nu^{-\mu}(\mathbf{p}) \cdot \mathbf{C}_r^q(\mathbf{p})\, d\Omega$$
$$= \mathcal{E}(\mathbf{C} \cdot \mathbf{C}) = \mathcal{E}(\mathbf{B} \cdot \mathbf{B}),$$
$$\mathcal{E}' = \mathcal{E}(\mathbf{B} \cdot \mathbf{C}) = -\mathcal{E}(\mathbf{C} \cdot \mathbf{B}). \tag{138}$$

Following the procedure used for the scalar case, we write $\mathbf{C} \cdot \mathbf{C}$ and $\mathbf{B} \cdot \mathbf{C}$ as sets of products of surface harmonics YY to reduce the present $\mathcal{E}$'s to sets of the E's of (3:42), and then use (3:43) to write $\mathcal{E}$ and $\mathcal{E}'$ in terms of h's and their derivatives times Y's. We illustrate this subsequently. (In the above, we have generated implicitly the addition theorems discussed by Stein.[16])

In particular, for spherically symmetric scatterers (137) reduces to

$$\mathbf{C}_{nm}^t = (-1)^m c_n^t\{\mathbf{C}_n^{-m}(\mathbf{i}) + \sum_{}{}' \sum [\mathbf{C}_{rq}^s \mathcal{E}(st; nm, rg) - \mathfrak{B}_{rq}^s \mathcal{E}']\},$$
$$\mathfrak{B}_{nm}^t = (-1)^m b_n^t\{\mathbf{B}_n^{-m}(\mathbf{i}) + \sum_{}{}' \sum [\mathbf{C}_{rq}^s \mathcal{E}' + \mathfrak{B}_{rq}^s \mathcal{E}]\}, \tag{139}$$

which we apply in detail to two scatterers in the next section.

4. TWO SCATTERERS

For two scatterers, we take the primary origin ($r = 0$) as the midpoint of the line joining the centers

of their circumscribed spheres. The centers are located at

$$\mathbf{b}_1 = \mathbf{b}(b, \tau, \beta) = b\mathbf{b}_+,$$
$$\mathbf{b}_2 = \mathbf{b}(b, \pi - \tau, \pi + \beta) = b\mathbf{b}_- = -\mathbf{b},$$

where b, τ, β are spherical coordinates; the local coordinates with respect to these centers are written as $\mathbf{r}_1 = \mathbf{r}_+$ and $\mathbf{r}_2 = \mathbf{r}_-$. For this case the scattered field reduces to

$$\tilde{\mathfrak{U}}(\mathbf{r}) = e^{i\delta}\tilde{U}_+(\mathbf{r}_+) + e^{-i\delta}\tilde{U}_-(\mathbf{r}_-),$$
$$\pm\delta = bk\mathbf{i}\cdot\mathbf{b}_\pm = \pm\mathbf{k}\cdot\mathbf{b}, \tag{140}$$

and the compound scattering amplitude equals

$$\mathfrak{G}(\mathbf{o}, \mathbf{i}) = e^{i(\delta-\Delta)}\tilde{G}_+(\mathbf{o}, i) + e^{-i(\delta-\Delta)}\tilde{G}_-(\mathbf{o}, \mathbf{i}),$$
$$\pm\Delta = kb\mathbf{o}\cdot\hat{b}_\pm = \pm k\mathbf{o}\cdot\mathbf{b}. \tag{141}$$

The plane wave representation yields

$$\tilde{U}_\pm = \frac{1}{2\pi}\int e^{ik\mathbf{p}\cdot\mathbf{r}_\pm}\tilde{G}_\pm(\mathbf{p})\, d\Omega_p, \tag{142}$$

where

$$\tilde{G}_\pm(\mathbf{o}) = \tilde{g}_\pm(\mathbf{o}, \mathbf{i}) + e^{\mp i2\delta}\int e^{i2kb\mathbf{b}\cdot\mathbf{p}\pm}\,\tilde{g}_\pm(\mathbf{o}, \mathbf{p})\cdot\tilde{G}_\pm(\mathbf{p})d\Omega_p/2\pi. \tag{143}$$

4.1. Inverse Separation Representation

For two scatterers, (132) reduces to

$$\tilde{G}_\pm(\mathbf{o}) = \tilde{g}_\pm(\mathbf{o}, \mathbf{i}) + \tilde{\mathcal{F}}_\pm\cdot\tilde{g}_\pm(\mathbf{o}, \mathbf{b}_\pm)\cdot\tilde{G}_\mp(\mathbf{b}_\pm),$$
$$\tilde{\mathcal{F}}_\pm = h(2kb)e^{\mp i2\delta}\mathfrak{D}_\pm, \tag{144}$$

where the subscripts on $\mathfrak{D}_\pm$ etc., indicate that the differentiations are to be performed with respect to the angles associated with the unit vectors $\mathbf{b}_\pm$. Replacing $\mathbf{o}$ by $\mathbf{b}_\pm$, we solve for

$$\tilde{G}_\mp(\mathbf{b}_\mp) = [\bar{I} - \tilde{\mathcal{F}}_\mp\cdot\tilde{g}_\mp(\mathbf{b}_\pm, \mathbf{b}_\mp)\cdot\tilde{\mathcal{F}}_\pm\cdot\tilde{g}_\pm(\mathbf{b}_\mp, \mathbf{b}_\pm)]^{-1}$$
$$\cdot[\tilde{g}_\mp(\mathbf{b}_\pm, \mathbf{i}) + \tilde{\mathcal{F}}_\mp\cdot\tilde{g}_\mp(\mathbf{b}_\pm, \mathbf{b}_\mp)\cdot\tilde{g}_\pm(\mathbf{b}_\mp, \mathbf{i})], \tag{145}$$

which when substituted into (144) gives a closed operational form for $\tilde{G}_\pm(\mathbf{o})$ in terms of the isolated scatterer functions $\tilde{g}_\pm$, i.e., the analog of (3:50). Since the inverse dyadic $[\bar{I} - \bar{X}]^{-1}$ equals $\bar{I} + \bar{X} + \bar{X}\cdot\bar{X} + \cdots$, we see that the expansion of the closed form in powers of b^{-1} yields the series (3:51d) to (3:54d), on replacing the previous scalars by our present functions.

4.2. Radially Symmetric Scatterers

For two spherically symmetric scatterers, we use isolated scattering amplitudes $\tilde{g}_\pm$ as in (134) in terms of $c_n^\pm$ and $b_n^\pm$, and $\tilde{G}_\pm$ as in (135) in terms of $\tilde{\mathbf{C}}_{nm}^\pm$ and $\mathfrak{B}_{nm}^\pm$. Specializing (139), we have

$$\mathbf{C}_{nm}^\pm = (-1)^m c_n^\pm\{\mathbf{C}_n^{-m}(\mathbf{i}) + \sum_{\nu\mu}[\mathbf{C}_{\nu\mu}^\mp\mathcal{E}_\pm(nm, \nu\mu) - \mathfrak{B}_{\nu\mu}^\mp\mathcal{E}'_\pm]\},$$
$$\mathfrak{B}_{nm}^\pm = (-1)^m b_n^\pm\{\mathbf{B}_n^{-m}(\mathbf{i}) + \sum_{\nu\mu}[\mathbf{C}_{\nu\mu}^\pm\mathcal{E}'_\pm + \mathfrak{B}_{\nu\mu}^\pm\mathcal{E}_\pm]\}, \tag{146}$$

where

$$\mathcal{E}_\pm(nm, \nu\mu) = e^{\mp i2\delta}\int e^{i2kb\hat{\mathbf{p}}\cdot\mathbf{b}_\pm}\mathbf{C}_n^{-m}(\mathbf{p})\cdot\mathbf{C}_\nu^\mu(\mathbf{p})\, d\Omega/2\pi \tag{147}$$

and similarly $\mathcal{E}'$ involves $\mathbf{B}_n^{-m}\cdot\mathbf{C}_\nu^\mu$.

To illustrate the above, we keep only the electric and magnetic dipole terms (b_1 and c_1), and suppress the arguments $2kb$ in h_n, and $\hat{b}_\pm$ in Y_n^m. We retain only the six equations of (146) involving $\mathbf{C}_{1m}$ and $\mathfrak{B}_{1m}$ for $m = 0, \pm1$. The integral $\mathcal{E}_\pm(10, 10)$ involves $\mathbf{C}_1^0\cdot\mathbf{C}_1^0 = \sin^2\tau = \frac{2}{3}(Y_0 - Y_2)$, and consequently, from (3:14), we have $\mathcal{E}_\pm(10, 10) = \frac{2}{3}(h_0 + Y_2h_2)$; similarly for the other integrals. Thus

$$\mathbf{C}_{10}^\pm/c_1^\pm = \mathbf{C}_1^0(\mathbf{i}) + e^{\mp i2\delta}[\tfrac{2}{3}(h_0 + Y_2h_2)\mathbf{C}_{10}^\mp + \tfrac{1}{3}Y_2^1h_2\mathbf{C}_{11}^\mp + Y_2^{-1}h_2\mathbf{C}_{1-1}^\mp - Y_1^1h_1\mathfrak{B}_{11}^\mp + Y_1^{-1}h_1\mathfrak{B}_{1-1}^\mp],$$
$$-\mathbf{C}_{11}^\pm/c_1^\pm = \mathbf{C}_1^{-1} + e^{\mp i2\delta}[Y_2^{-1}h_2\mathbf{C}_{10}^\mp - \tfrac{1}{3}(2h_0 - Y_2h_2)\mathbf{C}_{11}^\mp + 2Y_2^{-2}h_2\mathbf{C}_{1-1}^\mp - Y_1^{-1}h_1\mathfrak{B}_{10}^\mp - Y_1h_1\mathfrak{B}_{11}^\mp],$$
$$-\mathbf{C}_{1-1}^\pm/c_1^\pm = \mathbf{C}_1^1 + e^{\mp i2\delta}[\tfrac{1}{3}Y_2^1h_2\mathbf{C}_{10}^\mp + \tfrac{1}{3}Y_2^2h_2\mathbf{C}_{11}^\mp - \tfrac{1}{3}(2h_0 - Y_2h_2)\mathbf{C}_{1-1}^\mp + Y_1^1h_1\mathfrak{B}_{10}^\mp + Y_1h_1\mathfrak{B}_{1-1}^\mp \tag{148}$$

plus the analogous set for $\mathfrak{B}^\pm$ obtained by interchanging all forms of "B" and "C" in the above and replacing all Y_1^m by $-Y_1^m$.

If the axis ($\hat{b}$) of the pair of scatterers is taken along the polar axis ($\hat{z}$) [i.e., scatterers located at $z_\pm = \pm b$, $x = y = 0$] then $\tau = 0, \pi$ and $\beta = 0$. All Y's but $Y_2 = 1$ and $Y_1(\hat{b}_\pm) = \pm1$ vanish, and we may compress the remaining terms by using

$$\tfrac{2}{3}(h_0 + h_2) = 2h_1/\rho \equiv H,$$
$$\tfrac{1}{3}(2h_0 - h_2) = (\rho h_1)'/\rho \equiv \mathcal{K},$$
$$\rho = 2kb, \quad c_1 = c, \quad b_1 = b. \tag{149}$$

Thus

$$\mathbf{C}_{10}^\pm/c = \mathbf{C}_1^0(\mathbf{i}) + e^{\mp i2\delta}H\mathbf{C}_{10}^\mp,$$
$$\mathbf{C}_{11}^\pm/c = -\mathbf{C}_1^{-1}(\mathbf{i}) + e^{\mp i2\delta}(\mathcal{K}\mathbf{C}_{11}^\mp \pm h_1\mathfrak{B}_{11}^\mp),$$
$$\mathbf{C}_{1-1}^\pm/c = -\mathbf{C}_1^1(\mathbf{i}) + e^{\mp i2\delta}(\mathcal{K}\mathbf{C}_{1-1}^\mp \mp h_1\mathfrak{B}_{1-1}^\mp),$$
$$\mathfrak{B}_{10}^\pm/b = \mathbf{B}_1^0(\mathbf{i}) + e^{\mp i2\delta}H\mathfrak{B}_{10}^\mp,$$
$$\mathfrak{B}_{11}^\pm/b = -\mathbf{B}_1^{-1}(\mathbf{i}) + e^{\mp i2\delta}(\mathcal{K}\mathfrak{B}_{11}^\mp \pm h_1\mathbf{C}_{11}^\mp),$$
$$\mathfrak{B}_{1-1}^\pm/b = -\mathbf{B}_1^1(\mathbf{i}) + e^{\mp i2\delta}(\mathcal{K}\mathfrak{B}_{1-1}^\mp \pm h_1\mathbf{C}_{1-1}^\mp). \tag{150}$$

Since $\mathbf{C}_1^0 = \hat{\varphi} \sin\theta_i$ and $\mathbf{B}_1^0 = \hat{\theta} \sin\theta_i$ the components $\mathfrak{C}_{10}$ and $\mathfrak{B}_{10}$ vanish for incidence along the pair's axis ($\theta = 0$); for these "axial" components, there is no coupling between electric and magnetic moments. On the other hand, the "perpendicular" components $\mathfrak{B}_{1\pm1}$ and $\mathfrak{C}_{1\pm1}$ are coupled in general.

For the axial components, we iterate once and regroup terms to obtain

$$\mathfrak{C}_{10}^{\pm} = c^{\pm}\{\mathbf{C}_1^0(\mathbf{i}) + e^{\pm i2\delta}Hc^{\mp}[\mathbf{C}_1^0(\mathbf{i}) + e^{\mp i2\delta}H\mathfrak{C}_{10}^{\mp}]\}$$
$$= \frac{c^{\pm}[1 + c^{\mp}e^{\mp i2\delta}H]}{1 - c^{+}c^{-}H^2}\,\mathbf{C}_1^0(\mathbf{i}) \equiv c^{\pm}A(c^{\mp})\mathbf{C}_1^0(\mathbf{i}), \tag{151}$$

$$\mathfrak{B}_{10}^{\pm} = b^{\pm}A(b^{\mp})\mathbf{B}_1^0(\mathbf{i}), \tag{152}$$

where $A(b)$ is obtained on replacing c's by b's in $A(c)$.

For the perpendicular components, we first consider the cases when only the electric or only the magnetic effect exists.

Electric dipoles: If $c^{\pm} = 0$ the multiple-scattered field is fully specified by (152), and by the simplified form of the last two equations of (150),

$$\mathfrak{B}_{11}^{\pm} = -b^{\pm}[\mathbf{B}_1^{-1}(\mathbf{i}) - e^{\mp i2\delta}\mathcal{K}\mathfrak{B}_{11}^{\mp}], \tag{153}$$

plus the analogous equation for $\mathfrak{B}_{1-1}^{\pm}$ involving $\mathbf{B}_1^1(\mathbf{i})$. Thus

$$\mathfrak{B}_{11}^{\pm} = -b^{\pm}D(b^{\mp})\mathbf{B}_1^{-1}(\mathbf{i}), \quad \mathfrak{B}_{1-1}^{\pm} = -b^{\pm}D(b^{\mp})\mathbf{B}_1^{1}(\mathbf{i}),$$
$$D(b^{\mp}) \equiv (1 + b^{\mp}e^{\mp i2\delta}\mathcal{K})/(1 - b^{+}b^{-}\mathcal{K}^2), \tag{154}$$

where D differs from A of (151) only in that H is replaced by $\mathcal{K}$.

For this case,

$$\tilde{g}_{\pm}(\mathbf{o}, \mathbf{i}) = \tilde{B}(\mathbf{o}, \mathbf{i})b^{\pm} = (\tilde{B}^0 + \tilde{B}^1)b^{\pm}, \tag{155}$$

where the $\tilde{B}$'s are the corresponding $\tilde{B}_1$'s of (95). Similarly,

$$\tilde{G}^{\pm}(\mathbf{o}) = \tilde{G}^{\pm}(\mathbf{o}, \mathbf{i}) = b^{\pm}A^{\pm}(b^{\mp})\tilde{B}^0 + b^{\pm}D^{\pm}(b^{\mp})\tilde{B}^1. \tag{156}$$

Thus while each isolated scattering amplitude is an electric dipole determined essentially by the direction of incidence, the corresponding multiple-scattered amplitude is a sum of two uncoupled "compound dipoles": the compound axial term bA and perpendicular term bD are each the closed form of the corresponding geometrical progression of orders of scattering.

If the direction of incidence is along the dipole axis ($\mathbf{i} = \hat{z}$), then $\tilde{B}^0 = 0$, and if the incident polarization is $\boldsymbol{\epsilon} = \hat{x}$, we have

$$\tilde{B}^1 \cdot \hat{x} = (\hat{\theta}\cos\varphi\cos\theta - \hat{\varphi}\sin\varphi) \equiv \boldsymbol{\gamma}_1,$$
$$\tilde{g}(\mathbf{o}, \hat{z}) \cdot \hat{x} = \mathbf{g}(\mathbf{o}, \hat{z}\colon \hat{x}) = b^{\pm}\boldsymbol{\gamma}_1,$$
$$\tilde{G} \cdot \hat{x} = \mathbf{G} = b^{\pm}D(b^{\mp})\boldsymbol{\gamma}_1. \tag{157}$$

If the direction of incidence is perpendicular to the axes ($\mathbf{i} = \hat{x}$), and if the incident field is polarized parallel to the dipole's axis ($\boldsymbol{\epsilon} = \hat{z}$), then $\tilde{B}^1 \cdot \hat{z} = 0$ and

$$\tilde{B}^0 \cdot \hat{z} = -\hat{\theta}\sin\theta \equiv \boldsymbol{\gamma}_2, \quad \mathbf{g}(\mathbf{o}, \hat{x}\colon \hat{z}) = b_1^{\pm}\boldsymbol{\gamma}_2,$$
$$\mathbf{G} = b^{\pm}A(b^{\mp})\boldsymbol{\gamma}_2. \tag{158}$$

On the other hand, if the polarization is perpendicular to the axis ($\boldsymbol{\epsilon} = \hat{y}$) then $\tilde{B}^0 \cdot \hat{y} = 0$, and

$$\tilde{B}^1 \cdot \hat{y} = \hat{\varphi}\cos\varphi + \hat{\theta}\sin\varphi\cos\theta \equiv \boldsymbol{\gamma}_3,$$
$$\mathbf{g}(\mathbf{o}, \hat{x}\colon \hat{y}) = b^{\pm}\boldsymbol{\gamma}_3, \quad \mathbf{G} = b^{\pm}D(b^{\mp})\boldsymbol{\gamma}_3. \tag{159}$$

In all the above, the forwardscattered values of $\mathbf{g}$ and $\mathbf{G}$ have the same polarization as the incident wave. The same holds for arbitrary direction of incidence $\mathbf{i} = \mathbf{o}$ for which case we have $\tilde{B}^0 = \hat{\theta}\hat{\theta}\sin^2\theta$ and $\tilde{B}^1 = \hat{\theta}\hat{\theta}\cos^2\theta + \hat{y}\hat{y}$. If $\boldsymbol{\epsilon} = \hat{y}$, then $\mathbf{g} = b\boldsymbol{\epsilon}$ and $\mathbf{G} = bD\boldsymbol{\epsilon}$; similarly if $\boldsymbol{\epsilon} = \hat{\theta}$ (perpendicular to $\mathbf{i}$), then $\mathbf{g} = b\boldsymbol{\epsilon}(\sin^2\theta + \cos^2\theta) = b\boldsymbol{\epsilon}$, and

$$\mathbf{G} = b\boldsymbol{\epsilon}(A\sin^2\theta + D\cos^2\theta).$$

Although $\tilde{g}_+$ and $\tilde{g}_-$ satisfy (75), the theorem does not apply individually to the corresponding multiple-scattering functions $\tilde{G}_+$ and $\tilde{G}_-$ for the elements of the pair: the reciprocity relation applies only to the scattering amplitude for the configuration $\tilde{\mathfrak{G}}(\mathbf{o}, \mathbf{i})$ as in (141). From (141), (152), and (154), we write

$$\tilde{\mathfrak{G}} = \tilde{F}_+ + \tilde{F}_-,$$
$$\tilde{F}_+(\mathbf{o}, \mathbf{i}) = e^{i(\delta-\Delta)}\tilde{G}_+(\mathbf{o}, \mathbf{i}) = e^{ik\mathbf{b}\cdot(\mathbf{i}-\mathbf{o})}\tilde{K}_+ + e^{-ik\mathbf{b}\cdot(\mathbf{i}+\mathbf{o})}\tilde{K},$$
$$\tilde{F}_-(\mathbf{o}, \mathbf{i}) = e^{-i(\delta-\Delta)}\tilde{G}_-(\mathbf{o}, \mathbf{i}) = e^{-ik\mathbf{b}\cdot(\mathbf{i}-\mathbf{o})}\tilde{K}_- + e^{ik\mathbf{b}\cdot(\mathbf{i}+\mathbf{o})}\tilde{K};$$

$$\tilde{K}_+ = \frac{b_1^+\tilde{B}^0(\mathbf{o}, \mathbf{i})}{1 - b^+b^-H^2} - \frac{b_1^+\tilde{B}^1(\mathbf{o}, \mathbf{i})}{1 - b^+b^-\mathcal{K}^2}, \quad \tilde{K}_- = \tilde{K}_+\frac{b^-}{b^+};$$
$$\tilde{K} = \frac{b^+b^-H}{1 - b^+b^-H^2}\,\tilde{B}^0(\mathbf{o}, \mathbf{i}) - \frac{b^+b^-\mathcal{K}}{1 - b^+b^-\mathcal{K}^2}\,\tilde{B}^1(\mathbf{o}, \mathbf{i}). \tag{160}$$

From (160), we have

$$\tilde{F}_+^{\mathrm{T}}(-\mathbf{i}, -\mathbf{o}) = e^{ik\mathbf{b}\cdot(\mathbf{i}-\mathbf{o})}\tilde{K}_+ + e^{ik\mathbf{b}\cdot(\mathbf{i}+\mathbf{o})}\tilde{K} \neq \tilde{F}_+(\mathbf{o}, \mathbf{i}),$$
$$\tilde{F}_-^{\mathrm{T}}(-\mathbf{i}, -\mathbf{o}) = e^{-ik\mathbf{b}\cdot(\mathbf{i}-\mathbf{o})}\tilde{K}_- + e^{-ik\mathbf{b}\cdot(\mathbf{i}+\mathbf{o})}\tilde{K} \neq \tilde{F}_-(\mathbf{o}, \mathbf{i}),$$
$$\tilde{\mathfrak{G}}^{\mathrm{T}}(-\mathbf{i}, -\mathbf{o}) = \tilde{F}_+^{\mathrm{T}}(-\mathbf{i}, -\mathbf{o}) + \tilde{F}_-^{\mathrm{T}}(-\mathbf{i}, -\mathbf{o}) = \tilde{\mathfrak{G}}(\mathbf{o}, \mathbf{i}). \tag{161}$$

Thus, although the individual functions do not satisfy theorem (75) (because the phase of the $\tilde{K}$ term is not preserved) their sum does—and this is all that is required. See more general discussion in Appendix A.

In the forwardscattered direction,

$$\tilde{\mathfrak{G}}(\mathbf{i}, \mathbf{i}) = \tilde{G}_+(\mathbf{i}, \mathbf{i}) + \tilde{G}_-(\mathbf{i}, \mathbf{i})$$
$$= \tilde{K}_+ + \tilde{K}_- + 2\tilde{K}\cos(2k\mathbf{b}\cdot\mathbf{i}), \tag{162}$$

where we may use $\tilde{B}^0 = \epsilon_1\epsilon_1 \sin^2\theta$ and $\tilde{B}^1 = \epsilon_1\epsilon_1 \cos^2\theta + \epsilon_2\epsilon_2$. From theorem (80) the total cross section for a pair of scatterers equals

$$\begin{aligned}\mathfrak{Q}(\mathbf{i}:\boldsymbol{\epsilon}) &= -(4\pi/k^2)\,\mathrm{Re}\,(\boldsymbol{\epsilon}\cdot\tilde{\mathfrak{S}}\cdot\boldsymbol{\epsilon}) \\ &= -(4\pi/k^2)2\,\mathrm{Re}\,[\boldsymbol{\epsilon}\cdot(\tilde{K}_+ + \tilde{K}_- + 2\tilde{K}\cos 2\delta)\cdot\boldsymbol{\epsilon}].\end{aligned} \tag{163}$$

For the special case of a pair of identical scatterers ($b^+ = b^-$) under symmetrical excitation (i.e., $\delta = k\mathbf{b}\cdot\mathbf{i} = 2n\pi$; $n = 0, \pm1, \cdots$, with $n = 0$ corresponding to incidence perpendicular to the axis) we have

$$\begin{aligned}&\tilde{G}_- = \tilde{G}_+ = \tilde{G} = b(A\tilde{B}^0 + D\tilde{B}^1), \\ &A = 1/(1 - bH), \quad D = 1/(1 - b\mathcal{K}),\end{aligned} \tag{164}$$

and the total cross section follows from

$$\begin{aligned}\mathrm{Re}\,\tilde{\mathfrak{S}}(\mathbf{i},\mathbf{i}) &= 2\,\mathrm{Re}\,\tilde{G}(\mathbf{i},\mathbf{i}) \\ &= \tilde{B}^0 2\,\mathrm{Re}\,bA + \tilde{B}^1 2\,\mathrm{Re}\,bD.\end{aligned} \tag{165}$$

For lossless scatterers,

$$\mathrm{Re}\,bA = \frac{\mathrm{Re}\,b(1 - bH)^*}{|1 - bH|^2} = \frac{\mathrm{Re}\,b - |b|^2 J}{|1 - bH|^2},\ J = \frac{2j_1}{\rho},$$

and since the theorem for an isolated lossless dipole gives $-\mathrm{Re}\,b = \frac{2}{3}|b|^2$, we have

$$\mathrm{Re}\,bA = \frac{-\frac{2}{3}|b|^2(1 + \frac{3}{2}J)}{|1 - bH|^2} = -\tfrac{2}{3}(1 + \tfrac{3}{2}J)\,|bA|^2.$$

Similarly for $\mathrm{Re}\,Db$ we replace H, J by $\mathcal{K}$, $\mathcal{J}$ with $\mathcal{J} = \partial_\rho[\rho j_1]/\rho$. Thus

$$\begin{aligned}\mathrm{Re}\,\tilde{\mathfrak{S}}(\mathbf{i},\mathbf{i}) = &-\tfrac{4}{3}(1 + \tfrac{3}{2}\mathcal{J})\,|bA|^2\,\tilde{B}^0 \\ &+ \tfrac{4}{3}(1 + \tfrac{3}{2}\mathcal{J})\,|bD|^2\,\tilde{B}^1,\end{aligned} \tag{166}$$

from which we obtain $\mathfrak{Q}$ by dot multiplication as in (163).

For $\delta = 2n\pi$, it is simple to demonstrate that $\tilde{\mathfrak{S}}(\mathbf{i}_1, \mathbf{i}_2)$ satisfies the general theorem (79). For the present case $\tilde{\mathfrak{S}}^T(\mathbf{i}_2, \mathbf{i}_1) = \tilde{\mathfrak{S}}(\mathbf{i}_1, \mathbf{i}_2)$, and consequently, (79) reduces to

$$-\mathrm{Re}\,\tilde{\mathfrak{S}}(\mathbf{i}_1, \mathbf{i}_2) = \frac{1}{4\pi}\int \tilde{\mathfrak{S}}(\mathbf{i}_1, \mathbf{o})\cdot\tilde{\mathfrak{S}}^*(\mathbf{o}, \mathbf{i}_2)\,d\Omega_o;$$

equivalently,

$$\begin{aligned}&-2\,\mathrm{Re}\,\tilde{G}(\mathbf{i}_1, \mathbf{i}_2) \\ &\quad = \frac{1}{4\pi}\int \tilde{G}(\mathbf{i}_1, \mathbf{o})\cdot\tilde{G}^*(\mathbf{o}, \mathbf{i}_2)\,|e^{i\Delta} + e^{-i\Delta}|^2\,d\Omega_o,\end{aligned} \tag{167}$$

where $\Delta = kb\hat{b}\cdot\delta = kb\cos\theta$. Since $\tilde{G}^*$ is obtained by replacing bA and bD by their complex conjugates, we may show directly by proceeding as for (148) that (167) is satisfied. For example, in the right-hand side $|bD|^2\,\mathbf{B}_1^{-1}(\mathbf{i}_1)\mathbf{B}_1^1(\mathbf{i}_2)$ is multiplied by

$$\begin{aligned}&\int \mathbf{B}_1^1(\mathbf{o})\cdot\mathbf{B}_1^{-1}(\mathbf{o})[2 + e^{i2\Delta} + e^{-i2\Delta}]\,d\Omega_o \\ &\quad = 2\int \mathbf{B}_1^1\cdot\mathbf{B}_1^{-1}\,d\Omega - \int \frac{(2 + Y_2)}{3}(e^{i2\Delta} + e^{-i2\Delta})\,d\Omega;\end{aligned}$$

the first term gives $-\frac{16}{3}\pi$, and the second gives $-\frac{8}{3}\pi(2j_0 - j_2) = -8\pi\mathcal{J}$. Since $-\frac{4}{3}[1 + 3\mathcal{J}/2]\,|bD|^2 = 2\,\mathrm{Re}\,bD$ [from (165) and (166)], etc., we see that both sides of (167) yield identical terms.

Magnetic dipoles: Similarly if $b_1^\pm = 0$ in (150), we use (151) and the second and third equations of (150):

$$\mathbf{C}_{11}^\pm = -c^\pm[\mathbf{C}_1^{-1}(\mathbf{i}) - e^{\mp i2\delta}\mathbf{C}_{11}^\pm\mathcal{K}], \tag{168}$$

plus the analogous equation for $\mathbf{C}_{1-1}^\pm$ involving $\mathbf{C}_1^1(\mathbf{i})$. Thus, as previously,

$$\mathbf{C}_{11}^\pm = -c^\pm D(c^\mp)\mathbf{C}_1^{-1}(\mathbf{i}), \quad \mathbf{C}_{1-1}^\pm = -c^\pm D(c^\mp)\mathbf{C}_1^1(\mathbf{i}). \tag{169}$$

The single scattered amplitude is

$$\tilde{g}_\pm = C^\pm(\tilde{C}^0 + \tilde{C}^1), \tag{170}$$

with the $\tilde{C}$'s as in (95), and the corresponding multiple scattered values are

$$\tilde{G}^\mp(\mathbf{o}, \mathbf{i}) = c^\mp A(c^\pm)\tilde{C}^0 + c^\mp D(c^\pm)\tilde{C}^1. \tag{171}$$

The present case is completely analogous to the previous and corresponding results may be obtained by inspection.

For axial incidence $\mathbf{i} = \hat{z}$ we have $\tilde{C}^0 = 0$, and if the incident $\mathbf{E}$ is polarized parallel to $\hat{x}$, then corresponding to (157), we have

$$\begin{aligned}&\tilde{C}^1\cdot\hat{x} = \hat{\theta}\cos\varphi - \hat{\varphi}\sin\varphi\cos\theta \equiv \boldsymbol{\gamma}_1^1, \\ &\tilde{g}(\mathbf{o}, \hat{z})\cdot\hat{x} = \mathbf{g}(\mathbf{o}, \hat{z}:\hat{x}) = c^\mp\boldsymbol{\gamma}_1^1, \quad \mathbf{G} = c^\mp D(c^\mp)\boldsymbol{\gamma}_1^1.\end{aligned} \tag{172}$$

For normal incidence $\mathbf{i} = \hat{x}$, if the incident $\mathbf{E}$ is along the axis ($\hat{\boldsymbol{\epsilon}} = \hat{z}$), then $\tilde{C}^0\cdot\hat{z} = 0$ and corresponding to (158),

$$\begin{aligned}&\tilde{C}^1\cdot\hat{z} \equiv \boldsymbol{\gamma}_2^1 = -\boldsymbol{\gamma}_1^1, \quad \mathbf{g}(\mathbf{o}, \hat{x}:\hat{z}) = c^\pm\boldsymbol{\gamma}_2^1, \\ &\qquad \mathbf{G} = c^\pm D(c^\mp)\boldsymbol{\gamma}_2^1.\end{aligned} \tag{173}$$

If the polarization is along $\hat{y}$, then $\tilde{C}^1\cdot\hat{y} = 0$, and corresponding to (159),

$$\begin{aligned}&\tilde{C}^0\cdot\hat{y} = \hat{\varphi}\sin\theta = \boldsymbol{\gamma}_3^1, \quad \mathbf{g}(\mathbf{o}, \hat{x}:\hat{y}) = c^\pm\boldsymbol{\gamma}_3^1, \\ &\qquad \mathbf{G} = c^\pm A(c^\mp)\boldsymbol{\gamma}_3^1.\end{aligned} \tag{174}$$

One electric plus one magnetic dipole: The remaining elementary situation in (150) is that in which one scatterer (+) is an electric dipole and the other (−) is a magnetic dipole. For this case we set $b^- = c^+ = 0$ so that the required functions in (151) and (152) reduce

to the single-scattered values

$$\mathfrak{C}_{10}^{-} = c^{-}\mathbf{C}_1(\mathbf{i}), \quad \mathfrak{B}_{10}^{+} = b^{+}\mathbf{B}_1^0(\mathbf{i}), \tag{175}$$

which correspond to the first and fourth equations of (150). The second and fifth of (150) reduce to

$$\begin{aligned}\mathfrak{C}_{11}^{-} &= -c^{-}[\mathbf{C}_1^{-1} + e^{i2\delta}\mathfrak{B}_{11}^{+}h_1],\\ \mathfrak{B}_{11}^{+} &= -b^{+}[\mathbf{B}_1^{-1} + e^{-i2\delta}\mathfrak{C}_{11}^{-}h_1],\end{aligned} \tag{176}$$

plus the analogous set for $\mathfrak{C}_{1-1}^{-}$ and $\mathfrak{B}_{1-1}^{+}$ in terms of $\mathbf{C}_1^1$ and $\mathbf{B}_1^1$ with h_1 replaced by $-h_1$.

Solving (176) and its analog we obtain

$$\begin{aligned}\mathfrak{C}_{11}^{-} &= -c^{-}E\mathbf{C}_1^{-1}(\mathbf{i}) + e^{i2\delta}F\mathbf{B}_1^{-1}(\mathbf{i}),\\ \mathfrak{C}_{1-1}^{-} &= -c^{-}E\mathbf{C}_1^{1}(\mathbf{i}) - e^{i2\delta}F\mathbf{B}_1^{1}(\mathbf{i}),\\ \mathfrak{B}_{11}^{+} &= -b^{+}E\mathbf{B}_1^{-1}(\mathbf{i}) + e^{-i2\delta}F\mathbf{C}_1^{-1}(\mathbf{i}),\\ \mathfrak{B}_{1-1}^{+} &= -b^{+}E\mathbf{B}_1^{1}(\mathbf{i}) - e^{-i2\delta}F\mathbf{C}_1^{1}(\mathbf{i}),\\ E &\equiv 1/(1 - b^{+}c^{-}h_1^2), \quad F \equiv c^{-}b^{+}h_1E.\end{aligned} \tag{177}$$

The single-scattered amplitudes for this case are

$$\tilde{g}_{+} = b^{+}(\tilde{B}^0 + \tilde{B}^1), \quad \tilde{g}_{-} = c^{-}(\tilde{C}^0 + \tilde{C}^1), \tag{178}$$

and the corresponding multiple-scattered amplitudes equal

$$\begin{aligned}\tilde{G}_{+} &= b^{+}\tilde{B}^0 + b^{+}E\tilde{B}^1 + e^{-i2\delta}F\tilde{D},\\ \tilde{G}^{-} &= c^{-}\tilde{C}^0 + c^{-}E\tilde{C}^1 + e^{i2\delta}F\tilde{D}^1,\\ \tilde{D} &= \mathbf{B}_1^1(\mathbf{o})\mathbf{C}_1^{-1}(\mathbf{i}) - \mathbf{B}_1^{-1}(\mathbf{o})\mathbf{C}_1^1(\mathbf{i}) = i\,\mathrm{Im}\,\mathbf{B}_1^1(\mathbf{o})\mathbf{C}_1^{1*}(\mathbf{i})\\ &= i(\hat{\theta}\hat{\varphi}_i\cos\theta\cos\theta_i - \hat{\varphi}\hat{\theta}_i)\sin(\varphi - \varphi_i)\\ &\quad + i(\hat{\theta}\hat{\theta}_i\cos\theta + \hat{\varphi}\hat{\varphi}_i\cos\theta_i)\cos(\varphi - \varphi_i),\\ \tilde{D}^1 &= \mathbf{C}_1^1(\mathbf{o})\mathbf{B}_1^{-1}(i) - \mathbf{C}_1^{-1}(\mathbf{o})\mathbf{B}_1^1(\mathbf{i}) = -i\,\mathrm{Im}\,\mathbf{C}_1^1(\mathbf{o})\mathbf{B}_1^{1*}(\mathbf{i})\\ &= -i(\hat{\theta}\hat{\varphi} - \hat{\varphi}\hat{\theta}_i\cos\theta\cos\theta_i)\sin(\varphi - \varphi_i)\\ &\quad - i(\hat{\theta}\hat{\theta}_i\cos\theta_i + \hat{\varphi}\hat{\varphi}_i\cos\theta)\cos(\varphi - \varphi_i).\end{aligned} \tag{179}$$

The present case is much less symmetrical than the preceding ones, and provides a simple illustration of a scatterer containing cross terms [e.g., $\mathbf{C}_1^1(\mathbf{o})\mathbf{B}_1^{-1}(\mathbf{i})$] corresponding to coupling between electric and magnetic dipoles.

For axial incidence $\mathbf{i} = \hat{z}$, we have $\tilde{B}^0 = \tilde{C}^0 = 0$. If $\boldsymbol{\epsilon} = \hat{x}$ then

$$\begin{aligned}\tilde{D}\cdot\hat{x} &= i\gamma_1, \quad \tilde{D}^1\cdot\hat{x} = -i\hat{\gamma}_1^1,\\ \tilde{G}_{+}\cdot\hat{x} &= (b^{+}E + e^{-i2\delta}Fi)\gamma_1,\\ \tilde{G}_{-}\cdot\hat{x} &= (c^{-}E - e^{i2\delta}Fi)\gamma_1^1.\end{aligned} \tag{180}$$

In the forward direction,

$$\mathfrak{S} = \mathbf{G}_{+} + \mathbf{G}_{-} = (b^{+} + c^{-})E\hat{x} + 2\sin(2kb)F\hat{x}. \tag{181}$$

For normal incidence $\mathbf{i} = \hat{x}$, if $\boldsymbol{\epsilon} = \hat{z}$ then $\tilde{B}^1\cdot\hat{z} = \tilde{C}^0\cdot\hat{z} = \tilde{D}^1\cdot\hat{z} = 0$ and

$$\begin{aligned}\tilde{D}\cdot\hat{z} &= -i\gamma_1, \quad \mathbf{G}_{+} = b^{+}\gamma_2 - ie^{-i2\delta}F\gamma_1,\\ \mathbf{G}_{-} &= c^{-}E\gamma_2^1.\end{aligned} \tag{182}$$

In the forward direction,

$$\begin{aligned}\mathbf{G}_{+} &= b^{+}\hat{z}, \quad \mathbf{G}_{-} = c^{-}E\hat{z},\\ \mathfrak{S} &= \mathbf{G}_{+} + \mathbf{G}_{-} = (b^{+} + c^{-}E)\hat{z}.\end{aligned} \tag{183}$$

If $\boldsymbol{\epsilon} = \hat{y}$ then $\tilde{B}^0\cdot\hat{y} = \tilde{C}^1\cdot\hat{y} = \tilde{D}\cdot\hat{y} = 0$ and

$$\begin{aligned}\tilde{D}^1\cdot\hat{y} &= -i(\theta\sin\varphi + \varphi\cos\theta\cos\varphi) = -i\gamma_4,\\ \mathbf{G}_{+} &= b^{+}E\gamma_3, \quad \mathbf{G}_{-} = c^{-}\gamma_3^1 - i\gamma_4e^{i2\delta}F.\end{aligned} \tag{184}$$

In the forward direction

$$\mathfrak{S} = \mathbf{G}_{+} + \mathbf{G}_{-} = (b^{+}E + c^{-})\hat{y}. \tag{185}$$

More generally in the forward direction we write

$$\mathbf{g}_{+} = b^{+}\boldsymbol{\epsilon}, \quad \mathbf{g}_{-} = c^{-}\boldsymbol{\epsilon}, \quad \mathbf{G}_{\pm} = G_{\pm}\boldsymbol{\epsilon}. \tag{186}$$

If $\boldsymbol{\epsilon} = \hat{y}$ then

$$\begin{aligned}G_{+} &= b^{+}E + e^{-i2\delta}Fi\cos\theta,\\ G_{-} &= c^{-}\sin^2\theta + c^{-}E\cos^2\theta - e^{i2\delta}Fi\cos\theta.\end{aligned} \tag{187}$$

Similarly, if $\boldsymbol{\epsilon} = \hat{\theta}$ then

$$\begin{aligned}G_{+} &= b^{+}\sin^2\theta + b^{+}E\cos^2\theta + e^{-i2\delta}iF\cos\theta,\\ G_{-} &= c^{-}E - e^{i2\delta}iF\cos\theta.\end{aligned} \tag{188}$$

Thus, in all cases, the forwardscattered values have the incident polarization.

Electric plus magnetic dipoles: For the general situation of (150), each scatterer has both electric and magnetic dipole moments. The axial components are as in (151) and the corresponding perpendicular components follow from the remaining four equations of (150). Thus eliminating $\mathfrak{C}_{11}^{\mp}$ and $\mathfrak{B}_{11}^{\mp}$ from the second and fifth, we obtain

$$\begin{aligned}\mathfrak{C}_{11}^{\pm} &= -c_1^{\pm}\{\mathbf{C}_1^{-1}(\mathbf{i})R(b,c) \pm \mathbf{B}_1^{-1}(\mathbf{i})S(b,c)\}/\Delta,\\ \mathfrak{B}_{11}^{\pm} &= -b_1^{\pm}\{\mathbf{B}_1^{-1}(\mathbf{i})R(c,b) \mp \mathbf{C}_1^{-1}(\mathbf{i})S(c,b)\}/\Delta,\\ R(b,c) &\equiv 1 - b^{+}b^{-}\mathcal{K}^2 - c^{\mp}b^{\pm}h_1^2\\ &\quad + \mathcal{K}c^{\mp}e^{\mp i2\delta}[1 - b^{+}b^{-}(\mathcal{K}^2 + h_1)],\\ S(b,c) &\equiv h_1\mathcal{K}b^{\pm}(b^{\mp} - c^{\mp})\\ &\quad + h_1b^{\mp}e^{\mp i2\delta}[1 - b^{+}b^{-}(\mathcal{K}^2 + h_1^2)],\\ \Delta &\equiv 1 - \mathcal{K}^2(b^{+}b^{-} + c^{+}c^{-}) - h_1^2(c^{+}b^{-} + c^{-}b^{+})\\ &\quad + c^{+}c^{-}b^{+}b^{-}(\mathcal{K}^2 + h_1^2)^2,\end{aligned} \tag{189}$$

where $R(c, b)$ is obtained from $R(b, c)$ by interchanging b and c, and similarly for S. The corresponding coefficients $\mathfrak{C}_{1-1}^{\pm}$ and $\mathfrak{B}_{1-1}^{\pm}$ are obtained by replacing $\mathbf{C}_1^{-1}$ and $\mathbf{B}_1^{-1}$ by $\mathbf{C}_1^1$ and $\mathbf{B}_1^1$, and h_1 by $-h_1$.

Equations (151), (152), and (189) provide the coefficients for an explicit closed form for multiple scattering by two arbitrarily separated scatterers such that each is fully specified by its appropriate electric and magnetic dipoles when isolated. The set covers the special cases considered previously and allows us to obtain corrections, e.g., to (156) for the case where

the magnetic dipoles are not negligible. The present results apply to small spheres of different radii with both ϵ and μ different from unity, to two perfectly conducting spheres (for which case the electric and magnetic coefficients b, c are of the same order of magnitude), etc. Simple forms of the present results follow not only for the cases considered as illustrations but also for the limit of separations small compared to wavelength ($kb \approx 0$) for which region we use the origin expansions of the h's, as well as for separations large compared to wavelength ($kb \gg 1$) for which case we use the asymptotic form of the h's. From symmetry considerations, specializing the above two-scatterer results to identical scatterers, enables us to write down the corresponding solutions for one scatterer near a perfectly conducting plane, two protuberances on such a plane, and one protuberance on the wall of a perfectly conducting quadrant; see analogous expressions for two cylinders given previously.[23]

4.3. Small Scatterers of Arbitrary Shape

To construct analogous closed forms for two non-spherical scatterers small compared to wavelength, we base the development on (99)ff, and its generalization to an essentially arbitrary exciting electric field $\tilde{\Phi}$:

$$\tilde{u} \cdot \tilde{\Phi} = \tilde{h} \cdot \tilde{p} \cdot \tilde{\Phi} + (\nabla \times \tilde{h}) \cdot \tilde{m} \cdot (\nabla \times \tilde{\Phi}/k^2). \quad (190)$$

Electric dipoles: For a configuration of two arbitrary electric dipoles specified by $\tilde{p}_\pm$ excited by $\tilde{\varphi}$, the multiple-scattered fields $\tilde{U}_\pm$ may be written

$$\tilde{U}_\pm = \tilde{u}_\pm \cdot \tilde{\varphi}^\pm_\mp + \tilde{u}_\pm \cdot \tilde{U}^\pm_\mp \equiv \tilde{u}_\pm \cdot \tilde{\Phi}^\pm, \quad \tilde{u}_\pm = \tilde{h}(kr_\pm) \cdot \tilde{p}_\pm, \quad (191)$$

where $\tilde{\varphi}^+$ is the value of the source term at the scatterer located at $b\mathbf{b}_+ = b\hat{b}$, and where $\tilde{\Phi}^+$ is the corresponding total exciting field; similarly, for brevity, $\tilde{U}^\pm_\mp$ means the field of $\tilde{U}_-$ evaluated at $b\mathbf{b}_+$, etc. We have

$$\tilde{\Phi}^\pm = \tilde{\varphi}^\pm + \tilde{U}^\pm_\mp = \tilde{\varphi}^\pm + \tilde{u}^\pm_\mp \cdot \tilde{\Phi}^\mp; \quad (192)$$

consequently,

$$\tilde{\Phi}^\pm = (\tilde{I} - \tilde{u}^\pm_\mp \cdot \tilde{u}^\mp_\pm)^{-1} \cdot (\tilde{\varphi}^\pm + \tilde{u}^\pm_\mp \cdot \tilde{\varphi}^\mp),$$
$$\tilde{\varphi}^\pm = (\tilde{I} - \mathbf{ii})e^{\pm i\delta}, \quad \tilde{u}^\mp_\pm = \tilde{h}(2kb) \cdot \tilde{p}_\pm,$$
$$\tilde{h} = (\tilde{I} - \hat{b}\hat{b})\mathcal{K} + \hat{b}\hat{b}H, \quad (193)$$

where $\delta = \mathbf{k} \cdot \mathbf{b}$, and $\mathcal{K}$ and H are defined in (100).

We consider first the case corresponding to Sec. 4.2, for which $\tilde{p}_\pm = p_\pm \tilde{I}$ (small spheres, elementary model for oscillating electrons, etc.), where we have replaced the previous $\tilde{b}^\pm$ by $\tilde{p}_\pm$ to avoid confusion with the other b's. Since

$$\tilde{u}^\mp_\pm \cdot \tilde{u}^\mp_\pm = p_+p_-\tilde{h} \cdot \tilde{h} = p_+p_-[(\tilde{I} - \hat{b}\hat{b})\mathcal{K}^2 + \hat{b}\hat{b}H^2],$$

we may write

$$\tilde{I} - \tilde{u}^\mp_\pm \cdot \tilde{u}^\mp_\pm = (\tilde{I} - \hat{b}\hat{b})(1 - p_+p_-\mathcal{K}^2) + \hat{b}\hat{b}(1 - p_+p_-H^2),$$

and express the reciprocal as

$$(\tilde{I} - \tilde{u}^\pm_\mp \cdot \tilde{u}^\mp_\pm)^{-1} = \frac{\tilde{I} - \hat{b}\hat{b}}{1 - p_+p_-\mathcal{K}^2} + \frac{\hat{b}\hat{b}}{1 - p_+p_-H^2}. \quad (194)$$

We also have

$$\tilde{\varphi}^\pm - \tilde{u}^\pm_\mp \cdot \tilde{\varphi}^\pm = [(\tilde{I} - \hat{b}\hat{b})(1 + \mathcal{K}p_\mp e^{\mp i2\delta}) + \hat{b}\hat{b}(1 + Hp_\mp e^{\mp i2\delta})] \cdot \tilde{\varphi}^\pm. \quad (195)$$

Thus using (194) and (195), we reduce (193) to

$$\Phi^\pm = [(\tilde{I} - \hat{b}\hat{b})D_\mp + \hat{b}\hat{b}A_\mp] \cdot \tilde{\varphi}^\pm,$$
$$A_\mp = (1 + p_\mp e^{\mp i2\delta}H)/(1 - p_+p_-H^2) = A(H, p_\mp),$$
$$D_\mp = A(\mathcal{K}, p_\mp), \quad (196)$$

where A and D are essentially as defined in (151) and (154). The corresponding scattered waves from (191) are thus

$$\tilde{U}_\pm = \tilde{h}(r_\pm) \cdot \tilde{p}_\pm \cdot [(\tilde{I} - \hat{b}\hat{b})D_\mp + \hat{b}\hat{b}A_\mp] \cdot \tilde{\varphi}^\pm \equiv \tilde{h}(r_\pm) \cdot \tilde{P}_\pm \cdot \tilde{\varphi}^\pm, \quad (197)$$

where $\tilde{P}$ is the multiple-scattered moment. The asymptotic form of (197) for $kr \gg 1$, $r \gg b$ is $\tilde{U}_\pm \sim he^{\pm i(\delta - \Delta)}\tilde{G}_\pm$, with

$$\tilde{G}_\pm(\mathbf{o}, \mathbf{i}) = (\tilde{I} - \mathbf{oo}) \cdot \tilde{P}_\pm \cdot (\tilde{I} - \mathbf{ii}) = (\tilde{I} - \mathbf{oo}) \cdot p_\pm[(\tilde{I} - \hat{b}\hat{b})D_\mp + \hat{b}\hat{b}A_\mp] \cdot (\tilde{I} - \mathbf{ii}). \quad (198)$$

If we take $\hat{b} = \hat{z}$ (i.e., if we measure θ from $\hat{b}$), then the multiple-scattered amplitudes $\tilde{G}_\pm$ of (198) may be rewritten directly in the form (156) by using (107), i.e., for this choice of axis we have

$$\tilde{P}_+ = p_\pm[(\hat{x}\hat{x} + \hat{y}\hat{y})D_\mp + \hat{z}\hat{z}A_\mp],$$

and using (107) reduces (198) to the form (112) with $p_x = p_y = p_\pm D_\pm$ and $p_z = p_\pm A_\mp$.

For a small sphere of radius a with $\mu = 1$ and dielectric constant ϵ, we have $p \approx i(ka)^3(\epsilon - 1)/(\epsilon + 2)$. For small spacing $\rho = 2kb \ll 1$ we may use $\mathcal{K} \approx -H/2 \approx i/\rho^3$. Thus

$$\mathcal{K}p \approx -\tfrac{1}{2}Hp \to -\left(\frac{a}{2b}\right)^3\left(\frac{\epsilon - 1}{\epsilon + 2}\right)$$

in the static limit $k \to 0$. For this case we may also neglect $k\delta \to 0$. For identical scatterers we then have

$$\frac{\tilde{P}}{p} \to \frac{\tilde{I} - \hat{b}\hat{b}}{1 + R} + \frac{\hat{b}\hat{b}}{1 - 2R}, \quad R = \left(\frac{a}{2b}\right)^3\left(\frac{\epsilon - 1}{\epsilon + 2}\right). \quad (199)$$

[23] V. Twersky, J. Appl. Phys. **23**, 407 (1952).

On the other hand, for $\rho \gg 1$ we have $\mathcal{K} \sim h_0$ and $H \sim -2ih_0/\rho$; if we neglect H, then we get the "far-field multiple-scattering form"

$$\tilde{P} \sim p[D(\tilde{I} - \hat{b}\hat{b}) + \hat{b}\hat{b}], \quad \mathcal{K} \sim h_0(2kb); \quad (200)$$

more generally we use $\mathcal{K} = h[1 + (i/\rho) - 1/\rho^2]$, $H = -2h[(i/\rho) - 1/\rho^2]$ to convert $\tilde{P}$ to the analog of (144) plus (145).

The above dyadic dipoles are spherically symmetric in that $\tilde{p} = b\tilde{I}$ means the vector dipole $\tilde{p} \cdot \boldsymbol{\epsilon} = p\boldsymbol{\epsilon}$ has the direction of the incident polarization. We can also consider the case of a fixed vector dipole $\tilde{p} = p\hat{d}\hat{d}$ (i.e., a fixed metal wire oriented along $\hat{d}$), or the general dipole $\tilde{p} = \tilde{p}^{\mathrm{T}}$ of (102)ff. For simplicity we assume that the principal axes of the scatterers are parallel, and take $\hat{\xi} = \hat{b}$. Thus we may write

$$\tilde{u}_{\pm} = \tilde{h} \cdot \tilde{p}_{\pm} = \mathcal{K}(p_{\xi\pm}\hat{\xi}\hat{\xi} + p_{\eta\pm}\hat{\eta}\hat{\eta}) + Hp_{\zeta\pm}\hat{\zeta}\hat{\zeta}, \quad (201)$$

and obtain

$$\tilde{I} - \tilde{u}_{\mp}^{\pm} \cdot \tilde{u}_{\pm}^{\mp} = (1 - \mathcal{K}^2 p_{\xi+}p_{\xi-})\hat{\xi}\hat{\xi} + (1 - \mathcal{K}^2 p_{\eta+}p_{\eta-})\hat{\eta}\hat{\eta} + (1 - H^2 p_{\zeta+}p_{\zeta-})\hat{\zeta}\hat{\zeta},$$

$$(\tilde{I} - u_{\mp}^{\pm} \cdot u_{\pm}^{\mp})^{-1} = E_{\xi}(\mathcal{K}, p)\hat{\xi}\hat{\xi} + E_{\eta}(\mathcal{K}, p)\hat{\eta}\hat{\eta} + E_{\zeta}(H, p)\hat{\zeta}\hat{\zeta},$$

$$E_{\xi}(\mathcal{K}, p) = 1/(1 - \mathcal{K}^2 p_{\xi+}p_{\xi-}), \quad \text{etc.} \quad (202)$$

Substituting into $\tilde{p} \cdot \tilde{\Phi} = \tilde{P} \cdot \tilde{\varphi}$, we construct

$$\tilde{P}_{\pm} = p_{\pm} \cdot [\tilde{I} - \tilde{u}_{\mp}^{\pm} \cdot \tilde{u}_{\pm}^{\mp}]^{-1} \cdot [\tilde{I} + \tilde{u}_{\mp}^{\pm} e^{\mp i2\delta}] = p_{\xi\pm}D(p_{\xi\mp})\hat{\xi}\hat{\xi} + p_{\eta\pm}D(p_{\eta\mp})\hat{\eta}\hat{\eta} + p_{\zeta\pm}A(p_{\zeta\mp})\hat{\zeta}\hat{\zeta}. \quad (203)$$

Magnetic dipoles: Similarly for two magnetic dipoles, say each of the form

$$\tilde{v} \cdot \nabla \times \tilde{\varphi} = \nabla \times \tilde{h} \cdot (\tilde{m}/k^2) \cdot \nabla \times \tilde{\varphi} = (-h_1 \mathbf{o} \times \tilde{I}) \cdot \tilde{m} \cdot (\tilde{I} \times \mathbf{i}) \cdot \tilde{\varphi} i, \quad (204)$$

we may work with the electric functions

$$\tilde{U}_{\pm} = \tilde{v}_{\pm} \cdot \nabla \times \tilde{\Phi}^{\pm}, \quad (205)$$

$$\tilde{\Phi}^{\pm} = \tilde{\varphi}^{\pm} + \tilde{U}_{\mp}^{\pm} = \tilde{\varphi}^{\pm} + v_{\mp}^{\pm} \cdot \nabla \times \tilde{\Phi}^{\mp} \quad (206)$$

to obtain

$$\nabla \times \tilde{\Phi}^{\pm} = [\tilde{I} - (\nabla \times \tilde{v}_{\mp}^{\pm}) \cdot (\nabla \times \tilde{v}_{\pm}^{\mp})]^{-1} \cdot [\nabla \times \tilde{\varphi}^{\pm} + \nabla \times \tilde{v}_{\mp}^{\pm} \cdot \nabla \times \tilde{\varphi}^{\mp}]. \quad (207)$$

Since

$$\nabla \times \tilde{v} = \nabla \times \nabla \times \tilde{h} \cdot \tilde{m}/k^2 = \tilde{h} \cdot \tilde{m},$$

the function $[\ ]^{-1}$ is of the same form as for the electric case but with the previous p replaced by m.

For $\tilde{m}_{\pm} = m_{\pm}\tilde{I}$ corresponding to the spherically symmetric case of Sec. 4.2, we have from (207) and (196),

$$\nabla \times \tilde{\Phi}^{\pm} = [(\tilde{I} - \hat{b}\hat{b})D(m_{\mp}) + \hat{b}\hat{b}A(m_{\mp})] \cdot \tilde{\varphi}^{\pm} \times \mathbf{i}ik. \quad (208)$$

Thus using (208) in (205) with $\tilde{v}$ as in (204), we obtain

$$\tilde{U}_{\pm} = ih_1 m_{\pm}(\mathbf{o} \times \tilde{I}) \cdot [(\tilde{I} - \hat{b}\hat{b})D(m_{\mp}) + \hat{b}\hat{b}A(m_{\pm})] \cdot (\tilde{I} \times \mathbf{i})e^{\pm i\delta} \equiv (\nabla \times \tilde{h}_{\pm}) \cdot \tilde{M}_{\pm} \cdot \nabla \times \tilde{\varphi}_{\pm}/k^2, \quad (209)$$

where $\tilde{M}$ is the multiple-scattered moment. Since

$$h_1(kr_{\pm}) \sim -ih(kr_{\pm}) \sim -ih(kr)e^{\mp i\Delta},$$

the scattering amplitudes are

$$\tilde{G}_{\pm} = -(\mathbf{o} \times \tilde{I}) \cdot \tilde{M}_{\pm} \cdot (\tilde{I} \times \mathbf{i}) = -(\mathbf{o} \times \tilde{I}) \cdot m_{\pm}[(\tilde{I} - \hat{b}\hat{b})D(m_{\mp}) + \hat{b}\hat{b}A(m_{\mp})] \cdot (\tilde{I} \times \mathbf{i}). \quad (210)$$

If we take $\hat{b} = \hat{z}$ then (210) may be rewritten directly in the form (171) by using (109) to reduce $\tilde{G}$ to the form (113).

For a small sphere of radius a with $\epsilon = 1$ and permittivity μ, we replace ϵ by μ in the previous illustration. Similarly for the magnetic analog of the more general case (203) we obtain

$$\tilde{M}_{\pm} = m_{\xi\pm}D(m_{\xi\mp})\hat{\xi}\hat{\xi} + m_{\eta\pm}D(m_{\eta\mp})\hat{\eta}\hat{\eta} + m_{\zeta\pm}A(m_{\zeta\mp})\hat{\zeta}\hat{\zeta}. \quad (211)$$

See Appendix B for analogous results for scalar problem.

Electric plus magnetic: If we are dealing with one electric ($\tilde{u}_+$) and one magnetic ($\tilde{v}_-$) dipole, then we may work with

$$\tilde{U}_{+} = \tilde{u}_{+} \cdot (\tilde{\varphi}^{+} + \tilde{U}_{-}^{+}) = \tilde{u}_{+} \cdot \tilde{\Phi}^{+}, \quad (212)$$

$$\tilde{U}_{-} = \tilde{v}_{-} \cdot \nabla \times (\tilde{\varphi}^{-} + \tilde{U}_{+}^{-}) = \tilde{v}_{-} \cdot \nabla \times \tilde{\Phi}^{-}, \quad (213)$$

where

$$\tilde{\Phi}^{+} = \tilde{\varphi}^{+} + \tilde{v}_{-}^{+} \cdot \nabla \times \tilde{\Phi}^{-}, \quad (214)$$

$$\nabla \times \tilde{\Phi}^{-} = \nabla \times \tilde{\varphi}^{-} + (\nabla \times \tilde{u}_{+}^{-}) \cdot \tilde{\Phi}^{+}. \quad (215)$$

Solving (214) and (215) we obtain

$$\tilde{\Phi}^{+} = [\tilde{I} - \tilde{v}_{-}^{+} \cdot \nabla \times \tilde{u}_{+}^{-}]^{-1} \cdot [\tilde{\varphi}^{+} + \tilde{v}_{-}^{+} \cdot \nabla \times \tilde{\varphi}^{-}], \quad (216)$$

$$\nabla \times \tilde{\Phi}^{-} = [\tilde{I} - (\nabla \times \tilde{u}_{+}^{-}) \cdot \tilde{v}_{-}^{+}]^{-1} \cdot [\nabla \times \tilde{\varphi}^{-} + (\nabla \times \tilde{u}_{+}^{-}) \cdot \tilde{\varphi}^{+}]. \quad (217)$$

We have

$$\tilde{v}_{-}^{+} = \nabla \times \tilde{h} \cdot \tilde{m}/k^2 = -kh_1(\hat{b} \times \tilde{I}) \cdot \tilde{m}/k^2,$$

and similarly

$$\nabla \times \tilde{u}_{+}^{-} = -kh_1(-\hat{b} \times \tilde{I}) \cdot \tilde{p} = kh_1(\hat{b} \times \tilde{I}) \cdot \tilde{p}.$$

For the case of Sec. 4.2, we have $\tilde{p} = p\tilde{I}$ and $\tilde{m} = m\tilde{I}$ and

$$\tilde{v}_{-}^{+} \cdot \nabla \times \tilde{u}_{+}^{-} = -m_{-}p_{+}h_1^2(\hat{b} \times \tilde{I}) \cdot (\hat{b} \times \tilde{I}) = m_{-}p_{+}h_1^2(\tilde{I} - \hat{b}\hat{b}).$$

Thus

$$(\tilde{I} - \tilde{v}_-^+ \cdot \nabla \times \tilde{u}_+^-)^{-1} = [\tilde{I} - (\tilde{I} - \hat{b}\hat{b})m_- p_+ h_1^2]^{-1} = \hat{b}\hat{b} + (\tilde{I} - \hat{b}\hat{b})/(1 - p_+ m_- h_1^2) \equiv \hat{b}\hat{b} + (\tilde{I} - \hat{b}\hat{b})E,$$

and consequently

$$\tilde{\Phi}^+ = [\hat{b}\hat{b} + (\tilde{I} - \hat{b}\hat{b})E - im_- h_1(\hat{b} \times \tilde{I})e^{-i2\delta}(\tilde{I} \times \mathbf{i})E] \cdot \tilde{\varphi}^+, \quad E = 1/(1 - p_+ m_- h_1^2). \tag{218}$$

Similarly

$$\nabla \times \tilde{\Phi}^-/ik = \{[\hat{b}\hat{b} + (\tilde{I} - \hat{b}\hat{b})E] \cdot (\mathbf{i} \times \tilde{I}) - ip_+ h_1(\hat{b} \times \tilde{I})e^{i2\delta}\} \cdot \tilde{\varphi}_-. \tag{219}$$

Using (218) and (219) in (212) and (213) gives the corresponding electric dyadic fields. To obtain the scattering amplitude we use

$$\tilde{u}_+ = p_+ \tilde{h}(hr^+) \sim p_+ h_0(kr)e^{-i\Delta}(\tilde{I} - \mathbf{oo}),$$

and similarly

$$\tilde{v}_- = -kh_1(\mathbf{o} \times \tilde{I})m_-/k^2 \sim -h_0 e^{i\Delta} m_-(\mathbf{o} \times \tilde{I})/ik.$$

Thus

$$\tilde{G}_+ = (\tilde{I} - \mathbf{oo}) \cdot \{p_+[\hat{b}\hat{b} + (\tilde{I} - \hat{b}\hat{b})E] - i(\hat{b} \times \tilde{I}) \cdot (\mathbf{i} \times \tilde{I})Fe^{-i2\delta}\} \cdot (\tilde{I} - \mathbf{ii}), \quad F = p_+ m_- h_1 E, \tag{220}$$

$$\tilde{G}_- = -(\mathbf{o} \times \tilde{I}) \cdot \{m[\hat{b}\hat{b} + (\tilde{I} - \hat{b}\hat{b})E] + i(\hat{b} \times \tilde{I}) \cdot (\tilde{I} \times \mathbf{i})Fe^{i2\delta}\} \cdot (\mathbf{i} \times \tilde{I}), \tag{221}$$

where we replaced $\tilde{I} - \mathbf{ii}$ in the last term by $-(\tilde{I} \times \mathbf{i}) \cdot (\tilde{I} \times \mathbf{i})$ to stress the similarities of the form of the composite moments $\mathbf{P}_+$ and $\mathbf{M}_-$ corresponding to (220) and (221), respectively. To reduce (220) and (221) to the forms in (179), we take $\hat{b}\hat{b} = \hat{z}\hat{z}$ and use (107) and (109).

For the more general case of

$$\tilde{p}_+ = \sum p_x \hat{x}\hat{x}, \quad \tilde{m}_- = \sum m_x \hat{x}\hat{x}, \quad x = \xi, \eta, \zeta, \tag{222}$$

with $\hat{\zeta} = \hat{b}$, we have

$$\tilde{v}_-^+ = \nabla \times \tilde{h} \cdot \tilde{m}/k^2 = -h_1(\hat{\zeta} \times \tilde{I}) \cdot \tilde{m}/k = h_1(\hat{\xi}\hat{\eta}m_\eta - \hat{\eta}\hat{\xi}m_\xi)/k,$$

$$\nabla \times \tilde{u}_+^- = kh_1 \hat{\zeta} \times \tilde{I} \cdot \tilde{p} = -kh_1(\hat{\xi}\hat{\eta}p_\eta - \hat{\eta}\hat{\xi}p_\xi).$$

Consequently

$$\tilde{v}_-^+ \cdot \nabla \times \tilde{u}_+^- = h_1^2(\hat{\xi}\hat{\xi}p_\xi m_\eta + \hat{\eta}\hat{\eta}p_\eta m_\xi),$$

$$\nabla \times \tilde{u}_+^- \cdot \tilde{v}_-^+ = h_1^2(\hat{\xi}\hat{\xi}m_\xi p_\eta + \hat{\eta}\hat{\eta}m_\eta p_\xi).$$

Thus

$$\tilde{\Phi}^+ = [\hat{\zeta}\hat{\zeta} + \hat{\xi}\hat{\xi}E(p_\xi m_\eta) + \hat{\eta}\hat{\eta}E(p_\eta m_\xi)] \cdot [\tilde{I} + ih_1 e^{-i2\delta}(m_\eta \hat{\xi}\hat{\eta} - m_\xi \hat{\eta}\hat{\xi})(\tilde{I} \times \mathbf{i})] \cdot \tilde{\varphi}^+, \tag{223}$$

$$\tilde{P}_+ \equiv \tilde{p}_+ \cdot \tilde{\Phi}^+ \cdot (\tilde{\varphi}^+)^{-1} = p_\zeta \hat{\zeta}\hat{\zeta} + p_\xi E(p_\xi, m_\eta)\hat{\xi}\hat{\xi} + p_\eta E(p_\eta m_\xi)\hat{\eta}\hat{\eta} + ie^{-i2\delta}[F(p_\xi m_\eta)\hat{\xi}\hat{\eta} - F(p_\eta m_\xi)\hat{\eta}\hat{\xi}](\mathbf{i} \times \tilde{I}),$$

$$\tilde{G}_+ = (\tilde{I} - \mathbf{oo}) \cdot \tilde{P}_+ \cdot (\tilde{I} - \mathbf{ii}). \tag{224}$$

Similarly

$$\nabla \times \tilde{\Phi}_- = [\hat{\zeta}\hat{\zeta} + \hat{\xi}\hat{\xi}E(m_\xi p_\eta) + \hat{\eta}\hat{\eta}E(m_\eta p_\xi)] \cdot [\tilde{I} - ih_1(\hat{\xi}\hat{\eta}p_\eta - \hat{\eta}\hat{\xi}p_\xi)e^{i2\delta}(\mathbf{i} \times \tilde{I})] \cdot ik(\mathbf{i} \times \tilde{\varphi}^-), \tag{225}$$

$$\tilde{M}_- \equiv (\tilde{m}_- \cdot \nabla \times \tilde{\Phi}^-)(ik\mathbf{i} \times \boldsymbol{\varphi}^-)^{-1} = m_\zeta \hat{\zeta}\hat{\zeta} + m_\xi E(m_\xi p_\eta)\hat{\xi}\hat{\xi} + m_\eta E(m_\eta p_\xi)\hat{\eta}\hat{\eta} - ie^{i2\delta}[F(m_\xi p_\eta)\hat{\xi}\hat{\eta} - F(m_\eta p_\xi)\hat{\eta}\hat{\xi}](\mathbf{i} \times \tilde{I}),$$

$$\tilde{G}_- = -(\mathbf{o} \times \tilde{I}) \cdot \tilde{M} \cdot (\tilde{I} \times \mathbf{i}), \tag{226}$$

which differs from (224) in the interchange of m and p and the replacement of i by $-i$.

If each scatterer consists of an electric plus a magnetic dipole such that the isolated-scatterer values equal $\tilde{u}_\pm + \tilde{v}_\pm$ with $\tilde{u}$ and $\tilde{v}$ as defined in this section, then we write the scattered electric dyadics as

$$\tilde{U}_\pm = \tilde{u}_\pm \cdot \tilde{\Phi}^\pm + \tilde{v}_\pm \cdot \nabla \times \tilde{\Phi}^\pm = (\tilde{u}_\pm + \tilde{v}_\pm \cdot \nabla \times \tilde{I}) \cdot \tilde{\Phi}^\pm, \tag{227}$$

with

$$\tilde{\Phi}^\pm = \tilde{\varphi}^\pm + (\tilde{u}_\mp^\pm + \tilde{v}_\mp^\pm \cdot \nabla \times \tilde{I}) \cdot \tilde{\Phi}^\mp, \tag{228}$$

where $\nabla \times \tilde{I} \cdot \tilde{\Phi}$ is a temporary expedient for $\nabla \times \tilde{\Phi}$. Eliminating $\tilde{\Phi}^\mp$ from the right-hand side, we obtain

$$\tilde{\Phi}^\pm = (\tilde{\varphi}^\pm + \tilde{u}_\mp^\pm \cdot \tilde{\varphi}^\mp + \tilde{v}_\mp^\pm \cdot \nabla \times \tilde{\varphi}^\mp) + (\tilde{u}_\mp^\pm \cdot \tilde{u}_\pm^\mp + \tilde{v}_\mp^\pm \cdot \nabla \times \tilde{u}_\pm^\mp) \cdot \tilde{\Phi}^\pm + (\tilde{u}_\mp^\pm \cdot \tilde{v}_\pm^\mp + \tilde{v}_\mp^\pm \cdot \nabla \times \tilde{v}_\pm^\mp) \cdot \nabla \times \tilde{\Phi}^\pm, \tag{229}$$

plus the corresponding expression for $\nabla \times \tilde{\Phi}^\pm$ obtained by replacing all left-hand elements in the terms in parentheses by their curls, e.g., $\nabla \times \tilde{\Phi}^\pm$ involves

$$\nabla \times \hat{\varphi}^\pm + (\nabla \times \tilde{u}_\mp^\pm) \cdot \hat{\varphi}^\mp + (\nabla \times \tilde{v}_\mp^\pm) \cdot (\nabla \times \hat{\varphi}^\mp),$$

etc. Essentially as for the previous case, we may solve the simultaneous equations for $\tilde{\Phi}^\pm$ and $\nabla \times \tilde{\Phi}^\pm$ to reduce the above to the case considered previously by separations of variables.

APPENDIX A. RECIPROCITY RELATIONS

We should stress that $\tilde{G}$ of (131) does not in general satisfy the reciprocity relation of the form (75):

$$\tilde{g}_t(-\mathbf{i}_1, \mathbf{i}_2) = \tilde{g}_t^{\mathrm{T}}(-\mathbf{i}_2, \mathbf{i}_1). \tag{A1}$$

We may always write

$$\tilde{G}_t(-\mathbf{i}_1,\mathbf{i}_2)=\tilde{R}_t(-\mathbf{i}_2,\mathbf{i}_1)+\tilde{L}_t(-\mathbf{i}_1,\mathbf{i}_2)$$
$$=\tilde{R}_t(-\mathbf{i}_1,\mathbf{i}_2)+\sum{}'\tilde{R}'_{ts}(-\mathbf{i}_1,\mathbf{i}_2)e^{-ik\mathbf{i}_2\cdot\mathbf{b}_{ts}}, \tag{A2}$$

where $\tilde{R}_t$ (R for "reversible") includes only those "chains" of successive scattering processes which start and end with scatterer t, and L includes those that start with $s\neq t$ and end with t. Interchanging the directions we get

$$\tilde{G}^{\mathrm{T}}_t(-\mathbf{i}_2,\mathbf{i}_1)=\tilde{R}^{\mathrm{T}}_t(-\mathbf{i}_2,\mathbf{i}_1)+\sum{}'\tilde{R}'_{ts}(-\mathbf{i}_2,\mathbf{i}_1)e^{-ik\mathbf{i}_1\cdot\mathbf{b}_{ts}}$$
$$=\tilde{R}_t(-\mathbf{i}_1,\mathbf{i}_2)+\sum{}'\tilde{R}'_{ts}(-\mathbf{i}_1,\mathbf{i}_2)e^{-ik\mathbf{i}_1\cdot\mathbf{b}_{ts}}, \tag{A3}$$

so that

$$\tilde{G}_t(-\mathbf{i}_1,\mathbf{i}_2)-\tilde{G}^{\mathrm{T}}_t(-\mathbf{i}_2,\mathbf{i}_1)$$
$$=\sum{}'\tilde{R}'_{ts}(-\mathbf{i}_1,\mathbf{i}_2)[e^{-ik\mathbf{i}_2\cdot\mathbf{b}_{ts}}-e^{-ik\mathbf{i}_1\cdot\mathbf{b}_{ts}}], \tag{A4}$$

is not in general zero. We illustrate this explicitly for an elementary case in (160)ff. In the present Appendix we list additional theorems for G.

The compound scattering amplitude

$$\tilde{\mathfrak{G}}(-\mathbf{i}_1,\mathbf{i}_2)=\sum e^{ik(\mathbf{i}_2+\mathbf{i}_1)\cdot\mathbf{b}_{ts}}\tilde{G}_s(-\mathbf{i}_1,\mathbf{i}_2) \tag{A5}$$

of (117) satisfies the same theorems as $\tilde{g}$. Thus using (A4) in (75), we obtain

$$\tilde{G}_t(-\mathbf{i}_1,\mathbf{i}_2)=\tilde{G}^{\mathrm{T}}_t(-\mathbf{i}_2,\mathbf{i}_1)$$
$$-\sum{}'e^{-ik(\mathbf{i}_1+\mathbf{i}_2)\cdot\mathbf{b}_{ts}}[\tilde{G}_s(-\mathbf{i}_1,\mathbf{i}_2)-\tilde{G}^{\mathrm{T}}_s(-\mathbf{i}_2,\mathbf{i}_1)]. \tag{A6}$$

These equations follow essentially from $\{\tilde{\Psi}_1,\tilde{\Psi}_2\}=0$ over any surface bounding the collection. In addition, we have $\{\tilde{\psi}_1,\tilde{\Psi}_2\}_t=0$ which led to the "reciprocity" relation of (131), i.e.,

$$\tilde{G}_t(-\mathbf{i}_1,\mathbf{i}_2)=\tilde{g}^{\mathrm{T}}_t(-\mathbf{i}_2,\mathbf{i}_1)$$
$$+\sum{}'\int e^{ik(\mathbf{p}-\mathbf{i}_2)\cdot\mathbf{b}_{ts}}\tilde{g}^{\mathrm{T}}_t(-\mathbf{p},\mathbf{i}_1)\cdot\tilde{G}_s(\mathbf{p},\mathbf{i}_2)\,d\Omega/2\pi, \tag{A7}$$

as well as the result obtained by interchanging and transposition:

$$\tilde{G}^{\mathrm{T}}_t(-\mathbf{i}_2,\mathbf{i}_1)=\tilde{g}(-\mathbf{i}_1,\mathbf{i}_2)$$
$$+\sum{}'\int e^{ik(\mathbf{p}-\mathbf{i}_1)\cdot\mathbf{b}_{ts}}\tilde{G}^{\mathrm{T}}_s(\mathbf{p},\mathbf{i}_1)\cdot\tilde{g}_t(-\mathbf{p},\mathbf{i}_2)\,d\Omega/2\pi. \tag{A8}$$

Subtracting (A8) from (A7), we obtain

$$\tilde{G}_t(-\mathbf{i}_1,\mathbf{i}_2)=\tilde{G}^{\mathrm{T}}_t(-\mathbf{i}_2,\mathbf{i}_1)$$
$$+\sum{}'\int e^{ik\mathbf{p}\cdot\mathbf{b}_{ts}}[e^{-ik\mathbf{i}_2\cdot\mathbf{b}_{ts}}\tilde{g}^{\mathrm{T}}_t(-\mathbf{p},\mathbf{i}_1)\tilde{G}_s(\mathbf{p},\mathbf{i}_2)$$
$$-e^{-ik\mathbf{i}_1\cdot\mathbf{b}_{ts}}\tilde{G}^{\mathrm{T}}_s(\mathbf{p},\mathbf{i}_1)\cdot\tilde{g}_t(-\mathbf{p},\mathbf{i}_2)]\,d\Omega/2\pi. \tag{A9}$$

Similarly from $\{\tilde{\Psi}'_1,\tilde{\Psi}'_2\}_t=0$ we obtain

$$\tilde{G}_t(-\mathbf{i}_1,\mathbf{i}_2)=\tilde{G}^{\mathrm{T}}_t(-\mathbf{i}_2,\mathbf{i}_1)$$
$$+\sum{}'\int e^{ik\mathbf{p}\cdot\mathbf{b}_{ts}}[e^{-ik\mathbf{i}_2\cdot\mathbf{b}_{ts}}\tilde{G}^{\mathrm{T}}_t(-\mathbf{p},\mathbf{i})\tilde{G}_s(\mathbf{p},\mathbf{i}_2)$$
$$-e^{-ik\mathbf{i}_1\cdot\mathbf{b}_{ts}}\tilde{G}^{\mathrm{T}}_s(\mathbf{p},\mathbf{i}_1)\cdot\tilde{G}_t(-\mathbf{p},\mathbf{i}_2)]\,d\Omega/2\pi, \tag{A10}$$

which we may reduce to (A9) by substituting (A7) for $\tilde{G}^{\mathrm{T}}_t(-\mathbf{p})$.

The above "reciprocity relations" follow from

$$\{\tilde{\psi}_1,\tilde{\psi}_2\}_t=\{\tilde{\Psi}_1,\tilde{\Psi}_2\}_c=\{\tilde{\psi}_1,\tilde{\Psi}_2\}_t=\{\tilde{\Psi}_1,\tilde{\Psi}_2\}_t=0, \tag{A11}$$

where t indicates the surface that isolates scatterer t from the others, and c indicates a surface around the whole collection. We may also regard the theorems for lossless scatterers that follow from

$$\{\tilde{\psi}^*_1,\tilde{\psi}_2\}_t=\{\tilde{\Psi}^*_1,\tilde{\Psi}_2\}_c=\{\tilde{\psi}^*_1,\tilde{\Psi}_2\}_t=\{\tilde{\Psi}^*_1,\tilde{\Psi}_2\}_t=0, \tag{A12}$$

as "reciprocity relations."

The first form of (A12) yields (79), and the second yields (79) with $\tilde{g}$ replaced by the compound amplitude $\tilde{\mathfrak{G}}$. Using (A5) in theorem (79) for $\tilde{\mathfrak{G}}$, we obtain

$$\tilde{G}_t(\mathbf{i}^*_1,\mathbf{i}_2)+\tilde{G}^{\dagger}_t(\mathbf{i}^*_2,\mathbf{i}_1)$$
$$=-\sum{}'[\tilde{G}_s(\mathbf{i}^*_1,\mathbf{i}_2)-\tilde{G}^{\dagger}_s(\mathbf{i}_2,\mathbf{i}_1)]e^{-ik(\mathbf{i}_2-\mathbf{i}_1{}^*)\cdot\mathbf{b}_{ts}}$$
$$-\frac{1}{2\pi}\int[\tilde{G}^{\dagger}_t(\mathbf{o},\mathbf{i}_1)\cdot\tilde{G}_t(\mathbf{o},\mathbf{i}_2)$$
$$+\sum{}'\tilde{G}^{\dagger}_s\cdot\tilde{G}_s e^{-ik(\mathbf{i}_2-\mathbf{i}_1{}^*)\cdot\mathbf{b}_{ts}}$$
$$+\sum\sum{}'\tilde{G}^{\dagger}_m\cdot\tilde{G}_n e^{ik(\mathbf{o}\cdot\mathbf{b}_{mn}+\mathbf{i}_2\cdot\mathbf{b}_{nt}+\mathbf{i}_1{}^*\cdot\mathbf{b}_{tm})}]d\Omega, \tag{A13}$$

where $\mathbf{b}_{nm}=\mathbf{b}_n-\mathbf{b}_m$, etc. In the forward direction $\mathbf{i}^*_1=\mathbf{i}_2=\mathbf{i}_1=\mathbf{i}^*_2=\mathbf{i}$ for the class of scatterers such that $\tilde{G}^{\mathrm{T}}(\mathbf{o},\mathbf{i})=\tilde{G}(\mathbf{i},\mathbf{o})$ we have

$$-\mathrm{Re}\,\tilde{G}_t(\mathbf{i},\mathbf{i})=\mathrm{Re}\sum{}'\tilde{G}_s(\mathbf{i},\mathbf{i})$$
$$+\frac{1}{4\pi}\int[\tilde{G}^{*}_t(\mathbf{i},\mathbf{o})\cdot\tilde{G}_t(\mathbf{o},\mathbf{i})+\sum{}'\tilde{G}^{*}_s\cdot\tilde{G}_s$$
$$+\sum\sum{}'\tilde{G}^{*}_m\cdot\tilde{G}_n e^{ik(\mathbf{o}-\mathbf{i})\cdot\mathbf{b}_{mn}}]\,d\Omega. \tag{A14}$$

We consider a special case of (A14) corresponding to two simple scatterers in (163)ff, and have considered other special cases in the papers on periodic and random distributions cited in Refs. 4 and 5.

The third form in (A12) yields

$$\tilde{G}_t(\mathbf{i}^*_1,\mathbf{i}_2)+\tilde{g}^{\dagger}_t(\mathbf{i}_2,\mathbf{i}_1)=-\frac{1}{2\pi}\int\tilde{g}^{\dagger}_t(\mathbf{o},\mathbf{i}_1)\cdot\tilde{G}_t(\mathbf{o},\mathbf{i}_2)d\Omega_o$$
$$-\frac{1}{2\pi}\sum{}'\int e^{ik(\mathbf{p}-\mathbf{i}_2)\cdot\mathbf{b}_{ts}}\,\tilde{g}^{\dagger}_t(\mathbf{p}^*,\mathbf{i}_1)\cdot\tilde{G}_s(\mathbf{p},\mathbf{i}_2)\,d\Omega_p, \tag{A15}$$

in which we may specialize to forward scattering and

use (75) for $\tilde{g}$ to eliminate Re $\tilde{g}(\mathbf{i}, \mathbf{i})$. Finally the fourth form of (A12) yields

$$\begin{aligned}\tilde{G}_t(\mathbf{i}_1^*, \mathbf{i}_2) + \tilde{G}_t^\dagger(\mathbf{i}_2, \mathbf{i}_1) &= -\frac{1}{2\pi}\int \tilde{G}_t^\dagger(\mathbf{o}, \mathbf{i}_1)\cdot\tilde{G}_t(\mathbf{o}, \mathbf{i}_2)\, d\Omega_o \\ &\quad - \frac{1}{2\pi}\sum{}'\int e^{ik(\mathbf{p}-\mathbf{i}_2)\cdot\mathbf{b}_{ts}}\tilde{G}_t^\dagger(\mathbf{p}^*, \mathbf{i}_1)\cdot\tilde{G}_s(\mathbf{p}, \mathbf{i}_2)\, d\Omega \\ &\quad - \frac{1}{2\pi}\sum{}'\left[\int e^{ik(\mathbf{p}-\mathbf{i}_2)\cdot\mathbf{b}_{ts}}\tilde{G}_t^\dagger(\mathbf{p}^*, \mathbf{i}_2)\cdot\tilde{G}_s(\mathbf{p}, \mathbf{i}_1)\, d\Omega\right]^\dagger. \end{aligned} \tag{A16}$$

For forward scattering and $\tilde{G}^{\mathrm{T}}(\mathbf{o}, \mathbf{i}) = \tilde{G}(\mathbf{i}, \mathbf{o})$ we have

$$\begin{aligned}-\mathrm{Re}\,\tilde{G}_t(\mathbf{i}, \mathbf{i}) &= \frac{1}{4\pi}\int \tilde{G}^*(\mathbf{i}, \mathbf{o})\cdot\tilde{G}(\mathbf{o}, \mathbf{i})\, d\Omega_o \\ &\quad + \frac{1}{2\pi}\,\mathrm{Re}\sum{}'\int e^{ik(\mathbf{p}-\mathbf{i})\cdot\mathbf{b}_{ts}}\tilde{G}_t^*(\mathbf{i}, \mathbf{p}^*)\cdot\tilde{G}_s(\mathbf{p}, \mathbf{i})\, d\Omega. \end{aligned} \tag{A17}$$

We consider special cases of (A17) in the papers on periodic and random distributions cited in Refs. 4 and 5.

APPENDIX B. SCALAR DIPOLES

In the previous developments of the analogous scalar problems (Refs. 1 and 3), the case of two different monopoles was used as the simplest illustration. For the present electromagnetic case, our discussion of two electrical dipoles as in (191)ff provides the dyadic analog of the previous results for monopoles: i.e., if we replace the dyadics by appropriate scalars we again get the earlier results. Thus the scalar version of (191)ff is

$$\begin{aligned} U_\pm &= u_\pm\Phi^\pm, \quad u_\pm = a_0^\pm h(kr_\pm), \quad \varphi = e^{i\mathbf{k}\cdot\mathbf{r}}, \\ \Phi^\pm &= \varphi^\pm + u_\mp^\pm\Phi^\mp = \frac{(1 + u_\mp^\pm e^{\mp i2\delta})\varphi^\pm}{1 - u_+u_-} \\ &= \frac{[1 + a_0^\mp e^{\mp i2\delta}h(2kb)]\varphi^\pm}{1 - a_0^+a_0^-h^2}, \end{aligned} \tag{B1}$$

where $h = h_0^{(1)}$ for the three-dimensional problem [see Ref. 3, Eq. (63)], and $h = H_0^{(1)}$ for the two-dimensional problem [see Ref. 1, Eq. (71)].

The corresponding scalar problems of symmetrical dipoles,

$$u = iH_1^{(1)}(kr)a_1(2)\cos(\theta - \theta_i)\varphi, \quad \varphi = e^{ik\mathbf{i}\cdot\mathbf{r}}, \tag{B2}$$

$$u = ih_1^{(1)}(kr)a_1(3)\cos(\theta - \theta_i)\varphi, \tag{B3}$$

where (B2) and (B3) correspond to two dimensions and three dimensions, respectively, were considered by separations of variables.[1,3] The normalization of the scattering coefficients $a_1(2)$ and $a_1(3)$ is here chosen so that for lossless scatterers, we have $-\mathrm{Re}\,a_1(n) = n\,|a_1(n)|^2$, e.g., for $\psi = 0$ at the surface, we have $a_1 = -2J_1/H_1$ and $a_2 = -3j_1/h_1$. The present dyadic development for two magnetic dipoles (204)ff suggests an analogous development for generalizing the scalar results to two arbitrary dipoles.

We rewrite (B2) and (B3) in the single form

$$\begin{aligned} u \equiv \mathbf{v}\cdot\nabla\varphi &= -(a/k^2)(\nabla h)\cdot(\nabla\varphi) \\ &= -(a/k^2)[kh'\mathbf{o}]\cdot[ik\varphi\mathbf{i}] \\ &= -iah'\varphi(\mathbf{i}\cdot\mathbf{o}), \end{aligned} \tag{B4}$$

where h is either $H_0^{(1)}$ or $h_0^{(1)}$, and $h' = \partial_{kr}h(kr)$ is either $-H_0^{(1)}$ or $-h_0^{(1)}$; similarly $a = a_1(2)$, $a_1(3)$. We may now proceed essentially as for (205). Thus for two dipoles we use

$$U_\pm = \mathbf{v}_\pm\cdot\nabla(\varphi^\pm + U_\mp^\pm) = \mathbf{v}_\pm\cdot\nabla\Phi^\pm, \qquad \mathbf{v} = -(a/k^2)\nabla h, \tag{B5}$$

$$\Phi^\pm = \varphi^\pm + \mathbf{v}_\mp^\pm\cdot\nabla\Phi^\mp. \tag{B6}$$

Taking the gradient of (B6), and eliminating $\nabla\Phi^\pm$ from the right-hand side, we obtain the analog of (207):

$$\nabla\Phi^\pm = [\tilde{I} - \nabla\mathbf{v}_\mp^\pm\cdot\nabla\mathbf{v}_\pm^\mp]^{-1}\cdot[\nabla\varphi^\pm + \nabla\mathbf{v}_\mp^\pm\cdot\nabla\varphi^\mp]. \tag{B7}$$

We have

$$\begin{aligned}\nabla\mathbf{v} &= -(a/k^2)\nabla\nabla h \\ &= a[\hat{b}\hat{b}h_1' + (\tilde{I} - \hat{b}\hat{b})h_1/\rho] \\ &\equiv a[\hat{b}\hat{b}\mathcal{K} + (\tilde{I} - \hat{b}\hat{b})H], \quad \rho = 2kb, \end{aligned} \tag{B8}$$

where h is either $h_0^{(1)}$ or $H_0^{(1)}$, and h_1 is either $h_1^{(1)}$ or $H_1^{(1)}$. Thus $\nabla\mathbf{v}_\mp^\pm\cdot\nabla\mathbf{v}_\pm^\mp = a_+a_-[\hat{b}\hat{b}\mathcal{K}^2 + (\tilde{I} - \hat{b}\hat{b})H^2]$, and

$$\begin{aligned}[\tilde{I} - \nabla\mathbf{v}\cdot\nabla\mathbf{v}]^{-1} &= \hat{b}\hat{b}E(\mathcal{K}) + (\tilde{I} - \hat{b}\hat{b})E(H), \\ E(\mathcal{K}) &= 1/(1 - a_+a_-\mathcal{K}^2), \end{aligned} \tag{B9}$$

$$\begin{aligned}\nabla\Phi^\pm &= [\hat{b}\hat{b}A(\mathcal{K}) + (\tilde{I} - \hat{b}\hat{b})A(H)]\cdot\nabla\varphi^\pm \\ &\equiv \tilde{P}\cdot\nabla\varphi^\pm/a_\pm, \\ A(a^\pm, \mathcal{K}) &= \frac{1 + a^\mp\mathcal{K}e^{\mp i2\delta}}{1 - a_+a_-\mathcal{K}^2}, \end{aligned} \tag{B10}$$

where $\tilde{P}_\pm$ are the multiple-scattered moments.

Substituting (B10) into (B5), we obtain

$$U_\pm = \mathbf{v}_\pm\cdot\tilde{P}\cdot\nabla\varphi^\pm/a_\pm \sim h(kr)e^{\pm i(\delta-\Delta)}G_\pm, \tag{B11}$$

where the multiple-scattered amplitudes equal

$$\begin{aligned}G_\pm(\mathbf{o}, \mathbf{i}) &= \mathbf{o}\cdot\tilde{P}\cdot\mathbf{i} \\ &= \mathbf{o}\cdot[\hat{b}\hat{b}a_\pm A(a^\mp, \mathcal{K}) + (\tilde{I} - \hat{b}\hat{b})a_\pm A(a^\mp, H)]\cdot\mathbf{i}. \end{aligned} \tag{B12}$$

(Note the shift in location between H and $\mathcal{K}$ type functions from axial to perpendicular components as

compared to vector problems in the text proper; this is in accord with the relations between, e.g., acoustic pressure dipoles and the electric functions in the text.)

If we express (B12) in terms of the appropriate two- and three-dimensional special functions, we obtain the previous results [(85) of Ref. 1 for circular cylinders, and (69) of Ref. 3 for spheres].

To generalize the above to arbitrary dipoles, we replace $\mathbf{v}$ in the above by

$$\mathbf{v} = -(1/k^2)\nabla h \cdot \tilde{p}, \tag{B13}$$

so that

$$u = -(1/k^2)\nabla h \cdot \tilde{p} \cdot \nabla\varphi = ih_1\varphi \mathbf{o} \cdot \tilde{p} \cdot \mathbf{i}, \tag{B14}$$

where, e.g., $\tilde{p}$ may be constructed from the known approximations for elliptic cylinders and ellipsoids. For the case where the principal axis of $\tilde{p}_+$ and $\tilde{p}_-$ are parallel (i.e., essentially as in the text) we obtain (B11) and (B12) with $\tilde{P}$ replaced by

$$\tilde{P}_\pm = p_{\xi\pm}A(p_{\xi\mp}, H)\hat{\xi}\hat{\xi} + p_{\eta\pm}A(p_{\eta\mp}, H)\hat{\eta}\hat{\eta} + p_{\zeta\pm}A(p_{\zeta\mp}, \mathcal{K})\hat{\zeta}\hat{\zeta}, \tag{B15}$$

where the ξ term is to be dropped for two dimensions.

We could also extend the above to all moments by working with

$$u(n) = (L_n h) \otimes p_n \otimes (L_n\varphi); \quad n = 2, 3. \tag{B16}$$

Thus for monopoles

$$L = 1, \quad p = a. \tag{B17}$$

For dipoles

$$L = \nabla/ik, \quad \otimes = \cdot, \quad p = (\mathbf{ab}). \tag{B18}$$

For quadrupoles we have

$$L_n = \left(\tilde{I} + n\frac{\nabla\nabla}{k^2}\right), \quad \otimes = :, \quad p_n = (\mathbf{abcd})/(n-1)n, \tag{B19}$$

which represents u as the scalar resulting from double-dotting a tetradic[6] fore and aft by dyadics.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Representations of the Lie Algebra of the Homogeneous Galilei Group and Their Relation to the Representations of the Lorentz Algebra

JACQUES VOISIN
Department of Theoretical Physics, University of Bujumbura, Bujumbura, Burundi, and Department of Physics, Syracuse University, Syracuse, New York

(Received 23 May 1966)

The irreducible Hermitian representations of the Lie algebra of the homogeneous Galilei group were first constructed by using the method of the maximal compact subgroup. The same representations were then obtained by contracting the irreducible representations of the homogeneous Lorentz group. In the Appendix, the faithful representations are given with the diagonalized generators of the velocity transformations.

I. INTRODUCTION

THE homogeneous Galilei group $\mathcal{G}$ is a 6-parameter Lie group. The corresponding 6-dimensional Lie algebra consists of the linear combinations of the elements K_1, K_2, K_3, J_1, J_2, and J_3 with the commutation relations

$$[K_i, K_j] = 0, \tag{1.1}$$

$$[J_k, K_j] = i\epsilon_{kjm}K_m, \tag{1.2}$$

$$[J_k, J_j] = i\epsilon_{kjm}J_m. \tag{1.3}$$

In the following, we construct the Hermitian irreducible representations of this Lie algebra in a basis which diagonalizes J^2. We first use the method of the maximal compact subgroup[1]; we then apply the method of contraction to the irreducible representations of the Lorentz group.[2] In the Appendix, the faithful representations are expressed with the diagonalized K_1, K_2, and K_3.

II. HERMITIAN IRREDUCIBLE REPRESENTATIONS OF THE LIE ALGEBRA OF $\mathcal{G}$

In order to obtain the irreducible Hermitian representations of the infinitesimal elements of $\mathcal{G}$, we proceed in two steps.

First, the general form of the matrices K_1, K_2, and K_3 is such that

$$[J_k, P_j] = i\epsilon_{kjm}P_m$$

is determined from the known expressions of the irreducible representations of the J_k's.

Let us choose, in the representation space, a basis labeled by the spin eigenvalues l, and l is fixed by the eigenvalues m of J_3. We also introduce a third parameter τ to distinguish between the subspaces with the same l; however, such a parameter is not necessary in the present case, since we restrict ourselves to irreducible representations; for these representations, any l appears at most once. We then have

$$J_3\,|l, m\rangle' = m\,|l, m\rangle', \tag{2.1}$$

$$J_+\,|l, m\rangle' = \alpha^l_{m+1}\,|l, m+1\rangle', \tag{2.2}$$

$$J_-\,|l, m\rangle' = \alpha^l_m\,|l, m-1\rangle', \tag{2.3}$$

where

$$(\alpha^l_m)^2 = (l+m)(l-m+1), \tag{2.4}$$

$$J_\pm = J_1 \pm iJ_2. \tag{2.5}$$

The states $|l, m\rangle'$ are normalized to one; later on, we change this normalization.

Now, if we want to obtain the general matrices representing K_1, K_2, and K_3 submitted to conditions (1.2), we can use the Wigner–Eckart theorem,[3] according to which the matrix element

$$\langle j', m'|\, T(1, q)\,|j, m\rangle$$

of the qth component of the tensor of first rank (= vector) T (1) is given by

$$\langle j', m'|\, T(1, q)\,|j, m\rangle = (-)^{1-j+j'}\,[\langle 1, q, j, m \mid 1, j, j', m'\rangle/(2j'+1)^{\frac{1}{2}}]\,\langle j'\|\, T(1)\,\|j\rangle, \tag{2.6}$$

where q takes the values 1, 0, -1, and

$$\langle 1, q, j, m \mid 1, j, j', m'\rangle$$

is the usual Clebsch–Gordan coefficient. Equation (2.6) gives the expected result if we put

$$T_1 = K_+ = K + iK_2, \quad T_0 = K_3,$$

$$T_{-1} = K_- = K_1 - iK_2. \tag{2.7}$$

[1] We follow the way proposed, in the case of the Lorentz group, by I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon Press, Inc., London, 1963), Part I, Chap. II, Sec. 9. The method of the maximal compact subgroup has been further extended by the Syracuse University group, particularly by Jacob G. Kuriyan, Ph.D. thesis, Syracuse University (1966).

[2] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **39**, 510 (1953); E. J. Saletan, J. Math. Phys. **2**, 1 (1961).

[3] See, for example, A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1957).

However, we prefer to follow here the notation of Gel'fand, Minlos, and Shapiro[1] in order to make use of their results on the Lorentz group. These authors have shown that, if

$$K_3 |l, m\rangle' = c_{l',l;m',m} |l', m'\rangle', \tag{2.8}$$

we have

$$\begin{aligned} c_{l',l;m'm} &= c_{l',l;m}\delta_{m'm}, \\ c_{l-1,l;m} &= c_{l-1,l}(l^2 - m^2)^{\frac{1}{2}}, \\ c_{l+1,l;m} &= c_{l+1,l}[(l+1)^2 - m^2]^{\frac{1}{2}}, \\ c_{l',l;m} &= 0, \quad \text{if} \quad |l' - l| > 1, \\ c_{l,l;m} &= c_{l,l}m. \end{aligned} \tag{2.9}$$

Now, let $h(l)$ be given by

$$h(l) = \left(\prod_{r=l_0}^{l} c_{r+1,r}/c_{r-1,r}\right)^{\frac{1}{2}}, \tag{2.10}$$

with l_0 being the smallest weight involved in the representation. If we replace the states $|l, m\rangle'$ by the states

$$|l, m\rangle = h(l) |l, m\rangle', \tag{2.11}$$

and if we define A_l and C_l by

$$A_l = -ic_{l,l}, \quad C_l = [h(l)/h(l-1)](-i)c_{l-1,l}, \tag{2.12}$$

we obtain for K_3, K_+, and K_- the representation quoted in Eqs. (I.1), (I.4), (I.5), and (I.6) of Table I. J_3, J_+, J_- are still given by (2.1), (2.2), and (2.3) with $|l, m\rangle'$ replaced by $|l, m\rangle$.

What remains to be done is the determination of A_l and C_l by taking into account the relations that

$$[K_3, K_\pm] = 0, \quad [K_+, K_-] = 0,$$

and that the representation is Hermitian and irreducible (for K_j, J_l). This is the second step of our procedure.

By following the method of Gel'fand, Minlos, and Shapiro[1] for the Lorentz group, we obtained, after lengthy calculation, the results that the irreducible Hermitian representations are characterized by two numbers l_0 and l_1. l_0 is the smallest weight involved in the representation. $(-l_1^2)$ is the value of the Casimir operator $K_3^2 + K_+K_-$. In the representation (l_0, l_1),

$$A_l = il_0l_1/l(l+1), \quad C_l = -l^{-1}[l_1^2(l^2 - l_0^2)/(4l^2 - 1)]^{\frac{1}{2}},$$

where l_1 is either a pure imaginary number or zero.

Case I: l_1 is the pure imaginary number. In this case, the representation is infinite dimensional, l takes the values l_0, $l_0 + 1$, $l_0 + 2$, $\cdots$. The representation is single-valued (double-valued) if and only if l_0 is an integer (half odd integer).

Case II: $l_1 = 0$. The representation is finite. Since A_l and C_l are equal to zero, the translations are trivially represented and the representation is unfaithful. It coincides with the irreducible representation l_0 of the rotation algebra.

III. $\mathfrak{G}$ AS A CONTRACTION OF THE HOMOGENEOUS LORENTZ GROUP $\mathfrak{L}$

We are going to obtain the irreducible representations of $\mathfrak{G}$ by considering this latter group as the contraction of the homogeneous Lorentz group $\mathfrak{L}$ with respect to the rotations.

The group $\mathfrak{L}$ has six infinitesimal operators M_{ik}, K_i corresponding to rotations in the planes (i, k) $[i, k = 1, 2, 3]$ and $(i, 0)$ $[i = 1, 2, 3]$, respectively. The commutation relations, in the usual normal coordinate system, are as follows (with $J_+ = M_{23} + iM_{13}$, $J_- = M_{23} - iM_{13}$, $J_3 = M_{12}$, $K_+ = K_1 + iK_2$, $K_- = K_1 - iK_2$, and K_3):

$$\begin{gathered} [J_+, J_3] = -J_+, \quad [J_-, J_3] = J_-, \quad [J_+, J_-] = 2J_3, \\ [K_+, J_+] = [J_-, K_-] = [J_3, K_3] = 0, \\ [J_+, K_3] = -K_+, \quad [J_-, K_3] = K_-, \\ [J_+, K_-] = -[J_-, K_+] = 2K_3, \\ [K_+, J_3] = -K_+, \quad [K_-, J_3] = K_-, \\ [K_+, K_3] = J_+, \quad [K_-, K_3] = -J_-, \quad [K_+, K_-] = -2J_3. \end{gathered} \tag{3.1}$$

A. Irreducible Hermitian Representations of the Lie Algebra of $\mathfrak{L}$

The representations of the generators J_3, $J_\pm$, K_3, $K_\pm$ of $\mathfrak{G}$ and those of the generators J_3, $J_\pm$, K_3, $K_\pm$ of $\mathfrak{L}$ differ by nothing other than the values of A_l and C_l.[4] In both cases, Eqs. (I.1)–(I.6) of Table I are valid. But instead of (I.7) and (I.8), we have in the case of $\mathfrak{L}$ the following values for A_l and C_l

$$\begin{aligned} A_l &= il_0l_1/l(l+1), \\ C_l &= (i/l)[(l^2 - l_0^2)(l^2 - l_1^2)/(4l^2 - 1)]^{\frac{1}{2}}. \end{aligned}$$

The pair of invariants (l_0, l_1) can take three different types of values.

Case I: l_1 is a pure imaginary number, l_0 is an arbitrary integer or half-integer. The corresponding representations form the so-called main series of representations; they are infinite dimensional ($l = l_0$, $l_0 + 1$, $l_0 + 2$, $\cdots$).

Case II: $l_0 = 0$, l_1 is a real number such that $|l_1| < 1$. The corresponding representations are infinite dimensional too ($l = l_0, l_0 + 1, \cdots$). They belong to the so-called supplementary series.

Case III: $l_0 = 0$, $l = 1$. The corresponding representation is finite. It belongs to the supplementary series.

[4] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, Ref. 1.

TABLE I. Lie algebra of $\mathfrak{G}$.

$$J_3|l,m\rangle = m|l,m\rangle \quad \text{(I.1)}$$

$$J_+|l,m\rangle = \alpha^l_{m+1}|l,m+1\rangle \quad \text{(I.2)} \qquad (\alpha^l_m)^2 = (l+m)(l-m+1)$$

$$J_-|l,m\rangle = \alpha^l_m|l,m-1\rangle \quad \text{(I.3)}$$

$$K_3|l,m\rangle = C_l[(l^2-m^2)^{\frac{1}{2}}|l-1,m\rangle - A_l m|l,m\rangle - C_{l+1}[(l+1)^2-m^2]^{\frac{1}{2}}|l+1,m\rangle \quad \text{(I.4)}$$

$$K_+|l,m\rangle = C_l[(l-m)(l-m-1)]^{\frac{1}{2}}|l-1,m+1\rangle - A_l[(l-m)(l+m+1)]\ |l,m+1\rangle + C_{l+1}[(l+m+1)(l+m+2)]^{\frac{1}{2}}|l+1,m+1\rangle \quad \text{(I.5)}$$

$$K_-|l,m\rangle = -C_l[(l+m)(l+m-1)]^{\frac{1}{2}}|l-1,m+1\rangle - A_l[(l+m)(l-m+1)]^{\frac{1}{2}}|l,m-1\rangle - C_{l-1}[(l-m+1)(l-m+2)]^{\frac{1}{2}}|l+1,m-1\rangle \quad \text{(I.6)}$$

$$A_l = il_0l_1/l(l+1) \quad \text{(I.7)} \qquad C_l = -\frac{1}{l}[l_1^2(l^2-l_0^2)/(4l^2-1)] \quad \text{(I.8)}$$

Case I: l_1 = pure imaginary; $l = l_0, l_0+1, \cdots; l_0 =$ {positive integer, positive half odd integer.

Case II: $l_1 = 0$; $l = l_0$.

B. Contraction of $\mathfrak{L}$

The rotations form a subgroup of $\mathfrak{L}$ with respect to which contraction may be carried out in the sense of Inönü and Wigner.[2] If we make the substitution

$$M'_{ij} = M_{ij}, \quad J'_3 = J_3, J'_\pm = J_\pm,$$

or

$$K'_i = \epsilon K_i, \quad K'_3 = \epsilon K_3, K'_\pm = \epsilon K_\pm, \qquad (3.2)$$

the commutation relations (3.1) lead, in the limit $\epsilon \to 0$, to the commutation relations (1.1)–(1.3) characteristic of the Galilei group.

Now, let us consider the irreducible Hermitian representations of the algebra of $\mathfrak{L}$ and from these let us try to obtain representations for $\mathfrak{G}$ by using the contraction procedure. Relations (I.1)–(I.3) in Table I remain unaltered under the substitution (3.2), but the same is not true for (I.4)–(I.6), because A_l and C_l are replaced by

$$A'_l = il_0(\epsilon l_1)/l(l+1),$$

$$C'_l = (i/l)[(l^2-l_0^2)(\epsilon^2l^2-\epsilon^2l_1^2)/(4l^2-1)]. \qquad (3.3)$$

Now, let ϵ go to zero. For the representations belonging to the supplementary series ($|l_1| \leq 1$), A'_l and C'_l tend to zero. Hence, the generators K_3, $K_\pm$ are represented by the null matrix. For $J_3, J_\pm$ we have the representation (I.1)–(I.3) with either $l = 0$ if we contract the finite representation ($l_0 = 0$, $l_1 = 1$) of $\mathfrak{L}$ or $l = 0, 1, 2, \cdots$ if we start from the representation ($l_0 = 0$, $|l_1| < 1$). In the former case, the final representation is irreducible (trivial one-dimensional representation), in the latter case, it is reducible into the representations $l = 0$, $l = 1$, $\cdots$ of the rotation algebra.

On the other hand, in the case of the principal series, we have two possibilities: either we keep l_1 (and l_0) fixed, then A_l and C_l tend to zero with ϵ and we obtain a representation which is reducible into the representations $l = l_0$, $l = l_0+1$, $\cdots$ of the rotation algebra; or we let l_1 go to infinity (which is allowed since, here, any positive value may be accepted for $|l_1|$) in such a way that

$$\lim_{\substack{|l_1|\to\infty \\ \epsilon\to 0}} l_1\epsilon = \Xi \qquad (3.4)$$

with Ξ a fixed finite number; here we have again to distinguish between two cases—if $\Xi = 0$, A'_l and C'_l are zero at the limit and we are reduced to the case considered above; but if $\Xi \neq 0$, A'_l and C'_l tend to the nonzero limits

$$A'_l = \epsilon A_l \to i\Xi l_0/l(l+1), \qquad (3.5)$$

$$C'_l = \epsilon C_l \to (i/l)[-\Xi^2(l^2-l_0^2)/(4l^2-1)]^{\frac{1}{2}} \qquad (3.6)$$

with Ξ a pure imaginary number. l can take the values l_0, l_0+1, $\cdots$; the corresponding representation is obviously an irreducible representation of Type I for $\mathfrak{G}$ [infinite (faithful) representation].

APPENDIX. FAITHFUL IRREDUCIBLE REPRESENTATIONS OF THE ALGEBRA OF $\mathfrak{G}$ WITH THE DIAGONALIZED K_3, K_+, AND K_-

It can be derived from results obtained by Wightman[5] that the irreducible faithful representations of the homogeneous Galilei group can be written ($\mathbf{v}$ = change in velocity or acceleration, R = rotation) as

$$[U(\mathbf{v},R)\Psi](\mathbf{q}) = \exp iS\varphi \cdot \exp i\mathbf{q}\cdot\mathbf{v}\Psi(R^{-1}\mathbf{q}) \qquad \text{(A1)}$$

with $|\mathbf{q}|$ a constant, S the integer (half odd integer), and φ the angle of the rotation $R_{\mathbf{q}}^{-1}RR_{R^{-1}\mathbf{q}}$ about $\mathbf{e}_3 \cdot R_{\mathbf{q}}$ is the rotation about the axis $\mathbf{e}_3\wedge\mathbf{q}/q$ by the smallest positive angle between $\mathbf{q}$ and the unit vector $\mathbf{e}_3$ on the axis Ox_3.

[5] A. S. Wightman, Rev. Mod. Phys. **34**, 845 (1962).

It is then possible to obtain the corresponding representation for the generators by going to the infinitesimal elements. We find

$$\begin{aligned} K_i &= q_i, \\ iJ_1 &= q_2(\partial/\partial q_3) - q_3(\partial/\partial q_2) - iS[q_1/(q+q_3)], \\ iJ_2 &= q_3(\partial/\partial q_1) - q_1(\partial/\partial q_3) - iS[q_2/(q+q_3)], \\ iJ_3 &= q_1(\partial/\partial q_2) - q_2(\partial/\partial q_1) - iS. \end{aligned} \tag{A2}$$

The values of J_i are obtained by computing the angle of the rotation $R_{\mathbf{q}}^{-1}RR_{R^{-1}\mathbf{q}}$ for an infinitesimal R. We use the 2×2 representation of the rotations and the formulas[6]

$$R_{\mathbf{q}} = 2^{-\frac{1}{2}}[1 + (\mathbf{q}\cdot\mathbf{l}_3/q)]^{-\frac{1}{2}}\{1 + (\mathbf{q}\cdot\mathbf{l}_3/q) + i[\boldsymbol{\sigma}\cdot(\mathbf{l}_3\wedge\mathbf{q})/q]\}, \tag{A3}$$

$$R(\mathbf{n},\theta) = 1 + \tfrac{1}{2}i\theta\boldsymbol{\sigma}\cdot\mathbf{n} \tag{A4}$$

if θ is infinitesimal

$$\mathbf{q}' = R^{-1}\mathbf{q} = \mathbf{q}\cos\theta + (\mathbf{q}\cdot\mathbf{n})\mathbf{n}(1-\cos\theta) - (\mathbf{n}\wedge\mathbf{q})\sin\theta \cong \mathbf{q} - (\mathbf{n}\wedge\mathbf{q})\theta. \tag{A5}$$

[6] The general method used here is especially well explained by A. J. Macfarlane, J. Math. Phys. **3**, 6 (1962). See also J. M. Levy-Leblond, J. Math. Phys. **4**, 776 (1961).

Putting now

$$R_{\mathbf{q}}^{-1}RR_{R^{-1}\mathbf{q}} = 1 + \tfrac{1}{2}i\varphi\boldsymbol{\sigma}\cdot\boldsymbol{\omega} \tag{A6}$$

we have

$$\varphi \cong -\frac{\mathbf{n}\cdot\mathbf{l}_3 + \mathbf{n}\cdot(\mathbf{q}/q)}{1 + (\mathbf{q}\cdot\mathbf{l}_3)/q}\,\theta. \tag{A7}$$

Thus, we obtain the result (A2), since from (A1) we deduce

$$[U(\mathbf{0}, R)\Psi](\mathbf{q}) = (1 + iS\varphi)[1 + \theta(\mathbf{q}\wedge\nabla_{\mathbf{q}})\cdot\mathbf{n}]\Psi(\mathbf{q}) \tag{A8}$$

if R is infinitesimal.

ACKNOWLEDGMENTS

The author wishes to express his hearty thanks to Professor E. C. G. Sudarshan for his kind interest in this work and for critical discussions. The author thanks Professor M. Dehousse for the hospitality extended to him at the University of Bujumbura. He also thanks Dr. J. G. Nagel for assistance with the final form of the manuscript.

This research was supported in part by the U.S. Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

General Theory of Resistive Beam Instabilities

STEVEN WEINBERG*
Department of Physics, University of California, Berkeley, California

(Received 1 June 1966)

This paper considers all resistive instabilities of a self-pinched cylindrically symmetric beam of charged particles in a finite or an infinite Ohmic plasma channel. The problem is reduced to an ordinary second-order linear differential equation for the longitudinal component of the perturbed electric field. The equation can be solved for a uniform beam shape, yielding an implicit transcendental equation whose roots define the various modes. We find that for each azimuthal "quantum number" m there are two infinite sequences of modes and two exceptional modes, except that some of these modes are missing for $m = 0$, 1, and 2. In all modes we find stable oscillation at very low and very high frequencies, and instability at intermediate frequencies, the growth rates generally reaching maxima somewhat less than the betatron frequency ω_β. The largest maximum growth rate is in the "hose" mode (the only exceptional mode for $m = 1$), where it is approximately $0.29\,\omega_\beta$. For a general smooth beam shape, the catalog of modes is similar to that for a uniform beam, except that there also appears a continuous spectrum. It is also proved for general beam shape that at low frequencies the "hose" dispersion relation becomes the same as that derived earlier under the assumption of rigid beam displacement; this is not the case at higher frequencies.

I. INTRODUCTION

THIS article deals with the general resistive instabilities of a self-pinched cylindrically symmetric unmodulated beam of charged particles passing through an Ohmic plasma channel. The modes are characterized by the appearance in the fields and currents of exponential factors

$$e^{im\theta}e^{-i\omega t}e^{ikz} \tag{1.1}$$

multiplying various functions of r. The streaming modes, with ω of the order of the plasma channel conductivity σ, have been adequately treated in previous articles[1]; the present work deals only with the resistive instabilities, with $|\omega| \ll \sigma$.

What we hope to get for our trouble is a dispersion formula giving k (or ω) for general complex values of

* Morris Loeb Lecturer, Physics Department, Harvard University, Cambridge, Massachusetts.

[1] E. A. Frieman, M. L. Goldberger, K. M. Watson, S. Weinberg, and M. N. Rosenbluth, Phys. Fluids **5**, 196 (1962); earlier references are quoted therein.

It is then possible to obtain the corresponding representation for the generators by going to the infinitesimal elements. We find

$$\begin{aligned} K_i &= q_i, \\ iJ_1 &= q_2(\partial/\partial q_3) - q_3(\partial/\partial q_2) - iS[q_1/(q+q_3)], \\ iJ_2 &= q_3(\partial/\partial q_1) - q_1(\partial/\partial q_3) - iS[q_2/(q+q_3)], \\ iJ_3 &= q_1(\partial/\partial q_2) - q_2(\partial/\partial q_1) - iS. \end{aligned} \tag{A2}$$

The values of J_i are obtained by computing the angle of the rotation $R_{\mathbf{q}}^{-1}RR_{R^{-1}\mathbf{q}}$ for an infinitesimal R. We use the 2×2 representation of the rotations and the formulas[6]

$$R_{\mathbf{q}} = 2^{-\frac{1}{2}}[1 + (\mathbf{q}\cdot\mathbf{l}_3/q)]^{-\frac{1}{2}}\{1 + (\mathbf{q}\cdot\mathbf{l}_3/q) + i[\boldsymbol{\sigma}\cdot(\mathbf{l}_3\wedge\mathbf{q})/q]\}, \tag{A3}$$

$$R(\mathbf{n},\theta) = 1 + \tfrac{1}{2}i\theta\boldsymbol{\sigma}\cdot\mathbf{n} \tag{A4}$$

if θ is infinitesimal

$$\mathbf{q}' = R^{-1}\mathbf{q} = \mathbf{q}\cos\theta + (\mathbf{q}\cdot\mathbf{n})\mathbf{n}(1-\cos\theta) - (\mathbf{n}\wedge\mathbf{q})\sin\theta \cong \mathbf{q} - (\mathbf{n}\wedge\mathbf{q})\theta. \tag{A5}$$

[6] The general method used here is especially well explained by A. J. Macfarlane, J. Math. Phys. **3**, 6 (1962). See also J. M. Levy-Leblond, J. Math. Phys. **4**, 776 (1961).

Putting now

$$R_{\mathbf{q}}^{-1}RR_{R^{-1}\mathbf{q}} = 1 + \tfrac{1}{2}i\varphi\boldsymbol{\sigma}\cdot\boldsymbol{\omega} \tag{A6}$$

we have

$$\varphi \cong -\frac{\mathbf{n}\cdot\mathbf{l}_3 + \mathbf{n}\cdot(\mathbf{q}/q)}{1 + (\mathbf{q}\cdot\mathbf{l}_3)/q}\,\theta. \tag{A7}$$

Thus, we obtain the result (A2), since from (A1) we deduce

$$[U(\mathbf{0}, R)\Psi](\mathbf{q}) = (1 + iS\varphi)[1 + \theta(\mathbf{q}\wedge\nabla_{\mathbf{q}})\cdot\mathbf{n}]\Psi(\mathbf{q}) \tag{A8}$$

if R is infinitesimal.

ACKNOWLEDGMENTS

The author wishes to express his hearty thanks to Professor E. C. G. Sudarshan for his kind interest in this work and for critical discussions. The author thanks Professor M. Dehousse for the hospitality extended to him at the University of Bujumbura. He also thanks Dr. J. G. Nagel for assistance with the final form of the manuscript.

This research was supported in part by the U.S. Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

General Theory of Resistive Beam Instabilities

STEVEN WEINBERG*
Department of Physics, University of California, Berkeley, California

(Received 1 June 1966)

This paper considers all resistive instabilities of a self-pinched cylindrically symmetric beam of charged particles in a finite or an infinite Ohmic plasma channel. The problem is reduced to an ordinary second-order linear differential equation for the longitudinal component of the perturbed electric field. The equation can be solved for a uniform beam shape, yielding an implicit transcendental equation whose roots define the various modes. We find that for each azimuthal "quantum number" m there are two infinite sequences of modes and two exceptional modes, except that some of these modes are missing for $m = 0$, 1, and 2. In all modes we find stable oscillation at very low and very high frequencies, and instability at intermediate frequencies, the growth rates generally reaching maxima somewhat less than the betatron frequency ω_β. The largest maximum growth rate is in the "hose" mode (the only exceptional mode for $m = 1$), where it is approximately $0.29\,\omega_\beta$. For a general smooth beam shape, the catalog of modes is similar to that for a uniform beam, except that there also appears a continuous spectrum. It is also proved for general beam shape that at low frequencies the "hose" dispersion relation becomes the same as that derived earlier under the assumption of rigid beam displacement; this is not the case at higher frequencies.

I. INTRODUCTION

THIS article deals with the general resistive instabilities of a self-pinched cylindrically symmetric unmodulated beam of charged particles passing through an Ohmic plasma channel. The modes are characterized by the appearance in the fields and currents of exponential factors

$$e^{im\theta}e^{-i\omega t}e^{ikz} \tag{1.1}$$

multiplying various functions of r. The streaming modes, with ω of the order of the plasma channel conductivity σ, have been adequately treated in previous articles[1]; the present work deals only with the resistive instabilities, with $|\omega| \ll \sigma$.

What we hope to get for our trouble is a dispersion formula giving k (or ω) for general complex values of

* Morris Loeb Lecturer, Physics Department, Harvard University, Cambridge, Massachusetts.

[1] E. A. Frieman, M. L. Goldberger, K. M. Watson, S. Weinberg, and M. N. Rosenbluth, Phys. Fluids **5**, 196 (1962); earlier references are quoted therein.

ω (or k). It is helpful to have in mind a more specific problem of practical importance in experiments like those at the Astron: Suppose the beam is tickled at $z = 0$ with a perturbation having a single real frequency ω and with m value and radial dependence such that only a single mode is excited. The instability grows if $\mathrm{Im}\, k < 0$, and at a point z downstream the number of e foldings is $z\,|\mathrm{Im}\, k|$. Our primary aim is then to decide, for each mode, at which real value of ω there occurs a maximum in $-\mathrm{Im}\, k$ and to calculate this maximum value. However, there are conditions of "free growth" for which ω and k take complex values determined by a saddle-point condition[2]; hence we really solve our instability problem only when we understand the behavior of $k(\omega)$ throughout the complex ω plane.

One of the resistive instabilities with $m = 1$, the "hose" mode, has already been intensively studied[3,4] under the assumption that the beam moves rigidly from side to side. This assumption is expected to be valid for very small ω, or more precisely, when the skin depth $(c^2/4\pi\sigma\omega)^{\frac{1}{2}}$ is much larger than the beam radius. However, the growth rate is also expected to reach a maximum when the skin depth is of the order of the beam radius, and in this most important case we cannot rely on the rigid beam approximation. Furthermore, the "sausage" ($m = 0$) mode and all higher modes depend for their existence on beam deformation, so we could never hope to treat them all until we learned how to take into account the perturbations in the individual beam particle orbits. *In this paper we discuss all m values in a unified way, with no assumption of beam rigidity in any mode.*

The simplifying assumptions made here are:

(1) Beam particles undergo no collisions. This assumption is essential not only to our calculation but also to the very existence of the beam. (For a high-density beam moving very fast with respect to the plasma, we might have to take into account beam–beam collisions without worrying about beam–plasma collisions. In this case the instability could be treated hydro-dynamically without looking at individual beam particle orbits.)

(2) The unperturbed beam particles all have velocities with the same z-component v. This assumption is made for simplicity, and does not appear to be of critical importance in these modes.

(3) The particles in the unperturbed beam move in *circular* helices[5] around the z axis. This assumption is made so that a beam particle feels a simple harmonic force from the perturbed fields. It could be given up, but only at the cost of having to solve an additional nontrivial second-order differential equation for the perturbed particle orbits.

(4) The parameters of the beam and plasma are such that

$$c^2/4\pi\sigma a v \ll 1, \tag{1.2}$$

$$\omega_\beta a/v \ll 1, \tag{1.3}$$

and we only look for modes with

$$|\omega|\, a \ll v, \tag{1.4}$$

$$|k|\, a \ll 1. \tag{1.5}$$

Here, σ is the plasma conductivity, c is the speed of light, a is a characteristic beam radius, v is the z component of the unperturbed beam particle velocity, and ω_β is a typical value of the unperturbed beam particle gyration frequency around the z axis. All previous work on the hose and sausage modes has been based on the assumptions (1.4) and (1.5), because they lead to an enormous simplification in the derivation of the dispersion relation and in the dispersion relation itself. The point of the assumptions (1.2) and (1.3) is that we expect and do find the fastest growing modes to be such that

$$|\omega| \approx c^2/4\pi a^2\sigma, \tag{1.6}$$

$$|\omega - kv| \approx \omega_\beta, \tag{1.7}$$

and in this case (1.2) and (1.3) are necessary and sufficient for (1.4) and (1.5). [As an example, nominal values for the beam experiment at Astron[6] are $\sigma \simeq 10^{12}\ \mathrm{sec}^{-1}$, $a \simeq 1$ cm, $\omega_\beta \simeq 9 \times 10^8\ \mathrm{sec}^{-1}$, and $v \simeq c$, so the left-hand sides of (1.2) and (1.3) take the values 0.002 and 0.03, respectively, and our assumptions are well justified.] It should be noted that (1.4) and (1.6) yield

$$|\omega|/4\pi\sigma \approx |\omega|^2 a^2/c^2 \ll 1,$$

and, as already remarked, this excludes the streaming modes from our present consideration.

In Ref. 4 we first found the "exact" hose dispersion relation, and then applied (1.2)–(1.7), at the cost of considerable clarity and physical insight. Here we use

[2] This point is discussed in detail in Sec. XII.

[3] The original work on the "hose" instability was done by C. Longmire (unpublished) and M. Rosenbluth, Phys. Fluids 3, 932 (1960). See also Ref. 4, and unpublished work by H. Lewis, K. Brueckner, G. Ascoli, H. Chang, S. Yadavalli, H. Singhaus, and R. Briggs.

[4] The "hose" instability was discussed for general beam shape, with emphasis on the effects of beam modulation, by S. Weinberg, J. Math. Phys. 5, 1371 (1964).

[5] This assumption was used in unpublished work by H. Lewis to obtain the "sausage" dispersion relation for a uniform beam. Our result for this case [Eq. (8.12) with $m = 0$ and $R = \infty$] is identical with that previously obtained by Lewis.

[6] N. Christofilos, private communication.

(1.2)–(1.7) from the beginning; it is hoped the reader finds this an improvement.

The first half of the present paper (Secs. II–VI) is devoted to a derivation of the dispersion relation as the condition for the solubility of an ordinary second-order differential equation [Eq. (6.1)] for the z component of the electric field. In Sec. II we describe the unperturbed beam. In Sec. III we use assumptions (1.2)–(1.7) to calculate the perturbed electromagnetic fields and forces generated by a given perturbation in the beam charge density. The response of a beam particle to these forces is calculated in Sec. IV, and in Sec. V we use the results of Sec. IV to calculate the perturbed beam charge density. Putting this into the simplified field equation derived in Sec. III, we emerge in Sec. VI with our fundamental equation (6.1).

It becomes immediately apparent upon inspection of Eq. (6.1) that in each mode the dispersion relation has a basic scaling property, which can be expressed in the formula

$$\Omega^2 \equiv (\omega - kv)^2 = \omega_\beta^2 \lambda^2 (4\pi i \sigma a^2 \omega / c^2). \qquad (1.8)$$

The quantity λ^2 is for each mode a dimensionless function entirely determined by the *shape* of the beam particle number density $n(r)$ (and by the precise definitions chosen for a and ω_β) except that for a finite plasma channel radius R the λ^2 functions also depend on the fixed parameter a/R.

The second half (Secs. VII–XIII) of this paper is concerned with the construction of a catalog of modes and a detailed examination of the properties of the λ^2 functions in each mode, leading to an estimate of the forced and free growth rates. In Sec. VII we check[7] that the rigid beam "hose" dispersion relation derived in Ref. 4 for general beam shape emerges here as the limit of Eq. (1.8) for $\omega \ll c^2/4\pi\sigma a^2$ in the case $m = 1$. In Sec. VIII we specialize to the case of a uniform beam, and derive the dispersion relation as an implicit transcendental equation

$$\eta\left(\frac{J_m'(qa/\eta)}{J_m(qa/\eta)}\right) + \xi^2/qa = \frac{H_m^{(1)\prime}(qa)J_{|m-1|}(qR) - J_m'(qa)H_{|m-1|}^{(1)}(qR)}{H_m^{(1)}(qa)J_{|m-1|}(qR) - J_m(qa)H_{|m-1|}^{(1)}(qR)}, \qquad (1.9)$$

where a is the beam radius, R is the plasma channel radius, and

$$q^2 \equiv 4\pi i\sigma\omega/c^2 \quad (\mathrm{Im}q > 0),$$
$$\eta^2 \equiv 1 - \frac{2(4 - m^2 - \lambda^2)}{(4 - m^2 - \lambda^2)^2 - 4m^2\lambda^2},$$
$$\xi^2 \equiv \frac{2m^2(1 - \eta^2)}{\lambda^2 - m^2}\left(1 - \frac{2\lambda^2}{4 - m^2 - \lambda^2}\right),$$
$$\lambda^2 \equiv \Omega^2/\omega_\beta^2 \equiv (\omega - kv)^2/\omega_\beta^2.$$

Then in Sec. IX we catalog the solutions of this equation for λ^2 as a function of q^2. It is found that for each m there are generally two infinite sequences of modes A_{mn} and B_{mn} and two exceptional modes C_m and D_m, except that the modes B_{0n}, C_0, D_0, D_1, A_{2n}, C_2 are missing. The "hose" mode is C_1. The high- and low-frequency behaviors of λ^2 in all these modes is derived in Appendix B and by machine calculations,[8] and summarized in Sec. IX, Tables 1 to 3, and Figs. 1 to 5.

In Secs. X–XIII we return to the case of a general smooth beam shape. We show in Sec. X that there occurs here a continuous spectrum[9] similar to that found in Ref. 1 for the streaming modes. That is, there are real positive Ω^2 values [in bands given by Eqs. (10.12)–(10.16)] *each* of which allows a solution of Eq. (6.1) for *all* real or complex ω. The solutions contain logarithmic and step function singularities, but the location of the singularities depends on Ω^2, so that the fields are nonsingular when integrated over a range of Ω^2. The continuous spectrum includes only stable oscillation frequencies (i.e., Ω real) so it is of no practical importance here; this is in contrast with the streaming modes,[1] for which the continuous spectrum is unstable and is in fact the whole spectrum. In Sec. XI we treat the discrete modes, for which Ω *is* a function of ω, and we show that they correspond, more or less, to the A, B, C, and D modes found in Sec. IX for a uniform beam. The limits of Ω as $\omega \to \infty$ and (for A, B, and C_1 modes) as $\omega \to 0$ are shown to be the same as for a uniform beam, except that ω_β must be replaced with the particle gyration frequency at $r = 0$; however, the way that Ω approaches its limit as $\omega \to \infty$ depends sensitively on the beam shape. The analytic properties of the function $\Omega(\omega)$ are discussed in Sec. XII and used to construct approximate dispersion relations, which for the uniform beam compare rather well with exact computer results.

We conclude in Sec. XIII with a discussion of how

[7] The existence of a mode, for which $\Omega \to 0$ as $\omega \to 0$, can be inferred from the translational invariance of Maxwell's and Newton's equations, and the uniformity of the background plasma. The beam could be anywhere within the plasma channel, and its being at one position rather than another "breaks" this translational symmetry. There is a theorem, known in quantum field theory as the Goldstone theorem, which states that whenever the equations (i.e., the Lagrangian) of a system have a symmetry which is not shared by some solution of these equations (in our case, the unperturbed beam), this solution must admit perturbations (in our case, the hose mode) which allow arbitrarily small wavenumber and frequency. See J. Goldstone, Nuovo Cimento **19**, 154 (1961), and J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. **127**, 965 (1962).

[8] Computer results for the most important modes were very kindly supplied by S. C. Wright.

[9] We are very grateful to K. M. Case for pointing out the existence of a continuous spectrum in this problem. The continuous spectrum found here is very closely related to that of the streaming modes, discussed in Ref. 1.

the dispersion relation is to be used in calculating the number of e foldings under conditions of forced growth and free growth. In the former case ω is a fixed real tickling frequency, and λ^2 in Eq. (1.8) is complex, so there are two roots, one growing and one decaying. In all modes $\lambda^2(0)$ and $\lambda^2(\infty)$ are non-negative real numbers, so we may expect Im λ to reach its maximum when the argument of λ^2 is of order unity, at which point |Re λ| and |Im λ| are also of order unity. [Inspection of Eq. (1.8) then explains why (1.6) and (1.7) characterize the most rapidly growing modes.] Calculation[8] of the maximum growth rates for the uniform beam (with $R = 2a$) reveals that they all lie between $0.09\omega_\beta$ and $0.12\omega_\beta$, except that the $m = 1$ "A modes" have maximum growth rate $0.17\omega_\beta$, the $m = 0$ "sausage" modes have maximum growth rate $0.26\omega_\beta$, and last but not least, the $m = 1$ "hose" mode has maximum growth rate $0.29\omega_\beta$.

The saddle-point method is used at the end of Sec. XIII to treat the case of free growth. It is found that at a fixed distance $vt - z$ behind the disturbed part of the beam, the saddle point moves as $z \to \infty$ to $\omega \to \infty$, and the number of e foldings grows as

$$\# \propto z^{(N+1)/(2N+1)} \quad (N = 1, 2, \cdots)$$

if the beam density $n(r)$ approaches $n(0) + O(r^{2N})$ as $r \to 0$. This conclusion seems to indicate that for asymptotic free growth (or more generally, for all large ω) the instability is worse for a smooth beam shape (N small) than for a uniform beam (N large). However, it is necessary to take this conclusion with some reservations, because for a smooth beam shape the particles participating in the instability are limited for large ω to a small area near the beam axis, and the instability may not matter. A general treatment of this phenomenon (including the effects of noncircular orbits) seems called for, and may be the subject of a future article.

II. UNPERTURBED BEAM

The average velocity of the beam particles is in the z direction, with magnitude v, and their number density at a distance r from the beam axis is $n(r)$. Therefore the unperturbed magnetic field points in the $+\theta$ direction, with magnitude

$$B_{0\theta}(r) = \frac{4\pi ev}{rc}\int_0^r r'n(r')\,dr'. \tag{2.1}$$

We assume the plasma to maintain local charge neutrality, so the force on a beam particle is solely due to the magnetic field. Assuming the beam particles to move on circular helices, their velocity has no r component, so the magnetic force points in the $-r$ direction, with magnitude

$$-F_{0r}(r) = (ev/c)B_{0\theta}(r). \tag{2.2}$$

If the beam particles have angular frequency $\alpha(r)$, we must equate (2.2) to the relativistic centrifugal force $M\gamma\alpha^2(r)$, obtaining for the angular frequency

$$\alpha^2(r) = \frac{4\pi e^2v^2}{M\gamma c^2r^2}\int_0^r r'n(r')\,dr', \tag{2.3}$$

with

$$\gamma \equiv [1 - v^2/c^2]^{-\frac{1}{2}}.$$

For instance, if the beam density $n(r)$ is uniform, Eq. (2.3) gives just the familiar betatron frequency

$$\alpha^2(r) = \omega_\beta^2 \equiv 2\pi e^2v^2n/M\gamma c^2 \quad \text{(inside beam)}. \tag{2.4}$$

However, we do not restrict ourselves to the case of a uniform beam.

Equation (2.3) has two roots for $\alpha(r)$, corresponding to the possibility of both clockwise and anticlockwise helical orbits. We usually expect both to occur with equal probability, and so assume in this article, though it would be a trivial matter to extend our analysis to the case of a polarized beam.

III. PERTURBED FIELDS AND FORCES

Our unperturbed system has a translation symmetry in each of the coordinates θ, z, and t. Therefore we may assume that the cylindrical components of the perturbed fields $\mathbf{E}_1$, $\mathbf{B}_1$ and perturbed current $\mathbf{J}_1$ take the form

$$e^{im\theta}e^{ikz}e^{-i\omega t} \times \text{functions of } r. \tag{3.1}$$

Here, m is an integer, with $m = 0$ for the sausage mode, $m = \pm 1$ for the hose mode, etc.

The exact Maxwell equations can therefore be written as

$$(im/r)E_{1z} - ikE_{1\theta} = (i\omega/c)B_{1r}, \tag{3.2}$$

$$ikE_{1r} - E'_{1z} = (i\omega/c)B_{1\theta}, \tag{3.3}$$

$$(1/r)(rE_{1\theta})' - (im/r)E_{1r} = (i\omega/c)B_{1z}, \tag{3.4}$$

$$\frac{im}{r}B_{1z} - ikB_{1\theta} = \left(\frac{4\pi\sigma - i\omega}{c}\right)E_{1r} + \frac{4\pi}{c}J_{1r}, \tag{3.5}$$

$$ikB_{1r} - B'_{1z} = \left(\frac{4\pi\sigma - i\omega}{c}\right)E_{1\theta} + \frac{4\pi}{c}J_{1\theta}, \tag{3.6}$$

$$\frac{1}{r}(rB_{1\theta})' - \frac{im}{r}B_{1r} = \left(\frac{4\pi\sigma - i\omega}{c}\right)E_{1z} + \frac{4\pi}{c}J_{1z}. \tag{3.7}$$

(We have set the plasma current equal to $\sigma\mathbf{E}_1$; a prime means $\partial/\partial r$.) It is very convenient to use (3.2), (3.3),

(3.5), and (3.6) to express the transverse field components in terms of E_{1z}, B_{1z}, J_{1r}, and $J_{1\theta}$:

$$q^2E_{1r} = \frac{-m\omega}{rc}B_{1z} + ikE'_{1z} - \frac{4\pi i\omega}{c^2}J_{1r}, \quad (3.8)$$

$$q^2E_{1\theta} = \frac{-mk}{r}E_{1z} - \frac{i\omega}{c}B'_{1z} - \frac{4\pi i\omega}{c^2}J_{1\theta}, \quad (3.9)$$

$$q^2B_{1r} = \left(\frac{4\pi\sigma - i\omega}{c}\right)\left(\frac{im}{r}\right)E_{1z} + ikB'_{1z} + \frac{4\pi ik}{c}J_{1\theta}, \quad (3.10)$$

$$q^2B_{1\theta} = -\left(\frac{4\pi\sigma - i\omega}{c}\right)E'_{1z} - \frac{km}{r}B_{1z} - \frac{4\pi ik}{c}J_{1r}, \quad (3.11)$$

where q is the transverse wave number, defined by

$$q^2 = -k^2 + (i\omega/c^2)(4\pi\sigma - i\omega). \quad (3.12)$$

Equations (3.4) and (3.7) give decoupled equations for E_{1z} and B_{1z}:

$$\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r}E_{1z} - \frac{m^2}{r^2}E_{1z} + q^2E_{1z} = \frac{4\pi\omega}{c^2(q^2+k^2)}\left(-iq^2J_{1z} + \frac{imk}{r}J_{1\theta} + \frac{k}{r}\frac{\partial}{\partial r}rJ_{1r}\right), \quad (3.13)$$

$$\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r}B_{1z} - \frac{m^2}{r^2}B_{1z} + q^2B_{1z} = \frac{-4\pi}{rc}\left[-imJ_{1r} + \frac{\partial}{\partial r}rJ_{1\theta}\right]. \quad (3.14)$$

We now invoke the approximations discussed in the Introduction. Specifically, we look for modes with

$$|\omega|\, a \ll v, \quad (3.15)$$

$$|k|\, a \ll 1, \quad (3.16)$$

$$|q|\, a \sim 1, \quad (3.17)$$

where a is a characteristic beam radius and v is the unperturbed beam particle velocity z component. By Eq. (3.17) we do not necessarily mean to exclude the possibility that $|q|\, a \sim 0.01$ or $|q|\, a \sim 100$, but only to require that $|q|\, a$ is much closer to unity than is $|\omega|\, a/v$ or $|k|\, a$; however, the fastest growing modes in fact turn out to be those with $|q|\, a$ quite close to unity.

From (3.15)–(3.17) we see that $|q| \gg |k|$ and $|q| \gg |\omega/c|$, so (3.12) now gives q as

$$q^2 = 4\pi i\sigma\omega/c^2. \quad (3.18)$$

Note that (3.15) and (3.17) are consistent only if

$$c^2/4\pi\sigma av \ll 1. \quad (3.19)$$

We also find that $|\omega - kv|$ is of the order of the typical betatron frequency ω_β, so (3.15) and (3.16) are consistent only if

$$\omega_\beta a/v \ll 1. \quad (3.20)$$

We have already made assumptions (3.19) and (3.20) in Sec. I.

We next simplify the electrodynamic equations (3.8)–(3.14), using (3.15)–(3.20), and estimating

$$q \sim \partial/\partial r \sim 1/r \sim 1/a \quad (3.21)$$

everywhere in these equations. Our starting point is the tentative assumption that the perturbed beam particle velocity $\mathbf{v}_1$ has comparable r and θ components, and a z component at most comparable with these. This ansatz is verified later in this section; it actually turns out that v_{1z} is much less than v_{1r} and $v_{1\theta}$. The perturbed beam current components are

$$J_{1r} = env_{1r}, \quad (3.22)$$

$$J_{1\theta} = env_{1\theta}, \quad (3.23)$$

$$J_{1z} = env_{1z} + en_1v, \quad (3.24)$$

where n is the unperturbed beam density and n_1 is the perturbation to n. Charge conservation gives

$$+i\omega en_1 = \nabla \cdot \mathbf{J}_1$$

or

$$+i\Omega n_1 = \nabla \cdot (n\mathbf{v}_1), \quad (3.25)$$

where

$$\Omega \equiv \omega - kv. \quad (3.26)$$

Hence v_{1r} and $v_{1\theta}$ may be estimated as of order

$$v_{1\perp} \sim \Omega a n_1/n, \quad (3.27)$$

and so

$$J_{1z} \simeq evn_1, \quad (3.28)$$

the other term in (3.24) being smaller by at least a factor $|\Omega|\, a/v$. [Equations (3.15) and (3.16) imply $|\Omega|\, a \ll v$.] Also (3.27) lets us estimate the transverse perturbed beam currents (3.22) and (3.23) as being of order

$$J_{1r} \sim J_{1\theta} \sim e\Omega a n_1. \quad (3.29)$$

Formula (3.28) and the estimate (3.29) show that the J_{1r} and $J_{1\theta}$ terms on the right-hand side of Eq. (3.13) are smaller than the J_{1z} term by factors of order $(\Omega a/v)(ka)$, and hence may be neglected. Recalling that $|q| \gg |k|$, we may thus simplify Eq. (3.13) to read

$$\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r}E_{1z} - \frac{m^2}{r^2}E_{1z} + q^2E_{1z} = \frac{-4\pi iev\omega}{c^2}n_1. \quad (3.30)$$

We may also estimate E_{1z} as being of order

$$E_{1z} \sim (4\pi ev\omega a^2/c^2)n_1. \quad (3.31)$$

Using (3.29) and (3.14), we may estimate B_{1z} as being of order

$$B_{1z} \sim (4\pi e\Omega a^2/c)n_1. \tag{3.32}$$

Using (3.29), (3.31), and (3.32) in (3.8)–(3.11) yields the estimates

$$E_{1r} \sim E_{1\theta} \sim (4\pi e\omega\Omega a^3/c^2)n_1 \quad \text{or} \quad (4\pi e\omega kva^3/c^2)n_1, \tag{3.33}$$

$$B_{1r} \sim B_{1\theta} \sim (4\pi eva/c)n_1. \tag{3.34}$$

Also, (3.10) and (3.11) now simplify to

$$B_{1r} = (mc/\omega r)E_{1z}, \tag{3.35}$$

$$B_{1\theta} = (ic/\omega)E'_{1z}, \tag{3.36}$$

the terms neglected being smaller by a factor $ka(\Omega a/v)$.

We are now in a position to decide which are the important forces on a beam particle. A particle in an unperturbed orbit of radius r feels forces due to the perturbed electromagnetic fields, given by

$$F_{1r} = e[E_{1r} + (\alpha r/c)B_{1z} - (v/c)B_{1\theta}], \tag{3.37}$$

$$F_{1\theta} = e[E_{1\theta} + (v/c)B_{1r}], \tag{3.38}$$

$$F_{1z} = e[E_{1z} - (\alpha r/c)B_{1r}], \tag{3.39}$$

where α is the angular gyration frequency at r. We have already assumed that the typical value ω_β of α is much less than v/a. Our estimates (3.31)–(3.34) of the field magnitudes then lets us replace (3.37) and (3.38) by

$$F_{1r} = -(ev/c)B_{1\theta} = (-iev/\omega)E'_{1z}, \tag{3.40}$$

$$F_{1\theta} = (ev/c)B_{1r} = (evm/\omega r)E_{1z}, \tag{3.41}$$

the neglected terms being smaller by factors of order $(\omega a/c)(\Omega a/c)$, $(\omega a/c)(ka)(v/c)$, or $(\Omega a/v)(\omega_\beta a/v)$. And F_{1z} is smaller than F_{1r} and $F_{1\theta}$ by factors of order $(\omega a/v)$ or $(\omega_\beta a/v)$, justifying our previous statement that v_{1z} is smaller than v_{1r} and $v_{1\theta}$.

The important results of this section are the differential equation (3.30) for the E_{1z} generated by a given n_1, and the formulas (3.40) and (3.41) for the forces in terms of E_{1z}. To Eq. (3.30) we must also add a boundary condition on E_{1z} for $r \to \infty$. If the plasma conductivity σ stays constant everywhere, then outside the beam we must require

$$E_{1z} \propto H_m^{(1)}(qr), \quad \text{Im } q > 0. \tag{3.42}$$

The boundary condition for finite plasma channel radius is worked out in Appendix A.

IV. PERTURBED BEAM PARTICLE ORBITS

The unperturbed orbit of a beam particle can be characterized by three parameters, the radius r and the values φ and h of θ and z at $t = 0$. The corresponding perturbed orbit is then given by equations

$$\tilde{r}(t; r\varphi h) = r + D_r(t; r\varphi h), \tag{4.1}$$

$$\tilde{\theta}(t; r\varphi h) = \alpha(r)t + \varphi + r^{-1}D_\theta(t; r\varphi h), \tag{4.2}$$

$$\tilde{z}(t; r\varphi h) = vt + h. \tag{4.3}$$

Here D_r and D_θ are infinitesimal perturbations; we do not perturb z because we have already seen that the perturbed force is almost purely transverse.

The exact equations of motion are

$$\ddot{\tilde{r}} - \dot{\tilde{\theta}}^2\tilde{r} = F_r(\tilde{r}, \tilde{\theta}, \tilde{z}, t)/M\gamma, \tag{4.4}$$

$$\tilde{r}\ddot{\tilde{\theta}} + 2\dot{\tilde{r}}\dot{\tilde{\theta}} = F_\theta(\tilde{r}, \tilde{\theta}, \tilde{z}, t)/M\gamma, \tag{4.5}$$

with $M\gamma$ the relativistic beam particle mass. The zeroth-order solution has already been discussed in Sec. II. The first-order terms give

$$\ddot{D}_r(t, r\varphi h) - \alpha^2(r)D_r(t; r\varphi h) - 2\alpha(r)\dot{D}_\theta(t; r\varphi h) = -[r\alpha^2(r)]'D_r(t; r\varphi h) + (M\gamma)^{-1}F_{1r}[r, \alpha(r)t + \varphi, vt + h, t], \tag{4.6}$$

$$D_\theta(t; r\varphi h) + 2\alpha(r)\dot{D}_r(t; r\varphi h) = (M\gamma)^{-1}F_{1\theta}[r, \alpha(r)t + \varphi, vt + h, t]. \tag{4.7}$$

On the right-hand side of Eq. (4.6) we have used the fact that the unperturbed radial force is $-M\gamma r\alpha^2(r)$.

We have already observed that the perturbed forces have the functional form

$$F_{1n}(r, \theta, z, t) = c^{im\theta}e^{ikz}e^{-i\omega t}\mathcal{F}_n(r) \tag{4.8}$$

with $n = r$ or θ, and $\mathcal{F}_r(r)$ and $\mathcal{F}_\theta(r)$ two infinitesimal functions of r. Then

$$F_{1n}[r, \alpha(r)t + \varphi, vt + h, t] = e^{-i[\Omega - m\alpha(r)]t}e^{ikh}e^{im\varphi}\mathcal{F}_n(r), \tag{4.9}$$

where again

$$\Omega \equiv \omega - kv. \tag{4.10}$$

We may therefore conclude that the r and θ components of the beam displacement take the form

$$D_n(t; r, \varphi, h) = e^{-i[\Omega - m\alpha(r)]t}e^{ikh}e^{im\varphi}\mathcal{D}_n(r). \tag{4.11}$$

Inserting (4.9) and (4.11) in (4.6) and (4.7), we find

$$\{-[\Omega - m\alpha(r)]^2 + r\alpha^{2'}(r)\}\mathcal{D}_r(r) + 2i\alpha(r)[\Omega - m\alpha(r)]\mathcal{D}_\theta(r) = (M\gamma)^{-1}\mathcal{F}_r(r),$$

$$-2i\alpha(r)[\Omega - m\alpha(r)]\mathcal{D}_r(r) - [\Omega - m\alpha(r)]^2\mathcal{D}_\theta(r) = (M\gamma)^{-1}\mathcal{F}_\theta(r).$$

The solution of these equations is

$$M\gamma\mathcal{D}_r(r) = \frac{\mathcal{F}_r(r) + 2i\alpha(r)\mathcal{F}_\theta(r)/[\Omega - m\alpha(r)]}{4\alpha^2(r) - [\Omega - m\alpha(r)]^2 + r\alpha^{2'}(r)}, \tag{4.12}$$

$$M\gamma\mathcal{D}_\theta(r) = \frac{\left[\begin{array}{l}-2i\alpha(r)\mathcal{F}_r(r)/[\Omega - m\alpha(r)] \\ \quad + (1 - \{r\alpha^{2'}(r)/[\Omega - m\alpha(r)]^2\})\mathcal{F}_\theta(r)\end{array}\right]}{4\alpha^2(r) - [\Omega - m\alpha(r)]^2 + r\alpha^{2'}(r)}. \tag{4.13}$$

The forces are given in Sec. III in terms of the perturbed electric field E_{1z}, which takes the form

$$E_{1z}(r, \theta, z, t) = e^{im\theta}e^{ikz}e^{-i\omega t}\mathcal{E}(r). \tag{4.14}$$

Hence from (3.40), (3.41), (4.8), and (4.14) we have

$$\mathcal{F}_r(r) = (-iev/\omega)\mathcal{E}'(r), \tag{4.15}$$

$$\mathcal{F}_\theta(r) = (evm/\omega r)\mathcal{E}(r). \tag{4.16}$$

Using (4.15) and (4.16) in (4.12) and (4.13) gives

$$\mathfrak{D}_r(r) = \frac{\dfrac{-iev}{M\gamma\omega}\left(\mathcal{E}'(r) - \dfrac{2m\alpha(r)}{\Omega - m\alpha(r)}\, r^{-1}\mathcal{E}(r)\right)}{4\alpha^2(r) - [\Omega - m\alpha(r)]^2 + r\alpha^{2\prime}(r)}, \tag{4.17}$$

$$\mathfrak{D}_\theta(r) = \frac{\left[\dfrac{ev}{M\gamma\omega}\left\{\dfrac{-2\alpha(r)}{\Omega - m\alpha(r}\,\mathcal{E}'(r) + \left[1 - \dfrac{r\alpha^{2\prime}(r)}{(\Omega - m\alpha(r)]^2}\right]\left(\dfrac{m}{r}\right)\mathcal{E}(r)\right\}\right]}{4\alpha^2(r) - [\Omega - m\alpha(r)]^2 + r\alpha^{2\prime}(r)}. \tag{4.18}$$

V. PERTURBED BEAM DENSITY

The driving term for the perturbed fields was shown in Eq. (3.30) to be the perturbation n_1 in the beam particle density. We must now express n_1 in terms of the beam displacement functions found in the last section.

The total perturbed beam density $\tilde{n} = n + n_1$ is given by integrating the number $n(\imath)\imath\, d\imath\, d\varphi\, dh$ of beam particles between $\imath$, φ, h, and $\imath + d\imath$, $\varphi + d\varphi$, $h + dh$, times appropriate δ functions:

$$r\tilde{n}(r\theta zt) = \int n(\imath)\imath\, d\imath\, d\varphi\, dh\delta[\tilde{r}(t; \imath\varphi h) - r] \times \delta[\tilde{\theta}(t; \imath\varphi h) - \theta]\delta[\tilde{z}(t; \imath\varphi h) - z] \tag{5.1}$$

with $\tilde{r}$, $\tilde{\theta}$, and $\tilde{z}$ given by (4.1), (4.2), and (4.3). To first order in the $\mathfrak{D}$'s, the δ functions give

$$\imath = r - \mathfrak{D}_r[t; r, \theta - \alpha(r)t, z - vt], \tag{5.2}$$

$$\varphi = \theta - \alpha(r)t + \alpha'(r)t\mathfrak{D}_r[t; r, \theta - \alpha(r)t, z - vt] - r^{-1}\mathfrak{D}_\theta[t; r, \theta - \alpha(r)t, z - vt], \tag{5.3}$$

$$h = z - vt \tag{5.4}$$

or using (4.11)

$$\imath = r - e^{-i\omega t}e^{im\theta}e^{ikz}\mathfrak{D}_r(r), \tag{5.5}$$

$$\varphi = \theta - \alpha(r)t - e^{-i\omega t}e^{im\theta}e^{ikz} \times [r^{-1}\mathfrak{D}_\theta(r) - \alpha'(r)t\mathfrak{D}_r(r)]. \tag{5.6}$$

The Jacobian of the transformation $r\theta z \to \imath\varphi h$ is the determinant

$$\left|\frac{\partial(r\varphi h)}{\partial(r\theta z)}\right| = \begin{vmatrix} 1 - e^{(\ )}\mathfrak{D}_r'(r) & -ime^{(\ )}\mathfrak{D}_r(r) & -ike^{(\ )}\mathfrak{D}_\theta(r) \\ -\alpha'(r)t - e^{(\ )}\mathfrak{D}_{r\theta}'(r) & 1 - ime^{(\ )}\mathfrak{D}_{r\theta}(r) & -ike^{(\ )}\mathfrak{D}_{r\theta}(r) \\ 0 & 0 & 1 \end{vmatrix},$$

where we have used the abbreviations

$$e^{(\ )} = e^{-i\omega t}e^{ikz}e^{im\theta},$$

$$\mathfrak{D}_{r\theta}(r) = r^{-1}\mathfrak{D}_\theta(r) - \alpha'(r)t\mathfrak{D}_r(r).$$

Keeping only terms of first order in $\mathfrak{D}$, we find that the $t\alpha'(r)$ terms cancel, giving

$$\left|\frac{\partial(\imath\varphi h)}{\partial(r\theta z)}\right| = 1 - e^{ikz}e^{im\theta}e^{-i\omega t}\left\{\mathfrak{D}_r'(r) + \frac{im}{r}\,\mathfrak{D}_\theta(r)\right\}. \tag{5.7}$$

Applying (5.5)–(5.7) to (5.1), we find the beam density

$$\begin{aligned} r\tilde{n}(r\theta zt) &= \{rn(r) - [rn(r)]'e^{ikz}e^{im\theta}e^{-i\omega t}\mathfrak{D}_r(r)\}\left|\frac{\partial(\imath\varphi h)}{\partial(r\theta z)}\right| \\ &= rn(r) - e^{ikz}e^{im\theta}e^{-i\omega t} \\ &\quad \times \{[rn(r)\mathfrak{D}_r(r)]' + imn(r)\mathfrak{D}_\theta(r)\}. \end{aligned} \tag{5.8}$$

The perturbation in n is hence of the form

$$n_1(r\theta zt) = e^{ikz}e^{im\theta}e^{-i\omega t}\mathcal{N}(r) \tag{5.9}$$

with

$$\mathcal{N}(r) = -(1/r)[rn(r)\mathfrak{D}_r(r)]' - (im/r)n(r)\mathfrak{D}_\theta(r). \tag{5.10}$$

This result, that $n_1 = -\nabla \cdot (\mathbf{D}n)$, is certainly one that might have been guessed from conservation considerations. However, it is difficult to find any simpler proof of (5.10) than the one given here.

Using (4.17) and (4.18) in (5.10), we find

$$\mathcal{N}(r) = \frac{-iev}{M\gamma\omega}\left\{-\frac{1}{r}\frac{d}{dr}\frac{rn(r)}{F(r)}\frac{d}{dr}\mathcal{E}(r) + \frac{m^2}{r^2}\frac{n(r)}{F(r)}\mathcal{E}(r) + \frac{2m}{r[\Omega - m\alpha(r)]}\left(\frac{\alpha(r)n(r)}{F(r)}\right)'\mathcal{E}(r)\right\}, \tag{5.11}$$

where $F(r)$ is the denominator in (4.17) and (4.18):

$$F(r) \equiv 4\alpha^2(r) - [\Omega - m\alpha(r)]^2 + r\alpha^{2\prime}(r). \tag{5.12}$$

It is convenient to use the static equilibrium condition (2.3) to express $n(r)$ in terms of $\alpha(r)$:

$$(4\pi e^2v^2/M\gamma c^2)rn(r) = [r^2\alpha^2(r)]'. \tag{5.13}$$

This gives

$$\mathcal{N}(r) = \frac{ic^2}{4\pi ev\omega}\left\{\frac{1}{r}\frac{d}{dr}rf_+(r)\frac{d}{dr}\mathcal{E}(r) - \frac{m^2}{r^2}f_+(r)\mathcal{E}(r) - g_+(r)\mathcal{E}(r)\right\}, \tag{5.14}$$

where

$$f_+(r) \equiv \frac{2\alpha^2(r) + r\alpha^{2\prime}(r)}{4\alpha^2(r) - [\Omega - m\alpha(r)]^2 + r\alpha^{2\prime}(r)}, \tag{5.15}$$

$$g_+(r) \equiv \frac{2m}{r[\Omega - m\alpha(r)]} \times \left(\frac{\alpha(r)[2\alpha^2(r) + r\alpha^{2\prime}(r)]}{4\alpha^2(r) - [\Omega - m\alpha(r)]^2 + r\alpha^{2\prime}(r)}\right)'. \tag{5.16}$$

This answer is correct as it stands only in the unlikely event that all beam particles have the same sign for the rotational frequency $\alpha(r)$. It is much more realistic to suppose that clockwise and anticlockwise orbits are equally common (at each r), in which case the beam density perturbation $\mathcal{N}(r)$ must be averaged over the two signs of $\alpha(r)$. We then find

$$\mathcal{N}(r) = \frac{ic^2}{4\pi ev\omega}\left\{\frac{1}{r}\frac{d}{dr} rf(r)\frac{d}{dr}\,\mathcal{E}(r) - \frac{m^2}{r^2} f(r)\mathcal{E}(r) - g(r)\mathcal{E}(r)\right\}, \tag{5.17}$$

where

$$f(r) = \tfrac{1}{2}[f_+(r) + f_-(r)] = \frac{[r\alpha^{2'}(r) + 2\alpha^2(r)][(4-m^2)\alpha^2(r) - \Omega^2 + r\alpha^{2'}(r)]}{[(4-m^2)\alpha^2(r) - \Omega^2 + r\alpha^{2'}(r)]^2 - 4m^2\Omega^2\alpha^2(r)}, \tag{5.18}$$

$$g(r) = \tfrac{1}{2}[g_+(r) + g_-(r)] = \frac{-4\Omega^2m^2}{r[\Omega^2 - m^2\alpha^2(r)]} \times \left\{\frac{\alpha^2(r)f(r)}{(4-m^2)\alpha^2(r) - \Omega^2 + r\alpha^{2'}(r)}\right\}' + \frac{2m^2\alpha(r)}{r[\Omega^2 - m^2\alpha^2(r)]}\{\alpha(r)f(r)\}' \tag{5.19}$$

the functions $f_-(r)$ and $g_-(r)$ being given by replacing $\alpha(r)$ by $-\alpha(r)$ in (5.15) and (5.16). [It would also be possible to treat a "polarized" beam by using different weights for f_+, g_+ and f_-, g_- in (5.18) and (5.19).]

VI. FUNDAMENTAL EQUATIONS

Using (4.14) and (5.9) let us write the field equation (3.30) for E_{1z} as

$$\frac{1}{r}\frac{d}{dr} r\frac{d}{dr}\,\mathcal{E}(r) - \frac{m^2}{r^2}\,\mathcal{E}(r) + q^2\mathcal{E}(r) = -4\pi ie\omega\,\frac{v}{c^2}\,\mathcal{N}(r).$$

And (5.17) gives $\mathcal{N}(r)$ in terms of $\mathcal{E}(r)$, so the circle is closed, and we emerge with a second-order ordinary differential equation for $\mathcal{E}(r)$:

$$\frac{1}{r}\frac{d}{dr} r[1 - f(r)]\frac{d}{dr}\,\mathcal{E}(r) - \frac{m^2}{r^2}[1 - f(r)]\mathcal{E}(r) + q^2\mathcal{E}(r) + g(r)\mathcal{E}(r) = 0. \tag{6.1}$$

For the reader's convenience, we repeat that $f(r)$ and $g(r)$ are functions describing the beam response to the field $\mathcal{E}$:

$$f(r) = \frac{[r\alpha^{2'}(r) + 2\alpha^2(r)][(4-m^2)\alpha^2(r) - \Omega^2 + r\alpha^{2'}(r)]}{[(4-m^2)\alpha^2(r) - \Omega^2 + r\alpha^{2'}(r)]^2 - 4m^2\Omega^2\alpha^2(r)}, \tag{5.18}$$

$$g(r) = \frac{-4m^2\Omega^2}{r[\Omega^2 - m^2\alpha^2(r)]}\left\{\frac{\alpha^2(r)f(r)}{(4-m^2)\alpha^2(r) - \Omega^2 + r\alpha^{2'}(r)}\right\}' + \frac{2m^2\alpha(r)}{r[\Omega^2 - m^2\alpha^2(r)]}\{\alpha(r)f(r)\}', \tag{5.19}$$

with $\alpha(r)$ the beam particle gyration frequency.

The solutions of (6.1) behave like r^{-m} and r^{+m} for $r \to 0$ (or, for $m = 0$, like $\ln r$ and 1), and we must of course pick the solution that goes to $r^{|m|}$ for $r \to 0$. The boundary condition at the plasma channel radius R is shown in Appendix A to be

$$\mathcal{E}'(R)/\mathcal{E}(R) = -|m|/R. \tag{6.2}$$

[For $R = \infty$ we just pick the solution that behaves like $H^{(1)}_{|m|}(qr)$ outside the beam.] The consistency of these requirements on the behavior of $\mathcal{E}(r)$ at $r = 0$ and $r = R$ imposes a condition relating q^2 and Ω^2, which of course is the dispersion relation we seek.

Inspection of Eq. (6.1) reveals that the dispersion relation has the form promised in Sec. I:

$$\Omega^2 = \omega_\beta^2\lambda^2(q^2a^2), \tag{1.8}$$

where a is some characteristic beam radius and ω_β is a typical value of $\alpha(r)$. The circumstance that only Ω^2 rather than Ω appears in the dispersion relation (which implies that there is always one growing and one decaying mode) arises because we average over the two directions of beam polarization in Eq. (5.17). For the same reason the dispersion relation depends only upon m^2 rather than m. Had we allowed unequal numbers of clockwise and anticlockwise beam particle orbits we should have found the dispersion relation to depend not only on Ω^2 and m^2, but also on the *relative* sign of Ω and m; hence for each growing or decaying mode with $m > 0$ there would be a decaying or growing mode with opposite m and Ω.

Since the dispersion relation does not depend on the sign of m, we save writing below by always taking m positive, $m \geq 0$.

VII. RIGID BEAM LIMIT

There is one special case where (6.1) can be solved exactly for arbitrary beam shape. With $\Omega = 0$, $q = 0$, and $m = 1$ the functions $f(r)$ and $g(r)$ are given by (5.18), (5.19) as

$$f(r) = f_0(r) \equiv [r\alpha^{2'}(r) + 2\alpha^2(r)]/[r\alpha^{2'}(r) + 3\alpha^2(r)], \tag{7.1}$$

$$g(r) = g_0(r) \equiv [-2/r\alpha(r)][\alpha(r)f_0(r)]'. \tag{7.2}$$

It is straightforward to check that (6.1) is then satisfied exactly by

$$\mathcal{E}(r) = \mathcal{E}_0(r) \equiv Ar\alpha^2(r) \quad (A \text{ const}). \tag{7.3}$$

Note that outside the beam (2.3) and (7.3) give $\mathcal{E}(r) \propto 1/r$, in agreement with the requirement that

$$\mathcal{E}'(R)/\mathcal{E}(R) = -1/R.$$

Also, (7.3) obeys the condition that for $r \to 0$

$$\mathcal{E}(r) \propto r.$$

The meaning of the solution (7.3) can be understood by referring back to the formulas (4.17) and (4.18)

for the beam displacement functions $\mathfrak{D}_n(r)$. With $\Omega = 0$ and $m = 1$ they give

$$\mathfrak{D}_r(r) = \frac{-iev}{M\gamma\omega}\frac{\mathcal{E}'(r) + (2/r)\mathcal{E}(r)}{3\alpha^2(r) + r\alpha^{2\prime}(r)}, \tag{7.4}$$

$$\mathfrak{D}_\theta(r) = \frac{ev}{M\gamma\omega}\frac{2\mathcal{E}'(r) + [r^{-1} - \alpha^{2\prime}(r)/\alpha^2(r)]\mathcal{E}'(r)}{3\alpha^2(r) + r\alpha^{2\prime}(r)}. \tag{7.5}$$

Using (7.3) we find

$$\mathfrak{D}_r(r) = d, \tag{7.6}$$

$$\mathfrak{D}_\theta(r) = id, \tag{7.7}$$

where

$$d \equiv -ievA/(M\gamma\omega). \tag{7.8}$$

The Cartesian components of the beam particle displacements are then given by (7.6), (7.7), and (4.11) [with $\omega = k = 0$] as

$$\begin{aligned} D_x &= e^{i\alpha(r)t}e^{i\varphi}\{\mathfrak{D}_r(r)\cos[\alpha(r)t + \varphi] \\ &\quad + \mathfrak{D}_\theta(\theta)\sin[\alpha(r)t + \varphi]\} \\ &= d, \end{aligned} \tag{7.9}$$

$$\begin{aligned} D_y &= e^{i\alpha(r)t}e^{i\varphi}\{-\mathfrak{D}_r(r)\sin[\alpha(r)t + \varphi] \\ &\quad + \mathfrak{D}_\theta(\theta)\cos[\alpha(r)t + \varphi]\} \\ &= id. \end{aligned} \tag{7.10}$$

We therefore conclude that this solution corresponds to the rigid displacement[7] of the beam in the direction $(1, +i, 0)$. The mode $m = -1$ would correspond to displacement in the $(1, -i, 0)$ direction. Equation (7.8) just expresses the fact that the perturbed electric field vanishes like ω for $\omega \to 0$. [To be honest, the solution (7.3) was guessed by requiring that (7.4) and (7.5) give constants.]

Now that we have a zeroth-order solution we can do perturbation theory to obtain a relation between Ω^2 and q^2 when both are small. Let the solution in this case be written

$$\mathcal{E}(r) = \mathcal{E}_0(r) + \Omega^2\epsilon(r) + \cdots \tag{7.11}$$

and expand $f(r)$ and $g(r)$ in powers of Ω^2

$$f(r) = f_0(r) + \Omega^2\Phi(r) + \cdots, \tag{7.12}$$

$$g(r) = g_0(r) + \Omega^2\Gamma(r) + \cdots, \tag{7.13}$$

where $f_0(r)$, $g_0(r)$, and $\mathcal{E}_0(r)$ given by (7.1)–(7.3), and

$$\Phi(r) = \frac{[2\alpha^2(r) + r\alpha^{2\prime}(r)][7\alpha^2 + r\alpha^{2\prime}(r)]}{[3\alpha^2(r) + r\alpha^{2\prime}(r)]^3}, \tag{7.14}$$

$$\begin{aligned} \Gamma(r) = {} & \frac{4}{r\alpha^2(r)}\left[\frac{\alpha^2(r)[2\alpha^2(r) + r\alpha^{2\prime}(r)]}{[3\alpha^2(r) + r\alpha^{2\prime}(r)]^2}\right]' \\ & - \frac{2}{r\alpha^3(r)}\left[\frac{\alpha(r)[2\alpha^2(r) + r\alpha^{2\prime}(r)]}{[3\alpha^2(r) + r\alpha^{2\prime}(r)[}\right]' - \frac{2}{r\alpha(r)} \\ & \times \left[\frac{\alpha(r)[2\alpha^2(r) + r\alpha^{2\prime}(r)][7\alpha^2(r) + r\alpha^{2\prime}(r)]}{[3\alpha^2(r) + r\alpha^{2\prime}(r)]^3}\right]'. \end{aligned} \tag{7.15}$$

Keeping only terms of first order in Ω^2 and q_z, the differential equation (6.1) gives

$$\begin{aligned} & \frac{1}{r}\frac{d}{dr}r[1 - f_0(r)]\frac{d}{dr}\epsilon(r) - \frac{1}{r^2}[1 - f_0(r)]\epsilon(r) + g_0(r)\epsilon(r) \\ & = \frac{1}{r}\frac{d}{dr}r\Phi(r)\frac{d}{dr}\mathcal{E}_0(r) - \frac{1}{r^2}\Phi(r)\mathcal{E}_0(r) \\ & \quad - \Gamma(r)\mathcal{E}_0(r) - \left(\frac{q^2}{\Omega^2}\right)\mathcal{E}_0(r). \end{aligned} \tag{7.16}$$

Multiply by $r\mathcal{E}_0(r)$, and integrate from $r = 0$ to the plasma channel radius R. Since $\mathcal{E}_0(r)$ obeys the zeroth-order equation we find

$$\begin{aligned} & \int_0^R r\mathcal{E}_0(r)\left\{\frac{1}{r}\frac{d}{dr}r\Phi(r)\frac{d}{dr} - \frac{1}{r^2}\Phi(r) - \Gamma(r) - \left(\frac{q^2}{\Omega^2}\right)\right\}\mathcal{E}_0(r) \\ & = \int_0^R\left\{\mathcal{E}_0(r)\frac{d}{dr}[1 - f_0(r)]\frac{d}{dr}\epsilon(r)\right. \\ & \quad \left. - \epsilon(r)\frac{d}{dr}[1 - f_0(r)]\frac{d}{dr}\mathcal{E}_0(r)\right\}dr \\ & = [1 - f_0(R)]\{\mathcal{E}_0(R)\epsilon'(R) - \epsilon(R)\mathcal{E}_0'(R)\}. \end{aligned}$$

The right side vanishes because $\mathcal{E}(r)$ is subject to the boundary condition

$$\begin{aligned} -\frac{1}{R} &= \frac{\mathcal{E}_0'(R) + \Omega^2\epsilon'(R)}{\mathcal{E}_0(R) + \Omega^2\epsilon(R)} \\ &\cong -\frac{1}{R} + \frac{\Omega^2}{\mathcal{E}_0^2(r)}[\epsilon'(R)\mathcal{E}_0(R) - \epsilon(R)\mathcal{E}_0'(R)]. \end{aligned}$$

The dispersion relation thus becomes very simple

$$\Omega^2 = -q^2U^2, \tag{7.17}$$

where U^2 is a real constant defined by

$$U^2 = \frac{\int_0^R \mathcal{E}_0^2(r)r\,dr}{\int_0^R [\mathcal{E}_0'^2(r)\Phi(r) + r^{-2}\mathcal{E}_0^2(r)\Phi(r) + \Gamma(r)\mathcal{E}_0^2(r)]r\,dr}. \tag{7.18}$$

We have done an integration by parts, using the fact that $\Phi(r)$ vanishes outside the beam.

The integral of $\Gamma\mathcal{E}_0^2 r$ can be re-expressed using integration by parts as

$$\begin{aligned} \int_0^R \Gamma(r)\mathcal{E}_0^2(r)r\,dr &= A^2\int_0^R r\,dr \\ & \times \frac{[2\alpha^2(r) + r\alpha^{2\prime}(r)][40\alpha^6(r) + 38r\alpha^4(r)\alpha^{2\prime}(r) + 9r^2\alpha^2(r)\alpha^{2\prime}(r)^2 + r^3\alpha^{2\prime}(r)^3]}{[3\alpha^2(r) + r\alpha^{2\prime}(r)]^3}. \end{aligned}$$

The other terms in the denominator give

$$\begin{aligned} \int_0^R \Phi(r)\{\mathcal{E}_0'^2(r) + r^{-2}\mathcal{E}_0^2(r)\}r\,dr &= A^2\int_0^R r\,dr \\ & \times \frac{[\alpha^2(r) + r\alpha^{2\prime}(r)][14\alpha^6(r) + 16r\alpha^4(r)\alpha^{2\prime}(r) + 9r^2\alpha^2(r)\alpha^{2\prime}(r)^2 + r^3\alpha^{2\prime}(r)^3]}{[3\alpha^2(r) + r\alpha^{2\prime}(r)]^3}. \end{aligned}$$

A marvelous cancellation gives at last

$$U^2 = \frac{1}{2}\int_0^R \alpha^4(r) r^3\, dr \Big/ \int_0^R [r^2\alpha^2(r)]'\, dr. \quad (7.19)$$

This can be rewritten in a more useful form by using the fact that outside the beam

$$\alpha^2(r) = u^2/r^2 \text{ (outside)}, \quad (7.20)$$

where

$$u^2 \equiv \frac{4\pi e^2 v^2}{M\gamma c^2}\int_0^\infty r n(r)\, dr. \quad (7.21)$$

Therefore

$$\int_0^R [r^2\alpha^2(r)]'\, dr = u^2,$$

$$\int_0^R \alpha^4(r) r^3\, dr = u^2 \ln R - \int_0^R \ln r\, d(\alpha^4 r^4).$$

The last integral is actually R-independent since $d(r^4\alpha^4)$ vanishes outside the beam. We can thus define a characteristic beam radius r_c by

$$u^4 \ln r_c \equiv \int_0^\infty \ln r\, d(r^4\alpha^4), \quad (7.22)$$

so (7.19) becomes

$$U^2 = \tfrac{1}{2}u^2 \ln (R/r_c) \quad (7.23)$$

and the dispersion relation (7.17) reads

$$\Omega^2 = -\tfrac{1}{2}q^2u^2 \ln (R/r_c). \quad (7.24)$$

This is in complete agreement with the results of Ref. 4 [Eqs. (8.12) and (8.15)] if we note that the quantities called $\omega_\beta^2 a^2$ and $2r_0/C$ in Ref. 4 are identical with the u^2 and r_c introduced here. In Ref. 4 we *assumed* that the beam moves rigidly, and we are not surprised to see that this indeed gives the correct dispersion relation for small Ω and q, but is reassuring to have this result proved. Also, we are encouraged by this example to have faith in the rather complicated formulas (5.18), (5.19) for the coefficients in our fundamental equation (6.1).

VIII. UNIFORM BEAM: GENERAL DISPERSION RELATION

In order to get quantitative results in more general cases than discussed in the last section, it is necessary to choose the beam shape so as to simplify the function $f(r)$ and $g(r)$ as much as possible. The simplest choice is the uniform beam

$$n(r) = \begin{cases} n & \text{for } r < a, \\ 0 & r > a. \end{cases} \quad (8.1)$$

Then (2.3) gives the rotation frequency $\alpha(r)$ as

$$\alpha^2(r) = \begin{cases} \omega_\beta^2 & \text{for } r < a, \\ \omega_\beta^2 a^2/r^2 & r > a, \end{cases} \quad (8.2)$$

where ω_β^2 is the betatron frequency

$$\omega_\beta^2 = 2\pi n e^2 v^2/M\gamma c^2. \quad (8.3)$$

The functions $f(r)$ and $g(r)$ defined by Eqs. (5.18) and (5.19) are therefore

$$f(r) = \begin{cases} 1 - \eta^2 & \text{for } r < a, \\ 0 & r > a, \end{cases} \quad (8.4)$$

and

$$g(r) = -\xi^2 \delta(r - a)/a, \quad (8.5)$$

where

$$\eta^2 \equiv 1 - \frac{2(4 - m^2 - \lambda^2)}{(4 - m^2 - \lambda^2)^2 - 4m^2\lambda^2}, \quad (8.6)$$

$$\xi^2 \equiv \frac{2m^2(1 - \eta^2)}{\lambda^2 - m^2}\left(1 - \frac{2\lambda^2}{4 - m^2 - \lambda^2}\right), \quad (8.7)$$

and

$$\lambda^2 \equiv \Omega^2/\omega_\beta^2. \quad (8.8)$$

Inspection of Eq. (6.1) and its boundary conditions tells us immediately that the electric field here is

$$\mathcal{E}(r) \propto \begin{cases} J_m(qr/\eta), & \text{for } r < a, \\ H_m^{(1)}(qr) - \alpha_m(qR)J_m(qr), & r > a. \end{cases} \quad (8.9)$$

The coefficient α_m is determined by the properties of the plasma channel; if the channel is infinite then α_m is zero, while for a finite uniform plasma channel with radius R we show in Appendix A that

$$\alpha_m(qR) = H_{m-1}^{(1)}(qR)/J_{m-1}(qR) \quad \text{for} \quad m > 0, \quad (8.10)$$

or for $m = 0$

$$\alpha_0(qR) = H_1^{(1)}(qR)/J_1(qR).$$

Also, q is the root of $q^2 = 4\pi i\sigma\omega/c^2$ with positive imaginary part.

Our dispersion relation comes from the equations connecting the solutions for $r < a$ and $r > a$. We note first that $\mathcal{E}(r)$ must be continuous at $r = a$ because it is the tangential component of an electric field. The jump condition on $\mathcal{E}'(r)$ is given by integrating (6.1) from $r = a - \epsilon$ to $r = a + \epsilon$;

$$\mathcal{E}'(a + \epsilon) - \eta^2\mathcal{E}'(a - \epsilon) - \xi^2\mathcal{E}(a)/a = 0$$

or

$$\eta^2(\mathcal{E}'/\mathcal{E})_{a-\epsilon} + \xi^2/a = (\mathcal{E}'/\mathcal{E})_{a+\epsilon}. \quad (8.11)$$

Imposing this condition on (8.9) yields the dispersion relation

$$\eta\left(\frac{J_m'(qa/\eta)}{J_m(qa/\eta)}\right) + \frac{\xi^2}{qa} = \frac{H_m^{(1)\prime}(qa) - \alpha_m(qR)J_m'(qa)}{H_m^{(1)}(qa) - \alpha_m(qR)J_m(qa)}. \quad (8.12)$$

We remind the reader that ξ^2 and η^2 are given in terms of Ω^2 by (8.6)–(8.8), and $\alpha_m(qR)$ is given by (8.10).

IX. UNIFORM BEAM: CATALOG OF MODES

The uniform-beam dispersion relation (8.12) is analyzed in detail in Appendix B; this section gives only the results.

We find that for each m the modes may be conveniently classified into two infinite sequences, which we call the A_{mn} and B_{mn} modes, and two individual modes called C_m and D_m. The only exceptions are that the A modes are missing for $m = 2$, the B modes are missing for $m = 0$, the C mode is missing for $m = 0$ and $m = 2$, and the D mode is missing for $m = 1$. That is, the only modes with $m \leqq 2$ are

$$A_{0n}, A_{1n}, B_{1n}, C_1, B_{2n}, D_2$$

but all modes of types A, B, C, and D are present for $m \geqq 3$. By "mode" here we mean a value of Ω^2; for each such "mode" there are of course two roots for Ω, one growing and one decaying, or both oscillating. The "hose" mode is C_1.

A_{mn} ($m \neq 2$)

For $|qa| \ll 1$ we find

$$\frac{\Omega^2}{\omega_\beta^2} \to 3 + m^2 - (12m^2+1)^{\frac{1}{2}} - q^2a^2\left(1 - \frac{2m^2-1}{(12m^2+1)^{\frac{1}{2}}}\right)\Big/ j_{mn}^2 + \cdots \quad (9.1)$$

with j_{mn} the nth positive zero of $J_m(x)$. The only exception is for $m = 0$, where

$$\Omega^2/\omega_\beta^2 \to 2 - 2q^2a^2/y_n^2 + \cdots, \quad (9.2)$$

where y_n is the nth root of the equation

$$J_0'(y)/yJ_0(y) = \tfrac{1}{2}[(R/a)^2 - 1]. \quad (9.3)$$

For $|qa| \gg 1$ we find

$$\Omega^2/\omega_\beta^2 = [m - 2 - j_{m-1,n+1}^2/4q^2a^2 + \cdots]^2. \quad (9.4)$$

The $1/q^2a^2$ term in (9.4) is correct only for $m \geqq 3$; for $m = 0$ and $m = 1$, we have instead

$$A_{0n}: \quad \Omega^2/\omega_\beta^2 = [2 + j_{1,n}^2/2q^2a^2 + \cdots]^2, \quad (9.5)$$

$$A_{1n}: \quad \Omega^2/\omega_\beta^2 = [1 + x_{n+1}^2/4q^2a^2 + \cdots]^2, \quad (9.6)$$

with x_n the nth root of the equation

$$xJ_1'(x) = -3J_1(x). \quad (9.7)$$

The first three roots are

$$x_1 = 2.95, \quad x_2 = 5.84, \quad x_3 = 8.87.$$

B_{mn} ($m \neq 0$)

For $|qa| \ll 1$ we find

$$\frac{\Omega^2}{\omega_\beta^2} = 3 + m^2 + (12m^2+1)^{\frac{1}{2}} - q^2a^2\left(1 + \frac{2m^2-1}{(12m^2+1)^{\frac{1}{2}}}\right)\Big/ j_{mn}^2 + \cdots. \quad (9.8)$$

For $|qa| \gg 1$ we find

$$\Omega^2/\omega_\beta^2 = [m + 2 + j_{m+1,n}^2/4q^2a^2 + \cdots]^2. \quad (9.9)$$

There are no exceptions.

C_m ($m \neq 0, 2$)

For $|qa| \ll 1$ we find

$$\Omega^2/\omega_\beta^2 \to m^2 - 2m + \tfrac{3}{2} - (2m^2 - 4m + \tfrac{9}{4})^{\frac{1}{2}}. \quad (9.10)$$

The next terms are of order q^2, but it is difficult to give a general formula for their coefficients. We have done the calculation to order q^4 for $m = 1$ and q^2 for $m = 3$, and we find that, for $|qa| \ll 1$ *and* $|qR| \ll 1$,

$$C_1: \quad \frac{\Omega^2}{\omega_\beta^2} \to -\tfrac{1}{2}q^2a^2\left(\ln\frac{R}{a} + \frac{1}{4}\right) - \tfrac{1}{4}q^4a^4\left(\ln\frac{R}{a} - \frac{3}{8}\right) - \tfrac{1}{8}q^4a^2R^2 + \cdots, \quad (9.11)$$

$$C_3: \quad \frac{\Omega^2}{\omega_\beta^2} \to \tfrac{1}{2}[9 - (33)^{\frac{1}{2}}] + \left(\frac{17}{90} - \frac{3}{40}\left[\frac{a}{R}\right]^2\right)q^2a^2 + \cdots. \quad (9.12)$$

For $|qa| \gg 1$ the C mode is very much like the A mode:

$$\Omega^2/\omega_\beta^2 \to [m - 2 - (\chi_m^c)^2/4q^2a^2 + \cdots]^2, \quad (9.13)$$

where χ_1^c is the first root of Eq. (9.7)

$$\chi_1^c = x_1 = 2.95 \quad (m = 1),$$

and χ_m^c for $m \geq 3$ is the first zero of $J_{m-1}(x)$

$$\chi_m^c = j_{m-1,1} \quad (m \geq 3).$$

The vanishing of (9.11) for $q = 0$ identifies C_1 as the hose mode. Equation (9.11) may be compared with the results of the rigid-beam calculations made in earlier work. We found in Ref. 4 that

$$\Omega^2/\omega_\beta^2 = 1 - i\pi J_1(qa)\{H_1^{(1)}(qa) - [H_0^{(1)}(qR)/J_0(qR)]J_1(qa)\} \quad (9.14)$$

and to order q^4 this gives

$$\frac{\Omega^2}{\omega_\beta^2} = -\tfrac{1}{2}q^2a^2\left(\ln\frac{R}{a} + \frac{1}{4}\right) - \tfrac{1}{16}q^4a^4\left(\ln\frac{R}{a} + 4\right) + q^4a^2R^2\left(\ln\frac{R}{a} + \frac{5}{8}\right) - \tfrac{1}{32}q^4R^4 \quad \text{(rigid)}. \quad (9.15)$$

Note that (9.15) agrees with the correct result (9.11) to order q^2, as it must according to the general theorem of Sec. 7. [Equations (7.21) and (7.22) here give $u^2 = \omega_\beta^2a^2$ and $r_c = ae^{-\frac{1}{4}}$.] However, (9.11) and (9.15) begin to differ in the q^4 terms, which is not surprising

since there is no reason why the rigid-beam result (9.14) should give more than the q^2 terms correctly.

If R is very much greater than a, it is possible to have $|qa| \ll 1$ but $|qR| \gg 1$. In this case the correct (or the rigid beam) hose dispersion relation is given by (9.11) [or (9.15)] with $q^4a^2R^2$ (and q^4R^4) terms dropped, and with the replacement

$$\ln R/a \to \ln(-\tfrac{1}{2}iqa) + 0.577. \tag{9.16}$$

It is expected that similar rules work in all other C and D modes.

$$D_m\ (m \neq 0, 1)$$

For $|qa| \ll 1$ we find

$$\Omega^2/\omega_\beta^2 \to m^2 - 2m + \tfrac{3}{2} + (2m^2 - 4m + \tfrac{9}{4})^{\frac{1}{2}}. \tag{9.17}$$

The next terms are of order q^2 but (as in the C_m modes) rather difficult to calculate in general. We have done the work only for $m \leqq 3$; for $|qa| \ll 1$ and $|qR| \ll 1$ we find:

$$D_2\text{:}\quad \Omega^2/\omega_\beta^2 \to 3 - (\tfrac{1}{6} - \tfrac{1}{8}[a/R]^2)q^2a^2 + \cdots, \tag{9.18}$$

$$D_3\text{:}\quad \Omega^2/\omega_\beta^2 \to \tfrac{1}{2}[9 + (33)^{\frac{1}{2}}] - (\tfrac{17}{90} - \tfrac{3}{40}[a/R]^2)q^2a^2 + \cdots. \tag{9.19}$$

For $|qa| \gg 1$ we find a behavior quite different from that of the A, B, and C modes:

$$D_m\text{:}\quad \Omega^2/\omega_\beta^2 \to m^2\left\{1 + \frac{1}{iqa} \times \left[1 + \left(\frac{3m^2 - 2}{4(m^2 - 1)}\right)^{\frac{1}{2}}\right]^{-1} + \cdots\right\}. \tag{9.20}$$

There are no exceptions to Eq. (9.20).

It is important to note that in all modes Ω^2/ω_β^2 approaches a positive-definite constant λ_0^2 or λ_∞^2 for $q^2 \to 0$ or $q^2 \to \infty$, the only exception being that $\lambda_0^2 = 0$ in the hose mode. Hence all modes are stable for very small or very large ω, oscillating at the frequencies $\pm\lambda_0\omega_\beta$ or $\pm\lambda_\infty\omega_\beta$. Our formulas for the behavior of Ω^2/ω_β^2 as ω approaches zero or infinity show that Ω^2 is generally complex for finite real ω (finite imaginary q^2) so that in all modes there is one growing root with maximum growth at some finite ω; a method for estimating this maximum is presented in Sec. XII.

The stable oscillation frequencies for $\omega \to \infty$ have a simple physical interpretation. For $|qa| \gg 1$ we found the limiting behavior

$$\frac{\Omega}{\omega_\beta} \to \begin{cases} \pm(m-2) & A_{mn} \\ \pm(m+2) & B_{mn} \\ \pm(m-2) & C_m \\ \pm m & D_m \end{cases}.$$

A close look at Appendix B and Sec. IV shows that it is only the beam particles with gyration frequency $\pm\omega_\beta$ (rather than $\mp\omega_\beta$) that oscillate with these frequencies. According to Sec. IV such particles suffer a displacement proportional to

$$\exp(-i[\Omega \mp m\omega_\beta]t) \propto \exp\left(\pm i\left[\frac{\Omega}{\omega_\beta} \mp m\right]\theta\right) \to \begin{cases} e^{-2i\theta} & (A_{mn}) \\ e^{2i\theta} & (B_{mn}) \\ e^{-2i\theta} & (C_m) \\ e^{0} & (D_m) \end{cases}.$$

Thus the perturbed orbit is a slightly eccentric ellipse in modes A, B, and C and a slightly enlarged or diminished circle in mode D. This is very reasonable because for $|qa| \gg 1$ the electromagnetic fields are essentially frozen in place by the plasma conductivity, so the beam particle orbits can be independently perturbed into ellipses or circles of different radius. Tables I–III and Figs. 1–5 summarize our conclusions in numerical form for $m = 0$, 1, 2, and 3, and for $m \gg 1$.

TABLE I. Catalog of modes for the uniform beam. The behavior of Ω for $q \to 0$ ($\omega \to 0$) is given by $\Omega^2/\omega_\beta^2 \to \lambda_0^2 - q^2a^2\lambda_1^2 + \cdots$. Also, for $q \to \infty$ ($\omega \to \infty$), $\Omega^2/\omega_\beta^2 \to \lambda_\infty^2$. Note the relatively small changes in Ω^2 between these two limits.

Mode	λ_0^2	λ_1^2	λ_∞^2
$m = 0$: A_{0n}	2	0.344 ($n = 1$), 0.066 ($n = 2$), $\cdots$	4
$m = 1$: A_{1n}	0.39	0.049 ($n = 1$), 0.015 ($n = 2$), $\cdots$	1
B_{1n}	7.61	0.086 ($n = 1$), 0.026 ($n = 2$), $\cdots$	9
Hose $\to C_1$	0	$0.125 + \frac{1}{2}\ln R/a$	1
$m = 2$: B_{2n}	14	0.076 ($n = 1$), 0.028 ($n = 2$), $\cdots$	16
D_2	3	$0.167 - 0.125\ a^2/R^2$	4
$m = 3$: A_{3n}	1.55	−0.015 ($n = 1$), −0.0065 ($n = 2$), $\cdots$	1
B_{3n}	22.45	0.065 ($n = 1$), 0.028 ($n = 2$), $\cdots$	25
C_3	1.63	$-0.188 + 0.075\ a^2/R^2$	1
D_3	7.38	$0.083 - 0.34\ a^2/R^2$	9
$m \gg 1$: A_{mn}	$(m - 1.732)^2$		$(m-2)^2$
B_{mn}	$(m + 1.732)^2$		$(m+2)^2$
C_m	$(m - 1.707)^2$		$(m-2)^2$
D_m	$(m - 0.293)^2$		m^2

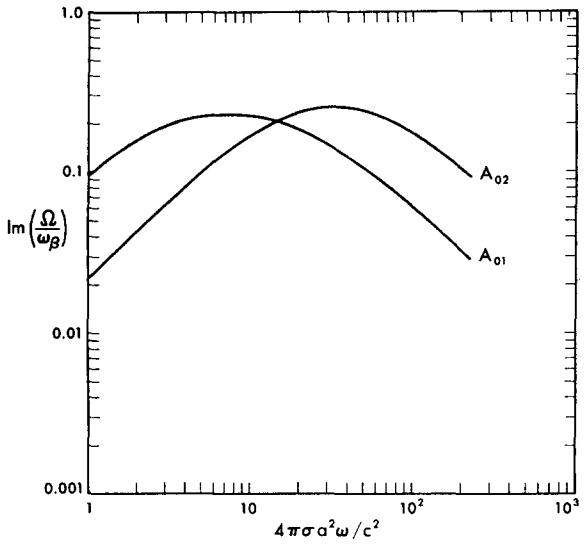


FIG. 1. Growth rate vs frequency for the A_{01} and A_{02} (sausage) modes computed (Ref. 8) for an uniform beam, with $R = 2a$.

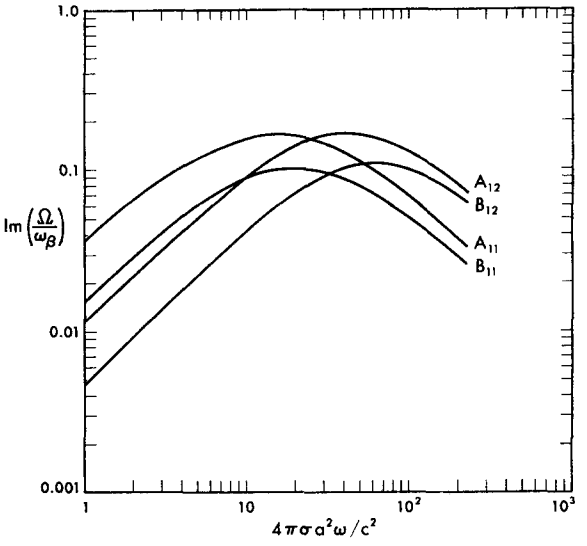


FIG. 2. Growth rate vs frequency for the A_{11}, A_{12}, B_{11}, B_{12} modes, computed (Ref. 8) for an uniform beam, with $R = 2a$.

X. CONTINUOUS SPECTRUM

We now return to the case of a general beam shape. It is assumed that the beam density $n(r)$ drops smoothly and monotonically from a value $n(0)$ at the beam axis to $n(a) = 0$ at a radius $r = a$. A catalog of modes is compiled in the next section, but first we must discuss a type of mode qualitatively different from any found for the uniform beam.

Suppose Ω^2 is such that $f(r)$ becomes unity at some radius r_0, with

$$1 - f(r_0) \to (r - r_0)/b \quad \text{for} \quad r \to r_0, \qquad (10.1)$$

the length b as well as r_0 depending on Ω^2 in a fairly complicated way. For $|r - r_0| \ll b$ the differential equation (6.1) becomes essentially

$$(d/dr)(r - r_0)(d\mathcal{E}/dr) - \kappa\mathcal{E} = 0, \qquad (10.2)$$

$$\kappa = -b[q^2 + g(r_0)]. \qquad (10.3)$$

Hence we may define two solutions of Eq. (6.1), which

TABLE II. Uniform beam modes at high frequency. For $q \to \infty$ ($\omega \to \infty$) the behavior of the A, B, and C modes is $\Omega^2/\omega_\beta^2 \to [\lambda_\infty - \chi^2/4q^2a^2 + \cdots]^2$. In the D modes we have instead $\Omega^2/\omega_\beta^2 \to [\lambda_\infty + \chi/2iqa + \cdots]^2$. The number of e foldings in free growth at a fixed distance μ behind the head of the beam is, for the A, B, and C modes: $\# = [\chi^2c^2\omega_\beta\mu z/8\pi\sigma a^2v^2]^{\frac{1}{2}}$ and for the D modes $\# = \frac{3}{8}\sqrt{3}[\chi^2c^2\omega_\beta^2\mu z^2/\pi\sigma a^2v^3]^{\frac{1}{3}}$.

Mode	λ_∞	x
$m = 0$: A_{0n}	-2	5.42 ($n = 1$), 9.90 ($n = 2$), $\cdots$
$m = 1$: A_{1n}	-1	5.84 ($n = 1$), 8.87 ($n = 2$), $\cdots$
B_{1n}	-3	5.14 ($n = 1$), 8.42 ($n = 2$), $\cdots$
Hose $\to C_1$	-1	2.95
$m = 2$: B_{2n}	-4	6.38 ($n = 1$), 9.76 ($n = 2$), $\cdots$
D_2	$+2$	1.05
$m = 3$: A_{3n}	$+1$	8.42 ($n = 1$), 11.62 ($n = 2$), $\cdots$
B_{3n}	-5	7.59 ($n = 1$), 11.06 ($n = 2$), $\cdots$
C_3	$+1$	5.14
D_3	$+3$	1.59

TABLE III. Parameters for fastest forced growth. The growth rate Im Ω reaches a maximum value equal to $(\mathrm{Im}\,\lambda)_M\omega_\beta$ when the real frequency ω takes the value $c^2Y_M/4\pi\sigma a^2$. The first two columns were computed[8] directly from the "exact" uniform beam dispersion relation (8.12) for $R = 2a$. The last column gives λ_M as estimated by a method mentioned in Sec. XII; apparently this method gives values about 15% too high.

Mode		Y_M	$(\mathrm{Im}\,\lambda)_M$	$(\mathrm{Im}\,\lambda)_M^{\mathrm{est}}$
$m = 0$: A_{01}		7	0.225	0.29
A_{02}		35	0.255	0.29
$\cdots$				
$m = 1$: A_{11}		14	0.165	0.19
A_{12}		30	0.167	0.19
$\cdots$				
B_{11}		20	0.10	0.12
B_{12}		62	0.11	0.12
$\cdots$				
	$R^2/a^2 = 1.25$	2.1	0.293	
	1.5	2.0	0.310	
Hose $\to C_1$	2	1.6	0.318	0.35
	4	0.85	0.297	
	8	0.55	0.260	
$m = 2$: B_{21}		30	0.11	0.13
B_{22}		78	0.12	0.13
$\cdots$				
D_2		6	0.09	0.13
$m = 3$: A_{3n}				0.12
B_{3n}				0.13
C_3		17	0.11	0.14
D_3		15	0.097	0.14
$m \gg 1$: A_{mn}				0.134
B_{mn}				0.134
C_m				0.146
D_m				0.146

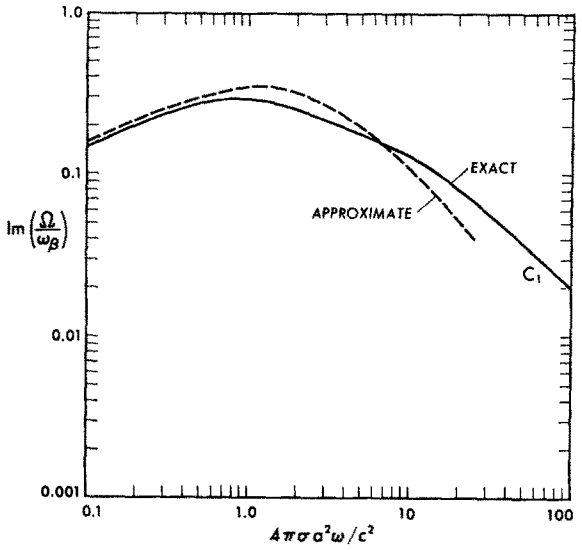


FIG. 3. Growth rate vs frequency for the C_1 (hose) mode, computed (Ref. 8) for an uniform beam, with $R = 2a$. The dashed curve shows the corresponding approximate result, based on the simple formula (12.25), choosing ω_0 to give $\Omega(\omega)$ the correct behavior as $\omega \to 0$. (A different choice of ω_0 would simply displace the dashed curve along the ω axis.)

near r_0 have the behavior

$$\mathcal{E}_{\mathrm{I}} \to I_0\{[\kappa(r - r_0)]^{\frac{1}{2}}\} \to 1 + \tfrac{1}{4}\kappa(r - r_0) + \cdots, \quad (10.4)$$

$$\mathcal{E}_{\mathrm{II}} \to K_0\{[\kappa(r - r_0)]^{\frac{1}{2}}\} \to -\tfrac{1}{2}[1 + \tfrac{1}{4}\kappa(r - r_0) + \cdots] \ln [\tfrac{1}{2}C^2\kappa(r - r_0)] + \tfrac{1}{4}\kappa(r - r_0) + \cdots. \quad (10.5)$$

Ordinarily we would have to avoid the logarithmic singularity at r_0, and take $\mathcal{E} = \mathcal{E}_{\mathrm{I}}$. The requirement that this solution also behave nicely when continued to $r = 0$ and $r = R$ imposes *two* conditions on Ω^2, which we do not expect to be satisfied by any choice of the *single* function $\Omega^2(q^2)$.

However there is a less obvious kind of instability. Suppose that the solution of Eq. (6.1) regular at $r = 0$ is

$$\mathcal{E}(r) \propto \mathcal{E}_{\mathrm{II}}(r) + \beta_-\mathcal{E}_{\mathrm{I}}(r), \quad r < r_0, \quad (10.6)$$

while the solution of Eq. (6.1) which behaves properly at $r = R$ is

$$\mathcal{E}(r) \propto \mathcal{E}_{\mathrm{II}}(r) + \beta_+\mathcal{E}_{\mathrm{I}}(r), \quad r > r_0. \quad (10.7)$$

We try as our solution

$$\mathcal{E}_\Omega(r) = \mathcal{E}_{\mathrm{II}}(r) + \theta(r - r_0)\beta_-\mathcal{E}_{\mathrm{I}}(r) + \theta(r_0 - r)\beta_+\mathcal{E}_{\mathrm{I}}(r). \quad (10.8)$$

This is admittedly singular at $r = r_0$ (though nice at $r = 0$ and $r = R$) but it nevertheless gives a perfectly nonsingular electric field if we consider, not just one Ω value, but a continuous range with a smooth weighting function $W(\Omega)$

$$\mathcal{E}(r, \theta, z, t) = \exp\left[-i\omega\left(\frac{t - z}{v}\right)\right] \exp[im\theta] \times \int d\Omega W(\Omega) \exp\left(-\frac{i\Omega z}{v}\right) \mathcal{E}_\Omega(r) \quad (10.9)$$

for, since r_0 depends on Ω^2, the Ω integration smooths out the logarithmic and step-function singularities in $\mathcal{E}_\Omega$.

We must still check that (10.8) solves the differential equation (6.1). Using the fact that $\mathcal{E}_{\mathrm{I}}$ and $\mathcal{E}_{\mathrm{II}}$ are

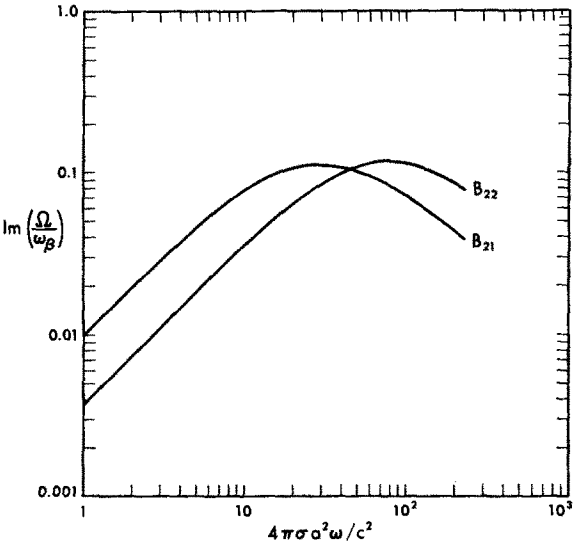


FIG. 4. Growth rate vs frequency for the B_{21} and B_{22} modes, computed (Ref. 8) for an uniform beam, with $R = 2a$.

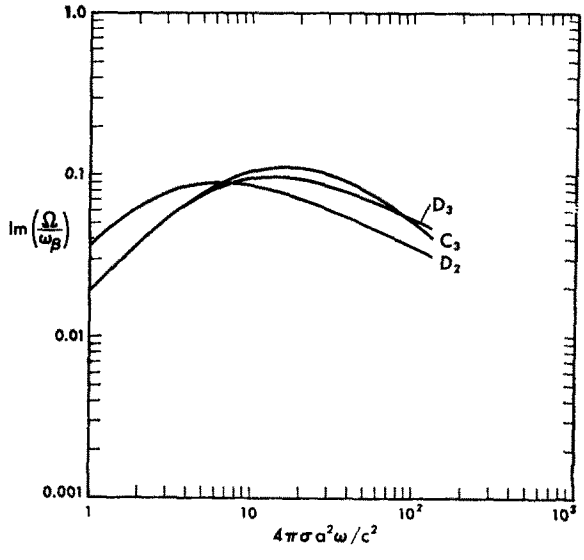


FIG. 5. Growth rate vs frequency for the D_2, C_3, and D_3 modes, computed (Ref. 8) for an uniform beam, with $R = 2a$.

defined to satisfy (6.1) for all r we find

$$\left[\frac{1}{r}\frac{d}{dr}r[1-f(r)]\frac{d}{dr}-\frac{m^2}{r^2}[1-f(r)]+q^2+g(r)\right]\mathcal{E}_\Omega(r)$$
$$=(\beta_- - \beta_+)\left[\frac{1}{r}\frac{d}{dr}r[1-f(r)]\delta(r-r_0)\mathcal{E}_I(r)\right.$$
$$\left.+[1-f(r)]\mathcal{E}_I'(r)\delta(r-r_0)\right]. \quad (10.10)$$

But $1-f(r)$ vanishes [and $\mathcal{E}_I(r)$ is regular] at r_0 so (10.10) gives no contribution in an integral over Ω. {The first term must be integrated by parts, and then both contain factors $\delta(r-r_0)[1-f(r)]$.} Hence (10.8) is a satisfactory solution for a range of Ω^2 values and any fixed ω. Such modes are usually said to belong to the *continuous spectrum* of the differential equation.[9]

What is the Ω^2 range of the continuous spectrum? We note from Eq. (5.18) that the condition $f(r_0)=1$ is satisfied by the Ω^2 values

$$\Omega^2=(3+m^2)\alpha^2(r)+\tfrac{1}{2}r\alpha^{2'}(r)\pm[(1+12m^2)\alpha^4(r)$$
$$+r\alpha^{2'}(r)\alpha^2(r)(1+2m^2)+\tfrac{1}{4}r^2\alpha^{2'2}(r)]^{\frac{1}{2}} \quad (10.11)$$

except that the "+" root is absent for $m=0$. The gyration frequency $\alpha(r)$ drops from $\alpha(0)$ at $r=0$ to some value $\alpha(a)<\alpha(0)$ at the beam edge, while $2\alpha^2(r)+r\alpha^{2'}(r)$ [which is proportional to $n(r)$] drops from $2\alpha^2(0)$ at $r=0$ to zero at $r=a$. A little work reveals that the continuous spectrum consists of the Ω^2 values

$m=0$,

$$2\alpha^2(0)\geq\Omega^2\geq 2\alpha^2(a); \quad (10.12)$$

$$[4-(13)^{\frac{1}{2}}]\alpha^2(0)\geq\Omega^2\geq(\sqrt{2}-1)^2\alpha^2(a), \quad (10.13)$$

$m=1$,

$$[4+(13)^{\frac{1}{2}}]\alpha^2(0)\geq\Omega^2\geq(\sqrt{2}+1)^2\alpha^2(a); \quad (10.14)$$

$m\geq 2$,

$$(3+m^2-[1+12m^2]^{\frac{1}{2}})\alpha^2(0)\leq\Omega^2$$
$$\leq(m-\sqrt{2})^2\alpha^2(a), \quad (10.15)$$

$$(3+m^2+[1+12m^2]^{\frac{1}{2}})\alpha^2(0)\geq\Omega^2$$
$$\geq(m+\sqrt{2})^2\alpha^2(a). \quad (10.16)$$

[For sufficiently large m the limits in (10.15) must be reversed.] The continuous spectrum thus consists only of stable oscillation frequencies Ω. We repeat that the above ranges of Ω^2 allow solutions for any ω, real or complex.

XI. GENERAL BEAM SHAPE: CATALOG OF MODES

We next examine the discrete modes for a general beam shape, which (unlike those of the last section) correspond to the A, B, C, and D modes found for the uniform beam in Sec. IX. We first center our attention on the limit $\omega\to\infty$, i.e., $|qa|\gg 1$, because this is by far the easiest way to catalog the modes, and because it is the high-frequency limit which will be seen in Sec. XIII to govern the asymptotic behavior of instabilities under conditions of free growth. Then we attempt to say something about the case of finite frequency, by examining the behavior of Ω^2 for $q^2<0$.

1. High Frequency

Inspection of Eq. (6.1) suggests that when $q^2\to\infty$, Ω^2 must approach a value Ω_∞^2, such that either $f(r)$ or $g(r)$ becomes infinite somewhere. To simplify the problem, it is assumed that the beam density $n(r)$ is flat at $r=0$, drops smoothly and monotonically to some value $n(a)>0$ at the beam edge $r=a$, and then drops very steeply to 0 just outside $r=a$. The functions $f(r)$ and $g(r)$ are then stationary only at $r=0$, where

$$f(0)=\frac{2\alpha^2(0)[(4-m^2)\alpha^2(0)-\Omega^2]}{[(4-m^2)\alpha^2(0)-\Omega^2]^2-4m^2\Omega^2\alpha^2(0)}, \quad (11.1)$$

$$g(0)=0. \quad (11.2)$$

At the beam edge, we have

$$g(r)\to-\xi^2\delta(r-a)/a, \quad (11.3)$$

$$\xi^2=\frac{2m^2\alpha^2(a)f(a)}{\Omega^2-m^2\alpha^2(a)}$$
$$\times\left[1-\frac{2\Omega^2}{(4-m^2)\alpha^2(a)-\Omega^2+a\alpha^{2'}(a)}\right]. \quad (11.4)$$

Hence our guess, based largely on our experience for the uniform beam, is that there are three kinds of modes for $|qa|\gg 1$:

$$A\ \&\ C:\quad f(0)\to\infty,\quad \Omega^2\to\alpha^2(0)(2-m)^2, \quad (11.5)$$

$$B:\quad f(0)\to\infty,\quad \Omega^2\to\alpha^2(0)(2+m)^2, \quad (11.6)$$

$$D:\quad \xi^2\to\infty,\quad \Omega^2\to m^2\alpha^2(a). \quad (11.7)$$

Note that the vanishing of the denominator

$$(4-m^2)\alpha^2(a)-\Omega^2+a\alpha^{2'}(a)$$

in Eq. (11.4) does not lead to another pole of ξ^2, since it is canceled by a zero in $f(a)$. The modes B_0, D_0, A_2, and C_2 are evidently absent.

ABC Modes

We suppose that for $|qa|\gg 1$ the value of Ω^2 is

$$\Omega^2=\alpha^2(0)(2\pm m-\delta)^2,\quad |\delta|\ll 1, \quad (11.8)$$

the sign $\pm$ being $+$ for B modes and $-$ for A and C modes. In the region of r close to the beam axis, the

beam density may presumably be well represented by

$$n(r) \simeq n(0)[1 - (r/l)^{2N}] \tag{11.9}$$

with N an integer [the exponent of r must be an even integer to avoid a singularity in $n(r)$ when expressed in Cartesian coordinates], and l a length of the order of the beam radius a. Integration of Eq. (2.3) yields

$$\alpha^2(r) \simeq \alpha^2(0)\left[1 - \frac{(r/l)^{2N}}{1+N}\right]. \tag{11.10}$$

Inserting (11.10) and (11.8) in (5.18) and (5.19), we find for $|qa| \gg 1$ and $r \ll l$

$$f(r) \simeq \left[4\delta - \frac{(4 \pm 2m + 2N)}{1+N}(r/l)^{2N}\right]^{-1}, \tag{11.11}$$

$$g(r) \simeq \pm m f'(r)/r. \tag{11.12}$$

[Some care is needed to show that Eq. (11.12) holds even for the A_1 and C_1 modes.] Equation (6.1) becomes approximately

$$\frac{1}{r}\frac{d}{dr} rf(r)\frac{d}{dr}\mathcal{E}(r) - m^2\frac{f(r)}{r^2}\mathcal{E}(r) \pm m\frac{f'(r)}{r}\mathcal{E}(r) - q^2\mathcal{E}(r) = 0. \tag{11.13}$$

We introduce a dimensionless variable

$$\rho = [4\delta(1+N)/(4 \mp 2m + 2N)]^{-\frac{1}{2}N}(r/l) \tag{11.14}$$

and write (11.13) in dimensionless form as

$$\frac{1}{\rho}\frac{d}{d\rho}\frac{\rho}{1-\rho^{2N}}\frac{d}{d\rho}\mathcal{E} - \frac{m^2}{\rho^2}\frac{1}{1-\rho^{2N}}\mathcal{E} \pm \frac{2mN\rho^{2N-2}}{(1-\rho^{2N})^2}\mathcal{E} + \mu^2\mathcal{E} = 0 \tag{11.15}$$

with

$$\mu^2 = -[(1+N)/(4 \pm 2m + 2N)]^{1/N} \times (4\delta)^{(N+1)/N}q^2l^2. \tag{11.16}$$

The boundary conditions on (11.15) are

$$\mathcal{E} \sim \rho^m \quad \text{for} \quad |\rho| \ll 1, \tag{11.17}$$

$$\mathcal{E} \sim \exp[-\mu\rho^{N+1}/(N+1)] \quad \text{for} \quad |\rho| \gg 1, \tag{11.18}$$

with μ defined as the root of (11.16) with $\operatorname{Re}\mu > 0$.

Solving (11.16) for δ and inserting the answer in (11.8) yields the dispersion relation for $|qa| \gg 1$

$$\Omega^2 = \alpha^2(0)\left[2 \pm m - \frac{1}{4}\left(\frac{4 \pm 2m + 2N}{1+N}\right)^{1/(N+1)} \times \left(\frac{-\mu^2}{q^2l^2}\right)^{N/(N+1)}\right]^2 \tag{11.19}$$

with the eigenvalues μ^2 to be found by solving Eq. (11.15) and imposing (11.17) and (11.18). It is interesting to note that for a beam density $n(r)$ very flat near the beam axis we must take N large, and (11.19) becomes

$$\Omega^2 = \alpha^2(0)[2 \pm m + (\mu^2/4q^2l^2)]^2, \tag{11.20}$$

which is of the same form as the result found earlier [see (9.4) and (9.9)] for the uniform beam, i.e.,

$$\Omega^2 = \alpha^2[2 \pm m + (j^2_{m\pm1,n}/4q^2a^2)]^2. \tag{11.21}$$

In analogy with the uniform beam we call the modes with $\Omega^2 \to (2+m)\alpha^2(0)$ the B_m modes, while the mode with $\Omega^2 \to (2-m)^2\alpha^2(0)$ corresponding to the smallest eigenvalue of (11.15) is called the C_m mode, and all other such modes are called the A_m modes, except that for $m = 0$ there is no C mode.

We make no attempt to discuss the low-lying eigenvalues of (11.15). However, comparison of (11.21) with (11.20) leads us to suspect that there is an infinite sequence of μ^2 values tending to infinity. This is verified by a WKB calculation in Appendix C. We show there that the nth eigenvalue for a given m has the asymptotic value for large n:

$$\mu_{mn} \to \left[2(N+1)\Gamma\left(\frac{1}{2} + \frac{1}{2N}\right)\Big/\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2N}\right)\right] \times (n + \tfrac{1}{2}m + \tfrac{1}{2})\pi. \tag{11.22}$$

In particular, for a nearly uniform beam we must take N large and (11.22) becomes

$$\mu_{mn} \to (n + \tfrac{1}{2}m + \tfrac{1}{2})\pi. \tag{11.23}$$

This may be compared with the uniform beam result (11.20), which in place of μ has

$$j_{m\pm1,n} \to (n + \tfrac{1}{2}m - \tfrac{1}{4} \pm \tfrac{1}{2})\pi. \tag{11.24}$$

There is thus not a great deal of difference between an uniform beam and a smooth beam nearly flat at the beam axis.

It is apparent from Eq. (11.19) that for a given very large frequency ($|ql| \gg 1$) the growth rate $\operatorname{Im}\Omega$ will be larger for finite N than for a uniform beam; it vanishes like $\omega^{-N/N+1}$ as $\omega \to 0$. To some extent this enhanced growth rate is illusory, however, for (11.14) and (11.18) show that, as $\omega \to \infty$, the fields and currents are limited to a region around the beam axis which vanishes as

$$\delta^{\frac{1}{2}N} \sim \omega^{-\frac{1}{2}(N+1)}.$$

D Modes

The discontinuity condition at the beam edge $r = a$ is, as in Eq. (8.11),

$$\{[1 - f(r)](\mathcal{E}'/\mathcal{E})\}_{a-\epsilon} + \xi^2/a = (\mathcal{E}'/\mathcal{E})_{a+\epsilon}. \tag{11.25}$$

A trivial WKB calculation shows that, for $r < a$, the solution of Eq. (6.1) with $q^2a^2 \gg 1$ is a linear combination of the two functions

$$\mathcal{E}_\pm(r) = r^{-\frac{1}{2}}[1 - f(r)]^{-\frac{1}{4}} \exp\left[\pm iq\int_0^r \frac{dr}{[1 - f(r)]^{\frac{1}{2}}}\right]. \tag{11.26}$$

The well-behaved solution of Eq. (6.1) near $r = 0$ can be seen by inspection to be

$$\mathcal{E}(r) \propto J_m(qr/[1 - f(0)]^{\frac{1}{2}}). \tag{11.27}$$

For sufficiently large q there is a range of r values with $|qr| \gg [1 - f(0)]$ but $f(r) \simeq f(0)$; in this region (11.26) gives

$$\mathcal{E}_\pm(r) \propto r^{-\frac{1}{2}} \exp(\pm iqr/[1 - f(0)]^{\frac{1}{2}}),$$

while (11.27) gives, for $\operatorname{Im} q > 0$,

$$\mathcal{E}(r) \propto r^{-\frac{1}{2}} \exp(-iqr/[1 - f(0)]^{\frac{1}{2}}),$$

so the correct solution for $r \to a$ is just $\mathcal{E}_-(r)$. Its logarithmic derivative can be seen from (11.26) to be

$$\mathcal{E}_-'(r)/\mathcal{E}_-(r) \simeq -iq/[1 - f(r)]^{\frac{1}{2}} \quad (r < a). \tag{11.28}$$

Outside the beam the field is, for $|qR| \gg 1$,

$$\mathcal{E}(r) \propto H_m^{(1)}(qr),$$

so

$$\mathcal{E}'(r)/\mathcal{E}(r) \simeq +iq \quad (r > a). \tag{11.29}$$

Using (11.28) and (11.29) in (11.25) gives the asymptotic dispersion relation

$$-i\eta_\infty + \xi^2/qa = i \tag{11.30}$$

with

$$\eta_\infty^2 \equiv 1 - f(a - \epsilon). \tag{11.31}$$

Equation (11.30) is in agreement with the uniform beam result (8.43), except that a finite slope of the beam density $n(r)$ just inside the beam edge affects the value of $f(a - \epsilon)$ and hence of η_∞^2; setting $\Omega^2 = m^2\alpha^2(a)$ in (5.18) gives

$$\eta_\infty^2 = 1 - \frac{[a\alpha^{2\prime}(a) + 2\alpha^2(a)][(4 - 2m^2)\alpha^2(a) + a\alpha^{2\prime}(a)]}{[(4 - 2m^2)\alpha^2(a) + a\alpha^{2\prime}(a)] - 4m^4\alpha^4(a)}.$$

The solution of Eqs. (11.4) and (11.30) takes the form

$$\Omega^2 \to m^2\alpha^2(a)\left[1 + \frac{\chi}{2iqa}\right], \tag{11.32}$$

where

$$\chi = \frac{2[\alpha^2(a) + \frac{1}{2}a^2\alpha^{2\prime}(a)^2][(1 - m^2)\alpha^2(a) + \frac{1}{4}a\alpha^{2\prime}(a)]}{(1 + \eta_\infty)[(1 - m^2)\alpha^4(a) + \frac{1}{4}a\alpha^{2\prime}(a)\alpha^2 \times (a)(2 - m^2) + \frac{1}{16}a^2\alpha^{2\prime}(a)^2]}. \tag{11.33}$$

The form of (11.32) is the same as for an uniform beam. Note that as $\omega \to \infty$, the field $\mathcal{E}_-(r)$ vanishes exponentially except within a skin depth of the beam edge.

It is not clear what happens to the D modes when we consider a smooth-edge beam, with $n(a) = 0$.

2. Finite Frequency

Equations (11.11) and (11.16) show that, for large negative real q^2, the function Ω^2 in the A, B, and C modes takes values such that $f(r)$ becomes infinite at some point r_∞ near the beam axis. Our experience with the uniform beam suggests that this is still the case for all $q^2 < 0$ in the A and B modes, and also for sufficiently large $-q^2$ in the C modes, except that r_∞ may be anywhere inside the beam. The D modes are not expected to have $f(r)$ infinite anywhere and are not discussed further here.

According to Eq. (5.18), $f(r_\infty)$ is infinite if $n(r_\infty) \neq 0$ and

$$\Omega^2 = \{m\alpha(r_\infty) \pm [4\alpha^2(r_\infty) + r_\infty\alpha^{2\prime}(r_\infty)]^{\frac{1}{2}}\}^2. \tag{11.34}$$

For simplicity, we now assume that $n(r)$ drops smoothly and monotonically from $n(0)$ at $r = 0$ to $n(a) = 0$ at $r = a$; then the Ω^2 values which allow solutions of (11.34) with $r_\infty < a$ are

$m = 0$

$$4\alpha^2(0) \geq \Omega^2 \geq 2\alpha^2(a); \tag{11.35}$$

$m = 1$

$$\alpha^2(0) \geq \Omega^2 \geq (\sqrt{2} - 1)^2\alpha^2(a), \tag{11.36}$$

$$9\alpha^2(0) \geq \Omega^2 \geq (\sqrt{2} + 1)^2\alpha^2(a); \tag{11.37}$$

$m \geq 2$

$$(m - 2)^2\alpha^2(0) \leq \Omega^2 \leq (m - \sqrt{2})^2\alpha^2(a), \tag{11.38}$$

$$(m + 2)^2\alpha^2(0) \geq \Omega^2 \geq (m + \sqrt{2})^2\alpha^2(a). \tag{11.39}$$

[The limits in (11.38) must be reversed for sufficiently large m.] We must however exclude Ω^2 values satisfying (10.12)–(10.16), for these lie in the continuous spectrum. This leaves the following bands:

$m = 0(A)$,

$$4\alpha^2(0) \geq \Omega^2 \geq 2\alpha^2(0); \tag{11.40}$$

$m = 1(A, C)$,

$$\alpha^2(0) \geq \Omega^2 \geq [4 - (13)^{\frac{1}{2}}]^2\alpha^2(0); \tag{11.41}$$

$m \geq 3(A, C)$,

$$(m - 2)^2\alpha^2(0) \leq \Omega^2 \leq (3 + m^2 - [1 + 12m^2]^{\frac{1}{2}})\alpha^2(0); \tag{11.42}$$

$m \geq 1(B)$,

$$(m + 2)^2\alpha^2(0) \geq \Omega^2 \geq (3 + m^2 + [1 + 12m^2]^{\frac{1}{2}})\alpha^2(0). \tag{11.43}$$

The labels A, B, and C have been attached in accordance with the evident correspondence of the left-hand limits in (11.40)–(11.43) with the known behavior of Ω^2 as $q^2 \to -\infty$. However, our work in Appendix B on the uniform beam suggests that in the C modes, Ω^2 lies in the real bands (11.41), (11.42) only for $-q^2$ greater than some value Q_2^2, at which Ω^2 has a singularity and becomes complex, returning to real values when $-q^2$ passes below another singularity Q_1^2. This is particularly clear in the case of the hose mode C_1, since in this case we know that $\Omega^2 \to 0$ for $q^2 \to 0$ (see Sec. VII), and the value $\Omega^2 = 0$ does *not* lie in the "allowed" band (11.41). We have not allowed for the possibility of a C mode with $m = 0$, because one can easily show that for $q^2 < 0$ there are no $m = 0$ modes for which Eq. (6.1) is nonsingular, i.e., except for the continuous spectrum, Ω^2 must lie in the range (11.40) for all $q^2 < 0$. (See Appendix D.) Also, the A mode is missing for $m = 2$ (just as it was missing in Sec. IX) because for $m = 2$ the range (11.36) precisely overlaps the continuous spectrum (10.15). Furthermore, all A and C modes are absent for m sufficiently large. This is because the derivation of (11.42) made use of (10.15) and (11.38), which, respectively, make sense only for m sufficiently small so that

$$(3 + m^2 - [1 + 12m^2]^{\frac{1}{2}})/(m - \sqrt{2})^2 \leq \alpha^2(a)/\alpha^2(0), \tag{11.44}$$

$$(m - 2)^2/(m - \sqrt{2})^2 \leq \alpha^2(a)\,\alpha^2(0). \tag{11.45}$$

These inequalities hold for all but very large m if $\alpha(a)$ is close to $\alpha(0)$, as is the case if $n(r)$ is nearly constant almost out to $r = a$ and then drops steeply to zero at $r = a$. But in any case (11.44) and (11.45) breaks down for sufficiently large m. If (11.45) holds but (11.44) does not, then the allowed range for A modes is given by

$$(m - 2)^2\alpha^2(0) \leq \Omega^2 \leq (m - \sqrt{2})^2\alpha^2(a), \tag{11.46}$$

while if m is so large that (11.44) as well as (11.45) is not valid, then there are no A modes.

It is very satisfying that the right-hand limits in (11.40)–(11.43) are precisely the same as the limits at $q^2 \to 0$ of the uniform-beam A and B modes discussed in Sec. IX, except that what was ω_β has now become $\alpha(0)$. We can therefore have some confidence in the assertion that for low frequency the function Ω^2 approaches the values

$$\Omega^2 \to (3 + m^2 - [1 + 12m^2]^{\frac{1}{2}})\alpha^2(0) \quad [A_m, q^2 \to 0], \tag{11.47}$$

$$\Omega^2 \to (3 + m^2 + [1 + 12m^2]^{\frac{1}{2}})\alpha^2(0) \quad [B_m, q^2 \to 0]. \tag{11.48}$$

We do not know the low-frequency limits in the C or D modes, except, of course, that the C_1 mode has been thoroughly examined at low frequency in Sec. VII.

XII. ANALYTIC INTERPOLATION

In this section we offer arguments to the effect that Ω^2 is analytic in q^2 (i.e., in ω) except for singularities on the positive-real q^2 axis (i.e., on the negative imaginary ω axis). This property is then used to guess at interpolation formulas for Ω^2 for general complex q^2.

1. Causality

The usual causality argument may be used to show that Ω is analytic for $\operatorname{Im}\omega > 0$. (See Sec. XIII.) But this argument breaks down in the presence of a continuous spectrum, where Ω is not a function of ω at all. Presumably this explains the singularity in Ω found in the uniform beam C modes in Appendix B; the singularity occurs at a value of Ω at which $1 - f(r)$ vanishes inside the beam, and hence which would lie in the continuous spectrum of Eq. (6.1) for a smooth beam shape. It would be well to understand in more detail what is going on in the C modes when they cross into the continuous spectrum, but for our present purposes little harm will be done if we simply forget the whole problem and accept the implication of naive causality that Ω is analytic in the upper-half ω plane.

2. Poles

The next step is to show that Ω can become infinite only at a series of poles on the positive real q^2 axis, the nth pole lying at

$$q_{mn}^2 = j_{|m-1|,n}^2/R^2, \tag{12.1}$$

where R is the plasma channel radius and $j_{p,n}$ is the nth zero of $J_p(x)$. We note that when $\Omega^2 \gg \alpha^2$ the beam response functions $f(r)$ and $g(r)$ become negligible, so Eq. (6.1) becomes the differential equation for an undriven electromagnetic wave

$$\frac{1}{r}\frac{d}{dr}\,r\,\frac{d}{dr}\,\mathcal{E}(r) - \frac{m^2}{r^2}\,\mathcal{E}(r) + q^2\mathcal{E}(r) = 0. \tag{12.2}$$

The solution regular at $r = 0$ is $\mathcal{E} = J_m(qr)$. This must join smoothly with the known solution in the plasma channel outside the beam

$$\mathcal{E}(r) \propto J_{|m-1|}(qR)H_m^{(1)}(qr) - H_{|m-1|}^{(1)}(qR)J_m(qr). \tag{12.3}$$

Thus we can have $\Omega^2 \to \infty$ only when

$$J_{|m-1|}(qR) \to 0 \tag{12.4}$$

yielding Eq. (12.1).

It is also easy to obtain the residues of Ω near the

poles (12.1) by keeping terms in Eq. (6.1) of order Ω^{-2}. To this order, Eqs. (5.18) and (5.19) yield the beam response functions

$$f(r) = -[r^2\alpha^2(r)]'/r\Omega^2, \tag{12.5}$$

$$g(r) = 0. \tag{12.6}$$

We also suppose that

$$\delta(r) = J_m(qr) + \epsilon(r) \tag{12.7}$$

with ϵ of order Ω^{-2}. Neglecting higher-order terms, Eq. (6.1) becomes

$$\frac{1}{r}\frac{d}{dr}\, r\frac{d}{dr}\,\epsilon(r) - \frac{m^2}{r^2}\,\epsilon(r) + q^2\epsilon(r) = \nu(r), \tag{12.8}$$

where

$$\nu(r) \equiv -(r\Omega^2)^{-1}\left\{\frac{d}{dr}\,[r^2\alpha^2(r)]'\,\frac{d}{dr}\,J_m(qr) - \frac{m^2}{r^2}\,[r^2\alpha^2(r)]'J_m(qr)\right\}. \tag{12.9}$$

The solution of (12.8) regular at $r = 0$ is well known,

$$\epsilon(r) = -\tfrac{1}{2}i\pi\int_0^\infty \mathrm{r}\nu(\mathrm{r})J_m(qr_<)H_m^{(1)}(qr_>)\,d\mathrm{r}, \tag{12.10}$$

with $r_<$ and $r_>$ the lesser and greater of r and r. [We are using the freedom evidently allowed us by (12.7) to choose $\epsilon(r)$ so that it does not contain a term proportional to $J_m(qr)$ in addition to the integral (12.10).] For r outside the beam we take $r_> = r$, $r_< = \mathrm{r}$, and (12.7) becomes

$$\delta(r) = J_m(qr) - \tfrac{1}{2}i\pi H_m^{(1)}(qr)\int_0^\infty \mathrm{r}\nu(\mathrm{r})J_m(q\mathrm{r})\,d\mathrm{r}. \tag{12.11}$$

Comparing with (12.3), we see that

$$\frac{J_{|m-1|}(qR)}{H_{|m-1|}(qR)} \to -\tfrac{1}{2}i\pi\int_0^\infty r\nu(r)J_m(qr)\,dr. \tag{12.12}$$

Letting q approach the values (12.1) on the left, and using (12.9) and integrating by parts on the right, we find from (12.12) and (2.2) that

$$\Omega^2 \to \Omega_{mn}^2 q_{mn}^2/(q^2 - q_{mn}^2), \tag{12.13}$$

where Ω_{mn}^2 are the positive numbers

$$\Omega_{mn}^2 = \frac{8\pi e^2v^2}{M\gamma c^2R^2J'^2_{|m-1|}(j_{|m-1|,n})} \times\int_0^\infty\left\{J_m'^2(q_{mn}r) + \frac{m^2}{q_{mn}^2r^2}\,J_m^2(q_{mn}r)\right\}n(r)r\,dr. \tag{12.14}$$

If the plasma channel is infinite, then instead of the poles (12.1) we find a cut on the whole positive real q^2 axis.

3. Other Singularities

When $q^2 \to x^2 \pm i\epsilon$ (with $x^2 > 0$) we must take q in Eq. (12.3) as the root with positive imaginary part, $q = \pm x + i\epsilon$, so (12.3) appears to have a cut on the positive real axis. This cut is actually present if the plasma channel is infinite, and leads to a cut in Ω^2 which takes the place of the poles (12.1). But for finite R the cut is a chimera, since when $q \to -q$ the linear combination (12.3) merely changes sign.

It is tempting at this point to guess that Ω^2 is meromorphic in q^2, but this cannot be true. We know that Ω^2 stays bounded for $|q^2| \to \infty$ provided $\mathrm{Arg}\,(q^2) \neq 0$, so if Ω^2 were meromorphic in q^2 it could be written

$$\Omega^2 = \Omega_0^2 + q^2\sum_n \frac{\Omega_{mn}^2}{q^2 - q_{mn}^2} \tag{12.15}$$

with Ω_0^2 an unknown constant. But a simple calculation shows that for $n \to \infty$

$$q_{mn}^2 \to n^2\pi^2/R^2, \tag{12.16}$$

$$\Omega_{mn}^2 \to \frac{4\pi e^2v^2}{M\gamma c^2R}\int_0^\infty n(r)\,dr \equiv \frac{2K}{R}. \tag{12.17}$$

Thus the sum (12.15) does converge, but for large nonpositive q^2 it behaves like $K(-q^2)^{\frac{1}{2}}$, violating our theorem that Ω^2 can become infinite only at the poles q_{mn}. [Also, letting $R \to \infty$ for fixed nonpositive q^2 we find that (12.15) becomes just $\Omega_0^2 + K(-q^2)^{\frac{1}{2}}$, which is certainly too good to be true.]

The failure of meromorphicity seems to arise because as q^2 increases from q_{mn}^2 to $q_{m,n+1}^2$ we would expect Ω^2 to drop from $+\infty$ to $-\infty$, passing through one or two bands of Ω^2 values lying in the continuous spectrum. Branch points may be expected at the onset of such bands, though we confess to having made no progress toward a detailed understanding of these cuts. It is only a guess that they lie on the positive real q^2 axis.

[This very complicated singularity structure is in sharp contrast with the $m = 1$ solution found in Ref. 4 under the assumption of rigid beam displacement. There, Ω^2 was explicitly given as a meromorphic function of q^2 which *can* be represented as in (12.15), the residues Ω_{1n}^2 vanishing exponentially for $n \to \infty$. This underscores the great qualitative change in the dispersion relation made by giving up the constraint of rigid beam displacement.]

The analyticity properties discussed above suggest that Ω may be written as a dispersion integral in q^2:

$$\Omega = \Omega(0) + q^2\int_0^\infty \frac{\rho(p)\,dp}{q^2 - p^2} \tag{12.18}$$

with $\rho(p)$ some weight function. Since Ω approaches a limit $\Omega(\infty)$ for $|q^2| \to \infty$, it must obey the sum-rule

$$\Omega(\infty) - \Omega(0) = \int_0^\infty \rho(p)\, dp. \tag{12.19}$$

The convergence of (12.19) then leads us to guess that $\rho(p)$ falls off fast enough so that all the singularities in Ω may be lumped together into a pole at $q^2 = q_0^2$, i.e., so that we may approximate

$$\rho(p) \cong \delta(p - q_0)[\Omega(\infty) - \Omega(0)]. \tag{12.20}$$

Inserting this in (12.18) yields the approximate formula

$$\Omega(q^2) \cong \Omega(0) + \frac{q^2}{q^2 - q_0^2}[\Omega(\infty) - \Omega(0)] \tag{12.21}$$

or

$$\Omega(\omega) \cong \Omega(0) + \frac{\omega}{\omega + i\omega_0}[\Omega(\infty) - \Omega(0)], \tag{12.22}$$

$$\omega_0 = q_0^2 c^2/4\pi\sigma. \tag{12.23}$$

However, the hose mode requires special consideration, for as $\omega \to 0$ the function Ω vanishes like $(q^2)^{\frac{1}{2}}$ or $(\omega)^{\frac{1}{2}}$, and this branch point is too strong a singularity to lump with all the others. Instead, we must apply the above arguments to Ω^2 itself, and write

$$\Omega^2(q^2) \simeq \Omega^2(\infty)[q^2/(q^2 - q_0^2)], \tag{12.24}$$

or

$$\Omega(\omega) \simeq \omega_\beta[\omega/(\omega + i\omega_0)]^{\frac{1}{2}} \quad [C_1], \tag{12.25}$$

where $\omega_\beta \equiv \alpha(0)$ and ω_0 is given by (12.23).

The characteristic frequencies ω_0 in (12.22) and (12.25) can be chosen either to give $\Omega^{2\prime}(0)$ the correct value, or to give $\Omega(\omega)$ the correct asymptotic behavior as $\omega \to \infty$, or any other way that seems best suited to the problem at hand.

In order to test the value of the approximate formulas (12.22) and (12.25) in a way that does not depend on the specific choice of ω_0, we may use them to compute the maximum value of the growth rate Im Ω achieved for real frequency ω. Equation (12.22) gives, for real ω,

$$\text{Im } \Omega(\omega) \simeq \omega\omega_0/(\omega^2 + \omega_0^2)[\Omega(\infty) - \Omega(0)]. \tag{12.26}$$

Modes always come in pairs differing in the sign of Ω, and for $\omega > 0$ it is the mode with $\Omega(\infty) - \Omega(0)$ positive that has positive growth rate, which reaches a maximum when $\omega = \omega_0$, where

$$(\text{Im } \Omega)_{\max} \simeq \tfrac{1}{2}[\Omega(\infty) - \Omega(0)]. \tag{12.27}$$

For the hose mode the growth rate is given by (12.25) as

$$\text{Im } \Omega(\omega) \cong \omega_\beta\left[\frac{\omega[-\omega + (\omega^2 + \omega_0^2)^{\frac{1}{2}}]}{2(\omega^2 + \omega_0^2)}\right]^{\frac{1}{2}}. \tag{12.28}$$

This has a maximum at $\omega = \omega_0/\sqrt{3}$, where

$$(\text{Im } \Omega)_{\max} = \frac{\omega_\beta}{2\sqrt{2}} = 0.35\omega_\beta. \tag{12.29}$$

Table III shows that the values (12.27) and (12.29) are not far from the true maximum growth rates for a uniform beam, being generally about 15% too high. In particular, for the hose mode the correct value of the maximum growth rate is $0.29\omega_\beta$.

XIII. GROWTH RATES AND SADDLE POINTS

At last, we come to the point. In this section we use the results obtained in the previous twelve sections to estimate the growth of various modes under various conditions of excitation.

It is essential to begin by distinguishing between *forced growth* at a fixed frequency and *free growth* at a saddle point. Suppose the beam is tickled at $z = 0$ with a disturbance (in a pure normal mode) having t dependence

$$\mathcal{E}(t) = \int f(\omega)e^{-i\omega t}\, d\omega, \tag{13.1}$$

where $f(\omega)$ is some smooth function reaching a maximum at a real ω_0, with width Γ; for example, if the disturbance is turned on and off gradually, we might take

$$\pi f(\omega) = \Gamma/[(\omega - \omega_f)^2 + \Gamma^2],$$
$$\mathcal{E}(t) = e^{-i\omega_f t}e^{-\Gamma|t|}, \tag{13.2}$$

while if the disturbance (or the beam itself) is turned on suddenly at $t = 0$ and then turned off gradually, we might take

$$-2\pi i f(\omega) = [\omega - \omega_f + i\Gamma]^{-1},$$
$$\mathcal{E}(t) = \begin{cases} 0, & t < 0, \\ e^{-i\omega_f t}e^{-\Gamma t}, & t > 0. \end{cases} \tag{13.3}$$

Then at any $z > 0$ the field has the z and t dependence:

$$\mathcal{E}(z, t) = \int_{-\infty}^{\infty} f(\omega) \exp\{-i\omega t + ik(\omega)z\}\, d\omega$$

or

$$\mathcal{E}(z, t) = \int_{-\infty}^{\infty} f(\omega) \exp\{-i\omega(t - z/v) - i\Omega(\omega)z/v\}\, d\omega. \tag{13.4}$$

[Note that in the example (13.3) of a suddenly turned-on beam, the analyticity of $\Omega(\omega)$ in the upper-half ω plane ensures that $\mathcal{E}(z, t)$ vanishes for $z > vt$, since then the contour of integration can be closed with a large semicircle in the upper-half ω plane, where $\Omega(\omega)$ and $f(\omega)$ are analytic. Thus there is no need to impose a special boundary condition at $z = vt$ in this case.]

We are in a condition of *forced growth* when

$$\Gamma\,|t - z/v + \Omega'(\omega_f)(z/v)| \ll 1. \tag{13.5}$$

In this case the exponential in (13.4) is essentially constant over the support of $f(\omega)$, and we may approximate

$$\mathcal{E}(z, t) \simeq \exp\,[-i\omega_f(t - z/v) - i\Omega(\omega_f)z/v]. \tag{13.6}$$

The number of e foldings is

$$\# = (z/v)\,\mathrm{Im}\,\Omega(\omega_f) \tag{13.7}$$

and the maximum growth is achieved by adjusting the real frequency ω_f to the value at which (13.7) is greatest. Table III gives values of this maximum growth rate and the frequency ω_f at which it is achieved for the various low A, B, C, and D modes; it seems that the hose mode is from this point of view the most serious instability. Incidentally, when (13.7) is a maximum we may define a real group velocity

$$u_f \equiv v[1 - \Omega'(\omega_f)]^{-1} = 1/k'(\omega_f)$$

and (13.5) is just the condition that $\Gamma\,|t - z/u_f| \ll 1$. In other words, we have forced growth at z, t if, at the time $t - z/u_f$ when the disturbance left the disturber, the disturbance had not yet been turned off.

In contrast, we are in the case of *free growth* when

$$\Gamma\,|t - z/v + \Omega'(\omega_f)(z/v)| \gg 1. \tag{13.8}$$

For, even if ω_f is chosen to maximize $\mathrm{Im}\,\Omega(\omega_f)$, the phase of the exponential in Eq. (13.4) undergoes so many oscillations in the frequency range $\omega_f - \Gamma \leq \omega \leq \omega_f + \Gamma$ that cancellations intervene to prevent the pure exponential growth found in Eq. (13.6). It is well known that such integrals can usually be estimated by the saddle-point method. Where this is valid (and we do not venture a rigorous justification in the present context) the growth is dominated by a function

$$\mathcal{E}(z, t) \propto \exp\,[-i\omega_s(t - z/v) - i\Omega(\omega_s)z/v], \tag{13.9}$$

where ω_s is a complex frequency, depending on z and t, and determined by the condition that the argument of (13.9) be stationary, i.e.,

$$\Omega'(\omega_s) \equiv 1 - vt/z. \tag{13.10}$$

Equation (13.8) can be viewed as just the requirement that the saddle point *not* be at ω_f; if it were, then the saddle-point method would give the same result (13.6) as in the case of forced growth.

Although (13.8) requires that z or t must be large, we may still distinguish different cases according to the relative magnitude of t and z. We recall the basic scaling law (1.8):

$$\Omega(\omega) = \omega_\beta\lambda(i\omega/\omega_1), \tag{13.11}$$

$$\omega_1 \equiv c^2/4\pi\sigma a^2, \tag{13.12}$$

where λ is a function whose value and derivatives are of order unity when its argument is of order unity. Hence, the two special cases of greatest interest are $|z - vt|\,\omega_1 \gg z\omega_\beta$ and $|z - vt|\,\omega_1 \ll z\omega_\beta$.

1. $|z - vt|\,\omega_1 \gg z\omega_\beta$

In this case $|\Omega'(\omega_f)|$ is much greater than ω_β/ω_1, so ω_s must be near one of the poles found in the last section. For ω near the nth pole, the function $\Omega^2(\omega)$ is given by Eq. (12.13) as

$$\Omega^2(\omega) \to i\Omega^2_{mn}\omega_{mn}/(\omega + i\omega_{mn}), \quad \text{for } \omega \to -i\omega_{mn},$$

where

$$\omega_{mn} \equiv c^2q^2_{mn}/4\pi\sigma = (c^2R^2/4\pi\sigma)j^{-2}_{|m-1|,n} \tag{13.13}$$

and the positive constants Ω^2_{mn} are given by (12.14). The saddle-point condition (13.10) then gives

$$\omega \to -i\omega_{mn} + O[(\Omega_{mn}\,|z - vt|/\omega_{mn}z)^{\frac{2}{3}}\omega_{mn}]$$

and the electric field (13.9) behaves like

$$\mathcal{E} \propto \exp\,(-\omega_{mn}t). \tag{13.14}$$

This just means that if we fix z and wait long enough, we eventually find ourselves watching the decay of the wave-guide modes excited in the plasma channel.

2. $|z - vt|\,\omega_1 \ll z\omega_\beta$

In this case $|\Omega'(\omega_f)|$ is much smaller than ω_β/ω_1, so ω_s must be in the asymptotic region $\omega_s \gg \omega_1$, where $\Omega(\omega_s)$ approaches a constant. In all cases the asymptotic behavior of $\Omega(\omega)$ is of the form

$$\begin{aligned}\Omega(\omega) &\to \omega_\beta[\lambda_\infty + (\chi/2iqa)^{2\nu} + \cdots] \\ &= \omega_\beta[\lambda_\infty + (i\chi^2c^2/16\pi\sigma a^2\omega)^\nu + \cdots],\end{aligned} \tag{13.15}$$

where λ, χ, and ν are real dimensionless numbers depending on the mode in question. For the A, B, and C modes

$$\nu = N/(N + 1) \quad (A, B, C), \tag{13.16}$$

it being assumed that the beam density $n(r)$ behaves like $n(0) + O(r^{2N})$ as $r \to 0$. (For the uniform beam we take $N = \infty$, so in this case $\nu = 1$.) For the D modes

$$\nu = \tfrac{1}{2} \quad (D). \tag{13.17}$$

The quantities λ_∞ are

$$\lambda_\infty = \begin{cases} -2 + m, & A \,\&\, C, \\ -2 - m, & B, \\ m, & D, \end{cases} \qquad [\omega_\beta \equiv \alpha(0)].$$

The numbers χ are given for the uniform beam in Table II and Sec. IX. (It is amusing that the rigid hose calculation of Ref. 4 gets this backwards, giving

$\nu = \frac{1}{2}$ for a uniform beam and $\nu = 1$ for any other shape!) The saddle-point condition (13.10) yields here

$$16\pi\sigma a^2\omega_s/i\chi^2c^2 \to (+16\pi\sigma a^2\omega_\beta\nu z/i\chi^2c^2[vt - z])^{1/(\nu+1)}, \tag{13.18}$$

$$\Omega_s \to \omega_\beta[\lambda_\infty + (i\chi^2c^2[vt - z]/16\pi\sigma a^2\omega_\beta\nu z)^{\nu/(\nu+1)}], \tag{13.19}$$

so the exponential (13.9) becomes

$$\mathcal{E} \propto \exp\left\{-i\omega_\beta\lambda_\infty(t - z/v) - i(i\chi^2c^2[t - z/v]/16\pi\sigma a^2)^{\nu/(\nu+1)} \times \left(\frac{\omega_\beta z}{v}\right)^{1/(\nu+1)}[\nu^{1/(\nu+1)} + \nu^{-\nu/(\nu+1)}]\right\}$$

and the number of e foldings is

$$\# = \mathrm{Im}\,(i^{\nu/(\nu+1)})(\chi^2c^2[t - z/v]/16\pi\sigma a^2)^{\nu/(\nu+1)} \times (\omega_\beta z/v)^{1/(\nu+1)}(1 + \nu)\nu^{-\nu/(\nu+1)}. \tag{13.20}$$

In the two most interesting cases, the fastest growing roots give

$$\nu = 1: \quad \# = [\chi^2c^2\omega_\beta(vt - z)z/8\pi\sigma a^2v^2]^{\frac{1}{2}}, \tag{13.21}$$

$$\nu = \tfrac{1}{2}: \quad \# = \tfrac{3}{8}\sqrt{3}[\chi^2c^2\omega_\beta^2(vt - z)z^2/\pi\sigma a^2v^3]^{\frac{1}{3}}. \tag{13.22}$$

The distinction between forced and free growth can also be drawn for a disturbance which at $t = 0$ extends over some finite range of z. In forced growth the number of e foldings is Im ωt, with ω evaluated at a fixed real k. In free growth the asymptotic behaviors of $\mathcal{E}$ is again dominated by a saddle point at which the exponential in (13.4) is stationary, and all results obtained above in the case of free growth hold also for these different initial conditions.

ACKNOWLEDGMENTS

The author is glad to have the opportunity to thank K. Brueckner, K. Case, N. Christofilos, H. W. Lewis, and K. M. Watson for many valuable conversations on beam stability problems. Also, he is very grateful to S. C. Wright for performing the machine computation quoted in Sec. IX.

This work was performed by the author as a member of the Jason Division of the Institute for Defense Analyses, Arlington, Virginia.

APPENDIX A. BOUNDARY CONDITIONS AT THE CHANNEL RADIUS

We suppose that outside the beam the perturbed current $\mathbf{J}_1$ drops to zero, while the plasma conductivity σ stays constant out to a channel radius R, where σ drops sharply to zero. We assume there to be a vacuum for $r > R$, though we shall also see that it would make no difference if the whole system were enclosed by a conducting cylinder of radius larger than R.

The boundary conditions at R are the usual ones at the surface of a cylinder of finite conductivity, i.e.,

$$\Delta E_{1z} = 0, \tag{A1}$$

$$\Delta(4\pi\sigma - i\omega)E_{1r} = 0, \tag{A2}$$

$$\Delta E_{1\theta} = 0, \tag{A3}$$

$$\Delta B_{1z} = 0, \tag{A4}$$

$$\Delta B_{1r} = 0, \tag{A5}$$

$$\Delta B_{1\theta} = 0. \tag{A6}$$

Using Eqs. (3.8) and (3.11) (setting $\mathbf{J}_1 = 0$) with (A4), we find that (A2) and (A6) are equivalent, both yielding at $r = R$:

$$\Delta[(4\pi\sigma - i\omega/q^2c)E'_{1z}] = -(km/R)B_{1z}\Delta q^{-2}. \tag{A7}$$

Using Eqs. (3.9) and (3.10) (setting $\mathbf{J}_1 = 0$) with (A1), we find that (A3) and (A5) are equivalent, both yielding at $r = R$:

$$\Delta(-i\omega/cq^2)B'_{1z} = (km/R)E_{1z}\Delta(q^{-2}). \tag{A8}$$

The relevant discontinuity equations are then (A1), (A4), (A7), and (A8). We first consider the case $m \neq 0$, returning later to the sausage mode $m = 0$.

Outside the plasma channel E_{1z} and B_{1z} are given by the exponentially decaying solutions of (3.13) and (3.14), with σ and $\mathbf{J}_1$ set equal to zero:

$$E_{1z} = \mathcal{E}H^{(1)}_{|m|}(q_0r), \tag{A9}$$

$$B_{1z} = \mathcal{B}H^{(1)}_{|m|}(q_0r), \tag{A10}$$

where q_0 is the value of q with $\sigma = 0$,

$$q_0^2 = -k^2 + \omega^2/c^2; \quad \mathrm{Im}\, q_0 \geq 0, \tag{A11}$$

and $\mathcal{E}$, $\mathcal{B}$ are unknown constants. Dividing (A7) by E_{1z} and (A8) by B_{1z} then gives at $r = R$

$$\left(\frac{4\pi\sigma - i\omega}{q^2c}\right)\left(\frac{E'_{1z}}{E_{1z}}\right)_< = \frac{-i\omega}{q_0c}\frac{H^{(1)'}_{|m|}(q_0R)}{H^{(1)}_{|m|}(q_0R)} - \frac{km}{R}\left(\frac{\mathcal{B}}{\mathcal{E}}\right)\left(\frac{1}{q^2} - \frac{1}{q_0^2}\right), \tag{A12}$$

$$\left(\frac{-i\omega}{q^2c}\right)\left(\frac{B'_{1z}}{B_{1z}}\right)_< = \frac{-i\omega}{q_0c}\frac{H^{(1)'}_{|m|}(q_0R)}{H^{(1)}_{|m|}(q_0R)} + \frac{km}{R}\left(\frac{\mathcal{E}}{\mathcal{B}}\right)\left(\frac{1}{q^2} - \frac{1}{q_0^2}\right). \tag{A13}$$

The subscript "$<$" means that the logarithmic derivatives are evaluated just inside the plasma channel.

We now employ the approximations of Eqs. (1.2)–(1.5). If $|qR| \gg 1$ then there is nothing to do, since in

this case we already know that E_{1z} and B_{1z} become proportional to $H^{(1)}_{|m|}(qr)$ outside the beam. Therefore consider the more challenging problem, where $|qR|$ is roughly of order unity. The logarithmic derivatives $(E'_{1z}/E_{1z})_<$ and $(B'_{1z}/B_{1z})_<$ are then of order $R^{-1} \sim |q|$. Also, since $|q_0R| \ll 1$ we have (for $m \neq 0$)

$$H^{(1)'}_{|m|}(q_0R)/H^{(1)}_{|m|}(q_0R) \simeq -|m|/q_0R. \quad \text{(A14)}$$

But then the left-hand side of (A13) is smaller than the first term on the right by a factor of order $|q_0R|$, and using $|q| \gg |q_0|$ lets us write (A13) as

$$(\mathcal{E}/\mathcal{B}) \simeq (i\omega/kc)(m/|m|). \quad \text{(A15)}$$

Using (A14) and (A15) in (A12) (and taking $|q| \gg |q_0|$ and $\sigma \gg |\omega|$) gives finally

$$(E'_{1z}/E_{1z})_< \simeq -|m|/R \quad (m \neq 0). \quad \text{(A16)}$$

The fact that (A16) as well as the differential equation (6.1) is independent of the sign of m allows us to restrict our attention to the case $m \geq 0$.

For $m = 0$ we see directly from (A7) that

$$\left(\frac{4\pi\sigma - i\omega}{q^2c}\right)\left(\frac{E'_{1z}}{E_{1z}}\right)_< = -\frac{i\omega}{q_0c}\frac{H^{(1)'}_0(q_0R)}{H^{(1)}_0(q_0R)}. \quad \text{(A17)}$$

Instead of (A14) we now have, for $|q_0R| \ll 1$,

$$\frac{H^{(1)'}_0(q_0R)}{H^{(1)}_0(q_0R)} \simeq \frac{1}{q_0R \ln(-\frac{1}{4}C^2q_0^2R^2)}, \quad \text{(A18)}$$

where $\ln C = 0.577\cdots$. Hence for large σ, (A17) gives

$$\left(\frac{E'_{1z}}{E_{1z}}\right)_< = \left(\frac{\omega^2}{q_0^2c^2}\right)[R \ln(-\tfrac{1}{4}C^2q_0^2R^2)]^{-1}. \quad \text{(A19)}$$

For simplicity we take the condition $|q_0R| \ll 1$ as a strict limit, so that (A19) is taken to vanish.

Knowing $(E'_{1z}/E_{1z})_<$, we can now write the form of the solution outside the beam but inside the channel. In this region E_{1z} is a linear combination of $J_m(q_0r)$ and $H^{(1)}_m(q_0r)$, and imposing condition (A16) at $r = R$ gives this linear combination as (for $m \geq 0$)

$$E_{1z} \propto H^{(1)}_m(qr) - \{H^{(1)}_{m-1}(qR)/J_{m-1}(qR)\}J_m(qr) \quad (m \neq 0). \quad \text{(A20)}$$

[For $m = 1$ the ratio in brackets is just $H^{(1)}_0(qR)/J_0(qR)$, a result already familiar from Ref. 4.] The corresponding result for the sausage modes is

$$E_{1z} \propto H^{(1)}_0(qr) - [H^{(1)}_1(qR)/J_1(qR)]J_0(qr) \quad (m = 0). \quad \text{(A21)}$$

If the system were surrounded with a conducting shell at $R_1 > R$ the function $H^{(1)}_m(qr)$ in Eqs. (A9) and (A10) would be replaced with some linear combination $F_m(qr)$ of $H^{(r)}_m(qr)$ and $J_m(qr)$. But (A14) would hold for $F_m(qr)$ as well as $H^{(1)}_m(q_0r)$, (unless the coefficient of $H^{(1)}_m$ were much less than that of J_m) since for $|q_0R| \ll 1$, F_m is dominated by its $H^{(1)}_m$ term. Thus, it makes no difference what kind of tube the system is in, except that we require a vacuum immediately outside the plasma channel.

APPENDIX B. ANALYSIS OF THE UNIFORM BEAM DISPERSION RELATION

In this Appendix we show how the results quoted in Sec. IX are derived from the dispersion relation (8.12). We first classify the modes according to their low-frequency behavior, then classify them according to their high-frequency behavior, and then show how the two classification schemes are connected.

Low Frequency: $|q| \to 0$

For $|qa| \ll 1$ the dispersion relation (8.12) may be written

$$\frac{\eta}{qa}\frac{J'_m(qa/\eta)}{J_m(qa/\eta)} = -\frac{(m + \xi^2)}{q^2a^2} + O(1) \quad (m \neq 0), \quad \text{(B1)}$$

$$\frac{\eta}{qa}\frac{J'_0(qa/\eta)}{J_0(qa/\eta)} = \frac{1}{2}\left[\frac{R^2}{a^2} - 1\right] + O(q^2) \quad (m = 0). \quad \text{(B2)}$$

[Equation (B2) holds only if $|qR| \ll 1$; if $|qR| \gg 1$ then the right-hand side is infinite.] We divide the modes into those for which $|qa/\eta| \not\to 0$ at low frequency (called A and B) and those for which $|qa/\eta| \to 0$ at low frequency (called C and D).

A & B: $|qa/\eta| \not\to 0$

Here η must vanish at least as fast as qa. Equation (8.6) may be written

$$\eta^2 = \frac{\lambda^4 - 2\lambda^2(3 + m^2) + (m^2 - 4)(m^2 - 2)}{[\lambda^2 - (m + 2)^2][\lambda^2 - (m - 2)^2]}. \quad \text{(B3)}$$

The numerator of (B3) vanishes at two λ^2 values which define the modes of Types A and B:

$$A: \quad \lambda^2 = 3 + m^2 - (12m^2 + 1)^{\frac{1}{2}}, \quad \text{(B4)}$$

$$B: \quad \lambda^2 = 3 + m^2 + (12m^2 + 1)^{\frac{1}{2}}. \quad \text{(B5)}$$

However, the vanishing of the denominator prevents η^2 from vanishing for the cases $m = O(B)$ and $m = 2(A)$, so these modes are absent.

Direct calculation shows that $m + \xi^2$ does not vanish for $m \geq 1$ when λ^2 takes one of the values (B4) or (B5). Hence (B1) shows that for $m > 0$

$$\eta \to qa/j_{mn}, \quad \text{(B6)}$$

where j_{mn} is the nth positive root of $J_m(x)$. Solving

(B3) then gives for $m > 0$

$$A_{mn}: \quad \lambda^2 \to 3 + m^2 - (12m^2+1)^{\frac{1}{2}} - q^2a^2\left(1 - \frac{2m^2-1}{(12m^2+1)^{\frac{1}{2}}}\right)\Big/ j_{mn}^2 + \cdots, \tag{B7}$$

$$B_{mn}: \quad \lambda^2 \to 3 + m^2 + (12m^2+1)^{\frac{1}{2}} - q^2a^2\left(1 + \frac{2m^2-1}{(12m^2+1)^{\frac{1}{2}}}\right)\Big/ j_{mn}^2 + \cdots. \tag{B8}$$

For $m = 0$, Eq. (B2) can be written

$$\eta \to qa/y_n, \tag{B9}$$

where y_n is the nth root of the equation

$$J_0'(y)/yJ_0(y) = \tfrac{1}{2}[(R^2/a^2) - 1]. \tag{B10}$$

Solving (B3) then gives for $m = 0$

$$A_{0n}: \quad \lambda^2 \to 2 - 2q^2p^2/y_n^2 + \cdots. \tag{B11}$$

If $R \gg a$ then $y_n \simeq j_{0n}$.

C & D: $|qa/\eta| \to 0$

The left-hand side of (B2) becomes $-\frac{1}{2}$ for $qa/\eta \to 0$, so there can be no $m = 0$ mode of this type, and we restrict ourselves to the case $m > 0$.

For $qa/\eta \to 0$, Eq. (B1) becomes

$$m\eta^2 + m + \xi^2 = 0. \tag{B12}$$

We may write the left-hand side as a function of λ^2

$$m\eta^2 + m + \xi^2 = \frac{2[\lambda^4 - \lambda^2(2m^2 - 4m + 3) + m(m-(2)m-1)^2]}{[\lambda^2 - m^2][\lambda^2 - (2-m)^2]}. \tag{B13}$$

There are two zeros of the numerator, defining the modes of Types C and D:

$$C: \quad \lambda^2 = m^2 - 2m + \tfrac{3}{2} - (2m^2 - 4m + \tfrac{9}{4})^{\frac{1}{2}}, \tag{B14}$$

$$D: \quad \lambda^2 = m^2 - 2m + \tfrac{3}{2} + (2m^2 - 4m + \tfrac{9}{4})^{\frac{1}{2}}. \tag{B15}$$

However, the vanishing of the denominator prevents (B13) from vanishing for the cases $m = 1(D)$ and $m = 2(C)$, so these modes as well as $m = 0(C \,\&\, D)$ are absent.

In the $m = 1$ mode of Type C, Eq. (B14) shows that $\lambda^2 \to 0$, so this is the hose mode. In order to obtain λ^2 to order q^4 we need the following expansions: For $|\lambda| \ll 1$ and $m = 1$, Eqs. (8.6) and (8.7) give

$$\eta^2 \to \tfrac{1}{3} - \tfrac{14}{27}\lambda^2 - \tfrac{122}{243}\lambda^4 + \cdots, \tag{B16}$$

$$\xi^2 \to -\tfrac{4}{3} - \tfrac{40}{27}\lambda^2 - \tfrac{364}{243}\lambda^4 + \cdots. \tag{B17}$$

For $|qa/\eta| \ll 1$

$$\frac{\eta}{qa}\frac{J_1'(qa/\eta)}{J_1(qa/\eta)} \to \frac{\eta^2}{q^2a^2}\left[1 - \frac{q^2a^2}{4\eta^2} - \frac{q^4a^4}{96\eta^4} + \cdots\right]. \tag{B18}$$

For $|qa| \ll 1$ and $|qR| \ll 1$

$$\frac{H_0^{(1)}(qR)J_1'(qa) - J_0(qR)H_1^{(1)\prime}(qa)}{H_0^{(1)}(qR)J_1(qa) - J_0(qR)H_1^{(1)}(qa)} \to -\frac{1}{qa}\left[1 - q^2a^2\ln\frac{R}{a} + \tfrac{1}{2}q^4a^4\ln^2\frac{R}{a} + \tfrac{1}{2}q^4a^4\ln\frac{R}{a} - \tfrac{1}{4}q^4a^2(R^2-a^2) + \cdots\right]. \tag{B19}$$

Putting these expansions into (8.12) gives an implicit dispersion relation

$$\lambda^2 + \lambda^4 + \cdots = -\tfrac{1}{2}q^2a^2\left(\ln\frac{R}{a} + \frac{1}{4}\right) + \tfrac{1}{4}q^4a^4 \times \left(\ln^2\frac{R}{a} + \ln\frac{R}{a} + \frac{7}{16}\right) - \tfrac{1}{8}q^4a^2R^2 + \cdots. \tag{B20}$$

Solving (B20) to order q^4 gives the explicit dispersion relation

$$C_1: \quad \lambda^2 = -\tfrac{1}{2}q^2a^2\left(\ln\frac{R}{a} + \frac{1}{4}\right) - \tfrac{1}{4}q^4a^4\left(\ln\frac{R}{a} - \frac{3}{8}\right) - \tfrac{1}{8}q^4a^2R^2 + \cdots. \tag{B21}$$

For the C and D modes with $m \geq 2$ we can evaluate the q^2 terms in λ^2 by using the expansions

$$\frac{\eta}{qa}\frac{J_m'(qa/\eta)}{J_m(qa/\eta)} \to \frac{m\eta^2}{q^2a^2}\left[1 - \frac{q^2a^2}{2m(m+1)\eta^2} + \cdots\right], \tag{B22}$$

$$\frac{H_{m-1}^{(1)}(qR)J_m'(qa) - J_{m-1}(qR)H_m^{(1)\prime}(qa)}{H_{m-1}^{(1)}(qR)J_m(qa) - J_{m-1}(qR)H_m^{(1)}(qa)} \to -\frac{m}{qa}\left[1 - \frac{q^2a^2}{2m} + \frac{q^2R^2}{2m(m-1)}\left(\frac{a^2}{R^2}\right)^m + \cdots\right]. \tag{B23}$$

The implicit dispersion relation is now, for $m > 1$,

$$C_m, D_m: \quad m(\eta^2+1) + \xi^2 = \frac{m+2}{2(m+1)}q^2a^2 - \frac{1}{2(m-1)}\left(\frac{a^2}{R^2}\right)^m q^2R^2 + \cdots. \tag{B24}$$

We do not attempt to make this explicit for general m; the numerical results for $m = 2$ and $m = 3$ are included in Table I.

High Frequency: $|q| \to \infty$

For $|qa| \gg 1$ the dispersion relation (8.12) may be written

$$\eta\left(\frac{J_m'(qa/\eta)}{J_m(qa/\eta)}\right) + \frac{\xi^2}{qa} = i + O(1/qa). \tag{B25}$$

We divide the modes into those for which $|qa/\eta| \nrightarrow \infty$ at high frequency (called Types $\mathcal{A}$ and $\mathcal{B}$) and those for which $|qa/\eta| \to \infty$ at high frequency (called $\mathcal{D}$).

$\mathcal{A}$ & $\mathcal{B}$: $|qa/\eta| \nrightarrow \infty$.

Here η must go to infinity at least as fast as qa. From (B3) we see that there are two possible limits for λ^2, which define the modes of Types $\mathcal{A}$ and $\mathcal{B}$:

$$\mathcal{A}: \quad \lambda^2 \to (2-m)^2, \tag{B26}$$

$$\mathcal{B}: \quad \lambda^2 \to (2+m)^2. \tag{B27}$$

For $m = 0$ the two limits are the same, and we call the mode Type $\mathcal{A}$. For $m = 2$ the Type $\mathcal{A}$ mode is absent because the vanishing of the numerator of (B3) prevents it from becoming infinite when $\lambda^2 \to 0$.

In order to obtain the $1/q^2$ term in λ^2 we write (B25) as

$$[J'_m(x)/xJ_m(x)] + (\xi^2/x^2\eta^2) \to 0, \tag{B28}$$

where

$$x \equiv qa/\eta. \tag{B29}$$

$\mathcal{A}$: When λ^2 approaches the limit (B26) we find

$$\eta^2 \to (2-m)/2[\lambda^2 - (m-2)^2], \tag{B30}$$

$$\xi^2/\eta^2 \to m, \tag{B31}$$

except that for $m = 0$

$$\eta^2 \to 2/(\lambda^2 - 4), \tag{B32}$$

and for $m = 1$

$$\xi^2/\eta^2 \to 3. \tag{B33}$$

Hence for $m \neq 0$ the dispersion relation is

$$\mathcal{A}_{mn}: \quad \lambda^2 \to (m-2)^2 - (m-2)x^2/2q^2a^2 + \cdots, \tag{B34}$$

while for $m = 0$

$$\mathcal{A}_{0n}: \quad \lambda^2 \to 4 + 2x^2/q^2a^2 + \cdots, \tag{B35}$$

where x is a root of Eq. (B28), which becomes

$$xJ'_m(x) = -mJ_m(x), \tag{B36}$$

except that for $m = 1$

$$xJ'_1(x) = -3J_1(x). \tag{B37}$$

$\mathcal{B}$: When λ^2 approaches the limit (B27) we find

$$\eta^2 \to (m+2)/2[\lambda^2 - (m+2)^2], \tag{B38}$$

$$\xi^2/\eta^2 \to -m. \tag{B39}$$

Hence the dispersion relation is

$$\mathcal{B}_{mn}: \quad \lambda^2 \to (m+2)^2 + (m+2)x^2/2q^2a^2 + \cdots, \tag{B40}$$

where x is a root of (B28), which becomes

$$xJ'_m(x) = +mJ_m(x). \tag{B41}$$

Some standard Besseling shows that the roots of (B36) and (B41) are, respectively,

$$\mathcal{A}: \quad x = j_{|m-1|,n}, \qquad \mathcal{B}: \quad x = j_{m+1,n}, \tag{B42}$$

where $j_{p,n}$ is the nth root of $J_p(x)$. The roots of (B37) must be calculated separately, and there are an infinite number of them.

$\mathcal{D}$: $|qa/\eta| \to \infty$

Suppose we choose η as the root of (8.6) such that qa/η goes to infinity with positive imaginary part. Then the dispersion relation (B25) for $|qa| \to \infty$ becomes

$$-i\eta + \xi^2/qa \to -i. \tag{B43}$$

If ξ^2 stays finite then $\eta \to -1$, which is impossible since we define q with $\mathrm{Im}\, q > 0$ and η with $\mathrm{Im}\,(qa/\eta) > 0$. Thus $\xi^2 \to \infty$ at least as fast as qa. If η also goes to infinity we have η^2/ξ^2 asymptotically constant for $m > 0$ (vide supra) so according to (B43) η must go like qa, which is impossible since $|qa/\eta|$ is assumed to diverge. Hence we must have $\xi^2 \to \infty$ but $\eta^2 \nrightarrow \infty$. Inspection of (8.6) and (8.7) shows that this happens when

$$\lambda^2 \to m^2, \tag{B44}$$

except that for $m = 1$ the vanishing of the numerator of (8.7) prevents ξ^2 from becoming infinite when $\lambda^2 \to 1$, and of course ξ^2 for $m = 0$ is identically zero, as the type $\mathcal{D}$ modes start with $m \geq 2$. When $\lambda^2 \to m^2$, Eqs. (8.6) and (8.7) give

$$\eta \to \eta_\infty = [(3m^2-2)/4(m^2-1)]^{\frac{1}{2}}, \tag{B45}$$

$$\xi^2 \to m^2/(\lambda^2 - m^2), \tag{B46}$$

so (B43) becomes

$$\mathcal{D}_m: \quad \lambda^2 \to m^2[1+\xi^2] \to m^2\left[1 + \frac{1}{iqa(1+\eta_\infty)}\right]. \tag{B47}$$

The derivation shows that the square root in (B45) must be taken positive.

Connection: $q^2 < 0$

It should be noted that for purely imaginary q, the high- and low-frequency formulas (B7), (B8), (B11), (B21), (B24), and (B34), (B35), (B40), (B47) all give λ real. The relative ordering of the values of λ^2 at low frequencies is

$$\begin{aligned}
m = 0: &\quad A_{01} > A_{02} > \cdots, \\
m = 1: &\quad B_{11} > B_{12} > \cdots > A_{11} \\
&\qquad > A_{12} > \cdots > C_1, \\
m = 2: &\quad B_{21} > B_{22} > \cdots > D_2, \\
m \geq 3: &\quad B_{m1} > B_{m2} > \cdots > D_m > C_m \\
&\qquad > \cdots > A_{m2} > A_{m1},
\end{aligned}$$

where ">" refers to the λ^2 values of the modes indicated. At high frequencies the relative ordering is

$$\begin{aligned} m=0:&\quad \mathcal{A}_{01} > \mathcal{A}_{02} > \cdots,\\ m=1:&\quad \mathcal{B}_{11} > \mathcal{B}_{12} > \cdots > \mathcal{A}_{11} > \mathcal{A}_{12} > \cdots,\\ m=2:&\quad \mathcal{B}_{21} > \mathcal{B}_{22} > \cdots > \mathcal{D}_2,\\ m\geq 3:&\quad \mathcal{B}_{m1} > \mathcal{B}_{m2} > \cdots > \mathcal{D}_m\\ &\qquad > \cdots > \mathcal{A}_{m2} > \mathcal{A}_{m1}. \end{aligned}$$

(The missing modes are B_{0n}, C_0, D_0, D_1, A_{2n}, C_2 at low frequency and $\mathcal{B}_{0n}$, $\mathcal{D}_0$, $\mathcal{D}_1$, $\mathcal{A}_{2n}$ at high frequency.) If one could be sure that for a given m and $q^2 < 0$ the various λ^2 are real continuous functions of $-q^2$ which do not cross, then we could immediately conclude that the modes $\mathcal{A}$, $\mathcal{B}$, $\mathcal{D}$ labeled by their behavior for $q^2 \to -\infty$ are to be identified with the corresponding A, B, D modes labeled by their behavior for $q^2 \to 0$. That is, we would guess that $\mathcal{A}_{mn} \equiv A_{mn}$, $\mathcal{B}_{mn} \equiv B_{mn}$, and $\mathcal{D}_m \equiv D_m$.

But the trouble with this guess is that the C_m modes are left out in the cold, there being no $\mathcal{C}_m$ modes they can hook onto at high frequency. A careful reinspection of the ordering of the λ^2 will convince the reader that, as we move away from zero on the negative q^2 axis, the C_m modes must either cross the others or reach a singularity.

In fact, what happens is a singularity. In order to see this without inessential complications, we take the plasma channel radius R infinite, and write the dispersion relation (8.12) for $q^2 < 0$ as

$$\eta \frac{I_m'(Qa/\eta)}{I_m(Qa/\eta)} + \frac{\xi^2}{Qa} = \frac{K_m'(Qa)}{K_m(Qa)}, \tag{B48}$$

where $q = iQ$, with $Q > 0$. For $Q \to 0$ it is easy to check from (B14) that $\eta^2 > 0$. As long as η^2 stays positive (B48) requires that

$$m\eta^2 + \xi^2 + m < 0, \tag{B49}$$

since $xI_m'(x)/I_m(x) > m$ and $xK_m'(x)/K_m(x) < -m$ for real positive x. However, it is easy to see that (B48) has at some finite Qa a solution with $\eta^2 \to 0_+$. When $\eta^2 \to 0$ (B49) requires that we take the root with

$$\lambda^2 \to 3 + m^2 - (12m^2 + 1)^{\frac{1}{2}} \equiv \lambda_1^2. \tag{B50}$$

[See (B3) and (B13).] In this case ξ^2 approaches a finite limit $\xi_1^2 < -m$, and (B48) gives $Q \to Q_1$, where

$$\xi_1^2 = Q_1 a K_m'(Q_1 a)/K_m(Q_1 a), \tag{B51}$$

an equation with precisely one solution. For instance, in the hose mode C_1 we have

$$\begin{aligned} \lambda_1^2 &= 4 - (13)^{\frac{1}{2}} = 0.39,\\ \xi_1^2 &= -6/[(13)^{\frac{1}{2}} - 1] = -2.29, \end{aligned} \tag{B52}$$

and (B51) gives

$$Q_1 a \simeq 2.3. \tag{B53}$$

However, it is not possible for η^2 to return above zero when Q increases past Q_1. For $\xi^2 - \xi_1^2$ is of order η^2 when $\eta \to 0$, so (B48) gives

$$\begin{aligned} |\eta| \to \left[\frac{d}{dx}\frac{1}{x}[-\xi_1^2 + xK_m'(x)/K_m(x)]\right]_{x=Q_1 a} (Qa - Q_1 a)\\ = \left(\frac{Qa}{Q_1 a} - 1\right)\left[\frac{d}{dx} xK_m'(x)/K_m(x)\right]_{x=Q_1 a}. \end{aligned} \tag{B54}$$

The derivative in (B54) is negative, so this is only possible for η^2 to go to zero from above when Q goes to Q_1 from below.

It is also not possible for η^2 to pass through zero to negative values as Q passes Q_1, since the ratio $I_m'(Q_1 a/\eta)/I_m(Q_1 a/\eta)$ oscillates between $+\infty$ and $-\infty$ as $\eta^2 \to 0_-$. We can only conclude that the function λ^2 simply comes to an end in the C_m modes when $-q^2$ reaches the value Q_1^2.

There is still the question of what happens to λ^2 in the C_m modes when $q \to \infty$ along some direction other than the imaginary axis. Here, λ^2 is complex and we do not expect to encounter a break in the C_m mode. From the way the λ^2 values are ordered for $q^2 < 0$, we might expect that when q^2 is only slightly above the negative real axis the C_m mode λ^2 will be nearly real for Re $(-q^2)$ increasing from zero to Q_1^2, when λ^2 becomes highly complex and passes *over* an infinite number of A_{mn} modes to hook on finally to a low $\mathcal{A}_{mn}$ mode, probably $\mathcal{A}_{m1}$. In fact, a machine calculation[8] shows that this is what happens in the C_1 ("hose") mode. Following the λ^2 function from its known behavior at $q = 0$ up to large imaginary values of q^2a^2 (i.e., ω large and real) shows the C_1 asymptotic behavior is what we have above called the $\mathcal{A}_{11}$ mode. Assuming the same to happen for $m \geq 3$, we finally conclude that the correct identifications must be

$$\begin{aligned} A_{m,n} \equiv \mathcal{A}_{m,n+1},\quad B_{m,n} \equiv \mathcal{B}_{m,n},\\ C_m \equiv \mathcal{A}_{m,1},\quad D_m \equiv \mathcal{D}_m. \end{aligned} \tag{B55}$$

APPENDIX C. WKB SOLUTION OF EQUATION (11.15)

We first write (11.15) in the Riccatti form

$$\begin{aligned} -\frac{1}{\rho}\frac{d}{d\rho}\left[\frac{\rho}{\rho^{2N}-1}\left(\frac{\mathcal{E}'}{\mathcal{E}}\right)\right] + \frac{m^2}{\rho^2}\frac{1}{\rho^{2N}-1}\\ \mp \frac{2mN\rho^{2N-2}}{(\rho^{2N}-1)^2} + \mu^2 = \frac{1}{\rho^{2N}-1}\left(\frac{\mathcal{E}'}{\mathcal{E}}\right)^2. \end{aligned} \tag{C1}$$

For large μ this gives $\mathcal{E}'/\mathcal{E} \simeq \mu[\rho^{2N}-1]^{\frac{1}{2}}$, which when inserted in the left-hand side of (6.1) yields

$$\mathcal{E}'/\mathcal{E} = \mu[\rho^{2N}-1]^{\frac{1}{2}} - \frac{[\rho^{2N}-1]^{\frac{1}{2}}}{2\rho} \times \frac{d}{d\rho}\frac{\rho}{[\rho^{2N}-1]^{\frac{1}{2}}} + O\left(\frac{1}{\mu}\right), \quad \text{(C2)}$$

either sign being taken for the square root of $\rho^{2N}-1$. Integrating (C2), we find our WKB solutions

$$\mathcal{E} \propto \rho^{-\frac{1}{2}}[\rho^{2N}-1]^{\frac{1}{4}} \exp\left(\mu\int d\rho[\rho^{2N}-1]^{\frac{1}{2}}\right). \quad \text{(C3)}$$

These solutions are valid except very near $\rho = 0$ and $\rho = 1$.

For $\rho \ll 1$, Eq. (11.15) becomes just Bessel's equation

$$\frac{1}{\rho}\frac{d}{d\rho}\rho\frac{d\mathcal{E}}{d\rho} - \frac{m^2}{\rho^2}\mathcal{E} + \mu^2\mathcal{E} = 0 \quad \text{(C4)}$$

with solution

$$\mathcal{E} \propto J_m(\mu\rho). \quad \text{(C5)}$$

If μ^2 is large we can find a region

$$\mu^{-1} \ll \rho \ll 1 \quad \text{(C6)}$$

within which the solution (C5) becomes

$$\mathcal{E} \propto \rho^{-\frac{1}{2}}\cos\left(\mu\rho - \tfrac{1}{2}m\pi - \frac{\pi}{4}\right). \quad \text{(C7)}$$

Comparing with (C3), we see that this may be written in the WKB form

$$\mathcal{E} \propto \rho^{-\frac{1}{2}}[1-\rho^{2N}]^{\frac{1}{4}} \times \cos\left(\mu\int_0^\rho d\rho[1-\rho^{2N}]^{\frac{1}{2}} - \tfrac{1}{2}m\pi - \tfrac{1}{4}\pi\right) \quad \text{(C8)}$$

Thus (C8) is a good solution for $\rho < 1$, except very close to the turning point at $\rho = 1$.

For ρ very close to unity we may write

$$\rho = 1 + x(2N\mu^2)^{-\frac{1}{3}}, \quad \text{(C9)}$$

and Eq. (11.15) becomes

$$\frac{d}{dx}\frac{1}{x}\frac{d\mathcal{E}}{dx} - \mathcal{E} = 0. \quad \text{(C10)}$$

We find two regular solutions

$$\mathcal{E}_\pm = xI_{\pm\frac{2}{3}}[\tfrac{2}{3}x^{\frac{3}{2}}] \quad (x > 0), \quad \text{(C11)}$$

$$\mathcal{E}_\pm = xJ_{\pm\frac{2}{3}}[\tfrac{2}{3}(-x)^{\frac{3}{2}}] \quad (x < 0). \quad \text{(C12)}$$

It is easy to check that these forms are correctly normalized to join smoothly at $\rho = 1$, since both (C11) and (C12) agree that for $x \to 0$

$$\mathcal{E}_+ \to [2^{\frac{2}{3}}/\Gamma(\tfrac{4}{3})(3)^{\frac{4}{3}}]x^2[1 + \tfrac{1}{14}x^3 + \cdots],$$
$$\mathcal{E}_- \to [2^{\frac{2}{3}}/\Gamma(\tfrac{1}{3})][1 + \tfrac{1}{3}x^3 + \cdots].$$

For sufficiently large μ^2 there is a region of $|x|$:

$$1 \ll |x| \ll \mu^{\frac{2}{3}} \quad \text{(C13)}$$

within which we may use the well-known asymptotic forms of (C11) and (C12)

$$\mathcal{E}_\pm \to (2\pi)^{-\frac{1}{2}}x^{\frac{1}{4}}[\exp\{-\tfrac{2}{3}x^{\frac{3}{2}} - \tfrac{1}{4}i\pi \mp \tfrac{2}{3}i\pi\} + \exp\{\tfrac{2}{3}x^{\frac{3}{2}} + \tfrac{1}{4}i\pi\}] \quad (1 \ll x \ll \mu^{\frac{2}{3}}), \quad \text{(C14)}$$

$$\mathcal{E}_\pm \to (2/\pi)^{\frac{1}{2}}(-x)^{\frac{1}{4}}\cos[\tfrac{2}{3}(-x)^{\frac{3}{2}} \mp \tfrac{1}{3}\pi - \tfrac{1}{4}\pi] \quad (1 \ll -x \ll \mu^{\frac{2}{3}}). \quad \text{(C15)}$$

The correct linear combination of the solutions (C14) is the one that decays exponentially, i.e.,

$$\mathcal{E} \propto \mathcal{E}_+ - \mathcal{E}_-, \quad \text{(C16)}$$

and for $x < 0$ this is given by (C15) as

$$\mathcal{E} \propto (-x)^{\frac{1}{4}}\cos[\tfrac{2}{3}(-x)^{\frac{3}{2}} - \tfrac{1}{4}\pi]. \quad \text{(C17)}$$

But when x is negative and in the region (C14) we have $\rho \simeq 1$, $(1-\rho^{2N})^{\frac{1}{4}} \propto (-x)^{\frac{1}{4}}$, and

$$\mu\int_\rho^1 [1-\rho^{2N}]^{\frac{1}{2}}\,d\rho \simeq \tfrac{2}{3}(-x)^{\frac{3}{2}}, \quad \text{(C18)}$$

so (C17) may be written in a WKB form as

$$\mathcal{E} \propto \rho^{-\frac{1}{2}}[1-\rho^{2N}]^{\frac{1}{4}}\cos\left(\mu\int_\rho^1 [1-\rho^{2N}]^{\frac{1}{2}}\,d\rho - \tfrac{1}{4}\pi\right). \quad \text{(C19)}$$

The two forms (C8) and (C19) agree if

$$\mu\int_0^\alpha [1-\rho^{2N}]^{\frac{1}{2}}\,d\rho = (n + \tfrac{1}{2}m + \tfrac{1}{2})\pi \quad \text{(C20)}$$

with n an integer. The integral is just a Beta function

$$\int_0^1 [1-\rho^{2N}]^{\frac{1}{2}}\,d\rho = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}N)}{2(N+1)\Gamma\left(\frac{1}{2}+\frac{1}{2N}\right)}. \quad \text{(C21)}$$

APPENDIX D. THE SAUSAGE MODES

The $m = 0$ case is sufficiently simple to allow the proof of some useful rigorous results. We show that for $m = 0$ the only possible modes are the A_0 modes and the continuous spectrum.

For $m = 0$ the functions $f(r)$ and $g(r)$ are

$$f(r) = [r\alpha^{2\prime}(r) + 2\alpha^2(r)]/[r\alpha^{2\prime}(r) + 4\alpha^2(r) - \Omega^2], \quad \text{(D1)}$$

$$g(r) = 0, \quad \text{(D2)}$$

and Eq. (6.1) becomes

$$\frac{1}{r}\frac{d}{dr} r[1 - f(r)] \frac{d}{dr} \mathcal{E}(r) + q^2 \mathcal{E}(r) = 0. \quad \text{(D3)}$$

Multiply (D3) by $r\mathcal{E}^*(r)$, multiply the complex conjugate of (D3) by $r\mathcal{E}(r)$, subtract, and integrate from 0 to R; this gives

$$\int_0^R r \operatorname{Im} f(r) \, |\mathcal{E}'(r)|^2 \, dr = -\operatorname{Im} q^2 \int_0^R r \, |\mathcal{E}(r)|^2 \, dr. \quad \text{(D4)}$$

But using (D1), this is

$$\operatorname{Im} \Omega^2 \int_0^R \frac{r\alpha^{2\prime}(r) + 2\alpha^2(r)}{|r\alpha^{2\prime}(r) + 4\alpha^2(r) - \Omega^2|^2} \, |\mathcal{E}'(r)|^2 r \, dr = -\operatorname{Im} q^2 \int_0^R r \, |\mathcal{E}(r)|^2 \, dr. \quad \text{(D5)}$$

Thus when q^2 is real, Ω^2 is real. For $q^2 < 0$ the boundary conditions on $\mathcal{E}(r)$ are also real, so $\mathcal{E}(r)$ is everywhere real.

Now, multiply Eq. (D3) by $r\mathcal{E}(r)$ and integrate from 0 to R. We find

$$\int_0^R r[1 - f(r)]\mathcal{E}'^2(r) \, dr = q^2 \int_0^R r\mathcal{E}^2(r) \, dr.$$

Hence for $q^2 < 0$, the real function $1 - f(r)$ must be negative somewhere. Outside the beam $1 - f(r)$ is unity, so we conclude that $1 - f(r)$ must pass below zero either by passing through zero or through infinity; in the former case we are in the continuous spectrum (10.12), while in the latter case we have an A_0 mode satisfying (11.40).

Note incidentally that (D4) does not imply that q^2 must be real for real Ω^2 in the continuous spectrum, for when (10.12) is satisfied the function $\mathcal{E}(r)$ has a logarithmic singularity which invalidates (D4).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Runge–Lenz Vector and the Coulomb Green's Function

LEVERE HOSTLER*
University of Pittsburgh, Pittsburgh, Pennsylvania and Yale University, New Haven, Connecticut

(Received 15 November 1965)

The fact that the Runge–Lenz vector is an extra constant of the motion for a charged particle moving in a Coulomb potential is found to account for the especially simple structure of the nonrelativistic coordinate space Coulomb Green's function. Also, the study of the consequences of this extra constant of the motion leads to a separation of variables in the differential equation for the coordinate space Coulomb Green's function, and hence to a new derivation of the closed-form expression for this Green's function which avoids the use of infinite series and the detailed properties of special functions.

I. INTRODUCTION

THE Runge–Lenz vector,

$$\mathbf{A} = \tfrac{1}{2}(\mathbf{p} \times \mathbf{L} - \mathbf{L} \times \mathbf{p}) - (mZe^2/4\pi)\mathbf{u}_r, \quad \mathbf{u}_r = \mathbf{r}/r \quad (1)$$

for particle motion in a Coulomb potential is a constant of the motion both classically and quantum mechanically.[1,2] Classically,[3,4] the constancy of $\mathbf{A}$ expresses the fact that the orbit of a particle moving in a Coulomb potential does not precess. The Runge–Lenz vector, $\mathbf{A}$, together with the orbital angular momentum vector, $\mathbf{L}$, provide a complete set of integration constants for the classical equations of motion excepting only for the initial position of the particle in its orbit. Also, the equation of the orbit follows from the constancy of $\mathbf{A}$ in a remarkably simple way. In the quantum-mechanical Kepler problem too, the Runge–Lenz vector has proved to be useful. With the help of this vector the Bohr energy levels can be deduced strictly within the framework of Heisenberg's matrix mechanics.[4,5] The existence of this extra constant of the motion—in addition to the orbital angular momentum vector (which is always a constant of the motion for a spherically symmetrical potential)—is equivalent to an additional symmetry of the Coulomb Hamiltonian beyond just invariance under spatial rotations.

We here wish to investigate the consequences of the existence of this extra vector constant of the motion for the Coulomb Green's function in coordinate space. This function may be defined by the differential equation[6]

$$\left(\nabla_2^2 + \frac{2kv}{r_2} + k^2\right) G(\mathbf{r}_2, \mathbf{r}_1, \omega) = \delta^3(\mathbf{r}_2 - \mathbf{r}_1),$$
$$k = \left(\frac{2m\omega}{\hbar}\right)^{\frac{1}{2}}, \quad \mathrm{Im}\,(k) > 0, \quad v = \frac{Ze^2m}{4\pi k\hbar^2}, \quad (2)$$

subject to certain regularity conditions at the origin and at infinity, or equivalently by the equation

$$G(\mathbf{r}_2, \mathbf{r}_1, \omega) = \langle \mathbf{r}_2 |\, G(\omega)\, | \mathbf{r}_1 \rangle,$$
$$G(\omega) = -\hbar^2/2m(H - \hbar\omega)]. \quad (3)$$

The quantity $\hbar\omega$ in (2) and (3) is any complex number not in the eigenvalue spectrum (discrete and continuous) of the Coulomb Hamiltonian,

$$H = p^2/2m - Ze^2/4\pi r.$$

That this investigation might lead to some interesting results is suggested by the specific form of the function $G(\mathbf{r}_2, \mathbf{r}_1, \omega)$. This is[6–8]

$$G(\mathbf{r}_2, \mathbf{r}_1, \omega) = -\frac{\Gamma(1 - iv)}{4\pi\,|\mathbf{r}_2 - \mathbf{r}_1|} \det \begin{bmatrix} W_{iv;\frac{1}{2}}(-iku)\mathcal{M}_{iv;\frac{1}{2}}(-ikv) \\ \dot{W}_{iv;\frac{1}{2}}(-iku)\mathcal{M}_{iv;\frac{1}{2}}(-ikv) \end{bmatrix},$$
$$u = r_2 + r_1 + |\mathbf{r}_2 - \mathbf{r}_1|, \quad v = r_2 + r_1 - |\mathbf{r}_2 - \mathbf{r}_1|. \quad (4)$$

The functions W and $\mathcal{M}$ here are Whittaker functions as defined in Buchholz.[9] The dots over the Whittaker functions denote differentiation with respect to the arguments of the Whittaker functions. The Green's function is seen to depend upon $\mathbf{r}_2$ and $\mathbf{r}_1$ only through the two variables u and v [the factor $|\mathbf{r}_2 - \mathbf{r}_1|^{-1}$ which occurs in (4) can be written $2(u - v)^{-1}$]. Now, on the basis of the spherical symmetry of the potential alone

* Present address: University of Minnesota, Minneapolis, Minnesota.

[1] Equation (1) is written in its quantum-mechanical form.

[2] Heaviside–Lorentz (equals rationalized Gaussian) units are used.

[3] C. Runge, *Vector Analysis* (E. P. Dutton and Company, Inc., New York, 1919), Chap. II, Sec. 5, p. 79.

[4] M. Born and P. Jordan, *Elementare Quantenmechanik* (Springer–Verlag, Berlin, 1930), Chap. 4, Sec. 35, p. 179.

[5] W. Pauli, Jr., Z. Physik **36**, 336 (1926).

[6] L. Hostler, J. Math. Phys. **5**, 591 (1964).

[7] K. Mano, J. Math. Phys. **5**, 505 (1964).

[8] L. Hostler and R. H. Pratt, Phys. Rev. Letters **10**, 469 (1963).

[9] H. Buchholz, *Die Konfluente Hypergeometrische Funktion* (Springer–Verlag, Berlin, 1953), p. 12, Eq. (7); p. 22, Eq. (25a).

one can only deduce that the Green's function must depend upon $\mathbf{r}_2$ and $\mathbf{r}_1$ through the *three* variables r_2, r_1, and $|\mathbf{r}_2 - \mathbf{r}_1|$, or three functionally independent combinations of these. The Green's function (4) therefore has a structure simpler than that required by invariance under spatial rotations alone. It is tempting to speculate that this special simplicity of the Coulomb Green's function may be a consequence of the additional symmetry of the Coulomb Hamiltonian, associated with the extra constant of the motion, **A**.

This speculation is confirmed by the work presented here (Sec. II). In addition (in Sec. III), we are led to a method of reducing the partial differential equation for the coordinate space Coulomb Green's function to a pair of uncoupled ordinary differential equations [Eqs. (27)] for the functional dependence on the remaining two variables u and v. Hence we can give a derivation of the closed-form expression (4) which avoids the use of infinite series and the detailed properties of special functions.[10,11]

II. CONSEQUENCES OF [A, H] = 0

To begin our investigation, we note that the constancy of **A** is equivalent to the commutability of **A** with the Coulomb Hamiltonian, H. But if **A** commutes with the Coulomb Hamiltonian, then it also commutes with any function $\mathcal{F}(H)$ of the Coulomb Hamiltonian,

$$\mathbf{A}\mathcal{F}(H) - \mathcal{F}(H)\mathbf{A} = 0. \tag{5}$$

Taking matrix elements of this equation relative to a basis of position eigenfunctions gives an identity which must be satisfied by the coordinate space representative,

$$\mathcal{F}(\mathbf{r}_2, \mathbf{r}_1) = \langle \mathbf{r}_2 | \mathcal{F}(H) | \mathbf{r}_1 \rangle, \tag{6}$$

of the operator $\mathcal{F}(H)$,

$$(-\mathbf{r}_2\nabla_2^2 + \mathbf{r}_2 \cdot \boldsymbol{\nabla}_2\boldsymbol{\nabla}_2 + \boldsymbol{\nabla}_2 - \mathbf{u}_2 a_1^{-1})\mathcal{F}(\mathbf{r}_2, \mathbf{r}_1) = (-\mathbf{r}_1\nabla_1^2 + \mathbf{r}_1 \cdot \boldsymbol{\nabla}_1\boldsymbol{\nabla}_1 + \boldsymbol{\nabla}_1 - \mathbf{u}_1 a_1^{-1})\mathcal{F}(\mathbf{r}_2, \mathbf{r}_1), \tag{7}$$

[10] This derivation is quite similar to the one given in Ref. 8, and the present work might be regarded as a refinement of this previous work.

[11] Actually, as far as the momentum space representative is concerned, the question of the connection between the Runge–Lenz vector and the Coulomb Green's function is already answered by a work of Schwinger [J. Schwinger, J. Math. Phys. **5**, 1606 (1964)]. As shown by Fock [V. Fock, Z. Physik **98**, 145 (1936); see, also, L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961)], the full symmetry of the Coulomb problem associated with the constants of the motion **A** and **L** is the symmetry of the four-dimensional rotation group. In Schwinger's treatment of the momentum space Coulomb Green's function, this four-dimensional rotational invariance plays a central role. He shows that the equation of the momentum space Coulomb Green's function can be written as an integral equation in a four-dimensional spherical space, and the four-dimensional rotational invariance of the equations is manifest throughout the calculation.

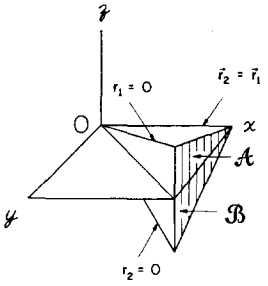


FIG. 1. This figure shows part of the infinite three-sided pyramid which is the domain of the variables x, y, and z of Eq. (8). The shaded area is a section through the pyramid at x = const. (The scale is distorted for ease of visualization.)

a_1, the radius of the first Bohr orbit, equals $4\pi\hbar^2/mZe^2$. Here $\mathbf{u}_{1,2}$ denote unit vectors in the directions $\mathbf{r}_{1,2}$. Choosing $\mathcal{F}(H) = -\hbar^2/2m(H - \hbar\omega)$, one finds that the coordinate space Coulomb Green's function, $G(\mathbf{r}_2, \mathbf{r}_1, \omega)$, satisfies this identity, but we need not specialize to this case until the end of the calculation.

The differential equation (7) involves the six variables $\mathbf{r}_1$ and $\mathbf{r}_2$. However, from the invariance of the Hamiltonian, H, under spatial rotations, it follows that $\mathcal{F}(\mathbf{r}_2, \mathbf{r}_1)$ can depend upon $\mathbf{r}_2$ and $\mathbf{r}_1$ only through the variables r_2, r_1, and $|\mathbf{r}_2 - \mathbf{r}_1|$, or three functionally independent combinations of these variables. This enables us to rewrite (7) as a differential equation involving only three independent variables. Writing $\mathcal{F}(\mathbf{r}_2, \mathbf{r}_1) = \mathcal{F}(x, y, z)$, where

$$x = r_1 + r_2, \quad y = |\mathbf{r}_2 - \mathbf{r}_1|, \quad z = r_2 - r_1, \tag{8}$$

Eq. (7) goes over into

$$\left[\frac{\partial^2\mathcal{F}}{\partial y \partial z}\frac{z^2 - y^2}{2y}(\mathbf{u}_2 + \mathbf{u}_1) + \frac{\partial^2\mathcal{F}}{\partial x \partial y}\frac{x^2 - y^2}{2y}(\mathbf{u}_2 - \mathbf{u}_1) - \frac{\partial\mathcal{F}}{\partial z}(\mathbf{u}_2 + \mathbf{u}_1) - \frac{\partial\mathcal{F}}{\partial x}(\mathbf{u}_2 - \mathbf{u}_1) - \mathcal{F}a_1^{-1}(\mathbf{u}_2 - \mathbf{u}_1)\right] = 0, \quad r_2 > 0, \quad r_1 > 0, \quad \mathbf{r}_2 \neq \mathbf{r}_1. \tag{9}$$

The domain of the variables x, y, z is the infinite three-sided pyramid of Fig. 1. As $\mathbf{r}_2$ and $\mathbf{r}_1$ run over all possible values; x varies over the range $0 \leq x < +\infty$. For fixed x, $y = |\mathbf{r}_2 - \mathbf{r}_1|$ varies over the range $0 \leq y \leq x$. For fixed x and y, $z = r_2 - r_1$ varies over the range $-y \leq z \leq +y$. The three edges of the pyramid are the three rays $y = z = 0$, $x \geq 0$; $y = x = z \geq 0$; and $y = x = -z \geq 0$. These correspond to $\mathbf{r}_2$, $\mathbf{r}_1$ values for which $\mathbf{r}_2 = \mathbf{r}_1$, $r_1 = 0$, and $r_2 = 0$, respectively.

The conditions $r_2 > 0$, $r_1 > 0$, $\mathbf{r}_2 \neq \mathbf{r}_1$ associated with Eq. (9) express the fact that the domain of validity of Eq. (9) is the pyramid less its three edges. This excludes the exceptional points $r_1 = 0$, $r_2 = 0$, and $\mathbf{r}_2 = \mathbf{r}_1$ which occur in the derivation of Eq. (9), and guarantees that the expressions entering Eq. (9) are

well defined. The points $r_2 = 0$, $r_1 = 0$, and $\mathbf{r}_2 = \mathbf{r}_1$ may also be singular points of the function $\mathcal{F}(x, y, z)$.

The xy plane cuts the region $r_2 > 0$, $r_1 > 0$, $\mathbf{r}_2 \neq \mathbf{r}_1$—i.e., the pyramid less the three edges—into two halves, denoted by $\mathcal{A}$ and $\mathcal{B}$ (see Fig. 1). These consist of the points corresponding to $r_2 > r_1 > 0$ and $r_1 > r_2 > 0$, respectively. In the following analysis we limit ourselves to the study of the function $\mathcal{F}(x, y, z)$ in the region $\mathcal{A}$. It is obvious that the same considerations apply also to the region $\mathcal{B}$. In the end we have to build up the function $\mathcal{F}(x, y, z)$ by piecing together its values in the two regions.

The vector equation (9) is equivalent to the two scalar equations

$$\tfrac{1}{2}(z^2 - y^2)(\partial^2\mathcal{F}/\partial y\partial z) - y(\partial\mathcal{F}/\partial z) = 0, \tag{10}$$

$$\tfrac{1}{2}(x^2 - y^2)(\partial^2\mathcal{F}/\partial x\partial y) - y(\partial\mathcal{F}/\partial x) - (y/a_1)\mathcal{F} = 0. \tag{11}$$

Of course, (10) and (11) follow from (9) only if $\mathbf{u}_2$ and $\mathbf{u}_1$ are linearly independent. In region $\mathcal{A}$ the vectors $\mathbf{u}_2$ and $\mathbf{u}_1$ fail to be linearly independent on the upper surface $y = z$ ($\mathbf{u}_2 = \mathbf{u}_1$) and on the side $y = x$ ($\mathbf{u}_2 = -\mathbf{u}_1$). It might seem, therefore, that Eqs. (10) and (11) would be subject to the additional conditions $y \neq z$ and $y \neq x$. However, the points on the surfaces $y = z$ and $y = x$ are either already excluded by our conditions $r_2 > r_1 > 0$, or they are regular points of the function $\mathcal{F}(x, y, z)$. In the latter case, Eqs. (10) and (11) continue to hold for $y = z$ and $y = x$, by continuity. Consequently, Eqs. (10) and (11) are valid throughout $\mathcal{A}$.

Equation (10) is equivalent to

$$\partial/\partial y(z^2 - y^2)\partial\mathcal{F}/\partial z = 0,$$

which implies that $(z^2 - y^2)\partial\mathcal{F}/\partial z$ is a function of x and z alone,

$$(z^2 - y^2)\partial\mathcal{F}/\partial z = \phi(x, z). \tag{12}$$

The function $\phi(x, z)$ occurring here can be evaluated by evaluating the left-hand side of the equation for a special value of y. For each pair of values of x and z which occurs in region $\mathcal{A}$, the point (x, y, z) with $y = z$ is a point in region $\mathcal{A}$ and, hence, a regular point of $\mathcal{F}(x, y, z)$. This choice of y makes the factor $(z^2 - y^2)$ in (12) vanish. Since $\partial\mathcal{F}/\partial z$ remains finite, the entire left-hand side of Eq. (12) vanishes. It follows that the function $\phi(x, z)$ vanishes throughout $\mathcal{A}$. Thus $(z^2 - y^2)\partial\mathcal{F}/\partial z = 0$ throughout $\mathcal{A}$. This in turn implies that

$$\partial\mathcal{F}/\partial z = 0 \tag{13}$$

and, hence, that $\mathcal{F}$ is a function of x and y alone. Extending this result to region $\mathcal{B}$ also, we can write

$$\mathcal{F}(x, y, z) = \theta(z)F_1(x, y) + \theta(-z)F_2(x, y),$$
$$\theta(z) = +1, \quad z > 0; \quad \theta(z) = 0, \quad z < 0. \tag{14}$$

Again assuming the only singularities of $\mathcal{F}(x, y, z)$ occur at the edges of the pyramid of Fig. 1, the right-hand side of Eq. (14) must be continuous across the xy plane, $z = 0$. This implies $F_1(x, y) = F_2(x, y) = F(x, y)$, say. From (14) it follows that $\mathcal{F}(x, y, z)$ is represented by the single expression $F(x, y)$ throughout its domain of definition,

$$\mathcal{F}(x, y, z) = F(x, y), \quad r_2 > 0, \quad r_1 > 0, \quad \mathbf{r}_2 \neq \mathbf{r}_1. \tag{15}$$

We have here (a slight generalization of) the result which we were looking for at the beginning: the coordinate space representative of any function of the Coulomb Hamiltonian can depend upon $\mathbf{r}_2$ and $\mathbf{r}_1$ only through the two variables $x = r_2 + r_1$ and $y = |\mathbf{r}_2 - \mathbf{r}_1|$—equivalently through the two variables $u = r_2 + r_1 + |\mathbf{r}_2 - \mathbf{r}_1|$ and $v = r_2 + r_1 - |\mathbf{r}_2 - \mathbf{r}_1|$. In particular, the coordinate space Coulomb Green's function

$$G(\mathbf{r}_2, \mathbf{r}_1, \omega) = -(\hbar^2/2m)\langle\mathbf{r}_2|\,(H - \hbar\omega)^{-1}\,|\mathbf{r}_1\rangle$$

must be a function of only the two variables u and v. However, in addition we learned that this function of only two variables must satisfy the identity (11), and we now want to explore the consequences of this identity.

Equation (11) [with $\mathcal{F}(x, y, z) = F(x, y)$] can be rewritten

$$\frac{\partial}{\partial y}\frac{x^2 - y^2}{2y}\frac{\partial}{\partial x}yF - a_1^{-1}yF = 0. \tag{16}$$

This equation can be integrated once by putting

$$yF(x, y) = \partial D(x, y)/\partial y. \tag{17}$$

We then find $\partial\Lambda/\partial y = 0$, where

$$\Lambda = [(\partial^2 D/\partial x\partial y)(x^2 - y^2)/2y - a_1^{-1}D].$$

This implies that Λ is a function of x only: $\Lambda = \lambda(x)$. Now for a given $F(x, y)$, the function $D(x, y)$ is not determined uniquely by Eq. (17). We can still add to $D(x, y)$ any function of x alone. In this way D can be adjusted such that $\Lambda = 0$,

$$(\partial^2 D/\partial x\partial y)(x^2 - y^2)/2y - a_1^{-1}D = 0. \tag{18}$$

The mixed derivatives in Eq. (18) can be eliminated by going over to the variables $u = x + y = r_1 + r_2 + |\mathbf{r}_2 - \mathbf{r}_1|$, $v = x - y = r_1 + r_2 - |\mathbf{r}_2 - \mathbf{r}_1|$. In these variables Eq. (18) becomes

$$\left(\frac{\partial^2}{\partial u^2} + \frac{1}{a_1 u}\right)D - \left(\frac{\partial^2}{\partial v^2} + \frac{1}{a_1 v}\right)D = 0, \tag{19}$$

and Eq. (17) goes over into

$$F = \frac{2}{u-v}\left(\frac{\partial D}{\partial u} - \frac{\partial D}{\partial v}\right). \tag{20}$$

These results "explain" the special significance of the variables u and v: For functions of u and v only, the identity (7) can be reduced to an equation [Eq. (19)] in which a separation of variables occurs.

III. DERIVATION OF COULOMB GREEN'S FUNCTION BY SEPARATION OF VARIABLES

We now want to apply these results to the specific case of the coordinate space Coulomb Green's function. In the domain of validity, $r_2 > 0$, $r_1 > 0$, $\mathbf{r}_2 \neq \mathbf{r}_1$, of the previous equations, the delta function source term in the differential equation, (2), vanishes, and the equation for the Green's function is homogeneous. Writing

$$G(\mathbf{r}_2, \mathbf{r}_1, \omega) = (1/y)[\partial D(x, y)/\partial y], \tag{21}$$

Eq. (2) goes over into

$$0 = \frac{\partial}{\partial y}\left(\frac{\partial^2 D}{\partial x^2} + \frac{\partial^2 D}{\partial y^2} + \frac{2x}{y}\frac{\partial^2 D}{\partial x \partial y} + k^2 D\right) - \frac{2}{r_2}\frac{\partial}{\partial y}\left(\frac{x^2 - y^2}{2y}\frac{\partial^2 D}{\partial x \partial y} - \frac{D}{a_1}\right). \tag{22}$$

As a consequence of the identity (18) this simplifies to

$$0 = \frac{\partial \Omega}{\partial y}, \quad \Omega = \left(\frac{\partial^2 D}{\partial x^2} + \frac{\partial^2 D}{\partial y^2} + \frac{4k\nu x}{x^2 - y^2} D + k^2 D\right),$$

$$a_1^{-1} = mZe^2/4\pi\hbar^2 = k\nu \text{ (as before)}. \tag{23}$$

From (23) we deduce that Ω is a function of x only: $\Omega = 4\psi(2x)$, for some function ψ. Changing over to the variables u and v, this statement becomes

$$O(u)D + O(v)D = 2\psi(u + v),$$
$$O(z) = \left(\frac{\partial^2}{\partial z^2} + \frac{k^2}{4} + \frac{k\nu}{z}\right). \tag{24}$$

The identity (19) can be rewritten in terms of the same differential operators $O(u)$ and $O(v)$ as occur in Eq. (24), and gives a relation,

$$O(u)D - O(v)D = 0, \tag{25}$$

involving these operators in an independent linear combination. Solving (24) and (25) for $O(u)D$ and $O(v)D$, find

$$O(u)D = \psi(u + v), \quad O(v)D = \psi(u + v). \tag{26}$$

It is now shown that the "integration constant" $\psi(u + v)$ vanishes. Operating on the first of Eqs. (26) with $O(v)$ and on the second with $O(u)$, we deduce that $O(v)\psi(u + v) = O(u)\psi(u + v)$. But for a function of $u + v$ only we have

$$\partial^2/\partial u^2 \psi(u + v) = \partial^2/\partial v^2 \psi(u + v).$$

Consequently, the relation

$$O(v)\psi(u + v) = O(u)\psi(u + v)$$

reduces to simply $k\nu(u - v)(uv)^{-1}\psi(u + v) = 0$, and this in turn implies $\psi(u + v) = 0$. Therefore Eqs. (26) become simply

$$O(u)D = 0, \quad O(v)D = 0. \tag{27}$$

We have here reduced the partial differential equation of the coordinate space Coulomb Green's function to a pair of uncoupled ordinary differential equations for the dependence of the generating function $D(u, v)$ on the two variables u and v.

The general solution of the equation $O(z)f(z) = 0$ can be expressed as a linear combination of the two Whittaker functions[12] $W_{i\nu;\frac{1}{2}}(-ikz)$ and $\mathcal{M}_{i\nu;\frac{1}{2}}(-ikz)$. Thus $D(u, v)$ may be expressed as a linear combination of (the four possible) products of one Whittaker function of argument $-iku$ times one of argument $-ikv$, all with the indices $i\nu$ and $\frac{1}{2}$. Exactly which products are allowed is dictated by the regularity conditions on $G(\mathbf{r}_2, \mathbf{r}_1, \omega)$ as a function of $\mathbf{r}_2$. At $r_2 = +\infty$, $u = +\infty$. Since the Green's function must vanish at $r_2 = +\infty$,[13] the products involving u through the $\mathcal{M}$ function are excluded.[14] As regards the variable v, this can be made to vanish in the region $r_2 > 0$, $r_1 > 0$, $\mathbf{r}_2 \neq \mathbf{r}_1$ of regularity of the Green's function by choosing $\mathbf{r}_2$ antiparallel to $\mathbf{r}_1$. But $G(\mathbf{r}_2, \mathbf{r}_1, \omega)$ would become infinite for this choice of $\mathbf{r}_2$ if products involving v through the W function were allowed [because $dW_{i\nu;\frac{1}{2}}(z)/dz$ becomes infinite for $z = 0$ (cf. Ref. 9, p. 28)]. We have now eliminated all but one possibility, namely,

$$D(u, v) \propto W_{i\nu;\frac{1}{2}}(-iku)\mathcal{M}_{i\nu;\frac{1}{2}}(-ikv).$$

Substituting into Eq. (21) [equivalently, Eq. (20)] we obtain

$$G(\mathbf{r}_2, \mathbf{r}_1, \omega) \propto \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} \times \det \begin{bmatrix} W_{i\nu;\frac{1}{2}}(-iku) & \mathcal{M}_{i\nu;\frac{1}{2}}(-ikv) \\ \dot{W}_{i\nu;\frac{1}{2}}(-iku) & \dot{\mathcal{M}}_{i\nu;\frac{1}{2}}(-ikv) \end{bmatrix}. \tag{28}$$

The homogeneous equations which we have been working do not, of course, determine the numerical factor which is missing here. To determine this factor one must go back to the inhomogeneous

[12] Reference 9, p. 25.
[13] Reference 6, Eqs. (1.2).
[14] Here the condition Im $(k) > 0$ [cf. Eq. (2)] comes in.

equation, (2), and match the singularity at $\mathbf{r}_2 = \mathbf{r}_1$ of the expression on the right-hand side of (28) with the amplitude of the delta function source term in (2). This is achieved by the requirement

$$G(\mathbf{r}_2, \mathbf{r}_1, \omega) \sim -1/4\pi\,|\mathbf{r}_2 - \mathbf{r}_1|, \quad (\mathbf{r}_2 \to \mathbf{r}_1). \tag{29}$$

Now when $\mathbf{r}_2$ approaches $\mathbf{r}_1$, u approaches v, and the determinant in (28) goes over into the Wronskian of the two functions W and $\mathcal{M}$. This Wronskian has the value[12] $1/\Gamma(1 - iv)$. Hence the missing numerical factor in (28) is $-\Gamma(1 - iv)/4\pi$. Supplying (28) with this factor gives the familiar closed form expression (4) for the coordinate space Coulomb Green's function.

ACKNOWLEDGMENTS

The author would like to thank Professor L. Brown for suggesting this problem and for a number of helpful discussions.

This research was supported in part by the United States Office of Naval Research and the United States Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Time-Dependent Green's Function for a Moving Isotropic Nondispersive Medium*

C. T. TAI
Radiation Laboratory, University of Michigan, Ann Arbor, Michigan

(Received 10 June 1966)

The time-dependent Green's function for a moving isotropic nondispersive medium is hereby obtained by taking the ω-integration of the time-harmonic solution which was previously obtained by means of an operational method and by making use of the known result of the two-dimensional Klein–Gordon equation.

INTRODUCTION

THE time-dependent Green's function for a moving isotropic medium was recently found by Compton.[1] He applied a four-fold Fourier transform to the pertinent differential equation to obtain the desired result. In evaluating the reciprocal Fourier transform, he considers the ω-integration first, followed by the k-integration. The steps involved in the calculation are rather long, comparable to those of Lee and Papas.[2]

In this paper, we show that the time-dependent solution can readily be obtained by taking the ω-integration of the time-harmonic solution[3] which was previously obtained by an operational method without a lengthy calculation.

THE BASIC EQUATION AND ITS SOLUTIONS

The time-dependent Green's function considered by Compton[4] satisfies the differential equation

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{1}{a}\frac{\partial^2}{\partial z^2} - \frac{2\Omega}{a}\frac{\partial^2}{\partial z \partial t} + \left(\frac{\Omega^2}{a} - \frac{n^2 a}{c^2}\right)\frac{\partial^2}{\partial t^2}\right] G(\bar{R}, \bar{R}'; t, t') = -\delta(\bar{R} - \bar{R}')\delta(t - t'), \tag{1}$$

where

$$a = \frac{1 - \beta^2}{1 - n^2\beta^2}, \quad \Omega = \frac{(n^2 - 1)\beta}{(1 - n^2\beta^2)c},$$

$$n = \left(\frac{\mu\epsilon}{\mu_0\epsilon_0}\right)^{\frac{1}{2}}, \quad \beta = \frac{v}{c}, \quad c = (\mu_0\epsilon_0)^{-\frac{1}{2}}, \quad \bar{v} = va_z.$$

We define the Fourier pair

$$F(\omega) = \int_{-\infty}^{\infty} G(t)e^{-j\omega t}\,dt, \tag{2}$$

$$G(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}\,d\omega. \tag{3}$$

* The research reported here was sponsored by the National Aeronautics and Space Administration under Grant NGR-23-005-107 with the Langley Research Center, Hampton, Virginia.

[1] R. T. Compton, Jr., J. Math. Phys. **7**, 2145 (1966).

[2] K. S. H. Lee and C. H. Papas, J. Math. Phys. **5**, 1668 (1964).

[3] C. T. Tai, Trans. IEEE Antennas Propagation, **AP13**, 322 (1965).

[4] Compton defines the Green's function with a positive sign attached to the delta function.

equation, (2), and match the singularity at $\mathbf{r}_2 = \mathbf{r}_1$ of the expression on the right-hand side of (28) with the amplitude of the delta function source term in (2). This is achieved by the requirement

$$G(\mathbf{r}_2, \mathbf{r}_1, \omega) \sim -1/4\pi\,|\mathbf{r}_2 - \mathbf{r}_1|, \quad (\mathbf{r}_2 \to \mathbf{r}_1). \tag{29}$$

Now when $\mathbf{r}_2$ approaches $\mathbf{r}_1$, u approaches v, and the determinant in (28) goes over into the Wronskian of the two functions W and $\mathcal{M}$. This Wronskian has the value[12] $1/\Gamma(1 - iv)$. Hence the missing numerical factor in (28) is $-\Gamma(1 - iv)/4\pi$. Supplying (28) with this factor gives the familiar closed form expression (4) for the coordinate space Coulomb Green's function.

ACKNOWLEDGMENTS

The author would like to thank Professor L. Brown for suggesting this problem and for a number of helpful discussions.

This research was supported in part by the United States Office of Naval Research and the United States Atomic Energy Commission.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Time-Dependent Green's Function for a Moving Isotropic Nondispersive Medium*

C. T. TAI
Radiation Laboratory, University of Michigan, Ann Arbor, Michigan

(Received 10 June 1966)

The time-dependent Green's function for a moving isotropic nondispersive medium is hereby obtained by taking the ω-integration of the time-harmonic solution which was previously obtained by means of an operational method and by making use of the known result of the two-dimensional Klein–Gordon equation.

INTRODUCTION

THE time-dependent Green's function for a moving isotropic medium was recently found by Compton.[1] He applied a four-fold Fourier transform to the pertinent differential equation to obtain the desired result. In evaluating the reciprocal Fourier transform, he considers the ω-integration first, followed by the k-integration. The steps involved in the calculation are rather long, comparable to those of Lee and Papas.[2]

In this paper, we show that the time-dependent solution can readily be obtained by taking the ω-integration of the time-harmonic solution[3] which was previously obtained by an operational method without a lengthy calculation.

THE BASIC EQUATION AND ITS SOLUTIONS

The time-dependent Green's function considered by Compton[4] satisfies the differential equation

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{1}{a}\frac{\partial^2}{\partial z^2} - \frac{2\Omega}{a}\frac{\partial^2}{\partial z \partial t} + \left(\frac{\Omega^2}{a} - \frac{n^2 a}{c^2}\right)\frac{\partial^2}{\partial t^2}\right] G(\bar{R}, \bar{R}'; t, t') = -\delta(\bar{R} - \bar{R}')\delta(t - t'), \tag{1}$$

where

$$a = \frac{1 - \beta^2}{1 - n^2\beta^2}, \quad \Omega = \frac{(n^2 - 1)\beta}{(1 - n^2\beta^2)c},$$

$$n = \left(\frac{\mu\epsilon}{\mu_0\epsilon_0}\right)^{\frac{1}{2}}, \quad \beta = \frac{v}{c}, \quad c = (\mu_0\epsilon_0)^{-\frac{1}{2}}, \quad \bar{v} = va_z.$$

We define the Fourier pair

$$F(\omega) = \int_{-\infty}^{\infty} G(t)e^{-j\omega t}\,dt, \tag{2}$$

$$G(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}\,d\omega. \tag{3}$$

* The research reported here was sponsored by the National Aeronautics and Space Administration under Grant NGR-23-005-107 with the Langley Research Center, Hampton, Virginia.

[1] R. T. Compton, Jr., J. Math. Phys. **7**, 2145 (1966).

[2] K. S. H. Lee and C. H. Papas, J. Math. Phys. **5**, 1668 (1964).

[3] C. T. Tai, Trans. IEEE Antennas Propagation, **AP13**, 322 (1965).

[4] Compton defines the Green's function with a positive sign attached to the delta function.

Thus, if Eq. (1) is multiplied by $e^{-j\omega t}$ and integrated with respect to t, we obtain

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{1}{a}\frac{\partial^2}{\partial z^2}-\frac{2j\omega\Omega}{a}\frac{\partial}{\partial z}+\omega^2\left(\frac{n^2a}{c^2}-\frac{\Omega^2}{a}\right)\right]F(\omega)=-e^{-j\omega t'}\delta(\bar{R}-\bar{R}'). \tag{4}$$

If we introduce an auxiliary function $f(\omega)$ such that

$$F(\omega)=e^{j\omega\Omega z}f(\omega), \tag{5}$$

then $f(\omega)$ satisfies the following equation:

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{1}{a}\frac{\partial^2}{\partial z^2}+\frac{\omega^2n^2a}{c^2}\right)f(\omega)=-e^{-j\omega(t'+\Omega z')}\delta(\bar{R}-\bar{R}'). \tag{6}$$

Except for the factor $e^{-j\omega(t'+\Omega z')}$, Eq. (6) is the same as Eq. (26) considered in Ref. 3. Hence, its solution is given by the following:

Case I: $n\beta<1$

$$f(\omega)=\frac{a^{\frac{1}{2}}\exp\{-j\omega[t'+\Omega z'+(n/c)a^{\frac{1}{2}}R_a]\}}{4\pi R_a}, \tag{7}$$

where

$$R_a=(a\xi^2+r^2)^{\frac{1}{2}},\quad r^2=(x-x')^2+(y-y')^2,\quad \xi=(z-z').$$

Case II: $n\beta>1$

$$f(\omega)=\left\{\begin{matrix}0, & |a|^{\frac{1}{2}}\xi<r\\ \dfrac{|a|^{\frac{1}{2}}\exp[-j\omega(t'+\Omega z')]\cos(\omega n/c)\,|a|^{\frac{1}{2}}R_a'}{2\pi R_a'}, & |a|^{\frac{1}{2}}\xi>r\end{matrix}\right\}, \tag{8}$$

where

$$R_a'=(|a|\,\xi^2-r^2)^{\frac{1}{2}}.$$

In view of Eqs. (3) and (5), we can obtain readily the solutions for $G(t)$; they are the following:

Case I: $n\beta<1$

$$G(t)=\frac{a^{\frac{1}{2}}}{8\pi^2R_a}\int_{-\infty}^{\infty}\exp\left[j\omega\left(\tau+\Omega\xi-\frac{n}{c}a^{\frac{1}{2}}R_a\right)\right]d\omega=\frac{a^{\frac{1}{2}}}{4\pi R_a}\delta\left(\tau+\Omega\xi-\frac{n}{c}a^{\frac{1}{2}}R_a\right), \tag{9}$$

where $\tau=t-t'$.

Case II: $n\beta>1$

$$G(t)=\frac{|a|^{\frac{1}{2}}}{4\pi^2R_a'}\int_{-\infty}^{\infty}\exp\left[j\omega(\tau+\Omega\xi)\right]\cos\frac{n}{c}|a|^{\frac{1}{2}}R_a'\,d\omega=\frac{|a|^{\frac{1}{2}}}{4\pi R_a'}\delta\left(\tau+\Omega\xi-\frac{n}{c}|a|^{\frac{1}{2}}R_a'\right),\quad |a|^{\frac{1}{2}}\xi>r, \tag{10}$$

$$G(t)=0,\quad |a|^{\frac{1}{2}}\xi<r. \tag{11}$$

Our expressions for $G(t)$ appear to be of slightly different form as compared to Compton's, but they are equivalent. In fact, the present ones are simpler in form and also put the time-dependent part explicitly in the delta function.

To discuss the locus of the wave front, we consider, for example, the case corresponding to $n\beta<1$. The impulsive wave front is described by

$$\tau+\Omega\xi-(n/c)a^{\frac{1}{2}}R_a=0. \tag{12}$$

The above equation can be written in the form

$$[(\xi-\xi_c)^2/A^2]+(r^2/B^2)=1, \tag{13}$$

where

$$\xi_c=\frac{(n^2-1)\beta c\tau}{n^2-\beta^2},\quad A=\frac{n(1-\beta^2)c\tau}{n^2-\beta^2},\quad B=\left(\frac{1-\beta^2}{n^2-\beta^2}\right)^{\frac{1}{2}}c\tau.$$

Equation (13) defines the same ellipsoid discussed by Compton. It can be shown that the same algebraic equation applies to the case $n\beta>1$. For the latter case, ξ_c is numerically smaller than A. The detached ellipsoidal wave front is therefore confined within the Mach cone defined by $|a|^{\frac{1}{2}}\xi-r=0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Relation between the Relativistic and Nonrelativistic S-Matrices

FRANCIS R. HALPERN
University of California, San Diego, La Jolla, California

(Received 24 January 1966)

It is shown that, given any Lorentz invariant S-matrix, it is possible to calculate from it, by purely algebraic operations, a Galilean invariant S-matrix, and vice versa.

I. INTRODUCTION

THE requirement of Lorentz invariance is more difficult to satisfy than the requirement of Galilean invariance in quantum mechanics. A formula is given below which shows that, if there is a theory that has a Lorentz invariant S-matrix, then there is a related theory that has a Galilean invariant S-matrix, and that the two S-matrices are related by only algebraic relations. If other conditions, such as crossing symmetry are required of the relativistic S-matrix, they may be imposed on the nonrelativistic one and the entire covariant problem reduced to the simpler noncovariant one. In Sec. II the formula and related definitions are given. In Sec. III the proof of the invariance and unitarity is given, and in Sec. IV a motivated derivation of the formula is given.

II. FORMULA RELATING THE RELATIVISTIC AND NONRELATIVISTIC S-MATRICES

The S-matrices are taken between plane-wave states $|p\sigma\rangle$. The relativistic S-operator is indicated by S and the nonrelativistic one by s. The general element of the S-matrix is

$$\langle \mathbf{p}_1\sigma_1 \cdots \mathbf{p}_n\sigma_n| \, S \, |\mathbf{q}_1\tau_1 \cdots \mathbf{q}_m\tau_m\rangle. \tag{1}$$

The invariance of S is contained in the statement $U^\dagger(L)SU(L) = S$, where $U(L)$ is the unitary operator that represents the Lorentz transformation L. The action of $U(L)$ on the state $|\mathbf{q}\tau\rangle$ is given by

$$U(L)\,|\mathbf{q}\tau\rangle = (\Omega_{L\mathbf{q}}/\Omega_{\mathbf{q}})^{\frac{1}{2}} D^{(T)}_{\tau'\tau}[R(L, q)]\,|L\mathbf{q}\tau'\rangle. \tag{2}$$

The vector $\mathbf{q}$ is the spatial part of a four vector whose time component $\Omega_{\mathbf{q}} = (\mathbf{q}^2 + m^2)^{\frac{1}{2}}$, where m is the mass of the particle. In the above formula $L\mathbf{q}$ means the spatial part of the transformed four vector $L(\mathbf{q}, \Omega_q)$. The rotation $R(L, q)$ is the Wigner rotation associated with the Lorentz transformation L and the four vector[1] q. The unitary matrix D is an irreducible representation of the rotation group of dimension $2T + 1$ when the particle in question has spin T. The T's and masses are usually dropped.

[1] E. P. Wigner, Ann. Math. **40**, 149 (1939).

The general element of the nonrelativistic s-matrix is

$$\langle \mathbf{p}_1\sigma_1 \cdots \mathbf{p}_n\sigma_n| \, s \, |\mathbf{q}_1\tau_1 \cdots \mathbf{q}_m\tau_m\rangle. \tag{3}$$

The nonrelativistic operator s has the invariance property $U^\dagger(G)s\,U(G) = s$, where $U(G)$ is the unitary operator that represents the Galilean transformation G. The action of $U(G)$ on the $|\mathbf{q}\tau\rangle$ is given by

$$U(R)\,|\mathbf{q}\tau\rangle = D_{\tau'\tau}(R)\,|R\mathbf{q}, \tau\rangle, \quad \text{for rotation } R; \tag{4a}$$

$$U(V)\,|\mathbf{q}\tau\rangle = |\mathbf{q} + m\mathbf{V}, \tau\rangle,$$

for a change to a moving coordinate system. (4b)

In terms of the S-matrix elements the requirements of Lorentz and rotational invariance become

$$\begin{aligned}\langle \mathbf{p}_1\sigma_1 \cdots| \, S \, |\mathbf{q}_1\tau_1 \cdots\rangle &= \sum_{\sigma_1' \cdots \tau_1' \cdots} D^{-1}_{\sigma_1\sigma_1'}[R(L, p_1)] \cdots \\ &\times (\Omega_{L\mathbf{p}_1} \cdots / \Omega_{\mathbf{p}_1} \cdots)^{\frac{1}{2}} \\ &\times \langle L\mathbf{p}_1\sigma_1' \cdots| \, S \, |L\mathbf{q}_1\tau_1' \cdots\rangle (\Omega_{L\mathbf{q}_1} \cdots / \Omega_{\mathbf{q}_1} \cdots)^{\frac{1}{2}} \\ &\times D_{\tau_1'\tau_1}[R(L, q_1)] \cdots .\end{aligned} \tag{5}$$

The dots indicate that the same factors are repeated for each of the m-incident and n-final particles as for the first. The analogous formula for the nonrelativistic s-matrix is

$$\begin{aligned}\langle \mathbf{p}_1\sigma_1 \cdots| \, s \, |\mathbf{q}_1\tau_1 \cdots\rangle = \sum_{\sigma_1' \cdots \tau_1' \cdots} & D^{-}_{\sigma_1\sigma_1'}(R) \cdots \\ &\times \langle R\mathbf{p}_1\sigma_1' \cdots| \, s \, |R\mathbf{q}_1\tau_1' \cdots\rangle \\ &\times D_{\tau_1'\tau_1}(R) \cdots .\end{aligned} \tag{6}$$

Given the set of initial four vectors $q_1 \cdots q_m$ for the initial state, there is a Lorentz transformation A that carries these four vectors to their center-of-mass values $\hat{q}_1 \cdots \hat{q}_m$ that is $Aq_i = \hat{q}_i$. The $\hat{q}$'s have the property $\sum \hat{\mathbf{q}}_i = \mathbf{0}$. The transformation A is to be a Lorentz transformation without rotation. That is, if $\mathbf{Q} = \sum \mathbf{q}_i$, then each $\mathbf{q}_i$ can be separated into its component parallel to $\mathbf{Q}$: $(\mathbf{q}_i \cdot \mathbf{Q})\mathbf{Q}/\mathbf{Q}^2$ and its component perpendicular to $\mathbf{Q}$: $\mathbf{q}_i - (\mathbf{q}_i \cdot \mathbf{Q})\mathbf{Q}/\mathbf{Q}^2$. Only the component of $\mathbf{q}_i$ parallel to $\mathbf{Q}$ is changed by A. Given the set of three vectors $\hat{\mathbf{q}}_1 \cdots \hat{\mathbf{q}}_m$, a Galilean transformation can be made on them so that they have the same total momentum $\mathbf{Q}$ as the original set of

vectors $\mathbf{q}_1 \cdots \mathbf{q}_m$. The Galilean transformation G applied to $\hat{\mathbf{q}}_i$ gives $G\hat{\mathbf{q}}_i = \bar{\mathbf{q}}_i = \hat{\mathbf{q}}_i + m_i\mathbf{Q}/M$, where m_i is the mass of the ith particle and $M = \sum m_i$. The components of $\mathbf{q}_i$ and $\bar{\mathbf{q}}_i$ perpendicular to $\mathbf{Q}$ are equal, and only the component parallel to $\mathbf{Q}$ is altered. The Jacobian of the transformation from $\mathbf{q}$ to $\bar{\mathbf{q}}$ is

$$d^3\mathbf{q}_1 \cdots d^3\mathbf{q}_m = (\Omega_1 \cdots \Omega_m/\omega_1 \cdots \omega_m)\, d^3\bar{\mathbf{q}}_1 \cdots d^3\bar{\mathbf{q}}_m. \tag{7}$$

The Ω's are the energies in the moving coordinate system, and the ω's are the energies in the center of mass. The Jacobian is easily calculated from the familiar facts that $d^3\mathbf{q}/\Omega$ is an invariant for Lorentz transformations and $d^3\mathbf{q}$ is an invariant for Galilean transformations. A similar set of vectors $\bar{\mathbf{p}}_1 \cdots \bar{\mathbf{p}}_m$ may be defined for the final state.[2]

The proposed connection between the relativistic and nonrelativistic S-matrices is

$$\begin{aligned}\langle \mathbf{p}_1\sigma_1 \cdots \mathbf{p}_n\sigma_n \mid S \mid \mathbf{q}_1\tau_1 \cdots \mathbf{q}_m\tau_m\rangle &= \sum_{\substack{\sigma_1' \cdots \sigma_n' \cdots \\ \tau_1' \cdots \tau_m' \cdots}} D^{-1}_{\sigma_1\sigma_1'}[R(A, p_1)] \cdots D^{-1}_{\sigma_n'\sigma_n}[R(A, p_n)] \\ &\times (\omega_{\mathbf{p}_1} \cdots \omega_{\mathbf{p}_n}/\Omega_{\mathbf{p}_1} \cdots \Omega_{\mathbf{p}_n})^{\frac{1}{2}} \\ &\times \langle \bar{\mathbf{p}}_1\sigma_1' \cdots \bar{\mathbf{p}}_n\sigma_n' \mid s \mid \bar{\mathbf{q}}_1\tau_1' \cdots \bar{\mathbf{q}}_m\tau_m'\rangle \\ &\times (\omega_{\mathbf{q}_1} \cdots \omega_{\mathbf{q}_m}/\Omega_{\mathbf{q}_1} \cdots \Omega_{\mathbf{q}_m})^{\frac{1}{2}} \\ &\times D_{\tau_1'\tau_1}[R(A, q_1)] \cdots D_{\tau_m'\tau_m}[R(A, q_m)]. \end{aligned} \tag{8}$$

III. UNITARITY AND LORENTZ INVARIANCE

By direct calculation the unitarity and Lorentz invariance of S are shown to follow from the unitarity and Galilean invariance of s. The unitarity of S is tested by evaluation the expression:

$$\begin{aligned}\sum_{\tau_1 \cdots \tau_m} \int d^3\mathbf{q}_1 \cdots d^3\mathbf{q}_m &\times \langle \mathbf{p}_1\sigma_1 \cdots \mathbf{p}_n\sigma_n \mid S \mid \mathbf{q}_1\tau_1 \cdots \mathbf{q}_m\tau_m\rangle \\ &\times \langle \mathbf{q}_1\tau_1 \cdots \mathbf{q}_m\tau_m \mid S^\dagger \mid \mathbf{r}_1\rho_1 \cdots \mathbf{r}_l\rho_l\rangle. \end{aligned} \tag{9}$$

The sums should be over all intermediate states. The outside sum in (9) indicates the sum over all possible particle types and number of particles. The inner sums and the integrals are the instructions for adding over all states with a given number and type of particle. If (8) is substituted for S in (9), the sum over τ_i involves $D_{\tau_1'\tau_1}[R(A, p_i)]$ from S and $D^*_{\tau_1''\tau_1}[R(Ap_i)]$ from $S^\dagger$. Since D is unitary and the arguments of the two D's are the same, the result is a $\delta_{\tau_1'\tau_1''}$. The factors $(\omega_{\mathbf{q}_1} \cdots \omega_{\mathbf{q}_m}/\Omega_{\mathbf{q}_1} \cdots \Omega_{\mathbf{q}_m})^{\frac{1}{2}}$ are the same for both S and $S^\dagger$, and

$$d^3\mathbf{q}_1 \cdots d^3\mathbf{q}_m(\omega_{\mathbf{q}_1} \cdots \omega_{\mathbf{q}_m}/\Omega_{\mathbf{q}_1} \cdots \Omega_{\mathbf{q}_m}) = d^3\bar{q}_1 \cdots d^3\bar{q}_m.$$

[2] A. Chakrabarti [J. Math. Phys. 5, 922 (1964)] has discussed similar coordinates.

With these remarks the unitarity sum (9) becomes

$$\begin{aligned}&\sum_{\substack{\sigma_1' \cdots \sigma_n' \\ \rho_1' \cdots \rho_l'}} D^{-1}_{\sigma_1\sigma_1'}[R(A, p_1)] \cdots D^{-1}_{\sigma_n\sigma_n'}[R(A, p_n)] \\ &\quad\times (\omega_{\mathbf{p}_1} \cdots \omega_{\mathbf{p}_n}/\Omega_{\mathbf{p}_1} \cdots \Omega_{\mathbf{p}_n})^{\frac{1}{2}} \\ &\quad\times \Big\{ \sum_{\substack{\tau_1' \cdots \tau_m' \\ \tau_1'' \cdots \tau_m''}} \int d^3\bar{\mathbf{q}}_1 \cdots d^3\bar{\mathbf{q}}_m \\ &\quad\times \langle \bar{\mathbf{p}}_1\sigma_1' \cdots \bar{\mathbf{p}}_n\sigma_n' \mid s \mid \bar{\mathbf{q}}_1\tau_1' \cdots \mathbf{q}_m'\tau_m'\rangle \delta_{\tau_1'\tau_1''} \cdots \delta_{\tau_m'\tau_m''} \\ &\quad\times \langle \mathbf{q}_1\tau_1'' \cdots \bar{\mathbf{q}}_m \mid s^\dagger \mid \bar{\mathbf{r}}_1\rho_1' \cdots \bar{\mathbf{r}}_l\rho_l'\rangle \Big\} \\ &\quad\times (\omega_{\mathbf{r}_1} \cdots \omega_{\mathbf{r}_l}/\Omega_{\mathbf{r}_1} \cdots \Omega_{\mathbf{r}_l})^{\frac{1}{2}} \\ &\quad\times D_{\rho_1'\rho_1}[R(A, r_1)] \cdots D_{\rho_l'\rho_l}[R(A, r_l)]. \end{aligned}$$

The expression in braces is just the nonrelativistic unitarity condition so that the unitarity sum (9) becomes

$$\begin{aligned}&\sum_{\substack{\sigma_1' \cdots \sigma_n' \\ \rho_1' \cdots \rho_n'}} D^{-1}_{\sigma_1\sigma_1'}[R(A, p_1)] \cdots D^{-1}_{\sigma_n\sigma_n'}[R(A, p_n)] \\ &\quad\times (\omega_{\mathbf{p}_1} \cdots \omega_{\mathbf{p}_n}/\Omega_{\mathbf{p}_1} \cdots \Omega_{\mathbf{p}_n})^{\frac{1}{2}} \\ &\quad\times \langle \bar{\mathbf{p}}_1\sigma_1' \cdots \bar{\mathbf{p}}_n\sigma_n' \mid 1 \mid \bar{\mathbf{r}}_1\rho_1' \cdots \bar{\mathbf{r}}_l\rho_l'\rangle \\ &\quad\times (\omega_{\mathbf{r}_1} \cdots \omega_{\mathbf{r}_l}/\Omega_{\mathbf{r}_1} \cdots \Omega_{\mathbf{r}_l})^{\frac{1}{2}} \\ &\quad\times D_{\rho_1'\rho_1}[R(A, r_1)] \cdots D_{\rho_l'\rho_l}[R(A, r_l)]. \end{aligned}$$

Unless the initial and final states have the same number and type of particles, the sum vanishes. When the initial and final states have the same number and type of particles, the spin indices σ' and ρ' must be equal and the D functions are summed to $\delta_{\sigma\rho}$. The momentum δ functions of $(\bar{\mathbf{p}} - \bar{\mathbf{q}})$ multiplied by the weight (ω/Ω) gives $\delta(\mathbf{p} - \mathbf{q})$, so the expression is just

$$\langle \mathbf{p}_1\sigma_1 \cdots \mathbf{p}_n\sigma_n \mid 1 \mid \mathbf{r}_1\rho_1 \cdots \mathbf{r}_l\rho_l\rangle,$$

and the unitarity of S follows from that of s.

Lorentz invariance of S requires that

$$\begin{aligned}\langle \mathbf{p}_1\sigma_1 \cdots \mathbf{p}_n\sigma_n \mid S \mid \mathbf{q}_1\tau_1 \cdots \mathbf{q}_m\tau_m\rangle &= \sum D^{-1}_{\sigma_1\sigma_1'}[R(L, p_1)] \cdots D^{-1}_{\sigma_n\sigma_n'}[R(L, p_n)] \\ &\times (\Omega_{L\mathbf{p}_1} \cdots \Omega_{L\mathbf{p}_n}/\Omega_{\mathbf{p}_1} \cdots \Omega_{\mathbf{p}_n})^{\frac{1}{2}} \\ &\times \langle L\mathbf{p}_1\sigma_1' \cdots L\mathbf{p}_n\sigma_n' \mid S \mid L\mathbf{q}_1\tau_1' \cdots L\mathbf{q}_m\tau_m'\rangle \\ &\times (\Omega_{L\mathbf{q}_1} \cdots \Omega_{L\mathbf{q}_m}/\Omega_{\mathbf{q}_1} \cdots \Omega_{\mathbf{q}_m})^{\frac{1}{2}} \\ &\times D_{\tau_1'\tau_1}[R(L, q_1)] \cdots D_{\tau_m'\tau_m}[R(L, q_m)]. \end{aligned} \tag{10}$$

The S-matrix elements on both sides can be evaluated in terms of the elements of s through (8). The Galilean invariance of s guarantees the Lorentz of S. The left-hand side of (10) is explicitly given by (8). The S-matrix element on the right-hand side of (10) can also be calculated from (8). The transformation that takes the vectors $Lq_1 \cdots Lq_m$ to their rest system is B. If the sum of the q's is Q, then A transforms Q to its rest system $AQ = (\mathbf{O}, M)$. Similarly, the transformation B transforms LQ to its rest system

$BLQ = (\mathbf{O}, M)$. These two results may be combined to yield $BLA^{-1}(\mathbf{O}, M) = (\mathbf{O}, M)$. The product $\mathbf{BLA}^{-1}$ is a rotation $R(L, Q)$, the Wigner rotation associated with L and Q, since it leaves $(\mathbf{O}, M)$ invariant. The set of vectors $\widehat{\mathbf{Lq}}_1 \cdots \widehat{\mathbf{Lq}}_m$ is given by

$$\begin{aligned}\mathbf{BLq}_1 \cdots \mathbf{BLq}_m &= R(L, Q)Aq_1 \cdots R(L, Q)Aq_m \\ &= R(L, Q)\hat{\mathbf{q}}_1 \cdots R(L, Q)\hat{\mathbf{q}}_m .\end{aligned}$$

The Galilean transformation to a moving coordinate system from the center of mass does not change the value of s. The matrix element

$$\langle L\mathbf{p}_1\sigma_1' \cdots L\mathbf{p}_n\sigma_n' | \, S \, | L\mathbf{q}_1\tau_1' \cdots L\mathbf{q}_m\tau_m' \rangle$$

that appears on the right-hand side of (10) is

$$\begin{aligned}&\sum_{\substack{\sigma_1'' \cdots \sigma_n'' \\ \tau_1'' \cdots \tau_m''}} D^{-1}_{\sigma_1'\sigma_1''}[R(B, Lp_1)] \cdots D^{-1}_{\sigma_n'\sigma_n''}[R(B, Lp_n)] \\ &\quad \times (\omega_{\mathbf{p}_1} \cdots \omega_{\mathbf{p}_n}/\Omega_{L\mathbf{p}_1} \cdots \Omega_{L\mathbf{p}_n})^{\frac{1}{2}} \\ &\quad \times \langle R(L, Q)\bar{\mathbf{p}}_1\sigma_1'' \cdots R(L, Q)\bar{\mathbf{p}}_n\sigma_n'' | \, s \, | R(L, Q)\bar{\mathbf{q}}_1\tau_1'' \cdots \\ &\quad \times R(L, Q)\bar{\mathbf{q}}_m\tau_m'' \rangle (\omega_{q_1} \cdots \omega_{q_m}/\Omega_{Lq_1} \cdots \Omega_{Lq_m})^{\frac{1}{2}} \\ &\quad \times D_{\tau_1''\tau_1'}[R(B, Lq_1)] \cdots D_{\tau_m''\tau_m'}[R(B, Lq_m)]. \qquad (11)\end{aligned}$$

To simplify this, it is necessary to use the inverse of (4) which gives

$$\begin{aligned}&\langle R\bar{\mathbf{p}}_1\sigma_1'' \cdots R\bar{\mathbf{p}}_n\sigma_n'' | \, s \, | R\bar{\mathbf{q}}_1\tau_1'' \cdots R\bar{\mathbf{q}}_m\tau_m'' \rangle \\ &\quad = \sum_{\substack{\sigma_1''' \cdots \sigma_n''' \\ \tau_1''' \cdots \tau_m'''}} D_{\sigma_1''\sigma_1'''}[R(L, Q)] \cdots D_{\sigma_n''\sigma_1'''}[R(L, Q)] \\ &\qquad \times \langle \bar{\mathbf{p}}_1\sigma_1''' \cdots \bar{\mathbf{p}}_n\sigma_n''' | \, s \, | \bar{\mathbf{q}}_1\tau_1''' \cdots \bar{\mathbf{q}}_m\tau_m''' \rangle D^{-1}_{\tau_1'''\tau_1''} \\ &\qquad \times [R(L, Q)] \cdots D^{-1}_{\tau_m'''\tau_m''}[R(L, Q)]. \qquad (12)\end{aligned}$$

When (12) is substituted into (11) and (11) is then substituted into (10), the result is products of the type

$$D^{-1}_{\sigma_1\sigma_1'}[R(L, p_1)]D^{-1}_{\sigma_1'\sigma_1''}[R(B, Lp_1)]D_{\sigma_1''\sigma_1''}[R(L, Q)] \qquad (13a)$$

on the left of the matrix element of s, and of the type

$$D^{-1}_{\tau_1'''\tau_1''}[R(L, Q)]D_{\tau_1''\tau_1'}[R(B, Lq_1)]D_{\tau_1'\tau_1}[R(L, q_1)] \qquad (13b)$$

on the right of the matrix element of s. These expressions can be simplified by recalling that $BLA^{-1} = R(L, Q)$, that $R(L_1L_2, v) = R(L_1L_2v)R(L_2, v)$, and that $R(R, v) = R$. The product in (13b) is $R(A, q_1) = R[R^{-1}(L, Q)BL, q_1]$, since

$$\begin{aligned}&D[R\{R^{-1}(L, Q)BL, q_1\}] \\ &\qquad = D[R\{R^{-1}(L, Q), BLq_1\}R(B, Lq_1)R(L, q_1)].\end{aligned}$$

The right-hand side of (10) becomes the same as the right-hand side of (8), demonstrating that the Galilean invariance of s implies the Lorentz invariance of S.

IV. DERIVATION OF THE FORMULA

The formula (8) can be derived in a straightforward way for the two-body to two-body S-matrix, and the generalization can be guessed.[3] The combination of the plane-wave states into irreducible unitary representations of the Lorentz group is achieved by the formula

$$\begin{aligned}|pJ\mu LS\mathbf{p}\rangle = \sum_{\sigma_1\sigma_2} \int d^3\mathbf{p}_1 \, d^3\mathbf{p}_2 \, |\mathbf{p}_1\sigma_1\mathbf{p}_2\sigma_2\rangle \\ \times \langle \mathbf{p}_1\sigma_1\mathbf{p}_2\sigma_2 \, | \, pJ\mu LS\mathbf{p}\rangle. \qquad (14)\end{aligned}$$

This formula is valid relativistically and nonrelativistically. The difference is in the transformation coefficients $\langle p_1\sigma_1p_2\rho_2 \, | \, pJ\mu LS\mathbf{p}\rangle$ which are

$$\begin{aligned}&\langle \mathbf{p}_1\sigma_1\mathbf{p}_2\sigma_2 \, | \, pJ\mu LS\mathbf{p}\rangle \\ &\quad = \sum_{\lambda\sigma} \delta(\mathbf{p} - \mathbf{p}_1 - \mathbf{p}_2)\delta[p - p(p_1p_2)]/p^2 \\ &\qquad \times C(LSJ \, | \, \lambda\sigma\mu)C(S_1S_2S \, | \, \sigma_1\sigma_2\sigma)\gamma_{L\lambda}(\Omega) \qquad (15a)\end{aligned}$$

nonrelativistically, and

$$\begin{aligned}&\langle \mathbf{p}_1\sigma_1\mathbf{p}_2\sigma_2 \, | \, pJ\mu LS\mathbf{p}\rangle \\ &\quad = \sum_{\lambda\sigma} \delta(\mathbf{p} - \mathbf{p}_1 - \mathbf{p}_2)\delta[p - p(p_1p_2)]/p^2 \\ &\qquad \times C(LSJ \, | \, \lambda\sigma\mu)C(S_1S_2S \, | \, \tau_1\tau_2\sigma) \\ &\qquad \times D_{\sigma_1r_1}[R(A, p_1)]D_{\sigma_2r_2}[(A, p_2)](\Omega\omega_1\omega_2/M\Omega_1\Omega_2)^{\frac{1}{2}} \\ &\qquad (15b)\end{aligned}$$

relativistically. The only difference in the formulas is the appearance of the D's and the factor $(\Omega\omega_1\omega_2/M\Omega_1\Omega_2)^{\frac{1}{2}}$ in the relativistic case. The weighting factor is different from that of (8), since the coordinates $\mathbf{p}$ and $\mathbf{P}$ have been used instead of $\bar{p}_1$ and $\bar{p}_2$. $\mathbf{p}(\mathbf{p}_1\mathbf{p}_2) = \hat{\mathbf{p}}_1 = -\hat{\mathbf{p}}_2$.

To achieve interacting states, the noninteracting states are multiplied by a radial-wavefunction $f_q(p)$ and integrated $p^2\,dp$. This radial-wavefunction is the same relativistically and nonrelativistically. The S-matrix elements

$$\langle q'J'\mu'L'S'\mathbf{p}'(-) \, | \, qJ\mu LSP(+)\rangle$$

are the same relativistically and nonrelativistically. Finally, the different transformation coefficients (15a) and (15b) lead to formula (8) when the S-matrix is transformed to rectangular coordinates.

ACKNOWLEDGMENT

This work was supported in part by the United States Atomic Energy Commission.

[3] R. Fong and J. Sucher [J. Math. Phys. 5, 456 (1964)] have studied the problem of producing a covariant interaction from the group theoretical point of view.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Existence of N and D Matrices

JAMES B. HARTLE* AND JOHN R. TAYLOR†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

(Received 23 May 1966)

This paper establishes the existence of N and D matrices with the property that the partial-wave T matrix has the form $T = ND^{-1}$. We consider the case of a finite number of two-body channels and prove that, if T is analytic with right- and left-hand cuts and is suitably bounded, then N and D can be constructed with all the usual properties—namely, N and D have the left- and right-hand cuts, respectively, N is finite at the bound-state poles, and D tends to one as the energy goes to infinity.

I. INTRODUCTION

THE matrix N/D method is a widely used technique for computing scattering amplitudes in bootstrap calculations.[1] The technique was originally proposed by Bjorken[2] as a generalization of the single channel N/D method of Chew and Mandelstam.[3] In those situations where the method is usually applied, one is concerned with the scattering between n two-body channels described by a symmetric $(n \times n)$ partial wave scattering matrix $T(s)$. This amplitude has a left-hand cut whose discontinuity is considered as given (corresponding to the forces of the problem) and, starting at s_0, a right-hand cut whose discontinuity is to be determined consistent with unitarity. The matrix N/D technique for doing this can be summarized as follows: One *assumes* that T can be written in the form

$$T = ND^{-1}, \tag{1.1}$$

where D satisfies the following conditions: (i) D is real analytic with a right-hand cut, but no left-hand cut. It may have a finite number of real CDD poles.[4] (ii) TD is real on the right, so that N has no right-hand cut.[5] (iii) $D \to 1$ as $|s| \to \infty$.[6] (iv) TD is finite at the bound-state energies. One now uses unitarity to derive the usual nonsingular integral equation for D (or N) in terms of the left-hand discontinuity of T. This equation is then solved to determine T.

The N/D method hinges on the assumption that T *can* be written as ND^{-1}. One might imagine a situation in which the T arrived at from a more complete calculation could not be decomposed into ND^{-1}. In this case solutions of the N/D equations might still exist, but they could not yield the correct T. In this paper we examine the possibility of decomposing T as ND^{-1} and conclude that for any T with the usual properties this can be done with D satisfying conditions (i)–(iv).

At first, one might think that proving the existence of N and D matrices should present no problem. In the case of a single channel, for example, the existence proof is direct and simple. One forms the Omnes expression,

$$D_0(s) = (s - s_0)^{-m} \exp\left(-\frac{1}{\pi}\int_{s_0}^{\infty} ds' \frac{\delta(s')}{s' - s - i\epsilon}\right), \tag{1.2}$$

where we have taken

$$\delta(\infty) = 0, \quad \delta(s_0) = m\pi.$$

The function D_0 has no left-hand cut, and $N = TD_0$ is real on the right, so that D_0 satisfies conditions (i) and (ii). Multiplying D_0 by a suitable rational function, one can ensure that (iii) and (iv) are also satisfied. This construction guarantees the existence of at least one D function with the properties (i)–(iv).

For the multichannel case, however, there seems to be no simple generalization of the Omnes formula. One can, of course, write the S matrix as $S = \exp(2i\Delta)$, where the matrix Δ is real and symmetric, and then D defined by Eq. (1.2) (with δ replaced by the matrix Δ) certainly has no left-hand cut.

* Present address: Physics Department, University of California, Santa Barbara, California.
† Present address: Department of Physics, University of Colorado, Boulder, Colorado.

[1] See for example, F. Zachariasen and C. Zemach, Phys. Rev. **128**, 849 (1962); E. Abers and C. Zemach, *ibid.* **131**, 2305 (1963); J. Fulco, G. Shaw, and D. Wong, *ibid.* **137**, B1242 (1965). Other references may be found in F. Zachariasen, Lectures at the Pacific International Summer School in Physics (1965) (unpublished).

[2] J. Bjorken, Phys. Rev. Letters **4**, 473 (1960).

[3] G. Chew and S. Mandelstam, Phys. Rev. **119**, 467 (1960).

[4] L. Castillejo, R. H. Dalitz, and F. J. Dyson, Phys. Rev. **101**, 453 (1956).

[5] Our analysis applies equally well to cases where D cancels only a finite portion of the right-hand cut, as in the N/D equation associated with the strip approximation.

[6] If T does not vanish fast enough at infinity, one does not expect to find a D which tends to one. These cases are not discussed here. The existence of a D matrix when T satisfies weaker conditions at large s than those used here is established in a forthcoming paper of R. L. Warnock, who treats the problem from a different point of view. We are indebted to Dr. Warnock for a helpful correspondence and for drawing our attention to his own work, an abstract of which appeared in Bull. Am. Phys. Soc. **9**, 116 (1964).

Unfortunately, the resulting $N = TD$ is *not* real on the right.[7]

Another obvious approach to the explicit construction of a D matrix would be to diagonalize the S matrix, $S = U^\dagger S_d U$, and to construct a diagonal D matrix, D_d, using the Omnes formula (1.2) for each element separately. This diagonal D_d obviously has the desired properties, but when one returns to the original representation, the resulting $D = U^\dagger D_d U$ is unsuitable because the matrix U has a left-hand cut which it communicates to D.

The situation in potential scattering (with a Yukawa potential for example) bears out our contention that the multichannel case needs special attention. In the one-channel case $S = f(k)/f(-k)$, where the Jost function $f(k)$ is analytic in the lower-half k plane, and D defined as $f(-k)$ has all the required properties. In the multichannel case $S = F(K)F^{-1}(-K)$, but the Jost matrix $F(-K)$ in general *does* have a left-hand cut.[8] The identification $D = F(-K)$ in fact satisfies conditions (ii)–(iv) but obviously not (i). Thus, here too the existence of a proper D is unproved.

In Sec. II we prove that one can, in fact, always construct a suitable D for any T with the following properties: (1) $T(s)$ is an $(n \times n)$ symmetric matrix, analytic in the usual cut plane except at a finite number of bound-state poles. (2) On the right-hand cut T is continuous and piecewise analytic (i.e., has only a countable number of branch point singularities). (3) On the right-hand cut, each element of T and its first two derivatives is bounded by some power of s as follows:

$$|\rho T| < C/s^\alpha, \quad C, \alpha > 0,$$
$$|\rho T'| \text{ and } |\rho T''| < C/s^\beta, \quad \beta > 1, \tag{1.3}$$

where ρ is the phase space matrix ($S = 1 + 2i\rho^{\frac{1}{2}} T \rho^{\frac{1}{2}}$) and primes denote derivatives with respect to s. (4) T satisfies extended unitarity on the right-hand cut; i.e., $T^{-1} = Y - i\rho$, where Y is a real matrix.[9]

Some comment on these assumptions is in order. First, it may be seen that, apart from the analyticity of T, we use only properties of T on the right-hand cut. Secondly, the bound (1.3) is chosen so that the dispersion relation for $(D - 1)$ is unsubtracted. The case where subtractions are needed is not considered here.[6] The bounds on the derivatives of T are needed to eliminate pathological oscillations when s is large. Finally, our use of extended unitarity is comparatively inessential. In order to prove that a certain equation—Eq. (4.5)—is of Fredholm type we need to know that Im T^{-1} is finite on the right-hand cut; this is conveniently guaranteed by extended unitarity, which implies that Im $T^{-1} = -\rho_0$, the open-channel part of ρ.

Our procedure is first to assume that D exists and to derive an integral equation which it must satisfy. This equation is a singular integral equation for D in terms of the physical values of T. We show first that if the equation has solutions at all, then it has solutions which satisfy conditions (i)–(iv) and secondly that it *does* have solutions. We present the whole of this analysis in Sec. II, omitting only some mathematical details which are given in Sec. IV.

In Sec. III we discuss the single-channel problem and compare our construction of D with the usual one. This discussion illustrates the role of CDD poles in our approach. In Sec. IV we repair the omissions of Sec. II; in particular, we outline the proof of the theorem on singular integral equations used in Sec. II, and we verify that the D matrix which we construct does satisfy condition (iii) that $D \to 1$ and $s \to \infty$.

II. EXISTENCE OF THE D MATRIX

In order to derive an integral equation for D, we first assume that $T = ND^{-1}$, where D has properties (i)–(iv) of Sec. I. It then follows that D satisfies a dispersion relation,

$$D(s) = 1 + \sum_{i=1}^{p} \frac{R_i}{s - c_i} + \frac{1}{\pi} \int_{s_0}^{\infty} ds' \frac{\text{Im } D(s')}{s' - s - i\epsilon}, \tag{2.1}$$

where R_i are real matrices and c_i are real numbers less than s_0. To evaluate Im D we note first that $D = T^{-1}N$, where N is real on the right, so

$$\text{Im } D = (\text{Im } T^{-1})N = (\text{Im } T^{-1})TD. \tag{2.2}$$

Extended unitarity implies that

$$T^{-1} = Y - i\rho = Y_0 - i\rho_0, \tag{2.3}$$

where ρ_0 is the open-channel part of ρ, so that both Y_0 and ρ_0 are real everywhere on the right. Equations (2.2) and (2.3) then give

$$\text{Im } D = -\rho_0 TD,$$

whence the dispersion relation for D becomes

$$D(s) = 1 + \sum_{i=1}^{p} \frac{R_i}{s - c_i} - \frac{1}{\pi} \int_{s_0}^{\infty} ds' \frac{\rho_0(s')T(s')D(s')}{s' - s - i\epsilon}. \tag{2.4}$$

This relation may now be regarded as a singular integral equation for D in terms of the values of T

[7] We should remark that we are not concerned here with methods that use the *determinant* of the D matrix. It is clear that the Omnes formula, (1.2), can always be used to construct the determinant of D—or rather, the function which would be the determinant of D if D exists—but the existence of this "determinant of D" does not guarantee the existence of D itself.

[8] See, for example, R. G. Newton, J. Math. Phys. **2**, 188 (1961).

[9] See, for example, J. R. Taylor, Nucl. Phys. **58**, 580 (1964).

on the right-hand cut. Its solution in the single-channel case can be found in Muskhelishvili[10] and Omnes[11] and is given essentially by the Omnes formula (1.2). Properties (i)–(iv) can then be verified directly from the solution. In the many-channel case an explicit solution cannot in general be found, and we proceed as follows: Any solution D of Eq. (2.4) clearly defines a function analytic in s with the right-hand cut. Multiplication of D by suitable factors $(s-b)/(s-c)$ can obviously guarantee that TD is finite at the bound states. That any solution D of Eq. (2.4) tends to one as $s \to \infty$ is certainly very plausible; we give a proof at the end of Sec. IV. Thus to guarantee the existence of a matrix D having properties (i)–(iv) it is sufficient to prove that: (a) If Eq. (2.4) has any solution, it has solutions for which TD is real on the right. (b) The equation does have solutions.

To discuss the first question we rewrite the integral in Eq. (2.4) as a δ function plus principal value and take the δ function term to the left-hand side. Then, since from unitarity [Eq. (2.3)],

$$1 + i\rho_0 T = Y_0 T, \tag{2.5}$$

the integral equation becomes

$$Y_0 T D = 1 + \sum_{i=1}^{p} \frac{R_i}{s - c_i} - \frac{P}{\pi} \int_{s_0}^{\infty} ds' \, \frac{\rho_0(s') T(s') D(s')}{s' - s}. \tag{2.6}$$

The two integral equations (2.4) and (2.6) are equivalent; i.e., any solution of the first provides a solution of the second, and vice versa. But Eq. (2.6) may be regarded as an equation for $TD = N$ and, since Y_0, R_i, c_i, and ρ_0 are all real, its solutions can always be chosen real.[12] Since Eqs. (2.4) and (2.6) are equivalent, this means that if Eq. (2.4) has any solutions, then it has solutions for which TD is real on the right.

We conclude that if Eq. (2.4) has solutions, they can be chosen to satisfy properties (i)–(iv), and it remains only to show that it does have solutions. The existence of solutions to equations of this type is discussed in Sec. IV; we quote here only the results. With the assumptions of Sec. I for $T(s)$ a solution for D exists provided that

$$\int_{s_0}^{\infty} ds \left(1 + \sum_{i=1}^{p} \frac{R_i^t}{s - c_i}\right) \phi_\alpha(s) = 0, \quad \alpha = 1, \cdots, k, \tag{2.7}$$

where t denotes transpose and the column vectors $\phi_1, \cdots, \phi_k$ are a complete set of linearly independent solutions of the homogeneous adjoint equation of Eq. (2.4),

$$\phi(s) = \rho_0(s) T(s) \int_{s_0}^{\infty} \frac{ds'}{s' - s + i\epsilon} \phi(s'). \tag{2.8}$$

This theorem is the analog of the existence theorem for Fredholm integral equations.[13] In general, if there are solutions to Eq. (2.8), conditions (2.7) cannot be satisfied without CDD poles. However, since (as is shown in Sec. IV) k is finite, we can always choose $R_1 \cdots R_p$ so that these conditions are satisfied by choosing p sufficiently large. Thus, if we allow a sufficient (but finite) number of CDD poles, a D matrix having properties (i)–(iv) can always be found. It is easy to see that the minimum number of poles necessary to satisfy Eq. (2.7) must be greater than k/n, but this number may be modified by the additional factors $(s-b)/(s-c)$ by which this D must be multiplied in order to ensure that it vanishes at the bound-state energies.

This completes our proof that one can construct a suitable D matrix. To recapitulate briefly, we first derived the integral equation, (2.4), which D must satisfy if it exists. We next argued that if Eq. (2.4) has solutions, then they can be chosen to satisfy conditions (i)–(iv)—that D is analytic with only a right-hand cut is obvious from Eq. (2.4), that TD can be chosen real on the right was seen by inspection of the equivalent Eq. (2.6), that TD is finite at the bound states can be guaranteed by multiplication with suitable factors $(s-b)/(s-c)$, and the proof that $D \to 1$ as $s \to \infty$ is given in Sec. IV. Finally, that there *are* solutions (if we allow enough CDD poles) is guaranteed by the conditions stated in Eqs. (2.7) and (2.8) and proved in Sec. IV.

III. SINGLE-CHANNEL CASE AS AN EXAMPLE

In this section we show that, in the single-channel case, our construction of the D function agrees with the usual solutions obtained from the Omnes formula. We dispense temporarily with the requirement (iv) that D vanish at the bound states [since in both cases this can be achieved by simple multiplication by factors of the form $(s-b)/(s-c)$].

To obtain D from the Omnes formula we define

$$\delta(s_0) - \delta(\infty) = m\pi$$

and adjust δ so that $\delta(\infty) = 0$. The Omnes expression,

$$D_0(s) = (s - s_0)^{-m} \exp\left(-\frac{1}{\pi} \int_{s_0}^{\infty} ds' \, \frac{\delta(s')}{s' - s - i\epsilon}\right),$$

[10] N. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953).

[11] R. Omnes, Nuovo Cimento **21**, 524 (1961).

[12] If N is any solution so is $\frac{1}{2}(N + N^*)$.

[13] See, for example, F. Smithies, *Integral Equations* (Cambridge University Press, New York, 1958); see, also, Ref. 10.

is real analytic with a right-hand cut, has the phase $-\delta(s)$ on the right, is everywhere finite and nonzero, and behaves like s^{-m} as $s \to \infty$. The usual construction of D then distinguishes two possibilities: (1) $m \geq 0$. To ensure that $D \to 1$ as $s \to \infty$, one must multiply D_0 by m factors $(s - a_i)$. This gives a D function without CDD poles. (2) $m < 0$. In this case D_0 must be divided by $|m| = M$ factors $(s - c_i)$; i.e.,

$$D(s) = D_0(s)\left(\prod_{i=1}^{M}(s - c_i)\right)^{-1}. \tag{3.1}$$

Thus, D has M CDD poles.

In our approach D is determined as a solution of Eq. (2.4), which in this case becomes

$$D(s) = 1 + \sum_{i=1}^{p}\frac{r_i}{s - c_i} - \frac{1}{\pi}\int_{s_0}^{\infty} ds' \frac{\exp[i\delta(s')]\sin\delta(s')D(s')}{s' - s - i\epsilon}. \tag{3.2}$$

As described in Sec. II, solutions exist if the inhomogeneous term is orthogonal to all solutions ϕ of the homogeneous adjoint equation, Eq. (2.8). The latter equation, written in terms of

$$\psi = \phi^* \exp(i\delta)/\sin\delta \tag{3.3}$$

is just

$$\psi(s) = \frac{1}{\pi}\int_{s_0}^{\infty} ds' \frac{\exp[-i\delta(s')]\sin\delta(s')\psi(s')}{s' - s - i\epsilon}, \tag{3.4}$$

which is the homogeneous form of Eq. (3.2) with δ replaced by $-\delta$. The analysis of Sec. II [in particular Eq. (2.6)] therefore guarantees that if homogeneous solutions ϕ exist they can all be chosen so that ϕ [or $\psi \exp(-i\delta)$] is real on the right; i.e., ψ has the phase δ on the right. Now Eq. (3.4) shows that ψ is real analytic with only a right-hand cut. So if we define

$$P(s) = \psi(s)D_0(s), \tag{3.5}$$

then $P(s)$ is real analytic, and real on the whole real axis; i.e., $P(s)$ has no singularities at all. Now since $D_0 \sim s^{-m}$ as $s \to \infty$ while, from Eq. (3.4), $\psi \to 0$ as $s \to \infty$, we can distinguish two possibilities for $P(s)$: (1) $m \geq 0$. In this case Eq. (3.5) shows that $P \to 0$ as $s \to \infty$. Since $P(s)$ is an entire function this means that $P \equiv 0$; i.e., there are no solutions ϕ to the homogeneous adjoint of Eq. (3.2). Thus one can solve Eq. (3.2) with any inhomogeneous term and, in particular, one can find a D function with no CDD poles. (2) $m < 0$. In this case Eq. (3.5) shows that $s^m P \to 0$ as $s \to \infty$ which means that P is a polynomial of degree less than $|m| = M$. There are therefore M independent homogeneous adjoint solutions $\phi_1, \cdots, \phi_M$ which can be chosen so that

$$\psi_\alpha = s^{\alpha-1}/D_0(s) \sim 1/s^{M-\alpha+1} \quad \text{as} \quad |s| \to \infty. \tag{3.6}$$

If we now seek a D function without CDD poles we must examine the conditions $\int_{s_0}^{\infty} ds\phi_\alpha = 0$ for $\alpha = 1, \cdots, M$. Since from Eq. (3.3)

$$\phi = \text{Im}\,\psi, \quad s_0 \leq s < \infty,$$

the integral of interest is

$$\int_{s_0}^{\infty} ds\phi_\alpha(s) = \int_{s_0}^{\infty} ds\ \text{Im}\ \psi_\alpha(s) = \frac{1}{2i}\int_C ds\psi_\alpha(s) = \frac{1}{2i}\int_{C'} ds\psi_\alpha(s) = -\pi\delta_{\alpha M}, \tag{3.7}$$

where the contours C and C' are shown in Fig. 1. The last step follows because of the asymptotic behavior (3.6), and it makes clear that the homogeneous adjoint solution ϕ_M is not orthogonal to 1. Thus there exist no solutions D without CDD poles; exactly as we would expect.

The condition that a solution with CDD poles exists is of course

$$\int_{s_0}^{\infty} ds\left(1 + \sum_{i=1}^{p}\frac{r_i}{s - c_i}\right)\phi_\alpha(s) = 0, \quad \alpha = 1, \cdots, M. \tag{3.8}$$

The first integral here is given by Eq. (3.7), while a similar calculation shows that

$$\int_{s_0}^{\infty} ds \frac{\phi_\alpha(s)}{s - c_i} = \frac{\pi(c_i)^{\alpha-1}}{D_0(c_i)}.$$

Thus condition (3.8) for the existence of D becomes

$$\sum_{i=1}^{p}(c_i)^{\alpha-1}\frac{r_i}{D_0(c_i)} = \delta_{\alpha M}. \tag{3.9}$$

If we regard the pole positions c_i as fixed, then exactly M poles are needed to satisfy Eq. (3.9) with the residues $r_1, \cdots, r_M$ uniquely determined. In fact, the

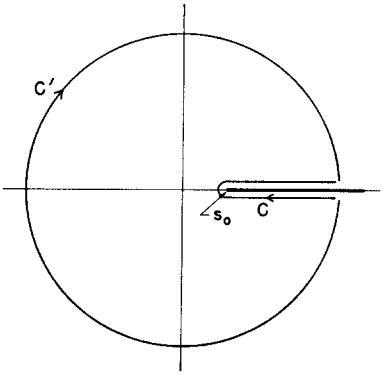


FIG. 1. The plane of the variable s with the contours C and C' used in Eq. (3.7).

matrix in Eq. (3.9) whose (αi) element is $(c_i)^{\alpha-1}$ is just a Vandermonde matrix[14] and can be explicitly inverted to give the residues,

$$r_j = D_0(c_j)\Big(\prod_{i\neq j}(c_j - c_i)\Big)^{-1},$$

in exact agreement with Eq. (3.1).

This means that the usual constructive method and the one based on the integral equation both yield the same D. CDD poles are required when $m < 0$ but not when $m \geq 0$. If D is then required to vanish at the n_B bound states, we need no CDD poles if $(n_B - m) \leq 0$ and exactly $(n_B - m)$ if this quantity is greater than zero.

IV. PROOF OF THE THEOREM ON SINGULAR INTEGRAL EQUATIONS

The main purpose of this section is to establish the conditions stated in Eqs. (2.7) and (2.8) for the existence of solutions to Eq. (2.4). These conditions follow from a theorem due to Noether and Giraud,[15] which we outline here for the sake of completeness. We first rewrite Eq. (2.4) as

$$Kf = g, \tag{4.1}$$

where K is the integral operator with kernel,

$$K(s, s') = 1\delta(s - s') + \frac{1}{\pi}\frac{\rho_0(s')T(s')}{s' - s - i\epsilon},$$

and we have written f for any column of the matrix $(D - 1)$ and g for the column of the corresponding inhomogeneous term from Eq. (2.4). The conditions (2.7) and (2.8) which we wish to prove can now be restated: Eq. (4.1) has solutions if and only if

$$\int_{s_0}^{\infty} ds(g, \phi) = 0 \tag{4.2}$$

for all ϕ which satisfy

$$K^{\dagger}\phi = 0, \tag{4.3}$$

where (g, ϕ) denotes the scalar product $\sum_i g_i^*(s)\phi_i(s)$

The necessity of the condition (4.2) follows exactly as in the proof of Fredholm's theorem and needs no discussion. To prove sufficiency, we introduce the "reducing operator" M with kernel

$$M(s, s') = 1\delta(s - s') - \frac{1}{\pi}\frac{\rho_0(s')T(s')}{s' - s + i\epsilon}.$$

This operator reduces the singular equation $Kf = g$ to a Fredholm equation,

$$MKf = Mg, \tag{4.4}$$

as we now verify. One can easily check that Eq. (4.4) has the form[16]

$$f(s) = h(s) - \int_{s_0}^{\infty} ds' L(s, s')f(s'), \tag{4.5}$$

where

$$L(s, s') = [1 - 2i\rho_0(s)T^*(s)]\frac{Q(s) - Q(s')}{s - s'}\rho_0(s')T(s'),$$

$$Q(s) = \frac{1}{\pi^2}\int_{s_0}^{\infty} ds' \frac{\rho_0(s')T(s')}{s' - s + i\epsilon},$$

and there is a corresponding expression for the inhomogeneous term $h(s)$. Now using the bound (1.3) on $T(s)$, it is a simple matter to verify that $Q(s)$ satisfies the bounds

$$|Q(s)| < C/s^{\alpha-\epsilon} \quad \text{and} \quad |Q'(s)| < C/s,$$

where ϵ is an arbitrary positive number.[17] From these it follows that the kernel $L(s, s')$ is L_2. The inhomogeneous term $h(s)$ is bounded by $C/s^{\alpha-\epsilon}$ and so, unless $\alpha > \frac{1}{2}$, is not L_2; however, by iterating Eq. (4.5) n times, we can rewrite it as an equation for

$$f_n(s) = f(s) - h(s) - h_2(s) - \cdots - h_{n-1}(s) \tag{4.6}$$

with the same kernel but with inhomogeneous term $h_n(s)$ satisfying

$$|h_n(s)| < C/s^{n\alpha-\epsilon}.$$

In this way we can obviously obtain a Fredholm equation as required.

A *necessary* condition that f satisfy $Kf - g = 0$ is that it satisfy the Fredholm equation $MKf - Mg = 0$. The necessary and sufficient conditions for solutions of a Fredholm equation are given by Fredholm's theorem; in our case,

$$\int_{s_0}^{\infty} ds(Mg, \psi_\beta) = 0, \quad \beta = 1, \cdots, l$$

or, equivalently,

$$\int_{s_0}^{\infty} ds(g, M^{\dagger}\psi_\beta) = 0, \quad \beta = 1, \cdots, l, \tag{4.7}$$

where $\psi_1, \cdots, \psi_l$ are independent solutions of

[14] See, for example, J. V. Uspensky, *Theory of Equations* (McGraw-Hill Book Company, Inc., New York, 1948), p. 214ff.

[15] F. Noether, Math. Ann. **82**, 42 (1921). G. Giraud, Ann. 'Ecole Norm. Suppl., **51**, 251 (1934). The exposition we follow is that of Ref. 10, to which the reader is referred for more detail.

[16] For more detail see S. Mandelstam [Phys. Rev. **140**, B375 (1965)], who treats this same equation from a somewhat different point of view.

[17] This can be checked directly by splitting the integral defining $Q(s)$ into three parts $(s_0, s - 1)$, $(s - 1, s + 1)$, and $(s + 1, \infty)$. Alternatively, it follows at once from the general theorem on dispersion integrals given by L. Lanz and G. M. Prosperi [Nuovo Cimento **33**, 201 (1964)]. The bound on the derivative $Q'(s)$ can be established by proving that $[Q(s + h) - Q(s)]/h$ is bounded by C/s, where C is independent of h.

$(MK)^\dagger\psi = 0$. If conditions (4.7) are satisfied, then the equation $MKf - Mg = 0$ has solutions of which the most general is

$$f = RMg + \sum_{\gamma=1}^{q} a_\gamma \chi_\gamma, \tag{4.8}$$

where $\chi_1, \cdots, \chi_q$ are the independent solutions of $MK\chi = 0$ and R is the resolvent of MK. This vector f does not necessarily satisfy $Kf - g = 0$, but since it does satisfy $M(Kf - g) = 0$, it is clear that

$$Kf - g = \sum_{\delta=1}^{r} b_\delta \xi_\delta, \tag{4.9}$$

where $\xi_1, \cdots, \xi_r$ are a complete orthonormal set of independent solutions of $M\xi = 0$. (The number r is finite since the ξ satisfy the Fredholm equation $KM\xi = 0$.)

We can now find conditions under which the arbitrary coefficients $a_1, \cdots, a_q$ in the definition (4.8) of f can be chosen so that all the $b_1, \cdots, b_r$ in Eq. (4.9) are zero. If this can be done then the resulting f satisfies $Kf - g = 0$. From Eqs. (4.9) and (4.8), it is easily seen that the b_δ depend on the a_γ as

$$b_\delta = \sum_{\gamma=1}^{q} A_{\delta\gamma} a_\gamma + c_\delta, \tag{4.10}$$

where the numbers $A_{\delta\gamma}$ are independent of the a_γ and g, and

$$c_\delta = \int_{s_0}^{\infty} ds(g, [KRM - I]^\dagger \xi_\delta). \tag{4.11}$$

The necessary and sufficient conditions that the a_γ can be chosen so that the numbers $b_1, \cdots, b_r$ in Eq. (4.10) are all zero have the general form

$$\sum_{\delta=1}^{r} B_{\sigma\delta} c_\delta = 0, \quad \sigma = 1, \cdots, t \text{ (for some } t),$$

which, by Eq. (4.11), can be written

$$\int_{s_0}^{\infty} ds(g, \zeta_\sigma) = 0, \quad \sigma = 1, \cdots, t, \tag{4.12}$$

where the exact form of the vectors $\zeta_1, \cdots, \zeta_t$ is uninteresting.

Thus the necessary and sufficient conditions that $Kf - g = 0$ have a solution are Eqs. (4.7) and (4.12), which can be rewritten as

$$\int_{s_0}^{\infty} ds(g, \varphi_\alpha) = 0, \quad \alpha = 1, \cdots, (l + t). \tag{4.13}$$

It remains only to show that conditions (4.13) are implied by the conditions of the theorem, namely,

$$\int_{s_0}^{\infty} ds(g, \phi) = 0 \tag{4.2}$$

for all ϕ satisfying $K^\dagger\phi = 0$. To prove this we consider the equation (for f) $Kf = Kh$, where h is arbitrary. Since this equation has a solution ($f = h$), the conditions (4.13) must hold; namely,

$$\int_{s_0}^{\infty} ds(Kh, \varphi_\alpha) = 0,$$

which implies

$$\int_{s_0}^{\infty} ds(h, K^\dagger \varphi_\alpha) = 0.$$

Since h is arbitrary, this implies that $K^\dagger\varphi_\alpha = 0$; i.e., all the φ_α of conditions (4.13) satisfy $K^\dagger\phi = 0$. Thus the conditions (4.2) imply the necessary and sufficient conditions (4.13), as required.

In order to apply these conditions as in Sec. II, it is essential that the number of independent solutions of the homogeneous equation $K^\dagger\phi = 0$ be finite. This follows from the fact that any solution of $K^\dagger\phi = 0$ also satisfies the Fredholm equation $(KM)^\dagger\phi = 0$.

Finally we must check that the D matrix determined by the integral equation, (4.1) or (2.4), satisfies $D \to 1$ as $s \to \infty$. Since in this section we have written f for any column of the matrix $(D - 1)$, we must show that $f \to 0$ as $s \to \infty$. Returning to the Fredholm equation, (4.5), and using the Schwartz inequality, one sees that

$$\|f_n(s)\| \leq \|h_n(s)\| + \left(\int_{s_0}^{\infty} ds' \, \|L(s, s')\|^2\right)^{\frac{1}{2}} \left(\int_{s_0}^{\infty} ds' \, \|f_n(s')\|^2\right)^{\frac{1}{2}}.$$

Now $f_n(s)$ is certainly L_2 and using the bounds already given, it is simple to check that both terms on the right are bounded by $C/s^{\alpha-\epsilon}$. It follows that every element of the matrix $(D - 1)$ tends to zero faster than $C/s^{\alpha-\epsilon}$ as $|s| \to \infty$.

ACKNOWLEDGMENTS

We wish to thank Dr. M. Froissart, Professor M. L. Goldberger, and Dr. C. Edward Jones, all of whom read drafts of this paper and made several helpful suggestions.

This work was supported by the United States Air Force Office of Research, Air Research and Development Command under Contract AF49(638)-1545.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Isoparity and Simple Lie Group

KOSAI TANABE
Department of Physics, Saitama University, Urawa, Japan

KAZUHISA SHIMA
Department of Physics, Nagoya University, Nagoya, Japan

(Received 26 July 1966)

The direct generalization of the isoparity (or G-parity), with the defining property that it is commutable with the referring internal symmetry group, is investigated on the basis of the theory of Lie algebra. This is one special problem of the group extension of a simple Lie group by an involution. It is shown that the isoparity of this type can be defined for the simple Lie groups $SU(2)(A_1$ type), $SO(2l+1)(B_l,\ l \geq 2)$, $Sp(2l)(C_l,\ l \geq 2)$, $SO(2l)(D_l,\ l \geq 3)$, G_2, F_4, E_7, and E_8, but not for the $SU(l+1)(A_l, l \geq 2)$. The relation between the inner automorphism group and the Weyl group of the simple Lie algebra concerned is available to construct the isoparity operator explicitly. Some illustrative examples are presented.

1. INTRODUCTION

As is well known, the operation of charge conjugation C does not commute with that of rotation in isotopic spin space, though the invariance of the strong interaction of elementary particles under charge conjugation leads the strict selection rules independent of the conservation of isotopic spin. However, Michel[1] resolved this apparent difficulty by introducing the isoparity (or G-parity[2]) G_p defined by

$$G_p = CR = RC, \tag{1}$$

where $R = \exp(i\pi I_2)$ is the rotation by angle π about the second axis of isotopic spin space in the usual representation. Then such an operator G_p becomes commutable with all the generators I_i of rotation in isotopic spin space,

$$[G_p, I_i] = 0 \quad (i = 1, 2, 3). \tag{2}$$

Accordingly, the existence of simultaneous eigenstates of isotopic spin and isoparity provides with an aid in deriving useful selection rules for many reactions. In this respect, the isoparity G_p may be interpreted as the redefined particle–antiparticle conjugation instead of the charge conjugation C when there exists an internal symmetry.

From the most general point of view, the isoparity problem may be regarded as that of the group extension[3,4] of the internal symmetry group by an involution. Biedenharn *et al.*[5] have exhausted all the possibilities of the extended groups. In the case of $SU(3)$, the similar treatment was already carried out by Dothan.[6]

In the present paper, the generalization of the isoparity to the internal symmetry built on any simple Lie group except for E_6 is considered on the basis of the theory of Lie algebra in such a way that the isoparity operator commutes with hyper-charge as well as isotopic spin. Hence, our discussions are restricted to the *strong isoparity*[5] which has an obvious physical meaning and is associated with the reflection operator described by inner automorphism. It seems, however, worthwhile to disclose in a simple and lucid way the general character of the isoparity of this type, and to derive the isoparity operator of the general form[7] by means of the Weyl group[8] of the simple Lie algebra concerned.

In Sec. 2, it is shown that our problem of isoparity is simply connected with that of group extension determined by a factor set. Such an approach to the problem makes clear the most general mathematical aspects of Dothan's discussion.

In Sec. 3, the structure of the automorphism group of every simple Lie algebra is fully investigated, and what simple Lie group suits itself to the isoparity of our type is examined. For the simple Lie group $SU(n)(n \geq 3)$, the isoparity cannot be defined with *inner automorphism*. On the other hand, it is possible to define it for the simple Lie groups $SU(2)(A_1$ type),

[1] L. Michel, Nuovo Cimento **10**, 319 (1953).
[2] T. D. Lee and C. N. Yang, Nuovo Cimento **3**, 749 (1956); C. Goebel, Phys. Rev. **130**, 258 (1956).
[3] H. Zassenhaus, *Lehrbuch der Gruppentheorie* (Springer-Verlag, Berlin, 1937), Vol. I, Chap. 3.
[4] L. Michel, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gursey, Ed. (Gordan and Breach Science Publishers, Inc., New York, 1964), p. 135; F. Kamber und N. Stautmann, Helv. Phys. Acta **37**, 563 (1965).
[5] L. C. Biedenharn, J. Nuyts, and H. Ruegg, Commun. Math. Phys. **2**, 231 (1966).
[6] Y. Dothan, Nuovo Cimento **30**, 399 (1963).
[7] K. Tanabe, Ph.D. thesis, University of Tokyo (1965).
[8] N. Jacobson, *Lie algebras* (Interscience Publishers, Inc., New York, 1962).

$SO(2l+1)(B_l, l \geq 2)$, $Sp(2l)(C_l, l \geq 2)$, $SO(2l) \times (D_l, l \geq 3)$, G_2, F_4, E_7, and E_8. The result for $SU(n)$ is essentially the same as that given by Okubo and Mukunda.[9]

Finally, in Sec. 4, an actual method is investigated to construct the isoparity operator. Some examples are given by means of this method.

2. GENERAL CONSIDERATION OF ISOPARITY

For the sake of self-contained description, we first make a brief survey of several definitions and a theorem on group extension.[3] Let there be given two groups N and F. A group E is called an extension of N by F if the following relation hold,

$$E/N \cong F. \tag{3}$$

Such a group E includes N as its normal subgroup and is expressed as the sum of the coset $\mathfrak{M}_x$ corresponding to every element x of F. An arbitrary representative u_x is selected from every coset

$$\mathfrak{M}_x(u_x N = N u_x = \mathfrak{M}_x).$$

Since $u_x u_y \in \mathfrak{M}_{xy}$, it follows that

$$u_x u_y = \omega(x, y) u_{xy}, \quad {}^{\exists}\omega(x, y) \in N. \tag{4}$$

If, for every element x of F, an automorphism S_x of N is defined by

$$S_x(a) = u_x a u_x^{-1} \quad \text{for} \quad {}^{\forall}a \in N, \tag{5}$$

the multiplication law of E is given in the following way:

$$(a u_x)(b u_y) = a S_x(b) \omega(x, y) u_{xy}. \tag{6}$$

Thus, the multiplication law of E is completely determined by the respective ones of N and F and by the set $\{S_x, \omega(x, y)\}$, which satisfies the following conditions:

$$S_x[S_y(a)] = \omega(x, y) S_{xy}(a) \omega(x, y)^{-1}, \tag{7}$$

$$\omega(x, y)\omega(xy, z) = S_x \omega(y, z) \omega(x, yz). \tag{8}$$

If, especially, u_1 is equated to 1, it follows that

$$S_1 = 1, \tag{9}$$

$$\omega(1, x) = \omega(x, 1) = 1 \quad \text{for} \quad {}^{\forall}x \in F. \tag{10}$$

In general, if there are given an automorphism S_x of N for every element x of F and an element $\omega(x, y)$ of N for every pair (x, y) of elements of F such that Eqs. (7) and (8) are satisfied, the set $\{S_x, \omega(x, y)\}$ is called a *factor set* of F with respect to N. Furthermore, if Eqs. (9) and (10) are satisfied, the factor set is said to be normalized.

[9] S. Okubo and N. Mukunda, Ann. Phys. (N.Y.) **36**, 311 (1966).

As has been seen above, a factor set is determined by a given extension. Conversely, if a factor set is given, the group extension can be constructed by means of the set.

Schreier's Theorem: Let the set $\{S_x, \omega(x, y)\}$ be a factor set of F with respect to N. If the following operation is introduced into the set $G = N \times F$:

$$(a, x)(b, y) = (a S_x(b) \omega(x, y), xy), \tag{11}$$

then E becomes an extension of N by F provided the normal subgroup $\{(a, 1); a \in N\}$ of E is identified with N. Furthermore, all the extensions of N by F are obtained in such a way.

This theorem is useful for our purpose. Since it is known that any factor set is equivalent to some normalized one (which means that the corresponding extended groups are isomorphic to each other), we hereafter use a normalized factor set.

Let $F = \{1, C\}$ be a cyclic group of order two ($C^2 = 1$), which represents the group of charge conjugation. Then a normalized factor set of F is as follows:

$$\{S_C, S_1 = 1; \omega(1, 1) = 1, \omega(1, C) = 1, \omega(C, 1) = 1, \omega(C, C)\}.$$

For the sake of simplicity, hereafter S_C is abbreviated to S and $\omega(C, C)$ to ω, and the former is called the *automorphism corresponding to charge conjugation.* These quantities S and ω satisfy the following relations:

$$S(\omega) = \omega, \tag{12}$$

$$S^2(a) = \omega a \omega^{-1} \quad \text{for} \quad {}^{\forall}a \in N. \tag{13}$$

Schreier's theorem determines the operation in an extension of N by F in the subsequent way:

$$(a, 1)(b, 1) = (ab, 1), \quad (a, 1)(b, C) = (ab, C),$$
$$(a, C)(b, 1) = (aS(b), C), \quad (a, C)(b, C) = (aS(b)\omega, 1). \tag{14}$$

Thus, the way of extending N by F depends on the choice of S and ω.

We are now in a position to define an isoparity of general form by using an extension E of N by F, where N is taken to be the representing group [meaning that $N = D(G)$, where G is the internal symmetry group and D is the referring irreducible representation of G] of an internal symmetry group by the referring irreducible representation. The automorphism corresponding to charge conjugation, S, is defined as

$$S(a) = \bar{a} \quad \text{for} \quad {}^{\forall}a \in N, \tag{15}$$

where $\bar{a}$ means the complex conjugate of a. [Needless to say, the definition of the operation of charge conjugation is not unique, because the right-hand side of Eq. (15) may include a phase factor. But this is not essential for the following discussions.] It follows from Eq. (13) that ω belongs to the center of N. Then ω is determined as 1 or -1 from Eq. (12). By an isoparity we understand an element G_p of E satisfying the equation

$$aG_pa^{-1} = G_p \quad \text{for} \quad {}^\forall a \in N. \tag{16}$$

Expressing G_p as (R, C), ${}^\exists R \in N$, we see that the equation above is equivalent to

$$R^{-1}aR = S(a) \quad \text{for} \quad {}^\forall a \in N, \tag{17}$$

because $(a, 1)(R, C)(a, 1)^{-1} = (aRS(a^{-1}), C)$. Therefore, a necessary and sufficient condition for the existence of the isoparity is that the automorphism corresponding to charge conjugation is expressed in terms of an inner automorphism of N.

Thus, we have constructed the most general framework to treat the isoparity within and derived a necessary and sufficient condition for its being well defined.

3. SIMPLE LIE ALGEBRAS FOR WHICH ISOPARITY IS ADAPTED

In this section, our consideration is restricted to the case when the internal symmetry group G is a simple Lie group, and what type of G adapts itself for our isoparity is investigated. In the previous section, we derived a necessary and sufficient condition for existence of the isoparity. The theory of Lie algebra helps us to reduce the condition to a more available one.

The Lie algebras of G and N are denoted by $\mathfrak{g}$ and $\mathfrak{n}$, respectively. Then the differential ρ,

$$\rho(X) = \lim_{t\to 0} [D\{\exp(tX)\} - E]/t \quad \text{for} \quad {}^\forall X \in \mathfrak{g},$$

of the referring irreducible representation D of G is also an irreducible representation of $\mathfrak{g}$, and $\mathfrak{n}$ is the image of $\mathfrak{g}$ by $\rho[\mathfrak{n} = \rho(\mathfrak{g})]$. Then the condition mentioned above, i.e., the condition (17) can be reduced to

$$R^{-1}\rho(X)R = -[\rho(X)]^t \quad \text{for} \quad {}^\forall X \in \mathfrak{g}. \tag{18}$$

Here, it is to be noticed that the regular matrix R satisfying Eq. (18) is either symmetric or antisymmetric if it exists. This is proved as follows. It can be easily verified from Eq. (18) that $R(R^t)^{-1}$ commutes with $\rho(X)$ for all X belonging to $\mathfrak{g}$. Since ρ is irreducible, it follows from Schur's lemma that $R(R^t)^{-1} = cE$, i.e., $R = cR^t$ for some complex number c. Hence,

$$R^t = \pm R. \tag{19}$$

Returning to the main course, we examine the relation between $\mathfrak{n}$ and $\mathfrak{g}$. It is easily proved that ρ is faithful, since $\mathfrak{g}$ is simple. Hence, $\mathfrak{n}$ is isomorphic to $\mathfrak{g}$. Therefore, $\mathfrak{n}$ can be identified with $\mathfrak{g}$ $[\rho(X) = X]$. In what follows we make such an identification. Then Eq. (18) can be rewritten as

$$R^{-1}XR = -X^t \quad \text{for} \quad {}^\forall X \in \mathfrak{g}. \tag{20}$$

It is easily seen that the mapping $X \to -X^t$ is an automorphism of $\mathfrak{g}$. Thus, we can state that a necessary and sufficient condition of the isoparity being adapted for $\mathfrak{g}$ is that the automorphism $X \to -X^t$ of $\mathfrak{g}$ is reduced to an inner automorphism of $\mathfrak{g}$.

Let us now turn our attention to the relation between the automorphism group[10] of $\mathfrak{g}$ and the corresponding inner automorphism group. Since the latter denoted by $I(\mathfrak{g})$ is a normal subgroup of the former denoted by $A(\mathfrak{g})$, it is important for our problem to disclose the character of the factor group $A(\mathfrak{g})/I(\mathfrak{g})$. Fortunately, the group has already been investigated by mathematicians. It can be determined from the group of the *automorphisms of the Dynkin diagram* associated with $\mathfrak{g}$. An automorphism of the Dynkin diagram means a one-to-one mapping $\alpha_i \to \alpha_{i'}$ in the diagram such that $(\alpha_i, \alpha_i) = (\alpha_{i'}, \alpha_{i'})$ and for any i, j the number of lines connecting α_i to α_j is equal to that connecting $\alpha_{i'}$ and $\alpha_{j'}$. It can be easily verified that all the automorphisms of each Dynkin diagram in Fig. 1 are reduced to the identity mapping $\alpha_i \to \alpha_i$ $(i = 1, \cdots, l)$ only. Hence it follows that $A(\mathfrak{g})/I(\mathfrak{g}) = 1$, i.e., $A(\mathfrak{g}) = I(\mathfrak{g})$ in this case. In other words, the automorphisms of $\mathfrak{g}$ are all inner ones. Thus, we arrive at the conclusion that our isoparity can be defined for the simple Lie algebras of types A_1, $B_l(l \geq 2)$, $C_l(l \geq 2)$, G_2, F_4, E_7, and E_8.

On the other hand, another automorphism exists for the Dynkin diagrams of types $A_l(l \geq 2)$, $D_l(l \geq 5)$, and E_6. As is easily seen, the mapping

$$\alpha_i \to \alpha_{l+1-i} \quad (i = 1, \cdots, l)$$

is an automorphism for $A_l(l \geq 2)$, the one

$$\alpha_i \to \alpha_i \quad (i \leq l-2),$$

$a_{l-1} \to \alpha_l$, $\alpha_l \to \alpha_{l-1}$ for $D_l(l \geq 5)$ and the one

$$\alpha_i \to \alpha_{6-i} \quad (i \leq 5),$$

$\alpha_6 \to \alpha_6$ for E_6. Hence we see that $A(\mathfrak{g})/I(\mathfrak{g})$ is a cyclic group of order two. There remains untouched the Dynkin diagram of type D_4. This diagram has automorphisms which permute α_1, α_3, α_4 and leave α_4

[10] N. Jacobson, Ref. 8, Chap. IX.

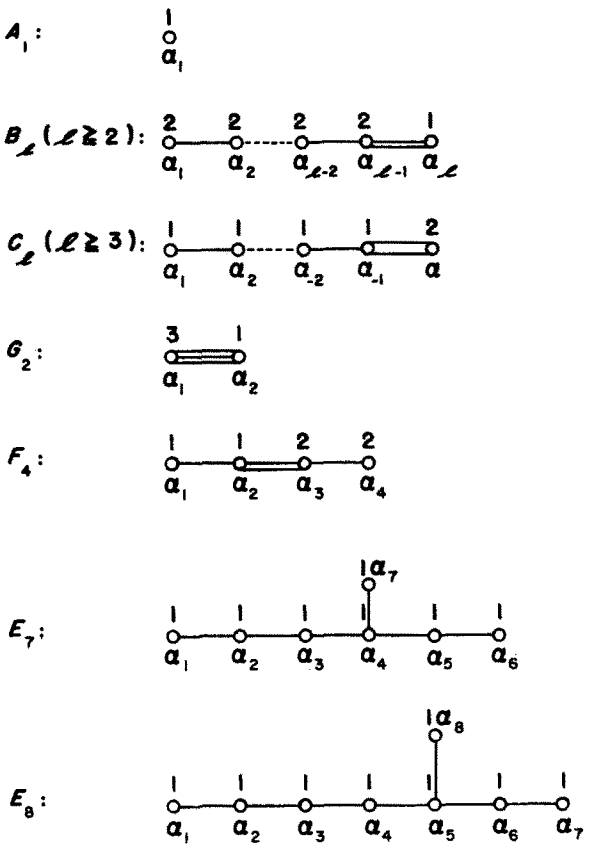


FIG. 1. Dynkin diagrams of simple Lie algebras of types A_l, $B_l(l \geq 2)$, $C_l(l \geq 3)$, G_2, F_4, E_7, and E_8.

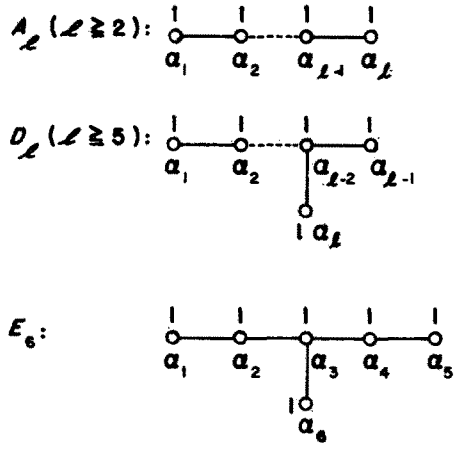


FIG. 2. Dynkin diagrams of simple Lie algebras of types $A_l(l \geq 2)$, $D_l(l \geq 5)$, and E_6.

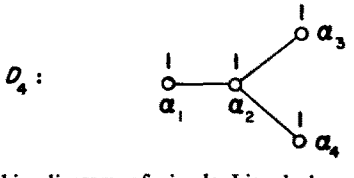


FIG. 3. Dynkin diagram of simple Lie algebra of type D_4.

fixed. Consequently, it can be inferred that $A(\mathfrak{g})/I(\mathfrak{g})$ is isomorphic to the symmetric group of degree three. Thus, not all the automorphisms of $\mathfrak{g}$ are reduced to inner ones of $\mathfrak{g}$ in case $\mathfrak{g}$ is the simple Lie algebra of type $A_l(l \geq 2)$, $D_l(l \geq 4)$, or E_6. Therefore, it becomes necessary to investigate directly whether the automorphism $X \to -X^t$ belongs to $I(\mathfrak{g})$ or not.

We first examine the case when $\mathfrak{g}$ is of type $A_l(l \geq 1)$. Then $\mathfrak{g}$ may be regarded as the Lie algebra of all the traceless matrices of degree $(l+1)$, and the dimension of $\mathfrak{g}$ is equal to $l(l+2)$. We here rewrite Eq. (20) as

$$R^{-1}X^tR = -X. \tag{21}$$

The set of all X satisfying this equation becomes an orthogonal or a symplectic Lie algebra according as R is symmetric or antisymmetric. The dimension of the Lie algebra is equal to $\frac{1}{2}(l+1)(l+2)$ or $\frac{1}{2}l(l+1)$. Since $\mathfrak{g}$ is of dimension $l(l+2)$, we must have either $l(l+2) = \frac{1}{2}(l+1)(l+2)$ or $l(l+2) = \frac{1}{2}l(l+1)$. Then we get the only solution $l = 1$. Hence the automorphism $X \to -X^t$ does not belong to $I(\mathfrak{g})$ if $l \geq 2$. Thus, we reach the conclusion that our isoparity cannot be defined for the simple Lie algebra of type $A_l(l \geq 2)$.

The simple Lie algebra of type $D_l(l \geq 4)$ is composed of matrices X which satisfy the equation $X = -X^t$, so that the matrix R satisfying Eq. (20) can be taken as the unit matrix. Therefore, our isoparity can be defined for this Lie algebra.

The simple Lie algebra which is not covered by the above considerations is that of type E_6 only. There remains unsolved the problem whether or not our isoparity is adapted for this Lie algebra.

4. EXPLICIT DETERMINATION OF ISOPARITY

This section is devoted to construct the reflection operator R defined by Eq. (20). An available formula to evaluate R by is given on the basis of the Weyl group[11] of $\mathfrak{g}$.

As seen later, the inner automorphism associated with R (which is nothing but the mapping $X \to -X^t$) maps the Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$ into $\mathfrak{h}$ itself. We examine beforehand the relationship between that subgroup $I_0(\mathfrak{g})$ of $I(\mathfrak{g})$ which makes $\mathfrak{h}$ invariant and the Weyl group W of $\mathfrak{g}$. For an element σ of $I_0(\mathfrak{g})$, let σ^* denote the dual transformation in $\mathfrak{h}^*$ of the restriction of σ onto $\mathfrak{h}$. Namely, σ^* is defined by

$$\{\sigma^*(\lambda)\}(H) = \lambda\{\sigma(H)\} \quad \text{for } {}^\forall\lambda \in \mathfrak{h}^* \text{ and } {}^\forall H \in \mathfrak{h}. \tag{22}$$

Then the mapping $\sigma \to (\sigma^*)^{-1}$ is a homomorphism of $I_0(\mathfrak{g})$ onto W. To see this, it is intended to investigate

[11] N. Jacobson, Ref. 8, Chap. VIII.

what element of $I_0(\mathfrak{g})$ corresponds to a given *Weyl reflection* S_α of W, where S_α is defined by

$$S_\alpha(\lambda) = \lambda - 2(\lambda, \alpha)/(\alpha, \alpha)\alpha \quad \text{for} \quad {}^\forall\lambda \in \mathfrak{h}^*.$$

For this purpose, use is made of Weyl's standard basis[12] $\{H_1, \cdots, H_l, E_\alpha; \alpha \in \Delta\}$ of $\mathfrak{g}$, where l is the rank of $\mathfrak{g}$, Δ is the set of all the nonzero roots of $\mathfrak{g}$, H_i is a basis of $\mathfrak{h}$, and E_α is an element of the eigensubspace of $\mathfrak{g}$ corresponding to a root α such that $B(E_\alpha, E_{-\alpha}) = -1$, where $B(\ ,\ )$ is the Killing form[13] of $\mathfrak{g}$, which is defined by

$$B(X, Y) = \mathrm{Tr}\,[\mathrm{ad}(X)\mathrm{ad}(Y)] \quad \text{for} \quad {}^\forall X,\ {}^\forall Y \in \mathfrak{g}.$$

This basis has the following properties.

$$[H_i, H_j] = 0, \tag{23}$$

$$[E_\alpha, E_{-\alpha}] = -H_\alpha, \tag{24}$$

$$[H, E_\alpha] = \alpha(H)E_\alpha \quad \text{for} \quad {}^\forall H \in \mathfrak{h}. \tag{25}$$

Here H_α is the element of $\mathfrak{h}$ defined by

$$B(H_\alpha, H) = \alpha(H) \quad \text{for} \quad {}^\forall H \in \mathfrak{h}. \tag{26}$$

By the use of this basis, the inner automorphism

$$\sigma_\alpha = \exp\left[\mathrm{ad}\left\{\frac{\pi}{[2(\alpha, \alpha)]^{\frac{1}{2}}}(E_\alpha + E_{-\alpha})\right\}\right] \tag{27}$$

is introduced for every positive root α. This mapping makes $\mathfrak{h}$ invariant and corresponds to S_α through the homomorphism mentioned above, as can be seen below. It is easily checked with the help of the mathematical induction with respect to p that

$$[\mathrm{ad}(X)]^{2p+1}H = (-1)^{p+1}\pi^{2p+1}\frac{\alpha(H)}{[2(\alpha, \alpha)]^{\frac{1}{2}}}(E_\alpha - E_{-\alpha}), \tag{28}$$

$$[\mathrm{ad}(X)]^{2p+2}H = (-1)^{p+1}\pi^{2p+2}\frac{\alpha(H)}{(\alpha, \alpha)}H_\alpha \quad \text{for} \quad {}^\forall H \in \mathfrak{h}, \tag{29}$$

where $X = \{\pi/[2(\alpha, \alpha)]^{\frac{1}{2}}\}(E_\alpha + E_{-\alpha})$. It follows from the straightforward calculation that

$$\begin{aligned}\sigma_\alpha(H) &= H + \sum_{p=0}^{\infty}\frac{1}{(2p+1)!}[\mathrm{ad}(X)]^{2p+1}H \\ &\quad + \sum_{p=0}^{\infty}\frac{1}{(2p+2)!}[\mathrm{ad}(X)]^{2p+2}H \\ &= H - \frac{\alpha(H)}{[2(\alpha, \alpha)]^{\frac{1}{2}}}(\sin\pi)(E_\alpha - E_{-\alpha}) \\ &\quad + \frac{\alpha(H)}{(\alpha, \alpha)}(\cos\pi - 1)H_\alpha \\ &= H - 2\frac{\alpha(H)}{(\alpha, \alpha)}H_\alpha \quad \text{for} \quad {}^\forall H \in \mathfrak{h}.\end{aligned} \tag{30}$$

It is immediately seen that σ_α makes $\mathfrak{h}$ invariant and that

$$\{\sigma_\alpha^*(\lambda)\}(H) = \lambda(H) - 2\frac{(\lambda, \alpha)}{(\alpha, \alpha)}\alpha(H) \quad \text{for} \quad {}^\forall\lambda \in \mathfrak{h}^* \quad \text{and} \quad {}^\forall H \in \mathfrak{h}. \tag{31}$$

[12] N. Jacobson, Ref. 8, Chap. IV.
[13] N. Jacobson, Ref. 8, Chap. III.

[Notice that $\lambda(H_\alpha) = (\lambda, \alpha)$.] It follows from Eq. (31) that

$$\sigma_\alpha^*(\lambda) = \lambda - 2\frac{(\lambda, \alpha)}{(\alpha, \alpha)}\alpha \quad \text{for} \quad {}^\forall\lambda \in \mathfrak{h}^*. \tag{32}$$

Hence, we obtain

$$\sigma_\alpha^* = (\sigma_\alpha^*)^{-1} = S_\alpha. \tag{33}$$

In conclusion, there exists in $I_0(\mathfrak{g})$ the subgroup $\{\sigma_\alpha; \alpha \in \Delta\}$ which is isomorphic to W. This group plays an important role in constructing the reflection operator R.

We have now come to the place to determine R explicitly. In what follows the internal symmetry group G is restricted to a compact simple Lie group. Then $\mathfrak{g}$ is a real simple Lie algebra whose Killing form is negative definite, and Weyl's standard basis $\{H_1, \cdots, H_l, E_\alpha; \alpha \in \Delta\}$ can be adjusted such that

$$(E_\alpha)^t = -E_{-\alpha}, \quad (H_i)^t = H_i. \tag{34}$$

[Strictly speaking, this equation should be expressed as $\rho(E_\alpha)^t = -\rho(E_{-\alpha})$, $\rho(H_i)^t = \rho(H_i)$.] Consequently, condition (20) can be rewritten as

$$R^{-1}E_\alpha R = E_{-\alpha}, \quad R^{-1}H_i R = -H_i. \tag{35}$$

This means that the inner automorphism I associated with R maps an eigenvector corresponding to a root α to the one corresponding to $-\alpha$, and that I not only makes $\mathfrak{h}$ invariant but also changes the sign of every element of $\mathfrak{h}$. Thus, we arrive at the conclusion that the inner automorphism I should be determined as the element of $I_0(\mathfrak{g})$ corresponding to the negative identity -1 of W. If -1 can be expressed as

$$-1 = \prod_{\alpha\in\delta} S_\alpha \tag{36}$$

with the help of some ordered set δ, then the inner automorphism I is represented as

$$\begin{aligned}I &= \prod_{\alpha\in\delta}\sigma_\alpha \\ &= \prod_{\alpha\in\delta}\exp\left[\mathrm{ad}\left\{\frac{\pi}{[2(\alpha, \alpha)]^{\frac{1}{2}}}(E_\alpha + E_{-\alpha})\right\}\right].\end{aligned} \tag{37}$$

In order to evaluate R from this I, we consider below the reduction of an operator of the form $\exp\{\mathrm{ad}(Z)\}$. The transformation $\mathrm{ad}(Z)$ can be expressed as

$$Z_L - Z_R(Z_L: X \to ZX, Z_R: X \to XZ).$$

Since $[Z_L, Z_R] = 0$, it follows that

$$\begin{aligned}\exp\{\mathrm{ad}(Z)\} &= \exp Z_L \exp(-Z_R) \\ &= (\exp Z)_L(\exp Z)_R^{-1} \\ &= \{\exp(-Z)\}_L^{-1}\{\exp(-Z)\}_R.\end{aligned} \tag{38}$$

This equation enables us to derive

$$R = \prod_{\alpha\in\bar{\delta}}\exp\left\{-\frac{\pi}{[2(\alpha, \alpha)]^{\frac{1}{2}}}(E_\alpha + E_{-\alpha})\right\}, \tag{39}$$

where $\bar{\delta}$ has the inverse order against δ.

So far we have identified $\mathfrak{n}$ with $\mathfrak{g}$. It is admitted as long as the referring representation ρ is fixed. Explicit description of the representation used leads us to the following expression:

$$R_\rho = \prod_{\alpha\in\delta} \exp\left[-\frac{\pi}{[2(\alpha,\alpha)]^{\frac{1}{2}}}\{\rho(E_\alpha)+\rho(E_{-\alpha})\}\right]. \quad (40)$$

This is the very equation that has been searched for. It is illustrated in the examples below.

(i) Compact Simple Lie Group $SU(2)(A_1$ type)

Since the corresponding Lie algebra $\mathfrak{g}$ is of rank one, the nonzero roots of $\mathfrak{g}$ consist of α and $-\alpha$ only. Then E_α, $E_{-\alpha}$, and H_α are abbreviated to E_+, E_-, and H, respectively. It follows that

$$[E_+, E_-] = -H, \quad (41)$$

$$[H, E_\pm] = \pm E_\pm. \quad (42)$$

The quantities E_+, E_- correspond to the charge-raising operation and the charge-lowering one, respectively. Since α is so normalized that $(\alpha,\alpha)=1$, the Weyl reflection S_α is equal to -1. Therefore, the reflection operator R is determined as

$$R_\rho = \exp\left[-\frac{\pi}{\sqrt{2}}\{\rho(E_+)+\rho(E_-)\}\right]. \quad (43)$$

Since $\rho(E_\pm)$ and $\rho(H)$ are connected with the isotopic spin operators I_1, I_2, I_3 satisfying the commutation relation $[I_i, I_j] = i\epsilon_{ijk}I_k$ such that

$$\rho(E_\pm) = \mp(I_1 \pm iI_2)/\sqrt{2}, \quad \rho(H) = I_3, \quad (44)$$

we obtain the usual expression

$$R = \exp(i\pi I_2). \quad (45)$$

In this case, R is antisymmetric.

(ii) Compact Simple Lie Group $Sp(6)(C_3$ type)

The fundamental root system[14] $(\alpha_1, \alpha_2, \alpha_3)$ is such that

$$(\alpha_1,\alpha_1) = (\alpha_2,\alpha_2) = 1, \quad (\alpha_3,\alpha_3) = 2,$$
$$(\alpha_1,\alpha_2) = -\tfrac{1}{2}, \quad (\alpha_2,\alpha_3) = -1,$$

and

$$(\alpha_1,\alpha_3) = 0. \quad (46)$$

Since $-1 = (S_1S_3S_2)^3$, we obtain

$$R_\rho = \left[\exp\left\{-\frac{\pi}{\sqrt{2}}[\rho(E_2)+\rho(E_{-2})]\right\} \times \exp\left\{-\frac{\pi}{2}[\rho(E_3)+\rho(E_{-3})]\right\} \times \exp\left\{-\frac{\pi}{\sqrt{2}}[\rho(E_1)+\rho(E_{-1})]\right\}\right]^3. \quad (47)$$

[14] M. Konuma, K. Shima, and M. Wada, Progr. Theoret. Phys. (Kyoto) Suppl. **28**, 1 (1963).

Especially, if we take ρ to be the six-dimensional representation, we get

$$R_6 = \begin{bmatrix} & & & 0 & 0 & 1 \\ & \mathrm{O} & & 0 & 1 & 0 \\ & & & 1 & 0 & 0 \\ 0 & 0 & -1 & & & \\ 0 & -1 & 0 & & \mathrm{O} & \\ -1 & 0 & 0 & & & \end{bmatrix}. \quad (48)$$

In this case also, R is antisymmetric.

(iii) Compact Simple Lie Group $SO(7)(B_3$ type)

In this case,[14] the fundamental root system $(\alpha_1, \alpha_2, \alpha_3)$ is as follows:

$$(\alpha_1,\alpha_1) = (\alpha_2,\alpha_2) = 2, \quad (\alpha_3,\alpha_3) = 1,$$
$$(\alpha_1,\alpha_2) = -\tfrac{1}{2}, \quad (\alpha_2,\alpha_3) = -1,$$

and

$$(\alpha_1,\alpha_3) = 0. \quad (49)$$

Since $-1 = (S_1S_3S_2)^3$ in this case also, we obtain

$$R_\rho = \left[\exp\{-\tfrac{1}{2}\pi[\rho(E_2)+\rho(E_{-2})]\} \times \exp\left\{-\frac{\pi}{\sqrt{2}}[\rho(E_3)+\rho(E_{-3})]\right\} \times \exp\{-\tfrac{1}{2}\pi[\rho(E_1)+\rho(E_{-1})]\}\right]^3. \quad (50)$$

In the eight-dimensional representation, we get

$$R_8 = \begin{bmatrix} & & & & 0 & 0 & 0 & 1 \\ & & \mathrm{O} & & 0 & 0 & -1 & 0 \\ & & & & 0 & 1 & 0 & 0 \\ & & & & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & & & & \\ 0 & 0 & 1 & 0 & & \mathrm{O} & & \\ 0 & -1 & 0 & 0 & & & & \\ 1 & 0 & 0 & 0 & & & & \end{bmatrix}, \quad (51)$$

seeing that R is symmetric in this case.

ACKNOWLEDGMENTS

The authors express their sincere gratitudes to Professor Y. Murai, Professor H. Shimodaira, and Professor Y. Ohnuki for valuable discussions. Thanks are also due to Professor H. Miyazawa for his valuable discussion and encouragement in the early stage of the work.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 3 MARCH 1967

Total Number of Particles and Fock Representation

G. F. DELL'ANTONIO* †
Courant Institute of Mathematical Sciences, New York University, New York, New York

S. DOPLICHER
Istituto di Fisica dell'Universita di Roma, Rome, Italy

(Received 26 July 1966)

It is proved that an average number of particles can be defined only in those representations of the canonical commutation (or anticommutation) relations which are multiples of the Fock–Cook representation.

WITH $\mathfrak{L}$ a real separable pre-Hilbert space, let φ be a representation of the canonical commutation relations (CCR) or of the canonical anticommutation relations (CAR) on $\mathfrak{L}$ by operators on a Hilbert space $\mathfrak{H}_\varphi$. In the case of CCR, we mean by this a Weyl system on $\mathfrak{L}$,[1] i.e., a map $f \to \{U(f), V(f)\}$ from $\mathfrak{L}$ to unitary operators $U(f)$, $V(f)$ on $\mathfrak{H}$ such that, for every $f, g \in \mathfrak{L}$,

(a) $U(f)U(g) = U(g)U(f)$, $V(f)V(g) = V(g)V(f)$, $V(f)V(g) = e^{-i(f,g)}V(g)U(f)$, where (f, g) is the natural scalar product on $\mathfrak{L}$.

(b) With t a real variable $U(tf)$ and $V(tf)$ are weakly continuous in t at the origin for all $f \in \mathfrak{L}$.

For $f \in \mathfrak{L}$, $a_\varphi(f)$ denotes the corresponding closed "destruction operator" and $N_\varphi(f)$ the number operator. One has[2]

$$N_\varphi(f) \smile N_\varphi(g)$$
$$a_\varphi(f)N_\varphi(f) \subset (N_\varphi(f) + 1)a_\varphi(f) \qquad (1)$$

for all $f, g \in \mathfrak{L}$, since φ is equivalent to a direct sum of Schrödinger representations when restricted to a finite-dimensional subspace of $\mathfrak{L}$.[1]

The Fock–Cook representation [i.e., the one for which there exists a cyclic vector $\Omega \in \mathfrak{H}$ such that $a_\varphi(f)\Omega = 0$, $\forall f \in \mathfrak{L}$] is denoted by φ_0.

In this note we want to prove some statements which are slight generalizations of results by Gårding and Wightman (see Ref. 3, where a characterization is given of all representations of CCR and CAR; also Ref. 4 for a direct proof in the irreducible case). These authors show that the Fock–Cook representation is the only one for which there exists a positive self-adjoint operator N_φ with integer eigenvalues such that its spectral projections $E_n(N_\varphi)$ are given by

$$E_n(N_\varphi) = \sum_{i=1}^{\infty} \sum_{n_i=1}^{\infty} \delta_{n, \sum_i n_i} \prod_i E_{n_i}[N_\varphi(f_i)]$$

with $\{f_i\}$ an arbitrary orthonormal complete set in $\mathfrak{L}$.

We state the main proposition as follows:

Theorem[5]: Let $\mathfrak{L}$ be the linear span of vectors $\{f_k, k = 1, 2, \cdots\}$ on which a scalar product is defined by $(f_k, f_j) = \delta_{kj}$ (that gives to $\mathfrak{L}$ a pre-Hilbert space structure). The following conditions are equivalent:

(a) There exists a cardinal number n such that $\varphi = n\varphi_0$.[6]

(b) The linear variety $\mathfrak{D}^0 \subset \mathfrak{H}_\varphi$ defined by[7]

$$\mathfrak{D}^0 \equiv \left\{\phi;\ \phi \in \bigcap_k D_{N_\varphi^{\frac{1}{2}}(f_k)},\ \sum_{k=1}^{\infty} \langle \phi N_\varphi(f_k)\phi\rangle < \infty\right\} \qquad (2)$$

is dense in $\mathfrak{H}_\varphi$.[8,9]

Proof: (a) $\Rightarrow$ (b) is well known.[4] To prove (b) $\Rightarrow$ (a) we need first a self-adjoint "number operator." It is

---

* On leave from the Istituto di Fisica Teorica, Universita di Napoli.

† Sponsored by the Sloan Foundation Program.

[1] J. Von Neumann, Math. Ann. **104**, 570 (1932).

[2] $A \smile B$ means that the spectral projections of the two operators commute; $A \subset B$ means that B is defined, and coincides with A, on D_A, the domain of A.

[3] L. Gårding and A. Wightman, Proc. Natl. Acad. Sci. U.S. **40**, 622 (1954).

[4] A. S. Wightman and S. Schweber, Phys. Rev. **98**, 812 (1955).

[5] After completion of this work; we have been informed by J. Chaiken (private communication and Cornell University preprint) that he has obtained a similar result for the CCR, and in fact given several other characterizations of the Fock–Cook representation, using the Weyl form of the CCR (and avoiding therefore the use of unbounded "destruction operators").

[6] $\varphi = n\varphi_0$ means that there exist Hilbert spaces $\mathfrak{H}_1$ and $\mathfrak{H}_2$, $\mathfrak{H}_2$ of dimension n, such that $\mathfrak{H} = \mathfrak{H}_1 \otimes \mathfrak{H}_2$ and $a_\varphi(f_k) = a(f_k) \otimes 1$.

[7] D_T is the domain of the operator T. Here and in the following, whenever $\phi, \psi \in D_{T^{\frac{1}{2}}}$, we write $\langle \psi T\phi\rangle$ for $\langle T^{\frac{1}{2}}\psi, T^{\frac{1}{2}}\phi\rangle$.

[8] One could equivalently require $\sum_{m=1}^{\infty}\langle \phi N_\varphi(f_m)\phi\rangle < \infty$ for one vector $\phi \in \bigcap_k D_{N^{\frac{1}{2}}(f_k)}$, cyclic in $\mathfrak{H}$ relative to the algebra generated by the partially isometric operators U_k, U_k^* where U_k is the extension to $\mathfrak{H}$ of the operators $[N_\varphi(f_k) + 1]^{-\frac{1}{2}}\, a_\varphi(f_k)$ defined on $D_{a_\varphi(f_k)}$.

[9] In the case of CAR there are obvious simplifications in the statement of the theorem (and in its proof), since for all $f \in \mathbf{L}$, $N_\varphi(f)$ is a bounded operator.

helpful to introduce the following:

Definitions: $N \equiv$ Friedrichs extension[10] of the positive bilinear form

$$[\Phi, \Psi] = \sum_{m=1}^{\infty} \langle \Phi N_\varphi(f_m)\Psi\rangle + \langle \Phi, \Psi\rangle.$$

$\mathcal{H}_0$ is the completion of $\mathfrak{D}^0$ in the norm defined by $[.,.]$. One has $\mathfrak{D}^0 \subseteq \mathcal{H}_0 \subseteq \mathcal{H}_\varphi$. N_k' is Friedrichs extension of the positive bilinear closable form $[\Phi, \Psi]_k = [\Phi, \Psi] - \langle \Phi N_\varphi(f_k)\Psi\rangle$.

We now want to prove a weak form of the commutation relation

$$Na_\varphi(f) \subset a_\varphi(f)(N-1). \qquad (**)$$

Observe that with $N_k = N_\varphi(f_k)$

$$D_{a_\varphi}(f_k) \supseteq D_{N^{\frac{1}{2}}};\ D_{N_k^{\frac{1}{2}}} \supseteq D_{N^{\frac{1}{2}}}, \qquad (3)$$

$$\Phi \in D_{N^{\frac{1}{2}}} \Rightarrow \langle \Phi N\Phi\rangle = \langle \Phi N_k\Phi\rangle + \langle \Phi N_k'\Phi\rangle \qquad (4)$$

are easy consequences of the definition of Friedrichs extension (see Appendix B) and of the inequalities $N_k \geq 0$, $k = 1, 2, \cdots$.

Remark next that, for $\Phi, \Psi \in \mathfrak{D}^0$, and real λ

$$\sum_m \langle e^{i\lambda N_k}\Phi, N_m e^{i\lambda N_k}\Psi\rangle = \sum_m \langle \Phi N_m \Psi\rangle$$

from which follows (see Appendix C)

$$N_k \backsim N. \qquad (5)$$

Let $E(\lambda)$, $E_k'(\lambda)$, $E_k(\lambda)$ be the spectral projections of N, N_k', N_k, respectively. Combining (4) and (5) one obtains

$$E_k(\lambda) \geq E(\lambda), \quad E_k'(\lambda) \geq E(\lambda). \qquad (6)$$

Define

$$\mathfrak{D} \equiv \bigcup_\lambda E(\lambda)\mathcal{H}_\varphi \subseteq D_N.$$

From (6) one concludes $\mathfrak{D} \subseteq D_{N_k}$, $\mathfrak{D} \subseteq D_{N_k}'$. The domain $\mathfrak{D}$ has the following properties:

Lemma 1:

$$a_\varphi(f_k)\mathfrak{D} \subseteq D_{N^{\frac{1}{2}}}; \quad N_k^{\frac{1}{2}}\mathfrak{D} \subseteq D_{N^{\frac{1}{2}}}.$$

We give the proof of the first relation. The second is proved along similar lines.

Let $\Phi \in \mathfrak{D}$. Then $E(\lambda)\Phi = \Phi$ for some real positive λ and, in view of (6), $E_k(\lambda)\Phi = \Phi$, $E_k'(\lambda)\Phi = \Phi$. Since $\mathfrak{D} \subseteq \mathcal{H}_0$, there exists a sequence $\Psi_n \in E(\lambda)\mathfrak{D}^0$ such that

$$|||\Phi - \Psi_n||| \xrightarrow[n\to\infty]{} 0,$$

where

$$|||\Psi|||^2 = [\Psi, \Psi].$$

[10] K. O. Friedrichs, Math. Ann. **109**, 465 (1934).

With $\Psi_{n,m} = \Psi_n - \Psi_m$,

$$|||a_\varphi(f_k)\Psi_{n,m}|||^2 = \langle \Psi_{n,m} N_k \Psi_{n,m}\rangle + \sum_{p=1}^{\infty} \langle a_\varphi(f_k)\Psi_{n,m}, N_p a_\varphi(f_k)\Psi_{n,m}\rangle$$
$$= \sum_{p=1}^{\infty} \langle \Psi_{n,m}, N_p N_k \Psi_{n,m}\rangle \leq \lambda\, |||\Psi_{n,m}|||^2.$$

Use has been made of the commutation relations between $a_\varphi(f_k)$ and N_p. There exists therefore a $\Psi \in \mathcal{H}_0$ such that $a_\varphi(f_k)\Psi_n \to \Psi$. Since $a_\varphi(f_k)$ is closed, this implies $\Psi = a_\varphi(f_k)\Phi$, i.e.,

$$a_\varphi(f_k)\Psi_n \to a_\varphi(f_k)\Phi. \qquad (7)$$

$\mathcal{H}_0$ is closed and $\mathcal{H}_0 = D_{N^{\frac{1}{2}}}$ (see Appendix B). Therefore

$$a_\varphi(f_k)\Phi \subseteq D_{N^{\frac{1}{2}}}. \qquad (8)$$

Q.E.D.

We are now in position to prove the following weak form of (**):

Lemma 2: If $\Phi \in \mathfrak{D}$, then

$$\langle a_\varphi(f_k)\Phi, Na_\varphi(f_k)\Phi\rangle = \langle N_k\Phi, (N-1)\Phi\rangle.$$

Proof: From Lemma 1 and Eq. (4)

$$\langle a_\varphi(f_k)\Phi, Na_\varphi(f_k)\Phi\rangle = \langle a_\varphi(f_k)\Phi, N_k a_\varphi(f_k)\Phi\rangle + \langle a_\varphi(f_k)\Phi, N_k' a_\varphi(f_k)\Phi\rangle$$
$$= \langle N_k\Phi, (N_k-1)\Phi\rangle + \langle N_k\Phi, N_k'\Phi\rangle$$
$$= \langle N_k\Phi, (N-1)\Phi\rangle.$$

Proof of the theorem. Let $\bar\lambda \geq 0$ be the greatest lower bound of the spectrum of N. Let $\Psi_0 \in \mathcal{H}_\varphi$ be such that $E(\bar\lambda + \frac{1}{2})\Psi_0 = \Psi_0$, $\|\Psi_0\| = 1$. Then

$$\bar\lambda\, \|a_\varphi(f_k)\Psi_0\|^2 \leq \langle a_\varphi(f_k)\Psi_0, Na_\varphi(f_k)\Psi_0\rangle \leq (\bar\lambda - \tfrac{1}{2})\, \|a_\varphi(f_k)\Psi_0\|^2, \qquad (9)$$

where the first inequality follows from the spectral decomposition

$$N^{\frac{1}{2}} = \int_{\bar\lambda^{\frac{1}{2}}}^{\infty} \lambda^{\frac{1}{2}} dE(\lambda),$$

and the second from Lemma 2. Inequality (9) implies

$$a_\varphi(f_k)\Psi_0 = 0, \quad k = 1, 2, \cdots. \qquad (10)$$

Let E_{Ψ_0} be the orthogonal projection on the cyclic subspace generated by application of the operators $a_\varphi(f_k)$ to the vector Ψ_0. From Eq. (10) one sees that the subrepresentation $\varphi_{\Psi_0} = E_{\Psi_0}^{\varphi} \circ \varphi$ is equivalent to the Fock–Cook representation in the subspace $E_{\Psi_0}^{\varphi}\mathcal{H}_\varphi$.

The subrepresentation $\varphi'_{\Psi_0} = (1 - E_{\Psi_0}^{\varphi}) \circ \varphi$ has the same properties as φ, relative to the subspace $(1 - E_{\Psi_0}^{\varphi})\mathcal{H}_\varphi$. One can therefore repeat for φ'_{Ψ_0} the

arguments which were given above for φ and prove the existence of a vector $\Psi_1 \in (1 - E^{\varphi}_{\Psi_0})\mathcal{H}_{\varphi}$ such that $E_{\Psi_1}{}^{\varphi'\Psi_0} \circ \varphi'_{\Psi_0}$ is equivalent to the Fock–Cook representation in the subspace $E_{\Psi_0}{}^{\varphi'\Psi_0}(1 - E^{\varphi}_{\Psi_0})\mathcal{H}_{\varphi}$. The proof of the theorem is now achieved by complete induction. To this theorem we add.

Corollary I: Let $\mathfrak{L}$ be a real separable pre-Hilbert space, $f \to a_{\varphi}(f)$ a representation of the canonical anticommutation relations on $\mathfrak{L}$ by operators on the Hilbert space $\mathcal{H}_{\varphi}$. The following conditions are equivalent:

(a) There exists a cardinal number n such that $\varphi = n\varphi_0$.

(b) There exists in $\mathfrak{L}$ an orthonormal basis f_k, $k = 1, 2, \cdots$ such that the linear variety $\mathfrak{D}^0 \in \mathcal{H}_{\varphi}$ defined by Eq. (2) is dense in $\mathcal{H}_{\varphi}$.

Proof: From the relations

$$a_{\varphi}(f)a^*_{\varphi}(g) + a^*_{\varphi}(g)a_{\varphi}(f) = (f, g)$$

it follows

$$\|a_{\varphi}(f)\| = \|f\|_{\mathfrak{L}} \tag{11}$$

for all nonzero representation φ. We may therefore assume that $\mathfrak{L}$ is complete.

The implication (a) $\Rightarrow$ (b) is again well known.

To prove (b) $\Rightarrow$ (a) notice first that, if $\tilde{\mathfrak{L}}$ is the linear span of $\{f_k\}$, $k = 1, 2, \cdots$, $\varphi\,|_{\tilde{\mathfrak{L}}}$ is a multiple of the Fock–Cook representation.

Let Ω be a vector in $\mathcal{H}$ which is a "vacuum" for $\varphi|_{\tilde{\mathfrak{L}}}$; we have $a_{\varphi}(f_k)\Omega = 0$, $k = 1, 2, \cdots$ and, from (11)

$$a_{\varphi}(f)\Omega = 0 \;\forall f \in \mathfrak{L}. \tag{12}$$

Therefore Ω is a "vacuum" for φ. Q.E.D.

Corollary II: Let $\mathfrak{L}$ be a real separable topological vector space which is a pre-Hilbert space for a weaker topology. Let $\mathfrak{L} \ni f \to a(f)$ be a representation of the canonical commutation relations on $\mathfrak{L}$ by operators on the Hilbert space $\mathcal{H}_{\varphi}$, such that there exists in $\mathcal{H}_{\varphi}$ a linear dense set D on which all the operators $a_{\varphi}(f)$, $a^*_{\varphi}(f)$, $f \in \mathfrak{L}$ are defined. Assume moreover that

(1) $\overline{a_{\varphi}(f)\,|_D} = a_{\varphi}(f)$; $\overline{a^*_{\varphi}(f)\,|_D} = a^*_{\varphi}(f)$.

(2) $\psi \in D$ implies that the maps

$$f \to a_{\varphi}(f)\psi, \quad f \to a^*_{\varphi}(f)\psi$$

are continuous from $\mathfrak{L}$ to $\mathcal{H}_{\varphi}$. Then the following conditions are equivalent:

(a) There exists a cardinal number n such that $\varphi = n\varphi_0$.

(b) The positive form

$$\phi \to [\phi, \phi] = \sum_k \langle \phi N_{\varphi}(f_k)\phi\rangle + \langle \phi, \phi\rangle$$

is densely defined in $\mathcal{H}$ and its closure does not depend on the orthonormal complete basis $\{f_k\}$.

(c) There exists in $\mathfrak{L}$ a total set $\{f_k\}$ which is orthonormal for the pre-Hilbert structure of $\mathfrak{L}$, such that the linear variety defined in Eq. (2) is dense in $\mathcal{H}_{\varphi}$.

Proof: a $\Rightarrow$ b is obvious, and b $\Rightarrow$ a follows immediately from the proof of the theorem and the remark that N is independent of the basis chosen for its definition.

a $\Rightarrow$ c is again well known. To be more precise (and in order that a $\Rightarrow$ c be not "logically trivial"), if $\varphi = n\varphi_0$ there exists a dense set D with the properties required in the statement of the corollary. One can choose for D the set $D_1^{n\varphi_0}$ obtained by applying to all vacua in $\mathcal{H}_{n\varphi_0}$ arbitrary polynomials in the creation operator $a^*_{n\varphi_0}(\mathfrak{L})$, $\forall f \in \mathfrak{L}$. Property (2) can be verified by computing the norm of $a^*_{n\varphi_0}(f)\psi$ with $f \in \mathfrak{L}$ and $\psi \in D_1^{n\varphi_0}$. Property (1) is true for $D_1^{n\varphi_0}$ if it is true for $D_1^{\varphi_0}$; for the latter case see, e.g., (Ref. 11, p. 48). We have now to prove c $\Rightarrow$ a. With the notations used in Corollary I, the theorem states that $\varphi\,|_{\tilde{\mathfrak{L}}}$ is a multiple of the Fock–Cook representation. If Ω is a vacuum for $\varphi|_{\tilde{\mathfrak{L}}}$, we have

$$\langle a_{\varphi}(f)\Phi, \Omega\rangle = 0 \tag{13}$$

for all $\Phi \in D$, $f \in \tilde{\mathfrak{L}}$. The continuity condition (2) implies that (13) holds for all $f \in \mathfrak{L}$. Therefore

$$(a^*_{\varphi}(f)|_D)^*\Omega = 0, \quad \forall f \in \mathfrak{L}$$

and condition (1) gives

$$\Omega \in D_{a_{\varphi}}(f), \quad a_{\varphi}(f)\Omega = 0, \quad \forall f \in \mathfrak{L}. \tag{14}$$

This proves the corollary.

In conclusion, we mention that a situation of interest in physics and to which Corollary II applies is, e.g., $\mathfrak{L} \equiv S^3$, the space of infinitely many times differentiable functions on R^3, decreasing at infinity faster than any inverse polynomial, provided with the Schwartz topology.

ACKNOWLEDGMENTS

It is a pleasure to thank D. Kastler for a very valuable discussion, and J. Chaiken and D. Ruelle for discussions on this and related subjects.

[11] R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society Publications, Providence, R.I., 1966).

APPENDIX A

We want to show that the bilinear form

$$[\Phi,\Psi] = \langle\Phi,\Psi\rangle + \sum_{m=1}^{\infty}\langle\Phi, N_m\Psi\rangle$$

is closable. Notation is

$$[\Psi,\Psi] = |||\Psi|||^2, \quad \langle\Psi,\Psi\rangle = \|\Psi\|^2.$$

We must prove that, if $|||\Phi_n - \Phi_m||| \to 0$ when $n, m \to \infty$, there exists an element $\Phi \in \mathcal{H}_\varphi$ such that $|||\Phi_n - \Phi||| \to 0$ for $n \to \infty$.

Since $\mathcal{H}_\varphi$ if complete and $|||\Psi||| \geq \|\Psi\|$, there exist $\Phi \in \mathcal{H}_\varphi$ and n_0 such that $\|\Phi_n - \Phi\| \leq 1$ for $n > n_0$.

Since $|||\Phi_n - \Phi_m||| \to 0$, there exist n', m' such that if $n > n'$, and $m > m'$ one has, for all M,

$$\sum_{k=1}^{M}\langle N_k^{\frac{1}{2}}(\Phi_n - \Phi_m), N_k^{\frac{1}{2}}(\Phi_n - \Phi_m)\rangle < 1. \quad \text{(A1)}$$

The operators $N_k^{\frac{1}{2}}$, $k = 1, 2, \cdots, M$ are closed, and $\Phi_n \to \Phi$ in $\mathcal{H}_\varphi$; we then have, for $n > n'$,

$$N_k^{\frac{1}{2}}(\Phi_n - \Phi_m) \xrightarrow[m\to\infty]{} N_k^{\frac{1}{2}}(\Phi_n - \Phi), \quad k = 1, 2, \cdots, M.$$

This shows that $\Phi \in D_{N_k^{\frac{1}{2}}}$, $k = 1, 2, \cdots, M$.

We also have

$$\sum_{k=1}^{M}\langle N_k^{\frac{1}{2}}(\Phi_n - \Phi), N_k^{\frac{1}{2}}(\Phi_n - \Phi)\rangle < 1 \quad \text{for} \quad n > n'.$$

This holds for all M and each term in the sum is positive. Therefore

$$|||\Phi_n - \Phi||| < 1 \quad \text{for} \quad n > n'. \quad \text{(A2)}$$

From the linearity of $\mathcal{H}_0$ it follows $\Phi \in \mathcal{H}_0$. Moreover, since (A2) holds for sequences of vectors Φ_n not necessarily normalized to 1, we have

$$|||\Phi_n - \Phi||| \to 0 \quad \text{when} \quad n \to \infty. \quad \text{(A3)}$$

Q.E.D.

APPENDIX B. A CHARACTERIZATION OF THE FRIEDRICHS EXTENSION

Let $\mathcal{H}$ be a Hilbert space and D_0 a linear manifold dense in $\mathcal{H}$, on which a bilinear closable form $[\Phi,\Psi]$ is defined with the property $[\Phi,\Phi] \geq \|\Phi\|^2$.

We denote by S any self-adjoint operator, bounded below by 1, which induces the form $[\Phi,\Psi]$

$$\langle\Phi, S\Psi\rangle = [\Phi,\Psi], \quad D_{S^{\frac{1}{2}}} \supseteq D_0.$$

S_0 is the Friedrichs extension of $[.,.]$ and $\mathcal{H}_0$ denotes the closure of D_0 in the norm $|||\Phi||| = [\Phi,\Phi]^{\frac{1}{2}}$. We have $D_{S_0} \subseteq \mathcal{H}_0 \subseteq \mathcal{H}$, $D_{S^{\frac{1}{2}}} \supseteq D_S$.

Define $R_S = S^{\frac{1}{2}}/D_0$. Evidently $D_{R_S^{**}} = \mathcal{H}_0$.

Theorem: The symmetric transformation R_S is essentially self-adjoint if and only if $S = S_0$.

Proof: (a) $R_{S_0}^{**}$ is self-adjoint. In fact, we have $D_{R_{S_0}^{**}} = \mathcal{H}_0 \supseteq D_{S_0}$, therefore $R_{S_0}^{**} \supseteq S_0^{\frac{1}{2}}/D_{S_0}$, and, taking closures, $R_{S_0}^{**} = S_0^{\frac{1}{2}}$. Q.E.D.

(b) Let R_S be essentially self-adjoint. This implies, by closure, $R_S^{**} = S^{\frac{1}{2}}$. Therefore $D_{S^{\frac{1}{2}}} = \mathcal{H}_0$ and $D_S \subseteq \mathcal{H}_0 = D_S^{\frac{1}{2}}$. Among the operators S, only S_0 satisfies this inequality. Q.E.D.

Corollary: $D_{S^{\frac{1}{2}}} = \mathcal{H}_0$ if and only if $S = S_0$.

APPENDIX C

Let $[\Phi,\Psi]$ be a positive bilinear closable form defined for vectors from a linear dense subspace D of a Hilbert space $\mathcal{H}$. Let B be a unitary operator on $\mathcal{H}$ such that

(a) $VD = D$.

(b) $[V\Phi, V\Psi] = [\Phi,\Psi]$ for all $\Phi,\Psi \in D$.

Then, if S is the Friedrichs extension of $[\Phi,\Psi]$, one has $VS = SV$ on D_S.

Proof: Let B be defined on $\mathcal{H}_0$ (the closure of D in the norm $[\Phi,\Phi]^{\frac{1}{2}}$) by $\langle\Phi,\Psi\rangle = [\Phi, B\Psi]$, $\Phi,\Psi \in \mathcal{H}_0$.

Our assumptions (a), (b) lead to

(1) $\langle\Phi,\Psi\rangle = \langle V\Phi, V\Psi\rangle = [V\Phi, BV\Psi]$,

(2) $\langle\Phi,\Psi\rangle = [\Phi, B\Psi] = [V\Phi, VB\Psi]$.

Comparing (1) and (2) we conclude $BV = VB$ on $\mathcal{H}_0$. Since B and V are bounded, it follows $BV = VB$ on $\mathcal{H}$, by continuity. Therefore, making use of $VD_S \subset D_S$, $BS/D_S = 1/D_S$, $SB = 1$, $0 = S(BV - VB)S = VS - SV$ on D_S. Q.E.D.