JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 5, NUMBER 9 September 1964

# Exact Solution of the Schrödinger Equation for Inverse Fourth-Power Potential

RICHARD M. SPECTOR
*Department of Physics and Astronomy, University of Rochester, Rochester, New York*
(Received 6 March 1964)

The solution of the Schrödinger equation for the potential $V(r) = V/(\mu r)^4$ and all angular momenta is given in terms of known functions, namely, the solutions to the modified Mathieu equation with complex parameters and complex argument. Scattering solutions for both attractive and repulsive potentials are given and in the latter case the explicit expression for the $S$ matrix is given.

## I. INTRODUCTION

HISTORICALLY,[1-3] interest in nonrelativistic potential theory has always been centered on solving the Schrödinger equation for potentials less singular than the inverse square. This is because the boundary conditions at the origin are easy to satisfy (the Jost functions are, in general, finite and nonzero), and there are discrete bound states with a lowest one: the ground state. However, Case,[4] some years ago, showed that potentials of the form $r^{-n}$ $n \geq 2$ do not cause as much trouble as might be expected and, in fact, do not have a divergent behavior at the origin for attractive potentials. With the adjustment of a single parameter, it is possible to form discrete bound states but there is no ground state.

Case pointed out two possible useful situations for such singular potentials: (1) the case where such a potential is not valid all the way to $r = 0$ but is joined to some less singular potential as $r$ becomes small in much the same way as the Coulomb potential is not really valid down to[5] $r = 0$; and (2) for repulsive singular potentials the study of scattering is mathematically well-defined and useful.

More recently, interest in potentials of the form $A/r^4 + B/r^3$, $A > 0$, have been investigated by Predazzi and Regge[6] in an attempt to shed light on the analyticity of the $S$ matrix in the complex $l$ plane for singular potentials. Their argument is that physical interactions seem more singular at close range than the potentials which one normally deals with. Therefore, if we expect the physical $S$ matrix to be meromorphic in $l$ we must study singular potential scattering to look for clues as to the proper behavior. They show that, indeed, making the potential singular removes the usual difficulties one has with the $S$ matrix in the $l$ plane that are found with the Coulomb potential, for instance, in the left half-plane. The authors are not able to solve the Schrödinger equation except for zero energy.

Other workers[7,8] have investigated $r^{-4}$ potentials in the Bethe–Salpeter equation so as to illuminate the behavior of this equation. However, in this case the solutions for $r^{-4}$ reveal a very similar behavior to the inverse-square potential for the Schrödinger equation. This is not surprising since the

[1] A. Bhattacharjie and E. C. G. Sudarshan, Nuovo Cimento **25**, 864 (1962).
[2] A. K. Bose, Phys. Letters **7**, 245 (1963).
[3] V. Bargmann, Rev. Mod. Phys. **21**, 488 (1949).
[4] K. M. Case, Phys. Rev. **80**, 797 (1950).
[5] Moreover, the Coulomb potential behaves very badly at infinity in terms of the standard phase shift analysis. The only reason we use it is because we must (it's physical).
[6] E. Predazzi and T. Regge, Nuovo Cimento **24**, 518 (1962).
[7] A. Bastai, L. Bertocchi, S. Fubini, G. Furlan, and M. Tonin, Nuovo Cimento **30**, 1512 (1963).
[8] A. Bastai, L. Bertocchi, G. Furlan, and M. Tonin, Nuovo Cimento **30**, 1532 (1963).



Copyright © 1964 by the American Institute of Physics

Bethe–Salpeter equation is in a sense a squared equation (in energy), while the Schrödinger equation is linear.

Though Case (see also Morse and Feshbach[9]) demonstrated the exact solutions to the Schrödinger equation for $r^{-2}$, he was able only to give the general behavior at the origin for more singular potentials. It is our intention here to give the exact solutions, for attractive and repulsive potentials, for the $r^{-4}$ potential in the nonrelativistic case for all angular momenta. The solutions turn out to be in the form of solutions of the modified Mathieu equation of complex argument and so are, in general, rather complicated in form. However, the final result for the $S$ matrix is remarkably simple, formally at least, and is amenable to a machine calculation of the scattering.

## II. GENERAL FORMULATION

We seek a solution to the separated radial equation

$$\varphi''(r) + \left[k^2 + \frac{V}{(\mu r)^4} - \frac{l(l+1)}{r^2}\right]\varphi(r) = 0 \qquad (1)$$

$V > 0$ attractive
$V < 0$ repulsive

where units are such that $E = k^2$, $V(\mathbf{r}) = V/(\mu\,|\mathbf{r}|)^4$ with $V$ having the dimension of energy and $\mu > 0$ the dimension of inverse length. It is possible (and in fact was so done) to find the solutions of (1) by using the methods of Refs. 1 and 2 by starting with the algebraic form of the modified Mathieu equation. It is easier and simpler to follow the reverse procedure and start with (1), making the correct transformations as we go. Let us put $\varphi(r) = r^{\frac{1}{2}}\psi(r)$ to get

$$\psi''(r) + \frac{1}{r}\psi'(r) + \left[k^2 + \frac{V}{(\mu r)^4} - \frac{l(l+1)+\frac{1}{4}}{r^2}\right]\psi(r) = 0,$$

and then putting $x = \lambda r$ we have

$$\psi''(x) + \frac{1}{x}\psi(x) + \left[\left(\frac{k}{\lambda}\right)^2 + \frac{V\lambda^2}{(\mu x)^4} - \frac{(l+\frac{1}{2})^2}{x^2}\right]\psi(x) = 0.$$

Making the further substitution

$$x = e^{-z} \qquad 0 \le |x| \le 1, \qquad (2)$$

$$x = e^{z} \qquad 1 \le |x| \le \infty, \qquad (3)$$

we arrive at

$$\psi''(z) + [(k/\lambda)^2 e^{\pm 2z} + (V\lambda^2/\mu^4)e^{\mp 2z} - (l+\tfrac{1}{2})^2]\psi(z) = 0, \qquad (4)$$

where we always have Re $(z) \ge 0$. Let us set

$$k^2/\lambda^2 = V\lambda^2/\mu^4 = q \quad \text{and} \quad a = (l+\tfrac{1}{2})^2$$

to get

$$\psi''(z) - [a - 2q\cosh 2z]\psi(z) = 0, \qquad (5)$$

which is the canonical form of the modified Mathieu equation, i.e., the form assumed by the Mathieu equation

$$\psi''(y) + [a + 2q\cos 2y]\psi(y) = 0$$

when $y = iz$. In what follows we draw mainly on the material in McLachlan[10] (hereafter referred to as M), whose notation we use, and Meixner and Schäfke[11] (hereafter MS) where $a = \lambda$ and $q = h^2$.

The standard treatment of the ordinary Mathieu equation is concerned with periodic and nonperiodic solutions and with regions of stable and unstable solutions for various choices of the pair $(a, q)$. However, it is obvious that these concerns are no longer of interest when we deal with the modified equation (5) because cosh $(2z)$ is not periodic for real $z$. Instead we are concerned with asymptotic properties of the solutions of (5).

In the relationship

$$k^2/\lambda_a^2 = V\lambda_a^2/\mu^4,$$

if we consider positive energy solutions for an attractive potential we must take

$$\lambda_a = \lambda = \mu k^{\frac{1}{2}}/V^{\frac{1}{4}} \quad \text{and} \quad q_a = q = kV^{\frac{1}{2}}/\mu^2 > 0$$

so that $\lambda_a$ and $q_a$ are real and $z = \pm\ln \lambda r$ is real. Though this is less interesting than the case of a repulsive potential, it is easier to deal with, since if $V < 0$ we must take

$$\lambda_r = \frac{1}{\sqrt{2}}(1-i)\frac{\mu k^{\frac{1}{2}}}{(-V)^{\frac{1}{4}}} = \frac{1}{\sqrt{2}}(1-i)\lambda,$$

$$q_r = i\frac{k(-V)^{\frac{1}{2}}}{\mu^2} = iq,$$

and $\lambda_r$ and $q_r$ are complex as is $z$.

[9] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), mainly Chaps. 5 and 11.

[10] N. W. McLachlan, *Theory and Applications of Mathieu Functions* (Oxford University Press, Oxford, 1947). The reader is cautioned that various passages in this book are incorrect, for example, Secs. 6.20 and 12.30 do not state the regions of validity of the expressions contained therein (these expressions are completely invalid for certain regions). Similarly parts 12 and 15 of the Additional Results are completely wrong. In all cases the correct expressions are found in Ref. 11.

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

### III. SOLUTIONS FOR ATTRACTIVE POTENTIAL

We consider first the case when $\lambda_a$ and $q_a$ are real for then it is easy to write down the solutions of (5) which are well known. Transforming the notation of MS to our own, we write the two independent solutions of (5) as[12]

$$M_\nu^{(3)(4)}(z, q^{\frac{1}{2}}) = [me_\nu(0, q)]^{-1} \cdot \sum_{r=-\infty}^{\infty} (-1)^r C_{2r}^\nu(q) H_{(\nu+2r)}^{(3)(4)}(2q^{\frac{1}{2}} \cosh z) \quad (6)$$

where $H_{\nu+2r}^{(3)(4)}$ are the Hankel functions of the first and second kind, respectively, and in the more conventional notation have the property[9]

$$H_\sigma^{(3)}(z) = H_\sigma^{(1)}(z) \to \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} e^{i(z-\frac{1}{2}\sigma\pi-\frac{1}{4}\pi)}$$

$$H_\sigma^{(4)}(z) = H_\sigma^{(2)}(z) \to \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} e^{-i(z-\frac{1}{2}\sigma\pi-\frac{1}{4}\pi)}$$

$$\text{as} \quad z \to +\infty. \quad (7)$$

In (6) $\nu$ is a parameter which is a very complicated function of $a$ and $q$, $\nu = \nu(a, q)$, discussed in more detail in the Appendix. The coefficients $C_{2r}^\nu(q)$ satisfy the recursion relations

$$[a - (\nu + 2r)^2]C_{2r}^\nu - q(C_{2r-2}^\nu + C_{2r+2}^\nu) = 0 \quad (8)$$

which unfortunately are of the three-term variety. We discuss (8) in the Appendix. The constant in (6) is

$$me_\nu(0, q) = \sum_{r=-\infty}^{\infty} C_{2r}^\nu(q). \quad (9)$$

It is obvious that having the two Hankel functions appearing in (6) means that we can form other solutions with different asymptotic behavior, for instance, behaving as the sine or cosine as do $J_\sigma(z)$ and $J_{-\sigma}(z)$ [or $N_\sigma(z)$ if $\sigma$ is an integer] the Bessel functions. We are obviously interested in the forms (7) so we will not discuss other possibilities. The recursion relations (8) are known[10,11] to give a converging series for the $C_{2r}^\nu$'s for the correct $\nu(a, q)$ (which always exists) and the series (6) can be proved convergent for $|\cosh z| \geq 1$, but uniformly convergent only when $|\cosh z| > 1$. Here $z$ takes on any complex value.

We write the general solution of (1) when the potential is attractive and energy positive as

$$\varphi(r) = Ar^{\frac{1}{2}}M_\nu^{(3)}(\pm \ln \lambda r, q^{\frac{1}{2}}) + Br^{\frac{1}{2}}M_\nu^{(4)}(\pm \ln \lambda r, q^{\frac{1}{2}}) \quad (10)$$

where the plus sign is taken if $|\lambda r| \geq 1$ and the negative sign if $|\lambda r| \leq 1$. In (10) $A$ and $B$ are arbitrary constants. It is easy to see that $|\cosh z| = 1$ occurs when $|\lambda r| = 1$, so that though the solution is continuous across $r = |\lambda|^{-1}$ the derivative is not [in fact it does not exist in the form (10)]. Here we make use of one of the interesting properties of the solutions of the modified Mathieu equation, namely, that there is a plurality of forms for the solutions. It can be shown that an equally suitable set of solutions to (5) are given by

$$Me_{\pm\nu}(z, q) = \sum_{r=-\infty}^{\infty} C_{2r}^\nu(q) e^{\pm(2r+\nu)z} \quad (11)$$

with the same $C_{2r}^\nu$! This solution converges uniformly for all finite $z$ in the complex plane and so can be used to join the two regions $r < |\lambda|^{-1}$ and $r > |\lambda|^{-1}$. In the next section we do this type of joining in detail to enable us to get the $S$ matrix.

The behavior of $\varphi(r)$ at the origin and infinity are both obtainable from the infinite asymptotic behavior alone of the $M_\nu^{(3)(4)}$. This is because we have

$$M_\nu(2q^{\frac{1}{2}} \cosh z) = M_\nu(q^{\frac{1}{2}}(\lambda r + 1/\lambda r)) = M_\nu(kr + V^{\frac{1}{2}}/\mu r)$$

and the argument of $M_\nu$ goes to infinity as $r \to 0$ and $r \to \infty$. Using the results proved in MS we have

$$\varphi(r) \to A\left(\frac{2\mu}{\pi V^{\frac{1}{2}}}\right)^{\frac{1}{2}} r \exp\left[i\left(\frac{V^{\frac{1}{2}}}{\mu r} - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\right] + B\left(\frac{2\mu}{\pi V^{\frac{1}{2}}}\right)^{\frac{1}{2}} r \exp\left[-i\left(\frac{V^{\frac{1}{2}}}{\mu r} - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\right] \quad r \to 0, \quad (12a)$$

$$\varphi(r) \to A'\left(\frac{2}{\pi k}\right)^{\frac{1}{2}} \exp\left[i\left(kr - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\right] + B'\left(\frac{2}{\pi k}\right)^{\frac{1}{2}} \exp\left[-i\left(kr - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\right] \quad r \to \infty. \quad (12b)$$

Several things should be noted about (12). First the correct wavefunction does not in general have the same coefficients at the origin and at infinity, in other words, $A \neq A'$ and $B \neq B'$. This is a result of the fact that if the $A$ term in (12a) is continued across the point $r = |\lambda|^{-1}$ as explained above, the resulting expression will contain some of each of the $A'$ and $B'$ terms; similarly for the $B$ term.

The infinite behavior of $\varphi(r)$ is exactly what we would expect and causes no difficulty. But at the origin, both solutions tend to zero while oscillating infinitely rapidly. To define a phase shift at infinity it would be necessary to choose the ratio $A/B$ as some arbitrary parameter. This is essentially Case's parameter.

[12] The two independent solutions are wrongly given in part 12 of Additional Results of Ref. 10. When $\nu$ is an integer the form of the solutions and various expressions we use are rather different but lead of course to the same conclusions.

We now turn our attention to the more interesting case of a repulsive potential.

## IV. SOLUTIONS FOR REPULSIVE POTENTIAL

We must now replace $q$ everywhere in Sec. III by $iq$ as can be seen from the relationships established at the end of II. These also show that we now have

$$z = \pm\ln(1/\sqrt{2})(1 - i)\lambda r = \pm\ln\lambda r \mp \tfrac{1}{4}i\pi.$$

Most of the results of III can be taken over directly since the convergence and asymptotic properties of the series given there have been proved for complex $C^r_{2r}$, $\nu$, and $z$ (see MS and Bateman[13]). The parameter $\nu = \nu(a, iq)$ is not well studied in the literature, though $\nu(a, q)$ has been extensively investigated. Besides this difference in $\nu$, we see that our solution of (1) is of the form

$$\varphi = Ar^{\frac{1}{2}}M^{(3)}_\nu[q^{\frac{1}{2}}(\lambda r + i/\lambda r)] + Br^{\frac{1}{2}}M^{(4)}_\nu[q^{\frac{1}{2}}(\lambda r + i/\lambda r)]$$
$$= Ar^{\frac{1}{2}}M^{(3)}_\nu(kr + iV^{\frac{1}{2}}/\mu r) + Br^{\frac{1}{2}}M^{(4)}_\nu(kr + iV^{\frac{1}{2}}/\mu r). \tag{13}$$

As $r \to \infty$ we still have the forms (12b), but as $r \to 0$ it is easy to see that we have a finite solution only in the one case that $B = 0$; then

$$\varphi(r) \to A\left(\frac{2\mu}{\pi V^{\frac{1}{2}}}\right)^{\frac{1}{2}} r\exp\left[-i\left(\frac{\nu\pi}{2} + \frac{\pi}{4}\right)\right]\exp\left(-\frac{V^{\frac{1}{2}}}{\mu r}\right) \quad \text{as} \quad r \to 0. \tag{14}$$

This solution is unique[6] in that any other solution finite at the origin is merely a constant multiple of it.

Now the condition[14] that $|\cosh z| > 1$ for uniform convergence means that when $z = -\ln(\lambda r) + \frac{1}{4}i\pi$ we can use (13) only when $\lambda r < (2 - \sqrt{3})^{\frac{1}{2}}$, and when $z = \ln(\lambda r) - \frac{1}{4}i\pi$ the solutions are valid only for $\lambda r > (2 + \sqrt{3})^{\frac{1}{2}}$.

In order to connect these two regions we must make use of the solutions $Me_{\pm\nu}$ given in (11). We choose some $r_1$ such that $0 < r_1 < \lambda^{-1}(2 - \sqrt{3})^{\frac{1}{2}}$ and write

$$M^{(3)}_\nu(r_1) = \alpha Me_\nu(r_1) + \beta Me_{-\nu}(r_1),$$
$$M^{(3)\prime}_\nu(r_1) = \alpha Me'_\nu(r_1) + \beta Me'_{-\nu}(r_1),$$

where $\alpha$ and $\beta$ are to be determined. Now at the point $r = \lambda^{-1}$ we switch from $z = -\ln(\lambda r) + \frac{1}{4}i\pi$ to $z = \ln(\lambda r) - \frac{1}{4}i\pi$, which means that the sign of $\partial z/\partial r$ changes from minus to plus. Thus at this point we must have

$$\alpha Me^{(-)}_\nu(\lambda^{-1}) + \beta Me^{(-)}_{-\nu}(\lambda^{-1}) = \gamma Me^{(+)}_\nu(\lambda^{-1}) + \delta Me^{(+)}_{-\nu}(\lambda^{-1}),$$
$$\alpha Me^{(-)\prime}_\nu(\lambda^{-1}) + \beta Me^{(-)\prime}_{-\nu}(\lambda^{-1}) = -\gamma Me^{(+)\prime}_\nu(\lambda^{-1}) - \delta Me^{(+)\prime}_{-\nu}(\lambda^{-1}), \tag{16}$$

where $M^{(-)}$ and $M^{(+)}$ refer to solutions for which $z = -\ln(\lambda r) + \frac{1}{4}i\pi$ and $z = \ln(\lambda r) - \frac{1}{4}i\pi$, respectively. Using (16) to determine $\gamma$ and $\delta$ we then have for some $r_2 > \lambda^{-1}(2 + \sqrt{3})^{\frac{1}{2}}$

$$\gamma Me_\nu(r_2) + \delta Me_{-\nu}(r_2) = A'M^{(3)}_\nu(r_2) + B'M^{(4)}_\nu(r_2),$$
$$\gamma Me'_\nu(r_2) + \delta Me'_{-\nu}(r_2) = A'M^{(3)\prime}_\nu(r_2) + B'M^{(4)\prime}_\nu(r_2). \tag{17}$$

We know from (12b) that the $S$ matrix is given by

$$S(k, l) = i[A'(k, l)/B'(k, l)]e^{-i\nu\pi}e^{il\pi} = e^{2i\delta_l(k)}.$$

To evaluate $A'$ and $B'$ it is necessary to do some very tedious algebra and make use of the fact that the Wronskian of two independent solutions of the Mathieu equation is a constant. With our normalization, MS gives these Wronskians as

$$W(M^{(3)}_\nu, M^{(4)}_\nu) = [3, 4] = -4i/\pi,$$
$$[1, 3] = -[1, 4] = 2i/\pi, \tag{19}$$
$$[1, 2] = -[2, 3] = -[2, 4] = 2/\pi,$$

where

$$M^{(1)}_\nu = \tfrac{1}{2}M^{(3)}_\nu + \tfrac{1}{2}M^{(4)}_\nu, \quad M^{(2)}_\nu = -\tfrac{1}{2}iM^{(3)}_\nu + \tfrac{1}{2}iM^{(4)}_\nu.$$

It is also necessary to use the relationships

$$Me_\nu = [me_\nu(0)/M^{(1)}_\nu(0)]M^{(1)}_\nu$$

and

$$Me_{-\nu} = [me_\nu(0)/M^{(1)}_{-\nu}(0)](\cos\nu\pi M^{(1)}_\nu - \sin\nu\pi M^{(2)}_\nu).$$

Making use of the explicit forms (11) we can finally write

$$S(k, l) = ie^{-i\nu\pi}e^{il\pi}[(R^2 + 1)/(R^2 + e^{-i2\nu\pi})] \tag{20}$$

with

$$R = M^{(1)}_{-\nu}(0)/M^{(1)}_\nu(0)$$

and

$$M^{(1)}_{\pm\nu}(0) = (C^{\pm\nu}_0)^{-1} \times \sum_{s=-\infty}^{\infty}(-1)^s C^{\pm\nu}_{2s}J_s((iq)^{\frac{1}{2}})J_{\pm\nu+s}((iq)^{\frac{1}{2}}). \tag{21}$$

In (21) the $J$'s are the usual Bessel functions. As $E \to 0$ (or $q \to 0$) we show in the Appendix that

[13] Bateman Manuscript Project (McGraw-Hill Book Company, Inc., New York, 1955), Vol. III.

[14] It is remarkable that in two books, Ref. 10 and 13, the condition $|\cosh z| < 1$ is stated to exclude only real $z$ such that $-1 \leq z \leq 1$. See p. 201 of MS for a diagram of the region in the complex $z$ plane for which $|\cosh z| \leq 1$.

$\nu \to l + \frac{1}{2}$ and it is shown in MS that[15] $C_{2S}^{\pm\nu}(0) = 0$ $S \neq 0$ while $C_0^{+\nu}(0) \neq 0$ and is finite. Now $J_0(0) = 1$ and $J_{\pm\nu}((iq)^{\frac{1}{2}}) \to q^{\pm\frac{1}{4}}$ as $q \to 0$. Hence we see that $R \to k^{-(l+\frac{1}{2})}$ as $k \to 0$ and conclude that $S(0, l) = 1$ as indeed it should be.

Unfortunately the determination of the behavior of the phase shifts as the energy becomes infinite is not such a simple problem. However it may be possible to investigate such behavior and the question is currently being studied.

The expressions comprising the $S$ matrix though complicated are tractable for a computer calculation of scattering from an inverse fourth-power repulsive potential. As has recently become apparent[16] the use of simple potentials such as the Yukawa gives a remarkable good fit to high energy $p$–$p$ scattering. It would be of great interest to investigate the present potential because it lies in between the Coulomb (too long a tail for high-energy scattering) and the Yukawa (very short range). Such an investigation would help shed light on exactly where the transition between acceptable short-range potentials and unacceptable long-range potentials comes.


## ACKNOWLEDGMENTS

The author would like to thank Dr. Hadi Aly for bringing the paper of Bose to his attention as well as for continued assistance, and also Professor Emil Wolf and Professor William Streifer for helpful discussions about the properties of the Mathieu functions.


## APPENDIX

As mentioned earlier in the paper, the solutions to the Mathieu equation take on a different form when the parameter $\nu$ is a real integer. All the properties we have proved in this paper can be shown to hold for these unusual cases as well; in fact the proofs are easier then. We have considered the more general case that $\nu$ is not a real integer.

Though the derivations of various formulas for $\nu$ are generally given in M and MS for real $q$, a study of these derivations easily shows that they hold for complex $q$ as well. The reason $\nu$ enters the problem at all is because the recursion relations will converge only for certain values of $\nu$.[9] We give two formulas useful for computing $\nu$. For small complex $q$ we have

$$a = \nu^2 + \frac{1}{2(\nu^2 - 1)} q^2 + \frac{5\nu^2 + 7}{32(\nu^2 - 1)^3(\nu^2 - 4)} q^4 + \frac{9\nu^4 + 58\nu^2 + 29}{64(\nu^2 - 1)^5(\nu^2 - 4)(\nu^2 - 9)} q^6 + O(q^8).$$

We make use of this formula when $q = 0$ just below (21), but in general it must be truncated and inverted to find $\nu$. More useful is

$$\cos \nu\pi = \cos a^{\frac{1}{2}}\pi + \frac{\pi \sin a^{\frac{1}{2}}\pi}{4a^{\frac{1}{2}}(a - 1)} q^2 + \left[\frac{15a^2 - 35a + 8}{64(a - 1)^3(a - 4)aa^{\frac{1}{2}}} \pi \sin a^{\frac{1}{2}}\pi - \frac{\pi^2 \cos a^{\frac{1}{2}}\pi}{32a(a - 1)^2}\right] q^4 + \left[\frac{105a^5 - 1155a^4 + 3815a^3 - 4705a^2 + 1652a - 288}{256(a - 1)^5(a - 4)^2(a - 9)a^{5/2}} \pi \sin a^{\frac{1}{2}}\pi - \frac{\pi^3 \sin a^{\frac{1}{2}}\pi}{384(a - 1)^3 a^{\frac{1}{2}}} - \frac{15a^2 - 35a + 8}{256a^2(a - 1)^4(a - 4)} \pi^2 \cos a^{\frac{1}{2}}\pi\right] q^6 + \cdots .$$

Having found $\nu$ by these means (or others available), the recursion relations are solved for the $C_{2r}^{\nu}$ by way of continued fractions as is explained in detail in Morse and Feshback or M where worked examples are given. When $\nu$ is an integer the entire process is usually reversed so that one looks for the corresponding value of $a$ that makes $\nu$ an integer (this is the Mathieu eigenvalue problem). For small $q$ the formula for $C_{2r}^{\nu}$ given in footnote 15 is useful, and there are many other formulas given in the literature for $\nu$ and $C_{2r}^{\nu}$ but usually only of use for special ranges of $a$ and $q$. In any given situation it is necessary to search various references for any useful expressions.

[15] More exactly we have

$$C_{2S}^{\nu} \to C_0^{\nu}\left(\frac{\Gamma(\nu + 1)}{2^{2S}S!\,\Gamma(\nu + S + 1)}\right) q^{2|S|}$$

as $q \to 0$. See MS, p. 121.

[16] R. Serber, Phys. Rev. Letters **10**, 357 (1963).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# Interaction of Nonrelativistic Particles with a Quantized Scalar Field

EDWARD NELSON
*Department of Mathematics, Princeton University, Princeton, New Jersey*
(Received 18 February 1964; final manuscript received 28 May 1964)

We demonstrate the mathematical existence of a meson theory with nonrelativistic nucleons. A system of Schrödinger particles is coupled to a quantized relativistic scalar field. If a cutoff is put on the interaction, we obtain a well-defined self-adjoint operator. The solution of the Schrödinger equation diverges as the cutoff tends to infinity, but the divergence amounts merely to a constant infinite phase shift due to the self-energy of the particles. In the Heisenberg picture, we obtain a solution in the limit of no cutoff. We use a canonical transformation due to Gross to separate the divergent self-energy term. It is shown that the canonical transformation is implemented by a unitary operator, and that the transformed Hamiltonian, with an infinite constant subtracted, can be interpreted as a self-adjoint operator.

## 1. INTRODUCTION

THE main mathematical problem presented by quantum field theory is to establish the existence, or nonexistence, of relativistic interactions of quantized fields. The free fields are well-understood, but the Hamiltonians of relativistic interactions fail to exist as well-defined operators. Here we study the interaction of a field of spinless "nucleons" with a scalar "meson" field, where the nucleons are treated nonrelativistically. For most of our discussion we may limit ourselves to a fixed number $N$ of nucleons, since $N$ does not change with time. In this model also the Hamiltonian $H$ fails to exist as a well-defined operator. Despite this, we are able to associate with $H$ a well-defined one-parameter group of unitary operators which describes in an unambiguous way the temporal development of states.

If we put a cutoff $\kappa$ on the interaction Hamiltonian, we obtain a self-adjoint operator $H_\kappa$. The operators $\exp(-itH_\kappa)$ have no limit as $\kappa \to \infty$, but the divergence is of a very simple kind. There is a family of real constants $E_\kappa$, which diverges logarithmically as $\kappa \to \infty$, such that $\exp(-it(H_\kappa - NE_\kappa))$ does converge strongly to the unitary operator $\exp(-it\hat{H})$, where $\hat{H}$ is a self-adjoint operator which is bounded below. The operator $\hat{H}$ has the interpretation of the Hamiltonian $H$ with the infinite self-energy of the nucleons subtracted. Since a state is given not by a point in Hilbert space but by a ray, this implies that the state determined by $\exp(-itH_\kappa)\,\Psi$ does converge as $\kappa \to \infty$. In the Heisenberg picture, if $A$ is any bounded self-adjoint operator then $\exp(itH_\kappa)A\exp(-itH_\kappa)$ converges strongly as $\kappa \to \infty$.

This model has been discussed by Gross[1] and others, and it was known that the only divergence in the theory was the divergent self-energy. Gross found a canonical transformation, a modification of a dressing transformation used by Greenberg and Schweber,[2] such that, formally, $e^T H e^{-T} = H' + NE$, where $E$ is an infinite constant and the self-energy terms for $H'$ are finite in all orders of a perturbation expansion. We show that $H'$ and $e^T$ have interpretations as well-defined operators on Hilbert space. The operator $\hat{H}$ referred to above is $e^{-T}H'e^T$, and once this is shown to exist the convergence theorem is not difficult.

The author first studied the problem from the point of view of Feynman path integrals, and described this approach to the Conference on Analysis in Function Space[3] held at M.I.T. in June 1963. It was there that he learned of the work of Gross which makes possible a much simpler and more complete treatment of the problem.

We use some operator theory which, although known, is not entirely standard. These results are summarized in an appendix.

## 2. STATEMENT OF THE THEOREM

Our Hilbert space $\mathcal{H}$ is the tensor product of the space of nucleon wavefunctions and Fock space for the meson field. The meson field is relativistic, but we shall use nonrelativistic notation throughout. An element $\Psi$ of $\mathcal{H}$ is a sequence $\{\Psi^{(n)}\}$ of functions

[1] E. P. Gross, Ann. Phys. 19, 219 (1962).

[2] O. W. Greenberg and S. S. Schweber, Nuovo Cimento 8, 378 (1958).

[3] E. Nelson, "Schrödinger Particles Interacting with a Quantized Scalar Field," in *Proceedings of a Conference on the Theory and Applications of Analysis in Function Space held at Endicott House in Dedham, Massachusetts June 9–13, 1963*. Edited by W. T. Martin and I. Segal (Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1964), p. 87.

on $\mathbf{R}^{3N+3n}$ with $||\Psi|| < \infty$, where

$$||\Psi||^2 = \sum_{n=0}^{\infty} \int \cdots \int d\mathbf{x}_1 \cdots d\mathbf{x}_N \, d\mathbf{k}_1 \cdots d\mathbf{k}_n \times |\Psi^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_N; \mathbf{k}_1, \cdots, \mathbf{k}_n)|^2,$$

and each $\Psi^{(n)}$ is symmetric in $\mathbf{k}_1, \cdots, \mathbf{k}_n$. It will not be necessary to specify the statistics of the nucleons, and $\mathfrak{H}$ could be replaced by the subspace of all $\Psi$ which are antisymmetric (or symmetric) in the nucleon variables $\mathbf{x}_1, \cdots, \mathbf{x}_N$.

Let $M$, the nucleon mass, be a strictly positive constant, and let

$$H_{\text{nuc}} = \frac{1}{2M} \sum_{m=1}^{N} \mathbf{p}_m^2,$$

where $-\mathbf{p}_m^2 = \nabla_{\mathbf{x}_m}^2$ is the Laplacian in the variable $\mathbf{x}_m$ with the usual domain which makes it a self-adjoint operator. Let $\mu$, the meson mass, be a strictly positive constant, and let

$$\omega(\mathbf{k}) = (\mathbf{k}^2 + \mu^2)^{\frac{1}{2}}.$$

Define $H_{\text{mes}}$ by

$$(H_{\text{mes}}\Psi)^{(n)} = \sum_{j=1}^{n} \omega(\mathbf{k}_j)\Psi^{(n)},$$

on the domain $\mathfrak{D}(H_{\text{mes}})$ of all $\Psi$ in $\mathfrak{H}$ such that $\{(H_{\text{mes}}\Psi)^{(n)}\}$ is again in $\mathfrak{H}$.

The symbolic annihilation operators $a_{\mathbf{k}}$ and creation operators $a_{\mathbf{k}}^*$ are given by

$$(a_{\mathbf{k}}\Psi)^{(n)}(\mathbf{k}_1, \cdots, \mathbf{k}_n) = (n+1)^{\frac{1}{2}}\Psi^{(n+1)}(\mathbf{k}, \mathbf{k}_1, \cdots, \mathbf{k}_n), \tag{1}$$

$$(a_{\mathbf{k}}^*\Psi)^{(n)}(\mathbf{k}_1, \cdots, \mathbf{k}_n) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \delta(\mathbf{k} - \mathbf{k}_j) \times \Psi^{(n-1)}(\mathbf{k}_1, \cdots, \hat{\mathbf{k}}_j, \cdots, \mathbf{k}_n), \tag{2}$$

where $\hat{\mathbf{k}}_j$ indicates that $\mathbf{k}_j$ is omitted. We have suppressed the nucleon variables in (1) and (2), and shall frequently do so when considering operators which do not affect them. The formal expression of $H_{\text{mes}}$ in terms of $a_{\mathbf{k}}$ and $a_{\mathbf{k}}^*$ is $\int \omega(\mathbf{k}) a_{\mathbf{k}}^* a_{\mathbf{k}} \, d\mathbf{k}$. Formally, $[a_{\mathbf{k}}, a_{\mathbf{k}'}^*] = \delta(\mathbf{k} - \mathbf{k}')$.

Let $\mathfrak{n}$ be the number of mesons operator defined by

$$(\mathfrak{n}\Psi)^{(n)} = n\Psi^{(n)}$$

on the domain $\mathfrak{D}(\mathfrak{n})$ of all $\Psi$ in $\mathfrak{H}$ such that $\{n\Psi^{(n)}\}$ is again in $\mathfrak{H}$.

If $f$ is in $\mathfrak{L}^2(\mathbf{R}^3)$, $||f||_2^2 = \int |f(\mathbf{k})|^2 \, d\mathbf{k} < \infty$, then $\int f(\mathbf{k})a_{\mathbf{k}} \, d\mathbf{k}$ and $\int f(\mathbf{k})a_{\mathbf{k}}^* \, d\mathbf{k}$, defined by

$$\left(\int f(\mathbf{k})a_{\mathbf{k}} \, d\mathbf{k}\Psi\right)^{(n)}(\mathbf{k}_1, \cdots, \mathbf{k}_n) = (n+1)^{\frac{1}{2}} \int f(\mathbf{k})\Psi^{(n+1)}(\mathbf{k}, \mathbf{k}_1, \cdots, \mathbf{k}_n) \, d\mathbf{k}, \tag{3}$$

and

$$\left(\int f(\mathbf{k})a_{\mathbf{k}}^* \, d\mathbf{k}\Psi\right)^{(n)}(\mathbf{k}_1, \cdots, \mathbf{k}_n) = n^{-\frac{1}{2}} \sum_{j=1}^{n} f(\mathbf{k}_j)\Psi^{(n-1)}(\mathbf{k}_1, \cdots, \hat{\mathbf{k}}_j, \cdots, \mathbf{k}_n), \tag{4}$$

are well-defined operators on $\mathfrak{D}(\mathfrak{n}^{\frac{1}{2}})$. By the Schwarz inequality,

$$\left\|\int f(\mathbf{k})a_{\mathbf{k}} \, d\mathbf{k}\Psi\right\| \leq ||f||_2 \, ||\mathfrak{n}^{\frac{1}{2}}\Psi||, \qquad \left\|\int f(\mathbf{k})a_{\mathbf{k}}^* \, d\mathbf{k}\Psi\right\| \leq ||f||_2 \, ||(\mathfrak{n}+1)^{\frac{1}{2}}\Psi||. \tag{5}$$

Also,

$$\left(\Psi_1, \int f(\mathbf{k})a_{\mathbf{k}} \, d\mathbf{k}\Psi_2\right) = \left(\int \bar{f}(\mathbf{k})a_{\mathbf{k}}^* \, d\mathbf{k}\Psi_1, \Psi_2\right)$$

for all $\Psi_1$ and $\Psi_2$ in $\mathfrak{D}(\mathfrak{n}^{\frac{1}{2}})$.

Notice that if $f$ is an arbitrary measurable function, then $\int f(\mathbf{k})a_{\mathbf{k}} \, d\mathbf{k}$ is a densely defined linear operator on the domain of all $\Psi$ in $\mathfrak{H}$ such that the integrals on the right-hand side of (3) exist and define an element of $\mathfrak{H}$. However, this operator will fail to have a closure if $f$ is not in $\mathfrak{L}^2$, so that the adjoint operator will not be densely defined. In fact, only $\Psi = 0$ is in the domain of $\int f(\mathbf{k})a_{\mathbf{k}}^* \, d\mathbf{k}$ if $f$ is not in $\mathfrak{L}^2$. Thus, annihilation operators are better behaved than creation operators. This is why it is advantageous to have creation operators to the left in an operator product, where they may be moved to the other side of an inner product as annihilation operators.

Let $\kappa \leq \infty$, and define

$$\chi_\kappa(\mathbf{k}) = \begin{cases} 1, & |\mathbf{k}| < \kappa, \\ 0, & |\mathbf{k}| \geq \kappa. \end{cases}$$

Let

$$\varphi_\kappa(\mathbf{x}) = [2(2\pi)^3]^{-\frac{1}{2}} \int \omega(\mathbf{k})^{-\frac{1}{2}} \times (a_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}} + a_{\mathbf{k}}^*e^{-i\mathbf{k}\cdot\mathbf{x}})\chi_\kappa(\mathbf{k}) \, d\mathbf{k}, \qquad H_{I\kappa} = g \sum_{m=1}^{N} \varphi_\kappa(\mathbf{x}_m), \tag{6}$$

where $g$ is a real constant. Except for the nonrelativistic notation, $\varphi_\infty(\mathbf{x})$ is the scalar field as given in Schweber.[4] Since $\omega(\mathbf{k})^{-\frac{1}{2}}$ is not in $\mathfrak{L}^2(\mathbf{R}^3)$, $\varphi_\infty(\mathbf{x})$ and $H_{I\infty}$ are merely formal expressions. The introduction of the cutoff $\kappa$ ensures that, for $\kappa < \infty$, $H_{I\kappa}$ is a well-defined operator on the domain $\mathfrak{D}(\mathfrak{n}^{\frac{1}{2}})$.

[4] S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson, and Company, Evanston, Illinois, 1961), p. 177.

To describe the interaction of the nucleons with the meson field we study the behavior of $\exp(-itH_\kappa)$ as $\kappa \to \infty$, where

$$H_\kappa = H_{\text{nuc}} + H_{\text{mes}} + H_{I\kappa}. \tag{7}$$

The integral

$$E_\kappa = -2Mg^2[2(2\pi)^3]^{-1} \times \int (2M\omega(\mathbf{k})^2 + \omega(\mathbf{k})\mathbf{k}^2)^{-1}\chi_\kappa(\mathbf{k})\, d\mathbf{k} \tag{8}$$

is divergent for $\kappa = \infty$, but if $\kappa < \infty$ then $E_\kappa$ is a well-defined real number.

*Theorem. For each $\kappa < \infty$, $H_\kappa$ is a self-adjoint operator on $\mathfrak{H}$. There is a unique self-adjoint operator $\hat{H}$ on $\mathfrak{H}$ such that, for all real $t$ and all $\Psi$ in $\mathfrak{H}$,*

$$\lim_{\kappa\to\infty} e^{-it(H_\kappa - NE_\kappa)}\Psi = e^{-it\hat{H}}\Psi. \tag{9}$$

*The operator $\hat{H}$ is bounded below.*

### 3. THE CANONICAL TRANSFORMATION OF GROSS

We shall proceed formally at first. Let $K < \infty$, and let the function $\beta$ and the operator $T_\kappa$ be given by

$$\beta(\mathbf{k}) = -\frac{g[2(2\pi)^3]^{-\frac{1}{2}}\omega(\mathbf{k})^{-\frac{1}{2}}}{\omega(\mathbf{k}) + \mathbf{k}^2/2M}(1 - \chi_K(\mathbf{k})), \tag{10}$$

$$T_\kappa = \sum_{m=1}^{N} \int \beta(\mathbf{k})(a_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}_m} - a_\mathbf{k}^* e^{-i\mathbf{k}\cdot\mathbf{x}_m})\chi_\kappa(\mathbf{k})\, d\mathbf{k}. \tag{11}$$

We shall compute $e^{T_\kappa}H_\kappa e^{-T_\kappa}$, which for $\kappa = \infty$ and $K = 0$ is the transformation considered by Gross.[1] For reasons which will become apparent in Sec. 4, we shall choose a fixed large value of $K$, and consider $T_\kappa$ for finite values of $\kappa$ tending to $\infty$ and for $\kappa = \infty$. We shall always assume that $\kappa > K$ (otherwise $T_\kappa = 0$).

Observe that

$$e^{T_\kappa}\mathbf{p}_m e^{-T_\kappa} = \mathbf{p}_m + \mathbf{A}_{m\kappa} + \mathbf{A}_{m\kappa}^*, \tag{12}$$

where

$$\mathbf{A}_{m\kappa} = \int \mathbf{k}\beta(\mathbf{k})a_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}_m}\chi_\kappa(\mathbf{k})\, d\mathbf{k},$$
$$\mathbf{A}_{m\kappa}^* = \int \mathbf{k}\beta(\mathbf{k})a_\mathbf{k}^* e^{-i\mathbf{k}\cdot\mathbf{x}_m}\chi_\kappa(\mathbf{k})\, d\mathbf{k}, \tag{13}$$

since $[T_\kappa, \mathbf{p}_m] = \mathbf{A}_{m\kappa} + \mathbf{A}_{m\kappa}^*$ and $[T_\kappa, [T_\kappa, \mathbf{p}_m]] = 0$. Similarly,

$$e^{T_\kappa}a_\mathbf{k} e^{-T_\kappa} = a_\mathbf{k} + \sum_{m=1}^{N} \beta(\mathbf{k})\chi_\kappa(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}_m},$$
$$e^{T_\kappa}a_\mathbf{k}^* e^{-T_\kappa} = a_\mathbf{k}^* + \sum_{m=1}^{N} \beta(\mathbf{k})\chi_\kappa(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}_m}. \tag{14}$$

Therefore, to find $e^{T_\kappa}H_\kappa e^{-T_\kappa}$ we need only substitute (12) and (14) into

$$H_\kappa = (2M)^{-1}\sum_m \mathbf{p}_m^2 + \int \omega(\mathbf{k})a_\mathbf{k}^* a_\mathbf{k}\, d\mathbf{k} + g[2(2\pi)^3]^{-\frac{1}{2}} \times \sum_m \int \omega(\mathbf{k})^{-\frac{1}{2}}(a_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}_m} + a_\mathbf{k}^* e^{-i\mathbf{k}\cdot\mathbf{x}_m})\chi_\kappa(\mathbf{k})\, d\mathbf{k}.$$

This yields

$$\begin{aligned} e^{T_\kappa}H_\kappa e^{-T_\kappa} &= (2M)^{-1}\sum_m (\mathbf{p}_m^2 + \mathbf{A}_{m\kappa}^2 + \mathbf{A}_{m\kappa}^{*2} \\ &+ 2\mathbf{A}_{m\kappa}^*\cdot\mathbf{A}_{m\kappa} + [\mathbf{A}_{m\kappa}, \mathbf{A}_{m\kappa}^*] + 2(\mathbf{p}_m\cdot\mathbf{A}_{m\kappa} + \mathbf{A}_{m\kappa}^*\cdot\mathbf{p}_m) \\ &+ \{[\mathbf{A}_{m\kappa}, \mathbf{p}_m] + [\mathbf{p}_m, \mathbf{A}_{m\kappa}^*]\}) + H_{\text{mes}} \\ &+ \left\{\sum_m \int \omega(\mathbf{k})\beta(\mathbf{k})(a_\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{x}_m} + a_\mathbf{k}^* e^{i\mathbf{k}\cdot\mathbf{x}_m})\chi_\kappa(\mathbf{k})\, d\mathbf{k}\right\} \\ &+ \sum_m \sum_l \int (\omega(\mathbf{k})\beta(\mathbf{k})^2 + g[2(2\pi)^3]^{-\frac{1}{2}}\omega(\mathbf{k})^{-\frac{1}{2}}\beta(\mathbf{k})) \\ &\times e^{i\mathbf{k}\cdot(\mathbf{x}_m - \mathbf{x}_l)}\chi_\kappa(\mathbf{k})\, d\mathbf{k} + \{H_{I\kappa}\}. \end{aligned} \tag{15}$$

Now

$$[\mathbf{A}_{m\kappa}, \mathbf{p}_m] = \int \mathbf{k}^2\beta(\mathbf{k})a_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}_m}\chi_\kappa(\mathbf{k})\, d\mathbf{k},$$
$$[\mathbf{p}_m, \mathbf{A}_{m\kappa}^*] = \int \mathbf{k}^2\beta(\mathbf{k})a_\mathbf{k}^* e^{-i\mathbf{k}\cdot\mathbf{x}_m}\chi_\kappa(\mathbf{k})\, d\mathbf{k}, \tag{16}$$

so that, by the definition (10) of $\beta(\mathbf{k})$, the terms of (15) enclosed in braces { } add up to $H_{IK}$. Similarly,

$$[\mathbf{A}_{m\kappa}, \mathbf{A}_{m\kappa}^*] = \int \mathbf{k}^2\beta(\mathbf{k})^2\chi_\kappa(\mathbf{k})\, d\mathbf{k},$$

so that $(2M)^{-1}\sum_m [\mathbf{A}_{m\kappa}, \mathbf{A}_{m\kappa}^*]$ and the terms $m = l$ in the double sum add up to $N(E_\kappa - E_K)$. Therefore,

$$e^{T_\kappa}H_\kappa e^{-T_\kappa} = H_\kappa' + N(E_\kappa - E_K), \tag{17}$$

where

$$\begin{aligned} H_\kappa' = H_0 &+ (2M)^{-1}\sum_m (\mathbf{A}_{m\kappa}^2 + \mathbf{A}_{m\kappa}^{*2} + 2\mathbf{A}_{m\kappa}^*\cdot\mathbf{A}_{m\kappa} \\ &+ 2(\mathbf{p}_m\cdot\mathbf{A}_{m\kappa} + \mathbf{A}_{m\kappa}^*\cdot\mathbf{p}_m)) + V_\kappa + H_{IK}, \end{aligned} \tag{18}$$

$$H_0 = H_{\text{nuc}} + H_{\text{mes}}, \tag{19}$$

$$V_\kappa = \sum_{m\neq l} \int (\omega(\mathbf{k})\beta(\mathbf{k})^2 + g[2(2\pi)^3]^{-\frac{1}{2}}\omega(\mathbf{k})^{-\frac{1}{2}}\beta(\mathbf{k})) \times e^{-i\mathbf{k}\cdot(\mathbf{x}_m - \mathbf{x}_l)}\chi_\kappa(\mathbf{k})\, d\mathbf{k}. \tag{20}$$

Now we turn to questions of mathematical existence. By (10), $\beta$ is in $\mathfrak{L}^2(\mathbf{R}^3)$, so that for $\kappa \leq \infty$, the right-hand side of (11) is a well-defined skew-symmetric operator on the domain $\mathfrak{D}(\mathfrak{n}^{\frac{1}{2}})$. We define $T_\kappa$ to be the closure of this operator.

*Lemma 1. For $\kappa \leq \infty$, the operator $T_\kappa$ is skew-*

*adjoint,* $T_\kappa^* = -T_\kappa$, *so that* $e^{T_\kappa}$ *and* $e^{-T_\kappa}$ *are well-defined unitary operators. For all* $\Psi$ *in* $\mathfrak{F}$,

$$\lim_{\kappa\to\infty} e^{T_\kappa}\Psi = e^{T_\infty}\Psi, \qquad \lim_{\kappa\to\infty} e^{-T_\kappa}\Psi = e^{-T_\infty}\Psi. \tag{21}$$

*Proof:* Let $\mathfrak{F}(n)$ be the subspace of all $\Psi$ in $\mathfrak{F}$ such that $\Psi^{(j)} = 0$ whenever $j > n$, and let

$$\mathfrak{F}(\infty) = \bigcup_{n=1}^{\infty} \mathfrak{F}(n).$$

By (5), if $\Psi$ is in $\mathfrak{F}(n)$ then

$$||T_\kappa\Psi|| \leq 2N\,||\beta||_2\,(n+1)^{\frac{1}{2}}\,||\Psi||, \tag{22}$$

and $T_\kappa\Psi$ is in $\mathfrak{F}(n+1)$. From this it follows that, for all real $s$ and $\Psi$ in $\mathfrak{F}(\infty)$,

$$\sum_{j=0}^{\infty} |s|^j\,||T_\kappa^j\Psi||/j! < \infty.$$

Since $\mathfrak{F}(\infty)$ is dense in $\mathfrak{F}$, this implies[5] that $T_\kappa^* = -T_\kappa$, and that, for $\Psi$ in $\mathfrak{F}(\infty)$,

$$e^{sT_\kappa}\Psi = \sum_{j=0}^{\infty} s^j T_\kappa^j\Psi/j!.$$

Since we have estimates on the tail of this series which are independent of $\kappa$, and since for each $j$, $T_\kappa^j\Psi \to T_\infty^j\Psi$ for $\Psi$ in $\mathfrak{F}(\infty)$, this implies that (21) holds for all $\Psi$ in $\mathfrak{F}(\infty)$, and consequently for all $\Psi$ in $\mathfrak{F}$.

*Lemma 2. For* $\kappa < \infty$, $H_\kappa$ *is self-adjoint, and* $\mathfrak{D}(H_\kappa) = \mathfrak{D}(H_0)$.

*Proof:* It is clear that $H_0$ [see (19)] is self-adjoint, since it is the sum of two positive commuting self-adjoint operators. [There is a measure space $X$ such that $\mathfrak{F}$ is unitarily equivalent to $\mathfrak{L}^2(X)$ in such a way that $H_{\text{nuc}}$ and $H_{\text{mes}}$ correspond to multiplication by positive measurable functions $h_{\text{nuc}}$ and $h_{\text{mes}}$, respectively. For $\psi$ in $\mathfrak{L}^2(X)$, $(h_{\text{nuc}} + h_{\text{mes}})\,\psi$ is in $\mathfrak{L}^2(X)$ if and only if $h_{\text{nuc}}\psi$ and $h_{\text{mes}}\psi$ are, since $h_{\text{nuc}}$ and $h_{\text{mes}}$ are positive, so that $H_{\text{nuc}} + H_{\text{mes}}$ is unitarily equivalent to multiplication by $h_{\text{nuc}} + h_{\text{mes}}$, and so is self-adjoint.]

By an argument due to Kato,[6] we need only show that there is an $a < 1$ and a $b < \infty$ such that

$$||H_{I\kappa}\Psi|| \leq a\,||H_0\Psi|| + b\,||\Psi|| \tag{23}$$

for all $\Psi$ in $\mathfrak{D}(H_0)$. By (5), $||H_{I\kappa}\Psi|| \leq C\,||(n+1)^{\frac{1}{2}}\Psi||$, where $C$ is a finite constant (depending on $\kappa$, $\kappa < \infty$). For all $\epsilon > 0$ there is a $b' < \infty$ such that $||(n+1)^{\frac{1}{2}}\Psi|| \leq \epsilon\,||n\Psi|| + b'\,||\Psi|| \leq \epsilon\mu^{-1}\,||H_0\Psi|| + b'\,||\Psi||$, so if we choose $\epsilon < C^{-1}\mu$, then (23) holds.

[5] See E. Nelson, Ann. Math. 70, 572 (1959), Lemma 5.1.
[6] T. Kato, Trans. Am. Math. Soc. 70, 195 (1951), proof of Lemma 5.

By Lemmas 1 and 2, if $\kappa < \infty$ then $e^{T_\kappa}H_\kappa e^{-T_\kappa}$ is a well-defined self-adjoint operator.

*Lemma 3. If* $\kappa < \infty$, (17) *is true.*

*Proof:* First, it should be remarked that there is something to prove: the computation made in the beginning of this section shows only that for $\Psi$ in a suitable domain on which $[T_\kappa, H_\kappa]$ and $[T_\kappa, [T_\kappa, H_\kappa]]$ are defined, say on $\mathfrak{F}(\infty) \cap \mathfrak{D}(H_0)$, that

$$(H_\kappa + [T_\kappa, H_\kappa])\Psi = (H_\kappa' + N(E_\kappa - E_K))\Psi, \qquad [T_\kappa, [T_\kappa, H_\kappa]]\Psi = 0. \tag{24}$$

Let us use $C$ to denote constants which are finite for $\kappa < \infty$. Different occurrences of $C$ are not necessarily the same. Analogously to (22), we find that for all $\Psi$ in $\mathfrak{F}(n) \cap \mathfrak{D}(H_0)$,

$$||H_0T_\kappa\Psi|| \leq C(n+1)^{\frac{1}{2}}(||H_0\Psi|| + ||\Psi||),$$

so that $H_0e^{-sT_\kappa}\Psi$ is an entire function of $s$, for $\Psi$ in $\mathfrak{F}(\infty) \cap \mathfrak{D}(H_0)$. Since $H_\kappa + s[T_\kappa, H_\kappa]$ sends $\mathfrak{F}(\infty) \cap \mathfrak{D}(H_0)$ into $\mathfrak{F}(\infty)$, $e^{-sT_\kappa}(H_\kappa + s[T_\kappa, H_\kappa])\Psi$ is an entire function of $s$, for $\Psi$ in $\mathfrak{F}(\infty) \cap \mathfrak{D}(H_0)$, by the proof of Lemma 1. By (24), these two functions of $s$ have the same Taylor series at $s = 0$, so they are identical, and (17) holds when applied to any $\Psi$ in $\mathfrak{F}(\infty) \cap \mathfrak{D}(H_0)$.

We shall show that $\mathfrak{D}(H_\kappa') = \mathfrak{D}(H_0)$ and

$$||H_\kappa'\Psi|| \leq C(||H_0\Psi|| + ||\Psi||) \tag{25}$$

for all $\Psi$ in $\mathfrak{D}(H_0)$. If we assume this, then by (17) for $\Psi$ in $\mathfrak{F}(\infty) \cap \mathfrak{D}(H_0)$,

$$||H_\kappa e^{-T_\kappa}\Psi|| = ||e^{T_\kappa}H_\kappa e^{-T_\kappa}\Psi|| \leq C(||H_0\Psi|| + ||\Psi||),$$

and so by (23),

$$||H_0e^{-T_\kappa}\Psi|| \leq C(||H_0\Psi|| + ||\Psi||)$$

for all $\Psi$ in $\mathfrak{F}(\infty) \cap \mathfrak{D}(H_0)$. Since $\mathfrak{F}(\infty) \cap \mathfrak{D}(H_0)$ is dense in $\mathfrak{D}(H_0)$ in the norm $||H_0\Psi|| + ||\Psi||$, it follows that $e^{-T_\kappa}$ maps $\mathfrak{D}(H_0)$ into itself. Similarly, $e^{T_\kappa}$ maps $\mathfrak{D}(H_0)$ into itself. (One way to see this is to replace $g$ by $-g$.) Therefore, $\mathfrak{D}(e^{T_\kappa}H_\kappa e^{-T_\kappa}) = \mathfrak{D}(H_0) = \mathfrak{D}(H_\kappa) = \mathfrak{D}(H_\kappa')$. Again, since $\mathfrak{F}(\infty) \cap \mathfrak{D}(H_0)$ is dense in $\mathfrak{D}(H_0)$ in the norm $||H_0\Psi|| + ||\Psi||$, (17) holds for all $\Psi$ in $\mathfrak{D}(H_0)$, and since this is the domain of the operators on the two sides of (17), the two operators are equal.

It remains to establish (25). Now

$$||\mathbf{A}_{m\kappa}^2\Psi|| \leq C\,||(n+2)^{\frac{1}{2}}(n+1)^{\frac{1}{2}}\Psi|| \leq C(||H_{\text{mes}}\Psi|| + ||\Psi||) \leq C(||H_0\Psi|| + ||\Psi||),$$

and similarly for $\mathbf{A}_{m\kappa}^{*2}$ and $\mathbf{A}_{m\kappa}^* \cdot \mathbf{A}_{m\kappa}$. Also, $||\mathbf{A}_{m\kappa}^* \cdot \mathbf{p}_m\Psi|| \leq$

$C\,||(\mathfrak{n}+1)^{\frac{1}{2}}\mathbf{p}_m\Psi||$, and since $(\mathfrak{n}+1)^{\frac{1}{2}}$ and $\mathbf{p}_m$ commute, this is $\leq C(||(\mathfrak{n}+1)\Psi|| + ||\mathbf{p}_m^2\Psi||) \leq C(||H_0\Psi|| + ||\Psi||)$. Similarly, if we use (16).

$$||\mathbf{p}_m\cdot\mathbf{A}_{m\kappa}\Psi|| \leq ||\mathbf{A}_{m\kappa}\cdot\mathbf{p}_m\Psi|| + ||[\mathbf{A}_{m\kappa}, \mathbf{p}_m]\Psi|| \leq C(||H_0\Psi|| + ||\Psi||).$$

By (23), $||H_{I\kappa}\Psi|| \leq C(||H_0\Psi|| + ||\Psi||)$, so we need only show that $||V_\kappa\Psi|| \leq C(||H_0\Psi|| + ||\Psi||)$. For later use, we prove a stronger result.

*Lemma 4. For all $\epsilon > 0$ there is a $b < \infty$ such that, for all $\kappa \leq \infty$,*

$$||V_\kappa\Psi|| \leq \epsilon\,||H_0^{\frac{1}{2}}\Psi|| + b\,||\Psi||.$$

*Proof:* By (20), $V_\kappa = \sum_{m\neq l} W_\kappa(\mathbf{x}_m - \mathbf{x}_l)$ where, even for $\kappa = \infty$, $W_\kappa(\mathbf{x})$ is in $\mathfrak{L}^q(\mathbf{R}^3)$ for $2 \leq q < \infty$, since $(\omega(\mathbf{k})\beta(\mathbf{k})^2 + g[2(2\pi)^3]^{-\frac{1}{2}}\omega(\mathbf{k})^{-\frac{1}{2}}\beta(\mathbf{k}))$ is in $\mathfrak{L}^p(\mathbf{R}^3)$ for $1 < p \leq 2$. (We use the Hausdorff–Young theorem.[7]) We need only show that for all $\epsilon > 0$ there is a $b < \infty$ such that

$$||W_\kappa u|| \leq \epsilon\,||\mathbf{p}u|| + b\,||u|| \tag{26}$$

for all $u$ in $\mathfrak{L}^2(\mathbf{R}^3)$ in the domain of $\mathbf{p} = (1/i)\nabla$, since we may make a linear change of variables which carries $\mathbf{x}_m - \mathbf{x}_l$ into $\mathbf{x}_1$. This is a standard Soboleff inequality, and may be proved as follows. Let $\hat{u}$ be the Fourier transform of $u$. By the Plancherel theorem, $(1 + |\mathbf{p}|)\hat{u} \in \mathfrak{L}^2(\mathbf{R}^3)$. But $(1 + |\mathbf{p}|)^{-1}$ is in $\mathfrak{L}^{3+\alpha}(\mathbf{R}^3)$ for all $\alpha > 0$, so that by the Hölder inequality, $\hat{u} = (1 + |\mathbf{p}|)^{-1}(1 + |\mathbf{p}|)\hat{u}$ in in $\mathfrak{L}^{6/5+\alpha}(\mathbf{R}^3)$ for all $\alpha > 0$. By the Hausdorff–Young theorem,[7] $u$ is in $\mathfrak{L}^{q'}(\mathbf{R}^3)$ for all $q'$, $2 \leq q' < 6$, and

$$||u||_{q'} \leq C_{q'}(||\mathbf{p}u|| + ||u||), \tag{27}$$

where $C_{q'}$ is a constant depending only on $q'$. Choose $q$ and $q'$ so that $2 \leq q < \infty$ and $2 \leq q' < 6$ and $1/q + 1/q' = \frac{1}{2}$; for example, $q = q' = 4$. We may write $W_\kappa = W_{\kappa 0} + W_{\kappa 1}$ where $||W_{\kappa 0}||_q\, C_{q'} \leq \epsilon$ and $||W_{\kappa 1}||_\infty \leq b'$, where $b'$ is independent of $\kappa$. Then, by the Hölder inequality and (27),

$$||W_\kappa u|| \leq ||W_{\kappa 0}u|| + b'\,||u|| \leq ||W_{\kappa 0}||_q\,||u||_{q'} + b'\,||u|| \leq \epsilon\,||\mathbf{p}u|| + (\epsilon + b')\,||u||.$$

This concludes the proof of Lemma 4, and consequently of Lemma 3.

### 4. THE TRANSFORMED HAMILTONIAN

Our main task now is to assign a meaning to the expression $H'_\infty$ as a self-adjoint operator. Notice that since $\mathbf{k}\beta(\mathbf{k})$ is not in $\mathfrak{L}^2$, $\mathbf{A}^*_{m\infty}$ is not well-defined and $H'_\infty$ is meaningless as it stands.

Consider an analogy from the theory of partial differential equations. The expression

$$L = \sum_{i,j} \frac{\partial}{\partial x^i}\, a^{ij}\, \frac{\partial}{\partial x^j} + \sum_i b^i\, \frac{\partial}{\partial x^i} - \sum_i \frac{\partial}{\partial x^i}\, \bar{b}^i$$

is meaningless as a partial differential operator unless differentiability conditions hold for the coefficients $a^{ij}$, $b^i$. Formally, however, $(u, Lu) = B\langle u, u\rangle$, where the Hermitian form $B$ is given by

$$B\langle u, u\rangle = -\sum_{i,j}\left(\frac{\partial u}{\partial x^i}\,,\, a^{ij}\,\frac{\partial u}{\partial x^j}\right) + 2\,\mathrm{Re}\sum_i\left(u,\, b^i\,\frac{\partial u}{\partial x^j}\right).$$

(More properly, the Hermitian form $B\langle u, v\rangle$ is obtained from $B\langle u, u\rangle$ by polarization.) This is well-defined for all $u$ in $\mathfrak{L}^2$ whose derivatives are in $\mathfrak{L}^2$, if the $a^{ij}$ and $b^i$ are merely bounded and measurable. If in addition the matrix $a^{ij}$ is uniformly positive-definite, one can construct a unique self-adjoint operator $L$ such that $(u, Lu) = B\langle u, u\rangle$ for all $u$ in $\mathfrak{D}(L)$. We shall do something quite similar for $H'_\infty$.

Let $\kappa \leq \infty$. Formally, by (18),

$$(\Psi, H'_\kappa\Psi) = (H_0^{\frac{1}{2}}\Psi, H_0^{\frac{1}{2}}\Psi) + B_\kappa\langle\Psi, \Psi\rangle, \tag{28}$$

where

$$B_\kappa\langle\Psi, \Psi\rangle = M^{-1}\,\mathrm{Re}\sum_m \{((\mathfrak{n}+1)^{\frac{1}{2}}\Psi, (\mathfrak{n}+1)^{-\frac{1}{2}}\mathbf{A}^2_{m\kappa}\Psi) + (\mathbf{A}_{m\kappa}\Psi, \mathbf{A}_{m\kappa}\Psi) + (\mathbf{p}_m\Psi, \mathbf{A}_{m\kappa}\Psi)\} + (\Psi, V_\kappa\Psi) + (\Psi, H_{I\kappa}\Psi). \tag{29}$$

*Lemma 5. For all $\kappa \leq \infty$, $B_\kappa\langle\Psi, \Psi\rangle$ is well-defined for all $\Psi$ in $\mathfrak{D}(H_0^{\frac{1}{2}})$. For all $\epsilon > 0$ there is a $K < \infty$ and a $b < \infty$ such that*

$$|B_\kappa\langle\Psi, \Psi\rangle| \leq \epsilon(H_0^{\frac{1}{2}}\Psi, H_0^{\frac{1}{2}}\Psi) + b(\Psi, \Psi) \tag{30}$$

*for all $\kappa \leq \infty$ and all $\Psi$ in $\mathfrak{D}(H_0^{\frac{1}{2}})$. Also,*

$$\lim_{\kappa\to\infty} B_\kappa\langle\Psi, \Psi\rangle = B_\infty\langle\Psi, \Psi\rangle \tag{31}$$

*uniformly on any set of $\Psi$ in $\mathfrak{D}(H_0^{\frac{1}{2}})$ for which $||H_0^{\frac{1}{2}}\Psi|| + ||\Psi||$ is bounded.*

The proof will be given later. By Theorems A and B of the Appendix, Lemma 5 has the following immediate consequence.

*Lemma 6. For all $\kappa \leq \infty$, there is a unique self-adjoint operator $H'_\kappa$ with $\mathfrak{D}(H'_\kappa) \subset \mathfrak{D}(H_0^{\frac{1}{2}})$ such that (28) holds for all $\Psi$ in $\mathfrak{D}(H'_\kappa)$. The operators $H'_\kappa$ are*

[7] See A. Zygmund, *Trigonometric Series* (Cambridge University Press, Cambridge, England, 1959), 2nd ed., Vol. II, p. 254.

*bounded below. For all real $t$ and all $\Psi$ in $\mathfrak{H}$,*

$$\lim_{\kappa\to\infty} e^{-itH_\kappa'}\Psi = e^{-itH_\infty'}\Psi.$$

Using this, we may prove the theorem (stated at the end of Sec. 2). By Lemma 2, if $\kappa < \infty$ then $H_\kappa$ is self-adjoint. By Lemma 3 and the uniqueness assertion of Lemma 6, if $\kappa < \infty$ then

$$H_\kappa' = e^{T_\kappa}(H_\kappa - N(E_\kappa - E_K))e^{-T_\kappa},$$

so that for all real $t$,

$$e^{T_\kappa}e^{-it(H_\kappa - NE_\kappa)}e^{-T_\kappa} = e^{-it(H_\kappa' - NE_K)}.$$

By Lemma 6, this converges strongly to $e^{-it(H_\infty' - NE_K)}$ as $\kappa \to \infty$. By Lemma 1, therefore, $e^{-it(H_\kappa - NE_\kappa)} = e^{-T_\kappa}e^{-it(H_\kappa' - NE_K)}e^{T_\kappa}$ converges strongly to

$$e^{T_\infty}e^{-it(H_\infty' - NE_K)}e^{-T_\infty} \quad \text{as} \quad \kappa \to \infty.$$

Let $\hat{H} = e^{-T_\infty}(H_\infty' - NE_K)e^{T_\infty}$. Then (9) holds, and $\hat{H}$ is bounded below since $H_\infty'$ is. The uniqueness assertion of the theorem is trivial. Therefore, to conclude the proof of the theorem we need only establish Lemma 5.

Let $C_1(K)^2 = \int \mathbf{k}^2\beta(\mathbf{k})^2\omega(\mathbf{k})^{-1}\,d\mathbf{k}$. [Recall that, by (10), $\beta(\mathbf{k})$ contains the factor $1 - \chi_K(\mathbf{k})$.] Then for all $\kappa \leq \infty$ and $\Psi$ in $\mathfrak{D}(H_0^{\frac{1}{2}})$,

$$||\mathbf{A}_{m\kappa}\Psi||^2 = \sum_{n=0}^{\infty}\int\cdots\int\left|\int \mathbf{k}\beta(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}_m}\omega(\mathbf{k})^{-\frac{1}{2}}\omega(\mathbf{k})^{\frac{1}{2}} \times (n+1)^{\frac{1}{2}}\Psi^{(n+1)}(\mathbf{k},\mathbf{k}_1,\cdots,\mathbf{k}_n)\chi_\kappa(\mathbf{k})\,d\mathbf{k}\right|^2 d\mathbf{k}_1\cdots d\mathbf{k}_n$$

$$\leq C_1(K)^2\sum_{n=0}\int\cdots\int\omega(\mathbf{k})(n+1) \times |\Psi^{(n+1)}(\mathbf{k},\mathbf{k}_1,\cdots,\mathbf{k}_n)|^2\,d\mathbf{k}\,d\mathbf{k}_1\cdots d\mathbf{k}_n$$

$$= C_1(K)^2\,||H_{\text{mes}}^{\frac{1}{2}}\Psi||^2 \leq C_1(K)^2\,||H_0^{\frac{1}{2}}\Psi||^2, \tag{32}$$

by the Schwarz inequality and the fact that, due to the symmetry of $\Psi^{(n+1)}$, we may replace $\omega(\mathbf{k})(n+1)$ by $\omega(\mathbf{k}) + \omega(\mathbf{k}_1) + \cdots + \omega(\mathbf{k}_n)$. Similarly,

$$||(\mathbf{A}_{m\infty} - \mathbf{A}_{m\kappa})\Psi||^2 \leq C_1(\kappa)^2\,||H_0^{\frac{1}{2}}\Psi||^2, \tag{33}$$

where $C_1(\kappa)$ is, of course, $C_1(K)$ with $K$ replaced by $\kappa$. Also,

$$||\mathbf{p}_m\Psi||^2 \leq 2M\,||H_0^{\frac{1}{2}}\Psi||^2. \tag{34}$$

The inequalities (32), (33), and (34) prove (30) and (31) as far as the terms $(\mathbf{A}_{m\kappa}\Psi, \mathbf{A}_{m\kappa}\Psi)$ and $(\mathbf{p}_m\Psi, \mathbf{A}_{m\kappa}\Psi)$ of $B_\kappa$ are concerned, since $C_1(K)$ is arbitrarily small for $K$ large enough.

Let $C_2(K)^2 = \int \mathbf{k}^2\beta(\mathbf{k})^2\omega(\mathbf{k})^{-\frac{1}{2}}\,d\mathbf{k}$. Then, for all $\kappa \leq \infty$ and $\Psi$ in $\mathfrak{D}(H_0^{\frac{1}{2}})$,

$$||(\mathfrak{n}+1)^{-\frac{1}{4}}\mathbf{A}_{m\kappa}^2\Psi||^2 = \sum_{n=0}^{\infty}\int\cdots\int\left|\iint \mathbf{k}'\beta(\mathbf{k}')e^{i\mathbf{k}'\cdot\mathbf{x}_m}\omega(\mathbf{k}')^{-\frac{1}{2}} \times \mathbf{k}\beta(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}_m}\omega(\mathbf{k})^{-\frac{1}{2}}\omega(\mathbf{k}')^{\frac{1}{2}}\omega(\mathbf{k})^{\frac{1}{2}}(n+2)^{\frac{1}{2}}\Psi^{(n+2)}(\mathbf{k}',\mathbf{k},\mathbf{k}_1,\cdots,\mathbf{k}_n)\chi_\kappa(\mathbf{k})\chi_\kappa(\mathbf{k}')\,d\mathbf{k}\,d\mathbf{k}'\right|^2 d\mathbf{k}_1\cdots d\mathbf{k}_n$$

$$\leq C_2(K)^4\sum_{n=0}^{\infty}\int\cdots\int\omega(\mathbf{k}')^{\frac{1}{2}}\omega(\mathbf{k})^{\frac{1}{2}}(n+2)\,|\Psi^{(n+2)}(\mathbf{k}',\mathbf{k},\mathbf{k}_1,\cdots,\mathbf{k}_n)|^2\,d\mathbf{k}'\,d\mathbf{k}\,d\mathbf{k}_1\cdots d\mathbf{k}_n$$

$$\leq C_2(K)^4\,||H_{\text{mes}}^{\frac{1}{2}}\Psi||^2 \leq C_2(K)^4\,||H_0^{\frac{1}{2}}\Psi||^2, \tag{35}$$

where we have used $\omega(\mathbf{k}')^{\frac{1}{2}}\omega(\mathbf{k})^{\frac{1}{2}} \leq \frac{1}{2}(\omega(\mathbf{k}') + \omega(\mathbf{k}))$ and symmetrized. Similarly,

$$||(\mathfrak{n}+1)^{-\frac{1}{4}}(\mathbf{A}_{m\infty}^2 - \mathbf{A}_{m\kappa}^2)\Psi||^2 \leq C_2(\kappa)^4\,||H_0^{\frac{1}{2}}\Psi||^2. \tag{36}$$

Also,

$$||(\mathfrak{n}+1)^{\frac{1}{4}}\Psi||^2 \leq \mu^{-1}\,||H_0^{\frac{1}{2}}\Psi||^2 + ||\Psi||^2. \tag{37}$$

The inequalities (35), (36), and (37) prove (30) and (31) as far as the term $((\mathfrak{n}+1)^{\frac{1}{4}}\Psi, (\mathfrak{n}+1)^{-\frac{1}{4}}\mathbf{A}_{m\kappa}^2\Psi)$ in $B_\kappa$ is concerned.

Lemma 4 proves (30) as far as the term $(\Psi, V_\kappa\Psi)$ is concerned, and the method of proof establishes (31) for it.

It remains to consider $(\Psi, H_{IK}\Psi)$. By (5), there is a constant $C$ such that $||H_{IK}\Psi|| \leq C\,||(\mathfrak{n}+1)^{\frac{1}{4}}\Psi||$, so that by (37) there is a constant $C$ such that $||H_{IK}\Psi|| \leq C(||H_0^{\frac{1}{2}}\Psi|| + ||\Psi||)$. Thus

$$|(\Psi, H_{IK}\Psi)| \leq C(||\Psi||\,||H_0^{\frac{1}{2}}\Psi|| + ||\Psi||^2). \tag{38}$$

For all $\epsilon > 0$, $2\,||\Psi||\,||H_0^{\frac{1}{2}}\Psi|| \leq \epsilon\,||H_0^{\frac{1}{2}}\Psi||^2 + \epsilon^{-1}||\Psi||^2$, so that (38) proves (30) for the term $(\Psi, H_{IK}\Psi)$. This term is independent of $\kappa$, so (31) holds trivially for it. This concludes the proof.

### 5. REMARKS

Let us now denote the Hilbert space which we have been calling $\mathfrak{H}$ by $\mathfrak{H}_N$, where $N$ is the number of nucleons, and let now

$$\mathfrak{H} = \sum_{N=0}^{\infty}\mathfrak{H}_N.$$

Consider the Hamiltonian

$$H = (2M)^{-1} \int \psi^*(\mathbf{x})\mathbf{p}^2\psi(\mathbf{x})\, d\mathbf{x} + \int \omega(\mathbf{k})a_{\mathbf{k}}^* a_{\mathbf{k}}\, d\mathbf{k} + g \int \psi^*(\mathbf{x})\varphi(\mathbf{x})\psi(\mathbf{x})\, d\mathbf{x},$$

where $\psi$ is a nonrelativistic complex scalar field (of fermions or bosons). Since the number of nucleons is conserved, this is essentially the same problem as the one we have studied. If we put a cutoff $\kappa$ on the interaction we obtain a self-adjoint operator $H_\kappa$, and $\exp[-it(H_\kappa - \mathfrak{N}E_\kappa)]\Psi$ converges as $\kappa \to \infty$, where $\mathfrak{N}$ is the number of nucleons operator $\mathfrak{N} = \int \psi^*(\mathbf{x})\psi(\mathbf{x})\, d\mathbf{x}$. If $\Psi$ is in $\mathfrak{H}_N$ for some $N$, then by the theorem, although $e^{-itH_\kappa}\Psi$ diverges as $\kappa \to \infty$, the state determined by it converges, in the sense that if $A$ is any bounded self-adjoint operator,

$$(e^{-itH_\kappa}\Psi,\ Ae^{-itH_\kappa}\Psi) \to (e^{-it\hat{H}}\Psi,\ Ae^{-it\hat{H}}\Psi). \tag{39}$$

However, if $\Psi$ has nonzero components in several $\mathfrak{H}_N$, this is no longer true, and the left-hand side of (39) diverges in general. It is possible to "renormalize" $H$ by subtracting infinite constants on the $\mathfrak{H}_N$, but there is an arbitrariness in this process in that each infinite constant may be changed by a finite constant, giving different rules for propagating states which have components in several $\mathfrak{H}_N$. If one takes the point of view that $H$, although it does not exist, is the true Hamiltonian, then the theorem may be interpreted as giving meaning to the state (ray) $e^{-itH}\Psi$, provided that $\Psi$ lies in $\mathfrak{H}_N$ for some $N$. Thus the dynamics of the interaction impose a superselection rule.

The case of a field interacting with a fixed point source—which is essentially the limiting case $M = \infty$—is formally simpler but is also more singular. In this case $\beta$ is no longer in $\mathfrak{L}^2(\mathbf{R}^3)$, and $e^{T^\infty}$ is not implementable by a unitary transformation. A rigorous discussion of this situation has been given by Shale.[8] The fact that in our case $e^{T^\infty}$ is unitarily implementable depends strongly on the fact that our particles are nonrelativistic, which produces the term $\mathbf{k}^2/2M$ in the denominator of $\beta$.

It would be interesting to have a direct description of the operator $\hat{H}$. Is $\mathfrak{D}(\hat{H}) \cap \mathfrak{D}(H_0^{\frac{1}{2}}) = 0$?


## ACKNOWLEDGMENTS

The author is grateful to W. T. Martin and I. E. Segal for organizing a most stimulating conference at Endicott House, to E. P. Gross for a conversation at the conference which was the origin of this paper, to A. Jaffe, O. Lanford, D. Ruelle, and A. S. Wightman for pointing out many errors in the first version of the manuscript and for helpful advice on how to correct them, to the Alfred P. Sloan Foundation for financial support, and to the Institute for Advanced Study for hospitality.


## APPENDIX: A VARIANT OF THE FRIEDRICHS EXTENSION THEOREM

There is a complete equivalence between bounded self-adjoint operators and bounded Hermitian forms. The Friedrichs extension theorem gives conditions under which an unbounded Hermitian form gives rise to an unbounded self-adjoint operator. We shall develop this theory in a form suitable for our purposes.[9]

*Theorem A. Let $H_0$ be a positive self-adjoint operator on a Hilbert space $\mathfrak{H}$. Let $B\langle\Psi_1, \Psi_2\rangle$ be a Hermitian form defined for all $\Psi_1$ and $\Psi_2$ in $\mathfrak{D}(H_0^{\frac{1}{2}})$, such that for some constants $a < 1$ and $b < \infty$,*

$$|B\langle\Psi, \Psi\rangle| \leq a(H_0^{\frac{1}{2}}\Psi, H_0^{\frac{1}{2}}\Psi) + b(\Psi, \Psi) \tag{A1}$$

*for all $\Psi$ in $\mathfrak{D}(H_0^{\frac{1}{2}})$. Then there is a unique self-adjoint operator $H'$ with $\mathfrak{D}(H') \subset \mathfrak{D}(H_0^{\frac{1}{2}})$ such that for all $\Psi$ in $\mathfrak{D}(H')$,*

$$(\Psi, H'\Psi) = (H_0^{\frac{1}{2}}\Psi, H_0^{\frac{1}{2}}\Psi) + B\langle\Psi, \Psi\rangle. \tag{A2}$$

*The operator $H'$ is bounded below by $-b$.*

*Proof:* Let $\mathfrak{D} = \mathfrak{D}(H_0^{\frac{1}{2}})$. Then $\mathfrak{D}$ is a Hilbert space with respect to the inner product

$$(\Psi_1, \Psi_2)_{\mathfrak{D}} = (1 - a)(H_0^{\frac{1}{2}}\Psi_1, H_0^{\frac{1}{2}}\Psi_2) + (\Psi_1, \Psi_2).$$

(The space $\mathfrak{D}$ is complete since $H_0^{\frac{1}{2}}$ is a self-adjoint, and hence closed, operator.) It is convenient to introduce the Banach space $\mathfrak{D}^*$ of all continuous linear functionals on $\mathfrak{D}$. If $\Psi_0$ is in $\mathfrak{H}$ then $\Psi \to (\Psi_0, \Psi)$ is an element of $\mathfrak{D}^*$, and is the zero element of $\mathfrak{D}^*$ if and only if $\Psi_0 = 0$ (since $H_0^{\frac{1}{2}}$ is self-adjoint and hence densely defined). If $\Psi_0$ is any element of $\mathfrak{D}^*$ and $\Psi$ is in $\mathfrak{D}$, we denote the value of $\Psi_0$ on $\Psi$ by $(\Psi_0, \Psi)$. Thus we may regard $\mathfrak{H}$ as a subspace of $\mathfrak{D}^*$. If $\Psi_1$ is in $\mathfrak{D}$, define $A\Psi_1$ to be the linear functional

$$\Psi \to (H_0^{\frac{1}{2}}\Psi_1, H_0^{\frac{1}{2}}\Psi) + B\langle\Psi_1, \Psi\rangle + (b + 1)(\Psi_1, \Psi).$$

Thus $A\Psi_1$ is in $\mathfrak{D}^*$. Note that $(A\Psi_1, \Psi_1) \geq (\Psi_1, \Psi_1)_{\mathfrak{D}}$, so that $||A\Psi_1||_{\mathfrak{D}^*} \geq ||\Psi_1||_{\mathfrak{D}}$. Consequently, the linear operator $A : \mathfrak{D} \to \mathfrak{D}^*$ has closed range. Since $\mathfrak{D}$ is a Hilbert space, it is reflexive, and so $\mathfrak{D}^*$ is reflexive. That is, every continuous conjugate-linear functional

[8] D. Shale, J. Math. Phys. 3, 915 (1962).

[9] For a more general result, see J. L. Lions, *Equations Différentielles Opérationnelles et Problèmes aux Limites* (Springer-Verlag, Berlin, 1961), Chap. II, Sec. 1.

on $\mathfrak{D}^*$ is of the form $\Psi_1 \to (\Psi_1, \Psi)$ for some $\Psi$ in $\mathfrak{D}$. Suppose $\Psi$ in $\mathfrak{D}$ is orthogonal to the range of $A$. Then in particular, $(A\Psi, \Psi) = 0$, so that $(\Psi, \Psi)_{\mathfrak{D}} = 0$ and $\Psi = 0$. Therefore the range of $A$ is both closed and dense, so that $A$ maps $\mathfrak{D}$ onto $\mathfrak{D}^*$. Let $A'$ be the restriction of $A$ to those $\Psi$ in $\mathfrak{D}$ such that $A\Psi$ is in $\mathfrak{K}$. Recall that $\mathfrak{D}^* \supset \mathfrak{K} \supset \mathfrak{D}$, so that $A'$ so defined is an operator on $\mathfrak{K}$ with domain contained in $\mathfrak{D}$. Since $B\langle\Psi_1, \Psi_2\rangle$ is a Hermitian form, $A'$ is a symmetric operator, and clearly $(A'\Psi, \Psi) \geq (\Psi, \Psi)$ for all $\Psi$ in the domain of $A'$. Since the range of $A'$ is all of $\mathfrak{K}$, $A'$ is self-adjoint. Define $H' = A' - (b+1)$. Then $H'$ is self-adjoint, $H'$ is bounded below by $-b$, $\mathfrak{D}(H') \subset \mathfrak{D}(H_0^{\frac{1}{2}})$, and for all $\Psi$ in $\mathfrak{D}(H')$, (A2) holds.

To see that $H'$ is unique, suppose that $H''$ also has these properties. By polarization,

$$(\Psi_1, H''\Psi_2) = (H_0^{\frac{1}{2}}\Psi_1, H_0^{\frac{1}{2}}\Psi_2) + B\langle\Psi_1, \Psi_2\rangle \qquad \text{(A3)}$$

for all $\Psi_1$ and $\Psi_2$ in $\mathfrak{D}(H'')$. For fixed $\Psi_2$ in $\mathfrak{D}(H'')$, $\Psi_1 \to (\Psi_1, H''\Psi_2)$ is continuous in $\Psi_1$ for $\Psi_1$ in $\mathfrak{K}$ and consequently for $\Psi_1$ in $\mathfrak{D}$, so that (A3) holds for all $\Psi_1$ in $\mathfrak{D}$. Therefore $H'' + (b + 1)$ is the restriction of $A$ to $\mathfrak{D}(H'')$. By the construction of $H'$, this means that $H'$ is an extension of $H''$, and since $H'$ and $H''$ are self-adjoint, $H'' = H'$.

*Theorem B. Let $H_0$ be a positive self-adjoint operator on a Hilbert space $\mathfrak{K}$, and let $B_\kappa\langle\Psi_1, \Psi_2\rangle$ for $\kappa \leq \infty$ be a family of Hermitian forms on $\mathfrak{D}(H_0^{\frac{1}{2}})$ satisfying* (A1) *with fixed $a < 1$ and $b < \infty$, and let $H_\kappa'$ be the corresponding self-adjoint operators. Suppose that*

$$\lim_{K\to\infty} B_\kappa\langle\Psi, \Psi\rangle = B_\infty\langle\Psi, \Psi\rangle$$

*uniformly on any set of $\Psi$ in $\mathfrak{D}(H_0^{\frac{1}{2}})$ for which $||H_0^{\frac{1}{2}}\Psi|| + ||\Psi||$ is bounded. Then*

$$\lim_{K\to\infty} e^{-itH_\kappa'}\Psi = e^{-itH_\infty'}\Psi$$

*for all real $t$ and all $\Psi$ in $\mathfrak{K}$.*

*Proof:* Let $\lambda$ be imaginary. Then $\lambda - H_\kappa'$ converges uniformly to $\lambda - H_\infty'$ as mappings of $\mathfrak{D}$ into $\mathfrak{D}^*$, so that $(\lambda - H_\kappa')^{-1}$ converges uniformly to $(\lambda - H_\infty')^{-1}$ as mappings of $\mathfrak{D}^*$ into $\mathfrak{D}$, and consequently of $\mathfrak{K}$ into $\mathfrak{K}$. The theorem now follows from a result of Trotter.[10]

[10] H. F. Trotter, Pacific J. Math. 8, 887 (1958), Theorem 5.1.

# On Multiple-Valued Random Functions*

J. L. LUMLEY†
*Department of Aeronautical Engineering, The Pennsylvania State University, University Park, Pennsylvania*
(Received 24 March 1964; final manuscript received 3 April 1964)

A technique is developed for the statistical description of multiple-valued functions. For the special case of single-valued functions, the relations developed are shown to reduce in a natural way to the customary ones. Transformation laws are developed relating the statistical descriptions of a multiple-valued function in two coordinate systems. The relationship of the present work to the Rice formula for the zeros of a random function is shown. The statistical description of homogeneous, isotropic, multiple-valued functions in two dimensions is developed. Previously known results regarding expected arc length of random functions are obtained using the present technique.

## 1. INTRODUCTION

SEVERAL recent papers[1,2] have dealt with the expected contour length and expected area of multiple-valued random functions. These papers have avoided the problem of the general statistical description of multiple-valued functions. Since many physical problems give rise to multiple-valued functions which should be described statistically, (e.g., the contours of constant concentration in a turbulent mixing problem, or the contours of constant density in the turbulent flow of a compressible medium), it is of interest to develop such a description, in order to provide tools with which to handle other questions arising from such functions.

The arguments presented here are intentionally heuristic, and the cases considered are intended to be the simplest possible. Generality has in no sense been attempted. No doubt many points of interest lie in the more subtle cases which are excluded; it is desired, however, to present the main ideas as clearly as possible.

The discussion in this paper will be confined to the plane; the generalization to higher dimensions will be clear. For simplicity we have in mind a line drawn on the plane, (Fig. 1) this line having a slope defined everywhere (in some Cartesian coordinate system) except at points where the line crosses itself. That is to say, the only points at which a slope cannot be defined by rotating to a different coordinate system, are points where the line crosses itself. At such points, as many slopes can be defined as there are branches present at the crossing. In order to avoid difficulties we will specify that the line has no ends in the finite part of the plane, except those which occur at crossings. In such a case the slopes at the crossings are to be defined as limits from one side. We wish specifically to exclude limit points of crossings and similar phenomena. A somewhat more formal way of specifying such a function is to say, that for a given coordinate system, at almost every value of the abscissa, the function will take on a denumerable set of values and this set will contain no limit points. The function is continuous and differentiable in the sense that the sets corresponding to adjacent values of the abscissa contain values whose differences (or the ratio of whose differences to the separation) will tend to the appropriate limit as the separation is reduced to zero. This will be true almost everywhere in some coordinate system, and the remaining points can be taken care of as one-sided limits.

We begin by selecting a particular coordinate system. In this system we consider an ensemble of multiple-valued functions $f(x, \beta)$ of the type described above, where $\beta$ is an ensemble parameter lying in the closed unit interval, i.e.,

$$\beta \in [0, 1]. \tag{1}$$

### THE SIMPLE DENSITY

Let us define a random indicator function

$$\varphi(u, \Delta u; x, \beta) = \begin{cases} 1, & u \leq f(x, \beta) < u + \Delta u, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Thus, for specified values of $x$ and $\beta$, $\varphi$ is unity if $f(x, \beta)$ takes on at least one value lying in the interval $[u, u + \Delta u)$. The function $\varphi$ is a single-valued function. It is sufficient for subsequent pur-

* This work was supported by the Bureau of Naval Weapons under contract Nord 16597.
† Professor of Aeronautical Engineering, Garfield Thomas Water Tunnel, Ordnance Research Laboratory, and Aeronautical Engineering Department, The Pennsylvania State University.
[1] S. Corrsin, Quart. Appl. Math. 12, 404 (1955).
[2] S. Corrsin and O. M. Phillips, J. Soc. Indust. Appl. Math. 9, 395 (1961).

poses to require that all such functions so defined be Lebesgue-measurable in $\beta$.

Consider the quantity

$$\int_0^1 \varphi(u, \Delta u; x, \beta)\, d\beta. \tag{3}$$

By definition for specified values of $x$ and $u$, this is a positive nondecreasing function of $\Delta u$. It is bounded below and above by zero and unity, respectively. It is, in fact, simply the probability of $f(x, \beta)$ taking on at least one value in the interval $[u, u + \Delta u)$, if $\beta$ is uniformly distributed, as we shall assume. We will not make the assumption that its limiting value is unity as $u \to -\infty$, $\Delta u \to +\infty$, since we may wish to consider, for example, functions which, with probability unity, "do not occur" (i.e., take on no values) at certain values of the abscissa.

Instead, we will assume that

$$\lim_{\Delta u \to 0} \frac{\int_0^1 \varphi(u, \Delta u; x, \beta)\, d\beta}{\Delta u} = P_f(u, x)$$

exists for all $u$ and $x$. (4)

Since the values assumed by $f(x, \beta)$ for fixed $x$ and $\beta$ are denumerable and have no limit points, $P_f(u, x)du$ is the probability that $f(x, \beta)$ take on exactly one value in $[u, u + du)$, for vanishingly small $du$.

Now, we must ask what will be the value of

$$\int_{-\infty}^{+\infty} P_f(u, x)\, du. \tag{5}$$

This will not necessarily be unity, as can be seen from the following argument. In the case of single-valued functions $f(x, \beta)$, integrals of the form

$$\int_{-\infty}^{+\infty} F(u) P_f(u, x)\, du \tag{6}$$

are expectations. For a single-valued function the events $A$: "$f(x, \beta)$ is close to $a$" and $B$: "$f(x, \beta)$ is close to $b$" ($a \neq b$) are mutually exclusive. For a multiple-valued function, they are not, since in a certain number of cases the events can occur simultaneously. Thus, in computing the probability of the occurrence of $A$ or $B$ (or both) we must use

$$P(A \vee B) = P(A) + P(B) - P(AB), \tag{7}$$

that is, the overlapping region must be subtracted, since it is counted twice. Thus the sum

$$P(A) + P(B) = P(A \vee B) + P(AB) \tag{8}$$

is a probability only in the case of mutually exclusive events. Such a sum, being of the form of (6) is, in general, an expectation. In fact, (5) is the expectation of the count—unity each time an event occurs. (5) is thus the expected "valuedness". The statement that the value of unity for (5) corresponds to certainty of occurrence (i.e., certainty that the function takes on some value) is erroneous. When one makes the initial assumption that one is dealing with a surely single-valued function, then the value of unity for (5) is simply a consequence—the expected valuedness of a surely single-valued function is unity. On the other hand, the value of unity for (5) does not imply single valuedness: for instance, a function which is exactly double valued with probability one-half, and does not occur with the same probability, will also give the value unity for (5).

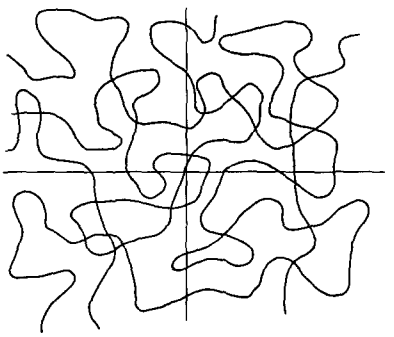

FIG. 1. A typical multiple-valued ensemble member.

We can examine this equation more formally. Imagine the $u$ axis to be divided by a mesh,

$$u_{k+1} = u_k + \Delta u_k. \tag{9}$$

Now consider

$$\sum_{k=p}^{q} \int_0^1 \varphi(u_k, \Delta u_k; x, \beta)\, d\beta = \int_0^1 \left\{ \sum_{n=p}^{q} \varphi(u_k, \Delta u_k; x, \beta) \right\} d\beta. \tag{10}$$

The integrand on the right can be interpreted as the total number of times that one or more values of $f$ have appeared in a $\Delta u$ window between $u_p$ and $u_{q+1}$ in the realization corresponding to $\beta$. Because of our assumption that the values assumed by $f(x, \beta)$ for fixed $x$ and $\beta$ are denumerable and have no limit points, as $||\Delta u_k|| \to 0$ the integrand will go to the total number of values of $f(x, \beta)$ between $u_p = u_1$, $u_q = u_2$ (say). We have simply

$$\int_{u_1}^{u_2} P_f(u, x)\, du = E\{n(x)\}_{u_1}^{u_2} \quad \text{(say)}, \tag{11}$$

the expected (ensemble average) number of values (or valuedness) of the function at $x$ between $u_1$ and $u_2$.

When $u_1 = -\infty$ and $u_2 = +\infty$, (11) is the expected valuedness of the function at $x$ and we will write it

$$E\{n(x)\}_{-\infty}^{+\infty} = E\{n(x)\} \tag{12}$$

when no confusion is possible. This may (and in many cases of interest will) be infinite, and we consider instead

$$\lim_{\substack{u_2 \to +\infty \\ u_1 \to -\infty}} \frac{E\{n(x)\}_{u_1}^{u_2}}{u_2 - u_1} \tag{13}$$

and similar quantities.

## THE SLOPE

We cannot build up a statistical description of the slope in quite such a straightforward way. This is because the valuedness of the slope will be at least as great as the valuedness of the function, and it is of interest to distinguish those cases in which the line can not cross itself, in which the valuedness of the slope is the same as that of the function, (simply being a reflection of this multiple-valuedness) and those cases in which the line can cross itself. In this latter case it seems worth while to define things in such a way that the multiple-valuedness of the slope is "per value of the function", i.e., in such a way that the valuedness of the function itself is not reflected.

It is perhaps easiest to begin by defining a two-point joint density

$$P_{ff}(u_1, x_1; u_2, x_2)\Delta u_1 \Delta u_2 \cong \int_0^1 \varphi(u_1, \Delta u_1; x_1, \beta)\varphi(u_2, \Delta u_2; x_2, \beta)\, d\beta. \tag{14}$$

From what we have already done, the normalization properties of this density are clear,

$$\int_{u_a}^{u_b} P_{ff}(u_1, x_1; u_2, x_2)\, du_1 = E\{n(x_1) \mid f(x_2) = u_2\}_{u_a}^{u_b} P_f(u_2, x_2), \tag{15}$$

where

$$E\{n(x_1) \mid f(x_2) = u_2\}_{u_a}^{u_b} \cong \frac{\int_0^1 n(x_1, \beta)|_{u_a}^{u_b} \varphi(u_2, \Delta u_2; x_2, \beta)\, d\beta}{\int_0^1 \varphi(u_2, \Delta u_2; x_2, \beta)\, d\beta} \tag{16}$$

is the expected number of values of $f(x_1)$, lying between $u_a$ and $u_b$, in those realizations in which $f(x_2)$ is approximately equal to $u_2$. $n(x_1, \beta)|_{u_a}^{u_b}$ is the number of values of $f(x_1, \beta)$ occurring between $u_a$ and $u_b$.

That is to say, (15) is of the form of a conditional expectation multiplied by the simple density; it is the conditional expectation of the function "1": a count of unity at each occurrence. (Note that we have now suppressed the ensemble index, to conform to customary usage).

As an aside, by setting $x_1 = x_2$ we can create a new type of density, the density for taking on two specific values simultaneously.

Now, to obtain the simultaneous value and *associated* slope, we have only to evaluate

$$\lim_{\Delta x \to 0} P_{ff}(u_1, x_1; u_1 + \alpha\Delta x, x_1 + \Delta x)\Delta x = P_{ff'}(u_1, \alpha; x_1), \tag{17}$$

which is the density for the value of $f$ to be close to $u_1$, while the *associated* slope is close to $\alpha$. We must obtain the normalization properties of this density.

From the relation

$$P_{ff'}(u_1, \alpha; x_1)\Delta u_1\, \Delta\alpha \cong \int_0^1 \varphi(u_1, \Delta u_1; x_1, \beta) \times \varphi(u_1 + \alpha\Delta x, \Delta\alpha\Delta x; x_1 + \Delta x, \beta)\, d\beta, \tag{18}$$

we can obtain, using the same techniques, the result

$$\int_{u_a}^{u_b} P_{ff'}(u_1, \alpha; x_1)\, du_1 = E\{n(x_1) \mid f'(x_1) = \alpha\}_{u_a}^{u_b} P_{f'}(\alpha, x_1), \tag{19}$$

where $E\{n(x_1) \mid f'(x_1) = \alpha\}_{u_a}^{u_b}$ is the expected number of values lying between $u_a$ and $u_b$ having *associated* slopes approximately equal to $\alpha$ considering only those realizations in which *some* value has an associated slope $\alpha$. $P_{f'}(\alpha, x_1)$ is the density function for the occurrence of slope $\alpha$ without regard for the associated value. That is, if we define

$$\psi(\alpha, \Delta\alpha; x, \beta) = \begin{cases} 1 & \text{if at least one value of} \\ & f'(x, \beta) \in [\alpha, \alpha + \Delta\alpha], \\ 0 & \text{otherwise,} \end{cases} \tag{20}$$

then

$$P_{f'}(\alpha, x)\Delta\alpha \cong \int_0^1 \psi(\alpha, \Delta\alpha; x, \beta)\, d\beta. \tag{21}$$

If we had chosen to define a joint density using

$$\int_0^1 \psi(\alpha, \Delta\alpha; x, \beta)\varphi(u, \Delta u; x, \beta)\, d\beta, \tag{22}$$

we would have gotten quite different answers, since we would have lost the information which associates a value with a slope: we would have statistics regarding the occurrence of a slope and a value without regard for whether the slope in question is the slope of $f(x)$ as it takes on that value. The

relation between the two types of densities could be obtained by defining a joint density for the occurrence of two values and the slope associated with one of them. Integration with respect to one or the other value could then give relations between the two types of densities. We will not pursue this further here, since the density defined in (17) is appropriate to our purposes.

Integrating (18) with respect to $\alpha$, we obtain

$$\int_{\alpha_1}^{\alpha_2} P_{ff'}(u, \alpha; x)\, d\alpha = E\{n'(x) \mid f(x) = u\}_{\alpha_1}^{\alpha_2} P_f(u, x), \tag{23}$$

where $E\{n'(x) \mid f(x) = u\}_{\alpha_1}^{\alpha_2}$ is the expected number of slopes having values between $\alpha_1$ and $\alpha_2$ associated with the value $u$, given that the value $u$ occurs [that is, in those realizations in which $f(x)$ taken on a value approximating $u$].

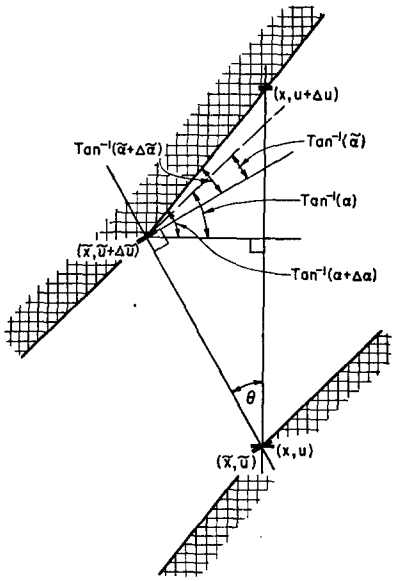


FIG. 2. Relations between amplitude windows under rotation.

## REDUCTION FOR SINGLE-VALUED FUNCTIONS

In the case of a surely single-valued function, it is clear that

$$\begin{aligned} &E\{n(x)\}_{-\infty}^{+\infty} = 1, \\ &E\{n(x) \mid f'(x) = \alpha\}_{-\infty}^{+\infty} = 1, \\ &E\{n'(x)\}_{-\infty}^{+\infty} = 1, \\ &E\{n'(x) \mid f(x) = u\}_{-\infty}^{+\infty} = 1, \\ &E\{n(x_1) \mid f(x_2) = u_2\}_{-\infty}^{+\infty} = 1, \end{aligned} \tag{24}$$

so that all our relations reduce to the customary ones.

As an aside, we might note that, for a multiple-valued function which can not cross itself, the number of slopes per value is unity, i.e.,

$$E\{n'(x) \mid f(x) = u\}_{-\infty}^{+\infty} = 1. \tag{25}$$

## GENERAL TRANSFORMATION PROPERTIES

Since we can now deal with multiple-valued functions as easily as (at least within the terms of our discussion) single-valued ones, we need no longer be tied to a particular coordinate system in defining our densities. If we can obtain the transformation properties of the densities, then we can examine the question of isotropic densities and similar problems. We can also obtain a check on our work, since the densities for a surely single-valued function, expressed in coordinates rotated through an angle of $\frac{1}{2}\pi$ radians from the usual, are the densities for a multiple-valued function, and should, in fact, give us the Rice formula for the probability of occurrence of zeros.

Let $P_{ff'}(u, \alpha; x)$ correspond to an $x$, $y$ coordinate system. Let $\tilde{P}_{ff'}(\tilde{u}, \tilde{\alpha}; \tilde{x})$ correspond to an $\tilde{x}$, $\tilde{y}$ system, rotated through a positive angle $\vartheta$.

Care must be taken in computing $\tilde{P}_{ff'}$, since a straightforward application of the usual transformation rules does not give the proper result. This is due to the fact that the rotating of axes causes an interaction between the slope of the function and the size of the amplitude "window". In Fig. 2, a simple diagram makes this clear. If a function appears in the $\tilde{u}$, $\tilde{u} + \Delta\tilde{u}$ window with slope lying in $\tilde{\alpha}$, $\tilde{\alpha} + \Delta\tilde{\alpha}$ (at constant $\tilde{x}$), then to appear in the $u$, $u + \Delta u$ window we must have (by the sine theorem)

$$\frac{\Delta u}{\sin\{\frac{1}{2}\pi - \vartheta + \tan^{-1}(\alpha + \Delta\alpha)\}} = \frac{\Delta\tilde{u}}{\sin\{\frac{1}{2}\pi - \tan^{-1}(\alpha + \Delta\alpha)\}}. \tag{26}$$

In the limit (as "window" size becomes smaller) we obtain

$$d\tilde{u} = du/(\cos\vartheta + \alpha\sin\vartheta). \tag{27}$$

Geometrically, it is clear that the size of the slope "window" does not interact with that of the amplitude window, so that we have

$$\tilde{\alpha} = \frac{\alpha - \tan\vartheta}{1 + \alpha\tan\vartheta}, \quad d\tilde{\alpha} = \frac{d\alpha}{(\cos\vartheta + \alpha\sin\vartheta)^2}. \tag{28}$$

Finally we can write (where we have introduced

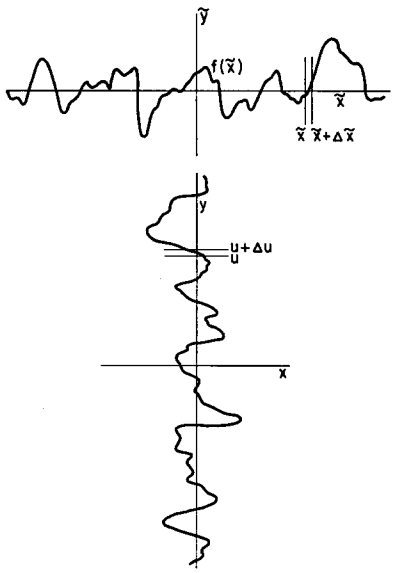


Fig. 3. The Rice example in two coordinate systems.

absolute value signs since only the magnitude of the transformation is of interest).

$$P_{ff'}(u, \alpha; x) = \tilde{P}_{ff'}\left(u \cos\vartheta - x \sin\vartheta, \frac{\alpha - \tan\vartheta}{1 + \alpha \tan\vartheta}; \times x \cos\vartheta + u \sin\vartheta\right) \frac{1}{|\cos\vartheta + \alpha \sin\vartheta|^3}. \quad (29)$$

## THE RICE FORMULA

Suppose that $f(x)$ was surely single-valued in $\tilde{x}$, $\tilde{y}$ coordinates (Fig. 3). We wish to obtain $P_f(u, 0)$. Integrating the joint density we obtain by (23)

$$\int_{-\infty}^{+\infty} P_{ff'}(u, \alpha; 0)\, d\alpha = E\{n'(0) \mid f(0) = u\} P_f(u, 0). \quad (30)$$

But the conditional expectation must equal unity, since the line can not cross itself. From (29), setting $\vartheta = \frac{1}{2}\pi$, we obtain

$$P_{ff'}(u, \alpha; x) = \tilde{P}_{ff'}\left(-x, -\frac{1}{\alpha}; u\right) \frac{1}{|\alpha|^3}. \quad (31)$$

Substituting in (27), we have

$$P_f(u, 0) = \int_{-\infty}^{+\infty} \tilde{P}_{ff'}\left(0, -\frac{1}{\alpha}; u\right) \frac{1}{|\alpha|^3}\, d\alpha, \quad (32)$$

and making a change of variable, $-1/\alpha = \tilde{\alpha}$, we obtain

$$P_f(u, 0) = \int_{-\infty}^{+\infty} \tilde{P}_{ff'}(0, \tilde{\alpha}; u)\, |\tilde{\alpha}|\, d\tilde{\alpha}, \quad (33)$$

which is exactly the Rice formula for the probability of occurrence of zeros.[3] Note that we have not used the assumption of single-valuedness of the function (though we have used the assumption of single-valuedness of the slope "per unit value"), so that (33) is equally valid for a multiple-valued function.[4]

The same technique can of course be used to obtain the well-known formula[5] for the simultaneous occurrence of two zeros.

## ISOTROPIC HOMOGENEOUS FUNCTIONS

We can examine isotropic functions—that is, functions for which $P_{ff'}$ is independent of $\vartheta$. The requirement for this is that (29) be valid if $P_{ff'}$ and $\tilde{P}_{ff'}$ have the same functional form. Differentiating, and requiring independence with respect to $\vartheta$, we obtain a first-order partial differential equation

$$-y_3 \frac{\partial g}{\partial y_1} + y_1 \frac{\partial g}{\partial y_3} - (1 + y_2^2) \frac{\partial g}{\partial y_2} - 3y_2 g = 0, \quad (34)$$

where we have set

$$y_1 = u \cos\vartheta - x \sin\vartheta,$$
$$y_2 = (\alpha - \tan\vartheta)/(1 + \alpha \tan\vartheta), \quad g = \tilde{P}_{ff'}, \quad (35)$$
$$y_3 = x \cos\vartheta + u \sin\vartheta.$$

If we apply the additional requirement of homogeneity, so that $P_{ff'}$ may not be a function of $x$ in any coordinate system, we find, of course, that it can not be a function of $u$ either, and by solving the associated equations corresponding to (34) we obtain

$$P_{ff'}(u, \alpha; x) = A(1 + \alpha^2)^{-\frac{3}{2}}, \quad (36)$$

where $A$ is a constant. If we restrict ourselves still further to a line that cannot cross itself, so that the conditional expectation in (23) is unity, then by integrating (36) with respect to $\alpha$, we obtain

$$P_f(u, x) = 2A. \quad (37)$$

Furthermore, using (11),

$$P_f(u, x) = E\{n(x)\}_{u_1}^{u_2}/(u_2 - u_1) = \bar{n}, \quad \text{say.} \quad (38)$$

That is, the expected number of occurrences per unit length. Finally, then, we have

$$P_{ff'}(u, \alpha; x) = \tfrac{1}{2}\bar{n}(1 + \alpha^2)^{-\frac{3}{2}}. \quad (39)$$

---

[3] S. O. Rice, Am. J. Math. **61**, 409 (1939).

[4] It is not necessary to assume that the lines do not cross, but only that the expected number of slopes per value is unity, or equivalently, that, of the members taking on a value near $u$ at $x$, only a set of measure zero has a crossing.

[5] S. O. Rice, Bell System Tech. J. **23**, 282 (1944).

## CONTOUR LENGTH OF ISOTROPIC HOMOGENEOUS FUNCTIONS

This can be applied immediately to the problem of the expected length of line per unit area as a function of the expected number of cuts per unit length on a line cast down arbitrarily in an isotropic homogeneous field of noncrossing lines.[4]

The expected length of line per unit area is just

$$\mathfrak{L} = \int_{-\infty}^{+\infty} (1 + \alpha^2)^{\frac{1}{2}} P_{ff'}(u, \alpha; x)\, d\alpha, \qquad (40)$$

since $P_{ff'}$ is independent of $u$ and $x$. Using (39), this is just

$$\mathfrak{L} = \frac{\bar{n}}{2} \int_{-\infty}^{+\infty} (1 + \alpha^2)^{-1}\, d\alpha = \frac{\bar{n}\pi}{2}. \qquad (41)$$

This is the result obtained by Corrsin[1] using a different method.

## CONTOUR LENGTH OF ANISOTROPIC FUNCTIONS

In Ref. 2, expressions are derived for the average contour length per unit length of axis for a homogeneous function, in both the single- and multiple-valued case. A way is presented of relating the results in the two cases. We present here another method of making this relation. The notation is that of Ref. 2.

Consider a multiple-valued homogeneous function of $x$, $f(x)$. If a line is passed through the $x$ axis at $x^*$, making an angle $\gamma$ with the positive side, the intersections of $f(x)$ with the line will be the zeros of the function

$$f(x) - \tan\gamma(x - x^*). \qquad (42)$$

Applying (33), the expected number of cuts made by this line is given by

$$M_\gamma = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} |\alpha|\, d\alpha \times P_{ff'}(\tan\gamma(x - x^*), \alpha + \tan\gamma), \qquad (43)$$

since $P_{ff'}(u, \alpha; x) = P_{ff'}(u, \alpha)$ by assumption. Carrying out the $x$ integration first, we obtain

$$M_\gamma = \int_{-\infty}^{+\infty} |\alpha|\, d\alpha \frac{1}{|\tan\gamma|} \times E\{n \mid f' = \alpha + \tan\gamma\} P_{f'}(\alpha + \tan\gamma), \qquad (44)$$

using (19). The average contour length per unit length of axis has been related to (44) by an integral over $\gamma$;[2] but the exact form of this integral need not concern us here. If the integration over $\gamma$ is carried out first, we obtain

$$L = \int_{-\infty}^{+\infty} d\alpha P_{f'}(\alpha) E\{n \mid f' = \alpha\}(1 + \alpha^2)^{\frac{1}{2}}. \qquad (45)$$

In the case of a surely single-valued function $E\{n \mid f' = \alpha\} = 1$, the density $P_{f'}(\alpha)$ is the usual one, and we have just

$$L = E\{(1 + \alpha^2)^{\frac{1}{2}}\}. \qquad (46)$$

Note that expression (45) could have been written down by inspection on the basis of the ideas presented above.

## ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. S. Corrsin and Dr. O. M. Phillips for their helpful comments.

# Linear Representation of Spinor Fields by Antisymmetric Tensors

JOHN R. KLAUDER
*Bell Telephone Laboratories, Incorporated,*
*Murray Hill, New Jersey*
(Received 8 May 1963; final manuscript received 8 May 1964)

A redundant, linear representation of four-component spinors $\psi(x)$ by antisymmetric tensors $f_{\mu\nu}(x)$ is proposed on the basis of the relation $\psi = \frac{1}{2}f_{\mu\nu}\gamma^{\mu\nu}v$, where $\gamma^{\mu\nu}$ is the skew product of gamma matrices and $v$ is a constant fiducial spinor. The redundant degrees of freedom in $f_{\mu\nu}$ as well as the particular choice of $v$ are both treated as unobservable "gauges," and the appropriate gauge groups are discussed. As part of this analysis, we demonstrate that the possibility that tensors can behave as spinors is intimately connected with the existence of two coordinate-invariant, gauge-covariant subsidiary conditions. A linear, tensorial reformulation of the Dirac equation is given, and shown to be the Euler–Lagrange equation of the conventional action functional for spinor fields. Finally covariance under arbitrary space-dependent coordinate and gauge transformations is discussed, and a generally covariant, tensorial form of the Dirac equation is proposed differing from the conventional generally covariant equation in the degree of arbitrariness in the (spin) connection.

## I. INTRODUCTION

INTEREST in tensorial descriptions of two- and four-component spinors and their equations of motion has existed almost since their inception[1–3] and persists to the present time.[4] In the usual approach, tensors are set in correspondence with a spin matrix, a two-index spinor, as opposed to a column spinor with but one index.[5] For this reason tensor equations equivalent to the Dirac equation lose the important property of linearity, at least explicitly.[3] It is our purpose to present a linear correspondence between single-index, four-component spinors and antisymmetric tensors, and discuss some of its consequences.

### A. Core of the Linear Representation of Spinors by Antisymmetric Tensors

The core of our linear correspondence is based on the following observations: Let $\gamma^\mu$ be four Dirac matrices[6] that satisfy

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu},$$

and define $\gamma^{\mu\nu}$ by

$$\gamma^{\mu\nu} \equiv \tfrac{1}{2}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu).$$

Let $v$ denote an arbitrary nonzero spinor—which we call the *fiducial spinor*—for which $\gamma_5 v \neq \pm iv$. *Then the six spinors* $\gamma^{\mu\nu}v$ *always span spin space,* and thus form a redundant "basis" for spin space. The expansion coefficients of an arbitrary spinor field $\psi(x)$ in this "basis" are complex, antisymmetric tensor fields:

$$\psi(x) = \tfrac{1}{2}f_{\mu\nu}(x)\gamma^{\mu\nu}v. \tag{1}$$

The tensorial transformation properties of $f_{\mu\nu}(x)$ under Lorentz transformations are dictated by (1) and the requirement that both $\psi(x)$ and $v$ are spinors.[7]

### B. Survey of Principal Topics

Equation (1) provides a linear map $f_{\mu\nu} \to \psi$ of antisymmetric tensors onto spinors, but because of the different dimensionality of these spaces this is necessarily a many–one mapping. Consequently, one of our initial tasks in Sec. II is to determine the

[1] E. T. Whittaker, Proc. Roy. Soc. (London) **A158**, 38 (1937).
[2] A. H. Taub, Ann. Math. **40**, 937 (1939).
[3] T. Takabayasi, Progr. Theoret. Phys. (Kyoto) **13**, 222 (1955); Phys. Rev. **102**, 297 (1956).
[4] Important issues involved in the relationship between tensors and spinors are put in particularly sharp focus by J. A. Wheeler, *Geometrodynamics* (Academic Press Inc., New York, 1962), pp. 88–94.
[5] An excellent account of the conventional formalism is given by E. M. Corson, *Introduction to Tensors, Spinors, and Relativistic Wave Equations* (Hafner Publishing Company, Inc., New York, 1953), Chap. II.
[6] We follow the notation of J. M. Jauch and F. Rohrlich, *Theory of Photons and Electrons* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1955). In particular, we adopt $g_{11} = g_{22} = g_{33} = -g_{00} = 1$; $\epsilon^{0123} = 1$; $\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3$. Additional relevant properties of the Dirac gamma matrices are given in Appendix A2 of the cited reference.
[7] An immediate algebraic consequence of (1) is

$$\gamma_5\psi(x) = -\tfrac{1}{2}\tilde{f}_{\mu\nu}(x)\gamma^{\mu\nu}v,$$

where

$$\tilde{f}_{\mu\nu}(x) \equiv \tfrac{1}{2}\epsilon_{\mu\nu\sigma\tau}f^{\sigma\tau}(x)$$

is the dual of $f_{\mu\nu}$. A $\gamma_5$ invariance of $\psi(\gamma_5\psi = \pm i\psi)$ is secured if $f_{\mu\nu}$ is chosen self-dual ($\tilde{f}_{\mu\nu} = \mp if_{\mu\nu}$). However, if $\gamma_5 v = \pm iv$, the six spinors $\gamma^{\mu\nu}v$ do not span spin space since all $\psi$ of the form (1) will satisfy $\gamma_5\psi = \pm i\psi$ independent of whether $f_{\mu\nu}$ is self-dual or not. We exclude this possibility by assuming $\gamma_5 v \neq \pm iv$.

set of $f_{\mu\nu}$ that map onto the same $\psi$.[8] For physical applications, we regard the difference between two antisymmetric tensors $f_{\mu\nu}$, which have the same image $\psi$, as an unobservable "gauge." As a consequence, our linear tensorial reformulation of the Dirac equation [Sec. II. I, Eq. (41)] is expressed in terms of redundant variables two of which correspond to gaugelike degrees of freedom. Of course, a redundant tensorial formulation is in no way incorrect, and indeed analogous redundant formulations with gauges permeate mathematical physics: e.g., the electromagnetic field, meson fields with nonzero integral spin, the gravitational field, etc.

A gaugelike property of a different nature, also discussed in Sec. II, arises from the arbitrary choice of the fiducial spinor $v$. If we adopt a different $v$, then Eq. (1) provides a different map $f_{\mu\nu} \to \psi$. An analysis of this gauge demonstrates that the tensors $f_{\mu\nu}$ that correspond to spinors undergo a gauge transformation isomorphic to an independent, spin-$\frac{1}{2}$ representation of the complex Lorentz group. The properties of this gauge group shed light on how tensors can behave as spinors, and show how tensors with the properties of spinors could have been introduced *ab initio*.

In Sec. III, we extend the linear tensorial representation of Eq. (1) to a general relativistic setting in which the gamma matrices and fiducial spinor become space dependent. General covariance requires the introduction of an affine connection—additional to the Christoffel connection—which is an analog of the well-known spin-affine connection.[9] The "geometrical" relation determining our affine connection differs from the usual one,[9] and leads to an affine with not just one, but ten arbitrary vector fields. As a consequence, our generally covariant linear tensorial formulation of the Dirac equation is capable of exhibiting Yukawa-like interactions with a number of independent vector fields.

The results which are developed in the following sections should demonstrate that a linear tensorial reformulation of any spinor relation may be readily obtained. It is ultimately hoped that a linear representation of spinors by tensors, for which we have an unquestionably greater intuitive feeling, may yield a deeper insight into the significance and meaning of spin and spinor fields in nature.

## II. ALGEBRAIC AND GAUGE PROPERTIES OF THE REDUNDANT TENSORIAL FORMULATION. REPRESENTATION OF THE SPECIAL RELATIVISTIC DIRAC EQUATION

### A. Fiducial Tensors Based on the Fiducial Spinor, and their Algebraic Identities

With the help of the spinor $v$ and its "transpose-conjugate" $\hat{v} \equiv \tilde{v}B\gamma_5$,[10] only two complex tensors arise, which are

$$c^{\mu} \equiv i\hat{v}\gamma^{\mu}v, \tag{2}$$

$$c^{\mu\nu} \equiv i\hat{v}\gamma^{\mu\nu}v, \tag{3}$$

all other covariants vanishing identically. Along with $c^{\mu\nu}$ we introduce its dual[11]

$$\tilde{c}_{\mu\nu} \equiv \tfrac{1}{2}\epsilon_{\mu\nu\sigma\tau}c^{\sigma\tau},$$

which is independent of $c^{\sigma\tau}$ since $\gamma_5 v \neq \pm iv$. These three tensors are fundamental for our purposes, and we shall hereafter refer to them as the *fiducial tensors*.

The fiducial tensors satisfy a number of algebraic relations[12] all of which are consequences of the basic identity

$$4iv\hat{v} = c_{\mu}\gamma^{\mu} - \tfrac{1}{2}c_{\mu\nu}\gamma^{\mu\nu}.$$

In particular, the spinor relations

$$c_{\mu}\gamma^{\mu}v = 0, \tag{4a}$$

$$c_{\mu\nu}\gamma^{\mu\nu}v = \tilde{c}_{\mu\nu}\gamma^{\mu\nu}v = 0, \tag{4b}$$

and

$$c^{\alpha}v = c^{\alpha}{}_{\beta}\gamma^{\beta}v = c_{\beta}\gamma^{\beta\alpha}v = \tilde{c}^{\alpha}{}_{\beta}\gamma_5\gamma^{\beta}v \tag{5}$$

hold. From these, follow the tensor relations

$$c_{\mu}c^{\mu} = 0, \tag{6a}$$

$$c_{\mu\nu}c^{\mu\nu} = \tilde{c}_{\mu\nu}c^{\mu\nu} = 0, \tag{6b}$$

and

$$c_{\mu\nu}c^{\nu} = \tilde{c}_{\mu\nu}c^{\nu} = 0, \tag{7}$$

$$c_{\mu\nu}c^{\nu\lambda} = \tilde{c}_{\mu\nu}\tilde{c}^{\nu\lambda} = -c_{\mu}c^{\lambda}, \tag{8a}$$

$$\tilde{c}_{\mu\nu}c^{\nu\lambda} = 0. \tag{8b}$$

---

[8] A description of spinors by redundant variables alternative to our Eq. (1) has been discussed by V. Hlavatý, *Geometry of Einstein's Unified Field Theory* (P. Noordhoff, Ltd., Groningen, The Netherlands, 1957), Appendix 2. We thank Professor Hlavatý for drawing this work to our attention.

[9] For general covariance under *coordinate* transformations, see any standard text covering general relativity, e.g., L. Landau and E. Lifshitz, *The Classical Theory of Fields*, translated by M. Hamermesh (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1951), Chap. 10. For general covariance under *spinor similarity* transformations see, e.g., W. L. Bade and H. Jehle, Rev. Mod. Phys. 25, 714 (1953); D. R. Brill and J. A. Wheeler, *ibid.* 29, 465 (1957).

[10] J. M. Jauch and F. Rohrlich, *Theory of Photons and Electrons* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1955).

[11] We denote the dual tensor by a tilde reserving the more conventional asterisk to denote the complex conjugate.

[12] Properties of biquadratic spinor identities have been frequently discussed. See, e.g., K. M. Case, Phys. Rev. 97, 810 (1955).

It may be noted that Eq. (4a) implies that the four spinors $\gamma^\mu v$ never span spin space.

The fiducial tensors are a complete replacement for the gamma matrices. Indeed, it suffices to begin with an antisymmetric tensor $c_{\mu\nu}$ both of whose invariants vanish as in (6b). These conditions imply the remaining algebraic constraints and determine $c_\mu$ up to an over-all sign from (8a).

We shall have use of the following consequence of (5): Let $b_\alpha$ be any four vector for which $b_\alpha c^\alpha \neq 0$. It then follows that

$$v = \tfrac{1}{2}E_{\beta\alpha}\gamma^{\beta\alpha}v, \qquad E_{\beta\alpha} \equiv (c_\beta b_\alpha - c_\alpha b_\beta)/(b_\mu c^\mu). \tag{9}$$

Thus $E_{\beta\alpha}$ is a tensor whose image spinor coincides with the fiducial spinor.

### B. Additive *c*-Gauge in the Mapping of Antisymmetric Tensors onto Spinors

The nature of the many–one linear mapping of antisymmetric tensors onto spinors is illuminated by Eq. (4b). This relation demonstrates that every tensor of the form

$$f_{\mu\nu}(x) + a(x)c_{\mu\nu} + b(x)\tilde{c}_{\mu\nu} \tag{10}$$

with $a(x)$ and $b(x)$ arbitrary is mapped by (1) into the same spinor $\psi(x)$. Equation (10) displays the additive "gauge" freedom of Eq. (1), which we shall call a *c*-gauge. Specifically we assert that (1) is invariant under a *c*-gauge transformation of $f_{\mu\nu}$ defined by

$$f_{\mu\nu} \to f_{\mu\nu}' = f_{\mu\nu} + ac_{\mu\nu} + b\tilde{c}_{\mu\nu}. \tag{11}$$

These transformations clearly form an Abelian gauge group.

In particular, the various tensors $E_{\mu\nu}$ that satisfy (9) differ from one another by *c*-gauges.

### C. The Inverse Map of Spinors onto Antisymmetric Tensors; Contragredient Tensors, and the Effective Metric

In analogy with (1) we can define a linear map of spinors $\varphi(x)$ onto antisymmetric tensors $y^{\mu\nu}(x)$, $\varphi \to y^{\mu\nu}$, according to

$$y^{\mu\nu}(x) \equiv i\hat{v}\gamma^{\mu\nu}\varphi(x). \tag{12}$$

The tensors defined by (12) differ from the $f^{\mu\nu}$ type in several respects: Firstly, they are *c*-gauge invariants, and secondly, they fulfill

$$c_{\mu\nu}y^{\mu\nu} = 0, \tag{13a}$$

$$\tilde{c}_{\mu\nu}y^{\mu\nu} = 0, \tag{13b}$$

which follow from the transpose adjoint of Eq. (4b). Equation (12) provides a tensorial expression of any spinor quantity, and Eq. (13) ensures that the tensorial expression $y^{\mu\nu}$ has the proper number of degrees of freedom.

Tensors of the type $y^{\mu\nu}$ may be said to be "contragredient" to those of the type $f_{\mu\nu}$. This is a consequence of the identity

$$\tfrac{1}{2}f_{\mu\nu}y^{\mu\nu} = i\hat{\psi}\varphi, \tag{14}$$

which suggests that the tensors $y^{\mu\nu}$ lie in a space $\mathfrak{Y}$ *dual* (as a linear vector space) to the space $\mathfrak{F}$ of tensors of the type $f_{\mu\nu}$. It is relation (13) that immunizes each $y^{\mu\nu}$ against the effect of any *c*-gauge transformation of $f_{\mu\nu}$.

Further insight into these two types of tensors may be won if we put $\varphi = \tfrac{1}{2}e_{\alpha\beta}\gamma^{\alpha\beta}v$, $e_{\alpha\beta} \in \mathfrak{F}$, in the manner of (1). Insertion of this relation into (12) leads to

$$y^{\mu\nu} = \tfrac{1}{2}M^{\mu\nu}{}_{\alpha\beta}e^{\alpha\beta}, \tag{15}$$

where

$$M^{\mu\nu}{}_{\alpha\beta} \equiv i\hat{v}\gamma^{\mu\nu}\gamma_{\alpha\beta}v = \delta^{[\mu}{}_{[\alpha}c^{\nu]}{}_{\beta]}, \tag{16}$$

which is expressed with help of the convention

$$A^{[\alpha}B^{\beta]} \equiv A^\alpha B^\beta - A^\beta B^\alpha.$$

Equation (15) gives a linear mapping of $\mathfrak{F}$ onto $\mathfrak{Y}$ with the help of the tensor $M^{\mu\nu}{}_{\alpha\beta}$. Furthermore,

$$\tfrac{1}{2}f_{\mu\nu}y^{\mu\nu} = \tfrac{1}{4}f_{\mu\nu}M^{\mu\nu\alpha\beta}e_{\alpha\beta} = i\hat{\psi}\varphi, \tag{17}$$

which enables us to interpret $M^{\mu\nu\alpha\beta} = -M^{\alpha\beta\mu\nu}$ as a kind of *metric* to form invariants from pairs of tensors in $\mathfrak{F}$, exactly in the sense that $B\gamma_5$ is a metric to form invariants from pairs of spinors.

### D. The *F*-Gauge Arising From Changes of the Fiducial Spinor

It is clear from (1) and (12) that the particular map $f_{\mu\nu} \to \psi$ and the particular map $\varphi \to y^{\mu\nu}$ depend very strongly on the particular choice made for $v$. Different fiducial spinors lead to different correspondences. But a change of $v$ and not of the gamma matrices is completely analogous to a change of representation of the gamma matrices without a corresponding change of $v$. Since no particular choice for a representation of the gamma matrices has any physical significance, it follows that any specific choice that is made for $\gamma^\mu$ must be treated as an unobservable "gauge." This result suggests that the choice of the fiducial spinor $v$ likewise corresponds to a choice of an unobservable gauge—let us say,

the $F$-gauge—and we now proceed to investigate this $F$-gauge freedom.

Under an $F$-gauge transformation,

$$v \rightarrow v_F, \tag{18a}$$

where $v_F$ denotes a new fiducial spinor (for which we require $\gamma_5 v_F \neq \pm i v_F$). An associated $F$-gauge transformation of $f_{\mu\nu}$,

$$f_{\mu\nu} \rightarrow f_{\mu\nu}{}^F, \tag{18b}$$

enables us to map $f_{\mu\nu}{}^F \rightarrow \psi$ according to

$$\psi = \tfrac{1}{2} f_{\mu\nu}{}^F \gamma^{\mu\nu} v_F = \tfrac{1}{2} f_{\mu\nu} \gamma^{\mu\nu} v. \tag{19}$$

Equation (19) is the fundamental condition that relates (18b) with the basic transformation (18a).

It is noteworthy that linearity of the basic mapping is preserved under an $F$-gauge; for if $f_{\mu\nu} \rightarrow \psi$ and $f'_{\mu\nu} \rightarrow \psi'$, then $f_{\mu\nu}{}^F + f'_{\mu\nu}{}^F$ and $(f_{\mu\nu} + f'_{\mu\nu})^F$ are mapped onto the same spinor and may be identified. (We ignore the simple $c$-gauge transformations here.)

The $F$-gauge transformation $v \rightarrow v_F$ induces a transformation in $y^{\mu\nu}$, Eq. (12), according to

$$y^{\mu\nu} \rightarrow y^{\mu\nu}{}_F \equiv i\bar{v}_F \gamma^{\mu\nu} \varphi. \tag{20}$$

As above, linearity of the map $\varphi \rightarrow y^{\mu\nu}$ is preserved under an $F$-gauge change.

Since the right side of (14) is composed of $F$-gauge invariant quantities, we find the important $F$-gauge invariant

$$f_{\mu\nu} y^{\mu\nu} = f_{\mu\nu}{}^F y^{\mu\nu}{}_F. \tag{21}$$

This equation implies that $f_{\mu\nu}$ and $y^{\mu\nu}$ undergo contragradient $F$-gauge transformations.

Finally, we consider the $F$-gauge transformation of the fiducial tensors for which

$$c^{\mu} \rightarrow c^{\mu}{}_{(F)} \equiv i\bar{v}_F \gamma^{\mu} v_F, \tag{22a}$$

$$c^{\mu\nu} \rightarrow c^{\mu\nu}{}_{(F)} \equiv i\bar{v}_F \gamma^{\mu\nu} v_F, \tag{22b}$$

while we set

$$\tilde{c}_{\mu\nu(F)} = \tfrac{1}{2} \epsilon_{\mu\nu\alpha\beta} c^{\alpha\beta}{}_{(F)}. \tag{23}$$

The distinction between (20) and (22b) should be carefully noted: the former involves a change of one spinor, while the latter involves a change of both spinors, a distinction for which the parentheses about $F$ serve as a reminder. Equations (22) and (23) clearly define new fiducial tensors that fulfill algebraic conditions similar to those in Eqs. (4)–(8).

### E. Spin Nature of the $F$-Gauge Group

The $F$-gauge transformations introduced above can clearly be regarded as a group of transformations with the various quantities, i.e., $v$, $f_{\mu\nu}$, $y^{\mu\nu}$, $c^{\mu\nu}$, $c^{\mu}$ transforming under different representations. To study this group it suffices to study the basic group of transformations $v \rightarrow v_F$, for this will determine the form of the remaining transformations.

We explicitly define the transformation of $v$ with the help of a $4 \times 4$ matrix $T$,

$$v_F = Tv. \tag{24}$$

It is actually possible to choose an eight (real) parameter family of transformations $\{T\}$ to describe (24), since every spinor is determined by eight real numbers. For example, $T$ may be labeled by the image fiducial spinor obtained after the transformation of a specific fiducial spinor. However, the formulation is simplified considerably by considering a group with additional parameters, all but eight of which are redundant.

In order to ensure $\gamma_5 v_F \neq \pm i v_F$ whenever $\gamma_5 v \neq \pm iv$, we choose a group of nonsingular transformations $\{T\}$ all of which commute with $\gamma_5$. The desired commutation is secured if the infinitesimal elements of these transformations commute with $\gamma_5$, which requires them to be linear sums of the eight matrices 1, $\gamma_{\mu\nu}$, and $\gamma_5$. To form such a group of transformations sufficiently large to carry any spinor into any other spinor, we must necessarily include as infinitesimal elements all the $\gamma_{\mu\nu}$. Since 1 and $\gamma_5$ would lead to an Abelian factor group with further redundant parameters, we include only the six $\gamma_{\mu\nu}$ as infinitesimal elements. And to obtain at least eight real parameters for $\{T\}$, we use six complex parameters for these infinitesimal elements.

The resultant family of transformations has a well-known form, and dictates *our choice of the basic $F$-gauge group as the spinor group $\mathcal{S}$ of complex Lorentz transformations.* A typical element, $T$, of this group has the form

$$T = \exp\left(\tfrac{1}{4} \epsilon_{\mu\nu} \gamma^{\mu\nu}\right), \tag{25}$$

where $\epsilon_{\mu\nu}$ is a complex, antisymmetric tensor. This group, indeed, has redundant parameters for our purposes as is made clear from the example

$$\exp\left[\tfrac{1}{4}(a c_{\mu\nu} + b \tilde{c}_{\mu\nu}) \gamma^{\mu\nu}\right] v = v.$$

Nevertheless, $\mathcal{S}$ is a convenient choice to define the $F$-gauge group of transformations.

### F. Detailed Nature of Various $F$-Gauge Transformations

#### *Homomorphic $F$-Gauge Transformations of the Fiducial Tensors*

Since $T$ has a form corresponding to a Lorentz transformation, it follows[10] that $\bar{v}_F = \bar{v} T^{-1}$, and thus

that

$$c^{\mu}{}_{(F)} = i\vartheta T^{-1}\gamma^{\mu}Tv = L^{\mu}{}_{\alpha}c^{\alpha}, \tag{26a}$$

where $L^{\mu}{}_{\alpha}$ are elements of the appropriate complex Lorentz transformation for vectors. In similar fashion, it follows that

$$c^{\mu\nu}{}_{(F)} = L^{\mu}{}_{\alpha}L^{\nu}{}_{\beta}c^{\alpha\beta}. \tag{26b}$$

These relations imply that $c^{\mu}$ and $c^{\mu\nu}$ transform in the manner of an *independent*, complex Lorentz transformation under an $F$-gauge transformation.

As a canonical example of such a transformation, consider the case $\epsilon_{12} = \theta$, $\theta$ complex, with all other independent $\epsilon_{\mu\nu}$ vanishing. This example corresponds to a (complex) rotation about the $z$ axis, and (26b) becomes

$$c^{12}{}_{(F)} = c^{12}, \tag{27a}$$

$$c^{13}{}_{(F)} = \cos(\theta)c^{13} + \sin(\theta)c^{23}, \tag{27b}$$

$$c^{23}{}_{(F)} = -\sin(\theta)c^{13} + \cos(\theta)c^{23}, \tag{27c}$$

together with the corresponding dual equations.

Since $b^{\mu}$, which enters the definition of $E^{\mu\nu}$ in (9), need only fulfill $b^{\mu}c_{\mu} \neq 0$, it follows that we can, if we choose, adopt a *new* $b^{\mu}$ in each gauge, $b^{\mu}{}_{(F)} = L^{\mu}{}_{\nu}b^{\nu}$, such that $b^{\mu}c_{\mu}$ actually remains an arbitrary, nonzero $F$-gauge invariant. In that case, $E^{\mu\nu}$ transforms as in (26b),

$$E^{\mu\nu}{}_{(F)} = L^{\mu}{}_{\alpha}L^{\nu}{}_{\beta}E^{\alpha\beta}, \tag{28}$$

i.e., under an independent but ordinary complex Lorentz transformation.

*Isomorphic F-Gauge Transformations*

The transformation of $y_{\mu\nu}$ reads $y^{\mu\nu}{}_{F} = i\vartheta \exp(-\frac{1}{4}\epsilon_{\alpha\beta}\gamma^{\alpha\beta})\gamma^{\mu\nu}\varphi$. To obtain a feel for this transformation let us first consider the canonical example of (27), where $\epsilon_{12} = \theta$, all other independent $\epsilon_{\mu\nu}$ vanishing. It follows immediately, for example, that

$$y^{12}{}_{F} = i\vartheta[\cos(\tfrac{1}{2}\theta)\gamma^{12} + \sin(\tfrac{1}{2}\theta)]\varphi,$$

$$y^{13}{}_{F} = i\vartheta \exp(-\tfrac{1}{4}\theta\gamma^{12})\gamma^{13}\exp(\tfrac{1}{4}\theta\gamma^{12})\varphi,$$

$$y^{23}{}_{F} = i\vartheta \exp(-\tfrac{1}{4}\theta\gamma^{12})\gamma^{23}\exp(\tfrac{1}{4}\theta\gamma^{12})\varphi,$$

together with the corresponding dual equations. The latter two equations again have the form of an ordinary rotation—but now by a (complex) angle of $\frac{1}{2}\theta$. If we employ the tensor $E_{\mu\nu}$, we can express these transformations as

$$y^{12}{}_{F} = \cos(\tfrac{1}{2}\theta)y^{12} - \sin(\tfrac{1}{2}\theta)(\tfrac{1}{2}E_{\mu\nu}y^{\mu\nu}), \tag{29a}$$

$$y^{13}{}_{F} = \cos(\tfrac{1}{2}\theta)y^{13} + \sin(\tfrac{1}{2}\theta)y^{23}, \tag{29b}$$

$$y^{23}{}_{F} = -\sin(\tfrac{1}{2}\theta)y^{13} + \cos(\tfrac{1}{2}\theta)y^{23}, \tag{29c}$$

together with their corresponding duals; specifically, the dual of the relation (29a) becomes

$$y^{30}{}_{F} = \cos(\tfrac{1}{2}\theta)y^{30} - \sin(\tfrac{1}{2}\theta)(\tfrac{1}{2}\tilde{E}_{\mu\nu}y^{\mu\nu}). \tag{30}$$

Clearly $y^{13}$ and $y^{23}$ transform as under a Lorentz transformation with the parameter $\frac{1}{2}\theta$.

Let us further specialize our example to gain insight into the form of $y^{12}{}_{F}$ and $y^{30}{}_{F}$ in this $F$-gauge transformation. We choose our fiducial tensors such that $c^{12} = c^{10} = 1$, and $c^{2} = c^{0} = 1$, all other independent components vanishing. For this example, $E_{20} = 1$, all other independent components vanishing. Thus, Eqs. (29a) and (30) become

$$y^{12}{}_{F} = \cos(\tfrac{1}{2}\theta)y^{12} - \sin(\tfrac{1}{2}\theta)y^{20}, \tag{31a}$$

$$y^{30}{}_{F} = -\sin(\tfrac{1}{2}\theta)y^{13} + \cos(\tfrac{1}{2}\theta)y^{30}. \tag{31b}$$

For our choice of $c^{\mu\nu}$, the constraints of (13) read $y^{12} = y^{10}$ and $y^{30} = -y^{23}$. In the new gauge, Eq. (27) and its dual indicate that $c^{12}{}_{(F)} = 1$, $c^{10}{}_{(F)} = \cos\theta$, and $c^{20}{}_{(F)} = -\sin\theta$, all other independent terms vanishing. These relations in (13) generate the constraints $y^{12}{}_{F} = \cos\theta\ y^{10}{}_{F} - \sin\theta\ y^{20}{}_{F}y^{30}{}_{F} = -\cos\theta\ y^{23}{}_{F} - \sin\theta\ y^{13}{}_{F}$ in the new $F$-gauge. When these two sets of constraints are inserted into (31), the resultant equations are seen to be consequences of (29b) and (29c) and their duals. In short, the curious transformations (31) are already contained in and implied by the simple half-angle rotations (29b) and (29c) and their duals. Thus the consistency and significance of (29a) and (30) are established.

We may extend the above example to a form appropriate to a general $\epsilon_{\mu\nu}$ only one of whose independent terms is nonvanishing. Let $\theta$ be the value of the single term in $\epsilon_{\mu\nu}$, and write $\epsilon_{\mu\nu} = \theta\zeta_{\mu\nu}$. We define $\xi = (\frac{1}{2}\zeta_{\mu\nu}\zeta^{\mu\nu})^{\frac{1}{2}}$, i.e., $\xi = 1$ for an ordinary rotation, while $\xi = i$ for a pure Lorentz rotation. Then

$$\begin{aligned} y^{\mu\nu}{}_{F} &= L^{\mu}{}_{\alpha}(\tfrac{1}{2}\xi\theta)L^{\nu}{}_{\beta}(\tfrac{1}{2}\xi\theta)y^{\alpha\beta} \\ &+ \tfrac{1}{2}[\cos(\tfrac{1}{2}\xi\theta) - 1](\zeta^{\mu\nu}\zeta_{\alpha\beta}y^{\alpha\beta} - \xi^{\mu\nu}\xi_{\alpha\beta}y^{\alpha\beta}) \\ &- \tfrac{1}{2}\xi^{-1}\sin(\tfrac{1}{2}\xi\theta)(\zeta^{\mu\nu}E_{\alpha\beta}y^{\alpha\beta} - \xi^{\mu\nu}\tilde{E}_{\alpha\beta}y^{\alpha\beta}), \end{aligned} \tag{32}$$

where $L^{\mu}{}_{\alpha}(\frac{1}{2}\xi\theta)$ denotes the appropriate Lorentz transformation with the half-angle parameter, $\frac{1}{2}\xi\theta$, explicitly stated.

The collection of $F$-gauge transformations determined by (32) constitute a generating set for all $F$-gauge transformations, and they imply that the tensors $y^{\mu\nu}$ undergo an $F$-gauge transformation isomorphic to an independent, spin-$\frac{1}{2}$ representation of the complex Lorentz group. The double-valued property characteristic of this representation is al-

ready contained in the canonical example presented above. To display this property, it suffices to set $\theta = 2\pi$; it follows from (29) that $y^{\mu\nu}{}_F = -y^{\mu\nu}$, while (27) leads to $c^{\mu\nu}{}_{(F)} = c^{\mu\nu}$.

The $F$-gauge transformation of $f_{\mu\nu}$ is contragredient to that of $y^{\mu\nu}$, and we first indicate its form for the canonical example discussed above. The result of our imposing the identity (21) and adopting a specific $c$-gauge leads to

$$f_{12}{}^F = \cos(\tfrac{1}{2}\theta) f_{12}, \tag{33a}$$

$$f_{13}{}^F = \cos(\tfrac{1}{2}\theta) f_{13} + \sin(\tfrac{1}{2}\theta) f_{23} + \sin(\tfrac{1}{2}\theta)\cos(\theta) f_{30}, \tag{33b}$$

$$f_{23}{}^F = -\sin(\tfrac{1}{2}\theta) f_{13} + \cos(\tfrac{1}{2}\theta) f_{23} - \sin(\tfrac{1}{2}\theta)\sin(\theta) f_{30}, \tag{33c}$$

together with the corresponding dual equations.

For an arbitrary $\epsilon_{\mu\nu} = \theta\zeta_{\mu\nu}$ only one of whose independent components is nonvanishing and has magnitude $\theta$, the generalization of (33) reads

$$\begin{aligned} f^{\mu\nu F} &= L^{\mu}{}_{\alpha}(-\tfrac{1}{2}\xi\theta) L^{\nu}{}_{\beta}(-\tfrac{1}{2}\xi\theta) f^{\alpha\beta} \\ &+ \tfrac{1}{2}[\cos(\tfrac{1}{2}\xi\theta) - 1](\zeta^{\mu\nu}\zeta_{\alpha\beta} f^{\alpha\beta} - \zeta^{\mu\nu}\zeta_{\alpha\beta} f^{\alpha\beta}) \\ &+ \tfrac{1}{2}\xi^{-1}(P^2 + \tilde{P}^2)^{-\frac{1}{2}} \sin(\tfrac{1}{2}\xi\theta)[E^{\mu\nu}{}_{(F)}(P\zeta_{\alpha\beta} f^{\alpha\beta} \\ &+ \tilde{P}\zeta_{\alpha\beta} f^{\alpha\beta}) + \tilde{E}^{\mu\nu}{}_{(F)}(\tilde{P}\zeta_{\alpha\beta} f^{\alpha\beta} - P\zeta_{\alpha\beta} f^{\alpha\beta})], \end{aligned} \tag{34}$$

where

$$P \equiv \tfrac{1}{2}(E_{\mu\nu} + c_{\mu\nu})\zeta^{\mu\nu}, \qquad \tilde{P} \equiv \tfrac{1}{2}(E_{\mu\nu} + c_{\mu\nu})\zeta^{\mu\nu}.$$

The set of transformations (34) form a generating set, and imply that $f^{\mu\nu}$ undergoes an $F$-gauge transformation isomorphic to an independent, complex, spin-$\frac{1}{2}$ Lorentz transformation. The double-valued property is exemplified by (33) when we set $\theta = 2\pi$, for then $f_{\mu\nu}{}^F = -f_{\mu\nu}$.

In order to examine the $c$-gauge freedom in connection with the $F$-gauge transformation (34), we turn our attention to a property of the combined $F$- and $c$-gauge groups.

### G. Connection Between Different $F$-Gauge Transformations of $c^{\mu\nu}$

In a fixed $F$-gauge, the tensor $c^{\mu\nu}$ may be regarded in three different ways: (i) as the fiducial tensor; (ii) as a member of $\mathcal{Y}$, since it fulfills the condition (13); or (iii) as a member of $\mathfrak{Y}$, the space dual to $\mathcal{Y}$. Whichever interpretation is chosen for $c^{\mu\nu}$ dictates its $F$-gauge transformation properties; and after an $F$-gauge transformation, these three tensors will no longer enjoy the numerical equality they possessed initially in view of their differing laws of transformation. We now wish to establish that

$$c_{\mu\nu(F)} = Ac_{\mu\nu}{}^F + B\tilde{c}_{\mu\nu}{}^F, \tag{35}$$

namely, that the transform of the fiducial tensor $c_{\mu\nu}$ is a linear combination of the transform of $c_{\mu\nu}$ and $\tilde{c}_{\mu\nu}$ viewed as members of $\mathfrak{Y}$. Here, $A$ and $B$ denote gauge-variant, Lorentz scalars.

We prove (35) most simply by assuming it to be false and showing contradiction. We assume that the right side of (35) is augmented by a term $CK_{\mu\nu}{}^F$, linearly independent from $c_{\mu\nu}{}^F$ and $\tilde{c}_{\mu\nu}{}^F$, and contract the modified relation with an *arbitrary* $y^{\mu\nu}{}_F$. If follows that

$$\begin{aligned} c_{\mu\nu(F)} y^{\mu\nu}{}_F &= (Ac_{\mu\nu}{}^F + B\tilde{c}_{\mu\nu}{}^F + CK_{\mu\nu}{}^F) y^{\mu\nu}{}_F \\ &= (Ac_{\mu\nu} + B\tilde{c}_{\mu\nu} + CK_{\mu\nu}) y^{\mu\nu} = 0, \end{aligned}$$

the last result following from (13), which in turn depends on (4b) with $v$ replaced by $v_F$. Since the coefficients of $A$ and $B$ are zero by assumption, we find that

$$CK_{\mu\nu} y^{\mu\nu} = 0 \tag{36}$$

for *all* $y^{\mu\nu} \in \mathcal{Y}$. We need only investigate four linearly independent $K^{\mu\nu}$, and it can be shown without difficulty that $E^{\mu\nu}$, defined by Eq. (9), $\tilde{E}^{\mu\nu}$, $J^{\mu\nu}$, and $\tilde{J}^{\mu\nu}$ form such a quartet, where $J^{\mu\nu} \equiv b^{[\mu} c^{\nu]}{}_{\alpha} b^{\alpha}$. When $K_{\mu\nu} = E_{\mu\nu}$, we find a violation of (36) for $y_{\mu\nu} = E_{\mu\nu} \in \mathcal{Y}$ since $\frac{1}{2}E_{\mu\nu}E^{\mu\nu} = -1$, unless $C = 0$. When $K_{\mu\nu} = \tilde{E}_{\mu\nu}$ a counter example is $y_{\mu\nu} = \tilde{E}_{\mu\nu} \in \mathcal{Y}$, unless $C = 0$. For $K_{\mu\nu} = J^{\mu\nu}$ we find a counter example in $y_{\mu\nu} = c_{\mu\nu} \in \mathcal{Y}$, since $\frac{1}{2}J^{\mu\nu}c_{\mu\nu} = (c_{\mu}b^{\mu})^2 \not\equiv 0$, unless $C = 0$. Finally, for $K_{\mu\nu} = \tilde{J}_{\mu\nu}$, $y_{\mu\nu} = \tilde{c}_{\mu\nu} \in \mathcal{Y}$ provides a counter example, unless $C = 0$. Thus no extra term can be consistently added to Eq. (35), and elementary examples show the necessity of both terms. The validity of (35) is therefore established.

Equation (35) sheds light on aspects of the $c$-gauge freedom of (34) associated with the nonuniqueness of $E_{\mu\nu}$, remarked on in Sec. II C. A different choice for $E_{\mu\nu(F)}$ augments Eq. (34) by terms proportional to $c_{\mu\nu(F)}$ and $\tilde{c}_{\mu\nu(F)}$. However, according to (35), these additional terms may, in turn, be expressed in terms of $c_{\mu\nu}{}^F$ and $\tilde{c}_{\mu\nu}{}^F$, and thus they correspond to a harmless, unobservable $c$-gauge transformation of $f_{\mu\nu}{}^F$.

We further note that Eq. (35) implies that the *subspace*, which is the image under an $F$-gauge transformation of the subspace

$$\mathfrak{C} \equiv \{ac_{\mu\nu} + b\tilde{c}_{\mu\nu} : a, b \text{ arbitrary complex numbers}\}$$

is the *same* whether $c_{\mu\nu}$ (and hence $\tilde{c}_{\mu\nu}$) transform as members of $\mathfrak{Y}$, or whether they transform as fiducial tensors. In an obvious notation we can

state this result as $\mathfrak{C}_{(F)} = \mathfrak{C}^F$. This observation is important in our answer to the question:

### H. How Can Tensors Behave as Spinors?

We are now in a position to gain additional insight into how, *beginning only with tensorial quantities, it would be possible to introduce tensors that behave as spinors.* Consider the collection $\mathfrak{R}$ of complex second-rank, skew-symmetric tensors, and single out of this family a fiducial tensor $c_{\mu\nu}$ and its dual $\tilde{c}_{\mu\nu}$ ($\not\equiv \pm ic_{\mu\nu}$), both of whose invariants vanish. We next use these tensors in order to define a subspace $\mathfrak{Y} \subset \mathfrak{R}$ as follows: Let $\mathfrak{Y}$ consist of all those tensors $y^{\mu\nu} \in \mathfrak{R}$ that fulfill the two relations

$$c_{\mu\nu}y^{\mu\nu} = 0, \tag{37a}$$

$$\tilde{c}_{\mu\nu}y^{\mu\nu} = 0. \tag{37b}$$

We now imagine that $\mathfrak{Y}$ formed the domain of tensors for some physical problem which itself in no way singles out any special antisymmetric tensors. If this is to be true, however, the *particular* tensors $c_{\mu\nu}$ and $\tilde{c}_{\mu\nu}$ we started with can not have any intrinsic physical significance, and the choice of these tensors must correspond to an arbitrary, unobservable gauge of the physical theory. We can pass to any other pair of acceptable fiducial tensors, $c_{\mu\nu(F)}$ and $\tilde{c}_{\mu\nu(F)}$, by means of an *independent*, complex Lorentz transformation, as indicated in Eq. (26b).

We might be tempted to define the transform of $y^{\mu\nu}$ by a contragredient (spin-1) complex Lorentz transformation. That would, in fact, be the *only* possiblity if we had imposed condition (37a) *or* (37b) *alone*. However, with the *two* conditions (37) an *alternate* choice exists for the transformation of $y^{\mu\nu}$; we could adopt a transformation for $y^{\mu\nu}$—call it $y^{\mu\nu}{}_F$—such that the vanishing of *each* of the equations

$$c_{\mu\nu(F)}y^{\mu\nu}{}_F = 0, \tag{38a}$$

$$\tilde{c}_{\mu\nu(F)}y^{\mu\nu}{}_F = 0, \tag{38b}$$

is a consequence of *both* Eqs. (37a) and (37b). In particular, if we employ (35), the validity of (38a) stems from

$$\begin{aligned} c_{\mu\nu(F)}y^{\mu\nu}{}_F &= (Ac_{\mu\nu}{}^F + B\tilde{c}_{\mu\nu}{}^F)y^{\mu\nu}{}_F \\ &= (Ac_{\mu\nu} + B\tilde{c}_{\mu\nu})y^{\mu\nu} = 0, \end{aligned}$$

which depends on both (37a) and (37b).

We may state these two transformation choices as follows: The tensors $y^{\mu\nu} \in \mathfrak{Y}$ are each orthogonal to the *subspace* $\mathfrak{C}$. The elements of the transformed $\mathfrak{Y}$ space must likewise be orthogonal to the transformed subspace $\mathfrak{C}$. But since $\mathfrak{C}_{(F)} = \mathfrak{C}^F$, we can transform $\mathfrak{Y}$ contragrediently to *either* of these choice for $\mathfrak{C}$ and still achieve the desired orthogonality in the new gauge; and it is those tensors which transform as $y^{\mu\nu}{}_F$ under the independent Lorentz transformation of $c_{\mu\nu}$ that behave as spinors. The *existence* of such tensors is made possible by the existence of *two* Lorentz-invariant subsidiary conditions (37), and *arises* when we regard these conditions as gauge-*co*variant (rather than gauge-*in*variant) under an independent, complex Lorentz transformation of the fiducial tensors.

To complete the present heuristic deduction of tensors which behave as spinors we need only introduce the space $\mathfrak{F}$ of tensors $f_{\mu\nu}$ as the space dual to the space $\mathfrak{Y}$. In so doing, we see that the $c$-gauge arises in a natural fashion.[13] That the tensors $f_{\mu\nu}$ likewise behave as spinors is ensured by our imposing on $f_{\mu\nu}$ a transformation contragredient to $y^{\mu\nu}$ under an $F$-gauge transformation.

### I. Redundant Tensorial Representation of the Dirac Equation

To convert the special relativistic Dirac spinor equation

$$(\gamma^\alpha \partial_\alpha + m)\psi = 0 \tag{39}$$

into a tensor equation, it suffices to carry out two operations: First, re-express $\psi(x)$ in terms of $f_{\mu\nu}(x)$ according to (1), and second, multiply Eq. (39) on the left by $i\hat{v}\gamma_{\sigma\tau}$, so as to express (39) itself in terms of the elementary "vectors" of the redundant basis. Since $\gamma^\mu$ and $v$ are independent of space, Eq. (39) becomes

$$\tfrac{1}{2}i(\hat{v}\gamma_{\sigma\tau}\gamma^\alpha\gamma^{\mu\nu}v)f_{\mu\nu,\alpha} + \tfrac{1}{2}mi(\hat{v}\gamma_{\sigma\tau}\gamma^{\mu\nu}v)f_{\mu\nu} = 0, \tag{40}$$

where the comma denotes ordinary partial derivative. If the products of gamma matrices are expanded in terms of the 16 basic matrices,[10] the two coefficients in (40) may be expressed in terms of $c^\mu$, $c^{\mu\nu}$, and the metric; indeed, the coefficient of $m$ has already been encountered in Eq. (16). It follows that Eq. (40) becomes

$$\begin{aligned} c^\mu(f_{\sigma\mu,\tau} + f_{\mu\tau,\sigma} + f_{\tau\sigma,\mu}) + c_\sigma f^\mu{}_{\tau,\mu} - c_\tau f^\mu{}_{\sigma,\mu} \\ + m(c_\sigma{}^\mu f_{\tau\mu} - c_\tau{}^\mu f_{\sigma\mu}) = 0, \end{aligned} \tag{41}$$

which is our redundant tensorial representation of

[13] The origin of the $c$-gauge discussed here may be compared with a related origin of the electromagnetic gauge in a standard introduction of the four-vector potential $A_\mu$ as a linear functional on currents defined by $\int j^\mu A_\mu \, dx$, wherein $j^\mu$ satisfies $j^\mu{}_{,\mu} = 0$.

the Dirac equation.[14] The first three terms of (41) may also be collected into the expression $\epsilon_{\sigma\mu\tau\beta}c^{\mu}\tilde{f}^{\beta\alpha}{}_{,\alpha}$.

Equation (41) is invariant under $c$-gauge transformations, and transforms covariantly as a member of $\mathfrak{Y}$ under $F$-gauge transformations. These properties are self-evident from the form (40) to which it is algebraically identical.

*Massless Dirac Equation and Duality Rotations*

If $m = 0$ and we employ the indicated alternate expression for the first three terms in (41), our tensorial form for the Dirac equation reads

$$\epsilon_{\sigma\mu\tau\beta}c^{\mu}\tilde{f}^{\beta\alpha}{}_{,\alpha} + c_{\sigma}f^{\mu}{}_{\tau,\mu} - c_{\tau}f^{\mu}{}_{\sigma,\mu} = 0. \tag{42}$$

The dual to this equation implies that $\tilde{f}_{\mu\nu}$ is also a solution whenever $f_{\mu\nu}$ is one. Linearity of the differential equation then leads to the result that the "duality rotated" tensor

$$(e^{\tilde{}\beta})f_{\mu\nu} \equiv \cos(\beta)f_{\mu\nu} + \sin(\beta)\tilde{f}_{\mu\nu} \tag{43}$$

is a solution of (42) for all constant $\beta$ if only $f_{\mu\nu}$ is a solution. This is simply our expression of the $\gamma_5$ invariance of the massless Dirac equation.

It is interesting to compare the solution degeneracy given in (43) with a well-known solution degeneracy for the source-free electromagnetic field $F_{\mu\nu}$. Specifically, Maxwell's source free equations admit a family of "duality rotated" solutions,

$$(e^{\tilde{}\alpha})F_{\mu\nu} \equiv \cos(\alpha)F_{\mu\nu} + \sin(\alpha)\tilde{F}_{\mu\nu}, \tag{44}$$

for all constant $\alpha$, provided only that $F_{\mu\nu}$ is a solution.

Our results in Eq. (43) may have bearing on a question raised by Wheeler,[15] whether there may exist a possible connection between $\gamma_5$ duality rotations for the massless Dirac equation and the duality rotations (44) for Maxwell's equations. The identity in form shared by (43) and (44) suggests that there is basically only one type of duality rotation, which, although we customarily view from different algebraic standpoints, can be made identical in form with the aid of the redundant tensorial formalism.[16]

[14] The redundant tensorial representation may be extended most directly to higher spin fields and their associated first-order equations by using the formalism of V. Bargmann and E. Wigner, Proc. Nat. Acad. Sci. U. S. **34**, 211 (1948).

[15] Reference 4, p. 91.

[16] It goes without mentioning that the similarity in form of degenerate solutions carries no implication of any similarity of the corresponding fields, e.g., between the neutrino and electromagnetic fields.

## J. Charge-Conjugate and Adjoint Tensors; Real Fiducial Tensors

Although we have confined our attention to spinors and tensors formed purely algebraically, physical applications require that we consider complex-conjugate quantities as well. For example, the charge-conjugate spinor[10]

$$\psi^{c}(x) = \tfrac{1}{2}f^{*}_{\mu\nu}(x)\gamma^{\mu\nu}v^{c}, \tag{45a}$$

$v^{c} \equiv C^{*}v^{*}$, can be expressed as

$$\psi^{c}(x) = \tfrac{1}{2}f^{c}_{\mu\nu}(x)\gamma^{\mu\nu}v \tag{45b}$$

in terms of the "charge-conjugate tensor"

$$f^{c}_{\mu\nu}(x) \equiv \tfrac{1}{2}f^{*}_{\alpha\beta}(x)C^{\alpha\beta}{}_{\mu\nu}, \tag{46}$$

where the tensor $C^{\alpha\beta}{}_{\mu\nu}$ satisfies

$$\gamma^{\alpha\beta}v^{c} = \tfrac{1}{2}C^{\alpha\beta}{}_{\mu\nu}\gamma^{\mu\nu}v. \tag{47}$$

Similarly, the Pauli adjoint of the spinor $\psi(x)$,[10]

$$\bar{\psi}(x) = -\tfrac{1}{2}f^{*}_{\mu\nu}(x)\bar{v}\gamma^{\mu\nu}, \tag{48a}$$

$\bar{v} \equiv v^{+}A$, can be expressed as

$$\bar{\psi}(x) = -\tfrac{1}{2}f^{c}_{\mu\nu}(x)\hat{v}\gamma^{\mu\nu}, \tag{48b}$$

where $f^{c}_{\mu\nu}$ is given by (46).

It is to be observed that $f^{c}_{\mu\nu}$ in (45b) and (48b) undergoes $F$-gauge transformations as a member of $\mathfrak{F}$, while $f^{*}_{\mu\nu}$ in (45a) and (48a) transforms as a member of $\mathfrak{F}^{*}$, i.e., according to the complex conjugate of (34).

Particularly noteworthy in the present connection are the special cases when the fiducial spinor $v$ is self-charge-conjugate, $v = v^{c}$, which lead to tensorial representations that are analogs of the Majorana representations of the gamma matrices. For self-charge-conjugate $v$, (47) is satisfied by $C^{\alpha\beta}{}_{\mu\nu} = \delta^{\alpha\beta}{}_{\mu\nu} \equiv \delta^{\alpha}{}_{[\mu}\delta^{\beta}{}_{\nu]}$, and therefore

$$f^{c}_{\mu\nu}(x) = f^{*}_{\mu\nu}(x), \quad \text{if} \quad v = v^{c}. \tag{49}$$

In addition to this simplification, the fiducial tensors $c^{\mu\nu}$, $\tilde{c}^{\mu\nu}$ and $c^{\mu}$ all become *real* tensors when $v = v^{c}$, [which, of course, must also satisfy the algebraic constraints, Eqs. (4)–(8)]. The $F$-gauge freedom that remains among self-charge-conjugate fiducial spinors corresponds to independent, spin-$\frac{1}{2}$ Lorentz transformations with *real* parameters $\epsilon_{\mu\nu}$.

Although self-charge-conjugate fiducial spinors constitute a subset of all possible fiducial spinors, they nevertheless form an important subset because of the simple relation between complex conjugates and charge conjugates (just as in a Majorana representation) and between complex conjugates and

Pauli adjoints in the redundant tensorial representation.[17]

### K. Ordinary Action Functional Applies to the Redundant Tensorial Representation of the Dirac Equation

The tensorial equation (41) can be derived as the Euler–Lagrange equation of the *usual* action functional[10]

$$I = -\int \bar{\psi}(\tfrac{1}{2}\gamma^{\mu}\overleftrightarrow{\partial}_{\mu} + m)\psi \, dx, \tag{50}$$

where $A\overleftrightarrow{\partial}_{\mu}B \equiv A(\partial_{\mu}B) - (\partial_{\mu}A)B$. We first insert into (50) the expressions for $\bar{\psi}$ and $\psi$ in Eqs. (48b) and (1), respectively. Straightforward algebraic manipulation leads to the form

$$I = -\tfrac{1}{2}i \int [\bar{f}^{e}{}_{\mu}{}^{\tau}c^{\mu}\overleftrightarrow{\partial}_{\alpha}\bar{f}_{\tau}{}^{\alpha} + f^{e}{}_{\mu}{}^{\tau}c^{\mu}\overleftrightarrow{\partial}_{\alpha}f_{\tau}{}^{\alpha} + 2mf^{e}{}_{\sigma}{}^{\tau}c_{\mu}{}^{\sigma}f_{\tau}{}^{\mu}] \, dx, \tag{51}$$

in which the invariance of the first two terms under a duality rotation ($\gamma_5$ rotation) is clear. If we now make a variation of $I$—not with respect to $\bar{\psi}$—but with respect to $f^{e\sigma\tau}$, then the resultant Euler–Lagrange equation is (41), as desired. Since the action Eq. (50) and thus Eq. (51) is invariant under both $c$- and $F$-gauge transformations, this leads, in the usual fashion, to equations of motion that undergo covariant gauge transformation, as we have already noted. A variety of other quantities can be deduced from the action, such as the stress energy tensor, the current, etc. All of these are simple translations of well-known spinor forms to a redundant tensorial form, as was the action functional itself.

### L. Fermi Nature of Quantum Theory

Although our principal concern here is directed toward $c$-number spinors, a few remarks may be in order regarding the quantum theory lest the reader take our tensorial description to imply that Bose quantization is required. This is, of course, not the case, and the present formalism is completely consistent with Fermi quantization with a positive-definite Hilbert-space metric.

For $x$–$y$ spacelike, the only nonvanishing anti-commutator is

$$\{\psi(x), \bar{\psi}(y)\gamma^{\mu}\} = i\delta^{\mu}(x, y), \tag{52}$$

the term $-\bar{\psi}(y)\gamma^{\mu}$ being the canonically conjugate spinor to $\psi(y)$.[10] Obviously it is $f_{\mu\nu}(x)$ that becomes the quantized operator in our representation, while $v$ and of course $\gamma^{\mu\nu}$ remain as in the $c$-number theory.

It would at first appear natural to set the anti-commutator

$$\{f_{\lambda\nu}(x), \Pi^{\alpha\beta\mu}(y)\}, \tag{53}$$

between $f_{\lambda\nu}(x)$ and its conjugate momentum

$$\Pi^{\alpha\beta\mu} = i(f^{c\alpha\beta}c^{\mu} + f^{e}{}_{\sigma}{}^{[\alpha}g^{\beta]\mu}c^{\sigma} + f^{c\mu[\alpha}c^{\beta]}), \tag{54}$$

proportional to

$$\delta_{\lambda\nu}{}^{\alpha\beta} \equiv \delta_{\lambda}{}^{[\alpha}\delta_{\nu}{}^{\beta]}.$$

However, the anticommutator (53) cannot rigorously be proportional to $\delta_{\lambda\nu}{}^{\alpha\beta}$ because the latter factor does not vanish on contraction with $c_{\alpha\beta}$ or $\bar{c}_{\alpha\beta}$ as does $\Pi^{\alpha\beta\mu} \equiv -\bar{\psi}\gamma^{\mu}\gamma^{\alpha\beta}v$ [see Eq. (4b)].

To overcome this difficulty we may exploit the $c$-gauge freedom in defining $f_{\lambda\nu}$ and adopt the expression

$$\{f_{\lambda\nu}(x), \Pi^{\alpha\beta\mu}(y)\} = -i\delta^{\mu}(x, y)[\delta_{\lambda\nu}{}^{\alpha\beta} + d^{-2}(c_{\lambda\nu}J^{\alpha\beta} - \bar{c}_{\lambda\nu}\bar{J}^{\alpha\beta})], \tag{55}$$

where $J^{\alpha\beta} = b^{[\alpha}c^{\beta]}{}_{\sigma}b^{\sigma}$, $d \equiv b^{\alpha}c_{\alpha} \neq 0$. The term in square brackets in (55) vanishes on contraction with either $c_{\alpha\beta}$ or $\bar{c}_{\alpha\beta}$, and is an idempotent matrix with trace four. In any diagonal form, the factor in square brackets has nonnegative terms, and thus a representation of the quantum operators in (55) exists in a Hilbert space with positive-definite metric. This would not be the case if terms of an intrinsically opposite sign had occurred on the right side of (55).[18]

It is of course not uncommon for the commutation relations between nonobservables to depend on gauge properties of the field. Such a situation occurs in the quantization of the four-vector potential of the radiation field, and the expression above provides yet another example of this phenomenon.

## III. GENERALLY COVARIANT GAUGE TRANSFORMATIONS; REPRESENTATIONS OF THE GENERAL RELATIVISTIC DIRAC EQUATION

As soon as general coordinate transformations are permitted, the metric—and, as a consequence, the Dirac gamma matrices—become space dependent. There is no longer any reason to require the fiducial

[17] The author thanks Professor A. H. Taub for suggesting the use of Lorentz-covariant fiducial tensors based on a general spinor $v$ rather than restricting attention to a subset thereof composed only of real fiducial tensors for which $v = v^{c}$. Complex conjugation and reality properties do not enter the representation properties of Eq. (1) at all, and only arise in the discussion of charge-conjugate equations, and the like.

[18] W. Pauli, Progr. Theoret. Phys. (Kyoto) 5, 526 (1950).

spinor to be constant, and the appropriate generalization of (1) reads

$$\psi(x) = \tfrac{1}{2} f_{\mu\nu}(x)\gamma^{\mu\nu}(x)v(x). \tag{56}$$

### A. Algebraic and Differential Covariants; The *F*-Affine Connection

It is clear that an analysis parallel to the preceding section can be carried out in which each *algebraic* expression verified there continues to hold true at *each point* $x$. Such a generalization extends directly to space-dependent $c$- and $F$-gauge transformations as well. In this way all algebraic relations of Sec. II become generally covariant under both coordinate and gauge transformations. However, an appropriate definition must be found for generally covariant differential operators. For the tensorial properties, it suffices to use the conventional covariant derivative based on the Christoffel symbols, and which we denote by a semicolon. To maintain covariance under space-dependent $F$-gauge transformations requires us to introduce another affine connection—say, the $F$-affine—analogous to the well-known spin affine.[19] We define the (complete) covariant derivative of $f_{\alpha\beta}$ by

$$\nabla_\mu f_{\alpha\beta} = f_{\alpha\beta;\mu} - \tfrac{1}{2}\Gamma_{\alpha\beta}{}^{\sigma\tau}{}_\mu f_{\sigma\tau}, \tag{57}$$

where $\Gamma_{\alpha\beta}{}^{\sigma\tau}{}_\mu$ is a tensor which undergoes an inhomogeneous $F$-gauge transformation in order that $(\nabla_\mu f_{\alpha\beta})^F = \nabla_\mu(f_{\alpha\beta}{}^F)$. Since $y^{\alpha\beta}$ transforms contragrediently to $f_{\alpha\beta}$ according to (21), it follows that

$$\nabla_\mu y^{\alpha\beta} = y^{\alpha\beta}{}_{;\mu} + \tfrac{1}{2} y^{\sigma\tau}\Gamma_{\sigma\tau}{}^{\alpha\beta}{}_\mu. \tag{58}$$

As an analog of the usual "geometrical" argument which leads to the Christoffel connection, we require that

$$\tfrac{1}{4} f_{\sigma\tau} M^{\sigma\tau}{}_{\alpha\beta} e^{\alpha\beta} = i\hat{\psi}\varphi$$

remain invariant under a parallel displacement of the tensors $f_{\alpha\beta}, e_{\alpha\beta} \in \mathfrak{F}$. There follows the relation

$$\nabla_\mu M^{\sigma\tau}{}_{\alpha\beta} = 0, \tag{59}$$

which on contraction leads, according to (16), to

$$\nabla_\mu c^{\sigma\tau} = 0; \tag{60}$$

and which, in turn, with the aid of (8a), implies that

$$\nabla_\mu c^\sigma = 0. \tag{61}$$

When we further require invariance of

$$\tfrac{1}{8} f^*_{\sigma\tau} C^{\sigma\tau}{}_{\nu\lambda} M^{\nu\lambda}{}_{\alpha\beta} f^{\alpha\beta} = i\bar{\psi}\psi$$

under parallel displacement of the tensor $f_{\alpha\beta} \in \mathfrak{F}$, we obtain

$$\nabla_\mu (C^{\sigma\tau}{}_{\nu\lambda} M^{\nu\lambda}{}_{\alpha\beta}) = 0,$$

which, with (59), leads to

$$\nabla_\mu C^{\sigma\tau}{}_{\alpha\beta} = 0. \tag{62}$$

Equations (59) and (62), which, respectively, state

$$M^{\sigma\tau\alpha\beta}{}_{;\kappa} + \tfrac{1}{2} M^{\mu\nu\alpha\beta}\Gamma_{\mu\nu}{}^{\sigma\tau}{}_\kappa + \tfrac{1}{2} M^{\sigma\tau\mu\nu}\Gamma_{\mu\nu}{}^{\alpha\beta}{}_\kappa = 0, \tag{63a}$$

$$C^{\sigma\tau}{}_{\alpha\beta;\kappa} + \tfrac{1}{2} C^{\mu\nu}{}_{\alpha\beta}\Gamma^*_{\mu\nu}{}^{\sigma\tau}{}_\kappa - \tfrac{1}{2} C^{\sigma\tau}{}_{\mu\nu}\Gamma_{\alpha\beta}{}^{\mu\nu}{}_\kappa = 0, \tag{63b}$$

enable us to examine the $F$-affine in a greater depth. For example, when the fiducial tensors are real, $C^{\mu\nu}{}_{\alpha\beta} = \delta^{\mu\nu}{}_{\alpha\beta}$; and (63b) reduces simply to

$$\Gamma^*_{\alpha\beta}{}^{\sigma\tau}{}_\kappa = \Gamma_{\alpha\beta}{}^{\sigma\tau}{}_\kappa, \quad \text{if} \quad v = v^c. \tag{64}$$

Thus (63b) can be viewed as a generalized reality condition on the $F$ affine. A solution of (63a) may be found with the aid of a tensor $U_{\alpha\beta\sigma\tau} = -U_{\sigma\tau\alpha\beta}$ that fulfills the relation

$$M_{\mu\nu}{}^{\theta\lambda} = \tfrac{1}{4} M_{\mu\nu}{}^{\alpha\beta} U_{\alpha\beta}{}^{\sigma\tau} M_{\sigma\tau}{}^{\theta\lambda}, \tag{65}$$

such as

$$U_{\alpha\beta}{}^{\sigma\tau} = d^{-2}\delta_{[\alpha}{}^{[\sigma}(b_{\beta]}c^{\tau]}{}_\mu - c_{\beta]\mu}b^{\tau]})b^\mu, \tag{66}$$

$d \equiv c_\alpha b^\alpha \neq 0$. It readily follows that

$$\Gamma_{\mu\nu}{}^{\alpha\beta}{}_\kappa = -\tfrac{1}{4} U_{\mu\nu}{}^{\sigma\tau}{}_{;\kappa} M_{\sigma\tau}{}^{\alpha\beta} - \tfrac{1}{2} U_{\mu\nu}{}^{\sigma\tau} M_{\sigma\tau}{}^{\alpha\beta}{}_{;\kappa} \tag{67}$$

is a particular solution of (63a). Any other solution differs from (67) by a homogeneous term, $H_{\mu\nu}{}^{\sigma\tau}{}_\kappa$, for which

$$M^{\sigma\tau\mu\nu} H_{\mu\nu}{}^{\alpha\beta}{}_\kappa = M^{\alpha\beta\mu\nu} H_{\mu\nu}{}^{\sigma\tau}{}_\kappa,$$

which expresses a transposition symmetry relative to $M^{\sigma\tau\alpha\beta}$. The general solution is of the form

$$H_{\mu\nu}{}^{\alpha\beta}{}_\kappa = \tfrac{1}{2} E_{\mu\nu}{}^{\theta\lambda}{}_\kappa M_{\theta\lambda}{}^{\alpha\beta}, \tag{68}$$

where

$$E^{\mu\nu\theta\lambda}{}_\kappa = E^{\theta\lambda\mu\nu}{}_\kappa; \tag{69}$$

there are thus ten such independent "vector" fields since (69) asks for the symmetric part of an essentially $4 \times 4$ matrix. In summary, our relation for the $F$ affine is given by (67) and arbitrary amounts of (68), all being subjected to the reality conditions of (63b), or (64) whenever the latter applies.[20]

The undetermined fields (68) may either be set

[19] W. L. Bade and H. Jehle, Ref. 9.

[20] We emphasize strongly that *none* of the ten vector fields of the form (68) corresponds to the electromagnetic potential. This field has no "geometrical origin" from the present point of view since $i\hat{\psi}\varphi$ can not remain invariant in the presence of the four-vector potential $A_\mu$. The electromagnetic field breaks this invariance (while the invariance of $i\bar{\psi}\psi$ is maintained), and must be added separately with a suitable strength for each field in question.

equal to zero, regarded as external fields other than the gravitational field, or given their own dynamical status and their own equations of motion as, e.g., in the Yang–Mills method. We shall not concern ourselves with the latter question here since analogous questions have been discussed in the literature for the conventional, generally covariant spinor Dirac equation.[21–24]

If the additional fields (68) are set equal to zero then a suitable $F$-affine connection can be readily constructed. It suffices to choose an example with real fiducial tensors and to pick a real $b_\alpha$ field in the formation of $U_{\alpha\beta}{}^{\sigma\tau}$; then the solution (67) satisfies both conditions (63).

### B. Redundant Tensorial Representation of the Generally Covariant Dirac Equation

Although Eq. (59) was determined by a straightforward "geometrical" argument of invariance under parallel transport, the very fundamental equations (60) and (61) were direct consequences thereof. Equation (61) permits us to immediately generalize the redundant tensorial representation of the Dirac equation to a generally covariant form simply as

$$\epsilon_{\sigma\mu\tau\beta}C^{\mu}\nabla_{\alpha}\bar{f}^{\beta\alpha} + c_{\sigma}\nabla_{\mu}f^{\mu}{}_{\tau} - c_{\tau}\nabla_{\mu}f^{\mu}{}_{\sigma} + m(c_{\sigma}{}^{\mu}f_{\tau\mu} - c_{\tau}{}^{\mu}f_{\sigma\mu}) = 0. \quad (70)$$

Indeed, Eq. (70) is the Euler–Lagrange equation obtained from the action functional (51) after the partial derivative is replaced by the generally covariant derivative and after the introduction of the invariant volume element. It is noteworthy that the Christoffel symbols do not explicitly appear in the generally covariant derivatives in (70) because of the simplification that arises for divergences of skew tensors; the Christoffel symbols do appear implicitly in the $F$-affine connection. Equation (60) would be important in discussing higher derivatives of (70) [such as the analog of the special relativistic equation $(\Box - m^2)\psi = 0$]. We emphasize that the redundant tensorial representation equations (60) and (61) play a simplifying role in the Dirac equation similar to the conventional assumption $\nabla_\mu\gamma^\sigma = 0$,[25] but which differ from this latter equation in that they arise from "geometrical" invariance arguments.

It would be interesting to trace out further the dynamics of the undetermined fields in the $F$ affine, e.g., in the manner of Leutwyler,[24] and see to what extent, if any, these boson fields of "geometrical origin" might bear on questions in elementary-particle physics.

#### ACKNOWLEDGMENTS

It is a pleasure to thank Professor J. A. Wheeler for the initial stimulus to study spinor fields with the ultimate hope of their incorporation into the general scheme of Geometrodynamics. While this paper does not solve so ambitious a problem, it may be regarded as a possible step in the right direction since it does demonstrate a simple and feasible tensorial description of spinor fields in a linear, algebraic fashion that maintains the time-tested equations of motion and interaction forms of spinor fields intact. A portion of this work was carried out during the author's very pleasant stay in Bern, and I sincerely thank Professor A. Mercier, not only for the kind hospitality shown me during my stay in the Institut für Theoretische Physik, but by his continued interest in this and in related questions of theoretical physics.

[21] O. Klein, Arkiv. Fysik. **17**, 517 (1960).
[22] A. M. Brodskii, D. Ivanenko, and G. A. Sokolik, Zh. Eksperim. i Teor. Fiz. **41**, 1307 (1961) [English transl.: Soviet Physics—JETP **14**, 930 (1962)].
[23] A. Peres, Nuovo Cimento Suppl. **24**, 389 (1962).
[24] H. Leutwyler, Nuovo Cimento **26**, 1066 (1962).
[25] See W. L. Bade and H. Jehle; D. R. Brill and J. A. Wheeler, Ref. 9.

# A Proof that the Free Energy of a Spin System is Extensive*

ROBERT B. GRIFFITHS†
*Department of Physics, University of California, San Diego, La Jolla, California*
(Received 17 March 1964; final manuscript received 5 May 1964)

The free energy obtained from the canonical partition function for a finite spin system possesses a certain convexity property, of which theorems by Peierls and Bogoliubov are particular applications. This property is used in proving the following result: Consider a regular lattice of spins in the form of a parallelepiped (in two dimensions a parallelogram, in one dimension a linear chain). The free energy of the system divided by the number of spins approaches a definite limit as the linear dimensions of the system become infinite. The limit is not influenced by certain common types of boundary conditions. A similar result, but with convergence understood in a weaker sense, holds for derivatives of the free energy such as entropy, magnetization, and specific heat. In the proof it is necessary to assume that the Hamiltonian has the translational symmetry of the spin system, and that long-range interactions decrease sufficiently rapidly with the distance $r$ between spins. (For example, as $r^{-3-\epsilon}$ with $\epsilon > 0$ for interactions between pairs of spins in 3 dimensions.)

## I. INTRODUCTION

SPIN systems, such as the Heisenberg and Ising models of ferromagnetism, provide some of the simplest applications of quantum statistical mechanics, and a study of them has already led to insights into the behavior of more complex systems such as classical gases.

In this paper we prove rigorously a property often assumed to be "intuitively obvious": that the free energy of a spin system in a regular lattice, as defined by the cononical partition function, is extensive, that is, proportional to the size of the system for a large system. We assume the system is in the shape of a parallelepiped, though other simple shapes could be handled by the same techniques. The proof requires that the Hamiltonian have the translation symmetry of the lattice and the interaction terms decrease sufficiently rapidly with distance between the spins. The same extensive property is possessed, though in a somewhat weaker sense, by quantities, such as the entropy and magnetization, which are expressed as derivatives of the free energy with respect to a parameter. The results are independent of boundary conditions of the usual kind employed in calculations.

The proof utilizes a convexity property of the free energy, of which theorems by Peierls and Bogoliubov are particular applications. This property, of some interest in itself, is discussed in Sec. II.

The actual proof that the free energy is extensive begins in Sec. IV where it is carried out in detail in Sec. IV.A for a linear chain. We feel that to first state and then prove the most general result for three dimensions (summarized in Sec. VI) would conceal the elementary character of the proof within a thicket of geometrical complications. These complications are introduced gradually in Sec. IV.B–D. Section V treats the derivatives of the free energy.

The previous work on this subject of which we are aware treats the partition function of a classical gas,[1] and the results also apply to the spin-$\frac{1}{2}$ Ising model, which may be regarded as a lattice gas.[2] Recently Ruelle[3] has extended his work on classical gases to quantum gases. Van Kampen has independently carried out a proof very similar to ours for the case of a spin system, and we are indebted to him for helpful correspondence.

## II. CONVEXITY OF THE FREE ENERGY

All operators discussed in this section are assumed to be $n \times n$ Hermitian matrices, where $n$ is a fixed finite integer. The "canonical" free energy $F$ associated with a Hermitian matrix $\mathcal{H}$ is defined by

$$F(\mathcal{H}) = -\beta^{-1} \log \operatorname{tr} [e^{-\beta\mathcal{H}}], \tag{1}$$

where tr stands for trace, and $\beta = 1/kT \geq 0$ is the inverse temperature.

Let $\mathcal{H}_0$ and $\mathcal{H}_1$ be any two Hermitian matrices, and $\lambda$ a number between 0 and 1. The free energy defined by (1) has the important *convexity property*:

$$F[(1-\lambda)\mathcal{H}_0 + \lambda\mathcal{H}_1] \geq (1-\lambda)F(\mathcal{H}_0) + \lambda F(\mathcal{H}_1). \tag{2}$$

A closely related inequality is Bogoliubov's theorem:

$$F(\mathcal{H}_0 + \mathcal{H}_1) \leq F(\mathcal{H}_0) + \operatorname{tr} [\mathcal{H}_1 e^{-\beta\mathcal{H}_0}]/\operatorname{tr} [e^{-\beta\mathcal{H}_0}]. \tag{3}$$

* This work was supported in part by the U. S. Office of Naval Research.
† National Science Foundation Postdoctoral Fellow. Present address: Department of Physics, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

[1] L. Van Hove, Physica **15**, 951 (1949); C. N. Yang and T. D. Lee, Phys. Rev. **87**, 404 (1952); L. Witten, *ibid.* **93**, 1131 (1954); D. Ruelle, Helv. Phys. Acta **36**, 183 (1963).
[2] T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
[3] D. Ruelle, Helv. Phys. Acta **36**, 789 (1963).

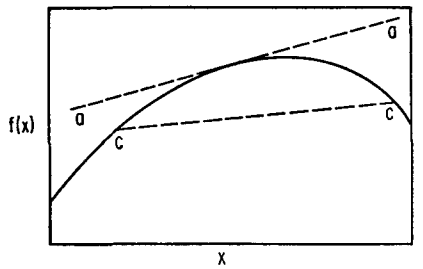


FIG. 1. A curve which is convex upwards has the properties (a) a tangent to the curve at some point lies above the curve; (b) the second derivative (where it exists) is negative; (c) a chord joining two points on the curve lies beneath the curve.

Also, if $x$ is a real parameter, the function

$$f(x) = F(\mathcal{H}_0 + x\mathcal{H}_1) \tag{4}$$

has the property

$$f''(x) \leq 0. \tag{5}$$

(Note that $f$ has a continuous second derivative since the matrices are finite-dimensional.) An immediate consequence of (3) is Peierls' theorem[4]: Let $\mathcal{H}_D$ be the diagonal part of $\mathcal{H}$ in some representation; that is, the matrix obtained by setting all the off-diagonal elements of $\mathcal{H}$ equal to zero. Then

$$F(\mathcal{H}) \leq F(\mathcal{H}_D). \tag{6}$$

Several proofs of Peierls' theorem (6) are found in the literature.[5,6] The result (3) is ascribed to Bogoliubov by Kvasnikov[7] and more than one proof has been published.[6,8] Ruelle[3] has proved (2) and (6) for a certain class of self-adjoint operators in infinite-dimensional Hilbert space. We shall not repeat these proofs, but merely point out the close connection between the inequalities (2), (3), (5), and (6).

From the definitions (1) and (4) it follows that

$$f'(x) = \langle \mathcal{H}_1 \rangle = \mathrm{tr}\,[\mathcal{H}_1 e^{-\beta(\mathcal{H}_0 + x\mathcal{H}_1)}]/\mathrm{tr}\,[e^{-\beta(\mathcal{H}_0 + x\mathcal{H}_1)}] \tag{7}$$

whether or not $\mathcal{H}_0$ and $\mathcal{H}_1$ commute, since the trace of an operator product is invariant under a cyclic permutation of the operators. If now in (3) we replace $\mathcal{H}_0$ by $\mathcal{H}_0 + x_0\mathcal{H}_1$ and $\mathcal{H}_1$ by $(x - x_0)\mathcal{H}_1$, the inequality may be written in the notation of (4) and (7) as

$$f(x) \leq f(x_0) + (x - x_0)f'(x_0), \tag{8}$$

or, in other words, the curve of the function $f(x)$ always lies below the tangent to the curve at $f(x_0)$ (see Fig. 1). This property is one characteristic of a function which is convex upwards,[9] and is equivalent to (5) in the case where the second derivative is continuous.

Another characteristic property of a convex-upwards function is that the chord joining two points of the curve of the function (see Fig. 1) always lies below the curve. In particular, the function

$$g(x) = F[\mathcal{H}_0 + x(\mathcal{H}_1 - \mathcal{H}_0)] \tag{9}$$

is convex upwards by (3) or (5) and therefore, for any $\lambda$ between 0 and 1,

$$g(\lambda) \geq (1 - \lambda)g(0) + \lambda g(1), \tag{10}$$

which is precisely the inequality (2).

The foregoing discussion shows that the inequalities (2), (3), and (5) are really equivalent, and closely correspond to the three properties of a real convex function illustrated in Fig. 1. Peierls' theorem (6) is an immediate consequence of (3) if in the latter inequality $\mathcal{H}_0$ is replaced by $\mathcal{H}_D$, $\mathcal{H}_1$ by $\mathcal{H} - \mathcal{H}_D$, and the traces on the right-hand side of (3) are evaluated in the representation where $\mathcal{H}_D$ is diagonal.

The entropy defined as $-\mathrm{tr}\,[\rho \log \rho]$, where $\rho$ is a density matrix, possesses a convexity property quite analogous to (2).[10]

The convexity property of the free energy has a number of interesting consequences; we shall point out two of them.

(a) The free energy is a convex function of the Hamiltonian and the temperature $T(= 1/k\beta)$ together. Let $T_1$, $T_2$ be nonnegative numbers, $\mathcal{H}_1$ and $\mathcal{H}_2$ Hermitian matrices, and $\lambda$ a number between 0 and 1. Define

$$T = (1 - \lambda)T_1 + \lambda T_2; \quad \mathcal{H} = (1 - \lambda)\mathcal{H}_1 + \lambda\mathcal{H}_2. \tag{11}$$

The inequality

$$F(\mathcal{H}, T) \geq (1 - \lambda)F(\mathcal{H}_1, T_1) + \lambda F(\mathcal{H}_2, T_2) \tag{12}$$

[4] R. E. Peierls, Phys. Rev. 54, 918 (1938).

[5] In addition to Ref. 4 see T. D. Schultz, Nuovo Cimento 8, 943 (1958); D. J. Thouless, *The Quantum Mechanics of Many-Body Systems* (Academic Press Inc., New York, 1961), p. 108; K. Huang, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1963), p. 220; D. Ruelle, Ref. 3.

[6] H. Falk, Physica 29, 1114 (1963); B. Mühlschlegel, Sitzber. Math. Naturw. Kl. Bayer. Akad. Wiss. München 1960, p. 123.

[7] J. Kvasnikov, Doklady Akad. Nauk SSSR 110, 755 (1956); see also V. V. Tolmachev, Doklady Akad. Nauk SSSR 134, 1324 (1960) [English transl.: Soviet Phys.—Doklady 5, 984 (1961)].

[8] J. Czerwonko, Bull. Acad. Polon. Sci. Cl. III 7, 639 (1959); M. Girardeau, J. Math. Phys. 3, 131 (1962).

[9] References on convex functions: R. Courant, *Differential and Integral Calculus* (Interscience Publishers, Inc., New York, 1936) Vol. II, p. 325; G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge University Press, Cambridge, England, 1959), 2nd ed., Chap. III; E. F. Beckenbach, Bull. Am. Math. Soc. 54, 439 (1948).

[10] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1955), p. 390; E. H. Wichmann, J. Math. Phys. 4, 884 (1963).

is easily verified by means of the relationship

$$F(\mathcal{H}, T) = TF(\mathcal{H}/T, 1) \tag{13}$$

together with the inequality (2).

(b) Let $\mathcal{H}_0$ and $\mathcal{H}_1$ be Hermitian matrices, let $\mathcal{H}$ be their sum, and let $\mu$ and $\nu$ be the maximum and minimum eigenvalues of $\mathcal{H}_1$, respectively. The bounds

$$F(\mathcal{H}_0) + \nu \leq F(\mathcal{H}) \leq F(\mathcal{H}_0) + \mu \tag{14}$$

are a consequence of Bogoliubov's theorem.[11] The upper bound is obtained by noting that the second term on the right side of (3) is a weighted average of the eigenvalues of $\mathcal{H}_1$ with positive weights, hence certainly less than $\mu$. The lower bound comes from the same argument upon interchanging the roles of $\mathcal{H}$ and $\mathcal{H}_0$.

The inequality (14) is at the heart of our proof that the free energy is extensive, and hence we shall give an alternate derivation. Let

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

be the $n$ eigenvalues of $\mathcal{H}_0$, and

$$\rho_1 \leq \rho_2 \leq \cdots \leq \rho_n$$

the eigenvalues of $\mathcal{H}$. The matrix $\mathcal{H}_1 - \nu I$, where $I$ is the identity, is clearly a positive matrix (i.e., none of its eigenvalues are negative). Hence the $m$th eigenvalue of

$$\mathcal{H} - \nu I = \mathcal{H}_0 + (\mathcal{H}_1 - \nu I), \tag{15}$$

which is $\rho_m - \nu$, must exceed $\lambda_m$, the $m$th eigenvalue of $\mathcal{H}_0$[12]:

$$\rho_m \geq \lambda_m + \nu. \tag{16}$$

Similarly, one may show that

$$\rho_m \leq \lambda_m + \mu, \tag{17}$$

and therefore

$$e^{-\beta\mu} \sum_{m=1}^{n} e^{-\beta\lambda_m} \leq \sum_{m=1}^{n} e^{-\beta\rho_m} \leq e^{-\beta\nu} \sum_{m=1}^{n} e^{-\beta\lambda_m}, \tag{18}$$

from which (14) follows upon taking logarithms and utilizing the definition (1).

For our purposes, a less precise form of (14) will suffice:

$$|F(\mathcal{H}_0 + \mathcal{H}_1) - F(\mathcal{H}_0)| \leq |\mathcal{H}_1|. \tag{19}$$

By $|\mathcal{H}_1|$ we denote the matrix norm, equal to the maximum of the absolute values of the eigenvalues for a Hermitian matrix, and possessing the important properties

$$|\mathcal{H}_0 + \mathcal{H}_1| \leq |\mathcal{H}_0| + |\mathcal{H}_1|, \tag{20}$$

$$|a\mathcal{H}| = |a|\ |\mathcal{H}|, \tag{21}$$

where $a$ is a real number.

## III. SPIN SYSTEMS: SOME NOTATION

Consider a system of $N$ atoms, each with spin $S$ (integral or half integral). The vector space of interest consists of all complex-valued functions of the $N$ arguments $X_1, X_2 \cdots X_N$ where each argument can take on only the $2S + 1$ values $-S, -S + 1, \cdots, S - 1, S$. The space has dimension $n = (2S + 1)^N$, and operators in the space are $n \times n$ matrices. The statement "$h(1, 2)$ acts only on the coordinates 1 and 2" means the operator has the following property: With every function $\Phi(X_1, X_2)$ there is associated a function $\Phi'(X_1, X_2)$ such that for *any* function $\Psi(X_3, X_4, \cdots, X_N)$ the relation

$$h(1, 2)[\Phi(X_1, X_2)\Psi(X_3, \cdots, X_N)] = \Phi'(X_1, X_2)\Psi(X_3, \cdots, X_N) \tag{22}$$

is satisfied.

Clearly, $h(1, 2)$ may be regarded as a $(2S + 1)^2$-dimensional matrix—which we call, for brevity, the "reduced matrix"—in the space spanned by all functions $\Phi(X_1, X_2)$. More generally, one may consider an operator "acting only on the coordinates $\mathfrak{A}$" where $\mathfrak{A}$ contains $m$ coordinates from the set $X_1, X_2, \cdots, X_N$. We shall also say the operator "involves" each of the $m$ coordinates in $\mathfrak{A}$. The corresponding reduced matrix is of dimension $(2S + 1)^m$.

Let $\mathfrak{A}$ and $\mathfrak{B}$ be two disjoint subsets of the $N$ coordinates, and $h_1$ and $h_2$ two operators acting only on the coordinates $\mathfrak{A}$ and $\mathfrak{B}$, respectively. By tr $[\mathfrak{A}; h_1]$ we mean the sum of the diagonal elements of the reduced matrix $h_1$. The relation[13]

$$\text{tr}\,[\mathfrak{A} \cup \mathfrak{B}; h_1h_2] = \text{tr}\,[\mathfrak{A}; h_1]\ \text{tr}\,[\mathfrak{B}; h_2] \tag{23}$$

will be of use in Sec. IV. If $\mathfrak{X}$ denotes the set of all $N$ coordinates, we shall sometimes write tr[h] in place of tr $[\mathfrak{X}; h]$ if no confusion is likely to arise.

## IV. PROOF THAT THE FREE ENERGY IS EXTENSIVE

### A. Linear Chain with Nearest-Neighbor Interactions. Prototype Hamiltonian and Translational Invariance

We use the term *prototype Hamiltonian* for an operator formally defined on an infinite lattice. For a

[11] We are indebted to N. G. van Kampen for sending us another derivation of (14) which is similar to, but not dependent on, the use of Bogoliubov's theorem.

[12] By an elementary application of the maximum modulus principle. See F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Frederick Ungar Publishing Company, New York, 1955), pp. 238f; R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill Book Company, Inc., New York, 1960), p. 115.

[13] In the space of all functions on $\mathfrak{A} \cup \mathfrak{B}$, $h_1h_2$ is the "direct product" of the two reduced matrices.

linear chain of atoms, each with spin $S$, a possible prototype Hamiltonian is

$$\mathcal{H} = \sum_{i=-\infty}^{\infty} h(i, i+1), \tag{24}$$

where $h(i, i+1)$ is an operator[14] acting only on the coordinates $X_i$ and $X_{i+1}$. The sum (24) may be understood in a formal sense, since the prototype Hamiltonian will be employed for conceptual, not computational, purposes.

Consider a translation of the lattice in which $X_i$ is carried into $X_{i+1}$. Such a translation may be thought of as generating a transformation on the summands in (24) with each $h(i, i+1)$ transformed into an operator $h'(i+1, i+2)$ acting on the coordinates $X_{i+1}$, $X_{i+2}$. If

$$\mathcal{H}' = \sum_{i=-\infty}^{\infty} h'(i, i+1) \tag{25}$$

is formally identical to $\mathcal{H}$ in the sense that to every term in the sum (24) there is one and only one identical term in (25)—or, in other words, if $h(i, i+1) = h'(i, i+1)$ for all $i$—then $\mathcal{H}$ will be called *translationally invariant.*[15]

The operator

$$\mathcal{H}(N) = \sum_{i=1}^{N-1} h(i, i+1) \tag{26}$$

is obtained by discarding from the prototype Hamiltonian (24) all terms which act on spin coordinates outside the set $X_1, X_2, \cdots X_N$. The norm[16]

$$\epsilon = |h(i, i+1)| \tag{27}$$

is independent of $i$ by translational invariance. Note that $\epsilon$ is the same whether one uses the complete or the reduced matrix for $h(i, i+1)$.

Let $\mathfrak{X}$ denote the coordinates $X_1, X_2, \cdots X_N$. The free energy associated with (26) is

$$F(N) = -\beta^{-1} \log \operatorname{tr} [\mathfrak{X}; e^{-\beta\mathcal{H}(N)}]. \tag{28}$$

For convenience, define the normalized free energy

$$f(N) = N^{-1}F(N). \tag{29}$$

We wish to show that $f(N)$ achieves a limiting value as $N$ becomes infinite, for every value of $\beta$ greater than zero.[17] Let the size of the system be doubled to include $2N$ spins. The new Hamiltonian is

$$\begin{aligned}\mathcal{H}(2N) &= \sum_{i=1}^{2N-1} h(i, i+1) \\ &= \mathcal{H}(N) + \mathcal{H}_s(N) + h(N, N+1) \\ &= \mathcal{H}_0 + h(N, N+1),\end{aligned} \tag{30}$$

where $\mathcal{H}(N)$ is defined by (28) and

$$\mathcal{H}_s(N) = \sum_{i=N+1}^{2N-1} h(i, i+1). \tag{31}$$

Let $\mathfrak{Y}$ denote the coordinates $X_{N+1}, \cdots, X_{2N}$. Define $F_0$ by

$$\begin{aligned}e^{-\beta F_0} &= \operatorname{tr} [\mathfrak{X} \cup \mathfrak{Y}; e^{-\beta\mathcal{H}_0}] \\ &= \operatorname{tr} [\mathfrak{X}; e^{-\beta\mathcal{H}(N)}] \operatorname{tr} [\mathfrak{Y}; e^{-\beta\mathcal{H}_s(N)}] \\ &= (\operatorname{tr} [\mathfrak{X}; e^{-\beta\mathcal{H}(N)}])^2;\end{aligned} \tag{32}$$

whence it follows that

$$F_0 = 2F(N). \tag{33}$$

Now since

$$F(2N) = -\beta^{-1} \log \operatorname{tr} [\mathfrak{X} \cup \mathfrak{Y}; e^{-\beta\mathcal{H}(2N)}], \tag{34}$$

one may show by means of (30), (27), (19), and (33) that

$$|F(2N) - 2F(N)| \leq \epsilon, \tag{35}$$

or, dividing both sides by $2N$,

$$|f(2N) - f(N)| \leq \epsilon/2N. \tag{36}$$

Next consider a chain of length $MN$, where $M$ is an integer.[18] The chain may be split into $M$ sections of length $N$, and the Hamiltonian written as

$$\mathcal{H}(MN) = \sum_{p=0}^{m-1} \mathcal{H}_p(N) + \sum_{p=1}^{m-1} h(pN, pN+1), \tag{37}$$

where

$$\mathcal{H}_p(N) = \sum_{i=1}^{N-1} h(pN+i, pN+i+1) \tag{38}$$

is the Hamiltonian for the $(p+1)$th section of length $N$. In analogy with (35) and by use of (20), one obtains the result

$$|F(MN) - MF(N)| \leq (M-1)\epsilon < M\epsilon, \tag{39}$$

or, dividing both sides by $MN$,

$$|f(MN) - f(N)| < \epsilon/N. \tag{40}$$

One may interchange the roles of $M$ and $N$:

$$|f(MN) - f(M)| < \epsilon/M, \tag{41}$$

[14] All operators are assumed to be Hermitian.

[15] In most applications, the translational invariance of the spin Hamiltonian (or lack thereof) is obvious from inspection.

[16] The norm exists provided that all the matrix elements involved are finite; that is, there is no "rigid" coupling between different spins.

[17] At $\beta = 0$ the free energy of a finite system is undefined. However, $\beta f(N)$ is well defined and continuous in the vicinity of $\beta = 0$ and converges to the well-defined function $\beta f$ as $N$ becomes infinite.

[18] I am indebted to Dr. V. Celli for suggesting an important simplification of the proof at this point.

and combine (40) and (41) to obtain

$$|f(M) - f(N)| < \epsilon(N^{-1} + M^{-1}). \quad (42)$$

Therefore $f(N)$ is a Cauchy sequence with a limit $f$, and

$$|f - f(N)| < \epsilon/N. \quad (43)$$

This completes the proof that $f(N)$ converges to a limit in the case of a linear chain with nearest-neighbor interaction. The proof, in essence, is nothing but a precise statement of the notion that "surface" energies may be neglected compared to "volume" energies for a large system. The generalization to more complex cases is almost obvious.

The Hamiltonian (26) corresponds to "periodic boundary conditions" if we add to it the term $h(N, 1)$. By means of (19), one may show that the new normalized free energy differs from (29) by at most $\epsilon/N$, a negligible quantity as $N$ becomes infinite. Similar arguments may be used for the boundary conditions where one or both of the end spins are held "fixed."

### B. Two-Dimensional Lattice: Interactions of Finite Range

Consider a regular, two-dimensional crystal arranged in a lattice with primitive translation vectors **a** and **b**, and a finite number of different spins (which may have different values of $S$) in each unit cell. The prototype Hamiltonian $\mathcal{H}$ shall have the following properties:

(a) $\mathcal{H}$ is translationally invariant;

(b) it consists of a sum of terms, each of which acts on a group of spins no two of which are separated by a distance greater than a constant $r$, where $r$ does not depend on the term considered;

(c) there are only a finite number of terms acting on a given spin.

Suppose there are $m$ spins in a particular unit cell Functions of the $m$ coordinates $X_1, X_2, \cdots, X_m$ may be regarded as functions of a single "super spin" coordinate $Y$ which takes on

$$\prod_{i=1}^{m} (2S_i + 1)$$

values if the coordinate $X_i$ takes on $(2S_i + 1)$ values. Hence without loss of generality we may suppose that there is only one spin per unit cell. The translational invariance of the Hamiltonian and also Properties (b) and (c) are preserved if we suppose the lattice to be distorted into one in which **a** and **b** are orthogonal unit vectors. The constant $r$ in (b) may have to be redefined, but still remains finite. In other words we may, for the purpose of the proof, assume that the prototype Hamiltonian is defined for a simple square lattice with one spin per lattice site.

Consider a rectangle of spins with sides $MP$ and $NQ$, where $M$, $N$, $P$, $Q$ are positive integers. The rectangle containing $MPNQ$ spins may be thought of as composed of $PQ$ smaller rectangles, each with sides $M$ and $N$. We define $\mathcal{H}(MP, NQ)$ as the sum of all terms in the prototype Hamiltonian which involve only spins located in the large rectangle. The Hamiltonian for the $i$th smaller rectangle, $\mathcal{H}_i(M, N)$, is similarly defined. We may write

$$\mathcal{H}(MP, NQ) = \sum_{i=1}^{PQ} \mathcal{H}_i(M, N) + \mathcal{H}', \quad (44)$$

where $\mathcal{H}'$ contains all terms in $\mathcal{H}(MP, NQ)$ which involve spins in two or more of the smaller rectangles.

Let $h_1, h_2, \cdots h_m$ [there are at most a finite number, by Property (c)] be all the terms in the prototype Hamiltonian which act on the first spin. Define[16]

$$\epsilon = \sum_{i=1}^{m} |h_i|. \quad (45)$$

We may find a bound for $\mathcal{H}'$ by noting that each term in $\mathcal{H}'$ must involve some spin within a distance $r-1$ of the boundary of one of the smaller rectangles. There are at most $2(M + N)PQr$ such spins in all the $PQ$ smaller rectangles. Hence, using a very liberal estimate, we find

$$|\mathcal{H}'| \leq 2\epsilon(M + N)PQr, \quad (46)$$

which combined with (44) and (19) yields the result

$$|f(MP, NQ) - f(M, N)| \leq 2\epsilon r(M^{-1} + N^{-1}). \quad (47)$$

By an argument analogous to (40) through (43) one easily shows that $f(M, N)$ approaches a limiting value $f$ as *both* $M$ and $N$ increase to infinity, and

$$|f - f(M, N)| \leq 2\epsilon r(M^{-1} + N^{-1}). \quad (48)$$

Once again it is clear that the boundary conditions have negligible effect for large rectangles if they only involve a modification of terms in the Hamiltonian near the boundary of the rectangle.

### C. Two-Dimensional Lattice: Interactions of Unlimited Range

As in part B of this section, we need only consider a simple square lattice with one spin per lattice site. The prototype Hamiltonian $\mathcal{H}$ shall have the following properties:

(a) Translational invariance.

(b) $\mathcal{H}$ consists of a sum of terms. Any given term $h$ acts on a finite set of spins, and the maximum distance between two spins in the set is $r(h)$, the "range" of $h$.

(c) For any number $R < \infty$, there are only a finite number of terms in $\mathcal{H}$ with range less than $R$ involving any given spin coordinate.

(d) Let $h_i$ for $i = 1, 2, 3, \cdots$ be all the terms in $\mathcal{H}$ involving the coordinate $X_1$. There is a number $\delta > 0$ and a constant $\mathcal{C} < \infty$ such that, given any positive number $R$,

$$\sum_{r(h_i) \geq R} |h_i| < \mathcal{C}R^{-\delta}, \tag{49}$$

where the sum extends over all $h_i$ with range exceeding $R$.

Conditions (a)–(d) may also be applied in the case of 1- or 3-dimensional lattices. Property (d) is the requirement that terms of long range go to zero sufficiently rapidly. A case often met in practice is where $h_i = h(1, i)$ acts only on the coordinates $X_1$ and $X_i$, in which case it is sufficient to require that

$$|h(1, i)| \leq D/r_i^{d+\delta}, \qquad i \neq 1, \tag{50}$$

where $r_i$ is the distance between the first and $i$th spins, $D$ a constant, and $d(= 1, 2, \text{or } 3)$ the dimensionality of the lattice.

As in part B of this section, let $\mathcal{H}(MP, NQ)$ consist of all terms in the prototype Hamiltonian which involve only spins located in an $MP$ by $NQ$ rectangle. The $\mathcal{H}_i(M, N)$ are similarly defined. The term $\mathcal{H}'$ in (44) has a bound of the form (see Appendix B):

$$(PQ)^{-1} |\mathcal{H}'| \leq \begin{cases} \mathcal{C}'(MN^{1-\delta} + NM^{1-\delta}) & \text{for } 0 < \delta < 1, \\ \mathcal{C}'(M \log N + N \log M) & \text{for } \delta = 1, \\ \mathcal{C}'(M + N) & \text{for } \delta > 1, \end{cases} \tag{51}$$

where $\mathcal{C}'$ depends on the constant $\mathcal{C}$ appearing in (49) and also on $\delta$, but not on $M$, $N$, $P$, or $Q$. Using the bound (51), it is easily shown that $f(M, N)$ approaches a limiting value as both $M$ and $N$ increase to infinity.

The use of periodic boundary conditions presents some difficulties when the range of interaction is not bounded, due to the terms in the Hamiltonian in which a spin "interacts with itself." Our method of proof is not directly applicable to this case, and we have not investigated under what circumstances the free energy approaches the limiting value obtained above using "nonperiodic" boundary conditions.

### D. Three-Dimensional Lattices

The techniques of parts A–C of this section may be applied immediately to three-dimensional lattices with only a few obvious alterations. Provided the prototype Hamiltonian satisfies Conditions (a)–(c) of part B (interactions of finite range) or (a)–(d) of part C (interactions with unbounded range), the free energy $f(M, N, P)$ of a parallelepiped $M$ by $N$ by $P$ unit cells on a side approaches a limiting value as all three integers approach infinity.

## V. DERIVATIVES OF THE FREE ENERGY

Suppose that the prototype Hamiltonian of a spin system is a function of a real parameter $x$ and satisfies the conditions of Sec. IV.B or C for all $x$ in some interval. Or, instead of a parameter in the Hamiltonian, $x$ could be the temperature. We have shown that the normalized free energy for a system of $N$ spins, $f(N, x)$, converges to a definite limit as $N$ becomes infinite[19]:

$$f(x) = \lim_{N\to\infty} f(N, x). \tag{52}$$

Consider the quantity (we assume the right-hand side is defined)

$$g_n(N, x) = (d/dx)^n f(N, x). \tag{53}$$

Is it true that

$$\lim_{N\to\infty} g_n(N, x) = (d/dx)^n f(x)? \tag{54}$$

For instance, let $x$ be the magnetic field and $-g_1(x)$ the magnetization per spin. Equation (54) (if true) asserts that the magnetization is extensive, and its limiting value for a large system is the derivative of the limiting value of the free energy.

We shall show that Eq. (54) holds when certain restrictions are placed on the Hamiltonian, but it must be understood in a weaker sense than (52); that is, the convergence is not necessarily pointwise. Slightly sharper results are possible for $n = 1$, which we discuss first.

Let the prototype Hamiltonian be of the form

$$\mathcal{H}(x) = \mathcal{H}_1 + x\mathcal{H}_2, \tag{55}$$

where both $\mathcal{H}_1$ and $\mathcal{H}_2$ are prototype Hamiltonians satisfying the conditions of Sec. IV.B or C. For a system of $N$ spins, the Hamiltonian is a linear

[19] We assume here and in the following discussion that the crystal is in the form of a parallelepiped (parallelogram or line) and that the linear dimensions become infinite as $N \to \infty$.

function of $x$ and $f(N, x)$ a convex (upwards) function of $x$ by (5). Pointwise convergence (52) guarantees that $f(x)$ is also convex, and therefore (see Appendix A)

$$\lim_{N\to\infty} df(N, x)/dx = df(x)/dx \tag{56}$$

at every point where the right side is continuous. Since $df/dx$ is monotone decreasing, it is discontinuous at most at a countable number of points.

Note that this argument also applies to the (normalized) energy and entropy, which may both be expressed in terms of the free energy and its first derivative with respect to the temperature. The free energy is a convex function of the temperature [Sec. II, Remark (a)].

Next consider the case where $\mathcal{H}(N, x)$ is not necessarily a linear function of $x$, but does possess a continuous first and piecewise continuous second derivative[20] for $x$ in the interval of interest, $[a, b]$, which we shall assume includes the origin. Further assume a bound for the second derivative of the form

$$|d^2\mathcal{H}(N, x)/dx^2| \leq \mathcal{C}N, \tag{57}$$

where $\mathcal{C}$ is independent of $x$ and $N$. When the range of interaction is bounded, a bound on the second derivative of each term in the prototype Hamiltonian is equivalent to (57).

Under the conditions of the preceding paragraph, the expansion

$$\mathcal{H}(N, x) = \mathcal{H}(N, 0) + \mathcal{H}'(N, 0) + \tfrac{1}{2}x^2\mathcal{H}_1(N, x) \tag{58}$$

(we denote the derivative with respect to $x$ by a prime) holds for all values of $x$ in $[a, b]$ and the correction term is bounded by

$$|\mathcal{H}_1(N, x)| \leq \mathcal{C}N. \tag{59}$$

An application of Bogoliubov's theorem (3) together with (59) yields the result

$$f(N, x) \leq f(N, 0) + xf'(N, 0) + \tfrac{1}{2}\mathcal{C}x^2, \tag{60}$$

which implies that the curve of the function

$$f^*(N, x) = f(N, x) - \tfrac{1}{2}\mathcal{C}x^2 \tag{61}$$

lies everywhere below a line tangent to the curve at $x = 0$. A similar result holds for $f^*(N, x)$ at all other points of $(a, b)$. Thus $f^*(N, x)$ is a convex (upwards) function to which one can apply the results of Appendix A with the same conclusion as before: Eq. (56) is satisfied at every point where the right side is continuous.

[20] That is, each matrix element (in some given representation) has these properties.

### Second and Higher-Order Derivatives

The essential problem, as is clear from (52) to (54), is to interchange $\lim (N \to \infty)$ with $d/dx$. Such an interchange is possible, in most cases, if $\lim (N \to \infty)$ is taken in the sense of "ideal functions" or "generalized functions," characterized by their inner products with a class of suitably chosen "test functions" rather than by their values at each point in an interval.[21]

For example, if $\mathcal{H}(N, x)$ is a continuous function of $x$ in the interval $[a, b]$ and satisfies the conditions of Sec. IV C with the constant $\mathcal{C}$ in (49) independent of $x$, then $f(N, x)$ (which is a continuous function of $x$) converges to $f(x)$ uniformly, which implies convergence in the sense of "ideal functions." If the test functions have continuous derivatives of arbitrary order and vanish, together with all their derivatives, at the end points of $[a, b]$, then the $n$th derivative of $f(N, x)$ converges to the $n$th derivative of $f(x)$ in the sense of ideal functions.[22]

The penalty for this freedom in interchanging limiting processes is, of course, that "convergence" and "function" must both be understood in a weaker (or broader) sense than is customary in elementary calculus. To give an example, the (normalized) magnetization $M$ as a function of the magnetic field $H$ is discontinuous at $H = 0$ for the two-dimensional Ising ferromagnet[23] in the limit $N \to \infty$, provided the temperature is less than the transition temperature. The susceptibility, $dM/dH$, is a continuous function of $H$ for finite $N$ but "converges" to an ideal function with the character of a Dirac delta function at $H=0$ in the limit $N \to \infty$.

## VI. SUMMARY

Consider a regular lattice of spins in one, two, or three dimensions with an arbitrary but finite number of spins per unit cell. If the prototype Hamiltonian defined (formally) for the infinite lattice has the translational symmetry of the lattice and also satisfies the other requirements of Sec. IV B or IV C, then the normalized free energy (free energy divided by $N$) for a crystal in the form of a parallelepiped (parallelogram or linear chain) containing $N$ unit cells converges to a limit as the linear dimen-

[21] M. J. Lighthill, *An Introduction to Fourier Analysis and Generalized Functions* (Cambridge University Press, Cambridge, England, 1959); R. Courant, *Methods of Mathematical Physics* (John Wiley & Sons, Inc., New York, 1962) 2nd ed., Vol. II, p. 766; A. Erdélyi in *Modern Mathematics for the Engineer, Second Series,* edited by E. F. Beckenbach (McGraw-Hill Book Company, Inc., New York, 1961), p. 5.

[22] The $n$th derivative of $f(N, x)$ or $f(x)$ may not exist in the usual sense.

[23] G. F. Newell and E. W. Montroll, Rev. Mod. Phys. **25**, 353 (1953).

sions become infinite. The proof was carried out for the case where the Hamiltonian for the finite system contains all terms of the prototype Hamiltonian involving only spins in the $N$ unit cells under consideration. However, other sensible boundary conditions lead to the same result: e.g., if the surface spins are held "fixed," or if periodic boundary conditions are employed when the interactions are of finite range.[24]

The convergence of the normalized free energy guarantees the convergence of its derivatives with respect to temperature or a parameter appearing in the Hamiltonian, but (in general) in a weaker sense than pointwise convergence. (For details and restrictions on the Hamiltonian, see Sec. V.)

## APPENDIX A. THEOREM ON SEQUENCES OF CONVEX FUNCTIONS

Let $f_n(x)$ be a sequence of functions defined on $[a, b]$ which are all convex upwards, and let $g_n(x)$ be the first derivative of $f_n(x)$. If there exists a function $f(x)$ such that

$$\lim_{n\to\infty} f_n(x) = f(x) \tag{A1}$$

for every point $x$ in $[a, b]$, then $f(x)$ is convex upwards, and, furthermore,

$$\lim_{n\to\infty} g_n(x) = g(x), \tag{A2}$$

where $g(x)$ is the first derivative of $f(x)$, and (A2) holds at every point where $g(x)$ is continuous.

The convexity of $f(x)$ follows immediately from (A1) and the definition of a convex function.[9] Therefore $g(x)$, together with each of the $g_n(x)$, is a monotone decreasing function with at most a countable number of jump discontinuities.

Suppose that $g(x)$ is continuous at $x = x_0$, but that (A2) does not hold at this point; in particular, assume that there is a number $\epsilon > 0$ such that

$$g(x_0) - g_m(x_0) > \epsilon \tag{A3}$$

is satisfied for arbitrarily large values of $m$. We shall show that this contradicts (A1). Since $g(x)$ is continuous at $x_0$, there exists a number $\delta > 0$ such that the points $x_0 \pm \delta$ are in $[a, b]$, and

$$|g(x_0) - g(x)| \leq \tfrac{1}{2}\epsilon \tag{A4}$$

provided that

$$|x - x_0| \leq \delta. \tag{A5}$$

Now since $g_m(x)$ as well as $g(x)$ is a monotone decreasing function, (A3) and (A4) imply, for $x$ between $x_0$ and $x_0 + \delta$,

$$g(x) - g_m(x) \geq \tfrac{1}{2}\epsilon. \tag{A6}$$

Integrate both sides of the inequality from $x_0$ to $x_0 + \delta$:

$$[f(x_0 + \delta) - f_m(x_0 + \delta)] + [f_m(x_0) - f(x_0)] \geq \tfrac{1}{2}\delta\epsilon. \tag{A7}$$

But if (A7) holds for arbitrarily large values of $m$, then either at $x_0$ or at $x_0 + \delta$, $f_n(x)$ does not converge to $f(x)$, in contradiction to (A1). A similar argument works when, in place of (A3), we have

$$g_m(x_0) - g(x_0) > \epsilon \tag{A8}$$

for arbitrarily large values of $m$.

The step from (A6) to (A7) uses the fact that a convex function is absolutely continuous[25] and hence equal to the indefinite integral of its derivative.[26]

## APPENDIX B. DERIVATION OF THE BOUND (51) FROM THE INEQUALITY (49)

Consider a particular spin $r$ located in the $j$th $M \times N$ rectangle a distance $m$ (the lattice constant is unity) from the nearest edge. Any term in $\mathcal{H}'$ which involves the spin $r$ must also (by definition) involve at least one spin outside the $j$th rectangle; therefore its range is at least $m + 1$. Thus the sum of the norms of all terms in $\mathcal{H}'$ involving the spin $r$ is less than $\mathcal{C}(m + 1)^{-\delta}$ by (49).

Now in all $PQ$ of the smaller rectangles there are not more than $2(M + N)PQ$ spins at a distance $m$ from the nearest edge of the rectangle to which each spin belongs. Hence a very liberal upper bound on the norm of $\mathcal{H}'$ is provided by

$$2(M + N)PQ \sum_{m=0}^{\frac{1}{2}M} \mathcal{C}(m + 1)^{-\delta} \tag{B1}$$

if we assume (without loss of generality) that $M \leq N$. For the case $0 < \delta < 1$, the corresponding integral provides for the sum in (B1) the bound

$$\mathcal{C}(1 - \delta)^{-1}(\tfrac{1}{2}M + 1)^{1-\delta} \leq \tfrac{1}{2}\mathcal{C}'M^{1-\delta} \tag{B2}$$

for a suitably defined $\mathcal{C}'$. From this the inequality (51) follows. (Note that $M^{2-\delta} \leq MN^{1-\delta}$, since we assumed $M \leq N$.) The cases $\delta = 1$ and $\delta > 1$ may be worked out in similar fashion.

[24] See the remark at the end of Sec. IV.C.

[25] As the reader may easily verify for himself; or see I. P. Natanson, *Theory of Functions of a Real Variable* (Frederick Ungar Publishing Company, New York, 1960), Vol. II, p. 230.

[26] F. Riesz and B. Sz.-Nagy, Ref. 12, p. 50.

# A Theorem on the Determinantal Solution of the Fredholm Equation

IRWIN MANNING
*Nucleonics Division, U.S. Naval Research Laboratory, Washington, D. C.*
(Received 3 April 1964)

For a Fredholm equation with Green's-function-type kernel, a new proof is given of a known theorem relating the Fredholm determinant with the behavior of the solution at the origin. The utility of this theorem in practical calculations is pointed out, as are some of its implications for potential scattering in quantum mechanics. The scattering phase shift is shown to have the property $(\partial/\partial\lambda)\delta(E, \lambda) = -\pi \langle\psi_E| V |\psi_E\rangle$.

## I. INTRODUCTION

CONSIDER the Fredholm equation

$$\psi(x) = f + \lambda K\psi \equiv f(x) + \lambda \int_a^b K(x, s)\psi(s)\, ds. \quad (1)$$

Following Brysk,[1] we shall say that the kernel $K$ is of the Green's function type if it has the form

$$K(x, s) = V(s)f(r_<)g(r_>), \quad (2)$$

where $r_<$ and $r_>$ are, respectively, the lesser and greater of $x$ and $s$. The main content of this paper is a new proof of a known theorem relating, for a kernel of this type, the Fredholm determinant and the quantity $\{\psi\}$, where we have introduced the notation

$$\{\psi\} \equiv \lim_{x\to a} \psi(x)/f(x).$$

The theorem is stated and discussed in Sec. II. A kernel of the Green's function type arises in the partial-wave analysis of quantum scattering, and in Sec. III the theorem is further discussed within that context. We conclude with the proof of the theorem, Sec. IV. The remainder of this section presents notation.

The solution of Eq. (1) is

$$\psi = f + \lambda\Gamma f. \quad (4)$$

The determinantal method[2] expresses the resolvent in the form

$$\Gamma(x, s; \lambda) = D(x, s; \lambda)/d(\lambda), \quad (5)$$

where

$$D(x, s; \lambda) = \sum_{n=0}^{\infty} \lambda^n D_n(x, s), \quad (6)$$

and

$$d(\lambda) = \sum_{n=0}^{\infty} \lambda^n d_n. \quad (7)$$

[1] H. Brysk, J. Math. Phys. 4, 1536 (1963).

In the above, $d_0 = 1$, $D_0 = K$, and for $n \neq 0$,

$$D_n(x, s) = (-)^n(n!)^{-1} \times \int_a^b \cdots \int_a^b \Delta_n(x, s; t_1, \cdots, t_n)\, dt_1 \cdots dt_n, \quad (8)$$

and

$$d_n = -n^{-1} \operatorname{Tr} D_{n-1} \equiv -n^{-1} \int_a^b D_{n-1}(s, s)\, ds, \quad (9)$$

where

$$\Delta_n(x, s; t_1, \cdots, t_n) = \begin{vmatrix} K(x, s) & K(x, t_1) & \cdots & K(x, t_n) \\ K(t_1, s) & K(t_1, t_1) & \cdots & K(t_1, t_n) \\ \cdots & \cdots & \cdots & \cdots \\ K(t_n, s) & K(t_n, t_1) & \cdots & K(t_n, t_n) \end{vmatrix}. \quad (10)$$

The functions $D(x, s; \lambda)$ and $d(\lambda)$ can be shown to be entire functions of $\lambda$ if tr $(K^\dagger K)$ is finite.[2] The particular expansions (8), (9), and (10) further require that tr $K$ exist.[3]

## II. THEOREM ON THE FREDHOLM DETERMINANT

The main result referred to is the theorem

*If $K$ is of the Green's function type, then*

$$d_n = -\{D_{n-1}f\}. \quad (11)$$

The proof presented in Sec. IV utilizes only the definitions (8), (9), and (10); the theorem is therefore valid whenever the right-hand side of (11) exists.

[2] See, for example, S. G. Mikhlin, *Integral Equations* (Pergamon Press, Inc., New York, 1957), or the book of the same title by F. Smithies (Cambridge University Press, New York, 1958).

[3] In quantum scattering, S. Weinberg, Phys. Rev. **131**, 443 (1963), has shown that tr $(K^\dagger K)$ is finite if and only if the potential satisfies $\int_0 |V(r)|^2 r^2\, dr < \infty$ and $\int^\infty |V(r)|^2\, dr < \infty$. We also note that tr $K = J(f)$ so, by Eq. (13), will exist if and only if $\{\psi^{(1)}\}$ is finite, where $\psi^{(1)}(x)$ is the first Born approximation.

This theorem is essentially contained in (or at least very strongly implied by) Eq. (22) of Ref. (1).[4] One easily extracts the theorem from that equation upon noticing that

$$\{K\psi\} = J(\psi) \equiv \int_a^b g(x)V(x)\psi(x)\,dx, \tag{12}$$

so that Eq. (1) gives

$$\{\psi\} = 1 + \lambda J(\psi). \tag{13}$$

When tr $(K^\dagger K)$ is finite the relevant power series are everywhere convergent, and on applying (11) we find

$$\lambda\{Df\} = -\sum_{n=0}^{\infty} \lambda^{n+1} d_{n+1} = -d(\lambda) + 1. \tag{14}$$

Equations (4) and (5) then yield the corollary

*If $K$ is of the Green's function type, and* tr $K$ *and* tr $(K^\dagger K)$ *are finite, then*

$$d(\lambda) = \{\psi\}^{-1}. \tag{15}$$

Upon taking the trace of Eq. (6) one finds[2]

$$\text{tr } \Gamma = -(d/d\lambda) \ln d(\lambda), \tag{16}$$

so that the corollary provides a connection between $\{\psi\}$ and the trace of the resolvent:

$$(d/d\lambda) \ln \{\psi\} = \text{tr } \Gamma. \tag{17}$$

Theorem (11) can be of considerable utility in practical calculations. Equivalent to the relation

$$\Gamma = K + \lambda K\Gamma, \tag{18}$$

the Fredholm coefficients obey the recursion relation[2]

$$D_n = d_n K + KD_{n-1}. \tag{19}$$

One therefore has the expansion[5]

$$\psi(x) = f(x) + \lambda\left[\sum_{n=0}^{\infty} \lambda^n u_n(x) \Big/ \sum_{n=0}^{\infty} \lambda^n d_n\right], \tag{20}$$

with $d_0 = 1$, $u_0 = Kf$, $d_1 = -\{u_0\} = -J(f)$; the other terms are obtained by the convenient recursion procedure

$$d_{n+1} = -\{u_n\} = -[J(u_{n-1}) + d_n J(f)], \tag{21}$$

and

$$u_{n+1} = Ku_n + d_{n+1}u_0. \tag{22}$$

[4] On account of an incorrect application of proof by induction made in the last paragraph of that work, Ref. 1 fails to show that the equations there—(22), (23), and (24)—are in fact the determinantal solution. When augmented by the demonstration of the present work, the proof of Ref. 1 can be easily completed.

[5] This expansion is closely related to (and is essentially implied by) Eq. (22) of Ref. 1. The comments associated with our Eqs. (12) and (13) apply here as well.

## III. QUANTUM SCATTERING

In the case of potential scattering, the theorem (11) has long been known, or at least very strongly implied. The Jost function $\mathfrak{F}(-k)$ has the properties[6]

$$\mathfrak{F}(-k) = d(\lambda), \tag{23}$$

and

$$\mathfrak{F}(-k)^{-1} = |\{\psi\}|\, e^{i\delta}, \tag{24}$$

where $\delta$ is the scattering phase shift. One need only add to this a demonstration that $\delta$ is also the phase of $\{\psi\}$ to obtain the corollary (15), and from it (by expanding a power series) the relation (11). We claim for the present proof the advantages of simplicity and generality. Indeed, as the above indicates, the corollary (15) would be of considerable heuristic utility in the presentation of the theory of Jost functions.

The corollary (15) may be of significant theoretical value as well. For example, for potential scattering at energy $E$ with the system Hamiltonian $H = H_0 + \lambda V$, the resolvent is[7]

$$\Gamma = [1/(E - H + i\epsilon)]V, \tag{25}$$

and Eq. (17) becomes

$$\frac{\partial}{\partial\lambda} \ln \{\psi_E\} = \sum_a \langle\psi_a| \frac{1}{E - H + i\epsilon} V |\psi_a\rangle$$

$$= \sum_a \frac{1}{E - E_a + i\epsilon} \langle\psi_a| V |\psi_a\rangle. \tag{26}$$

The imaginary part of this equation is

$$(\partial/\partial\lambda)\delta(E, \lambda) = -\pi\langle\psi_E| V |\psi_E\rangle. \tag{27}$$

In this result, there is implied a sum (partial trace) over all quantum numbers other than energy $(E)$ and angular momentum magnitude $(l)$[8].

## IV. PROOF OF THEOREM (11)

We first show the lemma:

$$D_n(x, s) = (-)^n$$

$$\times \int_{t_1 \le t_2 \le \cdots \le t_n} \cdots \int \Delta_n(x, s; t_1, \cdots, t_n)\, dt_1 \cdots dt_n, \tag{28}$$

[6] For $l = 0$, see R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951). For general $l$, see R. G. Newton, J. Math. Phys. **1**, 319 (1960).

[7] See, for example, M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).

[8] This result could also have been derived from the relation connecting, for a system with closely-spaced but discrete energy levels, the phase shift with the energy shift [J. Schwinger, Phys. Rev. **93**, 615 (1954); B. S. DeWitt, *ibid.* **103**, 1565 (1956); M. Baker, Ann. Phys. (N. Y.) **4**, 271 (1958).]:

$$\Delta E = -\pi^{-1}\delta(E)\, dE.$$

One differentiates with respect to $\lambda$.

and

$$d_{n+1} = (-)^{n+1} \times \int_{t_0 \le t_1 \le \cdots \le t_n} \cdots \int \Delta_n(t_0, t_0; t_1, \cdots, t_n)\, dt_0\, dt_1 \cdots dt_n, \tag{29}$$

where

$$\int_{t_1 \le t_2 \le \cdots \le t_n} \cdots \int dt_1 \cdots dt_n \equiv \int_a^b dt_n \int_a^{t_n} dt_{n-1} \cdots \int_a^{t_3} dt_2 \int_a^{t_2} dt_1.$$

From the definition (10) of $\Delta_n$, we note that interchanging any two of the variables

$$s, t_1, \cdots, t_n \tag{30}$$

results in simultaneously interchanging two rows and two columns, leaving the determinant unchanged. So $\Delta_n(x, s; t_1, \cdots, t_n)$ is invariant under any permutation of the variables (30). Utilizing this invariance, one can easily prove (28) by induction, or we can proceed as follows: Let $\mathcal{S}$ be the integral of $\Delta_n$ over the $n$-dimensional volume contained in the cube $a \le t_i \le b$, $i = 1, \cdots, n$, and $\mathcal{S}_p$ be the integral over the subvolume of this cube that satisfies the further restriction

$$t_{i_1} \le t_{i_2} \le \cdots \le t_{i_n}, \tag{31}$$

where (31) is a permutation of $t_1 \le t_2 \le \cdots \le t_n$. Clearly

$$\mathcal{S} = \sum_p \mathcal{S}_p, \tag{32}$$

where the sum is over all $n!$ permutations (31). Equation (28) follows from the definition (8) upon noting the invariance fo the integral to permutations (31), so that each of the terms $\mathcal{S}_p$ are equal. The demonstration of Eq. (29) follows from the definition (9) in an entirely similar manner.

Using the Green's function property (2), Eq. (28) yields ($s = t_0$),

$$\{D_n f\} = (-)^n \int_a^b dt_0 f(t_0) \int_{t_1 \le \cdots \le t_n} \cdots \int dt_1 \cdots dt_n \times \begin{vmatrix} V(t_0)g(t_0) & V(t_1)g(t_1) & \cdots & V(t_n)g(t_n) \\ K(t_1, t_0) & K(t_1, t_1) & \cdots & K(t_1, t_n) \\ \cdots & \cdots & \cdots & \cdots \\ K(t_n, t_0) & K(t_n, t_1) & \cdots & K(t_n, t_n) \end{vmatrix}. \tag{33}$$

Now, when $t_0 \ge t_1$, the first two lines in the determinant become

$$\begin{matrix} V(t_0)g(t_0) & V(t_1)g(t_1) & \cdots & V(t_n)g(t_n) \\ V(t_0)g(t_0)f(t_1) & V(t_1)g(t_1)f(t_1) & \cdots & V(t_n)g(t_n)f(t_1), \end{matrix}$$

and, because of this proportionality, the determinant vanishes. Bringing $f(t_0)$ into the determinant as a common factor of each of the elements in the first line, (33) thus becomes

$$\{D_n f\} = (-)^n \int_{t_0 \le t_1 \le \cdots \le t_n} \cdots \int dt_0\, dt_1 \cdots dt_n \times \begin{vmatrix} K(t_0, t_0) & K(t_0, t_1) & \cdots & K(t_0, t_n) \\ K(t_1, t_0) & K(t_1, t_1) & \cdots & K(t_1, t_n) \\ \cdots & \cdots & \cdots & \cdots \\ K(t_n, t_0) & K(t_n, t_1) & \cdots & K(t_n, t_n) \end{vmatrix}, \tag{34}$$

which, being just the negative of Eq. (29), completes the proof.


## ACKNOWLEDGMENT

I should like to thank S. Teitler for a helpful discussion.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# Characters of Irreducible Representations of the Simple Groups. I. General Theory

J.-P. ANTOINE AND D. SPEISER
*Université de Louvain, Centre de Physique Nucléaire, Héverlé, Belgium*
(Received 20 September 1963; final manuscript received 19 May 1964)

New formulas for the characters of irreducible representations of simple groups are presented. They yield the character directly as a sum instead of a quotient as does Weyl's formula. The procedure is purely geometrical and is based on the properties of the regular lattice associated with every simple group in the "global" theory of Lie groups due to Hopf and Stiefel.

## I. INTRODUCTION

THE conjectured global symmetries brought Lie groups again into the center of interest of theoretical physics. Nevertheless, to many physicists the mathematical questions associated with this seem still somewhat esoteric. This is partly due to the fact that the mathematical literature on Lie groups is focused on fundamental theorems rather than on detailed properties which the physicist needs. For instance, it is not easy to get a general view over the different representations of a particular group.

What makes the theory of representations richer and more complicated in the general than in the special case $SU_2 = O_3$ with which every physicist is so well acquainted, is the appearance of "multiple weights." Whereas within a particular representation of $SU_2$ all eigenvalues of $L_z$ are simple, the analogous statements for groups of rank $\geq 2$ are not true. Thus the determination of the "*multiplicities*" of the weights is a central problem to which this investigation is devoted.

An irreducible representation of a Lie group is completely determined, apart from equivalence, by its character, and the latter is given by the well-known formula obtained by Weyl[1] for (simple) classical groups:

$$\chi(D) = X(D)/\Delta, \tag{1}$$

$\chi(D)$ is the character of the representation $D$; the two expressions $X$ and $\Delta$ are homogeneous alternating sums of exponentials and may be obtained without great difficulty by using the prescriptions stated below. From this formula one easily derives the character of any particular element of the group. But if one wants to know the character in an explicit form, in particular if one wants to obtain the multiplicity of every weight, it is indispensable to carry out the division explicitly. The division of a polynomial of several variables by another one, however, is an extremely tedious affair and, if the dimension of the representation is sufficiently large, virtually impossible.

Another method is based on a theorem due to Cartan,[2] which says that every irreducible representation is completely characterized by its highest weight which is simple, and that it may be obtained as the highest weight of a direct product of two representations, provided the $l$ "fundamental" representations of the group are known ($l$ denotes the rank of the group, cf. below). This method is very simple indeed, if used in a purely geometric way, but it offers the disadvantage that in order to obtain the character of a certain representation, one previously has to compute a good many representations of a lower dimension.

For this reason, it seems useful to have a formula which presents the character of any representation in such a way that the multiplicities could be directly accessible.

Weyl's formula (1) and all the work derived from it are based on the infinitesimal method explained, e.g., in Racah's Princeton Lectures.[3] There one considers the Lie algebra of the group and constructs its representations using the well-known theorems of Weyl[1] and Cartan.[2] The "weights" appear as the solutions of eigenvalue equations.

But there exists another method, the so-called "*global method*" due to Hopf,[4] and extended by Stiefel[5] to the theory of representations. There the weights are the terms of the character of the representation; the latter is given yet by formula (1), but in a completely different approach: it is now derived from the properties of the regular lattice

[1] H. Weyl, Math. Z. **23**, 271 (1925); **24**, 328, 377, 789 (1926). [In *Selecta* (Birkhäuser, Basel, 1956)].

[2] E. Cartan, Thèse, Paris, 1894; Bull. Soc. Math. **41**, 53 (1913); Ann. Math. **4**, 209 (1929). [In *Oeuvres Complètes* (Gauthier-Villars, Paris, 1952), Vol. 1.]

[3] G. Racah, "Group Theory and Spectroscopy," in *Princeton Lecture Notes*, CERN reprint, 61–8.

[4] H. Hopf, Comm. Math. Helv. **13**, 119 (1940–1941); **15**, 59 (1942–1943).

[5] E. Stiefel, Comm. Math. Helv. **14**, 350 (1941–1942); **17**, 165 (1944–1945).

associated with every semisimple Lie group (cf. below, Sec. III B). This permits the use of an elementary geometrical calculus through which the theory becomes much more accessible and transparent. The two methods are in fact complementary and their comparison throws much light into the structure of Lie groups.

Moreover, the use of the Hopf–Stiefel method enables us to solve the problem stated before with elementary, i.e., essentially geometrical tools; we obtain a new formula, valid for any representation of any semisimple Lie group, which yields the character as a multiple sum rather than a quotient. The procedure described by the formula is entirely geometrical and deeply reflects the structure of the group; in that frame the multiplicities of the weights acquire a natural and simple significance and are readily obtained with the whole weight diagram.

In order to make the article reasonably self-contained, an outline of Hopf's and Stiefel's theory is given in Sec. II. (cf. also Ref. 6).

After having defined the notation in Sec. III A and having explained the geometric tools in Sec. III B, we shall compute the fundamental expression $1/\Delta$ in Sec. III C. This computation is done in the form of a geometric construction in the Cartan–Stiefel diagram $\Gamma$, which provides a logarithmic calculus. But at this point we shall transgress the frame of Cartan's and Stiefel's theory for a short while and work in a $m$- rather than an $l$-dimensional space. ($2m$ is the number of roots $\neq 0$, $l$ the rank of the group, $m \geq l$, cf. Sec. II). It is the step from $E_m$ back to $E_l$ which essentially yields the multiplicities of the weights. In Sec. III D we show how through multiplication with the "characteristic", $1/\Delta$ is cut back such that only a finite expression, namely the character, remains. In Sec. III E finally we shall justify the use of the divergent series $1/\Delta$.

In a forthcoming paper, the procedure proved here for all semisimple groups will be carried through in detail for the classical groups $A_l$, $B_l$, $C_l$, $D_l$, and $G_2$.

## II. SURVEY OF HOPF'S THEORY OF COMPACT LIE GROUPS

In the following, we work always within the frame of the global theory of compact (connected) groups as developed by Hopf.[4] This form of the theory is particularly well suited for the study of representations as was shown by Stiefel.[5,6]

An Abelian, connected, compact group is called a *toroid* (denoted by $T$). One proves that every toroid is the direct product of several groups $O_2$ (rotations in a 2-dimensional Euclidean plane $E_2$). A toroid therefore is completely characterized by its dimension.

A toroid which is a subgroup of a Lie group $G$, but is not a subgroup of a toroid of a higher dimension, is called *maximal toroid* of $G$. The fundamental theorem of Hopf says, that every element of a compact group is contained in (at least) one maximal toroid.

Let $T$ and $T'$ be two maximal toroids of a group $G$, then an element $g \in G$ exists such that $g^{-1}Tg = T'$. In other words, two maximal toroids are conjugated. This theorem justifies the selection of one particular maximal toroid as a tool for studying the group and shows that its dimension is an invariant, called the *rank* of the group. From this theorem follows that every representative of a compact group can be diagonalized.

If an element of a maximal toroid does not belong to any other maximal toroid, it is called *regular*, otherwise *singular*.

There exists a homeomorphism of a neighborhood of the identity $V(e)$ into a neighborhood $\tilde{V}$ of a point O of a $n$-dimensional Euclidean space $E_n$. By choosing O as the origin of a Cartesian coordinate system, one introduces *canonical coordinates* into the neighborhood $V(e)$. In particular, the intersection $V_T$ of the neighborhood $V(e)$ and the maximal toroid $T$ is mapped into an open set $\tilde{V}_T \subset \tilde{V}$ of dimension $l$.

First $\tilde{V}_T$ may be continued into the image of $T$; then, furthermore, into the image of the universal covering group of $T$, isomorphic to an $l$-dimensional Euclidean space $E_l$. Thus every element of $T$ will be represented by an infinite point lattice in $E_l$; in particular, the image of the unit element is a lattice $g^e$.

The image of the singular elements of $T$ consists of $m$ families of $(l-1)$-dimensional hyperplanes. Here $m = \frac{1}{2}(n-l)$, $n$ is the order of the group, $l$ its rank. One proves that no two singular hyperplanes may coincide. The set of all singular hyperplanes, i.e., the union of all singular points, is called the *diagram* $\Gamma$ *of Cartan and Stiefel.* The points of maximal intersection, i.e., the points which belong to one hyperplane of every set, represent the center of $G$.

The essential property of $\Gamma$ is the following: $\Gamma$ remains invariant under a reflection in any of

[6] D. Speiser, Lecture notes, *Istanbul Summer School in Theoretical Physics* (Gordon and Breach, New York, to be published).

the hyperplanes of which it is composed. Thus $\Gamma$ possesses very special symmetry properties.

If $G$ is *semisimple*, its center is a discrete group. Therefore, its image in $E_l$ is a point lattice $g^c$, generated by $l$ basis vectors. This lattice $g^c$ contains $g^e$ as a sublattice.

Consider the origin of $g^c$ and $g^e$ and all points into which it can be transformed through a series of successive reflections in hyperplanes of $\Gamma$. The set of these points is a sublattice $\gamma$ of $g^e$. Thus one has:

$$\gamma \subseteq g^e \subseteq g^c.$$

The finite discrete group, generated by the reflections in the hyperplanes passing through the origin of $g^c$, was called by Weyl the *group* $S$. $S$ is the crystal-class of $g^c$. The fundamental domains $D_i$ of $S$ are infinite pyramids whose corners are at the origin of $g^c$. $D_i$ has $l$ edges.

The $2m$ vectors, orthogonal to the $m$ hyperplanes passing through the origin of $g^c$ and twice as long as the distance to the next parallel hyperplane, are the *roots* of Cartan. Thus the set of roots, the *root diagram* also is invariant under the transformations of $S$. Whence follows immediately

$$2(\lambda, \mu)/(\lambda, \lambda) = n = \text{integer},$$

where $\lambda$, $\mu$ are two roots and $(\lambda, \mu)$ their scalar product. This relation is the source of van der Waerden's[7] classification of all simple groups, which was greatly simplified by Coxeter and Dynkin (cf. Ref. 8). (In the framework of the infinitesimal theory, the roots are the nonzero eigenvectors of the characteristic equation of the group. This equation is constructed in terms of the infinitesimal generators of the group (cf. Refs. 2 and 3).

The diagram $\Gamma$ of Cartan and Stiefel does not characterize a group completely, but rather the family of all groups which are locally isomorphic. Global properties are determined by the lattice $g^e$, or better by the way $g^e$ contains $\gamma$ and is itself contained in $g^c$.

For instance $g^e = g^c$ means that $e$ is the only element of the center of the group. This characterizes the adjoint group $G_A$. On the other hand, $g^e = \gamma$ characterizes the universal covering group. (For a proof of this statement cf. Ref. 8). In between these two extreme cases there may be room for intermediate possibilities.

As an example, take the unimodular unitary group $SU_n$, whose center C is the cyclic group $Z_n$. The same diagram is obtained for the following groups:

$-SU_n$ itself (universal covering group):

$$g^e = \gamma, \qquad C = Z_n;$$

$-SU_n/Z_p$ where $Z_p$ is a proper subgroup of $Z_n$: (i.e., $p$ divides $n$)

$$\gamma \subset g^e \subset g^c, \qquad C = Z_n/Z_p = Z_{n/p};$$

$-SU_n/Z_n$ (adjoint group):

$$g^e = g^c, \qquad C = e.$$

For $n = 3$, one has only the two extreme cases $SU_3$ and $SU_3/Z_3$, but for $n = 4$, there is one more possibility, namely $SU_4/Z_2$ (which is locally isomorphic to $SO_6$, the proper orthogonal group in 6 dimensions).

*Note.* A consequence of this situation is the classification of irreducible representations into classes which form a group isomorphic to the center. This will be studied in a forthcoming paper.

## III. THE CONSTRUCTION OF THE CHARACTERS OF IRREDUCIBLE REPRESENTATIONS

### A. Notations

Consider the space $E_l$ and the lattice $g^c$ contained in it. Both are divided into the fundamental domains $D_i$ of the group $S$. These are infinite pyramids limited by singular hyperplanes of $T$. By the operations of the group $S$ they are permuted among themselves. $S$ acts transitively on the $D_i$, i.e., for every pair $D_i$, $D_k$ there exists a $s \in S$ such that $D_i = sD_k$. Moreover, no $D_i$ remains invariant under any $s \in S$, $s \neq e$.[4]

#### *1. Definitions*

Introducing an orthonormal system $\{e_i\}$ into $E_l$ : $V = x^i e_i$, $i = 1, 2, \cdots l$, we say[2,3]:

A vector $V$ is *positive* if its first nonvanishing component is positive;

The vector $V$ is *greater* than $W$ if $V - W$ is positive;

The vectors $s_i V$ which are obtained by applying the operations $s_i$ of the group $S$ to a vector $V$ are called *equivalent to* $V$;

A vector greater than all its equivalents, is called *dominant.*

The $2m$ roots in particular fall into either of two classes: $m$ roots are positive (denoted $\alpha_i$), $m$ roots

[7] B. L. van der Waerden, Math. Zeit. **37**, 446 (1933).
[8] L. S. Pontrjagin, *Topologische Gruppen* (Teubner, Leipzig, 1957), 2nd ed.

are negative (denoted $\beta_i$, $\beta_i = -\alpha_i$). The two sets are separated by an $(l-1)$-dimensional hyperplane.

*2. The Vector $R_0$*

Let $R_0$ be half the sum of all positive roots,

$$R_0 = \tfrac{1}{2}\sum_{i=1}^{m}\alpha_i,$$

and call $D_0$ the fundamental domain to which $R_0$ belongs.

Let us now introduce an affine coordinate system with basic vectors $P_1 \cdots P_l \in g^c$, such that every $V \in g^c$ has the form

$$V = \sum_i p_i P_i \quad \text{with integers } p_i\ ,$$

and such that $D_0$ is defined by the inequalities

$$p_i \geq 0 \quad \text{for every} \quad i = 1 \cdots l,$$
$$\text{and for every} \quad V \in D_0. \tag{2}$$

For such a system Weyl has proved[1]:

$$R_0 = \tfrac{1}{2}\sum \alpha_i = P_1 + P_2 + \cdots P_l. \tag{3}$$

This shows that $R_0$ lies *inside* $D_0$ and not on its boundary.

*3. Outermost or Elementary Roots*

Let $\Pi_0$ be the hyperplane orthogonal to $R_0$, passing through the origin. $\Pi_0$ separates positive and negative roots, and none of either set lies on it. For if a root would lie on $\Pi_0$, the hyperplane orthogonal to it would contain $R_0$, and this is excluded by (3). In other words, the bundle of positive roots is convex, and so is the bundle of negative roots.

The $l$ surfaces of $D_0$ are the singular hyperplanes $\vartheta_i$ closest to $R_0$. The corresponding positive roots are therefore the roots closest to $\Pi_0$, i.e., the outermost roots of the bundle of positive roots. Since the outermost positive roots cannot be expressed as a sum of (2 or several) positive roots, they are also called *elementary*.

One easily shows (Ref. 8, Satz 113), that elementary roots are linearly independent whence there can be no elementary root besides the $l$ outermost ones.

Denoting the outermost roots $\alpha_1 \cdots \alpha_l$, one may write

$$\alpha_{l+i} = \alpha_j + \alpha_k, \qquad 1 \leq i \leq m-l, \quad 1 \leq j, k \leq m, \tag{4a}$$

and likewise for the negative roots:

$$\beta_{l+i} = \beta_j + \beta_k, \qquad 1 \leq i \leq m-l, \quad 1 \leq j, k \leq m. \tag{4b}$$

*4. Construction of the Coordinate System $p_i$*

One may proceed along the following line: the singular hyperplanes passing through the origin obey equations $\vartheta_i(x_j) = 0$, $i = 1 \cdots m$, which follows from the expressions for the roots $\alpha_i(e_j)$. They are well-known for all simple groups.[2,3,7] The $l$ hyperplanes closest to $R_0$ are the surfaces of $D_0$: $p_i = 0$, by construction. Thus,

$$(p_i = 0) \equiv (\vartheta_i(x_j) = 0),$$

whence

$$p_i = \lambda_i \vartheta_i(x_j).$$

The arbitrary constants $\lambda_i$ must be determined by (3) in both coordinate systems. One then gets $p_i = p_i(x_j)$ whence by inversion,

$$x_j = x_j(p_i). \tag{5}$$

## B. Formulation of the Problem and Geometrical Interpretation

*1. Weyl's Formula*

Weyl[1] showed that the character of an irreducible representation of a semisimple group may be written in the following form:

$$\chi = X/\Delta \tag{6}$$

where

$$X = X(K_0) = \sum_{s \in S} \delta_s e^{i(sK_0, \varphi)},$$
$$K_0 \in g^c \cap D_0. \tag{7}$$

The summation in this formula runs over all elements of $S$ and $\delta_s = \pm 1$, according to whether $s$ is a proper $(+1)$ or improper $(-1)$ rotation. The $\varphi^k$ are the group parameters introduced as coordinates into the toroid ($0 \leq \varphi^k \leq 2\pi$, $\varphi^i = 0$ or $2\pi$ represents the unit element of $T$). $X$ is called the *characteristic*[9] of the representation.

As a function of the $\varphi^k$, $|\Delta|^2$ is the Jacobian in the group integral, if the domain of integration is transformed from $G$ to $T$ with the help of the conjugation theorem of Sec. 2. This leads to the following form[10] for $\Delta$:

$$\Delta = \prod_{i=1}^{m}\{e^{\frac{1}{2}i(\alpha_i, \varphi)} - e^{-\frac{1}{2}i(\alpha_i, \varphi)}\}$$
$$= e^{i(R_0, \varphi)} \prod_{i=1}^{m}\{1 - e^{i(\beta_i, \varphi)}\}, \tag{8}$$

where the product is extended to the positive roots.

[9] The nomenclature is not uniform in the literature. Here we follow Stiefel, who derived formula (6) in the frame of Hopf's global theory,[5] whereas Weyl had used the infinitesimal method.

[10] We are indebted to Professor R. Brout and Professor Englert (University of Brussels) for having suggested to use this expression for $\Delta$.

On the other hand, $\Delta$ must be also the characteristic of the unit representation; indeed one has

$$\Delta = X(R_0) = \sum_{s \in S} \delta_s e^{i(sR_0, \varphi)}. \tag{9}$$

This expression contains only $\sigma$ terms ($\sigma$ = order of $S$), whereas the expression (8) contains $2^m$ terms and $2^m > \sigma$ for all simple groups: the additional terms cancel two by two, due to relations (4) between the roots.

With respect to $S$, $X$ and $\Delta$ are clearly alternating functions, $\chi$ an automorphic function:

$$\begin{aligned} sX &= \delta_s X, \\ s\Delta &= \delta_s \Delta, \qquad s \in S. \\ s\chi &= \chi, \end{aligned} \tag{10}$$

The fundamental theorem concerning irreducible representations then says:

There is a one-to-one correspondence between the irreducible representations of $G$ and the vectors $K_0 \in D_0$ (strictly inside $D_0$, not on its boundary) with integer $p$ components, i.e., lying on $g^c$. Every such lattice vector $K_0$ defines the characteristic $X(K_0)$ of an irreducible representation of $G$, and every characteristic can be obtained in this way.

The *character* has the following form:

$$\chi = \sum_M \gamma_M e^{i(M, \varphi)}, \tag{11}$$

where $\gamma_M$ is a positive integer and $M$ are vectors from $g^c$, such that the relation (10) is satisfied; these are the *weights* of the representation, $\gamma_M$ is the *multiplicity* of $M$. These weights, unlike the $K$'s, associated with $X$, do not all have the same length, since $\chi$ is not homogeneous. From (6), (7), and (9), one sees that the highest weight of the representation is the vector $L_0 = K_0 - R_0$. Already Cartan had proved that the highest weight of every irreducible representation is simple, i.e., $\gamma_{L_0} = 1$, and completely determines the representation.

We extend the notion of multiplicity in two directions: the multiplicity is simply a function defined on $g^c$ (i.e., on all lattice points or, equivalently, lattice vectors), whose range is the set of *all* integers, positive and negative.

### 2. Geometrical Interpretation

The term $e^{i(P, \varphi)}$ is geometrically represented in $T$ by the point (or vector) $P \in g^c$. Thus the character is represented by the set of points $M \in g^c$ of (8), with multiplicities $\gamma_M$. They form the *weight diagram* (WD); the WD is now a polyhedron invariant under the operations of $S$.

The characteristic $X(K_0)$ is represented by the points $sK_0$, with multiplicities $\delta_s$, which form a convex polyhedron, since all the vectors $sK_0$ have the same length. Two adjacent corners are symmetric with respect to a hyperplane $\vartheta_i$, therefore the edge joining them is parallel to the corresponding root $\alpha_i$.

For $\Delta$, this can be made somewhat more precise: every edge has the same length as the root parallel to it.

*Proof*: One obtains $s_q R_0$ from $R_0 (s_q \in R_0)$ by substituting for some of the $\alpha_i$'s the corresponding $\beta_i = -\alpha_i$, while the other $\alpha_i$'s permute among themselves:

$$\begin{aligned} R_0 &= \tfrac{1}{2}(\alpha_1 + \alpha_2 + \cdots + \cdots + \alpha_m), \\ s_q R_0 &= \tfrac{1}{2}(\alpha_1 + \alpha_2 \cdots + \alpha_k - \alpha_{k+1} \cdots - \alpha_m). \end{aligned}$$

Thus,

$$s_q R_0 - R_0 = -\alpha_{k+1} \cdots - \alpha_m = \beta_{k+1} + \cdots + \beta_m.$$

If $D_i = s_i D_0$ is a fundamental domain adjacent to $D_0$, then by definition, $s_i R_0 - R_0 = n\beta_i$, where $\beta_i$ is an *elementary* negative root; thus $n = 1$. This proves the assertion.

Our problem is the following one: given $X(D)$ and $\Delta$, compute the explicit form of $\chi(D)$, specially the $\gamma_M$'s; in other words, find the weight diagrams from the polyhedrons $X$ and $\Delta$.

The method used here consists first in multiplying $X$ by $1/\Delta$ instead of dividing $X$ by $\Delta$, and second in carrying out all the operations geometrically. In order to do so, $1/\Delta$ must be computed as an infinite sum.

## C. Construction of $1/\Delta$

### 1. Iterative Procedure

Formula (9) reads

$$\begin{aligned} \Delta &= \sum_S \delta_s e^{i(sR_0, \varphi)} \\ &= e^{i(R_0, \varphi)}[1 - \sum_S{}' (-\delta_s) e^{i(sR_0 - R_0, \varphi)}], \end{aligned}$$

($\sum'$ represents summation over all $s$ different from the identity). This can be written

$$\Delta = e^{i(R_0, \varphi)}(1 - Z)$$

with

$$Z = \sum_S{}' (-\delta_s) e^{i(sR_0 - R_0, \varphi)}, \tag{12}$$

whence, formally,

$$1/\Delta = e^{-i(R_0, \varphi)} 1/(1 - Z)$$

$$= e^{-i(R_0,\varphi)}\Big(\sum_0^\infty Z^k\Big). \qquad (13)$$

Clearly, this potential series converges only if $|Z| < 1$, which in general is not the case; therefore we shall consider it formally and justify our procedure afterwards. In fact, expression (13) should not be looked upon as a series to be summed up, but rather as a function defined on $g^c$, by interpreting it in the same way as the character (11): in a term $\gamma_p \exp i(P, \varphi)$, the multiplicity $\gamma_p$ at the point $P$ is the value of the function $1/\Delta$ at the point $P$. Divergence of the sum means only that $\gamma$ increases indefinitely when one departs further and further from the origin. But this, as we shall show, is of no importance since for computing the character; in fact only values of $1/\Delta$ lying in a bounded domain of $g^c$ are used. A detailed rigorous justification of the procedure will be given at the end (Sec. III E).

*Remark*: In (13), $1/\Delta$ no longer displays the $S$ symmetry explicitly. This is due to the distinction of the particular term $\exp i(R_0, \varphi)$. Clearly, any other of the $\sigma$ terms could have been taken in front of the bracket as well: there are in fact $\sigma$ different, but equivalent ways to construct $1/\Delta$, and the operations of $S$ simply exchange them pairwise.

It remains to express (13) in geometric language: $(1 - Z)$ is the same polyhedron as $\Delta$, shifted by the vector $-R_0$. This translation shifts the corner $R_0$ into the origin. $Z$ is obtained from $(1 - Z)$ by discarding the origin and reversing the signs at all other corners. Figure 1 shows the three polygonals $\Delta$, $(1 - Z)$, and $Z$ for the case of $A_2$. Thus, we may consider $Z$ as given by $Z \cdot 1$, i.e., the result of the application of the polyhedron $Z$ to the origin. In the same way, $Z^2$ is defined as $Z \cdot Z$, i.e., the superposition of the figures obtained when one applies the $Z$ polyhedron to every one of its own corners; and in general $Z^p$ is defined as $Z \cdot Z^{p-1}$. Figure 2 shows the construction of $Z^2$ in the case $A_2$.

The computation of the series (13) can be carried out step by step. Indeed consider in $g^c$ the hyperplane $\sigma_1$, containing the $l$ corners adjacent to the origin. It defines a family $\{\sigma^k\}$ of equidistant parallel hyperplanes. Let $\sigma_0$ be the neighbor of $\sigma_1$ passing through the origin, $\sigma_2$ the neighbor on the opposite side, $\sigma_3$ the neighbor of $\sigma_2$ $\cdots$ . One sees then immediately that $Z$ has corners only on the $\sigma_i$'s, with $i \geq 1$, $Z^2$ on the $\sigma_i$'s with $i \geq 2$, and in general, $Z^k$ on the $\sigma_i$'s with $i \geq k$. Thus the partial sum

$$\sum_k = 1 + Z + Z^2 + \cdots + Z^k$$

already yields the final multiplicities of all the relevant points of $\sigma_0, \sigma_1, \cdots, \sigma_k$. This is illustrated in Figs. 1 and 2 for the case $A_2$.

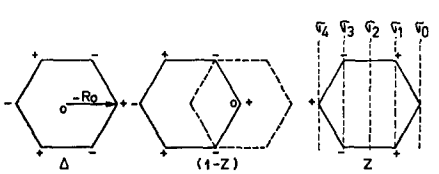


FIG. 1. The three polygonals $\Delta$, $1 - Z$, and $Z$ in the case $A_2$ ($l = 2$). The third one shows the family of hyperplanes $\{\sigma_i\}$ (straight lines in this case).

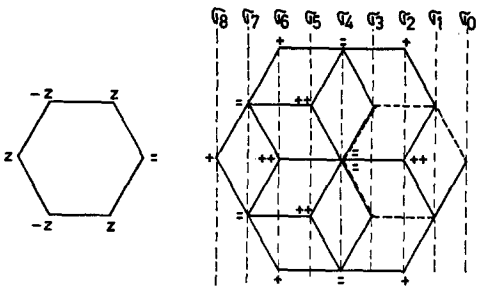


FIG. 2. Explicit construction of $Z^2 = Z \cdot Z$ in the case $A_2$.

In this way, one can calculate the series (13) as far as one wants for every group separately; for groups of rank $\leq 2$, this can be done graphically, for groups of rank $> 2$, one must use the coordinates of the different corners.

This method is simple, as the multiplication of the $Z^k$ is reduced to vector addition, i.e., geometric superposition in $g^c$. (This provides so to speak a logarithmic calculus!) But for bigger groups, the computation must be pushed rather far in order that one may be able to discover (empirically) the law which governs $1/\Delta$. It is therefore necessary to have a more direct method, valid for all simple groups, relying only on the structure of $\Delta$.

*2. Direct Procedure, Solution in the Space $E_m$*

From now on we use a shorter notation. We introduce the symbol

$$[P] \equiv e^{i(P,\varphi)}.$$

It has the obvious properties

$$[P][Q] = [P + Q], \qquad [P]^\alpha = [\alpha P].$$

We start then from formula (8), which can be written

$$\Delta = [R_0] \prod_{i=1}^{m} \{1 - [\beta_i]\}. \qquad (14)$$

Let us forget for a while the relations (4), and consider the $m$ negative roots $\beta_i$ as independent: they span an $m$-dimensional Euclidean space $E_m$;

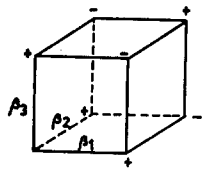


FIG. 3. The cube $\Delta/[R_0]$ in the case $A_2$.

in this space, $\Delta/[R_0]$ is now the hypercube defined by the $m$ vectors $\beta_i$: one corner is at the origin, with multiplicity $+1$, the other ones have alternatively multiplicity $-1$ and $+1$, starting from the origin. Figure 3 shows this figure in the case $A_2$ ($m = 3$).

We will now repeat on the expression (14) the manipulation performed in C1 of this Section on (12); one has formally

$$\begin{aligned}\frac{1}{\Delta} &= [R_0]^{-1} \prod_{i=1}^{m} \left\{ \frac{1}{1 - [\beta_i]} \right\} \\ &= [-R_0] \prod_{i=1}^{m} \left\{ \sum_{k_i=0}^{\infty} [\beta_i]^{k_i} \right\} \\ &= [-R_0] \prod_{i=1}^{m} \left\{ \sum_{k_i=0}^{\infty} [k_i \beta_i] \right\} \\ &= [-R_0] \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \left\{ \prod_{i=1}^{m} [k_i \beta_i] \right\} \\ &= [-R_0] \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \left[ \sum_{i=1}^{m} k_i \beta_i \right]. \end{aligned} \tag{15}$$

In the final result (15), the bracket $[\sum_{i=1}^{m} k_i \beta_i]$ clearly represents an arbitrary point of the lattice constructed on $\beta_1 \cdots \beta_m$, with positive, integer coefficients. The sum of this expression with respect to $k_1, k_2, \cdots, k_m$ from 0 to $\infty$ represents all points of one of the $2^m$ "octants" of this lattice in $E_m$. Each of them has multiplicity $+1$. $1/\Delta$ is the same figure translated by the vector $-R_0$.

This result may be expressed in a simpler way if we define the operation "summation along a vector". Consider a lattice with basic vectors $V_1 \cdots V_p$, a point $P$ of this lattice and a figure $F(P)$ attached to this point $P$: we shall call "sum of $F(P)$ along $V_k$" the figure obtained by superposition of all figures congruent to $F$ modulo $(+V_k)$. In a formula,

$$\sum_{V_k} F(P) \equiv \sum_{m=0}^{\infty} F(mV_k + P).$$

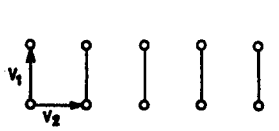


FIG. 4. The figure $\sum_{V_2} F(0)$, where $F(0)$ is the segment [(0,0)–(1,0)].

The summation sign on the rhs denotes the superposition just explained.

For instance, let $F(P)$ be the origin and $V_1$ the vector with coordinates $(1, 0, 0, \cdots)$, then $\sum_{V_1} [0]$ is the set of all the points $(m, 0, 0, \cdots)$, $m = 0, 1, 2, \cdots$.

As another example (Fig. 4), take $p = 2$, let $P$ be the origin and $F(0)$ the segment $[(0, 0) - (1, 0)]$; $\sum_{V_2} F(0)$ then represents the set of all the segments $[(0, m) - (1, m)]$, $m = 0, 1, 2 \cdots$.

With the help of this notation, (15) may be written in compact form,

$$\frac{1}{\Delta} = [-R_0] \sum_{\beta_1} \cdots \sum_{\beta_m} [0]. \tag{16}$$

### 3. *Solution in the Space $E_l$*

This gives the solution in $E_m$, where the problem of the multiplicities is trivial. But it remains to step back into $E_l$! As we have seen, this amounts to take into account relations (4) between the roots; the effect thereof is that in (14) some terms will cancel pairwise. This means that the corresponding corners of the hypercube with opposite multiplicities coincide. In other words, this operation defines an affine (singular) correspondence between $E_m$ and $E_l$ which preserves coincidence relations: $E_m$ is projected onto $E_l$, the hypercube in $E_m$ onto the $\Delta$ polyhedron in $E_l$ (see Fig. 5 for an example) and the $1/\Delta$ of $E_m$, as defined by (15) or (16), onto the $1/\Delta$ of $E_l$. Since these two formulas express only coincidence properties which still hold after this projection, they will thus remain valid in $E_l$. But since the projection is singular (from a $m$-dimensional onto an $l$-dimensional space), there will be new coincidences; they will cause some multiplicities becoming greater than 1. We can now write our result in $E_l$ under the following form, if we use the very definition of $R_0$:

$$\frac{1}{\Delta} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \left[ \sum_{i=1}^{m} (k_i + \tfrac{1}{2}) \beta_i \right], \tag{17}$$

or

$$\frac{1}{\Delta} = \sum_{\beta_1} \cdots \sum_{\beta_m} [-R_0]. \tag{18}$$

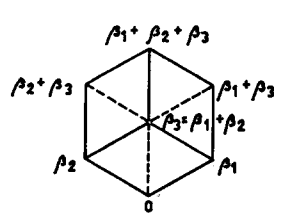


FIG. 5. In the case $A_2 (l = 2, m = 3)$, the cube of $E_3$ is projected onto the hexagon $\Delta$ of $E_2$, when one takes into account the relation $\beta_3 = \beta_1 + \beta_2$.

From these formulas, one sees immediately:

$1/\Delta$ is an infinite pyramid with $l$ surfaces and its corner is at $-R_0$;

Its surfaces are parallel to the surfaces of $\Delta$ which intersect at $R_0$;

At all points of $g^\circ$ on or inside the pyramid the multiplicity of $1/\Delta$ is greater or equal to 1, since it is equal to the number of points of the $1/\Delta$ of $E_m$ which coincide after the projection on $E_l$, i.e., the number of different combinations $k_1 \cdots k_m$ which correspond to this point;

In particular, the multiplicity is equal to 1 at all points on the edges of the pyramid; they correspond to the set of values

$k_i = m$, positive integer, $\beta_i$ = elementary root,
$k_i = 0, \quad i \neq j, \quad (j = 1, 2, \cdots l)$;

On each surface of $1/\Delta$ [$(l-1)$-dimensional hyperplane], the multiplicity depends only upon the angles between the elementary roots which span it.

### D. Construction of the Characters

#### 1. General Formulas

Now that the expression for $1/\Delta$ is under control, we may easily construct the characters of all irreducible representations.

Let a characteristic be given,

$$X(K_0) = \sum_S \delta_s e^{i(sK_0, \varphi)} = \sum_S \delta_s [sK_0].$$

The corresponding character then is

$$\begin{aligned}\chi &= X(K_0)\frac{1}{\Delta} \\ &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \left\{ X(K_0) \left[ \sum_{i=1}^{m} (k_i + \tfrac{1}{2})\beta_i \right] \right\} \\ &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \left\{ \sum_S \delta_s \left[ \sum_{i=1}^{m} (k_i + \tfrac{1}{2})\beta_i + sK_0 \right] \right\} \\ &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \left\{ \sum_S \delta_s \left[ \sum_{i=1}^{m} k_i\beta_i + sK_0 - R_0 \right] \right\}, \end{aligned} \tag{19}$$

$$\chi = \sum_{\beta_1} \cdots \sum_{\beta_m} \{X(K_0)[-R_0]\}. \tag{20}$$

The expression between curly brackets in (20) is the character translated by the vector $(-R_0)$.

#### 2. Geometric Interpretation

From this form, one recognizes immediately that $\chi$ contains only a finite number of terms, or better,

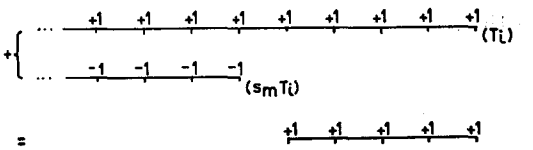


FIG. 6. Construction of $\sum_{\beta_m} [-R_0]$ as explained in the text.

only a finite number of terms of $1/\Delta$ contribute to $\chi$. Indeed,

$$\begin{aligned}\chi &= X \sum_{\beta_1} \cdots \sum_{\beta_m} [-R_0] \\ &= \sum_{\beta_1} \cdots \sum_{\beta_{m-1}} \{X \sum_{\beta_m} [-R_0]\} = \sum_{\beta_1} \cdots \sum_{\beta_{m-1}} \{\omega\}.\end{aligned}$$

$\sum_{\beta_m} [-R_0]$ is a "ray" of the lattice $g^\circ$ parallel to $\beta_m$, which begins at the point $(-R_0)$. All lattice points which it carries, have multiplicity $+1$. But to every corner $[T_i]$ of $X$ corresponds a second one $[s_m T_i]$ symmetric to $T_i$ with respect to $\vartheta_m$, therefore bearing the opposite sign. These two corners therefore yield identical rays in the product $\omega$ but with opposite signs. One of them is translated with respect to the other by a finite amount. Thus their superposition yields a finite segment parallel to $\beta_m$ all points of which have multiplicity $+1$. (Fig. 6). Thus $\omega$ consists of $\frac{1}{2}\sigma$ finite similar segments ($\sigma$ = order of $S$). If the $(m-1)$ other summations along $\beta_1 \cdots \beta_{m-1}$ are carried out, these segments are shifted such that they remain parallel to their original position. But they are never shifted in the direction $\beta_m$, and therefore the diameter of the set of points of $g^\circ$ in the direction of $\beta_m$ is finite. To put it exactly: every line parallel to $\beta_m$ contains only a finite number of points. However, the order of summation in (20) is just as arbitrary as in (18). Therefore what was said with respect to $\beta_m$ is also valid for the other $\beta_i$'s, whence follows that the diameter of $\chi$ is finite in the direction of all $\beta_i$'s. Thus it contains only a finite number of points, i.e., only a finite number of points of $1/\Delta$ really contribute to $\chi$.

#### 3. Simplification of the Formulas

Let us first note the following. Formulas (19) and (20) yield the entire character. But much less is needed; it is quite enough to know the part $\chi_0$ of $\chi$ which is contained in $D_0$ (inside and on the boundary); the other parts then will be obtained through the operations of $S$,

$$\chi_0 = \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \left\{ \sum_S \delta_s \left[ \sum_{i=1}^{m} k_i\beta_i + sK_0 - R_0 \right] \right\}, \tag{21}$$

with the condition

$$\sum_{i=1}^{m} k_i\beta_i + sK_0 - R_0 \in D_0,$$

and

$$\chi = \sum_s s\chi_0.$$

In the form (20), $\chi_0$ is obtained if one restricts oneself to the points of $D_0$; these may be obtained by starting from $X(K_0)[-R_0]$ and summing up along every one of the $\beta_i$'s, as usual.

Note also: The same remark which was made for $1/\Delta$ is valid also for $\chi$. Formulas (19) and (20) are not manifestly symmetrical with respect to $S$. This results from having designated $m$ arbitrary (adjacent) roots as negative ones, i.e., from the particular orientation of the orthogonal coordinate system with respect to $g^0$, implied by this choice.

The symmetry of the root diagram would have permitted to choose any other set of adjacent roots, i.e., roots which are all on the same side of an arbitrary $(l-1)$-dimensional hyperplane passing through the origin.

If

$$\{\beta_{i_i}\} = s_i\{\beta_i\}, \qquad s_i \in S,$$

one obtains

$$\chi_i = \sum_{\beta_{i_1}} \cdots \sum_{\beta_{i_m}} \{X(K_0)[-R_0']\},$$

with

$$R_0' = -\tfrac{1}{2}\sum_i \beta_{i_i}$$

$$= s_i\chi = \chi.$$

### E. Justification of the Procedure

In our direct procedure, we used the relation

$$\frac{1}{\Delta} = [-R_0] \prod_{j=1}^{m} \left\{\frac{1}{1-[\beta_j]}\right\}.$$

We consider now the $j$th factor and expand it formally:

$$\frac{1}{1-[\beta_j]} = \frac{1}{1-e^{i(\beta_j,\varphi)}} = \sum_{k_j=0}^{\infty} e^{i(k_j\beta_j,\varphi)}.$$

If we put $Z_j = [\beta_j] = e^{i(\beta_j,\varphi)}$, we have $|Z_j| = 1$ and the series obviously diverges.

But if we substitute for $\varphi$, $\varphi' = \varphi + i\psi$ (i.e., $\varphi'^k = \varphi^k + i\psi^k$, $k = 1 \cdots l$), with the condition $(\beta_j, \psi) > 0$, we have $|Z_j'| = |e^{i(\beta_j,\varphi)}e^{-(\beta_j,\psi)}| = e^{-(\beta_j,\psi)} < 1$ and the series $\sum_{k_j=0}^{\infty} Z'^{k_j}$ converges absolutely.

We choose now a $\psi$ such that for every $j = 1 \cdots m$, one has

$$(\beta_j, \psi) > 0. \tag{22}$$

This is always possible, since all vectors $\beta_j$ are on the same side of a hyperplane $\Pi_0$ passing through the origin: every vector $\psi$ on the same side of $\Pi_0$ as the $\beta_j$'s will satisfy the requirement. We have then, for every $j = 1 \cdots m$, an absolutely convergent expansion:

$$\frac{1}{1-Z_j'} = \sum_{k_j=0}^{\infty} (Z_j')^{k_j}.$$

As a consequence, we may now take the product of these $m$ expansions and arbitrarily change the order of the summations, as we did above:

$$\prod_{j=1}^{m} \frac{1}{1-Z_j'} = \prod_{j=1}^{m} \left\{\sum_{k_j=0}^{\infty} (Z_j')^{k_j}\right\}$$

$$= \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} Z_1'^{k_1} \cdots Z_m'^{k_m}.$$

This will give us formulas (15)–(18), but expressed in $\varphi'$ instead of $\varphi$.

We may write $1/\Delta = \Delta_f + R$, where $\Delta_f$ contains only the finite number of terms which contribute to $\chi$ and $R$ is the rest of the series.

Thus:

$$\chi = X(1/\Delta) = X\Delta_f + XR$$

$$= \chi' + \Omega.$$

The function $\Omega$ is analytic since it contains only exponentials. For every $\chi$ belonging to the domain of $E_l$ defined by (22), $\Omega$ vanishes identically. Therefore, if analytically continued, it vanishes identically everywhere, in particular also for $\psi = 0$, $\varphi' = \varphi$ (which is a limit point of the domain in question).

Thus, we have always and everywhere $\chi' = \chi$; this justifies the use of the diverging series (15)–(18) for computing the character.

This shows that the procedure which we called "direct" is in fact valid and so are the resulting formulas for $\chi$. The other procedure (the "iterative" one), is completely equivalent to the first one, so that no independent proof is needed.

#### ACKNOWLEDGMENTS

Most of this work was performed while one of the authors (J.-P. A.) was a guest of the "Institut de Physique Théorique de l'Université de Genève"; he acknowledges gratefully the hospitality offered to him by Professor J. M. Jauch and also a travel grant from the "Institut Interuniversitaire des Sciences Nucléaries" (Belgium). The other author (D. S.) expresses his gratitude to the Swiss National Science Foundation for financial support. Both of them thank Dr. G. Targonski and Dr. G. Barton for helpful discussions.

# Nonrelativistic Coulomb Green's Function in Momentum Space*

LEVERE HOSTLER
*Yale University, New Haven, Connecticut*
(Received 24 April 1964)

The nonrelativistic Coulomb Green's function in momentum space is obtained in closed form by Fourier transforming the known expression for the coordinate-space Green's function. Also, an integral representation for the momentum-space Green's function is obtained which looks rather attractive from the point of view of applications.

## I. INTRODUCTION

IN a previous paper[1] the author has given an expression in closed form for the nonrelativistic Coulomb Green's function in coordinate space. The coordinate-space Green's function was defined there as the solution $G(\mathbf{r}_2, \mathbf{r}_1, \omega)$ of the differential equation[2]

$$\{\nabla_2^2 + (2k\nu)/r_2 + k^2\}G(\mathbf{r}_2, \mathbf{r}_1, \omega) = \delta^3(\mathbf{r}_2 - \mathbf{r}_1), \quad (1)$$
$$k = (2m\omega/\hbar)^{\frac{1}{2}}, \quad \mathrm{Im}\,(k) > 0; \quad \nu = Ze^2m/4\pi k\hbar^2$$

subject to certain regularity conditions at the origin and at infinity. The quantity $\hbar\omega$ is any complex number not in the (discrete or continuous) eigenvalue spectrum of the Coulomb Hamiltonian. The Green's function $G(\mathbf{r}_2, \mathbf{r}_1, \omega)$ is analytic in the complex $\hbar\omega$ plane except for a branch cut along the positive real axis, $0 \leq \hbar\omega < +\infty$, and except for simple poles at the Bohr energy levels,

$$\hbar\omega = -mZ^2e^4/2(4\pi)^2\hbar^2n^2, \quad n = 1, 2, 3, \cdots .$$

For applications, it would be desirable to know the Green's function in momentum space. This function is the Fourier transform in $\mathbf{r}_1$ and $\mathbf{r}_2$ of the coordinate-space Green's function:

$$G(\mathbf{k}_2, \mathbf{k}_1, \omega) = \int \frac{d^3r_2}{(2\pi)^{\frac{3}{2}}} \int \frac{d^3r_1}{(2\pi)^{\frac{3}{2}}} e^{-i\mathbf{k}_2\cdot\mathbf{r}_2 + i\mathbf{k}_1\cdot\mathbf{r}_1} G(\mathbf{r}_2, \mathbf{r}_1, \omega). \quad (2)$$

It satisfies the integral equation

$$G(\mathbf{k}_2, \mathbf{k}_1, \omega) = \frac{\delta^3(\mathbf{k}_2 - \mathbf{k}_1)}{k^2 - k_2^2} - \frac{2k\nu}{k^2 - k_2^2} \int \frac{d^3k_3}{(2\pi)^3} \frac{4\pi}{|\mathbf{k}_2 - \mathbf{k}_3|^2} G(\mathbf{k}_3, \mathbf{k}_1, \omega), \quad (3)$$

which is just the momentum-space counterpart of Eq. (1).

We here derive an expression in closed form [Eq. (17) in conjunction with Eqs. (22), (23), (27), and (29)] for $G(\mathbf{k}_2, \mathbf{k}_1, \omega)$. In the course of this derivation, we obtain an integral representation [Eq. (24)] for $G(\mathbf{k}_2, \mathbf{k}_1, \omega)$ which looks rather attractive from the point of view of applications.

*Note added in proof*: Essentially the same integral representation has been obtained independently by S. Okubo and D. Feldman, Phys. Rev. **117**, 292 (1960).

These results are derived from the integral representation

$$G(\mathbf{r}_2, \mathbf{r}_1, \omega) = \frac{ik}{8\pi} \frac{e^{-\pi\nu}}{\sinh \pi\nu} \int_{+\infty;\,\mathrm{arc}(\zeta\pm1)=0}^{(1+)} d\zeta \left(\frac{\zeta + 1}{\zeta - 1}\right)^{i\nu} e^{ik(r_1+r_2)\zeta} \times I_0(-2ik(r_1r_2)^{\frac{1}{2}} \cos \tfrac{1}{2}\theta(\zeta^2 - 1)^{\frac{1}{2}}), \quad 0 < \mathrm{arc}\,(k) < \pi, \quad (4)$$

for the coordinate-space Green's function given in Ref. 1 [Eq. (1.13)]. Here $\theta$ is the angle between the vectors $\mathbf{r}_2$ and $\mathbf{r}_1$, and $I_0$ denotes the Bessel function of imaginary argument, as defined in Watson.[3] The integration contour begins at $\zeta = +\infty$ on the positive real axis, runs down the positive real axis to a point on the right of $\zeta = +1$, circles the point $\zeta = +1$ in the positive (counterclockwise) sense, and then returns along the positive real axis to $\zeta = +\infty$. The phases of $(\zeta \pm 1)$ are determined along the contour by continuity, their initial values at $\zeta = +\infty$ being arc $(\zeta \pm 1) = 0$.

The integral representation (4) has the merit of isolating the $Z$ dependence of the integrand in the simple factor $[(\zeta + 1)/(\zeta - 1)]^{i\nu}$. As regards the $\mathbf{r}_2$ and $\mathbf{r}_1$ dependence of the integrand, this is the same as we would have in the free-particle limit. Hence the integrand of (4) should be fairly easy to Fourier-transform in $\mathbf{r}_2$ and $\mathbf{r}_1$, and upon taking this Fourier transform, we will obtain directly an integral representation for the momentum-space Green's func-

* This research was supported in part by the U. S. Atomic Energy Commission.

[1] Levere Hostler, J. Math. Phys. **5**, 591 (1964).

[2] Heaviside–Lorentz (= rationalized Gaussian) units are used.

[3] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1962), 2nd ed., p. 77.

tion $G(\mathbf{k}_2, \mathbf{k}_1, \omega)$. The latter integral representation will clearly inherit from (4) the property of isolating the $Z$ dependence of the integrand in the single factor $[(\zeta + 1)/(\zeta - 1)]^{i\nu}$. Our program is, then, to first Fourier-transform the integrand of (4). This will give the integral representation of the momentum-space Green's function that we are looking for. This is done in Sec. II. We then find (Sec. III) that the parameter ($\zeta$) integral can be performed in terms of hypergeometric functions, giving the desired closed-form expression for the momentum-space Green's function.

## II. INTEGRAL REPRESENTATION FOR THE MOMENTUM-SPACE GREEN'S FUNCTION

Proceeding with our program as outlined above, we write $G(\mathbf{k}_2, \mathbf{k}_1, \omega)$ in the form

$$G(\mathbf{k}_2, \mathbf{k}_1, \omega) = \frac{1}{e^{2\pi\nu} - 1} \int_{+\infty;\arg(\zeta\pm1)=0}^{(1+)} d\zeta \left(\frac{\zeta + 1}{\zeta - 1}\right)^{i\nu} \times D(\mathbf{k}_2, \mathbf{k}_1, \omega), \tag{5}$$

where

$$D(\mathbf{k}_2, \mathbf{k}_1, \omega) = \frac{ik}{4\pi} \int \frac{d^3r_2}{(2\pi)^{3/2}} \int \frac{d^3r_1}{(2\pi)^{3/2}} e^{-i\mathbf{k}_2\cdot\mathbf{r}_2 + i\mathbf{k}_1\cdot\mathbf{r}_1} \times e^{ik(r_1+r_2)\zeta} I_0[-2ik(r_2r_1)^{\frac{1}{2}} \cos\tfrac{1}{2}\theta(\zeta^2 - 1)^{\frac{1}{2}}]. \tag{6}$$

In order to achieve convergence of the integral (6) for $D(\mathbf{k}_2, \mathbf{k}_1, \omega)$, we choose the $\zeta$ integration contour such that, for all $\zeta$ values on the contour, both inequalities

$$\operatorname{Im}[k(\zeta \pm (\zeta^2 - 1)^{\frac{1}{2}})] > 0 \tag{7}$$

hold. Using the method of the appendix, it can be shown that the two inequalities (7) restrict the $\zeta$ integration contour to the interior of the region on the right of the right-hand branch of the hyperbola

$$\zeta_1^2/\cos^2\delta - \zeta_2^2/\sin^2\delta = 1, \qquad \zeta_1 = \operatorname{Re}(\zeta), \quad \zeta_2 = \operatorname{Im}(\zeta), \quad \delta = \arg(k). \tag{8}$$

The real part of the contour, as previously described, already lies in this region. In order to satisfy (7), it is only necessary to take the loop about the point $\zeta = +1$ sufficiently small as to also lie in this region. (The integration contour could still be continuously deformed in any way, so long as the process of deformation does not involve going outside the permitted region and does not involve passing over the point $\zeta = +1$.)

The integral (6) can be evaluated by using the integral representation[4]

$$I_0(z) = \frac{1}{2\pi i} \int_{c-i\infty;c>0}^{c+i\infty} dt t^{-1} e^{(t + z^2/4t)} \tag{9}$$

of the Bessel function. If we choose

$$c > r_2 |k|^2 |\zeta^2 - 1|/\operatorname{Im}(k\zeta), \tag{10}$$

then the $\mathbf{r}_1$ integration can be performed before the $t$ integration. [Here we require the condition $\operatorname{Im}(k\zeta) > 0$, which follows from (7).] The $\mathbf{r}_1$ integration can be worked out with the help of the following integral (assumed convergent):

$$\int d^3r e^{-Ar - \mathbf{B}\cdot\mathbf{r}} = 8\pi A(A^2 - \mathbf{B}\cdot\mathbf{B})^{-2}, \tag{11}$$

and we find

$$D(\mathbf{k}_2, \mathbf{k}_1, \omega) = \int d^3r_2 e^{-i\mathbf{k}_2\cdot\mathbf{r}_2 + ikr_2\zeta} \times \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} dt e^t \frac{ik}{(2\pi)^3} \frac{-2ik\zeta t + k^2r_2(\zeta^2 - 1)}{(k^2\zeta^2 - k_1^2)^2(t - t_p)^2}, \tag{12}$$

$$t_p = \frac{ik^2(\zeta^2 - 1)(\mathbf{k}_1\cdot\mathbf{r}_2 - kr_2\zeta)}{(k^2\zeta^2 - k_1^2)}.$$

The $t$ integration contour in (12) can be closed on the left by an arc at infinity, and the $t$ integral evaluated by the residue theorem. The only pole of the integrand of (12) is a second-order pole at $t = t_p$. As a consequence of the inequality (10), this pole is found to lie inside the (now closed) integration contour. On evaluating the residue at the pole, (12) leads to

$$D(\mathbf{k}_2, \mathbf{k}_1, \omega) = \int d^3r_2 e^{-Ar_2 - \mathbf{B}\cdot\mathbf{r}_2} \times \frac{ik}{(2\pi)^3} \frac{1}{(k^2\zeta^2 - k_1^2)^3} [-2ik\zeta(k^2\zeta^2 - k_1^2) + 2\mathbf{r}_2\cdot\mathbf{k}_1k^3\zeta(\zeta^2 - 1) - r_2k^2(\zeta^2 - 1)(k^2\zeta^2 + k_1^2)],$$

$$A = -ik\zeta \frac{k^2 - k_1^2}{k^2\zeta^2 - k_1^2}, \qquad \mathbf{B} = i\mathbf{k}_2 - i\mathbf{k}_1 \frac{k^2(\zeta^2 - 1)}{k^2\zeta^2 - k_1^2}. \tag{13}$$

The $\mathbf{r}_2$ integration in (13) can be performed using (11) and the two formulas

$$\int d^3r\, re^{-Ar - \mathbf{B}\cdot\mathbf{r}} = 8\pi(3A^2 + \mathbf{B}\cdot\mathbf{B})(A^2 - \mathbf{B}\cdot\mathbf{B})^{-3}$$

and

$$\int d^3r\, \mathbf{r}e^{-Ar - \mathbf{B}\cdot\mathbf{r}} = -32\pi A\mathbf{B}(A^2 - \mathbf{B}\cdot\mathbf{B})^{-3}, \tag{14}$$

obtained from (11) by differentiating with respect to the parameters $A$ and $\mathbf{B}$, respectively. We then

[4] This may be obtained from Eq. (8), p. 177 of Watson[3] by using the relation (p. 77) $I_0(z) = J_0(iz)$.

obtain the following expression for $D(\mathbf{k}_2, \mathbf{k}_1, \omega)$:

$$D(\mathbf{k}_2, \mathbf{k}_1, \omega) = \frac{(ik)^3}{\pi^2} \times \frac{(k^2 - k_1^2)(k^2 - k_2^2)(3\zeta^2 - 1) + |\mathbf{k}_2 - \mathbf{k}_1|^2 k^2(\zeta^4 - 1)}{[(k^2 - k_1^2)(k^2 - k_2^2) - |\mathbf{k}_2 - \mathbf{k}_1|^2 k^2(\zeta^2 - 1)]^3}. \quad (15)$$

Equation (15), when used in conjunction with Eq. (5), would give an integral representation for the momentum-space Green's function. Although this integral representation might be convenient for some applications, we prefer not to leave it in this form for the following reason. In the free-particle case the momentum-space Green's function is known to have the form $\delta^3(\mathbf{k}_2 - \mathbf{k}_1)(k^2 - k_2^2)^{-1}$ and is a distribution rather than an ordinary function. Now we will find that the momentum-space Green's function contains a distribution also in the Coulomb case, and we prefer to rewrite the integral representation for the Coulomb case so as to exhibit this distribution. To this end, we observe that $D(\mathbf{k}_2, \mathbf{k}_1, \omega)$ can be rewritten in the form

$$D(\mathbf{k}_2, \mathbf{k}_1, \omega) = \frac{(ik)^3}{2\pi^2} \frac{1}{(k^2 - k_2^2)(k^2 - k_1^2)} \frac{d}{d\zeta}(\zeta^2 - 1) \frac{d}{d\zeta}(\zeta^2 - 1) \times \frac{1}{(k^2 - k_2^2)(k^2 - k_1^2) - k^2 |\mathbf{k}_2 - \mathbf{k}_1|^2 (\zeta^2 - 1)}. \quad (16)$$

This identity "explains" the rather complicated structure of (15) and suggests that when we substitute into (5) for $G(\mathbf{k}_2, \mathbf{k}_1, \omega)$ we should integrate by parts two times. Doing this, we obtain for $G(\mathbf{k}_2, \mathbf{k}_1, \omega)$

$$G(\mathbf{k}_2, \mathbf{k}_1, \omega) = G_0 + G_1 + \mathcal{G}_2, \quad (17)$$

where

$$G_0 = \left(\frac{\zeta + 1}{\zeta - 1}\right)^{i\nu} \frac{(ik)^3}{\pi^2} \times \frac{\zeta(\zeta^2 - 1)}{[(k^2 - k_2^2)(k^2 - k_1^2) - k^2 |\mathbf{k}_2 - \mathbf{k}_1|^2 (\zeta^2 - 1)]^2}\Bigg|_{\zeta=\infty}, \quad (18)$$

$$G_1 = i\nu\left(\frac{\zeta + 1}{\zeta - 1}\right)^{i\nu} \frac{(ik)^3}{\pi^2} \times \frac{(\zeta^2 - 1)}{[(k^2 - k_2^2)(k^2 - k_1^2) - k^2 |\mathbf{k}_2 - \mathbf{k}_1|^2 (\zeta^2 - 1)]}\Bigg|_{\zeta=\infty}, \quad (19)$$

and

$$\mathcal{G}_2 = (i\nu)^2 \frac{(ik)^3}{\pi^2} \frac{1}{(k^2 - k_2^2)(k^2 - k_1^2)} \times \frac{2}{e^{2\pi\nu} - 1} \int_{+\infty;\,\arg(\zeta\pm1)=0}^{(1+)} d\zeta \left(\frac{\zeta + 1}{\zeta - 1}\right)^{i\nu} \times \frac{1}{(k^2 - k_2^2)(k^2 - k_1^2) - k^2 |\mathbf{k}_2 - \mathbf{k}_1|^2 (\zeta^2 - 1)}. \quad (20)$$

The first two terms of (17) are the "surface terms" from the two integrations by parts. The factor $[(\zeta + 1)/(\zeta - 1)]^{i\nu}$ occurring in these terms [see Eqs. (18) and (19)] can be replaced by unity in the limit as $\zeta \to \infty$. The remaining expression in (18) can be evaluated by appealing to the free-particle limit. In this limit, $G_1 = \mathcal{G}_2 = 0$, so that $G(\mathbf{k}_2, \mathbf{k}_1, \omega) = G_0\,|_{\nu=0}$. Inserting the known expression for the momentum-space Green's function on the left-hand side of this equation, and using (18) we find

$$\frac{\delta^3(\mathbf{k}_2 - \mathbf{k}_1)}{k^2 - k_2^2} = \frac{(ik)^3}{\pi^2} \times \frac{\zeta(\zeta^2 - 1)}{[(k^2 - k_2^2)(k^2 - k_1^2) - k^2 |\mathbf{k}_2 - \mathbf{k}_1|^2 (\zeta^2 - 1)]^2}\Bigg|_{\zeta=\infty}. \quad (21)$$

We can apply this result to evaluate $G_0$ for general $Z$. By writing the limit of the product in (18) as the product of the limits, using (21) and the fact that the limit of the first factor, $[(\zeta + 1)/(\zeta - 1)]^{i\nu}$, equals unity (as pointed out above), we find

$$G_0 = \delta^3(\mathbf{k}_2 - \mathbf{k}_1)/(k^2 - k_2^2), \quad (22)$$

i.e., the limit (18) equals the free-particle Green's function also for finite $Z$. The limit (19) is more straightforward, and we find simply

$$G_1 = i\nu \frac{ik}{\pi^2} \frac{1}{|\mathbf{k}_2 - \mathbf{k}_1|^2 (k^2 - k_2^2)(k^2 - k_1^2)}. \quad (23)$$

Putting these results together, we finally obtain the momentum-space Green's function in the form

$$G(\mathbf{k}_2, \mathbf{k}_1, \omega) = \frac{\delta^3(\mathbf{k}_2 - \mathbf{k}_1)}{k^2 - k_2^2} + i\nu \frac{ik}{\pi^2} \frac{1}{|\mathbf{k}_2 - \mathbf{k}_1|^2 (k^2 - k_2^2)(k^2 - k_1^2)} + (i\nu)^2 \frac{(ik)^3}{\pi^2} \frac{1}{(k^2 - k_2^2)(k^2 - k_1^2)} \times \frac{2}{e^{2\pi\nu} - 1} \int_{+\infty;\,\arg(\zeta\pm1)=0}^{(1+)} d\zeta \left(\frac{\zeta + 1}{\zeta - 1}\right)^{i\nu} \times \frac{1}{(k^2 - k_2^2)(k^2 - k_1^2) - k^2 |\mathbf{k}_2 - \mathbf{k}_1|^2 (\zeta^2 - 1)}, \quad \mathrm{Im}\,[k(\zeta \pm (\zeta^2 - 1)^{\frac{1}{2}})] > 0. \quad (24)$$

This is the integral representation for the momentum-space Coulomb Green's function in its final form. In this form the Green's function is exhibited as a distribution plus ordinary functions. The conditions Im $[k(\zeta \pm (\zeta^2 - 1)^{\frac{1}{2}})] > 0$ here [carried over from (7)] exclude the points $\zeta$ at which the integrand has a pole from the interior of the $\zeta$ integration contour.[5]

It can be verified by direct substitution that the expression on the right-hand side of (24) actually satisfies the integral equation (3). We have already pointed out that the first term of (24) (which we have called $G_0$) is the free particle Green's function. The second term (called $G_1$) is found to be precisely the first Born approximation. In the Feynman language, the first term corresponds to no action of the potential, the second term corresponds to the potential acting exactly once, and the last term ($\mathcal{G}_2$) contains the effects of the potential acting two or more times. The momentum-space Coulomb Green's function thus appears to have a natural decomposition into a no-potential term plus a one-potential term plus a many-potential term. Also, the $\mathbf{k}_2$ and $\mathbf{k}_1$ dependence of the expressions occurring in the integral representation (24) is quite simple. It was for both of these reasons that it was suggested earlier that the integral representation (24) may be convenient for applications.

### III. MOMENTUM-SPACE COULOMB GREEN'S FUNCTION IN CLOSED FORM

The parameter integral for the many-potential term, $\mathcal{G}_2$, can be evaluated explicitly in terms of hypergeometric functions, giving a closed form expression for the momentum-space Green's function. We first make the change of variables

$$t = [(\zeta - 1)/(\zeta + 1)], \tag{25}$$

obtaining

$$\mathcal{G}_2 = (i\nu)^2 \frac{(ik)^3}{\pi^2} \frac{1}{(k^2 - k_2^2)^2(k^2 - k_1^2)^2} \tag{26}$$

$$\times \frac{4}{e^{2\pi\nu} - 1} \int_{+1;\arg(t)=0}^{(0+)} dt\, t^{-i\nu}$$

$$\times \left[1 - t\frac{1-q}{1+q}\right]^{-1}\left[1 - t\frac{1+q}{1-q}\right]^{-1},$$

where

$$q^2 = \frac{k^2 |\mathbf{k}_2 - \mathbf{k}_1|^2}{k^4 - 2k^2\mathbf{k}_2\cdot\mathbf{k}_1 + k_1^2k_2^2}.$$

On making a partial fraction expansion of the denominator of (26), we obtain

$$\mathcal{G}_2 = (i\nu)^2 \frac{ik}{\pi^2} \frac{\Gamma(1 - i\nu)}{(k^2 - k_2^2)(k^2 - k_1^2)\,|\mathbf{k}_2 - \mathbf{k}_1|^2}$$

$$\times q\left[F_{i\nu}\left(\frac{1+q}{1-q}\right) - F_{i\nu}\left(\frac{1-q}{1+q}\right)\right], \tag{27}$$

where

$$F_{i\nu}(z) = z\Gamma(i\nu)\frac{e^{-\pi\nu}}{2\pi i}\int_{+1,\arg(t)=0}^{(0+)} dt\, t^{-i\nu}\frac{1}{1 - tz}. \tag{28}$$

The integral (28) gives a hypergeometric function,

$$F_{i\nu}(z) = \frac{-z}{\Gamma(2 - i\nu)}\,{}_2F_1(1,\, 1 - i\nu;\, 2 - i\nu,\, z). \tag{29}$$

Equations (27) and (29) in conjunction with (17), (22), and (23) give the desired closed-form expression for the momentum-space Green's function. The notation ${}_2F_1(1,\, 1 - i\nu;\, 2 - i\nu;\, z)$ in (29) denotes that function of $z$ which is analytic in the whole $z$ plane, cut along the segment $+1 \leq z < +\infty$ of the positive real axis and which reduces to the usual hypergeometric series inside the unit circle $|z| < 1$. The function $F_{i\nu}(z)$ has the simple differentiation property

$$\frac{dF_{i\nu}(z)}{dz} = \frac{i\nu}{z}F_{i\nu}(z) - \frac{1}{\Gamma(1 - i\nu)(1 - z)}, \tag{30}$$

and reduces to just a logarithm in the free-particle case, $i\nu = 0$:

$$F_0(z) = \ln(1 - z). \tag{31}$$

The function $F_{i\nu}(z)$ is an analytic function of $\hbar\omega$ in the whole cut plane. Consequently, the only poles of $G(\mathbf{k}_2, \mathbf{k}_1, \omega)$ on the cut $\hbar\omega$ plane are the poles of the gamma function factor $\Gamma(1 - i\nu)$ of (27). But these are simple poles at the Bohr energy levels, as we know they must be.

### ACKNOWLEDGMENTS

The author wishes to acknowledge many helpful discussions with Dr. Lowell Brown and with Dr. Loyal Durand, III.

### APPENDIX

Of course, the fact that the conditions Im $[k(\zeta \pm (\zeta^2 - 1)^{\frac{1}{2}})] > 0$ exclude the points $\zeta$ at which the integrand of (24) has a pole from the interior of the $\zeta$ integration contour, follows from the fact that these inequalities guarantee the convergence (finiteness) of the integral (6) for $D(\mathbf{k}_2, \mathbf{k}_1, \omega)$. However, it may be of interest to check this by direct calculation, and that is the object of this appendix.

[5] This is verified directly in the Appendix.

We make the transformation $\zeta = \cosh\theta$. The branch of the (many-valued) function $\theta = \cosh^{-1}\zeta$ is determined by the conditions that $\theta$ shall vary continuously with $\zeta$ and shall start out real and positive when $\zeta$ is on the initial part (near $\zeta = +\infty$ on the positive real axis) of the $\zeta$ integration contour. The $\zeta$ integration contour is traced out by letting $\theta$ run down the positive real axis from $\theta = +\infty$ to a point on the right of the origin $\theta = 0$, describing an arc above the origin which comes down on the negative real axis at a point on the left of the origin, and then going off to $\theta = -\infty$ along the negative real axis. Thus only $\theta$ values for which Im $(\theta) \geq 0$ are involved. On the first part of the $\theta$ integration path, the quantities $\zeta$, $\theta$, and $(\zeta^2 - 1)^{\frac{1}{2}}$ are real and positive, so $(\zeta^2 - 1)^{\frac{1}{2}} = \sinh\theta$. This relation is preserved along the whole contour by continuity. Thus the conditions Im $[k(\zeta \pm (\zeta^2 - 1)^{\frac{1}{2}})] > 0$ become Im $[ke^{\pm\theta}] > 0$. Writing $\theta = \theta_1 + i\theta_2$ and $k = |k|\, e^{i\delta}$, $0 < \delta < \pi$, the conditions read

$$\sin(\delta \pm \theta_2) > 0. \tag{A1}$$

Since we start out with $0 < \delta \pm \theta_2 < \pi$ ($\theta_2 = 0$ along the first part of the contour), the two angles $\delta \pm \theta_2$ must remain between 0 and $\pi$ in order not to violate (A1). This leads to the condition

$$0 \leq \theta_2 < \min[\delta, \pi - \delta] \tag{A2}$$

on the contour in the $\theta$ plane. (As explained above, negative values of $\theta_2$ may be excluded.)

The conditions (A2) describe the region in the $\theta$ plane in which the $\theta$ integration contour must lie in order to satisfy the inequalities Im $[k(\zeta \pm (\zeta^2 - 1)^{\frac{1}{2}})] > 0$ at all points on the contour. Let us next map this region onto the $\zeta^2$ plane. Write $\zeta^2 = w = w_1 + iw_2 = \cosh^2\theta = \cosh^2(\theta_1 + i\theta_2)$. We find

$$\begin{aligned} w_1 - \tfrac{1}{2} &= \tfrac{1}{2}(\cosh 2\theta_1 \cos 2\theta_2), \\ w_2 &= \tfrac{1}{2}(\sinh 2\theta_1 \sin 2\theta_2). \end{aligned} \tag{A3}$$

We now distinguish two cases:

Case 1:

$$0 < \min[\delta, \pi - \delta] < \tfrac{1}{4}\pi; \tag{A4}$$

Case 2:

$$\tfrac{1}{4}\pi \leq \min[\delta, \pi - \delta] < \tfrac{1}{2}\pi.$$

In Case 1 we find that the integration contour in the $\zeta^2$ plane must fall in the interior of the region to the right of the right-hand branch of the hyperbola

$$(w_1 - \tfrac{1}{2})^2/\cos^2 2\delta - w_2^2/\sin^2 2\delta = \tfrac{1}{4} \tag{A5}$$

(the shaded area, $\mathcal{U}$, of Fig. 1). As min $[\delta, \pi - \delta]$ approaches $\frac{1}{4}\pi$ the boundary of $\mathcal{U}$ approaches coincidence with the vertical line $w = \frac{1}{2}$. When min $[\delta, \pi - \delta]$ increases over the value $\frac{1}{4}\pi$—i.e., as we go over into Case 2—cos $2\delta$ becomes negative and it is seen from Eq. (A3) that the boundary of $\mathcal{U}$ goes over (continuously) into the *left*-hand branch of the hyperbola (A5). Thus in Case 2 the permitted region in the $\zeta^2$ plane for the integration contour is the interior of the region to the right of the *left*-hand branch of the hyperbola (A5).

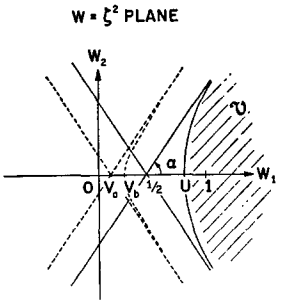


FIG. 1. The shaded area, $\mathcal{U}$, is the image in the $\zeta^2$ plane of the permitted region for the integration contour of (24). The dotted curve is the locus of possible values, $V$, of $\zeta^2$ at which the denominator of the integrand of (24) vanishes. The drawing is for Case 1. For $\alpha$, $U$, $V_b$, and $V_a$, see Eqs. (A9)–(A12).

Now the denominator of the integrand of (24) will vanish only if

$$\zeta^2 = \frac{k^4 - 2k^2\mathbf{k}_2\cdot\mathbf{k}_1 + k_2^2k_1^2}{k^2\,|\mathbf{k}_2 - \mathbf{k}_1|^2}.$$

We write

$$V = V_1 + iV_2 = \frac{k^4 - 2k^2\mathbf{k}_2\cdot\mathbf{k}_1 + k_2^2k_1^2}{k^2\,|\mathbf{k}_2 - \mathbf{k}_1|^2}, \tag{A6}$$

and plot the locus of possible $V$ values on the $\zeta^2$ plane. We have

$$\begin{aligned} V_1 + \frac{2\mathbf{k}_2\cdot\mathbf{k}_1}{|\mathbf{k}_2 - \mathbf{k}_1|^2} &= \frac{|k|^2 + k_2^2k_1^2/|k|^2}{|\mathbf{k}_2 - \mathbf{k}_1|^2}\cos 2\delta, \\ V_2 &= \frac{|k|^2 - k_2^2k_1^2/|k|^2}{|\mathbf{k}_2 - \mathbf{k}_1|^2}\sin 2\delta. \end{aligned} \tag{A7}$$

In Case 1 we have cos $2\delta > 0$ and the possible $V$ values lie on the right-hand branch of the hyperbola

$$\left(V_1 + \frac{2\mathbf{k}_2\cdot\mathbf{k}_1}{|\mathbf{k}_2 - \mathbf{k}_1|^2}\right)^2(\cos^2 2\delta)^{-1} - \frac{V_2^2}{\sin^2 2\delta} = \frac{4k_2^2k_1^2}{|\mathbf{k}_2 - \mathbf{k}_1|^2} \tag{A8}$$

(the dotted curve in Figure 1). In the limit as $\delta$ increases toward $\frac{1}{4}\pi$, this locus goes over into the vertical straight line $V = -2\mathbf{k}_2\cdot\mathbf{k}_1/|\mathbf{k}_2 - \mathbf{k}_1|^2$. As $\delta$ increases beyond $\frac{1}{4}\pi$—i.e., as we go into Case 2—cos $2\delta$ becomes negative and the locus of possible $V$ values goes (continuously) over into the *left*-hand branch of the hyperbola (A8).

We will show that the locus of possible $V$ values always lies outside (or at worst, coincides with the

boundary of) the permitted region in the $\zeta^2$ plane for the integration contour. When we take the square root, forming $\zeta = \cosh\theta$, the region $\mathcal{U}$ will go into the permitted region for the integration contour in the $\zeta$ plane. This region (as noted in the text) is the interior of the region to the right of the right-hand branch of the hyperbola (8). The locus of possible values of $V^{\frac{1}{2}}$ will split up into two disjoint pieces, corresponding to the two roots of $V$, but these pieces will still lie outside (or at worst coincide with the boundary of) the permitted region for the integration contour. Now these $V^{\frac{1}{2}}$ values are just the possible $\zeta$ values at which the denominator of (24) has a pole. Thus all $\zeta$ values at which the denominator of (24) has a pole will lie outside the permitted region for the $\zeta$ integration contour and hence will lie outside the contour itself.

We will establish this result here only for Case 1. Case 2 can be argued in the same way. We first note that the asymptotes of the two hyperbola (A5) and (A8) are parallel. In both cases the asymptotes are inclined to the real axis at angles $\pm\alpha$, where

$$\alpha = \min\,[2\delta,\, 2\pi - 2\delta]. \tag{A9}$$

Hence to show that the right-hand branch of the hyperbola (A8) lies outside the region $\mathcal{U}$ it suffices to show that (a) the intercept on the real axis of the right-hand branch of the hyperbola (A8) does not lie to the right of the intercept on the real axis of the right-hand branch of the hyperbola (A5), and (b) the "center" of the hyperbola (A8) does not lie to the right of the "center" of the hyperbola (A5) (see Fig. 1). The intercept on the real axis of the right-hand branch of the hyperbola (A5) occurs at

$$U = \cos^2\delta, \tag{A10}$$

and the center occurs at $w = \frac{1}{2}$. Now the intercept of the right-hand branch of the hyperbola (A8) occurs at

$$\begin{aligned} V_b &= \frac{-2\mathbf{k}_2\cdot\mathbf{k}_1 + 2k_2k_1\cos 2\delta}{k_2^2 - 2\mathbf{k}_2\cdot\mathbf{k}_1 + k_1^2} \\ &= 1 - \frac{k_2^2 - 2\mathbf{k}_2\cdot\mathbf{k}_1\cos 2\delta + k_1^2}{k_2^2 - 2\mathbf{k}_2\cdot\mathbf{k}_1 + k_1^2} \\ &\le 1 - \frac{k_2^2 - 2\mathbf{k}_2\cdot\mathbf{k}_1\cos 2\delta + k_1^2}{(k_2 + k_1)^2} \\ &\le \frac{4k_2k_1}{(k_2 + k_1)^2}\cos^2\delta \le \cos^2\delta = U, \end{aligned} \tag{A11}$$

i.e., this intercept does not lie to the right of the intercept of the boundary of $\mathcal{U}$. This establishes (a). The center of the hyperbola (A8) is located on the real axis at

$$\begin{aligned} V_a &= -\frac{2\mathbf{k}_2\cdot\mathbf{k}_1}{|\mathbf{k}_2 - \mathbf{k}_1|^2} = \frac{k_2^2 - 2\mathbf{k}_2\cdot\mathbf{k}_1 + k_1^2 - (k_2^2 + k_1^2)}{k_2^2 - 2\mathbf{k}_2\cdot\mathbf{k}_1 + k_1^2} \\ &= 1 - \frac{k_2^2 + k_1^2}{k_2^2 - 2\mathbf{k}_2\cdot\mathbf{k}_1 + k_1^2} \le 1 - \frac{k_2^2 + k_1^2}{(k_2 + k_1)^2} \\ &\le \frac{2k_2k_1}{(k_2 + k_1)^2} \le \tfrac{1}{2}, \end{aligned} \tag{A12}$$

and so does not lie to the right of the center of the hyperbola (A5). This establishes (b). The conclusion that the conditions $\mathrm{Im}\,[k(\zeta \pm (\zeta^2 - 1)^{\frac{1}{2}})] > 0$ exclude the points $\zeta$ at which the denominator of the integrand of (24) has a pole from the interior of the $\zeta$ integration contour now follows.

# A Model of Interacting Radiation and Matter*

CHARLES R. WILLIS
*Boston University, Boston, Massachusetts*
(Received 31 March 1964)

We investigate the long-time behavior of a model consisting of $N$ two-level atoms in a lossless cavity. The Hamiltonian of our system contains the radiation oscillators in addition to the matter Hamiltonian and the usual $\int \mathbf{j} \cdot \mathbf{A} dv$ interaction term. In order to treat the system perturbatively, it would be necessary to remove the tremendous degeneracy of the system. Since this is prohibitively difficult, and since we are interested in the long-time behavior of the system, we solve the quantum mechanical Liouville equation directly for a wide class of physically important initial distribution functions. We show the effective expansion parameter is $\bar{\gamma}N\bar{\gamma}$ where $\bar{\gamma}$ is a dimensionless atomic dipole moment and $N$ is the number of atoms. In the lowest order we find the self-consistent field approximation. In the next order, particle-field correlations appear. We explicitly solve the equations of motion for the particle-field correlations in terms of the average quantities that appear in the self-consistent field approximation. We show the self-consistent field approximation consists of five first-order differential equations. Next we show the equations of motion for the density matrix of the system correct to order $(\bar{\gamma}N\bar{\gamma})^2$ are equivalent to eight first-order differential equations. The three additional equations are needed to describe the three second moments of the density matrix of the electromagnetic field that appear in second order. Our lowest-order microscopic equations are equivalent to semiphenomenological theories and our higher-order equations contain only the measurable second-order moments of the electromagnetic field in addition to the variables that appear in semiphenomenological theories.

## I. INTRODUCTION

IN this paper we derive the equations of motion satisfied by the quantum mechanical density matrix of a model of interacting atoms and radiation. We explicitly solve for the particle-field correlations in terms of the one-particle density matrix and first and second moments of the electromagnetic field. We thus obtain for the one-particle density matrix a nonlinear equation which to lowest order satisfies the self-consistent field approximation.

Our model of interacting radiation and matter consists of $N$ distinguishable two-level atoms interacting with the cavity modes in a lossless cavity. The model of the matter system and the interaction Hamiltonian is essentially the same as the model introduced by Dicke.[1] Our model differs in one essential respect from Dicke's model in that we include the Hamiltonian of the radiation field in our system Hamiltonian. We do this because we wish to find the long-time behavior of the system including saturation effects corresponding to repeated absorptions and subsequent re-emissions. Dicke asked questions about driven systems and spontaneous emission which he answered by the help of the introduction of a constant of the motion whose eigenvalues he called the "cooperation number." Although we do not make explicit use of this conservation law our final results are consistent with it.

For a careful discussion of the matter system alone we refer the reader to Dicke's paper. However, because the degeneracy is the single most important property of the system we will present here a brief account of the effect that inclusion of the radiation field in the Hamiltonian has on the order of the degeneracy. Assume we have $N$ two-level atoms with the same two-level energy difference, $\hbar\omega_0$, which is essentially equivalent to the energy of a quantum of one of the cavity modes, $\hbar\Omega$. The lowest eigenstate of the total system is nondegenerate and the radiation oscillator and all the atoms are in their ground states. The unperturbed energy, $\hbar\omega_0$, is $(N+1)$-fold degenerate. The degenerate states are one state with the radiation oscillator in its first excited state and all $N$ atoms in their ground state, and $N$ states with the radiation oscillator in its ground state and one atom at a time in its excited state. The distinguishability which leads to the degeneracy arises from the fact that wavefunctions of the separate atoms do not overlap. The unperturbed energy $2\hbar\omega_0$ is $\{1+N+\frac{1}{2}N(N-1)/2\}$-fold degenerate. The degenerate states are one state with the radiation oscillator in its second excited state and the atoms all in their ground state, $N$ states with the radiation oscillator in its first excited state and one atom at a time in its excited state, and

* The research reported in this paper was sponsored in part by the Air Force Cambridge Research Laboratories, Office of Aerospace Research.
[1] R. H. Dicke, Phys. Rev. 93, 99 (1954).

$N(N-1)/2$ states corresponding to the number of ways that two atoms can be singly excited while the radiation oscillator is in its ground state.

The degeneracy of the unperturbed energy eigenvalue, $m\hbar\omega_0$, is

$$\sum_{j=0}^{m} \frac{N!}{j!\,(N-j)!}. \tag{1.1}$$

In order to apply perturbation theory to the system we would first have to remove the degeneracy by diagonalizing a matrix for each integer, $m \leq N$, whose dimension is given by Eq. (1.1).

The degeneracy arises physically because "each atom sees all the $(N-1)$ other atoms through the electromagnetic field." To anticipate our resolution of the problem we replace the preceding statement by the statement "each atom sees all the $(N-1)$ other atoms to the lowest order through the average electromagnetic field." The self-consistent field approximation which we shall refer to as SCFA is nonperturbative and leads to equations that are meaningful for all time. Our treatment of the problem is purely quantum mechanical and it is valid for systems containing a single quantum. The condition required for the validity of the SCFA is that there be an upper limit to the particle–field correlations present initially. However, if the particle–field correlations are initially zero they will grow to nonzero values by higher-order corrections to the SCFA which we shall explicitly calculate. Our lowest-order equations are essentially equivalent to the semiphenomenological theories derived by Jaynes and Cummings[2] and Tang.[3]

In Sec. II we introduce the Hamiltonian of our model and we discuss further the degeneracy of the system. We show in Sec. III that the solution to lowest order of the quantum mechanical Liouville equation is the SCFA.

In Secs. IV and V we find the equations of motion satisfied by the quantum mechanical Liouville equation to second order and we solve these equations in terms of the variables which appear in the lowest order which is the SCFA.

We show in Sec. VI the relationship between the lowest-order equations of our theory and the semiphenomenological theories,[2,3] and we express the higher order corrections to the SCFA in terms of phenomenological variables.

In the Appendix we find the equations of motion satisfied by the electromagnetic field fluctuations.

[2] E. T. Jaynes and F. W. Cummings, Proc. IEEE **51**, 89 (1963).
[3] C. L. Tang, J. Appl. Phys. **34**, 2935 (1964).

## II. HAMILTONIAN OF THE MODEL

We consider $N$ two-level atoms with energy levels $E_a = \frac{1}{2}(\hbar\omega_0)$ and $E_b = -\frac{1}{2}(\hbar\omega_0)$. The Hamiltonian for the atoms is

$$h(N) = \hbar\omega_0\left\{\sum_{\alpha}^{N} (\sigma_\alpha^\dagger\sigma_\alpha - \tfrac{1}{2})\right\}, \quad H_{\rm em} = \sum_{\alpha}^{N} \frac{P_\alpha^2}{2M}, \tag{2.1}$$

where

$$[\sigma_\alpha, \sigma_\alpha^\dagger]_+ = 1, \quad [\sigma_\alpha, \sigma_\alpha]_+ = [\sigma_\alpha^\dagger, \sigma_\alpha^\dagger]_+ = 0,$$

and

$$[\sigma_\alpha^\dagger, \sigma_\beta^\dagger] = [\sigma_\alpha^\dagger, \sigma_\beta] = [\sigma_\alpha, \sigma_\beta] = 0 \quad \text{for} \quad \alpha \neq \beta.$$

The plus subscript indicates an anticommutator. The vanishing of the commutators for different molecules represents the fact that we are treating the atoms as distinguishable. We are assuming the density of atoms is sufficiently low that the overlap of the wavefunctions is negligible and thus the effects of symmetry may be neglected. The momentum of the center of mass of the $\alpha$th particle is $P_\alpha$.

A convenient representation of the matter system is a Kronecker product of single-particle spaces. In the single-atom space we may represent the algebra of the operators in the following form

$$\sigma^\dagger = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \sigma = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

$$\sigma^\dagger\sigma = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \sigma^c = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{2.2}$$

where

$$\sigma^c \equiv [\sigma, \sigma^\dagger] \quad \text{and} \quad \psi_a = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \psi_b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The Hamiltonian for the electromagnetic field is

$$H_f = \tfrac{1}{2} \sum_k p_k^2 + \Omega_k^2 q_k^2 \tag{2.3}$$

where

$$[q_k, p_{k'}] = i\hbar\delta_{kk'} \quad \text{and} \quad [q_k, q_{k'}] = [p_k, p_{k'}] = 0.$$

The vector potential and the electromagnetic field in the cavity are given by

$$\mathbf{A}(xt) = c(4\pi)^{\frac{1}{2}} \sum_k \boldsymbol{\varepsilon}_k q_k(t) E_k(x),$$

$$\mathbf{E}(xt) = -(4\pi)^{\frac{1}{2}} \sum_k \boldsymbol{\varepsilon}_k p_k(t) E_k(x),$$

$$\mathbf{B}(xt) = -c(4\pi)^{\frac{1}{2}} \sum_k (\boldsymbol{\varepsilon}_k \times \nabla) q_k(t) E_k(x). \tag{2.4}$$

The $E_k(x)$ are the eigenfunctions of the cavity and $\boldsymbol{\varepsilon}_k$ is a unit vector in the plane of polarization.

Hamilton's equations of motion for $H_f$ give the free field Maxwell's equations. It is often convenient to represent the electromagnetic fields in terms of creation and annihilation operators,

$$H_f = \sum_k (\hbar\Omega_k a_k^\dagger a_k + \tfrac{1}{2}), \tag{2.5}$$

where

$$[a_k, a_{k'}^\dagger] = \delta_{kk'},$$
$$a_k = (2\hbar\Omega_k)^{-\frac{1}{2}}(p_k - i\Omega_k q_k),$$
$$a_k^\dagger = (2\hbar\Omega_k)^{-\frac{1}{2}}(p_k + i\Omega_k q_k).$$

The interaction Hamiltonian, $H_i$, is

$$H_i = \frac{1}{c}\int j\cdot A\, dx = \sum_k \sum_\alpha^N \{\sigma_\alpha \langle b\,|j_k(X_\alpha)|\,a\rangle + \sigma_\alpha^\dagger \langle a\,|j_k(X_\alpha)|\,b\rangle\} q_k E_k(X_\alpha), \tag{2.6}$$

where $X_\alpha$ is the center of mass of the $\alpha$th particle and

$$\langle a\,|j_k(X_\alpha)|\,b\rangle \equiv (e/2m)[E_k(X_\alpha)]^{-1}\int \psi_a^*(r) \times \{\boldsymbol{\varepsilon}_k\cdot\mathbf{p}E_k(\mathbf{X}_\alpha + \mathbf{r}) + E_k(\mathbf{X}_\alpha + \mathbf{r})\boldsymbol{\varepsilon}_k\cdot\mathbf{p}\}\psi_b(r)\, d\mathbf{r}$$

where $\mathbf{p} = (\hbar/i)\nabla_r$.

Equation (2.6) is the exact expression for the off-diagonal components of the particle current vector. The usual $A^2$ term is diagonal in the particle operators and constitutes a perturbation of the free field which is quadratic in annihilation and creation operators and can be removed to any order by a suitable canonical transformation. Since the $A^2$ term is usually small we neglect it. Combining Eqs. (2.1), (2.3), and (2.6), we obtain

$$H = h(N) + H_f + H_{cm} + H_i = H_0 + H_i. \tag{2.7}$$

Since the effect of the electromagnetic interactions on the center-of-mass motion is usually negligible we might consider dropping $H_{cm}$. However, $H_i$ depends on the center of mass through $X_\alpha$ which means the center-of-mass motion can and does have an appreciable effect on the internal degrees of freedom.[4]

In this paper we study the effect of correlations on the collective behavior of the system. In order to temporarily avoid the complications of the center-of-mass motion, we consider waves that are large compared to the dimensions of the cavity. Then $E_k(X_\alpha) \sim V^{-\frac{1}{2}}$ is independent of $X_\alpha$ and $H_i$ becomes

$$H_i = \sum_k \sum_\alpha^N q_k\{\Gamma_k\sigma_k + \Gamma_k^*\sigma_k^\dagger\}, \tag{2.8}$$

where

$$\Gamma_k = V^{-\frac{1}{2}}(e/m)\langle a\,|\boldsymbol{\varepsilon}_k\cdot\mathbf{p}|\,b\rangle.$$

To avoid an excess of subscripts we will consider a single mode. When the calculations are complete, the subscript $k$ can be reinserted by inspection. In order to compare our results with semiphenomenological theories we replace the matrix elements of the current operator by the matrix elements of the dipole operator and we obtain

$$H_i = \gamma p \sum_\alpha (\sigma_\alpha + \sigma_\alpha^\dagger) \equiv \gamma p \sum_\alpha \mu_\alpha, \tag{2.9}$$

where $\gamma = eV^{-\frac{1}{2}}\langle a|\,\boldsymbol{\varepsilon}\cdot\mathbf{r}\,|b\rangle$, and where we have used $\langle a\,|\boldsymbol{\varepsilon}\cdot\mathbf{p}|\,b\rangle = im\omega_0\langle a\,|\boldsymbol{\varepsilon}\cdot\mathbf{r}|\,b\rangle$ and $i\omega_0 q = p$ for $\omega_0 = \Omega$.

The reduction of the rigorous interaction Eq. (2.6) to Eq. (2.8) involves only one essential assumption, namely the neglect of the motion of the center of mass. Thus we neglect the effects of the Doppler broadening in the present paper. Note that this assumption is valid in any case when the wavelength of the radiation is greater than the dimensions of the container.

When we express $p$ in terms of annihilation and creation operators using Eq. (2.5), we observe that $H_i$ consists of two types of terms,

$$H_i = \gamma(\tfrac{1}{2}\hbar\Omega)^{\frac{1}{2}}(a + a^\dagger)\sum_\alpha^N (\sigma_\alpha + \sigma_\alpha^\dagger) = H_d + H_{nd}, \tag{2.10}$$

where

$$H_d = \gamma(\tfrac{1}{2}\hbar\Omega)^{\frac{1}{2}}\sum_\alpha^N (a^\dagger\sigma_\alpha + a\sigma_\alpha^\dagger),$$
$$H_{nd} = \gamma(\tfrac{1}{2}\hbar\Omega)^{\frac{1}{2}}\sum_\alpha^N (a^\dagger\sigma_\alpha^\dagger + a\sigma_\alpha).$$

Since $H_{nd}$ creates (destroys) excited atomic states at the same time as it creates (destroys) photons, it has no matrix elements that conserve the unperturbed energy. Consequently, it can be diagonalized to order $\gamma^2$ with ease. Actually semiphenomenological theories implicitly neglect $H_{nd}$ when they discard terms whose frequency dependence is proportional to $\pm 2\omega_0$ compared to $\omega_0$. We retain $H_{nd}$ because it is inconsistent to neglect it when we go beyond the SCFA as we do in Secs. IV and V.

To see how the problem of degeneracy discussed in the Introduction arises we investigate the equation of motion satisfied by $H_d$,

[4] W. E. Lamb, Jr., Phys. Rev. 134, A 1429 (1964).

$$\dot{H}_d = [H_d, H] \approx [H_d, H_0]$$
$$= \gamma(\tfrac{1}{2}\hbar\Omega)^{\frac{1}{2}} \sum_{\alpha}^{N} \sum_{\beta}^{N} [a^\dagger\sigma_\alpha + a\sigma_\alpha^\dagger, \hbar\omega_0\sigma_\beta^\dagger\sigma_\beta + \hbar\Omega a^\dagger a]$$
$$= \gamma(\tfrac{1}{2}\hbar\Omega)^{\frac{1}{2}} \sum_{\alpha} \{\hbar\omega_0 a[\sigma_\alpha^\dagger, \sigma_\alpha^\dagger\sigma_\alpha] + \hbar\omega_0 a^\dagger[\sigma_\alpha, \sigma_\alpha^\dagger\sigma_\alpha]$$
$$+ \hbar\Omega\sigma_\alpha^\dagger[a, a^\dagger a] + \hbar\Omega\sigma_\alpha[a^\dagger, a^\dagger a]\}$$
$$= \gamma(\tfrac{1}{2}\hbar\Omega)^{\frac{1}{2}} \sum_{\alpha}^{N} \{\hbar\omega_0(a^\dagger\sigma_\alpha^c\sigma_\alpha - a\sigma_\alpha^\dagger\sigma_\alpha^c)$$
$$+ \hbar\Omega(a\sigma_\alpha^\dagger - \sigma_\alpha a^\dagger)\}$$
$$= \gamma(\tfrac{1}{2}\hbar\Omega)^{\frac{1}{2}}(\hbar\omega_0 - \hbar\Omega) \sum_{\alpha}^{N} (a^\dagger\sigma_\alpha - a\sigma_\alpha^\dagger). \qquad (2.11)$$

We neglected the effect of $H_{nd}$ on $H_d$ because it is of order $\gamma^2$ and can not mitigate in any way the problem of removing the lowest-order degeneracy. Thus for $\omega_0 = \Omega$, $H_d$ is a constant of motion to order $\gamma^2$. This means the lowest-order perturbation theory is completely degenerate and removing the degeneracy is equivalent to diagonalizing for each unperturbed energy eigenvalue, $m\hbar\omega_0$, a matrix whose dimension is given by Eq. (1.1).

Since the lowest-order perturbation theory requires the diagonalization of so many extremely large dimensional matrices, we should investigate why we want the eigenvalues and eigenfunctions and what we would do with them if we had them. First we want to discuss the long-time behavior of a system with absorptions and repeated re-emissions. Ordinary perturbation theory is not valid for these times. The second reason we would like both eigenfunctions and eigenvalues is to construct the unitary transformation that diagonalizes the quantum mechanical Liouville equation for the density matrix of the system.

We conclude this section with another aspect of the degeneracy of the system. Consider once again the Hamiltonian, $H = H_0 + \gamma H_d$, where we have neglected the unimportant $H_{nd}$. The unperturbed energy, $E^0 = \hbar\omega_0$, is $(N+1)$-fold degenerate and the perturbation, $H_d$, has constant matrix elements, $\gamma$ and 0, between the degenerate states. With the use of Wigner–Brillouin perturbation theory we can show one of the energy eigenvalues is $-\gamma N$ where $N$ is the number of atoms. This phenomenon arises because each atom sees all the other atoms through the electric field. If we change this statement to read "each atom sees all the other atoms to the lowest order through their average behavior" we are able to solve the problem as we shall show in the next section.

## III. SELF-CONSISTENT FIELD APPROXIMATION

We are led to consider the density matrix for two reasons; first, we expect the fact that each atom sees the averaged behavior only can be represented statistically, and second, many of the interesting cases are problems of statistical mechanics which require ensembles rather than single quantum states.

The density matrix $F_N(1, 2, \cdots N, q)$ for $N$ atoms plus the radiation oscillator satisfies the quantum mechanical Liouville equation

$$i\hbar(\partial F_N/\partial t) + [F_N, H_N] = 0, \qquad (3.1)$$

where

$$H_N = h(N) + H_f + H_i = H_0 + H_i;$$
$$\mathrm{tr}_{1,2\cdots N,q} F_N = 1.$$

Our notation indicates $F_N$ is an operator in the Hilbert space of the first, second, $\cdots$, $N$th particle and in the Hilbert space denoted by $q$ of the radiation oscillator. We do not need an explicit representation of a matrix element of $F_N$; however, a typical one has the form

$$\langle +-++\cdots, q\,|F_N|\,\bar{q}, +--+\cdots\rangle, \qquad (3.2)$$

where a $+$ or $(-)$ in the $j$th position indicates the $j$th atom is in its excited (unexcited) state. For given $q$ and $\bar{q}$ there are $(2^N)^2$ matrix elements corresponding to all the ways of writing $+$'s and $-$'s. In the Hilbert space of the radiation oscillator the matrix is infinite. It is denumerably infinite if we choose the number representation or continuously infinite if we use the $q$ representation. We are able to express all our results as operators or as traces, both of which are independent of representation. Thus the argument of an operator denotes not matrix elements but operator functional dependence.

When we take the trace of Eq. (3.1) over the coordinates of the $N$th atom we obtain

$$i\hbar(\partial F_{N-1}/\partial t) + [F_{N-1}, H_{N-1}] = \gamma\,\mathrm{tr}_N[p\mu_N, F_N], \qquad (3.3)$$

and after $(N-s)$ similar operations we obtain

$$i\hbar(\partial F_s/\partial t) + [F_s, H_s]$$
$$= (N-s)\gamma\,\mathrm{tr}_{(s+1)}[p\mu_{(s+1)}, F_{(s+1)}]. \qquad (3.4)$$

In deriving Eq. (3.4) we used the fact that $F$ is a symmetric function of its arguments.

We obtain the equation for the density matrix of $s$ atoms by taking the trace of Eq. (3.4) over the oscillator coordinates

$$i\hbar(\partial\rho_s/\partial t) + [\rho_s, h(s)] + \gamma\,\mathrm{tr}_q\left[F_s, p\sum_{i=1}^{s}\mu_i\right]$$
$$= (N-s)\gamma\,\mathrm{tr}_{q,(s+1)}[p\mu_{(s+1)}, F_{(s+1)}] = 0, \qquad (3.5)$$

where $\rho_s \equiv \mathrm{tr}_q F_s$. The trace vanishes since $p\mu_{(s+1)}$ depends only on the coordinates of the $(s+1)$th atom and the radiation oscillator, and the trace is over both variables.

We add the term

$$\gamma\langle p\rangle\Big[\rho_s, \sum_i^s \mu_i\Big]$$

to both sides of Eq. (3.5) and we obtain

$$i\hbar(\partial\rho_s/\partial t) + [\rho_s, h(s)] + \gamma\langle p\rangle\Big[\rho_s, \sum_i^s \mu_i\Big] = \gamma\Big[\sum_i^s \mu_i, \Pi_1(1, 2, \cdots, s) - \rho_s\langle p\rangle\Big], \quad (3.6)$$

where

$$\langle p\rangle \equiv \mathrm{tr}_q\, pR(q) \equiv \mathrm{tr}_{1,q}\, pF_1(1, q).$$

The expression

$$\Pi_i(1, 2, \cdots s) = \mathrm{tr}_q\, p^i F_s(1, 2, \cdots s, q)$$

is an operator in the space of $s$ atoms and is a measure of particle–field correlations.

To solve Eq. (3.6) we assume a solution of the form[5]

$$F_s = \rho_1(1)\rho_1(2)\cdots\rho_1(s)R(q) + (\gamma N\gamma)\sum_i^s \chi(i, q)\prod_{j\neq i}^s \rho_1(j) + (\gamma N\gamma)^2\cdots. \quad (3.7)$$

In this section we are concerned with only the zeroth-order term which represents a lack of particle–particle and particle–field correlations. In the next section we derive the equation satisfied by $\chi(i, q)$ and explain the choice of the expansion parameter $\gamma N\gamma$.

When the first term of Eq. (3.7) is substituted in Eq. (3.6), we obtain

$$i\hbar(\partial\rho_s/\partial t) + [\rho_s, h(s)] + \gamma\langle p\rangle\Big[\rho_s, \sum_i^s \mu_i\Big] = 0,$$

or equivalently,

$$i\hbar(\partial\rho_1/\partial t) + [\rho_1, h(1)] + \gamma\langle p\rangle[\rho_1, \mu_1] = 0. \quad (3.8)$$

We need the operator equation of motion for $R(q)$ to obtain an equation of motion for $\langle p\rangle$. Since $R(q)$ is defined as $\mathrm{tr}_1 F_1(1, q)$, we take the trace of Eq. (3.4) for $(s = 1)$ over the particle coordinate and we obtain

$$i\hbar(\partial R/\partial t) + [R, H_f] = (N-1)\gamma\,\mathrm{tr}_1[p\mu_1, F_1(1, q)]. \quad (3.9)$$

[5] N. N. Bogoliubov, "Problems of a Dynamical Theory in Statistical Physics" (translated by E. K. Gora), in *Studies in Statistical Mechanics*, edited by J. De Boer and G. E. Uhlenbeck (North-Holland Publishing Company, Amsterdam, 1962), pp. 5–118.

Eq. (3.9) is completely rigorous. To find the equation of motion satisfied by $R(q)$ to lowest order, we replace $F_1(1, q)$ by $\rho_1(1)R(q)$,

$$i\hbar(\partial R/\partial t) + [R, H_f] = (N-1)\gamma\langle\mu\rangle[p, R], \quad (3.10)$$

where

$$\langle\mu(t)\rangle \equiv \mathrm{tr}_1\, \mu_1\rho_1(1, t).$$

The quantity $\langle\mu(t)\rangle$ is the dimensionless dipole moment per atom. The total dipole moment of the system is $N\gamma V^{\frac{1}{2}}\langle\mu\rangle$. For convenience in keeping track of the powers of $\gamma$ and $N$ we use dimensionless atomic quantities. In Sec. VI we express all of our equations in terms of macroscopic quantities.

To find the equation of motion to lowest order satisfied by $\langle p\rangle$ we multiply Eq. (3.10) by the operator $p$ and take the trace over the oscillator variables. We obtain

$$i\hbar(\partial\langle p\rangle/\partial t) + \mathrm{tr}_q\, p[R, \tfrac{1}{2}(q^2\Omega^2)] = 0. \quad (3.11)$$

We may simplify Eq. (3.11) to read

$$\partial\langle p\rangle/\partial t + \Omega^2\langle q\rangle = 0, \quad (3.12)$$

where we have used

$$\mathrm{tr}_q\, p[R, \tfrac{1}{2}(q^2\Omega^2)] = -\Omega^2\,\mathrm{tr}_q\, q[q, p]R = i\hbar\Omega^2\langle q\rangle, \quad \langle q\rangle \equiv \mathrm{tr}_q\, qR.$$

We obtain the equation for $\langle q\rangle$ by multiplying Eq. (3.10) by $q$ and taking the trace over the oscillator variable. The equation is

$$i\hbar(\partial\langle q\rangle/\partial t) + \mathrm{tr}_q\, q[R, \tfrac{1}{2}p^2] = N\gamma\langle\mu\rangle\,\mathrm{tr}_q\, q[p, R],$$

which may be written

$$\partial\langle q\rangle/\partial t - \langle p\rangle = N\gamma\langle\mu\rangle. \quad (3.13)$$

We obtain an equation of motion for $\langle p\rangle$ alone by substituting Eq. (3.13) in Eq. (3.12),

$$\partial^2\langle p\rangle/\partial t^2 + \Omega^2\langle p\rangle = -N\gamma\Omega^2\langle\mu\rangle. \quad (3.14)$$

Equations (3.8) and (3.14) together constitute a complete theory which is the SCFA.

The SCFA must fulfill two essential requirements to be a good solution of a physical problem. The first requirement is dynamical and states that a degree of freedom is more influenced by a large number of degrees of freedom than it is by a few nearest neighbors. In a plasma this requirement is met by the long-range nature of the Coulomb force. In the present problem the $N$ dependence of the right-hand side of Eq. (3.14) indicates that all atoms contribute equally to the electric field $\langle p\rangle$, and each atom sees all the other atoms through

$\langle p\rangle$. The second requirement is statistical and states that initially there can be no zeroth-order particle–particle or particle–field correlations. If these two requirements are met then the SCFA is a good approximation. These two conditions are independent of the classical or quantal nature of the problem. Our treatment of all degrees of freedom throughout the paper is purely quantum mechanical. To emphasize this we point out that, even if the total energy of the system were a single quantum, the SCFA would be a good solution as long as the initial density matrix had no particle–particle or particle–field correlations. The $N$ that appears in Eq. (3.14) and throughout the paper is the constant total number of atoms not the number of excited atoms. In other words, the SCFA is a particular solution of our quantum mechanical problem. This discussion has been necessary because the SCFA represented by Eqs. (3.8) and (3.14) has often been referred to erroneously as a "semiclassical" theory. This has happened because of the semiphenological nature of some derivations of the SCFA.

The SCFA has replaced the unmanageable degeneracy of the exact problem by a nonlinear equation for the density matrix of a single atom. Although this equation is difficult to solve in general, it represents a great advance in the tractability of our problem. Originally we found that the lowest-order perturbation theory required almost an exact solution of the full problem. The SCFA has the important property that it makes qualitative sense (remains bounded, etc.) for long times even if we neglect the higher-order terms. We now show that even if particle–field correlations are initially zero they grow to nonzero values.

## IV. PARTICLE-FIELD CORRELATIONS

It is not necessary to solve for $\chi(1, q)$ in detail to go beyond the SCFA, since we need to know only the first conditional moment $\Pi_1(1, t)$. In this section we find the equation of motion satisfied by $\Pi_1(1, t)$ and we solve the equation in terms of variables that appear in the SCFA. First, however, we will justify the use of $\gamma N\gamma$ as the expansion parameter. From Eq. (3.14) we observe that $\langle p\rangle$ is proportional to $N\gamma$; therefore, the expansion parameter in Eq. (3.8) is proportional to $\gamma N\gamma$. We show in this section that the corrections to the SCFA are proportional to $(\gamma N\gamma)^2$.

We rewrite Eq. (2.9) to obtain an expression for the dimensionless coupling constant,

$$\frac{H_i}{\hbar\Omega} = \gamma(2\hbar\Omega)^{-\frac{1}{2}}(a + a^\dagger)\sum_\alpha^N (\sigma_\alpha + \sigma_\alpha^\dagger)$$

$$= \tilde{\gamma}(a + a^\dagger)\sum_\alpha^N (\sigma_\alpha + \sigma_\alpha^\dagger).$$

The condition for the validity of our expansion procedure is

$$\tilde{\gamma}N\tilde{\gamma} \equiv (N/V)e^2\,|\langle a\,|\boldsymbol{\varepsilon}\cdot\mathbf{r}|\,b\rangle|^2\,(2\pi/\hbar\Omega) \ll 1,$$

where we have used the dimensionless parameter $\tilde{\gamma} = \gamma(2\hbar\Omega)^{-\frac{1}{2}}$. Since $\gamma N\gamma$ is just $2\hbar\Omega(\tilde{\gamma}N\tilde{\gamma})$, it is not necessary to go over to dimensionless variables. Thus, $\gamma N\gamma$ "small" means small compared with $2\hbar\Omega$.

We obtain the equation of motion for $\Pi_1(1, t)$ by multiplying Eq. (3.4) for $(s = 1)$ by $p$ and taking the trace over the oscillator variable

$$i\hbar\frac{\partial\Pi_1(1, t)}{\partial t} + [\Pi_1(1), h(1)] + \mathrm{tr}_q\, p[F_1, H_f] + \gamma[\Pi_2(1, t), \mu_1]$$
$$= (N - 1)\gamma\,\mathrm{tr}_{q,2}\, p[p\mu_2, F_2]$$
$$= (N - 1)\gamma\,\mathrm{tr}_2\,[\mu_2, \Pi_2(2, t)] = 0. \qquad (4.1)$$

We can simplify the third term of Eq. (4.1),

$$\mathrm{tr}_q\, p[F_1, H_f] = -\mathrm{tr}_q\,[p, H_f]F_1 = -\tfrac{1}{2}\Omega^2\,\mathrm{tr}_q\,[p, q^2]F_1$$
$$= i\hbar\Omega^2\,\mathrm{tr}_q\, qF_1 \equiv i\hbar\Omega^2\Pi^1(1, t), \qquad (4.2)$$

where we have used the definition

$$\Pi^i(1, 2, \cdots s) \equiv \mathrm{tr}_q\, q^i F_s(1, 2, \cdots s, q).$$

When Eq. (4.2) is substituted in Eq. (4.1), we obtain

$$i\hbar\frac{\partial\Pi_1(1, t)}{\partial t} + [\Pi_1(1, t), h(1)] + i\hbar\Omega^2\Pi^1(1, t) + \gamma[\Pi_2(1, t), \mu_1] = 0. \qquad (4.3)$$

We introduce our expansion

$$F_1 = \rho(1)R(q) + (\gamma N\gamma)\chi + (\gamma N\gamma)^2\cdots,$$

where $\rho$ without a subscript is $\rho_1$, into Eq. (4.3) and we obtain

$$i\hbar(\partial/\partial t)(\rho\langle p\rangle + \gamma N\gamma\pi_1) + \langle p\rangle[\rho, h(1)] + \gamma N\gamma[\pi_1, h(1)] + i\hbar\Omega^2(\rho\langle q\rangle + \gamma N\gamma\pi^1) + \gamma\langle p^2\rangle[\rho, \mu] + \gamma(\gamma N\gamma)[\pi_2^1, \mu] = 0, \qquad (4.4)$$

where we have used the definitions

$$\Pi_1(1, t) = \langle p\rangle\rho(1) + \gamma N\gamma\pi_1(1, t), \qquad \pi_i(1, t) \equiv \mathrm{tr}_q\, p^i\chi(1, q),$$
$$\Pi^1(1, t) = \langle q\rangle\rho(1) + \gamma N\gamma\pi^1(1, t), \qquad \pi^i(1, t) \equiv \mathrm{tr}_q\, q^i\chi(1, q).$$

The equation for the one-particle density matrix to order $(\gamma N\gamma)^2$ in the new variables now takes the form

$$i\hbar(\partial\rho/\partial t) + [\rho, h(1)] + \gamma\langle p\rangle[\rho, \mu] = \gamma(\gamma N\gamma)[\mu, \pi_1(1, t)]. \quad (4.5)$$

Since we will show that $\pi_1(1, t)$ is of order $N\gamma$, the right-hand side of Eq. (4.5) is proportional to $(\gamma N\gamma)^2$.

When we regroup the terms in Eq. (4.4), we obtain

$$\langle p\rangle\left(i\hbar\frac{\partial\rho}{\partial t} + [\rho, h(1)]\right) + i\hbar\rho\left(\frac{\partial\langle p\rangle}{\partial t} + \Omega^2\langle q\rangle\right) + \gamma N\gamma\left(i\hbar\frac{\partial\pi_1(1, t)}{\partial t} + [\pi_1(1), h(1)] + i\hbar\Omega^2\pi^1(1, t) + \frac{\langle p^2\rangle}{N\gamma}[\rho, \mu]\right) = 0, \quad (4.6)$$

where we have dropped the last term of Eq. (4.4) since it is of order $(\gamma N\gamma)^3$.

The second bracket of Eq. (4.6) vanishes because of Eq. (3.12). When Eq. (4.5) is substituted in Eq. (4.6) ,we obtain

$$i\hbar\frac{\partial\pi_1(1, t)}{\partial t} + [\pi_1, h(1)] + i\hbar\Omega^2\pi^1(1) + (N\gamma)^{-1}(\langle p^2\rangle - \langle p\rangle^2)[\rho, \mu] = 0\ . \quad (4.7)$$

We need to know $\langle p^2\rangle$ and $\pi^1(1)$ to solve this equation for $\pi_1(1)$. We derive the equation satisfied by $\langle p^2\rangle$, Eq. (A3), in the Appendix. The quantity $(\langle p^2\rangle - \langle p\rangle^2)$ is positive and proportional to $(N\gamma)^2$. Consequently, the right-hand side of Eq. (4.5) is proportional to $(\gamma N\gamma)^2$.

We derive the equation of motion $\pi^1(1)$ in the same manner as we used to obtain Eq. (4.7). We multiply Eq. (3.4) for $(s = 1)$ by $q$ and take the trace over the oscillator variable

$$\begin{aligned} i\hbar\frac{\partial\Pi^1(1, t)}{\partial t} &+ [\Pi^1(1), h(1)] \\ &- i\hbar\Pi_1(1) + \gamma\,\mathrm{tr}_q\, q[F_1, p\mu] \\ &= (N - 1)\,\mathrm{tr}_{2,q}\, q[p\mu_2, F_2] \\ &= (N - 1)\gamma i\hbar\langle\mu\rangle\rho(1) + \gamma N\gamma I(t), \end{aligned} \quad (4.8)$$

where

$$I(t) \equiv \gamma N\,\mathrm{tr}_{2,q}\, q[p\mu_2, \rho(1)\chi(2, q) + \rho(2)\chi(1, q)].$$

We used

$$\mathrm{tr}_q\, q[F_1, H_t] = -\mathrm{tr}_q\,[q, H_t]F_1 = -\tfrac{1}{2}\,\mathrm{tr}_q\,[q, p^2]F_1 = -i\hbar\,\mathrm{tr}_q\, pF_1 \equiv -i\hbar\Pi_1(1).$$

The expression $I(t)$ is complicated and we analyze it in the next section. When we introduce our expansion for $F_1$ and regroup terms, we obtain

$$i\hbar\rho\left(\frac{\partial\langle q\rangle}{\partial t} - \langle p\rangle - \gamma N\langle\mu\rangle\right) + \langle q\rangle\left(i\hbar\frac{\partial\rho}{\partial t} + [\rho, h(1)]\right) + \gamma N\gamma\left(i\hbar\frac{\partial\pi^1}{\partial t} + [\pi^1, h(1)] - i\hbar\pi_1 + (N\gamma)^{-1}\langle qp\rangle[\rho, \mu]\right) = \gamma N\gamma I(t). \quad (4.9)$$

The first bracket vanishes because of Eq. (3.13). When we substitute Eq. (4.5) in the second bracket, we obtain

$$i\hbar(\partial\pi^1/\partial t) + [\pi^1, h(1)] + i\hbar\pi_1 + (N\gamma)^{-1}(\langle qp\rangle - \langle q\rangle\langle p\rangle)[\rho, \mu] = I(t). \quad (4.10)$$

Eqs. (4.7) and (4.10) constitute two first-order equations for $\pi_1$ and $\pi^1$ which we now solve for $\pi_1$. We multiply Eq. (4.10) by $i\Omega$ and add it to Eq. (4.7),

$$i\hbar(\partial/\partial t)[\pi_1(1) + i\Omega\pi^1(1)] + [\pi_1(1) + i\Omega\pi^1(1), h(1)] + \hbar\Omega[\pi_1(1) + i\Omega\pi^1(1)] + [\rho, \mu]B(t) = i\Omega I(t), \quad (4.11)$$

where

$$B(t) \equiv B_R(t) + iB_I(t) = (N\gamma)^{-1}[(\langle p^2\rangle - \langle p\rangle^2) + i\Omega(\langle qp\rangle - \langle q\rangle\langle p\rangle)].$$

We define two new variables $\eta^\dagger$ and $\eta$ in the following manner:

$$\eta^\dagger = \pi_1(1) + i\Omega\pi^1(1), \qquad \eta = \pi_1(1) - i\Omega\pi^1(1).$$

When we substitute $\eta^\dagger$ and $\eta$ in Eq. (4.11) we obtain

$$i\hbar(\partial\eta^\dagger/\partial t) + [\eta^\dagger(1), h(1)] + \hbar\Omega\eta^\dagger + [\rho, \mu]B(t) = i\Omega I(t). \quad (4.12)$$

If we now multiply Eq. (4.10) by $-i\Omega$ and add it to Eq. (4.7) we obtain

$$i\hbar(\partial\eta/\partial t) + [\eta, h(1)] - \hbar\Omega\eta + [\rho, \mu]B^*(t) = -i\Omega I(t). \quad (4.13)$$

Eqs. (4.12) and (4.13) are inhomogeneous first-order equations which we can solve for any initial conditions. For definiteness we assume $\eta$ and $\eta^\dagger$ vanish in the infinite past. For these initial conditions we obtain

$$\hbar\eta^\dagger = \int_0^\infty e^{i\Omega\tau}e^{-i\omega\tau}[\mu, \rho(t - \tau)]e^{i\omega\tau}B(t - \tau)\,d\tau + i\Omega\int_0^\infty e^{i\Omega\tau}e^{-i\omega\tau}I(t - \tau)e^{i\omega\tau}\,d\tau, \quad (4.14)$$

$$\hbar\eta(t) = \int_0^\infty e^{-i\Omega\tau}e^{-i\omega\tau}[\mu, \rho(t-\tau)]e^{i\omega\tau}B^*(t-\tau)\,d\tau - i\Omega\int_0^\infty e^{-i\Omega\tau}e^{-i\omega\tau}I(t-\tau)e^{i\omega\tau}\,d\tau, \quad (4.15)$$

where $\hbar\omega(1) = h(1)$ is the one-atom Hamiltonian operator. When we add Eqs. (4.14) and (4.15) and divide by two we obtain

$$\hbar\pi_1(1, t) = \int_0^\infty \cos\Omega\tau e^{-i\omega\tau}[\rho(t-\tau), \mu]e^{i\omega\tau}B_R(t-\tau)\,d\tau - \int_0^\infty \sin\Omega\tau e^{-i\omega\tau}[\rho(t-\tau), \mu]e^{i\omega\tau}B_I(t-\tau)\,d\tau - \Omega\int_0^\infty \sin\Omega\tau e^{-i\omega\tau}I(t-\tau)e^{i\omega\tau}\,d\tau. \quad (4.16)$$

If we set $I = 0$, then Eqs. (4.16) and (4.5) constitute a complete theory correct to order $(\gamma N\gamma)^2$ for $\rho$ where the moments of the electromagnetic field satisfy Eqs. (3.12), (3.13), and Eqs. (A3), (A4), and (A5). We analyze these equations in Sec. VI. We show in the next section that $I(t)$ vanishes except for special initial conditions.

## V. SOLUTION FOR $I(t)$

The definition of $I(t)$ given in Eq. (4.8) may be written

$$\begin{aligned} I &= N\gamma\,\mathrm{tr}_{2,q}\,q[p\mu_2, \rho(1)\chi(2, q) + \rho(2)\chi(1, q)] \\ &= N\gamma\{\rho(1)\,\mathrm{tr}_q\,q[p, \mathrm{tr}_2(\mu_2\chi(2, q))] \\ &\qquad + \langle\mu\rangle\,\mathrm{tr}_q\,q[p, \chi(1, q)]\} \\ &= i\hbar N\gamma\{\rho(1)\bar{\mu}(t) + \langle\mu\rangle\pi_0(1, t)\}, \end{aligned} \quad (5.1)$$

where

$$\bar{\mu}(t) \equiv \mathrm{tr}_{1,q}\,\mu_1\chi(1, q), \qquad \pi_0(1, t) \equiv \mathrm{tr}_q\,\chi(1, q).$$

We now show the equations of motion satisfied by the new moments of $\chi$ are

$$i\hbar[\partial\pi_0(1, t)/\partial t] + [\pi_0(1, t), h(1)] = 0 \quad (5.2)$$

and

$$\partial^2\bar{\mu}/\partial t^2 + \omega_0^2\bar{\mu} = 0. \quad (5.3)$$

If $\pi_0(1)$ and $\bar{\mu}$ are zero initially, they will remain zero; thus, as we stated earlier, $\pi_0(1, t)$ and $\bar{\mu}(t)$ are important for only a few special sets of initial conditions.

We find the operator equation of motion satisfied by $\pi_0(1, t)$ by taking the trace of Eq. (3.4) for $F_1$ over $q$ after the substitution

$$F_1 = R(q)\rho(1) + \gamma N\gamma\chi(1, q),$$

and we obtain, after some regrouping of terms,

$$\begin{aligned} &i\hbar(\partial\rho/\partial t) + [\rho, h(1)] + \rho(1)\,\mathrm{tr}_q\,[R, H_f] \\ &+ \gamma\,\mathrm{tr}_q\,[\rho R, \mu_1 p] + \gamma N\gamma(i(\partial\pi_0/\partial t) + [\pi_0(1), h(1)] \\ &+ \mathrm{tr}_q\,[\chi, H_f] + \gamma\,\mathrm{tr}_q\,[\chi, \mu_1 p]) = N\gamma\,\mathrm{tr}_{2,q}\,[p\mu_2, F_2], \end{aligned}$$

where we have used $\mathrm{tr}_q\,R(q) = 1$. We may simplify the above equation to

$$\gamma N\gamma(i\hbar(\partial\pi_0/\partial t) + [\pi_0(1), h(1)]) = \gamma N\,\mathrm{tr}_{2,q}\,[p\mu_2, F_2(1, 2, q)], \quad (5.4)$$

where we have used Eq. (4.5) and the following relations:

$$\gamma\,\mathrm{tr}_q\,[\rho(1)R, \mu_1 p] \equiv \gamma\langle p\rangle[\rho(1), \mu_1],$$

$$\mathrm{tr}_q\,[\chi, \mu_1 p] = [\pi_1(1, t), \mu_1], \qquad \mathrm{tr}_q\,[R, H_f] = 0.$$

We now see Eq. (5.2) follows from Eq. (5.4). The trace vanishes identically because the operator $p\mu_2$ does not depend on atom-one operators. The general solution for Eq. (5.2) is

$$\pi_0(1, t) = e^{-i\omega t}\pi_0(1, 0)e^{i\omega t}. \quad (5.5)$$

In the process of finding the equation of motion for $\bar{\mu}$ we will find the equation of motion for $\langle\mu\rangle$. We multiply Eq. (3.4) for $F_1$ by $\mu$ and take the trace over atom one and the oscillator coordinate,

$$\begin{aligned} &i\hbar(\partial/\partial t)(\mathrm{tr}_{1,q}\,\mu F_1) + \mathrm{tr}_{1,q}\,\mu[F_1, h(1)] + \mathrm{tr}_{1,q}\,\mu[F_1, H_f] \\ &+ \gamma\,\mathrm{tr}_{1,q}\,\mu[F_1, p\mu_1] = N\gamma\,\mathrm{tr}_1\,\mu_1\,\mathrm{tr}_{2,q}\,[p\mu_2, F_2]. \end{aligned} \quad (5.6)$$

The third, fourth, and fifth terms of Eq. (5.6) vanish since

$$\mathrm{tr}_{1,q}\,\mu[F_1, H_f] = -\mathrm{tr}_{1,q}\,[\mu, H_f]F_1 = 0,$$

$$\mathrm{tr}_{1,q}\,\mu[F_1, p\mu] = -\mathrm{tr}_{1,q}\,[\mu, p\mu]F_1 = 0,$$

$$\mathrm{tr}_{1,2,q}\,\mu_1[p\mu_2, F_2] = -\mathrm{tr}_{1,2,q}\,[p\mu_2, \mu_1]F_2 = 0.$$

We evaluate the second term of Eq. (5.6) with the help of the anticommutation relations in Eq. (2.1),

$$[\mu, h(1)] = \hbar\omega_0[\sigma + \sigma^\dagger, \sigma^\dagger\sigma] = \hbar\omega_0(\sigma - \sigma^\dagger) \equiv -\hbar\omega_0\delta. \quad (5.7)$$

When we substitute our expansion for $F_1$ and Eq. (5.7) in Eq. (5.6), we obtain

$$i\hbar(\partial/\partial t)(\langle\mu\rangle + \gamma N\gamma\bar{\mu}) + \hbar\omega_0(\langle\delta\rangle + \gamma N\gamma\bar{\delta}) = 0, \quad (5.8)$$

where

$$\langle\delta\rangle \equiv \mathrm{tr}_1\,(\sigma^\dagger - \sigma)\rho(1),$$

$$\delta \equiv \mathrm{tr}_{1,q}\,(\sigma^\dagger - \sigma)\chi(1, q) \equiv \mathrm{tr}_1\,(\sigma^\dagger - \sigma)\pi_0(1).$$

In order to find the equations of motion satisfied

by $\langle\delta\rangle$ and $\bar{\delta}$ we multiply Eq. (3.4) for $F_1$ by the operator $\delta$ and take the trace over atom one and the oscillator coordinate,

$$i\hbar(\partial/\partial t)(\mathrm{tr}_{1,a}\ \delta F_1) + \mathrm{tr}_{1,a}\ \delta[F_1, h(1)] + \mathrm{tr}_{1,a}\ \delta[F_1, H_f] + \gamma\ \mathrm{tr}_{1,a}\ \delta[F_1, \mu p] = \mathrm{tr}_{1,2,a}\ \delta_1[p\mu_2, F_2]. \quad (5.9)$$

The third and fifth terms vanish because

$$\mathrm{tr}_{1,a}\ \delta[F_1, H_f] = -\mathrm{tr}_{1,a}\ [\delta, H_f]F_1 = 0,$$

$$\mathrm{tr}_{1,2,a}\ \delta_1[p\mu_2, F_2] = -\mathrm{tr}_{1,2,a}\ [p\mu_2, \delta_1]F_2 = 0.$$

We need the following commutator relations to evaluate the remaining terms:

$$\begin{aligned}[\delta, h(1)] &= \hbar\omega_0[\sigma^\dagger - \sigma, \sigma^\dagger\sigma] \\ &= \hbar\omega_0(\sigma^\dagger[\sigma^\dagger, \sigma] - [\sigma, \sigma^\dagger]\sigma) \\ &= \hbar\omega_0(-\sigma^\dagger\sigma^c - \sigma^c\sigma) = -\hbar\omega_0\mu, \end{aligned} \quad (5.10)$$

$$\begin{aligned}[\delta, \mu] &= [(\sigma^\dagger - \sigma), (\sigma^\dagger + \sigma)] = [\sigma^\dagger, \sigma] - [\sigma, \sigma^\dagger] \\ &= -2[\sigma, \sigma^\dagger] \equiv -2\sigma^c, \end{aligned} \quad (5.11)$$

where we used

$$\sigma^\dagger\sigma^c = \sigma^\dagger, \qquad \sigma^c\sigma = \sigma.$$

When Eqs. (5.10) and (5.11) are substituted in Eq. (5.9), we obtain

$$\begin{aligned}&i\hbar(\partial/\partial t)(\mathrm{tr}_{1,a}\ \delta F_1) \\ &\quad = -\hbar\omega_0\ \mathrm{tr}_{1,a}\ \mu F_1 - 2\gamma\ \mathrm{tr}_{1,a}\ \sigma^c p F_1. \end{aligned} \quad (5.12)$$

We expand $F_1$ to first order in $\gamma N\gamma$ and we obtain

$$\begin{aligned}&i\hbar(\partial/\partial t)(\langle\delta\rangle + \gamma N\gamma\bar{\delta}) = -\hbar\omega_0(\langle\mu\rangle + \gamma N\gamma\bar{\mu}) \\ &\quad - 2\gamma(\langle p\rangle\langle\sigma^c\rangle + \gamma N\gamma\ \mathrm{tr}_1\ \sigma^c\pi_1(1, t)). \end{aligned} \quad (5.13)$$

We find a single second-order equation for $\mu$ by eliminating $\delta$ from Eqs. (5.8) and (5.13),

$$\begin{aligned}&\partial^2\langle\mu\rangle/\partial t^2 + \omega_0^2\langle\mu\rangle + \gamma N\gamma(\partial^2\bar{\mu}/\partial t^2 + \omega_0^2\bar{\mu}_1) \\ &\quad = -2\gamma\hbar^{-1}\omega_0(\langle p\rangle\langle\sigma^c\rangle + \gamma N\gamma\ \mathrm{tr}_1\ \sigma^c\pi_1(1)). \end{aligned} \quad (5.14)$$

When we multiply Eq. (4.5) by $\mu$ and repeat the same steps that led to Eq. (5.13) we obtain

$$\begin{aligned}&\partial^2\langle\mu\rangle/\partial t^2 + \omega_0^2\langle\mu\rangle \\ &\quad = -2\gamma\hbar^{-1}\omega_0(\langle p\rangle\langle\sigma^c\rangle + \gamma N\gamma\ \mathrm{tr}_1\ \sigma^c\pi_1(1, t)). \end{aligned} \quad (5.15)$$

We finally obtain the equation of motion for $\bar{\mu}$ by substracting Eq. (5.15) from Eq. (5.14),

$$\partial^2\bar{\mu}/\partial t^2 + \omega_0^2\bar{\mu} = 0. \quad (5.3)$$

We have now determined $I(t)$ in terms of $\rho$ and $\langle\mu\rangle$. We calculate the contribution of $I(t)$ to $\pi_1(1, t)$ by substituting Eqs. (5.2) and (5.3) in the last term of Eq. (4.16),

$$\begin{aligned}&\Omega\int_0^\infty \sin\Omega\tau e^{-i\omega\tau}I(t-\tau)e^{i\omega\tau}\,d\tau \\ &= i\hbar\Omega N\gamma\int_0^\infty \sin\Omega\tau[\bar{\mu}(t-\tau)e^{-i\omega\tau}\rho(t-\tau)e^{i\omega\tau} \\ &\quad + \langle\mu(t-\tau)\rangle e^{-i\omega\tau}\pi_0(1, t-\tau)e^{i\omega\tau}]\,d\tau \\ &= i\hbar\Omega N\gamma\Big[\int_0^\infty \sin\Omega\tau\bar{\mu}(t-\tau)e^{-i\omega\tau}\rho(t-\tau)e^{i\omega\tau}\,d\tau \\ &\quad + \pi_0(1, t)\int_0^\infty \sin\Omega\tau\langle\mu(t-\tau)\rangle\,d\tau\Big]. \end{aligned} \quad (5.16)$$

Since $\bar{\mu}(t)$ and $\pi_0(1, t)$ satisfy homogeneous equations, the only time $I(t)$ is nonzero is when there are nonzero values of $\bar{\mu}$ and $\pi_0$ present initially. Practically, in most cases $I(t)$ is of little consequence and we will usually neglect its contribution to $\pi_1(1, t)$. We conclude this section with a discussion of perhaps the only important case where it is not permissible to ignore $I(t)$. An examination of the SCFA shows that if $\rho$ is initially diagonal it remains diagonal and there is no interaction. When Eq. (4.16) with $I(t) = 0$ is substituted in Eq. (4.5), we see that if $\rho$ is diagonal initially it remains diagonal and there is still no interaction. For instance, if we start the system with a population inversion but with the off-diagonal matrix elements zero, nothing happens. However, if initially there is a small particle–field correlation present, we can get the interaction started. We then have

$$\pi_1(1, t) = -\hbar^{-1}\Omega\int_0^t \sin\Omega\tau e^{-i\omega\tau}I(t-\tau)e^{i\omega\tau}\,d\tau,$$

$$I(t) = i\hbar N\gamma\bar{\mu}(t)\rho(t),$$

and thus

$$\begin{aligned}&\mathrm{tr}_1\ \sigma^c\pi_1(1, t) \\ &\quad = -iN\gamma\Omega\int_0^t \sin\Omega\tau\bar{\mu}(t-\tau)\langle\sigma^c(t-\tau)\rangle\,d\tau \\ &\quad \approx -iN\gamma\Omega\langle\sigma^c(t)\rangle\int_0^t \sin\Omega\tau\bar{\mu}(t-\tau)\,d\tau. \end{aligned} \quad (5.17)$$

The last line of Eq. (5.17) results from the fact that $\langle\sigma^c(t)\rangle$ is slowly varying compared with $\omega_0$. When Eq. (5.17) is substituted in Eq. (5.15) with $\langle p\rangle = 0$, we obtain

$$\begin{aligned}\partial^2\langle\mu\rangle/\partial t^2 + \omega_0^2\langle\mu\rangle &= +2i(\gamma N\gamma)^2\Omega\omega_0(\hbar)^{-1}\langle\sigma^c(t)\rangle \\ &\quad \times\int_0^t \sin\Omega\tau\bar{\mu}(t-\tau)\,d\tau. \end{aligned} \quad (5.18)$$

In Eq. (5.18) we showed how an initial correlation $\bar{\mu}(0)$ gives rise to off-diagonal matrix elements of $\rho$; i.e., $\langle\mu\rangle$. It is important to observe that the right-

hand side of Eq. (5.18) depends on only diagonal elements of $\rho$. Once $\langle\mu\rangle$ starts to grow, the SCFA becomes the valid description of the process.

## VI. REPRESENTATION BY MOMENTS

We derived the equations of motion satisfied by our model correct to order $(\gamma N\gamma)^2$, and we explicitly solved the equations for particle–field correlations which appear in second-order in terms of the one-particle density matrix $\rho$ and the first and second moments of the electromagnetic field. For convenience we collect the equations representing the complete theory with $I(t) = 0$,

$$i\hbar(\partial\rho/\partial t) + [\rho, h(1)] + \gamma\langle p\rangle[\rho, \mu] = \gamma(\gamma N\gamma)[\mu, \pi_1(1, t)], \quad (4.5)$$

$$\pi_1(1, t) = \hbar^{-1}\int_0^\infty \cos\Omega\tau e^{-i\omega\tau}[\rho(t-\tau), \mu]e^{i\omega\tau}B_R(t-\tau)\,d\tau - \hbar^{-1}\int_0^\infty \sin\Omega\tau e^{-i\omega\tau}[\rho(t-\tau), \mu]e^{i\omega\tau}B_I(t-\tau)\,d\tau,$$

$$B_R = (N\gamma)^{-1}(\langle p^2\rangle - \langle p\rangle^2) - \hbar\omega_0/2,$$

$$N\gamma B_I = \Omega(\tfrac{1}{2}\langle pq + qp\rangle - \langle p\rangle\langle q\rangle), \quad (4.16)$$

$$\partial\langle p\rangle/\partial t + \Omega^2\langle q\rangle = 0, \quad (3.12)$$

$$\partial\langle q\rangle/\partial t - \langle p\rangle = N\gamma\langle\mu\rangle, \quad (3.13)$$

$$\partial\langle p^2\rangle/\partial t + \Omega^2\langle qp + pq\rangle = 0, \quad (A3)$$

$$\partial\langle q^2\rangle/\partial t - \langle qp + pq\rangle = 2N\gamma\langle\mu\rangle\langle q\rangle, \quad (A4)$$

$$\partial\langle qp + pq\rangle/\partial t - 2(\langle p^2\rangle - \Omega^2\langle q^2\rangle) = 2N\gamma\langle\mu\rangle\langle p\rangle. \quad (A5)$$

Since the $\mathrm{tr}_1\,\rho = 1$, and $\rho = \rho^\dagger$, Eq. (4.5) represents three first-order differential equations for the three functions needed to specify $\rho$. The SCFA consists of five first-order differential equations, three for $\rho$, one for $\langle p\rangle$, and one for $\langle q\rangle$. The correct second-order theory represented by the above equations consists of eight first-order equations, three for $\rho$, two for the first moments $\langle p\rangle$ and $\langle q\rangle$, and three for the second moments $\langle p^2\rangle$, $\langle q^2\rangle$, and $\langle qp\rangle$. The nonlinearity has increased from the quadratic $\langle p\rangle\rho$ of the SCFA to cubic terms such as $\langle p\rangle^2\rho$ in Eq. (4.16).

We find it useful to represent the operator equation, Eq. (4.5), as three first-order differential equations for three $c$-number functions. The most natural choices are the traces of the operators, $\mu$, $\delta$, and $\sigma^c$,

$$\mathrm{tr}_1\,\mu\rho = \rho_{ab}(t) + \rho_{ba}(t), \quad \mathrm{tr}_1\,\delta\rho = \rho_{ab} - \rho_{ba}, \quad \mathrm{tr}_1\,\sigma^c\rho = -(\rho_{aa} - \rho_{bb}). \quad (6.1)$$

The equations of motion for $\langle\mu\rangle$ and $\langle\delta\rangle$ are given by Eqs. (5.8) and (5.13) with $\bar{\mu} = 0$,

$$i\hbar(\partial\langle\mu\rangle/\partial t) + \hbar\omega_0\langle\delta\rangle = 0, \quad (5.8')$$

$$i\hbar(\partial\langle\delta\rangle/\partial t) + \hbar\omega_0\langle\mu\rangle = 4(\hbar\omega_0)^{-1}(\langle p\rangle\langle h(t)\rangle + \gamma N\gamma\,\mathrm{tr}_1\,h(1)\pi_1(1, t)), \quad (5.13')$$

where $(\tfrac{1}{2}\hbar\omega_0)\sigma^c \equiv -h(1)$. The two first-order equations are equivalent to a single second-order equation for $\langle\mu\rangle$,

$$\partial^2\langle\mu\rangle/\partial t^2 + \omega_0^2\langle\mu\rangle = (4\gamma/\hbar^2)(\langle p(t)\rangle\langle h(t)\rangle + \gamma N\gamma\,\mathrm{tr}_1\,h(1)\pi_1(1, t)). \quad (6.2)$$

We obtain the first-order equation for $\langle h(t)\rangle$ when we multiply Eq. (4.5) by the operator $h(1)$ and take the trace over atom one,

$$\begin{aligned} i\hbar(\partial\langle h(t)\rangle/\partial t) + \mathrm{tr}_1\,h(1)[\rho, h(1)] &= -\gamma\langle p\rangle\,\mathrm{tr}_1\,h(1)[\rho, \mu_1] - \gamma(\gamma N\gamma)\,\mathrm{tr}_1\,h(1)[\mu_1, \pi_1(1, t)], \\ i\hbar(\partial\langle h(t)\rangle/\partial t) &= \gamma\langle p\rangle\,\mathrm{tr}_1\,([h(1), \mu_1]\rho) - \gamma(\gamma N\gamma)\,\mathrm{tr}_1\,([h(1), \mu]\pi_1(1, t)), \\ i\hbar(\partial\langle h(t)\rangle/\partial t) &= \gamma\hbar\omega_0\langle p\rangle\langle\delta\rangle - \gamma(\gamma N\gamma)\hbar\omega_0\,\mathrm{tr}_1\,\delta\pi_1(1, t), \end{aligned} \quad (6.3)$$

where $[h, \mu] = \hbar\omega_0\delta$.

The SCFA can be written as five first-order differential equations or two second and one first-order differential equation,

$$\begin{aligned} \partial^2\langle\mu\rangle/\partial t^2 + \omega_0^2\langle\mu\rangle &= (4\gamma/\hbar^2)\langle p\rangle\langle h(t)\rangle, \\ \partial^2\langle p\rangle/\partial t^2 + \Omega^2\langle p\rangle &= -N\gamma\Omega^2\langle\mu\rangle, \\ \partial\langle h\rangle/\partial t &= -\gamma\langle p\rangle(\partial\langle\mu\rangle/\partial t), \end{aligned} \quad (6.4)$$

where we used Eq. (5.8). Since the time variation of $\langle h(t)\rangle$ is slow compared to $\omega_0$ we can regard the SCFA intuitively as two coupled oscillators, one of which has a slowly varying coupling constant.

To complete our representation of the theory correct to order $(\gamma N\gamma)^2$ in terms of moments, we must evaluate $\mathrm{tr}_1\,h(1)\pi_1(1, t)$ in terms of moments. We multiply Eq. (4.16) by the operator $h(1)$ and take the trace over atom one,

$$\mathrm{tr}_1\,h(1)\pi_1(1, t) = \hbar^{-1}\int_0^\infty \cos\Omega\tau\{\mathrm{tr}_1\,h[\rho(t-\tau), \mu]\}B_R(t-\tau)\,d\tau - \hbar^{-1}\int_0^\infty \sin\Omega\tau\{\mathrm{tr}_1\,h[\rho(t-\tau), \mu]\}B_I(t-\tau)\,d\tau = -\hbar^{-1}\int_0^\infty \cos\Omega\tau(\hbar\omega_0)\,\mathrm{tr}_1\,(\delta\rho(t-\tau))B_R(t-\tau)\,d\tau$$

$$+ \hbar^{-1} \int_0^\infty \sin \Omega\tau(\hbar\omega_0)\, \mathrm{tr}_1\, (\delta\rho(t-\tau))B_I(t-\tau)\, d\tau$$

$$= i \int_0^\infty \cos \Omega\tau B_R(t-\tau) \frac{\partial}{\partial t} \langle\mu(t-\tau)\rangle\, d\tau$$

$$- i \int_0^\infty \sin \Omega\tau B_I(t-\tau) \frac{\partial}{\partial t} \langle\mu(t-\tau)\rangle\, d\tau. \qquad (6.5)$$

We finally obtain the following equation for $\langle\mu(t)\rangle$ correct to order $(\gamma N\gamma)^2$:

$$\frac{\partial^2\langle\mu\rangle}{\partial t^2} + \omega_0^2\langle\mu\rangle = \frac{4\gamma}{\hbar^2}\Bigg[\langle p\rangle\langle h\rangle + \gamma N\gamma i$$

$$\times \Bigg(\int_0^\infty \cos \Omega\tau B_R(t-\tau) \frac{\partial}{\partial t} \langle\mu(t-\tau)\rangle\, d\tau$$

$$- \int_0^\infty \sin \Omega\tau B_I(t-\tau) \frac{\partial}{\partial t} \langle\mu(t-\tau)\rangle\, d\tau\Bigg)\Bigg]. \qquad (6.6)$$

We evaluate $\hbar\omega_0\ \mathrm{tr}_1\delta\pi_1(1, t)$ in a similar fashion and we obtain

$$\frac{\partial\langle h\rangle}{\partial t} + \gamma\langle p\rangle \frac{\partial\langle\mu\rangle}{\partial t} = -8\hbar^{-1}\gamma(\gamma N\gamma)i$$

$$\times \Bigg[\int_0^\infty \cos \Omega\tau \cos\omega_0\tau\langle h(t-\tau)\rangle B_R(t-\tau)\, d\tau$$

$$- \int_0^\infty \sin \Omega\tau \cos\omega_0\tau\langle h(t-\tau)\rangle B_I(t-\tau)\, d\tau\Bigg]. \qquad (6.7)$$

The second-order differential equation, Eq. (6.6), and the first-order differential equation, Eq. (6.7), are completely equivalent to the three first-order equations represented by the operator Eq. (4.5), and are frequently easier to treat.

We obtain the equations of motion for macroscopic quantities from our microscopic variables with the following table:

| | |
|---|---|
| Electric field intensity | $E(t) \sim -V^{-\frac{1}{2}}\langle p(t)\rangle$ |
| Electric dipole moment | $M(t) \sim N\gamma V^{\frac{1}{2}}\langle\mu(t)\rangle$ |
| Energy of the atoms | $W(t) \sim N\langle h(t)\rangle$. |

The equations of motion of the SCFA in macroscopic variables are identical to those studied by Jaynes and Cummings.[2] If we express Eqs. (6.6) and (6.7) in macroscopic variables we see the only variables that appear in addition to $E(t)$, $M(t)$, and $W(t)$ are the second moments of the electromagnetic field, $\langle p^2\rangle$, $\langle q^2\rangle$, and $\langle pq\rangle$, which are also proportional to $N$.

## VII. DISCUSSION

We have shown that the problem of $N$ two-level atoms interacting through the electromagnetic field in a lossless cavity requires the diagonalization of matrices of tremendous dimensionality even in the lowest order of perturbation theory. If the initial particle–field correlations are of order $\gamma N\gamma$ or less, we find the solution of the problem to order $(\gamma N\gamma)^2$ is the SCFA plus higher terms. The SCFA consists of five first-order differential equations. The equation of motion for $\rho$ correct to order $(\gamma N\gamma)^2$ is given by eight first-order differential equations. The three additional equations are required because of the occurrence of the three second moments of the electromagnetic field. The maximum nonlinearity in the SCFA is quadratic, while the higher order theory is cubic.

Both the SCFA and the present theory are invariant under time reversal; i.e., no assumptions are made that wipe out any dynamical information. Our only assumption is that there is an initial lack of correlation. In a future publication we will show that some of the terms of order $(\gamma N\gamma)^2$ display an irreversible behavior that corresponds to a frequency-dependent damping correction to the SCFA. This term might be larger than the small phenomenological damping constants that are frequently introduced *ad hoc* to represent losses. This is why we avoided introducing any damping constants and used a lossless cavity. In addition, the SCFA is bounded and perfectly meaningful in a lossless cavity.

We are now carrying out the Doppler broadening case which corresponds to considering wavelengths that are of the dimension of the cavity and smaller. A generalized SCFA which includes the center-of-mass motion is still valid.

## APPENDIX

We now derive the equations of motion for $\langle p^2\rangle$, $\langle q^2\rangle$, and $\langle pq\rangle$ which appear in the second-order equation of motion for $\rho$.

To find the equation of motion for $\langle p^2\rangle$ we multiply Eq. (3.10) by $p^2$ and take the trace over the reservoir variable,

$$i\hbar(\partial/\partial t)\langle p^2\rangle + \tfrac{1}{2}\Omega^2\, \mathrm{tr}_a\, p^2[R, q^2] = 0. \qquad (A1)$$

The trace is

$$\mathrm{tr}\, p^2[R, q^2] = -\mathrm{tr}_a\, [p^2, q^2]R$$

$$= -2\, \mathrm{tr}_a\, \{pq[p, q] + [p, q]qp\}$$

$$= 2i\hbar\, \mathrm{tr}_a\, (pq + qp)R. \qquad (A2)$$

When Eq. (A2) is substituted in Eq. (A1) we obtain

$$(\partial/\partial t)\langle p^2\rangle + \Omega^2\langle qp + pq\rangle = 0. \qquad (A3)$$

We proceed in the same manner to find the equa-

tion of motion for $\langle q^2 \rangle$ and we obtain

$$i\hbar(\partial/\partial t)\langle q^2\rangle + \tfrac{1}{2}\,\mathrm{tr}_a\, q^2[R, p^2] = N\gamma\langle\mu\rangle\,\mathrm{tr}_a\, q^2[p, R],$$

$$i\hbar(\partial/\partial t)\langle q^2\rangle - i\hbar\langle qp + pq\rangle = N\gamma\langle\mu\rangle\,\mathrm{tr}_a\, q^2[p, R],$$

which becomes

$$(\partial/\partial t)\langle q^2\rangle - \langle qp + pq\rangle = 2N\gamma\langle\mu\rangle\langle q\rangle. \qquad \text{(A4)}$$

In Sec. IV we need the equation of motion for $\langle pq\rangle$ which can be obtained from $\langle qp + pq\rangle$ since

$$\langle qp + pq\rangle = i\hbar + 2\langle qp\rangle.$$

We find the equation of motion for $\langle pq + pq\rangle$ by multiplying Eq. (3.10) by $(pq + qp)$ and taking the trace. The result is

$$(\partial/\partial t)\langle qp + pq\rangle - 2(\langle p^2\rangle - \Omega^2\langle q^2\rangle) = 2N\gamma\langle\mu\rangle\langle p\rangle. \qquad \text{(A5)}$$

Equations (A3), (A4), and (A5) constitute three first-order, linear inhomogeneous equations for the three second moments of the electromagnetic field, $\langle p^2\rangle$, $\langle q^2\rangle$, and $\langle pq\rangle$. In general, if we go to $n$th order in $\gamma N\gamma$ we find a set of first-order inhomogeneous linear equations for the $n$th moments. The inhomogeneities depend on moments of lower order than the $n$th.

It is important to note that even if all the second moments are zero initially they will grow to nonzero values because the inhomogeneous terms depend on $\langle p\rangle$, $\langle q\rangle$, and $\langle\mu\rangle$.

The eigenfrequencies of the homogeneous equations for $\langle q^2\rangle$, $\langle p^2\rangle$, and $\langle pq\rangle$ are 0, $\pm 2i\Omega$. Since these frequencies are prominent in the inhomogeneous terms, the second moments are strongly coupled to the SCFA quantities $\langle p\rangle$, $\langle q\rangle$, and $\langle\mu\rangle$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# Comments on Nonlinear Wave Equations as Models for Elementary Particles

G. H. Derrick
*Applied Mathematics Department, The University of New South Wales, Kensington, N.S.W., Australia*
(Received 3 April 1964)

It is shown that for a wide class of nonlinear wave equations there exist no stable time-independent solutions of finite energy. The possibility is considered whether elementary particles might be oscillating solutions of some nonlinear wave equation, in which the wavefunction is periodic in the time though the energy remains localized.

## 1. INTRODUCTION

In an attempt to find a model for *extended* elementary particles, as opposed to singular *point* particles, Enz[1] has recently considered the nonlinear equation

$$\nabla^2\theta - (1/c^2)(\partial^2\theta/\partial t^2) = \tfrac{1}{2}\sin 2\theta, \qquad (1)$$

which is derived from the variation principle

$$\delta\int\left[\frac{1}{c^2}\left(\frac{\partial\theta}{\partial t}\right)^2 - (\nabla\theta)^2 - \sin^2\theta\right]d^3\mathbf{r}\,dt = 0. \qquad (2)$$

$\theta(\mathbf{r}, t)$ is a $c$-number wavefunction which is required to be free of singularities for all $\mathbf{r}$ and $t$. In the one-dimensional case ($\nabla^2$ replaced by $\partial^2/\partial x^2$) Enz showed that (1) has time-independent solutions where the energy is localized about a point on the $x$ axis; if we further require that the solution be stable with respect to small deformations then only certain discrete energy values are permitted. In addition these one-dimensional solutions possess certain symmetry and topological properties which Enz suggests might correspond in the three-dimensional case to such discrete quantum numbers as charge or parity.

These suggestive results of Enz for the one-dimensional case then lead us to consider the following problem: Can (1) or some similar nonlinear equation have stable, time-independent, localized solutions in three dimensions? If such solutions exist then it would be an attractive hypothesis that the allowed energies correspond to the rest energies of elementary particles.

The answer given to the above question by this

[1] U. Enz, Phys. Rev. **131**, 1392 (1963). We have taken Enz's constants $K$ and $A$ both equal to 1, which amounts to a suitable choice of units of length and energy.

tion of motion for $\langle q^2 \rangle$ and we obtain

$$i\hbar(\partial/\partial t)\langle q^2\rangle + \tfrac{1}{2}\,\mathrm{tr}_a\, q^2[R, p^2] = N\gamma\langle\mu\rangle\,\mathrm{tr}_a\, q^2[p, R],$$

$$i\hbar(\partial/\partial t)\langle q^2\rangle - i\hbar\langle qp + pq\rangle = N\gamma\langle\mu\rangle\,\mathrm{tr}_a\, q^2[p, R],$$

which becomes

$$(\partial/\partial t)\langle q^2\rangle - \langle qp + pq\rangle = 2N\gamma\langle\mu\rangle\langle q\rangle. \qquad \text{(A4)}$$

In Sec. IV we need the equation of motion for $\langle pq\rangle$ which can be obtained from $\langle qp + pq\rangle$ since

$$\langle qp + pq\rangle = i\hbar + 2\langle qp\rangle.$$

We find the equation of motion for $\langle pq + pq\rangle$ by multiplying Eq. (3.10) by $(pq + qp)$ and taking the trace. The result is

$$(\partial/\partial t)\langle qp + pq\rangle - 2(\langle p^2\rangle - \Omega^2\langle q^2\rangle) = 2N\gamma\langle\mu\rangle\langle p\rangle. \qquad \text{(A5)}$$

Equations (A3), (A4), and (A5) constitute three first-order, linear inhomogeneous equations for the three second moments of the electromagnetic field, $\langle p^2\rangle$, $\langle q^2\rangle$, and $\langle pq\rangle$. In general, if we go to $n$th order in $\gamma N\gamma$ we find a set of first-order inhomogeneous linear equations for the $n$th moments. The inhomogeneities depend on moments of lower order than the $n$th.

It is important to note that even if all the second moments are zero initially they will grow to nonzero values because the inhomogeneous terms depend on $\langle p\rangle$, $\langle q\rangle$, and $\langle\mu\rangle$.

The eigenfrequencies of the homogeneous equations for $\langle q^2\rangle$, $\langle p^2\rangle$, and $\langle pq\rangle$ are 0, $\pm 2i\Omega$. Since these frequencies are prominent in the inhomogeneous terms, the second moments are strongly coupled to the SCFA quantities $\langle p\rangle$, $\langle q\rangle$, and $\langle\mu\rangle$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# Comments on Nonlinear Wave Equations as Models for Elementary Particles

G. H. Derrick
*Applied Mathematics Department, The University of New South Wales, Kensington, N.S.W., Australia*
(Received 3 April 1964)

It is shown that for a wide class of nonlinear wave equations there exist no stable time-independent solutions of finite energy. The possibility is considered whether elementary particles might be oscillating solutions of some nonlinear wave equation, in which the wavefunction is periodic in the time though the energy remains localized.

## 1. INTRODUCTION

In an attempt to find a model for *extended* elementary particles, as opposed to singular *point* particles, Enz[1] has recently considered the nonlinear equation

$$\nabla^2\theta - (1/c^2)(\partial^2\theta/\partial t^2) = \tfrac{1}{2}\sin 2\theta, \qquad (1)$$

which is derived from the variation principle

$$\delta\int\left[\frac{1}{c^2}\left(\frac{\partial\theta}{\partial t}\right)^2 - (\nabla\theta)^2 - \sin^2\theta\right]d^3\mathbf{r}\,dt = 0. \qquad (2)$$

$\theta(\mathbf{r}, t)$ is a $c$-number wavefunction which is required to be free of singularities for all $\mathbf{r}$ and $t$. In the one-dimensional case ($\nabla^2$ replaced by $\partial^2/\partial x^2$) Enz showed that (1) has time-independent solutions where the energy is localized about a point on the $x$ axis; if we further require that the solution be stable with respect to small deformations then only certain discrete energy values are permitted. In addition these one-dimensional solutions possess certain symmetry and topological properties which Enz suggests might correspond in the three-dimensional case to such discrete quantum numbers as charge or parity.

These suggestive results of Enz for the one-dimensional case then lead us to consider the following problem: Can (1) or some similar nonlinear equation have stable, time-independent, localized solutions in three dimensions? If such solutions exist then it would be an attractive hypothesis that the allowed energies correspond to the rest energies of elementary particles.

The answer given to the above question by this

[1] U. Enz, Phys. Rev. **131**, 1392 (1963). We have taken Enz's constants $K$ and $A$ both equal to 1, which amounts to a suitable choice of units of length and energy.

paper is *no*. The equation

$$\nabla^2\theta - (1/c^2)(\partial^2\theta/\partial t^2) = \tfrac{1}{2}f'(\theta), \tag{3}$$

derived from the variation principle

$$\delta \int \left[\frac{1}{c^2}\left(\frac{\partial\theta}{\partial t}\right)^2 - (\nabla\theta)^2 - f(\theta)\right] d^3\mathbf{r}\, dt = 0, \tag{4}$$

will be proved to have no stable, time-independent, localized solutions for any $f(\theta)$. In particular, Enz's equation (1) with $f(\theta) = \sin^2\theta$ has no such solutions. By "localized" solution we shall mean one where $\int (\nabla\theta)^2\, d^3\mathbf{r}$ and $\int f(\theta)\, d^3\mathbf{r}$ converge when the integrals are taken over all space.

## 2. PROOF

If $\theta$ is a function of $\mathbf{r}$ only, we can replace (4) by $\delta E = 0$ with the energy $E$ given by

$$E = \int [(\nabla\theta)^2 + f(\theta)]\, d^3\mathbf{r}.$$

A necessary condition for the solution to be stable is that the second-order variation $\delta^2 E \geq 0$. Suppose $\theta(\mathbf{r})$ is a localized solution of $\delta E = 0$. Define $\theta_\lambda(\mathbf{r}) = \theta(\lambda\mathbf{r})$ where $\lambda$ is an arbitrary constant, and write $I_1 = \int (\nabla\theta)^2\, d^3\mathbf{r}$, $I_2 = \int f(\theta)\, d^3\mathbf{r}$.
Then

$$E_\lambda = \int [(\nabla\theta_\lambda)^2 + f(\theta_\lambda)]\, d^3\mathbf{r}$$

$$= I_1/\lambda + I_2/\lambda^3$$

on changing the variable of integration from $\mathbf{r}$ to $\lambda\mathbf{r}$; whence

$$(dE_\lambda/d\lambda)_{\lambda=1} = -I_1 - 3I_2,$$

$$(d^2E_\lambda/d\lambda^2)_{\lambda=1} = 2I_1 + 12I_2.$$

Since $\theta_\lambda$ is a solution of $\delta E = 0$ for $\lambda = 1$, we must have

$$(dE_\lambda/d\lambda)_{\lambda=1} = 0, \qquad I_2 = -\tfrac{1}{3}I_1,$$

$$(d^2E_\lambda/d\lambda^2)_{\lambda=1} = -2I_1 < 0.$$

That is, $\delta^2 E < 0$ for a variation corresponding to a uniform stretching of the "particle." Hence the solution $\theta(\mathbf{r})$ is unstable, proving the theorem.

In the above proof no restriction was placed on the sign of $f(\theta)$. In Enz's equation (1) we have $f(\theta) = \sin^2\theta \geq 0$, which means that the energy density has the desirable feature of being everywhere positive. However it is interesting to note that if $f(\theta) \geq 0$ then $\delta E = 0$ has no nontrivial localized solutions *at all*, either stable or unstable. For in this case both $I_1$ and $I_2$ are necessarily nonnegative so that $I_1 + 3I_2 = 0$ has only the trivial solution $I_1 = I_2 = 0$, giving $\theta = 0$. (Our result here is not applicable to the one-dimensional case where Enz does obtain stable solutions. In one dimension we obtain $E_\lambda = \lambda I_1 + I_2/\lambda$ yielding $I_1 = I_2$ on differentiation, which gives no contradiction.)

We can easily extend the above proof to certain cases where we have a complex, multicomponent wavefunction $\psi^A$ rather than the real scalar function $\theta$; the superscript $A$ denotes some tensor or spinor index. For example we can carry through the above proof for wave equations derived from the variation principle

$$\delta \int [\sum_{\substack{AB \\ \iota\kappa}} c_{AB} g^{\iota\kappa} (\partial\psi^{*A}/\partial x^\iota)(\partial\psi^B/\partial x^\kappa) - f(\psi^{*A}, \psi^B)]\, d^4x = 0,$$

where $c_{AB}$ is an arbitrary positive definite Hermitian matrix, and $g^{\iota\kappa}$ the usual metric tensor ($\iota, \kappa = 0, 1, 2, 3$). If $c_{AB}$ is not definite, or if the coefficients of $(\partial\psi^{*A}/\partial x^\iota)(\partial\psi^B/\partial x^\kappa)$ are not of the simple product form $c_{AB}g^{\iota\kappa}$, then the condition for stability is no longer $\delta^2 E \geq 0$ and the proof fails.

## 3. DISCUSSION

We are thus faced with the disconcerting fact that no equation of type (4) has any time-independent solutions which could reasonably be interpreted as elementary particles. Some possible ways out of this difficulty are:

(a) We could take a Lagragian in which the derivatives occur in higher powers than the second. For example, with the form $[(\nabla\theta)^2 - (1/c^2)(\partial\theta/\partial t)^2]^n$ the nonexistence proof of Sec. 2 fails for $n > \frac{3}{2}$. Such a Lagrangian, however, leads to a very complicated differential equation.

(b) We could consider first-order spinor equations, such as

$$\delta \int [i\psi^\dagger(\partial\psi/\partial t + c\boldsymbol{\alpha}\cdot\nabla\psi) - f(\psi^\dagger, \psi)]\, d^3\mathbf{r}\, dt = 0, \tag{5}$$

where $\psi$ is a Dirac 4-component spinor and $\psi^\dagger$ its Hermitian conjugate, $\boldsymbol{\alpha}$ is the usual Dirac matrix, and $f(\psi^\dagger, \psi)$ is an arbitrary Lorentz-invariant function. With a first-order equation of this type, the condition for stability is no longer $\delta^2 E \geq 0$, but is now very complicated, and the author has been unable to prove or disprove the existence of stable time-independent solutions of (5) for general functions $f(\psi^\dagger, \psi)$.

(c) Quantization of the field equations by replacing the wavefunction by an operator satisfying some postulated commutation relations. Quantized equations of type (5) have been investigated extensively by Heisenberg *et al.*,[2] who find particle-like solutions.

(d) Elementary particles might correspond to stable, localized solutions which are *periodic* in time, rather than time-independent.

We shall confine ourselves here to a consideration of Possibility (d), that elementary particles are oscillating localized concentrations of energy. We know experimentally[3] that a particle of momentum $\mathbf{p}$ has an associated de Broglie[4] wavevector $\mathbf{k} = \mathbf{p}/\hbar$; relativistic invariance then suggests that a particle of mass $m$ at rest should have a de Broglie frequency $\omega = mc^2/\hbar$. If elementary particles correspond to stable periodic solutions of some nonlinear wave equation, then we could possibly identify the frequency of this oscillation with the de Broglie frequency.

A particularly simple form of periodic solution is one where the structure rotates at a constant angular velocity $\omega$ about a fixed direction, say the $Z$ axis; i.e., the wavefunction is a function of $x'$, $y'$, $z'$, where

$$x' = x \cos \omega t + y \sin \omega t,$$

$$y' = -x \sin \omega t + y \cos \omega t,$$

$$z' = z.$$

Then the variation principle (4) is equivalent to

$$\delta \int \left[ (\nabla' \theta)^2 - \frac{\omega^2}{c^2} |L\theta|^2 + f(\theta) \right] d^3\mathbf{r}' = 0, \qquad (6)$$

where

$$L = -i[x'(\partial/\partial y') - y'(\partial/\partial x')].$$

However the condition for stability of solutions is now very complicated, and the author has been unable to demonstrate either the existence or nonexistence of stable solutions of Eq. (6).

ACKNOWLEDGMENTS

The author is grateful to Professor J. M. Blatt and to Dr. N. Takimoto for their valuable comments.

[2] H. P. Duerr, W. Heisenberg, H. Mitter, S. Schlieder, and K. Yamazaki, Z. Naturforsch. **14**, 441 (1959); W. Heisenberg, *Proceedings of the 1960 Annual International Conference on High-Energy Physics at Rochester* (Interscience Publishers, Inc., New York, 1960), p. 851.

[3] C. Davisson and L. H. Germer, Phys. Rev. **30**, 705 (1927).

[4] L. de Broglie, Phil. Mag. **47**, 446 (1926); Ann. Phys. (Paris) **3**, 22 (1925).

# On Some Topological Properties of Feynman Graphs and Their Application to Formulas Related to the Feynman Amplitudes*

YUTZE CHOW†
*Department of Physics, Brandeis University, Waltham, Massachusetts*
(Received 21 April 1964)

The purpose of this note is to demonstrate the usefulness of the topological concept of the *tree sets* introduced into any Feynman graph. We demonstrate here a relationship between certain functions appearing in two different forms of parametrized Feynman amplitudes by using the properties of the *tree sets* and some purely determinantal manipulations. As another exposition, we also present an alternative proof of a theorem due to Nakanishi by using only the concept of *tree sets* with almost no algebraic manipulation involved; the proof in this case is seen to be particularly simple and lucid.

## I. INTRODUCTION

BY considering an arbitrary Feynman graph consisting of some external lines, internal lines, and the vertices, one can write down the corresponding Feynman amplitude in the form of a multiple integral with respect to the basic momenta. That the integral is, in fact, independent of the particular choice of the basic momenta is well known. These sets of possible basic momenta correspond to topological choices of different sets of independent loops (or closed paths) each conducting a circular flow of momentum. The integral, for mathematical convenience, is usually parametrized by the well-known Feynman formula of integration; the parametrized form is very useful, particularly for the study of analyticity properties of Feynman amplitudes. The functions that appear in the integral, as a consequence of parametrization, have some interesting topological properties. This is, of course, what one should expect since a Feynman graph determines uniquely a Feynman integral, and such a one–one correspondence would naturally appear in the algebraic forms even after some integral transformations. If the integral transformation is an appropriate one—Feynman parametrization certainly is—then the topological properties may appear in a very transparent manner in the functions involved in the expression. Thus, in handling a Feynman amplitude, and especially in the parametrized form, one should be able to take advantage of the topological properties imbedded in the graph. This does not mean that the topological concepts are, in general, more concise or transparent than the ordinary algebraic manipulations or transformations. However, a proper, combined use of these techniques is sometimes very profitable.

Among the topological properties of Feynman graphs, the concept of *tree sets*, which are to be defined is a very useful one. The purpose of this paper is to show how the topological properties of the *tree sets* can be used to establish relationships between functions of the differently parametrized Feynman amplitudes; we note that although, in the past, the *tree sets* have sometime been used to write down certain functions in the parametrized Feynman amplitude[1] the topological properties of *tree sets* have never been explicitly used as a tool to establish or to show any relationships between different functions. In this note, the terminologies used follow closely those of Nakanishi,[2] though our notations are somewhat different from his.

## II. SOME DEFINITIONS FOR FEYNMAN GRAPHS

A Feynman graph is by topological definition an *oriented linear graph*. For any given oriented linear graph, $\Gamma$, an *incidence matrix*[3,4] $\vec{\epsilon}$, was introduced by Poincare as follows:

$\epsilon_{jn} = 0$, if the internal line $j$ does not initiate or terminate at the vertex $n$,

$\epsilon_{jn} = 1$, if the internal line $j$ initiates from the vertex $n$,

and

$\epsilon_{jn} = -1$, if the internal line $j$ terminates at the vertex $n$.

A one–one correspondence therefore exists be-

* Research supported by ONR under the grant No. R167703.

† Present address: Department of Physics, University of Wisconsin at Milwaukee, Milwaukee, Wisconsin.

[1] K. Symansik, Progr. Theoret. Phys. (Kyoto) **20**, 690 (1958); we note here that "tree" was called "skeleton" by Symanzik.

[2] Y. Nambu, Nuovo Cimento **6**, 1064 (1957).

[3] A. Logunov, I. T. Todorov, and N. A. Chernikov, Zh. Eksperim. i Teor. Fiz. **42**, 1285 (1962) [English transl.: Soviet Phys.—JETP **15**, 891 (1962)].

[4] I. T. Todorov, Doctoral Dissertation, Joint Institute of Nuclear Research Report P-1205, Dubna, 1963.

tween the incidence matrix $\overleftrightarrow{\epsilon}$ and the given graph $\Gamma$. In dealing with a Feynman graph, a set of *basic momenta* $\{k_l\}$ is introduced[5]; each basic momentum flows along a *closed path* (or *loop*) due to conservation. The set $\{k_l\}$ will be referred to hereafter as the *base set*. Now, if we denote the $j$th internal momentum (i.e., the momentum of the $j$th internal line) by $q_j$ and the set of external momenta by $\{p_i\}$ then we can write $q_j$ as a linear combination of $\{k_l\}$ and $\{p_i\}$:

$$q_j = \sum_{l=1}^{L} \epsilon_{jl} k_l + \sum_{i=1}^{N} \lambda_{ji} p_i = K_j + P_j,$$

or simply, in a compact notation,

$$\vec{q} = \overleftrightarrow{\epsilon} \cdot \vec{k} + \overleftrightarrow{\lambda} \cdot \vec{p},$$

where

$\overleftrightarrow{\epsilon}_{jl} = 0$, if the line $j$ does not belong to the loop $l$ along which a $k_l$ flows,

$\overleftrightarrow{\epsilon}_{jl} = 1$, if the line $j$ belongs to and is parallel to the loop $l$,

and

$\overleftrightarrow{\epsilon}_{jl} = -1$, if line $j$ belongs to and is antiparallel to the loop $l$,

while $\lambda_{ji}$ can take any arbitrary value provided momentum conservation at every vertex is satisfied (thus $\overleftrightarrow{\lambda}$ is not unique). It is clear that, unlike $\overleftrightarrow{\epsilon}$, the matrix $\overleftrightarrow{\epsilon}$ introduced here is not uniquely defined for a given $\Gamma$, since the choice of the base set $\{k_l\}$ is not unique while the incidence matrix $\overleftrightarrow{\epsilon}$ is base-independent. However, once $\{k_l\}$ is chosen, then $\overleftrightarrow{\epsilon}$ is uniquely fixed. The different choices of a base set $\{k_l\}$ correspond to the possible ways of putting (two sets differed only by a permutation are considered as equivalent):

$$\{k_1, \cdots, k_L\} = \{q_{\nu_1}, \cdots, q_{\nu_L}\}, \tag{1}$$

where $\{\nu_1, \cdots, \nu_L\}$ is a subset of $\{1, \cdots, J\}$ when topologically allowed and $L$ is the total number of independent loops in a given graph $\Gamma$. Conservation of momenta imposes the following condition:

$$L = J - N + 1,$$

where $J$ is the total number of internal lines and $N$ the total number of vertices involved in the graph.

The *tree set*, $T$, is now defined as the *complement* of the corresponding *base set*. That is,

$$T_\nu = {}_c\{k_l\}_\nu,$$

where the presubscript $c$ is the usual notation for the *complement* of a set. $\nu$ denotes the particular choice of a base set.

[5] N. Nakanishi, Suppl. Progr. Theoret. Phys. (Kyoto), No. 18, 1 (1961).

## III. THE PARAMETRIZED FEYNMAN AMPLITUDES

There are different parametrized forms of Feynman amplitudes due to different ways of parametrization. We shall first consider the form used by Logunov *et al.*[4] and Todorov[5] (originally derived by Nambu and Symanzik[1,2] )but written here in a different notation:

$$F = C \int_0^1 \cdots \int_0^1 \prod_{j=1}^{J} \frac{d\alpha_j}{\alpha_j^2} \cdot \frac{\delta(1 - \tilde{\alpha})}{h^2(\alpha)[Q(\alpha, p) + i0]^{2N-J-2}}, \tag{2}$$

where $C$ is a constant and $\{\alpha_j\}$ is the set of Feynman parameters, and

$$\tilde{\alpha} \equiv \sum_{j=1}^{J} \alpha_j, \qquad \forall j : \alpha_j \geq 0.$$

The function $h(\alpha)$ is defined by

$$h(\alpha) = \det \overleftrightarrow{h},$$

with

$$\overleftrightarrow{h} = \begin{bmatrix} h_{11} & \cdots & h_{1,N-1} \\ & h_{22} & \\ \vdots & & \vdots \\ h_{N-1,1} & \cdots & h_{N-1,N-1} \end{bmatrix}$$

and

$$h_{nn'} = \sum_{j=1}^{J} \alpha_j^{-1} \epsilon_{jn} \epsilon_{jn'}. \tag{3}$$

The function $Q$ is defined by

$$Q = -\sum_{j=1}^{J} \alpha_j m_j^2 - \frac{1}{h} \begin{vmatrix} 0 & p \\ p^T & \overleftrightarrow{h} \end{vmatrix},$$

where

$$p = \{p_1, \cdots, p_{N-1}\}.$$

That the definition (3) used by Logunov *et al.*[4] is precisely the one given by Symanzik[1] can be seen easily as follows:

Let $\mathcal{S}_n$ be the set of all the internal lines attached to the vertex $n$, thus for $n \neq n'$ we have

$$j \notin \mathcal{S}_n \cap \mathcal{S}_{n'} \quad \text{implies} \quad \epsilon_{jn} \cdot \epsilon_{jn'} = 0 \tag{4}$$

and

$$j \in \mathcal{S}_n \cap \mathcal{S}_{n'} \quad \text{implies} \quad \epsilon_{jn} \cdot \epsilon_{jn'} = -1, \tag{5}$$

since if $j$ is an *incoming* line with respect to $n$, then $j$ must be an *outgoing* line with respect to $n'$, and vice versa. Therefore, (3) becomes simply

$$h_{nn'} = -\sum_{j \in \mathcal{S}_n \cap \mathcal{S}_{n'}} \alpha_j^{-1}, \qquad n \neq n' \tag{6}$$

and

$$h_{nn} = \sum_{j \in \mathcal{S}_n} \alpha_j^{-1}, \tag{7}$$

which is the form used by Symanzik.

However, a more popular form of the parametrized Feynman integral is the one derived first by Chisholm.[6] In the form used by Nakanishi, we have

$$F = C \int_0^1 \cdots \int_0^1 \prod_{j=1}^{J} d\alpha_j \frac{\delta(1-\tilde{\alpha})}{U^2(V - i0)^{J-2L}}, \tag{8}$$

with

$$U = \sum_{\forall\{\nu\}} \prod_{l=1}^{L} \alpha_{\nu_l}, \tag{9}$$

where $\{\nu\} = \{\nu_1, \cdots, \nu_L\}$ has the meaning of choosing the set $\{q_{\nu_1}, q_{\nu_2}, \cdots, q_{\nu_L}\}$ as the basic momenta (or simply the *base*). The summation in (9) runs over all the possible choices of the basic momenta for a given Feynman graph. The $V$ function may be expressed in several different ways,[3,7] however, since it is not of our present concern we give only the one due to Nakanishi:

$$V = \sum_{j=1}^{J} \alpha_j m_j^2 - \sum_{\forall \mathfrak{C} \in \Gamma} \left[ U_{c\mathfrak{C}} \left( \sum_{j=1}^{J} \alpha_j \varepsilon_{j\mathfrak{C}} P_j \right)^2 \right],$$

where $\mathfrak{C}$ denotes an arbitrary, closed loop not necessary belonging to a single basic momentum. The $U_{c\mathfrak{C}}$ denotes the $U$-function, as defined by (9), for the graph with $\mathfrak{C}$ reduced (i.e., $\alpha_j = 0$ for $\forall j \in \mathfrak{C}$).

## IV. THE DERIVATION

Originally, (2) was derived by Symanzik by means of Fourier transformation, while (8) was obtained by integrating out the basic momentum variables of the parametrized Feynman integral. They have been shown to be equivalent, e.g., by Nakanishi.[7] Here we attempt to show that such an equivalence can be also demonstrated by using the topological concept of *tree sets*. In the following proof only some determinantal manipulations are used together with the properties of tree sets:

*Proposition*:

$$U = h \prod_{j=1}^{J} \alpha_j \tag{10}$$

*Proof:* Let us label the set of internal lines of the graph $\Gamma$ by

$$\{\nu_i\} = \{\nu_1, \cdots, \nu_L, \nu_{L+1}, \cdots, \nu_J\}$$

such that the subset $\{\nu_l\} = \{\nu_1, \cdots, \nu_L\}$ is a *base* and therefore the subset $\{\nu_i\} = \{\nu_{L+1}, \cdots, \nu_J\}$ is a tree (with respect to this *base*).

From (9), we have immediately,

$$U\left(\prod_{j=1}^{J} \alpha_j\right)^{-1} = \sum_{\forall\{\nu_l\}\in\Gamma} \frac{\prod_{l=1}^{L} \alpha_{\nu_l}}{\prod_{j=1}^{J} \alpha_{\nu_j}}$$

$$= \sum_{\forall\{\nu_i\}\in\Gamma} \prod_{i=L+1}^{J} \alpha_{\nu_i}^{-1}.$$

But, $J - L = N - 1$, thus one obtains:

$$U\left(\prod_{j=1}^{J} \alpha_j\right)^{-1} = \sum_{\text{all trees}} \prod_{n=1}^{N-1} \alpha_{\mu_n}^{-1}. \tag{11}$$

On the other hand, determinant theory gives:

$$h = \det \tilde{h} = \sum_{\forall\{\gamma\}} \varepsilon_{\{\gamma\}} \prod_{n=1}^{N-1} h_{\gamma_n n}, \tag{12}$$

where $\{\gamma\} = \{\gamma_1, \cdots, \gamma_{N-1}\} = \mathsf{P}\{1, \cdots, N-1\}$, $\mathsf{P}$ = Permutation, $\varepsilon_{\{\gamma\}} = (-)^{P(\gamma)}$, where $P$ is the number of transpositions involved in $\mathsf{P}$.

Substituting (2) into (12) gives:

$$h = \sum_{\forall\{\gamma\}} \varepsilon_{\{\gamma\}} \prod_{n=1}^{N-1} \sum_{j=1}^{J} \alpha_j^{-1} \epsilon_{j\gamma_n} \epsilon_{jn}$$

$$= \sum_{\forall\{\gamma\}} \varepsilon_{\{\gamma\}} \sum_{\forall\{\mu\}} \prod_{n=1}^{N-1} \alpha_{\mu_n}^{-1} \epsilon_{\mu_n \gamma_n} \epsilon_{\mu_n n}, \tag{13}$$

where $\{\mu\} = \{\mu_1, \cdots, \mu_{N-1}\}$ with $\mu_n = 1, \cdots, J$. We note that, in principle, any kind of repetition of elements is allowed for $\{\mu\}$, e.g., one might have

$$\{\mu\} = \{4, 10, 6, 3, 6, 3, 3\}$$

for $J = 11$ and $N = 8$, say. However, it will be shown that there is actually no repetition of elements in $\{\mu\}$.

Let us rewrite (13) into the form,

$$h = \sum_{\forall\{\mu\}} \left[ \left( \prod_{n=1}^{N-1} \alpha_{\mu_n} \right) \cdot \prod_{n=1}^{N-1} \epsilon_{\mu_n n} \sum_{\forall\{\gamma\}} \left( \varepsilon_{\{\gamma\}} \prod_{n=1}^{N-1} \epsilon_{\mu_n \gamma_n} \right) \right]$$

$$= \sum_{\forall\{\mu\}} \tilde{h}_{\{\mu\}}, \tag{14}$$

therefore we conclude:

$$\tilde{h}_{\{\mu\}} \begin{cases} \neq 0, & \text{if} \quad \forall n : \mu_n \in S_n, \\ = 0, & \text{if} \quad \exists n : \mu_n \notin S_n. \end{cases} \tag{15}$$

This means that the part of the graph $\Gamma$ which has a nonzero contribution to $h$ is the totality of trees provided if we can show that there is no repetition of elements in $\{\mu\}$ and further that $h$ is independent of our choice of the $N - 1$ vertices. First, since the product:

$$\prod_{n=1}^{N-1} \epsilon_{\mu_n}$$

implies topologically a pairing of

$$\mu_1 \quad \text{for} \quad n = 1,$$

$$\mu_2 \quad \text{for} \quad n = 2,$$

$$\cdots$$

$$\mu_{N-1} \quad \text{for} \quad n = N - 1,$$

[6] R. Chisholm, Proc. Cambridge Phil. Soc. **48**, 300 (1952).
[7] N. Nakanishi, Progr. Theoret. Phys. (Kyoto), **26**, 337 (1961).

any repetition of elements in $\{\mu\}$ is impossible. Because, if there is a repetition, say,

$$\mu_{n'} = \mu_{n''},$$

then in the summation

$$\sum_{\forall\{\gamma\}} \left( \varepsilon_{\{\gamma\}} \prod_{n=1}^{N-1} \epsilon_{\mu_n \gamma_n} \right)$$

there are two terms:

$$\varepsilon_{\{\gamma'\}} \prod_{n=1}^{N-1} \epsilon_{\mu_n \gamma_{n'}} \quad \text{and} \quad \varepsilon_{\{\gamma''\}} \prod_{n=1}^{N-1} \epsilon_{\mu_n \gamma_{n''}}$$

which cancel each other exactly because

$$\{\gamma'\} = \mathsf{P}_{n'n''}\{\gamma''\},$$

where $\mathsf{P}_{n'n''}$ is the transposition operator for vertices $n'$ and $n''$. Further, repetitions more than twice are definitely impossible since an internal line is attached to *two* vertices only. This proves that there is no repetition of elements in $\{\mu\}$.

Rewrite (14):

$$h = \sum_{\forall\{\mu\}} \left[ \prod_{n=1}^{N-1} \alpha_{\mu_n}^{-1} \sum_{\forall\{\gamma\}} \left( \varepsilon_{\{\gamma\}} \prod_{n=1}^{N-1} \epsilon_{\mu_n} \epsilon_{\mu_n \gamma_n} \right) \right]. \tag{16}$$

By introducing a matrix

$$\vec{\vec{\sigma}}_{\{\mu\}} = \begin{pmatrix} \epsilon_{\mu_1 1}\epsilon_{\mu_1 1} & \epsilon_{\mu_1 1}\epsilon_{\mu_1 2} & \cdots & \epsilon_{\mu_1 1}\epsilon_{\mu_1 N-1} \\ \epsilon_{\mu_2 2}\epsilon_{\mu_2 1} & \epsilon_{\mu_2 2}\epsilon_{\mu_2 2} & \cdots & \\ & \cdots & & \\ \epsilon_{\mu_{N-1} N-1}\epsilon_{\mu_{N-1} 1} & \cdots & & \epsilon_{\mu_{N-1} N-1}\epsilon_{\mu_{N-1} N-1} \end{pmatrix} \tag{17}$$

we have from (16)

$$h = \sum_{\forall\{\mu\}} \left( \prod_{n=1}^{N-1} \alpha_{\mu_n}^{-1} \right) \cdot (\det \vec{\vec{\sigma}}_{\{\mu\}}).$$

Since (15) shows already that only "tree sets" give nonzero contributions to $\det \vec{\vec{\sigma}}_{\{\mu\}}$, thus by considering now only the "tree" graphs, we have

$$\prod_{n=1}^{N-1} (\epsilon_{u_n n})^2 = 1$$

for the diagonal product of (17) while the off-diagonal products are all zero since we are *dealing with only tree sets*. This leads to

$$\det \vec{\vec{\sigma}}_{\{\mu\}} = 1, \quad \text{for a } \textit{tree set}, \quad \{\mu\}. \tag{18}$$

Therefore, for each different tree set we have a different set of elements involved, yet (18) will hold just the same. Thus, we have

$$h = \sum_{\substack{\forall \text{ tree,} \\ \{\mu\}\in\Gamma}} \prod_{n=1}^{N-1} \alpha_{\mu_n}^{-1}, \tag{19}$$

then (11) and (19) lead directly to the proposition.

However, we note here that since the product sign in (19) runs from $n = 1$ to $N - 1$ only, the proof is not quite complete without showing $h$ is, indeed, independent of the choice of the $N - 1$ vertices out of the total $N$ vertices in the graph. This can be shown[5] as follows:

$$h = \begin{vmatrix} h_{11} & \cdots & h_{1,N-1} \\ \vdots & & \vdots \\ h_{N-1,1} & \cdots & h_{N-1,N-1} \end{vmatrix}$$

$$= \left| \begin{array}{ccc|c} h_{11} & \cdots & h_{1,N-1} & 0 \\ & & & 0 \\ \vdots & & \vdots & \vdots \\ h_{N-1,1} & & h_{N-1,N-1} & 0 \\ \hline 1 & \cdots & 1 & 1 \end{array} \right|$$

$$= \frac{1}{N} \lim_{\delta \to 0} \left| \begin{array}{cccc|c} h_{11}+\delta & h_{12} & \cdots & h_{1,N-1} & \delta \\ h_{21} & h_{22}+\delta & \cdots & & \\ \vdots & & & \vdots & \vdots \\ h_{N-1,1} & & \cdots & h_{N-1,N-1}+\delta & \delta \\ \hline 1 & 1 & \cdots & 1 & N\delta \end{array} \right|$$

$$= \frac{1}{N}\lim_{\delta\to 0}\left|\begin{array}{cccc|c} h_{11}+\delta & h_{12} & & h_{1,N-1} & -\sum_{i=1}^{N} h_{1i} \\ h_{21} & h_{22}+\delta & \cdots & & \\ \vdots & & & \vdots & \vdots \\ h_{N-1,1} & & \cdots & h_{N-1,N-1}+\delta & -\sum_{i=1}^{N} h_{N-1,i} \\ \hline 1 & 1 & \cdots & 1 & 1 \end{array}\right|$$

$$= \frac{1}{N}\lim_{\delta\to 0}\begin{vmatrix} h_{11}+\delta & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22}+\delta & \cdots & h_{2N} \\ \vdots & & & \vdots \\ h_{N-1,1} & & \cdots & h_{N-1,N} \\ 1 & 1 & \cdots & 1 \end{vmatrix}$$

$$= \frac{1}{N}\lim_{\delta\to 0}\frac{1}{\delta}\begin{vmatrix} h_{11}+\delta & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22}+\delta & \cdots & \\ \vdots & & & \vdots \\ \sum_{i=1}^{N} h_{i1}+\delta & \sum_{i=1}^{N} h_{i2}+\delta & \cdots & \sum_{i=1}^{N} h_{iN}+\delta \end{vmatrix}$$

$$= \frac{1}{N}\lim_{\delta\to 0}\frac{1}{\delta}\begin{vmatrix} h_{11}+\delta & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22}+\delta & \cdots & h_{2N} \\ \vdots & & & \vdots \\ h_{N1} & & \cdots & h_{NN}+\delta \end{vmatrix}$$

where in the last three steps we have used the properties:

$$\forall i : \sum_{j=1}^{N} h_{ij} = 0 \quad \text{and} \quad \sum_{j=1}^{N} h_{ji} = 0.$$

Therefore, the function $h(\alpha)$ is seen to be independent of the choice of the $N-1$ vertices involved in (19). This completes the proof. Q.E.D.

## V. AN ALTERNATIVE PROOF OF A THEOREM OF NAKANISHI[3]

As another example of exposition of the tree sets, we present here an alternative proof of a theorem due to Nakanishi (the notation used below is the same as that of Ref. 3, except we use $\mathcal{P}$ to denote a *path*).

*Proposition:* For three arbitrary external lines $A$, $B$, $C$, in a Feynman graph $\Gamma$, the following relation holds:

$$\eta^{(AC)} + \eta^{(BC)} - \eta^{(AB)} = 2\sum_{k} \eta^{C}(\mathcal{P}_k(n_A n_B)) \qquad (20)$$

[and two other equations obtained by cyclic permutation of the indices of (20)], where $n_A$ and $n_B$ are the vertices of the external lines $A$ and $B$, respectively. The $\eta$ functions are defined by

$$\eta^{(AB)} = W^{(AB)}/U,$$

where $W^{(AB)}$ is the $U$-function of the modified graph obtained from $\Gamma$ by identifying the two external vertices $n_A$ and $n_B$ of $\Gamma$, but leaving other things unchanged.

*Proof:* To make the proof more transparent, we introduce here a notation $\downarrow$ to denote the operation "to shrink . . ." which means to reduce a line to a point (or more generally, to *delete* a set of lines and identify their end vertices, respectively). Thus, the right-hand side of (20) can be written as

$$2\sum_{\forall \text{ possible } k} \eta^{C}(\downarrow \mathcal{P}_k(n_A n_B)). \qquad (21)$$

Further, by definition, we have

$$W^{(AB)}(\downarrow \mathcal{P}_k(n_B n_C)) = W^A(\downarrow \mathcal{P}_k(n_B n_C)) = [W^{(AB)}]_{\mathcal{P}_k(n_B n_C)=\phi}, \quad (22)$$

which is very convenient for the following purpose. Now, in terms of *tree graphs* (i.e., subgraphs of $\Gamma$ that correspond to the different tree sets) we will show that the function $W^{(AB)}$ can be written as

$$W^{(AB)} = \sum_k [W^{(AB)}]_{\mathcal{P}_k(n_B n_C)=\phi} + \sum_k [W^{(AB)}]_{\mathcal{P}_k(n_A n_C)=\phi}. \quad (23)$$

The proof proceeds as follows. Since $W^{(AB)}$ is just the $U$ function[3] of $\Gamma(n_A = n_B)$ obtained from $\Gamma$ by merely identifying the two external vertices $n_A$ and $n_B$, $W^{(AB)}$ contains a term formed by the product of all the elements of a *tree* obtained from $\Gamma(n_A = n_B)$. The summation over all the tree sets yields precisely the function $W^{(AB)}$.

For a given tree (*any* tree), $T$, there are two possibilities:

either

(i) $\mathcal{P}_k(n_B n_C) \cap {}_cT \neq \phi$, ($c$: complement)

or

(ii) $\mathcal{P}_k(n_B n_C) \subset T$, [$T$: a tree of $\Gamma(n_A = n_B)$].

In Case (i), the contribution to $W^{(AB)}$ is zero. In Case (ii), the condition $\mathcal{P}_k(n_B n_C) = \phi$ does not effect the terms contributed to $W^{(AB)}$. Further, here, a "tree"guarantees

$$\nexists\, \mathcal{P}(n_A n_C) \subset T$$

because we are presently dealing with $\mathcal{P}_k(n_B n_C)$ and $\Gamma(n_A = n_B)$. Thus,

$$\exists\, \mathcal{P}(n_A n_C) \Rightarrow \exists \text{ a closed path inside } T \quad (24)$$

which obviously contradicts the definition of a *tree*.

Similarly, we have

$$\exists\, \mathcal{P}(n_A n_C) \subset T \Rightarrow \quad \mathcal{P}(n_B n_C) \subset T.$$

This means the two summations on the right-hand side of (23) give exactly $W^{(AB)}$ provided we can establish

$$\forall T \Rightarrow \exists\, \mathcal{P}(n_B n_C) \subset T \text{ or } \exists\, \mathcal{P}(n_A n_C) \subset T. \quad (25)$$

But (25) is a self-evident property of a *tree set*, thus Q.E.D.

Finally, rewrite (23) into the form:

$$\eta^{(AB)} = \sum_k \eta^A(\downarrow \mathcal{P}_k(n_B n_C)) + \sum_k \eta^B(\downarrow \mathcal{P}_k(n_A n_C)) \quad (26)$$

and two other similar equations (obtained by cyclic permutation of the indices). That (26) leads directly to (20) is obvious. Q.E.D.

*Note added in proof*: Using Eq. (23) one can immediately obtain the following interesting theorem:

$$\forall n_1 \neq n_2 : U = \sum_{\forall k} U(\downarrow \mathcal{P}_k(n_1 n_2)),$$

which will be discussed with some other theorems in a preprint jointly with Dr. D. J. Kleitman.


## ACKNOWLEDGMENTS

I am grateful to Professor S. S. Schweber for his constant encouragement and his kind interest during the course of this investigation. I am also indebted to him and Dr. D. J. Kleitman for reading the manuscript and for making several suggestions. To Professor Noboru Nakanishi I express my indebtness for his guidance and personal communications.

# Scattering of Electromagnetic Waves by a Ferrite in a Waveguide

KALMAN KALIKSTEIN*
*U. S. Naval Applied Science Laboratory, Naval Base, Brooklyn, New York*
AND
LARRY SPRUCH†‡
*Physics Department, New York University, New York, New York*
*and Physics Department, University College, London, England*
(Received 1 May 1964)

As for any multichannel scattering problem, variational techniques can be utilized in the determination of the elements of the scattering matrix or of the equivalent network elements for a gyromagnetic obstacle in a waveguide. As always, however, it can be quite difficult to interpret numerical results which in general are neither upper nor lower bounds. A variational bound originally developed for the determination of the phase shift for a given angular momentum in a quantum mechanical central potential scattering problem is here adapted to the solution of a transversally magnetized, lossless ferrite slab in a rectangular waveguide propagating only one mode, the $TE_{10}$ mode. With a simple trial function and with the aid of a comparison scattering problem which need not be tensor in character (so that the determination of upper and lower bounds is not really difficult), close bounds are obtained on cot $\eta_e$ and cot $\eta_0$, the cotangent of the real uncoupled phase shifts associated with the even and odd standing waves, respectively. The bounds obtained on cot $\eta_e$ and cot $\eta_0$ determine bounds on the equivalent $\pi$ network. A second variational bound, which can be simpler to apply and which can be applied to a wider class of problems, is also developed. This too is an adaption of a formalism originally introduced in quantum mechanical scattering problems, and depends upon a consideration of the spectrum of the fundamental operator of the theory, the Hamiltonian in the quantum mechanical case and an analogue thereof in the electromagnetic case.

## I. INTRODUCTION

VARIATIONAL techniques for scattering problems were first developed by Schwinger[1] in connection with the scattering of electromagnetic waves by isotropic obstacles in waveguides. These techniques have subsequently been extended in a variety of ways; they have been applied in other fields of physics, and in particular to quantum mechanical scattering problems, and the domain of waveguide problems to which the variational approach is applicable has been broadened. With regard to waveguide problems, the extensions include the development by Hauser[2] of a variational expression for the elements of the scattering matrix for lossy anisotropic obstacles characterized by non-Hermitian electric permittivity and magnetic permeability tensors. Independently, Nikolskii[3] found extreme functionals for the determination of the reflection and transmission coefficients for a nonabsorptive anisotropic medium. There has also been a development of variational bounds as opposed to simply variational principles. Variational bounds for scattering parameters, as for eigenvalue problems (where they have long been known), give errors which are not only of second order but are of known sign. We will consider two variational bounds.

The first is that due to Kato,[4] who developed it in a form useful in the partial wave analysis of quantum mechanical potential scattering problems, a variational bound being obtained on cot $(\eta - \theta)$, where $\eta$ is the phase shift for the angular momentum under consideration and $\theta$ is a parameter chosen for convenience. This approach has been extended and applied[5,6] to the scattering of electromagnetic waves by isotropic lossless obstacles in a lossless waveguide for the case for which only one mode propagates and for which the obstacle is described by a scalar permittivity $\epsilon_0\epsilon(x, y, z)$ and a free-space permeability $\mu_0$. The case[5] in which the obstacle is symmetric with respect to a plane perpendicular to the axis of the waveguide can be reduced to two one-channel processes, but the unsymmetric case[6] is a true two-channel process which can not be so reduced.

* Present address: Fundamental Methods Associates, Inc., 31 Union Square West, New York, N. Y.

† National Science Foundation Senior Post-Doctoral Fellow; permanent address, New York University.

‡ Jointly supported by the Geophysics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Command, under Contract No. AF 19 (604) 4555, Project No. 7635, Task No. 76361, and the Office of Ordnance Research under Contract No. DA-30-69-ORD-2581, Project No. 2360, and the Office of Naval Research and the Advanced Research Projects Agency under Contract Nonr-285 (49), NR 012-109.

[1] D. S. Saxon, "Notes on Lectures by J. Schwinger: Discontinuities in Waveguides," 1945 (unpublished).

[2] W. Hauser, Mech. Appl. Math. 9, 427 (1958).

[3] V. V. Nikolskii, Radio Engineering and Electronics (USSR) 3, 146 (1958).

[4] T. Kato, Progr. Theoret. Phys. (Kyoto) 6, 394 (1951).

[5] L. Spruch and R. Bartram, J. Appl. Phys. 31, 905 (1960); R. Bartram and L. Spruch, *ibid.* 31, 913 (1960).

[6] R. Bartram and L. Spruch, J. Math. Phys. 3, 287 (1962).

The Kato formalism is limited to relatively simple scattering problems because it requires that there be some solvable problems of a difficulty roughly comparable with that of the original problem, though the solvable problems can sometimes be single-channel problems even when the original problem is multichannel.[6] A second disadvantage is that one must calculate matrix elements of the square of the basic operator, which for quantum mechanical problems is the Hamiltonian. To distinguish this formalism from the second variational bound, it will therefore be referred to as the quadratic variational bound. The advantage of the quadratic formalism is that, where applicable, it can provide *both* bounds on the scattering parameters under consideration.

We will show that the scattering of electromagnetic waves by a transversally magnetized lossless ferrite obstacle that is not too thick and that is located in a lossless rectangular waveguide propagating only the fundamental mode, and is symmetric about a plane perpendicular to the direction of propagation, and is also symmetric about a plane containing the axis of symmetry and the transverse dc magnetic field (a magnetically symmetric configuration with the scattering independent of the direction of propagation) is a problem sufficiently simple for the quadratic formalism to be very usefully applied. The adjective "useful" seems appropriate for the problem under consideration since, with specific and reasonable parameters, the upper and lower bounds obtained on the elements of the $\pi$ network differ from one another by only a few percent; whether or not the adjective "simple" is appropriate, other than to indicate that the quadratic formalism can in fact be used to obtain rather accurate results, is largely a matter of taste, but we note that the waveguide problem corresponds to the one-dimensional quantum mechanical scattering of a particle by a second particle bound to a center of force, where the potential interaction is a $3 \times 3$ matrix and the wavefunction has 3 components.

There is also available a variational bound formulation[7-10] which can be very much simpler to apply than the Kato formulation, especially at zero energy; the simplicity originates in part from the fact that the basic operator appears only linearly. This formalism was developed within the context of quantum mechanics and is applicable to a very much wider class of problems, including scattering by a compound system, largely because there need not be any related solvable problems. The relative disadvantage of this second (linear) variational bound, which will sometimes be referred to simply as the variational bound, is that one can obtain only one bound on the scattering parameter (or parameters) of interest. At zero incident energy, this second variational bound formulation supersedes the original variational principles.

The fact that the variational bound is applicable to quantum mechanical scattering by systems with internal degrees of freedoms suggests that it may also be applicable to waveguide problems, for the essence of the method lies in the possibility of decomposing the scattering problem, by a partial wave analysis for example, into scattering problems for which the scattering is completely characterized by a finite number of parameters; the quantum mechanical aspect of the problem is not the significant feature. In the quantum mechanical case the formulation is particularly simple at zero incident kinetic energy, that is, for $k = 0$, where $k$ is the incident wave number. To simplify the discussion we will therefore limit our formal analysis of the waveguide problem to $k = 0$, where $k$ is here the incident wavenumber along the axis of propagation. It is a relatively straightforward matter having established the connection at $k = 0$ to establish the connection for $k \neq 0$, even for $k$ sufficiently large such that more than one mode propagates.

We will begin with a few comments on the ferrite problem. We will then develop the quadratic variational bound in a form applicable to this problem, and apply it to the case of a uniform ferrite slab. We then present the variational bound formalism for $k = 0$. Finally, we briefly discuss the limitations of the two variational bound formulations.

## 2. FERRITE PROPERTIES AND BOUNDARY-VALUE PROBLEMS

Before proceeding, it will be helpful to outline the properties of ferrites and the difficulties encountered in the solution of boundary-value problems in which ferrite media are involved. Neglecting damping, a ferrite region magnetized and saturated by a dc magnetic field, $H_{dc}$, in the $y$ direction, and subjected to rf fields, $\mathbf{H}$, will contain rf flux densities, $\mathbf{B}$, given by

[7] L. Spruch and L. Rosenberg, Phys. Rev. **116**, 1034 (1959); **117**, 1095 (1960).
[8] L. Rosenberg, L. Spruch, and T. F. O'Malley, Phys. Rev. **118**, 184 (1960).
[9] L. Spruch, in *Lectures in Theoretical Physics, Boulder*, 1961, edited by W. E. Brittin and B. W. Downs (Interscience Publishers, Inc., New York, 1962), Vol. 4.
[10] Y. Hahn, T. F. O'Malley, and L. Spruch, Phys. Rev. **128**, 932 (1962); **130**, 381 (1963).

$$\mathbf{B} = \mu_0 \mu \mathbf{H}, \tag{1a}$$

where

$$\mu = \begin{vmatrix} \mu_1 & 0 & -ik_1 \\ 0 & 1 & 0 \\ ik_1 & 0 & \mu_1 \end{vmatrix}, \tag{1b}$$

and where

$$\mu_0 \mu_1 = \mu_0 + |\gamma| M\omega_0/(\omega_0^2 - \omega^2),$$
$$\mu_0 k_1 = |\gamma| M\omega/(\omega_0^2 - \omega^2).$$

$\mu_0$ is the permeability of free space, $\gamma$ the gyromagnetic ratio of the electron, $M$ the saturation moment, $\omega$ the operating circular frequency, and $\omega_0 = \gamma H_{dc}$. The tensor form of the equations indicates that an rf **H** field applied along the $x$ or the $z$ axis will give rise to rf components of **B** along both the $x$ and $z$ axes.

The solution of boundary-value problems in which ferrite media are involved is a difficult task. In the case of the rectangular waveguide completely filled with ferrite where the magnetizing field $H_{dc}$ is perpendicular (pointing in the $y$ direction) to the broad face of the waveguide and transverse to the direction of propagation, $z$ (see Fig. 1), the modes which are independent of $y$ have only an electric field, **E**, component in the $y$ direction and **H** components transverse to **E**.[11] These modes are the same as if the waveguide were filled with an isotropic, homogenous medium of permeability $\mu_0 (\mu_1^2 - k_1^2)/\mu_1$, insofar as **E** and **B** are concerned. The components of **H** in terms of **B**, obtained from Eqs. (1), are

$$H_x = (\mu_1 B_x + ik_1 B_z)[\mu_0(\mu_1^2 - k_1^2)]^{-1},$$
$$H_z = (-ik_1 B_x + \mu_1 B_z)[\mu_0(\mu_1^2 - k_1^2)]^{-1}.$$

As a consequence of the distortion of the **H** field in the ferrite medium if a normal waveguide mode is incident on a semi-infinite ferrite filled section of waveguide, an infinite number of modes must be excited in order to satisfy the boundary conditions on the incident, reflected, and transmitted waves,

$$E_y(\text{inc}) + E_y(\text{ref}) = E_y(\text{trans})$$
$$H_x(\text{inc}) + H_x(\text{ref}) = H_x(\text{trans}),$$

at the air–ferrite interface. The representation in terms of a superposition of all possible $TE_{n0}$ modes leads to an infinite number of equations of slow convergence. (For a finite section, the fields will consist of forward and backward waves in the slab.)

[11] A. A. Th. Van Trier, Appl. Sci. Res. **3B**, 305 (1953).

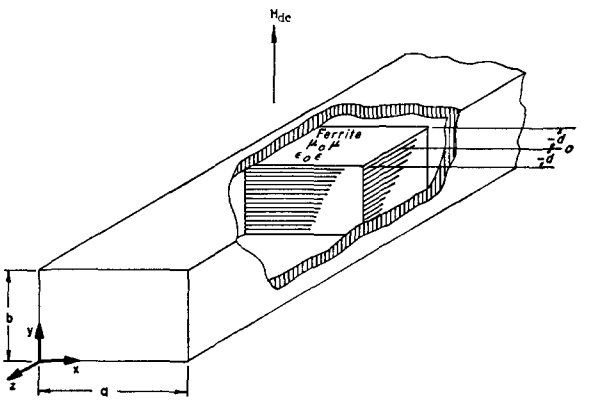


FIG. 1. Magnetized ferrite filled section of waveguide.

The computation of the reflection and transmission coefficients therefore meets with great difficulties.[12] Sharpe and Heim[13] have found an approximate solution for the semi-infinite problem in the case where no higher modes than the first propagate in the ferrite. The solution for the electric field is expressed in terms of a Neumann series obtained by iteration of a singular integral equation which satisfies the boundary conditions at the interface.

It has been shown[14] that the scattering matrix of microwave junctions containing ferrites satisfy the same requirements as that of junctions containing ordinary scalar media except that the ferrite scattering matrix is in general nonreciprocal. For scalar media, the scattering matrix is symmetric, $S_{nm} = S_{mn}$. For ferrite media, the asymmetry of the permeability tensor describing the medium induces a nonsymmetric scattering matrix, $S_{mn} \neq S_{nm}$, for an asymmetric distribution of obstacles. A distribution is considered asymmetrical even in the case where there is geometrical symmetry, if the geometrically similar portions are oppositely magnetized. If the ferrite is symmetrically distributed, the relationship between incident and transmitted fields is independent of the direction of propagation. For an $N$ channel waveguide containing lossless anisotropic material, since the scattering matrix is unitary and in general nonsymmetric, the asymptotic scattering effects can be described by $N^2$ real independent quantities. If we restrict ourselves to the two-channel symmetric case, that is, to a uniform waveguide propagating a single mode with obstacles symmetric with respect to a plane perpendicular to the direction of propagation and symmetric with respect to the dc magnetic field, the

[12] P. S. Epstein, Rev. Mod. Phys. **28**, 3 (1956).
[13] C. B. Sharpe and D. S. Heim, IRE Trans. Microwave Theory Tech. **6**, 42 (1958).
[14] G. S. Heller, Proc. IRE, **44**, 1386 (1956).

number of independent variables is reduced from four to two. Then as in the isotropic case, the solution can be represented in terms of even and odd standing waves, each characterized by one real number, the phase shift $\eta$. (Subscripts e and o will be used to denote quantities associated with the even and odd solutions, respectively, but will be omitted wherever the formalism is the same for both solutions.)

### 3. THE BASIC IDENTITY

Limiting ourselves to the ferrite system mentioned in the Introduction, which we can state more formally as the conditions that $\mu(x, z) = \mu(x, -z)$ and $\mu(x, z) = \mu(a - x, z)$ ($a$ is the wide dimension of the waveguide), with similar conditions on $\epsilon(x, z)$, we will derive our basic relations in terms of the magnetic field intensity, $\mathbf{H}(\mathbf{r})$, which has two nonvanishing components, $H_x$ and $H_z$. It might seem preferable to use the electric-field formulation, since $\mathbf{E}(\mathbf{r})$ has only one nonvanishing component, $E_y$, but unfortunately, because of the presence of the tensor permeability, the boundary conditions at the air–ferrite interface cannot then be satisfied with a trial function consisting of a finite number of modes.

Assuming a time dependence exp $(-i\omega t)$, Maxwell's equations are

$$\nabla \times \mathbf{H} = -i\omega\epsilon_0\epsilon\mathbf{E},$$

$$\nabla \times \mathbf{E} = i\omega\mu_0\mu\mathbf{H},$$

$$\nabla \cdot \mathbf{B} = 0, \qquad \nabla \cdot \mathbf{D} = 0,$$

where the relative permittivity $\epsilon$ is a scalar and the relative permeability $\mu$ is a Hermitian tensor in the ferrite, given by Eq. (1b), and $\epsilon = 1$ and $\mu = 1$, the three by three unit matrix, in the rest of the waveguide medium. It follows from Maxwell's equations that

$$\mathfrak{L}\mathbf{H} \equiv (-\nabla \times \epsilon^{-1}\nabla \times + \mu\omega^2/c^2)\mathbf{H} = 0. \tag{2}$$

Equation (2) can be rewritten as

$$\mathfrak{L}\mathbf{H} = (\omega^2/c^2 - \mathfrak{IC})\mathbf{H} = (\omega^2/c^2 - T - V)\mathbf{H} = 0, \tag{3a}$$

where

$$T = -\nabla^2, \tag{3b}$$

$$V = -W = \nabla\nabla\cdot + (1/\epsilon - 1)\nabla \times \nabla \times + [\nabla(1/\epsilon)] \times \nabla \times + (1 - \mu)\omega^2/c^2. \tag{3c}$$

With $\mathfrak{IC}$, $\omega^2/c^2$, $T$, and $V$ symbolically identified with the Hamiltonian and with the total, the kinetic, and the potential energy, respectively, we have a (purely formal) connection with the Schrödinger equation. This purely formal connection—the various quantities do not even have the dimensions of energy—simplifies the adaptation of the quantum mechanical variational bounds to the waveguide case.

As it must, $V$ vanishes asymptotically since $\epsilon$ and $\mu$ approach 1 asymptotically and since the fields on which $V$ operates become divergenceless. The fact that $V$ is nonlocal and $\omega$ dependent, which corresponds quantum mechanically to an energy dependent nonlocal potential, will cause no difficulty whatever. In fact, we will never need to use the explicit form (3c), but we note for later reference with regard to a monotonicity theorem that $V$ has a quite simple $\mu$ dependence.

For a $TE_{10}$ mode in a hollow rectangular waveguide, the components of the magnetic intensity are represented by the following relations for propagation in the $+z$ direction,

$$H_x = \sin(\pi x/a) \exp[-i(\omega t - kz)],$$

$$H_z = (\pi/ak) \cos(\pi x/a) \exp[-i(\omega t - kz - \tfrac{1}{2}\pi)],$$

where

$$k = [(\omega/c)^2 - (\pi/a)^2]^{\frac{1}{2}}.$$

Thus, for the space-dependent parts of $\mathbf{H}$, the even and odd standing wave solutions have the asymptotic forms for $z \to \infty$,

$$\begin{aligned}
\mathbf{H}_e \to\ & \mathbf{a}_x \sin(\pi x/a)[-\sin(kz + \theta) \\
& + \cot(\eta_e - \theta)\cos(kz + \theta)] \\
& + \mathbf{a}_z(\pi/ak)\cos(\pi x/a)[-\sin(kz + \theta + \tfrac{1}{2}\pi) \\
& + \cot(\eta_e - \theta)\cos(kz + \theta + \tfrac{1}{2}\pi)], \\
\mathbf{H}_o \to\ & \mathbf{a}_x \sin(\pi x/a)[\cos(kz + \theta) \\
& + \cot(\eta_0 - \theta)\sin(kz + \theta)] \\
& + \mathbf{a}_z(\pi/ak)\cos(\pi x/a)[\cos(kz + \theta + \tfrac{1}{2}\pi) \\
& + \cot(\eta_0 - \theta)\sin(kz + \theta + \tfrac{1}{2}\pi)],
\end{aligned} \tag{4}$$

where $0 \leq \theta < \pi$ ($\theta$ is an arbitrary parameter) and $\mathbf{a}_x$ and $\mathbf{a}_z$ are unit vectors in the $x$ and $z$ direction, respectively.

Consider $H_{ex}$, for example; we want $H_{ex} \to N \sin(\pi x/a)\cos(kz + \eta_e)$, where $N$ is an arbitrary normalization factor. The choice $N = [\sin(\eta_e - \theta)]^{-1}$ gives the asymptotic form of $H_{ex}$ of Eq. (4).

We now introduce a trial magnetic field $\mathbf{H}_t$ which is required to behave asymptotically in a fashion defined by Eq. (4), but with $\eta_e$ and $\eta_0$ replaced by $\eta_{et}$ and by $\eta_{ot}$, respectively. Consider for the

moment the case for which $\mu$ and $\epsilon$ are continuous functions. We then require that $\mathbf{H}_t$ be continuous and that it have a finite first derivative and an integrable second derivative. The further conditions to be imposed upon $\mathbf{H}_t$ will be arrived at below.

Consider the quantity

$$\int [\mathbf{H}^* \cdot \mathfrak{L}\mathbf{H}_t - \mathbf{H}_t \cdot (\mathfrak{L}\mathbf{H})^*]\, d\tau,$$

with the star representing the Hermitian adjoint. We can on the one hand substitute for $\mathfrak{L}$ explicitly; the terms involving $\mu$ vanish since $\mu^* = \mu$. We can on the other hand, by Eq. (2), drop the $\mathfrak{L}\mathbf{H}$ term. We thereby arrive at

$$\int \mathbf{H}^* \cdot \mathfrak{L}\mathbf{H}_t\, d\tau = -\int [\mathbf{H}^* \cdot \nabla \times (\epsilon^{-1} \nabla \times \mathbf{H}_t) - \mathbf{H}_t \cdot \nabla \times (\epsilon^{-1} \nabla \times \mathbf{H}^*)]\, d\tau, \quad (5a)$$

where the range of integration is over the interior of the waveguide. The right-hand side can be transformed to give

$$\int \mathbf{H}^* \cdot \mathfrak{L}\mathbf{H}_t\, d\tau = \int [\mathbf{H}^* \times (\epsilon^{-1} \nabla \times \mathbf{H}_t) - \mathbf{H}_t \times (\epsilon^{-1} \nabla \times \mathbf{H}^*)] \cdot d\mathbf{S}. \quad (5b)$$

Imposing the requirement on $\mathbf{H}_t$ that $(\nabla \times \mathbf{H}_t)_{\text{tang}} = 0$ on conducting surfaces, the surface integral vanishes on the conducting surfaces, and the contribution from the asymptotic region is twice the contribution from either asymptotic face. Thus, we have

$$\int \mathbf{H}^* \cdot \mathfrak{L}\mathbf{H}_t\, d\tau = 2k[1 + (\pi/ak)^2](\tfrac{1}{2}ab)[\cot(\eta_t - \theta) - \cot(\eta - \theta)], \quad (6)$$

where $b$ is the narrow dimension of the waveguide. In order to conform to our previous notation, let

$$\mathbf{u}_\theta \equiv (2/ab)^{\frac{1}{2}}\mathbf{H}, \quad \mathbf{u}_{\theta t} \equiv (2/ab)^{\frac{1}{2}}\mathbf{H}_t, \quad \mathbf{w}_\theta \equiv \mathbf{u}_{\theta t} - \mathbf{u}_\theta.$$

Since

$$\mathfrak{L}\mathbf{w}_\theta = \mathfrak{L}\mathbf{u}_{\theta t},$$

Eq. (6) can now be written

$$k[1 + (\pi/ak)^2] \cot(\eta - \theta) = k[1 + (\pi/ak)^2] \cot(\eta_t - \theta) - \tfrac{1}{2} \int \mathbf{u}_{\theta t}^* \cdot \mathfrak{L}\mathbf{u}_{\theta t}\, d\tau + \tfrac{1}{2} \int \mathbf{w}_\theta^* \cdot \mathfrak{L}\mathbf{w}_\theta\, d\tau. \quad (7)$$

Consider now the more realistic case of a discrete ferrite obstacle, so that $\epsilon$ and $\mu$ are not continuous. The volume integral in Eq. (5a) must now be split into a volume integral over the interior of the obstacle and a second volume integral over the remainder of the interior of the waveguide. The surface integral of Eq. (5b) now includes not only integrals of the conducting surfaces and of the end surface, but two integrals over the surface of the obstacle, with oppositely directed surface elements $d\mathbf{S}$. Since the tangential components of $\mathbf{H}^*$ and of $(1/\epsilon)(\nabla \times \mathbf{H}^*)$ (equal to $i\omega\epsilon_0\mathbf{E}^*$) are continuous, these latter two surface integrals cancel if the tangential components of $\mathbf{H}_t$ and of $\epsilon^{-1}\nabla \times \mathbf{H}_t$ are continuous at the ferrite–air interface, and we arrive again at Eqs. (6) and (7). The significant point is that the continuity conditions on $\mathbf{H}_t$ can be relaxed at the interface with regard to the requirement on the normal component of $\mu\mathbf{H}_t$ and of $\nabla \times \mathbf{H}_t$.

The first two terms on the right-hand side of the basic identity [Eq. (7)] constitute a variational approximation for $\cot(\eta - \theta)$; the third term is the error term and is of the order of the square of the error in the trial function. Equation (7) is the starting point for the development of both variational bounds. We consider the Kato formulation first.

## 4. THE "QUADRATIC" VARIATIONAL BOUND

### A. The Formalism

Kato's method for obtaining rigorous bounds on the error term can with some modification be applied to the ferrite case. We obtain upper and lower bounds on the error term $\int \mathbf{w}_\theta^* \cdot \mathfrak{L}\mathbf{w}_\theta\, d\tau$, namely,

$$-\alpha_\theta^{-1} \int (\mathfrak{L}\mathbf{u}_{\theta t})^* \cdot (\epsilon\rho^{-1}\mathfrak{L}\mathbf{u}_{\theta t})\, d\tau \leq \int \mathbf{w}_\theta^* \cdot \mathfrak{L}\mathbf{w}_\theta\, d\tau \leq \beta_\theta^{-1} \int (\mathfrak{L}\mathbf{u}_{\theta t})^* \cdot (\epsilon\rho^{-1}\mathfrak{L}\mathbf{u}_{\theta t})\, d\tau. \quad (8)$$

$\alpha_\theta$ is the smallest positive eigenvalue and $\beta_\theta$ the smallest (in absolute value) negative eigenvalue of the associated eigenvalue problem

$$\mathfrak{L}\psi_n(\mathbf{r}) + \sigma_n\epsilon^{-1}\rho(\mathbf{r})\psi_n(\mathbf{r}) = -\nabla \times \epsilon^{-1}\nabla \times \psi_n + \epsilon^{-1}[(\omega/c)^2\mu\epsilon + \sigma_n\rho]\psi_n = 0. \quad (9)$$

The matrix weight function $\rho(\mathbf{r})$ is positive-definite, an even function of $z$, magnetically symmetric, and lossless. The eigenfunctions $\psi_n$ have the asymptotic phase

$$\delta(\sigma_n) = \theta + n\pi, \qquad n = 0, \pm 1, \cdots, \quad (10)$$

corresponding to the eigenvalues $\sigma_n$. (The question of the completeness of the $\psi_n$ set of functions is discussed in Ref. 5. The conclusions reached there apply also to our case.) If $\int (\mathfrak{L}\mathbf{u}_{\theta t})^* \cdot (\epsilon\rho^{-1}\mathfrak{L}\mathbf{u}_{\theta t})\, d\tau$ is small, rough lower bounds on $\alpha_\theta$ and $\beta_\theta$ will provide close bounds on the phase shift.

From Eqs. (7) and (8), we obtain

$$-\tfrac{1}{2}\alpha_\theta^{-1}\int (\mathfrak{L}\mathbf{u}_{\theta t})^*\cdot(\epsilon\rho^{-1}\mathfrak{L}\mathbf{u}_{\theta t})\,d\tau$$

$$\leq k[1+(\pi/ak)^2]\cot(\eta-\theta)$$

$$-k[1+(\pi/ak)^2]\cot(\eta_t-\theta)+\tfrac{1}{2}\int \mathbf{u}_{\theta t}^*\cdot\mathfrak{L}\mathbf{u}_{\theta t}\,d\tau$$

$$\leq \tfrac{1}{2}\beta_\theta^{-1}\int (\mathfrak{L}\mathbf{u}_{\theta t})^*\cdot(\epsilon\rho^{-1}\mathfrak{L}\mathbf{u}_{\theta t})\,d\tau. \tag{11}$$

### B. Lower Bounds on $\alpha_\theta$ and on $\beta_\theta$

In the analysis of scalar problems, lower bounds on $\alpha_\theta$ and on $\beta_\theta$ can be obtained[4,5] by using a comparison scalar potential for which the scattering problem defined by Eqs. (9) and (10) can be solved exactly, and applying the monotonicity theorem which states, for scalar potentials, that the phase shift increases monotonically as the scalar potential $W$ increases. The quantum mechanical analog of $W$ is the negative of the potential $V$. The extension[6] of the monotonicity theorem to scattering by an $N$ by $N$ matrix potential when there are $N$ open channels is the statement that the $N$ eigenphase shifts each increase monotonically as the matrix potential $W$ increases; the matrix potentials $W_1$ and $W_2$ will be said to satisfy $W_1 \geq W_2$ if each of the eigenvalues of $W_1 - W_2$ is positive. This extension was used[6] in the analysis of one-dimensional scattering by an asymmetric potential, which was reformulated as the scattering by a two by two matrix potential, with two channels open. It is simple to extend[10] the monotonicity theorem to include the case in which there are $M$ closed channels as well as $N$ open channels, where $M$ can be infinite; this case arises in the quantum mechanical scattering of a particle by a compound system at energies below the threshold for breakup.

The form of the monotonicity theorem that is needed for our present application is a special case of the general theorem just quoted, but it is so simple a version that one can readily derive it directly. We have scattering by a $3 \times 3$ matrix potential, with two open channels and an infinite number of closed channels, but the assumption of symmetry with respect to $z \leftrightarrow -z$ reduces the problem to one in which the eigenmodes can immediately be decoupled beforehand by symmetry considerations, being simply the even and the odd waves. The eigenphase shifts are therefore simply the even and odd phase shifts, and the proof of the form of the monotonicity theorem that we need follows from the usual proof of the monotonicity theorem for scalar potentials by simply replacing scalar potentials by matrix potentials.

Thus, let $\eta$ be the even or the odd phase shift associated with the matrix potential $W$, defined by Eq. (3c), and let $\eta + d\eta$ be the corresponding phase shift associated with the matrix potential $W + dW$, where the correspondence is established by continuity considerations. Assume that $dW(\mathbf{r})$ is positive definite for all $\mathbf{r}$. To determine $\eta + d\eta$ we choose as our trial solution the exact corresponding (even or odd) eigenmode solution $\mathbf{u}$ associated with $W$ for the choice $\theta = 0$. For $dW$ sufficiently small, we find from Eq. (7), since

$$\mathfrak{L}\mathbf{u}_{\theta t} = \mathfrak{L}(W+dW)\mathbf{u}_0(W) = dW\mathbf{u}_0(W)$$

and since

$$\cot(\eta+d\eta)-\cot\eta = -\csc^2\eta\,d\eta,$$

that

$$d\eta = \tfrac{1}{2}\sin^2\eta(1/k)[1+(\pi^2/k^2a^2)]^{-1} \times \int \mathbf{u}^*\cdot dW\mathbf{u}\,d\tau \geq 0,$$

where the last step follows from the assumed positive definiteness of $dW$. The monotonicity theorem, for a finite difference between the matrix potentials, follows immediately. The conditions under which $dW(\mathbf{r}) \geq 0$ in terms of conditions on $\epsilon(\mathbf{r})$ and $\mu(\mathbf{r})$ are derived in Appendix A.

As opposed to the scalar case, the matrix potentials for which the scattering problem can be solved exactly and readily are far and few between. As noted previously,[6] it is therefore of the greatest practical importance that the comparison potential can be a multiple of the unit matrix, that is, effectively a scalar.

We know then from the monotonicity theorem that the phase shift corresponding to the ferrite slab which fills the space $-d < z < d$ out to the conducting boundaries of the waveguide is less than the phase shift which would result if the slab were replaced by a dielectric slab of the same dimensions and same permittivity (in what follows $\epsilon$ denotes only the permittivity of the ferrite) as the ferrite one, and whose permeability, $\mu_0 L$, where $L$ can be a function of $\mathbf{r}$ but is to be a scalar, is such that $\mu_0(L1-\mu) \geq 0$, or $L1 - \mu \geq 0$. The eigenvalues $\lambda$ are determined by solving the secular equation

$$\begin{vmatrix} (L-\mu_1)-\lambda & 0 & +ik_1 \\ 0 & (L-1)-\lambda & 0 \\ -ik_1 & 0 & (L-\mu_1)-\lambda \end{vmatrix} = 0.$$

Thus, we must choose $L$ such that

$$L \geq \mu_1 \pm k_1, \quad \text{and} \quad L \geq 1,$$

for all $\mathbf{r}$.

Let $\delta(\sigma)$ be the phase shift corresponding to the ferrite slab whose potential is equal to

$$W + \sigma\rho = (\omega^2/c^2)(\mu\epsilon - 1) + \sigma(\omega^2/c^2)\mu\epsilon,$$

where we have made the choice $\rho = (\omega^2/c^2)\mu\epsilon$. Let $\delta'(\sigma)$ be the phase shift corresponding to the dielectric slab whose potential is equal to $W_L + \sigma\rho_L$, where

$$W_L = (\omega^2/c^2)(L\epsilon - 1), \qquad \rho_L = (\omega^2/c^2)L\epsilon.$$

We then have

$$(W_L + \sigma\rho_L) - (W + \sigma\rho) = (\omega/c)^2\epsilon(L1 - \mu)(1 + \sigma).$$

It follows that $\delta'(\sigma) > \delta(\sigma)$ for $\sigma > -1$, and $\delta(\sigma) > \delta'(\sigma)$ for $\sigma < -1$.

The phase shift $\delta'(\sigma)$ is obtained by applying the boundary conditions at the air–dielectric interface. The even and odd phase shifts are

$$\delta_e'(\sigma) = -kd + \tan^{-1}\left[\frac{[k' + (\pi/a)^2/k']}{\epsilon[k + (\pi/a)^2/k]} \tan k'd\right],$$

$$\delta_0'(\sigma) = -kd + \tan^{-1}\left[\frac{\epsilon[k + (\pi/a)^2/k]}{[k' + (\pi/a)^2/k']} \tan k'd\right], \tag{12}$$

where

$$k' = k'(\sigma) = [k^2 + (\omega^2/c^2)(\epsilon L - 1) + \sigma(\omega^2/c^2)\epsilon L]^{\frac{1}{2}}.$$

It can be seen from Fig. 2, since $\delta'(\sigma) > \delta(\sigma)$ for $\sigma > -1$, that a lower bound on $\alpha_\theta$ is given by the value of $\sigma$, denoted by $\alpha_\theta'$ and defined by $\delta'(\alpha_\theta') \equiv \theta$, provided that $\eta' \equiv \delta(0) < \theta$ and $\eta > \theta - \pi$.

A lower bound on $\beta_\theta$ can be obtained as follows. As $\sigma$ approaches $-\infty$, it is evident that the phase shift $\delta'(\sigma)$ approaches $-kd$ in the odd case and $-kd - \frac{1}{2}\pi$ in the even case. Since $\delta(\sigma) > \delta'(\sigma)$ for $\sigma < -1$, it will be noticed, if $-kd > \theta - \pi$ in the odd case or $-kd - \frac{1}{2}\pi > \theta - \pi$ in the even case and if $\eta < \theta$, that there can be no eigenvalues and $\beta_\theta$ can therefore be regarded as infinite.

From the rough bounds already obtained on $\eta$,

$$-kd - \tfrac{1}{2}\pi < \eta_e < \eta_e', \qquad -kd < \eta_0 < \eta_0', \tag{13}$$

it follows that a sufficient condition for obtaining lower bounds on $\alpha_\theta$ and $\beta_\theta$ by the above procedure is that $d[k^2 + W_L]^{\frac{1}{2}}$ be less than $\frac{1}{2}\pi$ in the even case and less than $\pi$ in the odd case; this means that $2d$ must be less than $\frac{1}{2}\lambda_g$, where $\lambda_g$ is the smallest guide wavelength in the dielectric, if we are to use the above form of analysis for both the even and odd case. This restriction is not fundamental, since there are other possibilities for obtaining lower bounds on $\alpha_\theta$ and $\beta_\theta$.

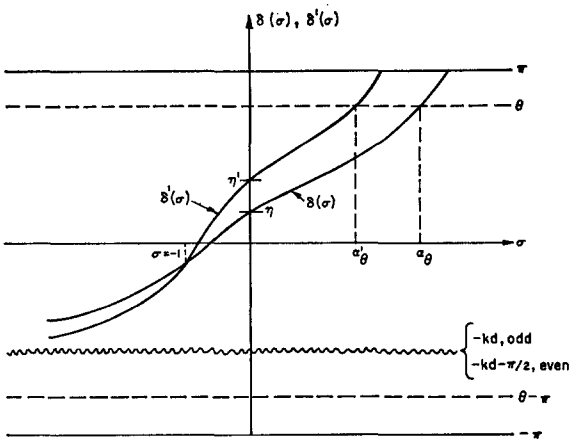


FIG. 2. Determination of lower bound on $\alpha_\theta$ and $\beta_\theta$. The curve $\delta'(\sigma)$ corresponds to a dielectric slab of permittivity $\epsilon_0\epsilon$ and permeability $\mu_0 L$ filling the guide in the range $-d < z < d$, and whose potential equals $(\omega^2/c^2)[L\epsilon(1 + \sigma) - 1]$. The curve $\delta(\sigma)$ corresponds to a ferrite slab of permittivity $\epsilon_0\epsilon$ and permeability $\mu_0\mu$ filling the guide in the range $-d \leq z \leq d$, and whose potential equals $(\omega^2/c^2)[\mu\epsilon(1 + \sigma) - 1]$. ($\mu$ is the relative tensor permeability.) It is assumed that $\eta' < \theta$, $\eta > \theta - \pi$, and $-kd - \frac{1}{2}\pi < \theta - \pi$.

## 5. APPLICATION TO A FERRITE SLAB IN A RECTANGULAR WAVEGUIDE

### A. Trial Function

The exact solution of a dielectric slab, having the same dimensions and permittivity as the ferrite one, is introduced as a trial function. The scalar permeability of the slab is retained as a parameter which can be varied to improve the bounds. The maximum axial extent of the slab is $2d \leq \frac{1}{2}\lambda_g$. This enables us to use the lower bounds on the associated eigenvalues $\alpha_\theta$ and $\beta_\theta$ developed in Section (4). The normalization constant $\theta$ can also be varied to obtain more accurate bounds on the even and odd phase shifts. However, for convenience, $\theta = \pi - kd$ is picked in the determination of $\alpha_\theta$ and $\beta_\theta$ in the bounds on cot $(\eta_0 - \theta)$, and $\theta = \pi - kd$ and $\theta = \frac{1}{2}\pi - kd$ in the determination of $\alpha_\theta$ and $\beta_\theta$, respectively, in the bounds on cot $(\eta_e - \theta)$. With this choice of $\theta$, $\beta_\theta \to \infty$, while a lower bound on $\alpha_\theta$ can be determined from Eq. (12), namely,

$$\alpha_\theta > \alpha_\theta' = \frac{\pi^2 - k^2d^2 - (\omega^2/c^2)(\epsilon L - 1)}{(\omega^2/c^2)\epsilon L d^2}. \tag{14}$$

The trial functions for $-d < z < d$ with the above choices of $\theta$ are

odd, upper and lower bound:

$$\mathbf{u}_{\theta t} = -\left(\frac{2}{ab}\right)^{\frac{1}{2}}\left(\frac{1}{\sin Kd}\right)\left[\mathbf{a}_x \sin\left(\frac{\pi x}{a}\right)\sin Kz + \mathbf{a}_z\left(\frac{\pi}{aK}\right)\cos\left(\frac{\pi x}{a}\right)\cos Kz\right]; \quad (15a)$$

even, lower bound:

$$\mathbf{u}_{\theta t} = -\left(\frac{2}{ab}\right)^{\frac{1}{2}}\frac{\epsilon[k + (\pi/a)^2 k^{-1}]}{[K + (\pi/a)^2 K^{-1}]}\frac{1}{\sin Kd} \times \left[\mathbf{a}_x \sin\left(\frac{\pi x}{a}\right)\cos Kz - \mathbf{a}_z\left(\frac{\pi}{aK}\right)\cos\left(\frac{\pi x}{a}\right)\sin Kz\right]; \quad (15b)$$

even, upper bound:

$$\mathbf{u}_{\theta t} = -\left(\frac{2}{ab}\right)^{\frac{1}{2}}\left(\frac{1}{\cos Kd}\right)\left[\mathbf{a}_x \sin\left(\frac{\pi x}{a}\right)\cos Kz - \mathbf{a}_z\left(\frac{\pi}{aK}\right)\cos\left(\frac{\pi x}{a}\right)\sin Kz\right]. \quad (15c)$$

The parameter $K$ is

$$K = [k^2 + (\omega^2/c^2)(\mu_t\epsilon - 1)]^{\frac{1}{2}},$$

where $\mu_t$ is the trial (scalar) permittivity.

The normalization of the trial functions in Eqs. (15) and the determination of the trial phase shifts is accomplished by matching the tangential components of $\mathbf{E}_t$ and $\mathbf{H}_t$ at $z = \pm d$ to the asymptotic expressions of Eqs. (4).

Substitution of the trial functions and trial phase shifts in Eq. (11) (where the range of integration is over the volume of the slab, $V = 2\,abd$) yields as the bounds on $\cot\eta_e$ and $\cot\eta_o$

$$-\frac{1}{2}\csc^2(Kd)P\left[Q^- + \left(\frac{\pi}{aK}\right)^2 Q^+\right]\alpha_\theta^{-1} \le \left[k + \left(\frac{\pi}{a}\right)^2 k^{-1}\right]\cot(\eta_0 - \pi + kd) - \frac{1}{\epsilon}\left[K + \left(\frac{\pi}{a}\right)^2 K^{-1}\right]\cot Kd + \frac{1}{2}\csc^2(Kd)R\left[Q^- + \left(\frac{\pi}{aK}\right)^2 Q^+\right] \le 0, \quad (16a)$$

$$\left[k + \left(\frac{\pi}{a}\right)^2 k^{-1}\right]\cot\left(\eta_e - \frac{\pi}{2} + kd\right) \le -\frac{1}{\epsilon}\left[K + \left(\frac{\pi}{a}\right)^2 K^{-1}\right]\tan(Kd) - \frac{1}{2}\sec^2(Kd)R\left[Q^+ + \left(\frac{\pi}{aK}\right)^2 Q^-\right], \quad (16b)$$

$$\left[k + \left(\frac{\pi}{a}\right)^2 k^{-1}\right]\frac{1}{\epsilon}\left[K + \left(\frac{\pi}{a}\right)^2 K^{-1}\right]\cot(Kd) - \frac{1}{2}\left[k + \left(\frac{\pi}{a}\right)^2 k^{-1}\right]\csc^2(Kd)R\left]Q^+ + \left(\frac{\pi}{aK}\right)^2 Q^-\right] - \frac{1}{2}\left[k + \left(\frac{\pi}{a}\right)^2 k^{-1}\right]\csc^2(Kd)P\left[Q^+ + \left(\frac{\pi}{aK}\right)^2 Q^-\right]\alpha_\theta^{-1} \le \left(\frac{1}{\epsilon}\right)^2\left[K + \left(\frac{\pi}{a}\right)^2 K^{-1}\right]^2\cot(\eta_e - \pi + kd), \quad (16c)$$

where

$$Q^{+,-} = (1 \pm \sin 2Kd/2Kd)d,$$

$$\alpha_\theta^{-1} = (\omega/c)^2\epsilon L d^2/[\pi^2 - k^2d^2 - (\omega/c)^2(L\epsilon - 1)d^2], \quad (L = 1.65 \geq \mu_1 \pm k_1),$$

$$R = -\epsilon^{-1}[(\pi/a)^2 + K^2] + (\omega/c)^2\mu_1,$$

$$P = \frac{\mu_1(c/\omega)^2}{\epsilon^2(\mu_1^2 - k_1^2)}\left[\left(\frac{\pi}{a}\right)^2 + K^2\right]^2 + \left(\frac{\omega}{c}\right)^2\mu_1 - \frac{2}{\epsilon}\left[\left(\frac{\pi}{a}\right)^2 + K^2\right].$$

### B. Numerical Example

A numerical example with convenient parameters,

$$(\omega^2/c^2) = 2\pi^2/a^2,$$

$$\epsilon = 10, \qquad \mu_1 = 1.35,$$

$$\mu(\text{eff}) = (\mu_1^2 - k_1^2)/\mu_1 = 1.3, \ (k_1 \approx -0.26),$$

$$d = \tfrac{1}{12}a, \qquad b = \tfrac{1}{2}a,$$

is presented in this section. With this choice of the parameters, only the dominant mode propagates in the free space portion of the waveguide. Also, a half guide wavelength in the ferrite is $(32)^{-\frac{1}{2}}a$, which exceeds $2d = \frac{1}{6}a$.

Intuitively, the parameter $K$ might be expected to yield the closest bounds for some value of $\mu_t$ [$K^2 = \mu_t\epsilon\omega^2/c^2 - (\pi/a)^2$] between $\mu_1 - k_1 = 1.61$ and $\mu_1 + k_1 = 1.09$, corresponding to the relative right circular polarized and left circular polarized permeability, respectively. Different choices of $K$ corresponding to $\mu_t$ between 1.61 and 1.09 were tried in obtaining the best bounds on $\eta_o$ and $\eta_e$. (Since we have a variational bound rather than simply a variational principle, it is not necessary to find the value $\mu_t$ which gives a stationary estimate of the phase shift.) It is more conventional to present the data in terms of the elements of the equivalent $\pi$ circuit shown in Fig. 3. The susceptances $B_1$ and $B_2$ are relative to the characteristic admittance of the waveguide, and are related to the phase shifts $\eta_o$ and $\eta_e$ by

$$B_1 = \tan\eta_o, \qquad B_2 = \tfrac{1}{2}(\cot\eta_e + \tan\eta_o). \quad (17)$$

The best upper and lower bounds on $\eta_e$ and $\eta_o$, and on $B_1$ and $B_2$ are listed in Table I.

It would be possible to obtain closer bounds by introducing a more sophisticated trial function (containing higher-order modes) with additional parameters. As opposed to the situation that obtains for the usual variational principle, the inclusion of additional parameters into the trial function will always give better results, but this will increase the calculational effort.

While the above results are presumably interesting, the question remains as to how wide a class of problems can fruitfully be attacked by the above approach. This point will be considered in the discussion.

## 6. THE VARIATIONAL BOUND

As noted above, we will present the details of the connection between the waveguide and quantum mechanical scattering formalisms[7-10] only for $k = 0$. To go to this limit, it is simplest to set $\theta = \frac{1}{2}\pi$. Furthermore, since the phase shift has the same energy dependence for the even case for the waveguide problem as for zero-angular-momentum quantum mechanical scattering, we will consider the even case. Thus, it can be readily shown that $\tan \eta_e$ is proportional to $k$ for $k$ sufficiently small, and we therefore define the scattering length $A_e$ and the trial scattering length $A_{et}$ by

$$A_e \equiv \lim_{k\to 0}\left(-\frac{\tan \eta_e}{k}\right), \qquad A_{et} \equiv \lim_{k\to 0}\left(-\frac{\tan \eta_{et}}{k}\right).$$

We of course also have $\omega^2/c^2 \to \pi^2/a^2$ as $k \to 0$. The scattering function $\mathbf{u}_e$ then has the asymptotic behavior

$$\mathbf{u}_e \to -\left(\frac{2}{ab}\right)^{\frac{1}{2}}\left[\mathbf{a}_x \sin\left(\frac{\pi x}{a}\right) - \mathbf{a}_z \frac{\pi}{a}\left(\cos\frac{\pi x}{a}\right)(z - A_e)\right]$$

as $z \to \infty$. $\mathbf{u}_e$ also satisfies certain continuity conditions, as well as the condition $(\nabla \times \mathbf{u}_e)_{\tan g} = 0$ on conducting surface. Similar continuity and boundary conditions will be imposed upon $\mathbf{u}_{et}$, and the difference function $\mathbf{w}_e$ will therefore satisfy the same continuity conditions as well as the boundary

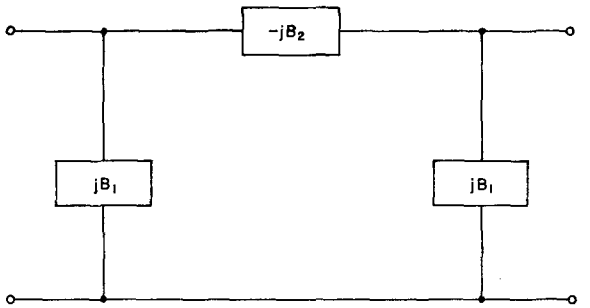


FIG. 3. Equivalent $\pi$ network for describing the far-field effects of scattering by symmetric ferrite obstacles in waveguide.

conditions

$$\mathbf{w}_e \to \left(\frac{2}{ab}\right)^{\frac{1}{2}}(A_e - A_{et})\frac{\pi}{a}\cos\left(\frac{\pi x}{a}\right)\mathbf{a}_z, \qquad z \to \infty$$

$$(\nabla \times \mathbf{w}_e)_{\tan g} = 0 \text{ on conducting surface.}$$

Equation (7) then reduces to

$$\frac{\pi^2}{a^2}A_e = \frac{\pi^2}{a^2}A_{et} + \frac{1}{2}\int \mathbf{u}_{et}\cdot\left(\mathfrak{K} - \frac{\pi^2}{a^2}\right)\mathbf{u}_{et}\, d\tau - \frac{1}{2}\int \mathbf{w}_e\cdot\left(\mathfrak{K} - \frac{\pi^2}{a^2}\right)\mathbf{w}_e\, d\tau, \qquad (18)$$

where $\mathfrak{K}$ is given by Eq. (3) but with $\omega^2/c^2$ set equal to $\pi^2/a^2$. In making the connection with the quantum problem, $x$ and $y$ are to be thought of as the "target" coordinates while $z$ is to be thought of as the coordinate of the "incident particle". The "target Hamiltonian", that part of $\mathfrak{K}$ which is independent of the coordinates of the incident particle, is then given by

$$\mathfrak{K}_T = -\partial^2/\partial x^2 - \partial^2/\partial y^2,$$

and the incident kinetic energy operator is given by

$$T(\text{incident}) = -\partial^2/\partial z^2.$$

The operator that corresponds to $\mathfrak{K} - \pi^2/a^2$ in the quantum mechanical case is $\mathfrak{K} - E_{BT}$, where $\mathfrak{K}$ is here the total Hamiltonian and $E_{BT}$, the binding energy of the target, is the lowest energy eigenvalue of the target Hamiltonian, $\mathfrak{K}_T$. The entire procedure in the quantum mechanical case for obtaining a bound on the analog of the last term of

TABLE I. The best upper and lower bounds on the even and odd phase shifts, and on the susceptance of the equivalent $\pi$ network.

| | $\eta_o$ | $\eta_e$ | $B_1$ | $B_2$ |
|---|---|---|---|---|
| Upper bound | ($\mu_t = 1.3$) 71°29′ | ($\mu_t = 1.35$) 34°08′ | 2.986 | 2.267 |
| Lower bound | ($\mu_t = 1.3$) 71°21′ | ($\mu_t = 1.35$) 32°52′ | 2.963 | 2.219 |

Eq. (18), the error term, then centers about the fact that the continuous spectrum of $\mathcal{H} - E_{BT}$ for the class of functions allowed is bounded from below by zero. This is a consequence of the fact that for the class of functions allowed both $\mathcal{H}_T - E_{BT}$ and $T$ (incident) are nonnegative operators, and that the interaction of the incident particle with the target does not affect the end points of the continuous spectrum though it may introduce discrete eigenvalues. The argument for obtaining a bound in the ferrite case is identical in form. The only point that warrants discussion is the proof that $\mathcal{H} - \pi^2/a^2$ is a nonnegative operator with respect to the functions $\mathbf{w}_e$. The proof is relagated to Appendix B.

We are therefore in a position to take over the zero-incident-energy quantum mechanical variational bound in its entirety. In particular, if $\mathcal{H} - \pi^2/a^2$, has no negative eigenvalues, or in waveguide language if $\mathcal{H} - \pi^2/a^2$ does not have any evanescent modes, we obtain our variational bound by simply dropping the error term, while if it does have negative eigenvalues we have to go through the usual "subtraction" procedures.[10] For the ferrite problem it would generally be difficult to determine from first principles whether or not there are in fact negative eigenvalues, and one would simply add terms until one felt sure that no further jumps would occur. For a dielectric obstacle, on the other hand, it might often well be possible to prove from first principles that negative eigenvalues do not exist by showing that the generalization of the Bargmann–Schwinger[15] necessary conditions for the existence of negative eigenvalues is not satisfied. We might note incidentally, as observed by Dr. B. Lippmann, that there are no negative eigenvalues for a plasma since $\epsilon < 1$ represents a repulsive potential.

## 7. DISCUSSION

We will consider briefly the class of problems to which the two variational bounds can be applied.

To begin with, the requirement that $\mu$ and $\epsilon$ be even functions of $z$ can be dropped for each of the two variational bounds. As noted above the quadratic variational bound has in fact been applied to unsymmetric dielectric obstacles[6,16] and an extension of the present formalism to unsymmetric gyromagnetic obstacles is in progress. We also remark that there are geometrical waveguides other than rectangular waveguides which can be attacked.

The requirement that the thickness of the obstacle be small will be difficult to circumvent in the quadratic variational bound because of the necessity of obtaining the eigenvalues $\alpha_\theta$ and $\beta_\theta$ or bounds on these eigenvalues. There is no difficulty in utilizing the linear variational bound for thick obstacles and thereby in studying the effects of surface scattering as distinct from volume scattering.

As noted earlier, the linear variational bound can be extended quite readily to multichannel scattering.[17] As apart from the formal aspect of the problem, however, it remains to be seen how practical this approach is, for it must be recalled that the method requires the (numerical) solution of a scattering problem, albeit a problem much simpler than the one under consideration.

Finally, we make a few brief comments on problems in which geometric as well as anisotropic diffraction is significant. Consider the case of an obstacle $O_1$, of rectangular parallelopiped shape, symmetrically placed in the waveguide and characterized by $\epsilon_1(x, z)$ and $\mu_1(x, z)$. The difficulty is not one of principle but of practice; it is to find simple but realistic trial functions which satisfy the necessary boundary conditions that the tangential components of $\mathbf{H}_t$ and of $(1/\epsilon_1)\nabla \times \mathbf{H}_t$ be continuous at the interface. Since these boundary conditions are independent of $\mu_1$, it is variations of $\epsilon_1$ and not of $\mu_1$ which are difficult to handle. The case for which $\epsilon_1$ is 1 but $\mu_1$ is a constant other than 1 or a varying function could be readily attacked. A more interesting but more difficult problem is one for which $\epsilon_1$ is not everywhere equal to 1. One approach to the problem would be to choose a trial function for which $(\mathbf{H}_t)_{tang}$ is continuous at the interface, and for which $(\nabla \times \mathbf{H}_t)_{tang} = 0$ at the interface so that $(1/\epsilon)(\nabla \times \mathbf{H}_t)_{tang}$ is continuous at the interface in spite of the variation of $\epsilon$. Such a trial function would however be neither simple nor realistic and would lead to a poor (if rigorous) bound, and then only at the cost of great labor. For simple variations of $\epsilon$, it may be possible to choose more realistic trial functions. A more general approach but one of limited accuracy would be to consider an obstacle $O_2$, characterized by $\epsilon_2(x, z)$ and $\mu_2(x, z)$, which is also a rectangular parallelopiped but which touches the four walls of the waveguide and which contains $O_1$. Choosing $\epsilon_2$ to be a constant, equal to the maximum value of $\epsilon_1(x, z)$, and choosing

[15] V. Bargmann, Proc. Natl. Acad. Sci. U. S. **38**, 961 (1952); J. Schwinger, *ibid.* **47**, 122 (1961). See also Ref. 9, Sec. IIA5.

[16] K. Kalikstein and B. Schuldiner, IEEE Trans. Microwave Theory and Tech. **12**, 252 (1964).

[17] The quantum mechanical case is considered by Y. Hahn, T. F. O'Malley, and L. Spruch, Phys. Rev. **134**, B397 (1964).

$\mu_2(x, z) \geq \mu_1(x, z)$, the monotonicity theorem tells us that $\eta_2 > \eta_1$. The particular choice of $\epsilon_2$ generates a problem amenable to a variational bound approach, and this upper (variational) bound on $\eta_2$ will serve as an upper bound on $\eta_1$; if the difference between $\epsilon_2$ and $\epsilon_1(x, z)$ is large, the bound will of course be crude.


### ACKNOWLEDGMENT

We would like to thank Dr. Ralph Bartram and Mr. Junn Troy for a helpful discussion. One of us (L. S.) was at University College, London, during the latter stages of the current research, and would like to thank Professor Massey and Professor Seaton for their various courtesies during that time.


### APPENDIX A: A MONOTONICITY THEOREM

In Sec. 4B, we showed that if $W_2(x, z) \geq W_1(x, z)$, that is, if for all $x$ and $z$ each of the eigenvalues of $W_2(x, z) - W_1(x, z)$ is nonnegative, then $\eta_2 > \eta_1$. $W_i(x, z)$ is given in terms of $\epsilon_i(x, z)$ and $\mu_i(x, z)$ by Eq. (3c), and we would here like to re-express the monotonicity theorem in terms of the $\epsilon_i(x, z)$ and $\mu_i(x, z)$, where $i = 1$ and 2. Assuming small differences, a restriction which can always ultimately be dropped, the monotonicity theorem goes through if we can prove that

$$J \equiv \int \mathbf{u}_0^{\star} \cdot [W_2(x, z) - W_1(x, z)]\mathbf{u}_0 \, d\tau \geq 0.$$

It is convenient to revert to the operator $\mathfrak{L}$, and use Eqs. (2) and (3a) to write

$$W_2 - W_1 = \mathfrak{L}_2 - \mathfrak{L}_1 = -\nabla \times \frac{1}{\epsilon_2} \nabla \times + \nabla \times \frac{1}{\epsilon_1} \nabla \times + (\mu_2 - \mu_1) \frac{\omega^2}{c^2}.$$

With some trivial manipulation and the use of the fact that certain surface terms vanish, we can write $J$ as

$$J = \int \left(\frac{1}{\epsilon_1} - \frac{1}{\epsilon_2}\right) |\nabla \times \mathbf{u}_0|^2 \, d\tau + \frac{\omega^2}{c^2} \int (\mu_2 - \mu_1) \, |\mathbf{u}_0|^2 \, d\tau.$$

It follows that $J \geq 0$, and therefore that $\eta_2 \geq \eta_1$, if for all $x$ and $z$, $\epsilon_2 \geq \epsilon_1$ and $\mu_2 \geq \mu_1$.

### APPENDIX B: NONNEGATIVENESS OF $\mathfrak{K}_T - \pi^2/a^2$ AND OF $T$ (INCIDENT)

We here prove that $\mathfrak{K}_T - \pi^2/a^2$ and $T$ (incident) are nonnegative with respect to the class of functions which satisfy the boundary conditions imposed upon $\mathbf{w}_e$. We begin by noting that in the absence of any obstacle, the free ($F$) even magnetic field is given by

$$\mathbf{H}_e^F(x, z) = -\mathbf{a}_x \sin \frac{\pi x}{a} + \mathbf{a}_z \left(\frac{\pi z}{a}\right) \cos \left(\frac{\pi x}{a}\right).$$

(The subscript e, it will be recalled, refers to the fact that $\mathbf{H}_e^F$ is even under $z \to -z$, for which $\mathbf{a}_z \to -\mathbf{a}_z$ of course.) The proof will depend upon a symmetry argument related to the $x$ coordinate, and it will be convenient to introduce the coordinates $X = x - \frac{1}{2}a$, $Y = y - \frac{1}{2}b$, and $Z = z$, whose origin is centrally (and therefore symmetrically) located in the waveguide. We then have

$$\mathbf{H}_e^F(X, Z) = -\mathbf{a}_X \cos \frac{\pi X}{a} - \mathbf{a}_Z \left(\frac{\pi Z}{a}\right) \sin \left(\frac{\pi X}{a}\right).$$

Denoting the transformation $X \to -X$, $Y \to Y$, $Z \to Z$ by $R$, it is clear since $\mathbf{a}_X \to -\mathbf{a}_X$ under $R$ that $\mathbf{H}_e^F$ is odd under $R$. Since $\mu(X, Z)$ and $\epsilon(X, Z)$ are by assumption even functions of $X$, and since

$$\nabla = \mathbf{a}_X \frac{\partial}{\partial X} + \mathbf{a}_Z \frac{\partial}{\partial Z}$$

is even under $R$, if follows that $\nabla$ and therefore $V(X, Z)$, as given by Eq. (3c), are even under $R$. To first order in $V(X, Z)$, only states of $\mathbf{H}_e(X, Z)$ that are odd under $R$ are therefore excited by the presence of the dielectric. Since $V$ cannot couple states that are even and odd, respectively, under $R$, the oddness of $\mathbf{H}_e$ is maintained to all orders, and the most general form that $\mathbf{H}_e$ can assume is therefore

$$\mathbf{H}_e(X, Z) = \sum_{n=0}^{\infty} \mathbf{S}(f_n, g_n), \tag{A1}$$

where

$$\mathbf{S}(f_n, g_n) \equiv \mathbf{a}_X \cos \frac{n\pi X}{a} f_n(Z) + \mathbf{a}_Z \sin \frac{n\pi X}{a} g_n(Z); \tag{A2}$$

the ($Z$-dependent) coefficients of the terms $\mathbf{a}_X \sin (n\pi X/a)$ and of $\mathbf{a}_Z \cos (n\pi X/a)$ that could appear in the Fourier expansion of an arbitrary function in the interval $-\frac{1}{2}a < X \leq \frac{1}{2}a$ must vanish for $\mathbf{H}_e$ since they are even under $R$. Restrictions can also be imposed upon the $f_n(Z)$. Since the normal component of $\mathbf{B}$ and therefore of $\mathbf{H}$ must vanish at a metallic surface, we must have $H_{eX}(\pm\frac{1}{2}a, Z) = 0$, which requires that $f_n(Z) = 0$ for $n$ even. The particular point of interest is that there is therefore no $n = 0$ term in Eq. (A1). We also note that

$$\mathbf{H}_e(X, Z) \to -\mathbf{a}_X \cos\frac{\pi X}{a} - \mathbf{a}_Z \frac{\pi(Z - A_e)}{a} \sin\frac{\pi X}{a}$$

as $Z \to \infty$. This follows from Eq. (4) on setting $\theta = \frac{1}{2}\pi$, letting $k \to 0$, and using the definition of $A_e$. We can therefore write

$$\mathbf{H}_e(X, Z) = \sum_{n=1}^{\infty} \mathbf{S}(f_n, g_n),$$

where, as $Z \to \infty$,

$$f_1(Z) \to -1,$$

$$g_1(Z) \to -\pi(Z - A_e)/a,$$

$$f_n(Z) \to 0, \qquad g_n(Z) \to 0, \qquad n > 1.$$

(Some of these coefficients vanish identically but that is of no concern here.) If $\mathbf{H}_{et}(X, Z)$ is chosen to be of the same form as $\mathbf{H}_e$, it follows that

$$\mathbf{w}_e(X, Z) = \sum_{n=1}^{\infty} \mathbf{S}(F_n, G_n)$$

where, as $Z \to \infty$,

$$F_n(Z) \to 0, \quad \text{all} \quad n,$$

$$G_1(Z) \to \frac{\pi(A_{et} - A_e)}{a},$$

$$G_n(Z) \to 0, \qquad n > 1.$$

It is then trivial to show that

$$\mathfrak{K}_T - \frac{\pi^2}{a^2} = \left(-\frac{\partial^2}{\partial X^2} - \frac{\pi^2}{a^2}\right)$$

and that

$$T(\text{incident}) = (-\partial^2/\partial Z^2)$$

are separately nonnegative with respect to functions $\mathbf{w}_e$ of the above form, which proves the theorem. Since

$$\left(-\frac{\partial^2}{\partial X^2} - \frac{\partial^2}{\partial Z^2}\right)\mathbf{H}_e^F = \frac{\pi^2}{a^2}\mathbf{H}_e^F,$$

the above theorem can be roughly interpreted as the statement that of all fields that satisfy the same sort of boundary conditions as $\mathbf{H}_e^F$, it is $\mathbf{H}_e^F$ which has the least "kinetic energy". This suggests, incidentally, that there is probably a completely trivial proof of the theorem.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# Kinetic Theory of a Weakly Coupled Gas

C. H. Su
*Massachusetts Institute of Technology*
*Cambridge, Massachusetts*
(Received 21 April 1964)

Using the multiple-time-scale method on the BBGKY hierarchy, the weak coupling expansion is carried out to higher orders. It is found that there are two local breakdowns of the expansion. One occurs at small relative velocities between particles. The correct asymptotic representation for the small relative velocity region is given. The second breakdown occurs for particles having a large separation at $t$ with their relative velocity oriented in such a way that they were in collision at $t = 0$. Such a breakdown indicates that in contrast to the Bogoliubov functional assumption, the higher-order correlation functions should vary on the kinetic time scale in their own right. A sufficient condition on the smoothness of the initial correlation functions is given such that one obtains the Fokker–Planck equation at the lowest-order approximation in the expansion. The connection between irreversibility and the requirement of nonsecularity in the multiple-time-scale formulation is also indicated.

## 1. INTRODUCTION

IN recent years the problem of the approach to thermal equilibrium has been investigated from the dynamical point of view.[1] The analysis is deductive in nature. Starting from the equation of motion of an assembly of particles, one tries to derive the kinetic equation (an autonomous or Markoffian equation for the one-particle distribution function) as purely deductively as possible. The objective of such analysis is twofold:

(1) To find the missing link between the reversible dynamical equation and the irreversible kinetic equation.

(2) To find the correction terms to the known collision integrals by use of a systematic expansion procedure. Also, to put the Boltzmann equation, etc., on as rigorous a basis as possible.

The methods of approach in the literature are very diversified. However, they have in common the following two features:

(1) The limit of large system. The system under consideration is asymptotic in the sense that the total number of particles $N \to \infty$ and the total volume $V \to \infty$, while the average density $n = N/V$ remains constant. Such a limit, usually called the bulk limit, is introduced so that surface effects of the system can be ignored and the Poincaré cycle becomes infinitely long (a necessary condition for the irreversibility, if we purport to describe the behavior of a single system rather than merely the average behavior of an ensemble).

(2) A clear distinction in the different time scales involved in the kinetic problem. It is a well-accepted point of view that the kinetic states (states governed by the kinetic equations) are the long-time asymptotic states on the (short) time scale of the order of the interaction time of the particles.

The kinetic equations which have been derived so far are as follows:

(1) Low-density system: the leading term in the asymptotic expansion gives the Boltzmann equation.[1(a),2]

(2) Weak-interaction system: the leading term in the asymptotic expansion gives the Fokker–Planck equation.[1(b),3]

(3) Plasma case: the leading term gives the Lenard–Balescu equation.[4,5]

Although the leading behaviors in the expansions of the above three cases were obtained through various methods by various investigators, the nature of the expansions carried to higher orders has not been analyzed properly in the literature. In this

[1] There are an enormous number of articles in the literature; we give only a few as follows (by no means inclusively): (a) N. N. Bogoliubov, J. Phys. (USSR) 10, 256(1946). [English transl.: E. K. Gora in *Studies in Statistical Mechanics*, edited by J. DeBoer and G. D. Uhlenbeck (North Holland Publishing Co., Amsterdam, 1962), Vol. 1]. (b) I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962). (c) Review article by S. Rice and H. Frisch, Ann. Rev. Phys. Chem. 11, 187 (1960), contained the work by Kirkwood and his co-workers. (d) M. Green, Natl. Bur. Std. Rept. No. 3327 (1955). (e) G. E. Uhlenbeck, lecture notes, "The Statistical Mechanics of Non-equilibrium Phenomena," Higgins lecture, Princeton University. (f) E. G. D. Cohen, *Fundamental Problem in Statistical Mechanics* (North-Holland Publishing Co., Amsterdam, 1962).

[2] S. T. Choh and G. E. Uhlenbeck, *The Kinetic Theory of Dense Gases*, (University of Michigan, Ann Arbor, Michigan, 1958).

[3] E. A. Frieman, J. Math. Phys. 4, 410 (1963).

[4] R. Balescu, Phys. Fluids 3, 52 (1960).

[5] A. Lenard, Ann. Phys. 10, 390 (1960).

paper, we shall carry out the series solution of the weak coupling system to higher orders. In addition to obtaining the higher correction terms, we shall investigate in detail the asymptotic nature of the series solution, i.e., the uniformity of each term in the series with respect to both space and velocity variables as well as the time variable. The choice of the weak coupling system is based on the relatively simple mathematical structure of the solution for this system. Explicit forms for the solutions to all orders can be obtained in principle and the uniformity analysis of these solutions can be carried out without too much complication.

We shall adopt in the present paper the multiple-time-scale method in the derivation of the Fokker–Planck equation and its correction equations from the BBGKY hierarchy. This method has its origin in the solution for a nearly periodic system in the field of nonlinear mechanics.[6,7] The use of the method in connection with the kinetic equation (nonperiodic) was first developed by Frieman[3] and Sandri.[8] In this method, the slow process in the problem (the evolution which is governed by the kinetic equation) is assumed to be describable by an independent "slow time variable," in contrast to the "fast time variable" which is of the order of the collision time. In the formal solution of the problem as far as the time dependence of the system is concerned, one formally replaces the single time variable by a "space of times," i.e., one replaces $f(t)$ by

$$f(t_0, t_1, t_2, \cdots), \tag{1}$$

where $f$ is any function describing the system and $t_0, t_1, t_2, \cdots$ are treated as independent variables as far as the formal solution is concerned. The rates of the various times are characterized by the simple differential equations

$$dt_i/dt = \beta_i(\epsilon) \qquad i = 0, 1, 2, \cdots \tag{2}$$

where $\epsilon$ is the small expansion parameter in the problem. The $\beta_i$ are appropriately chosen functions of $\epsilon$ satisfying $\beta_{i+1}(\epsilon) \ll \beta_i(\epsilon)$. For the weak coupling system, $\epsilon$ is taken to be the ratio of a typical potential to the average kinetic energy, which is assumed to be small. In the kinetic problem, it seems appropriate to assume that the $\beta_i$ are just simple powers in $\epsilon$, i.e.,

$$\beta_i(\epsilon) = \epsilon^i. \tag{3}$$

It is seen that a change of $f$ specified originally in the $t$ direction is now replaced by a change in the direction given by the characteristics (2). The variation along such a characteristic direction shall be chosen in such a way that the series representation of $f$ will be valid uniformly over the "space of times." In other words, the variation of $f$ on a slow time scale (say $t_i$) is determined by the condition that the representation of $f$ behave well asymptotically as a function of all faster times $t_0, t_1, \cdots t_{i-1}$, in particular as they become infinite. In practice, this amounts to requiring that no secular terms appear in the representation—terms containing positive powers of a time variable. As we shall see presently, in the kinetic problem this condition of nonsecularity on the fast time scale is just the desired kinetic equation.

We have so far discussed the extension of the time variable only in connection with the governing equation (variation of $f$). If now we pose an initial value problem for $f$, we shall need initial data as functions of space and velocity at $t = 0$ for the original unextended problem. We shall refer to such initial data as arbitrary initial data in our later analysis. It is seen that the arbitrary initial data determine $f$ only on one of the characteristic lines in Eq. (2). Without loss of generality, we may assume this line passes through the origin of the time axes, $t_0, t_1, \cdots t_i$. In order to determine $f$ completely over a cylindrical region with the $t_0$ axis as its axis, we must have the initial data prescribed on a hypersurface of the "space of times," say, with the normal of the hypersurface lying in the direction of the $t_0$ axis. The initial data on the hypersurface are unknown when the physical problem is first posed. (Actually only the derivatives of $f$ in the various directions on the hypersurface are required, since we know the value of $f$ at the origin.) However, they are not arbitrary in the sense that the original initial data are, since they are determined by requiring that the resulting solution for $f$ be uniformly valid over a long time. To avoid confusion with the original arbitrary initial data we shall call such initial data on the hypersurface in the "times space" the "unknown initial functions."

The foregoing discussion, of course, deals with only a formal procedure for the extension of the time variable. It is still an open question in general what is an appropriate definition for "nonsecularity" such that the extended problem becomes well-defined. In the Appendix this question is discussed for several simple examples. In one example, a definition of nonsecularity is satisfied by the use

[6] N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic Theory of Nonlinear Oscillations* (Gordon and Breach Science Publishers, New York, 1961).
[7] M. D. Kruskal, J. Math. Phys. **3**, 806 (1962).
[8] G. Sandri, Ann. Phys. (N. Y.) **24**, 332, 380 (1963).

of the unknown initial functions, while in another the problem becomes underdetermined because of the extra freedom introduced by the unknown initial functions. No such underdeterminism appears in the weak coupling expansion to the orders we shall treat in the present work.

In this paper, we shall not include the unknown initial functions for the weak coupling system.[9] It is found for this particular expansion that there are two local nonuniformities:

(1) When the relative velocity between particles becomes too small. We recall that in the weak coupling approximation the interaction force between particles is considered only as a small perturbation. When the relative velocity between particles becomes small, their interaction time increases and the change of velocity during the interaction may become even greater than their initial velocity of approach. Under such circumstances, the interaction between the particles cannot be treated as a small perturbation. In our formal weak coupling expansion, the inappropriate treatment of the force term causes a local singularity of the higher order correlation functions at zero relative velocity. This local singularity first appears in the second-order two-particle correlation function and it becomes worse and worse as we go to the higher orders. A boundary layer analysis is applied to a small local region in relative velocity space (with a thickness of the order $\epsilon^{\frac{1}{2}}$). The dynamical behavior in this region is found to be short-range interaction with a mixing time of the order of $\epsilon^{-\frac{1}{2}}$ in contrast to unity for the particles outside the layer. The mixing process for particles of small relative velocity is essential in the study of the following two important problems:

(a) Determining the transient behavior of the system in establishing the kinetic state (state satisfying the kinetic equation).

(b) Obtaining a smoothness condition on the arbitrary initial correlation functions such as to ensure the establishment of the kinetic state.

Both these processes are controlled essentially by the phase mixing of the particles with small relative velocity. The mixing process provided by the phase mixing integral in the relative velocity space becomes more and more ineffective as the relative velocity of the particles approaches zero. However, when the relative velocity between two particles becomes small, the transfer of momentum due to their interaction, even though weak, can no longer be neglected, because the orbits now deviate appreciably from straight lines. Such exchange of momentum provides enough phase mixing for the system. The results, after proper treatment in the region of small relative velocity, of the above two phenomena are

(a) The transient is an exponential decay on the time scale $\epsilon^{-\frac{1}{2}}$.

(b) The kinetic equation will be independent of the arbitrary initial conditions on the correlation functions as long as the latter are not exponentially large at the small relative velocity.

The smoothness condition just given is merely a sufficient one. It is far from necessary. Since the initial conditions are functions of several vectors, a necessary condition on the smoothness is very difficult to obtain.

(2) The second local breakdown of the ordinary weak coupling expansion occurs for particles of large separation at time $t$ with such relative velocity that they were in collision at $t = 0$. Since this singularity of the pair correlation functions is very localized in the phase space, its effect on the one-particle kinetic equation comes out only at higher orders of approximation. Furthermore, if the pair potential has a finite range and is small compared with the mean free path, we shall show that the effect of the second singularity on the one-particle equation can be neglected for all orders of approximation. It will be established, however, that the second kind of local singularity has its origin in that the lowest order approximation to the pair correlation function has an infinite range in this special local region. It is recalled that the lowest-order pair-correlation function is a binary interaction approximation. Now if the range of such pair correlation is infinite in a certain region, the presence of other particles cannot be ignored. This indicates that there will be a different asymptotic representation of the hierarchy equations in this special local region.

In our extension of the time variable to a space of times, there arises the delicate question whether one needs time so slow that it has no apparent physical significance. In the weak-coupling expansion, for example, the fast time $t_0 = t$ is characterized by the collision time and the physically meaningful slow time $t_2 = \epsilon^2 t$ by the mean free time. The question is whether we should expect to find any dependence on $t_3 = \epsilon^3 t$. Under the condition of nonsecularity on the $t_2$ scale, it is found

[9] For an analysis including the unknown functions, see C. H. Su, "The Kinetic Equation for a Weak-Coupling Gas," Ph.D. thesis, Princeton University, 1964.

that the time scales do terminate at the $t_2$ scale. This result can also be demonstrated in the second example in the Appendix.

Finally, we come to the question of irreversibility. There are various explanations of irreversibility in the literature. In what follows we shall restrict ourselves to pointing out the relevant properties of the hierarchy itself and the crucial operations which we shall apply to it. Such properties and operations which give rise to the irreversibility of the resulting kinetic equation are:

(1) Large-system limit (operation): We need only mention that after this limiting process the equations are still time reversible, though their solutions, in general, exhibit irreversible behavior (infinite Poincaré cycle).

(2) Phase mixing (property of the system): This phenomenon washes away the effect on the kinetic equation of the arbitrary initial data for the correlation functions. However, it must be noted that such mixing does not by itself single out a preferred direction of time. In fact, the phase mixing seems to correspond to the Ehrenfests' coarse-graining which, by itself, is not sufficient to make the system evolve in a preferred direction of time.[10]

(3) Distinction between fast and slow processes (property of the system) and requirement of nonsecularity in the (fast time) limit $t_0 \to +\infty$ (operation): It is this last operation which picks out a time-irreversible subclass of solutions of the hierarchy; and it is found that the individual solutions of this subclass not only exhibit irreversible behavior but satisfy the time-asymmetric kinetic equation. Thus, within the framework of the multiple-time-scale method, it is clear how and where the irreversibility is introduced. However, we believe that the more philosophical question of "why" still remains.

In the next section, we shall give a brief discussion of the physical parameters in the problem. In Sec. III we shall solve the hierarchy equations on the fast time scale under the weak interaction approximation and there the occurrence of the two local secularities will be pointed out. In Sec. IV the analysis for small relative velocity will be given. In Sec. V we shall give the higher-order kinetic equations. Finally, the smoothness condition on the arbitrary initial data and the question of irreversibility will be discussed.

[10] M. Kac, *Probability and Related Topics in Physical Sciences*, (Interscience Publishers, Inc., New York, 1957).

## 2. PHYSICAL PARAMETERS

The characteristic physical quantities which are associated with any kinetic problem of a homogeneous, one-species system are

$\langle\phi\rangle$, the typical strength of the interparticle potential;

$v_{av}$, average speed of the particles;

$r_0$, effective range of the interparticle potential, in the present analysis, this is assumed to be finite;

$n$, average spatial density of the particles in the system;

$t$, characteristic time;

$m$, mass of the particle.

Out of these six quantities can be formed three dimensionless parameters. They are

$$\Phi = \langle\phi\rangle/mv_{av}^2, \qquad D = nr_0^3, \qquad T = v_{av}t/r_0.$$

The first one, $\Phi$, measures the effective strength of the interaction between particles. The second one, $D$, characterizes the density of the system (number of particles within the interaction sphere). The last one, $T$, is a time parameter. In applying the multiple-time-scale method, the fastest time scale in the problem is taken to be the characteristic time. One expects the change on the slower time scales to be picked up by the method automatically. In the kinetic problem, the fastest time scale is the time for a particle to cross the range of force. Using this time as the characteristic time, we note that the time parameter $T$ above is of order one. If now we use the characteristic quantities to nondimensionalize the hierarchy,[11] we find

(a) Every time derivative term will be associated with a time factor $T$, which is of order unity.

(b) All convective terms, such as $(\mathbf{v}_1 - \mathbf{v}_2)\cdot\partial/\partial\mathbf{x}$ will have coefficient unity.

(c) All force terms (terms involving $\phi$) will have a potential factor $\Phi$.

(d) All the integral terms (terms on the right side of the equations) will have a potential factor $\Phi$ and a density factor $D$.

The nature of the weak coupling expansion is best described by the order of the magnitude of $\Phi$ and $D$

$$\Phi = \langle\phi\rangle/mv_{av}^2 = \epsilon \ll 1, \qquad D = nr_0^3 \sim 1. \tag{4}$$

The interaction is weak. We shall expand in the small parameter $\epsilon$. With the choice (4), it is simple to determine the orders of the various terms in the

[11] The first three members of the BBGKY hierarchy are given in the next section.

hierarchy equations. In fact, it is clear that all terms involving the potential $\phi$ are of order $\epsilon$, while all other terms are of order unity.

The hierarchy equations pose an initial-value problem. However, as usual in statistical mechanics, we do not have detailed initial data for the problem. In the sixth section we shall show that the system of hierarchy equations has the important property that the effect of any reasonably smooth initial data of the correlation functions on the one-particle distribution function will be mixed and washed away in a time of order of the duration of a collision. The influence of the initial data is therefore limited to this "initial time layer" only. As far as the kinetic behavior of the system is concerned, the initial correlations are unimportant. In this and the next three sections, we shall restrict our considerations only to correlations which are created during the initial time layer.

We have chosen the characteristic time of the problem to be the time for a particle to cross the range of force of another particle. We further mentioned that the change on the slower time scales was expected to be picked up automatically by the method of multiple time scales. It is interesting to see what the distinct time scales are in the kinetic theory of weak coupling systems before we go into a detailed study of the hierarchy equations. We take the fast time scale as the characteristic time scale, i.e.,

$$\tau_f = r_0/v_{\text{av}},$$

and the slow time scale as the time between collision, i.e., the effective mean free time. The change in momentum for a thermal particle crossing the range $r_0$ of the potential of a given particle is, in order of magnitude, the product of the force and the transient time, or

$$(\phi/r_0)\tau_f = \phi/v_{\text{av}}.$$

The change of angle in velocity is the relative (logarithmic) change in the momentum, or

$$(\phi/v_{\text{av}})/mv_{\text{av}} \sim \epsilon.$$

However, the change of angle averaged over collision orientations is zero; therefore it is the square of the change in angle that sets the scale for diffusion in velocity space. Thus the slow time is given by

$$\tau_s \sim \tau_f(1/\epsilon^2).$$

In the hierarchy equations, $t$ is the time for fast variation, and so we expect that the kinetic evolution will take place on the $\epsilon^2 t$ time scale.

## 3. WEAK-COUPLING EXPANSION

### 3.1 The Lowest-Order Approximation

The first three members of the hierarchy under the assumption of weak interaction are given as follows:

$$\frac{\partial f}{\partial t} = \epsilon \frac{n}{m} \int d\mathbf{x}\, d\mathbf{v}_2 \frac{d\phi}{d\mathbf{x}} \cdot \frac{\partial g}{\partial \mathbf{v}_1} \tag{5}$$

$$\frac{\partial g}{\partial t} + \mathbf{v}_{12} \cdot \frac{\partial g}{\partial \mathbf{x}} - \frac{\epsilon}{m} \frac{d\phi}{d\mathbf{x}} \cdot \left( \frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2} \right) [g(12) + f(1)f(2)]$$
$$= \epsilon \frac{n}{m} \int d\xi\, d\mathbf{v}_3 \frac{d\phi}{d\xi} \cdot \left\{ \frac{\partial}{\partial \mathbf{v}_1} [f(1, t)g(\xi - \mathbf{x}, 23, t) + h(\mathbf{x}, \xi, 123, t)] - \frac{\partial}{\partial \mathbf{v}_2} [f(2, t)g(\mathbf{x} - \xi, 13, t) + h(\mathbf{x}, \mathbf{x} - \xi, 123, t)] \right\}, \tag{6}$$

$$\left( \frac{\partial}{\partial t} + \mathbf{v}_{12} \cdot \frac{\partial}{\partial \mathbf{x}} + \mathbf{v}_{13} \cdot \frac{\partial}{\partial \xi} \right) h(\mathbf{x}, \xi, 123, t)$$
$$- \frac{\epsilon}{m} \frac{d\phi}{d\mathbf{x}} \cdot \left( \frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2} \right) [h + f(1)g(23) + f(2)g(13)]$$
$$- \frac{\epsilon}{m} \frac{d\phi}{d\xi} \cdot \left( \frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_3} \right) [h + f(1)g(23) + f(3)g(12)]$$
$$- \frac{\epsilon}{m} \frac{\partial \phi(\xi - \mathbf{x})}{\partial \xi} \cdot \left( \frac{\partial}{\partial \mathbf{v}_2} - \frac{\partial}{\partial \mathbf{v}_3} \right)$$
$$\times [h + f(2)g(13) + f(3)g(12)] \tag{7}$$
$$= \text{Integral terms of order } \epsilon,$$

where

$$f = F_1,$$
$$g = F_2 - f(1)f(2),$$
$$h = F_3 - f(1)g(23) - f(2)g(13) - f(3)g(12) - f(1)f(2)f(3),$$
$$\mathbf{v}_{12} = \mathbf{v}_1 - \mathbf{v}_2, \qquad \mathbf{v}_{13} = \mathbf{v}_1 - \mathbf{v}_3,$$
$$\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2, \qquad \xi = \mathbf{x}_1 - \mathbf{x}_3.$$

We consider the following series solution:

$$f = f^{(0)} + \epsilon f^{(1)} + \epsilon^2 f^{(2)} + \cdots,$$
$$g = g^{(0)} + \epsilon g^{(1)} + \epsilon^2 g^{(2)} + \cdots,$$
$$h = h^{(0)} + \epsilon h^{(1)} + \epsilon^2 h^{(2)} + \cdots.$$

Within the framework of multiple-time-scale formulation, the time derivative is also formally expanded as a power series in $\epsilon$, i.e.,

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial t_0} + \epsilon \frac{\partial}{\partial t_1} + \epsilon^2 \frac{\partial}{\partial t_2} + \cdots.$$

Using these series in Eqs. (5)–(7), one obtains a set of equations for different order of approximation. The solution in the lowest significant order has been carried out by Frieman.[3] We shall give here only a brief account. With the assumption of zero initial correlations, it can be shown that

$$f^{(0)} = f^{(0)}(t_2, t_3, \cdots), \quad g^{(0)} = h^{(0)} = h^{(1)} = 0,$$

and

$$g^{(1)}(12, t_0 \cdots) = \frac{1}{m}\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right) f^{(0)} f^{(0)} \cdot \int_0^{t_0} d\tau \, \frac{\partial \phi}{\partial \mathbf{x}'} \, [\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}\tau]. \tag{8}$$

Passing to the limit $t_0 \to \infty$, we see that the variation of $g^{(1)}$ in time occurs only through $f^{(0)}(1)$ and $f^{(0)}(2)$, which can at most vary on the $t_2$ scale.

The first nontrivial one-particle equation is

$$\frac{\partial f^{(2)}}{\partial t_0} + \frac{\partial f^{(1)}}{\partial t_1} + \frac{\partial f^{(0)}}{\partial t_2} = \frac{n}{m} \int d\mathbf{x}\, d\mathbf{v}_2 \, \frac{\partial \phi}{\partial \mathbf{x}} \cdot \frac{\partial g^{(1)}}{\partial \mathbf{v}_1}.$$

By requiring that the solutions for $f^{(2)}$, $f^{(1)}$ are non-secular on the time scales $t_0$, $t_1$, respectively, one obtains the following three equations:

$$\frac{\partial f^{(0)}}{\partial t_2} = \frac{n}{m^2} \int d\mathbf{x}\, d\mathbf{v}_2 \, \frac{\partial \phi}{\partial \mathbf{x}} \cdot \frac{\partial}{\partial \mathbf{v}_1}\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right) f^{(0)} f^{(0)} \times \int_0^{\infty} d\tau \, \frac{\partial \phi}{\partial \mathbf{x}'} \, [\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}\tau], \tag{9}$$

$$\partial f^{(1)}/\partial t_1 = 0, \tag{10}$$

$$\frac{\partial f^{(2)}}{\partial t_0} = -\frac{n}{m^2} \int d\mathbf{x}\, d\mathbf{v}_2 \, \frac{\partial \phi}{\partial \mathbf{x}} \cdot \frac{\partial}{\partial \mathbf{v}_1}\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right) f^{(0)} f^{(0)} \times \int_{t_0}^{\infty} d\tau \, \frac{\partial \phi}{\partial \mathbf{x}'} \, [\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}\tau]. \tag{11}$$

The first equation can be reduced to the familiar form of the Fokker–Planck equation.[3] Equation (11) describes the transient behavior of the one-particle distribution towards the kinetic regime. It can be shown[9] that

$$f^{(2)}(t_0) \sim 1/t_0^2 \quad \text{as} \quad t_0 \to \infty. \tag{12}$$

It should be noted that this result rests very heavily on the behavior of $g^{(1)}$ at small $|\mathbf{v}_{12}|$. We shall see in the next section that the solution for $g^{(1)}$ as given above is not valid for $\mathbf{v}_{12} \to 0$. With the correct $g^{(1)}$ in the small relative velocity region, we shall show that the decay of $f^{(2)}$ in time is actually much faster than that given in Eq. (12).

### 3.2 Solution for $g^{(2)}(x, v_1, v_2, t_0)$

The equation for the second-order $g$ function is

$$\left(\frac{\partial}{\partial t_0} + \mathbf{v}_{12} \cdot \frac{\partial}{\partial \mathbf{x}}\right) g^{(2)} = \frac{1}{m} \frac{d\phi}{d\mathbf{x}} \cdot \left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right) \times [f^{(0)}(1) f^{(1)}(2) + f^{(0)}(2) f^{(1)}(1) + g^{(1)}(12, t_0)] + \frac{n}{m} \int d\xi \, d\mathbf{v}_3 \, \frac{d\phi}{d\xi} \cdot \left[\frac{\partial f^{(0)}}{\partial \mathbf{v}_1} \, g^{(1)}(\xi - \mathbf{x}, 23, t_0) - \frac{\partial f^{(0)}}{\partial \mathbf{v}_2} \, g^{(1)}(\mathbf{x} - \xi, 13, t_0)\right]. \tag{13}$$

This can be written in the following form:

$$\left(\frac{\partial}{\partial t_0} + \mathbf{v}_{12} \cdot \frac{\partial}{\partial \mathbf{x}}\right) g^{(2)} = Q(\mathbf{x}, t_0), \tag{14}$$

where $Q$ is a known function. From the right-hand side of Eq. (13), it is seen that $Q$ consists of three parts:

(1) Correction terms to $f^{(0)}$, i.e., terms containing $f^{(1)}$.
(2) Iteration term containing $g^{(1)}$.
(3) Three-particle terms.

The solution of Eq. (14) is as follows:

$$g^{(2)} = \int_0^{t_0} d\tau Q(\mathbf{x} - \mathbf{v}_{12}\,\tau, t_0 - \tau). \tag{15}$$

In examining the asymptotic time behavior of $g^{(2)}$, i.e., its behavior in the limit $t_0 \to \infty$, it is noted that the solution due the first group of terms is similar to that in $g^{(1)}$, which is well behaved. The solution in Eq. (15) due to the iteration terms can be put into the following two parts:

$$(1) \quad \frac{1}{m^2}\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right)^2 f^{(0)} f^{(0)} : \int_0^{t_0} d\tau \, \frac{d\phi}{d\mathbf{x}'} \, [\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}\tau] \times \int_0^{t_0 - \tau} d\tau' \, \frac{d\phi}{d\mathbf{x}'} \, [\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}(\tau + \tau')] \tag{16}$$

[where $\mathbf{ab} : \mathbf{cd} = (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})$]. For any given time $t_0$, the domain of the double time integral in the $\tau\tau'$-plane is an isosceles triangle as shown in Fig. 1(a). As $t_0 \to \infty$, this domain of integra-

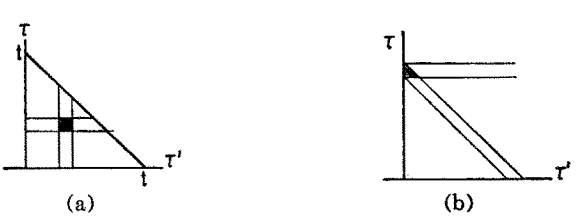


FIG. 1.

tion goes into the first-quarter plane. Because of the assumption of a repulsive potential and the finiteness of the range of the potential, the integrand is different from zero only in a certain region on this quarter plane. The integral in Eq. (16) will be finite as $t_0 \to \infty$ (nonsecular) if this region of nonzero integrand is finite. For any given value of $\mathbf{x}$ and $\mathbf{v}_{12}$ ($\neq 0$), the region of nonzero integrand is easily determined. It is shown in Fig. 1(a) as the shaded area. The width of the stripes shown in Fig. 1(a) is determined by the value $|\mathbf{v}_{12}|$, range of potential and the angle between $\mathbf{x}$ and $\mathbf{v}_{12}$, while the location of the strips is determined by $|\mathbf{x}|$. As $\mathbf{v}_{12} \to 0$, the region of nonzero integrand expands and fills up the entire quarter plane. The integral in Eq. (16) becomes doubly infinite.

To see the nature of the singularity at $\mathbf{v}_{12} = 0$ as $t_0 \to \infty$, we introduce, for any small but finite $\mathbf{v}_{12}$, the following transformation:

$$|\mathbf{v}_{12}|\,\tau = s, \qquad |\mathbf{v}_{12}|\,\tau' = s'. \tag{17}$$

The double integral over time becomes

$$\frac{1}{|\mathbf{v}_{12}|^2}\iint ds\,ds'\,\frac{d\phi}{d\mathbf{x}'}\,[\mathbf{x}' = \mathbf{x} - \hat{e}_{12}s] \times \frac{d\phi}{d\mathbf{x}'}\,[\mathbf{x}' = \mathbf{x} - \hat{e}_{12}(s + s')],$$

where $\hat{e}_{12} = \mathbf{v}_{12}/|\mathbf{v}_{12}|$ and the domain of integration is over the first quarter of the $ss'$-plane. The integral now is finite and we see that the singularity at $\mathbf{v}_{12} = 0$ is like $1/|\mathbf{v}_{12}|^2$. We substitute this result into Eq. (16), noting that

$$\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right) f^{(0)}(1) f^{(0)}(2) \sim \mathbf{v}_{12} \quad \text{for small} \quad |\mathbf{v}_{12}|.$$

We then see that the solution for $g^{(2)}(\mathbf{x}, \mathbf{v}_1, \mathbf{v}_2 t_0)$ has singularities of the form

$$1/|\mathbf{v}_{12}|^2, \qquad 1/|\mathbf{v}_{12}|$$

at $\mathbf{v}_{12} = 0$ in the limit of $t_0 \to \infty$.

The term we are considering now is part of the first iteration on the force of interaction. It is not difficult to see that on each further iteration, we shall introduce the operator

$$\int d\tau \exp\left[-\tau \mathbf{v}_{12} \cdot \frac{\partial}{\partial \mathbf{x}}\right] \frac{d\phi}{d\mathbf{x}} \cdot \left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right)$$

onto the known result for the $g$ function. This operator brings in further factors of $1/|\mathbf{v}_{12}|^2$ and $1/|\mathbf{v}_{12}|$ for $\mathbf{v}_{12} \to 0$. The formal series solution for the $g$ function is then not asymptotic in character[12] near $|\mathbf{v}_{12}| = 0$. It is readily seen that for the series solution to be asymptotic, we must have

$$|\mathbf{v}_{12}| \gg \epsilon^{\frac{1}{3}}. \tag{18}$$

In the remaining part of the present section we shall consider the solution for the correlations in such a region of the relative velocity space that Eq. (18) holds. For the relative velocity of the order $\epsilon^{\frac{1}{3}}$, a different asymptotic representation will be given in the following section.

$$(2)\quad \frac{1}{m^2}\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right) f^{(0)}(1) f^{(0)}(2) \times \int_0^{t_0} d\tau \int_0^{t_0 - \tau} d\tau'\,\tau' \frac{d^2\phi}{d\mathbf{x}'\,d\mathbf{x}'}\,[\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}(\tau + \tau')] \times \frac{d\phi}{d\mathbf{x}'}\,[\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}\tau]. \tag{19}$$

The domain of nonzero integrand is shown in Fig. 1(b) as the shaded area. It is seen that the integral in Eq. (19) is not secular for any finite $\mathbf{v}_{12}$. Near $|\mathbf{v}_{12}| = 0$, we may use the same transformation as before, Eq. (17). It is found that the whole term behaves like $|\mathbf{v}_{12}|^{-1}$, $|\mathbf{v}_{12}|^{-2}$ as $\mathbf{v}_{12} \to 0$.

The solution for $g^{(2)}$ due to one of the three-particle terms is as follows:

$$\frac{n}{m^2}\int d\xi\,d\mathbf{v}_3\,\frac{d\phi}{\partial \xi}\cdot\frac{\partial f^{(0)}}{\partial \mathbf{v}_1}\left(\frac{\partial}{\partial \mathbf{v}_2} - \frac{\partial}{\partial \mathbf{v}_3}\right) f^{(0)} f^{(0)} \times \int_0^{t_0} d\tau \int_0^{t_0 - \tau} d\tau'\,\frac{d\phi}{d\mathbf{x}'}\,[\mathbf{x}' = \xi - \mathbf{x} + \mathbf{v}_{12}\tau - \mathbf{v}_{23}\tau']. \tag{20}$$

For any given values of $\mathbf{x}$ and $\mathbf{v}_{12}$ it is seen that the region of nonzero integrand is limited in the $\tau\tau'$ plane for general values of $\xi$ and $\mathbf{v}_{23}$ [Fig. 2(a)].

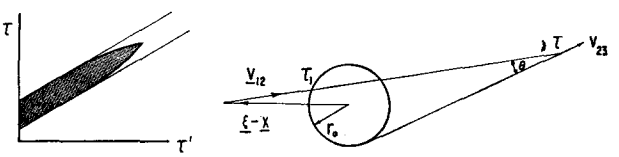


FIG. 2.

However, as the direction of $\mathbf{v}_{23}$ comes closer and closer to that of $\mathbf{v}_{12}$, this region stretches out to become longer and longer. Finally at $\mathbf{v}_{23} \parallel \mathbf{v}_{12}$ the region becomes a strip extending to infinity. The $\tau\tau'$-integral in Eq. (20) is thus singly infinite in the limit as $t_0 \to \infty$. However, this secularity is integrable in the $\mathbf{v}_3$ integration. In fact, as is shown in Fig. 2(b), the length of the above strip varies like $1/\theta$ as $\theta \to 0$, where $\theta$ is the angle between $\mathbf{v}_{23}$ and $\mathbf{v}_{12}$. Integration over the solid angle in the

[12] That is, each term will be bigger than its preceding term in the series at this singular point.

velocity ($\mathbf{v}_{23}$) space makes this singularity disappear, i.e.,

$$\int \frac{1}{\theta} \sin\theta \, d\theta \, d\phi \quad \text{is finite.}$$

We mention in passing that Eq. (20) has a singularity $|\mathbf{v}_{12}|^{-1}$ at $\mathbf{v}_{12} = 0$ in the limit of $t_0 \to \infty$. However, in the solution for $g^{(2)}$ $(\mathbf{x},12,\ t_0)$ of Eq. (15), the two three-particle terms cancel each other in this limit.

We conclude the present section with the following statement: The solution for $g^{(2)}(\mathbf{x},\ 12,\ t_0)$ with zero initial value is finite in the limit $t_0 \to \infty$ as long as we keep $\mathbf{v}_{12} \gg \epsilon^{\frac{1}{3}}$. Calling the region $\mathbf{v}_{12} \gg \epsilon^{\frac{1}{3}}$ the outer region, we see that up to second order the two-particle correlation function, after a time of order of that between collisions, becomes a functional of the one-particle distribution function as far as its time dependence is concerned.

### 3.3 Solution for $g^{(3)}(x,\ 12,\ t)$

The equation governing the third-order $g$ function is

$$\left(\frac{\partial}{\partial t_0} + \mathbf{v}_{12}\cdot\frac{\partial}{\partial \mathbf{x}}\right) g^{(3)}(\mathbf{x},\ \mathbf{v}_1,\ \mathbf{v}_2,\ t_0) + \frac{\partial g^{(1)}}{\partial t_2}$$

$$= \frac{1}{m}\frac{d\phi}{d\mathbf{x}}\cdot\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right)[g^{(2)}(12) + f^{(0)}(1)f^{(2)}(2,\ t_0)$$

$$+ f^{(1)}(1)f^{(1)}(2) + f^{(0)}(2)f^{(2)}(1,\ t_0)]$$

$$+ \frac{n}{m}\int d\xi\, d\mathbf{v}_3 \frac{d\phi}{d\xi}\cdot\left\{\frac{\partial f^{(1)}}{\partial \mathbf{v}_1} g^{(1)}(\xi - \mathbf{x},\ 23,\ t_0)\right.$$

$$- \frac{\partial f^{(1)}}{\partial \mathbf{v}_2} g^{(1)}(\mathbf{x} - \xi,\ 13,\ t_0) + \frac{\partial f^{(0)}}{\partial \mathbf{v}_1} g^{(2)}(\xi - \mathbf{x},\ 23,\ t_0)$$

$$- \frac{\partial f^{(0)}}{\partial \mathbf{v}_2} g^{(2)}(\mathbf{x} - \xi,\ 13,\ t_0) + \frac{\partial}{\partial \mathbf{v}_1} h^{(2)}(\mathbf{x},\ \xi,\ 123,\ t_0)$$

$$\left. - \frac{\partial}{\partial \mathbf{v}_2} h^{(2)}(\mathbf{x},\ \mathbf{x} - \xi,\ 123,\ t_0)\right\}. \tag{21}$$

We carry the two-particle correlation function up to this order for the following reasons:

(1) The multiple-time-scale theory first comes into effect in the solution for the $g$ function through the term $\partial g^{(1)}/\partial t_2$. Note that the dependence of $g^{(1)}$ on $t_2$ is completely known through the one-particle function $f^{(0)}$.

(2) $f^{(2)}$ is fast-time-dependent.

(3) The three-particle correlation function appears.

We shall leave out the analysis of the solution for $g^{(3)}$ due to all terms except those three groups of terms listed above, i.e.,

(1) $\partial g^{(1)}/\partial t_2$; its contribution to $g^{(3)}$ is

$$-\frac{1}{m}\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right)\frac{\partial}{\partial t_2} f^{(0)}f^{(0)}$$

$$\times \int_0^{t_0} d\tau \int_0^{t_0-\tau} d\tau' \frac{d\phi}{d\mathbf{x}'}\,[\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}(\tau + \tau')]. \tag{22}$$

For any finite $\mathbf{x}$ and $\mathbf{v}_{12}$, this remains finite in the limit $t_0 \to \infty$. However, it is secular if we choose

$$\mathbf{x} = [(1/\alpha)\mathbf{v}_{12}t_0 + \gamma] \to \infty, \tag{23}$$

where $\alpha \geq 0$ and $\gamma$ is a constant related to the range of the potential. The nature of this kind of secularity will be considered in more detail in the next paragraph.

(2) The solution of $g^{(3)}$ due to one of the terms containing $f^{(2)}$ is as follows:

$$-\frac{n}{m^3}\int_0^{t_0} d\tau \frac{d\phi}{d\mathbf{x}'}\,[\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}\tau]\cdot\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right)$$

$$\times \int d\xi\, d\mathbf{v}_3 \frac{d\phi}{d\xi}\cdot\frac{\partial}{\partial \mathbf{v}_2}\left(\frac{\partial}{\partial \mathbf{v}_2} - \frac{\partial}{\partial \mathbf{v}_3}\right) f^{(0)}f^{(0)}f^{(0)}$$

$$\times \int_0^{t_0-\tau} d\tau' \int_{\tau'}^{\infty} d\tau'' \frac{d\phi}{d\mathbf{x}'}\,[\mathbf{x}' = \xi - \mathbf{v}_{23}\tau'']. \tag{24}$$

The domain of the triple time integration is shown in Fig. 3(a). As $t_0 \to \infty$, this becomes a wedge-shaped space as indicated in Fig. 3(b). Since

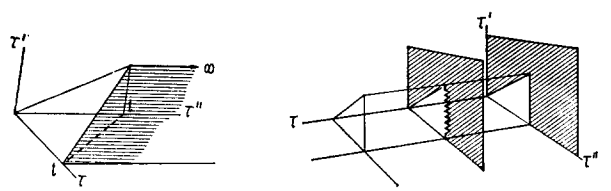


FIG. 3.

$\xi$ is bonded (finite) because of the factor $d\phi/d\xi$, it is seen that the region of nonzero integrand is finite as shown by the zigzag path in Fig. 3(b).

(3) The contribution from the three-particle correlation is

$$\frac{n}{m}\int d\xi\, d\mathbf{v}_3 \frac{d\phi}{d\xi}\cdot\frac{\partial}{\partial \mathbf{v}_1}$$

$$\times \int_0^{t_0} d\tau h^{(2)}(\mathbf{x} - \mathbf{v}_{12}\tau,\ \xi,\ 123,\ t_0 - \tau). \tag{25}$$

In order to analyze the asymptotic behavior of this expression, we have first to solve for $h^{(2)}$ through Eq. (7). This, in principle, is simple. However the expression for $h^{(2)}$ is rather lengthy, and we shall

not include it here. Suffice to say that, for any finite value of $\mathbf{x}$ and $\mathbf{v}_{12}$, all terms in Eq. (25) are nonsecular. However, if we let

$$\mathbf{x} = [(1/\alpha)\mathbf{v}_{12}t_0 + \gamma] \to \infty \tag{23}$$

one of the terms in Eq. (25), which has the following form, becomes secular:

$$\frac{n}{m^3}\int d\xi\, d\mathbf{v}_3 \frac{d\phi}{d\xi}\cdot\frac{\partial}{\partial \mathbf{v}_1} \times \int_0^{t_0} d\tau \int_0^{t_0-\tau} d\tau' \frac{d\phi}{d\mathbf{x}'}\,[\mathbf{x}' = \xi - \mathbf{v}_{13}\tau'] \times \left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right)\left\{\left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2}\right) f^{(0)}f^{(0)}f^{(0)} \times \int_0^{t_0-\tau-\tau'} d\tau'' \frac{d\phi}{d\mathbf{x}'}\,[\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}(\tau + \tau' + \tau'')]\right\}. \tag{26}$$

Graphically, the nonzero integrand in Eq. (26) is given by the zigzag path in Fig. 4. In the region

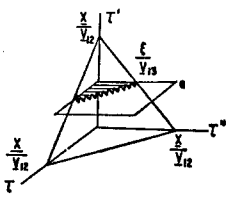

FIG. 4.

where Eq. (23) holds, we see that Eq. (26) becomes singly infinite. This local secularity has exactly the same nature as the one caused earlier by $\partial g^{(1)}/\partial t_2$. Making use of the Fokker–Planck equation, it can be shown that the region of nonzero integrand in Eq. (22) is the zigzag path in Fig. 5. It is to be noted

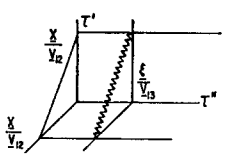

FIG. 5.

that the nonzero regions in the two figures are different; therefore the two secular terms in the solution for $g^{(3)}$ do not cancel each other exactly.

The failure of exact cancellation of the secularities can also be seen by carrying out Eq. (22) fully and comparing it with Eq. (26).

### 3.4 Secularity at $x = [(1/\alpha)v_{12}t_0 + \gamma]$

We have just shown that $g^{(3)}$ has a secular term (besides the secularity at small relative velocity) at

$$\mathbf{x} = [(1/\alpha)\mathbf{v}_{12}t_0 + \gamma] \to \infty, \quad \text{with} \quad \alpha \geq 1.$$

In order to have this kind of secularity, we see that we need a very careful arrangement of the interacting particles (both in velocities and positions). The particles have to be of large separation at time $t_0$ and with their relative velocity oriented in such a way that they are in collision at the initial moment $t_0 = 0$. It is not difficult to see that the angle between $\mathbf{x}$ and $\mathbf{v}_{12}$ should go as $1/|\mathbf{x}|$ in the limit $|\mathbf{x}| \to \infty$. If we focus our interest on the one-particle kinetic equation, which contains a velocity integration over the two-particle correlation, we see that up to $g^{(3)}$, the resulting one-particle equation is still well behaved. Moreover, since the pair potential has been assumed to have finite range, in the integral term of the one-particle equations, i.e.,

$$\int d\mathbf{x}\, d\mathbf{v}_2 \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g}{\partial \mathbf{v}_1},$$

the contribution of $g$ for any $|\mathbf{x}|$ greater than the range of the potential is strongly de-emphasized. We therefore conclude that as far as the one-particle equation is concerned, the second kind of secularity discussed in the present section can be completely ignored.

To analyze the nature of the second secularity, we re-examine the behavior of $g^{(1)}$, the lowest order nontrivial correlation function, we see that for the above arrangement in velocity and position, $g^{(1)}$ is not secular but it does not vanish for large separation between particles. We recall that $g^{(1)}$ is a solution based on two interacting particles only. If $g^{(1)}$ is infinite in range, the presence of a third particle cannot be neglected. However, as we have seen, the region where $g^{(1)}$ has an infinite range is very limited; the angle between the relative velocity and the relative position vector goes like $1/\mathbf{x}$ as $\mathbf{x} \to \infty$. If we average over any small but finite angle in the relative velocity space, it is easily seen that $g^{(1)}$ becomes finite in range, in fact,

$$g^{(1)} \sim 1/|\mathbf{x}|^2 \quad \text{as} \quad |\mathbf{x}| \to \infty. \tag{27}$$

The averaging process can be justified in the following way: Our ultimate interest is in the one-particle distribution function, which is affected by the correlation function only through the latter's gross properties. This is true because in obtaining the equation for the one-particle function, we always integrate over the relative velocity space, including the angular integration needed for the above averaging process. With this averaging, we see that $g^{(3)}$ is nonsecular and

$$g^{(3)} \sim 1/|\mathbf{x}| \quad \text{as} \quad |\mathbf{x}| \to \infty. \tag{28}$$

These asymptotic formulas for the $g$'s as $x \to \infty$

provide an indication of the validity of our formal expansion, i.e.,

$$\epsilon^3 g^{(3)}/\epsilon g^{(1)} \ll 1, \quad \text{or} \quad |x| \ll 1/\epsilon^2. \tag{29}$$

The corresponding inequality for the time variable is

$$t_0 \ll 1/\epsilon^2 \quad \text{for any finite} \quad |\mathbf{v}_{12}|. \tag{30}$$

It is recalled from our nondimensionalization that the mean free path and the mean free time in the system are of the order of $1/\epsilon^2$. From the inequalities in Eqs. (29) and (30) we see that in order to obtain a uniformly asymptotic representation of the correlation function, we have to rescale the hierarchy again in this local region by

$$\mathbf{x} = (1/\epsilon^2)X, \qquad t = (1/\epsilon^2)T.$$

The new equations are complicated and will not be considered in any detail in view of our interest in the one-particle kinetic equation. However, a conclusion one can draw at this state is that the correlation functions will now change in the kinetic time scale $(1/\epsilon^2)$ in their own right.

In what follows we shall ignore the second secularity and concentrate our attention on the secularity at small relative velocities only. It is true that the velocity integration in the one-particle equation will also smooth out some singularities at small $\mathbf{v}_{12}$. However, as we go to a higher order, e.g., $g^{(3)}$, the effect of the singularity at small $\mathbf{v}_{12}$ will definitely come in. The detailed analysis at small $|\mathbf{v}_{12}|$ and the correct higher order one-particle kinetic equation will be given in the following two sections.

## 4. EXPANSION FOR SMALL RELATIVE VELOCITIES

The weak coupling expansion we have developed in the previous section breaks down when the relative velocities among a group of particles under consideration become small. We have shown that the size of the relative velocities has to be much larger than $\epsilon^{\frac{1}{2}}$ in order that the series obtained be asymptotic in character. This criterion of size of the relative velocity can also be obtained by a simple argument based on the orbit of two interacting particles. Using the nondimensional form given in the last section, the equation of motion of two particles is

$$\ddot{\mathbf{x}} = \epsilon \mathbf{f}$$

where $\mathbf{f}$ is the force of order unity. The gain of velocity due to the interaction is of order $\epsilon t$, while the distance traversed by the particle due to this velocity is $\epsilon t^2$. Thus, for a particle with sufficiently small initial relative velocity, the time to traverse the range of the potential is of order $\epsilon^{-\frac{1}{2}}$, and the change of its velocity during this flight is of order $\epsilon^{\frac{1}{2}}$. Therefore, the weak-interaction approximation breaks down when the initial relative velocity is of order $\epsilon^{\frac{1}{2}}$. We see also from this simple argument that the fast time in the problem gets "slower" for particles having smaller relative velocity. Nevertheless, compared to the kinetic time scale $\epsilon^{-2}$, this "slower" fast time is still fast enough for the development of "kineticity."

To study the solution of the hierarchy equation for small relative velocity, we shall use the technique of the boundary layer analysis in fluid mechanics. As shown in the last paragraph, the region where the ordinary weak coupling expansion becomes invalid is very small (of order $\epsilon^{\frac{1}{2}}$). In order to investigate the structure inside the layer, we open up this thin layer by the following transformation[13]:

$$\mathbf{v}_{12} = \epsilon^{\frac{1}{2}}\mathbf{w}_{12} \quad \text{and thus} \quad \frac{\partial}{\partial \mathbf{v}_{12}} = \frac{1}{\epsilon^{\frac{1}{2}}}\frac{\partial}{\partial \mathbf{w}_{12}}, \tag{31}$$

where $\mathbf{w}_{12}$ is of order unity for $\mathbf{v}_{12}$ of order $\epsilon^{\frac{1}{2}}$. There should be corresponding stretching for other relative velocities. However, we shall concentrate here on the lower order approximation of the two-particle correlation functions only. For a detailed analysis to higher order, we refer to Ref. 9.

Using the scaling of Eq. (31) in the two-particle equation (6), it is easy to get the ordering of various terms in Eq. (6), except for the following two terms:

$$(\partial/\partial \mathbf{v}_1 - \partial/\partial \mathbf{v}_2)f(1)f(2), \qquad \partial f/\partial \mathbf{v}_2.$$

We have assumed that $f(2)$ is Taylor expandable around $\mathbf{v}_1$ for small $\mathbf{v}_{12}$; thus we see that the first term is of order $\epsilon^{\frac{1}{2}}$, while the second one is of order unity for $\mathbf{v}_{12}$ of order $\epsilon^{\frac{1}{2}}$.

The two-particle equation appropriate for small $\mathbf{v}_{12}$ is as follows:

$$\left\{\frac{\partial}{\partial t_1} + H(12)\right\} g(\mathbf{x}, \mathbf{w}_{12}, \mathbf{v}_1, t_1) - \frac{1}{m}\frac{d\phi}{d\mathbf{x}}$$
$$\times \left(\epsilon^{\frac{1}{2}}\frac{\partial}{\partial \mathbf{v}_1} + 2\frac{\partial}{\partial \mathbf{w}_{12}}\right) f(1)f(2) = \epsilon^{\frac{1}{2}}\frac{n}{m}\int d\xi\, d\mathbf{v}_{13}\frac{d\phi}{\partial \xi}$$
$$\times \left[\frac{\partial f^{(1)}}{\partial \mathbf{v}_1} g(\xi - \mathbf{x}, 23) + \frac{\partial f^{(2)}}{\partial \mathbf{v}_2} g(\mathbf{x} - \xi, 13)\right]$$
$$- \frac{n}{m}\int d\xi\, d\mathbf{v}_{13}\frac{d\phi}{d\xi}\cdot\left[\left(\epsilon^{\frac{1}{2}}\frac{\partial}{\partial \mathbf{v}_1} + \frac{\partial}{\partial \mathbf{w}_{12}}\right) h(\mathbf{x}, \xi, 123)\right.$$
$$\left. + \frac{\partial}{\partial \mathbf{w}_{12}} h(\mathbf{x}, \mathbf{x} - \xi, 123)\right], \tag{32}$$

[13] K. O. Friedrichs, Bull. Am. Math. Soc., **61**, 485 (1955).

where

$$t_{\frac{1}{2}} = \epsilon^{\frac{1}{2}} t_0,$$

and

$$H(12) = \mathbf{w}_{12} \cdot \frac{\partial}{\partial \mathbf{x}} - \frac{1}{m} \frac{d\phi}{d\mathbf{x}} \cdot \left( \epsilon^{\frac{1}{2}} \frac{\partial}{\partial \mathbf{v}_1} + 2 \frac{\partial}{\partial \mathbf{w}_{12}} \right).$$

We recall that the second term on the left side of Eq. (32) is of order $\epsilon$.

Instead of expanding all functions in series of $\epsilon$ we shall now expand them in powers of $\epsilon^{\frac{1}{2}}$, e.g.,

$$g = g^{(0)} + \epsilon^{\frac{1}{2}} g^{(\frac{1}{2})} + \epsilon g^{(1)} + \epsilon^{\frac{3}{2}} g^{(\frac{3}{2})} + \cdots.$$

Under the assumption of zero initial correlation functions, the lowest significant approximation is

$$\left\{ \frac{\partial}{\partial t_{\frac{1}{2}}} + H(12) \right\} g^{(1)}(\mathbf{x}, \mathbf{v}_1, \mathbf{w}_{12}, t_{\frac{1}{2}})$$

$$= \frac{1}{\epsilon} \frac{1}{m} \frac{d\phi}{d\mathbf{x}} \cdot \left( \epsilon^{\frac{1}{2}} \frac{\partial}{\partial \mathbf{v}_1} + 2 \frac{\partial}{\partial \mathbf{w}_{12}} \right) f^{(0)}(1) f^{(0)}(2). \qquad (33)$$

Note the term on the right side is of order unity. This equation differs from the corresponding one for finite $\mathbf{v}_{12}$. First, the interaction between particles is no longer treated as a small perturbation. Second, the fast time in the problem gets slower. These are what one would expect from the simple illustration given at the beginning of the section.

The solution of Eq. (33) is formally as follows:

$$g^{(1)}(\mathbf{x}, \mathbf{v}_1, \mathbf{w}_{12}, t_{\frac{1}{2}}) = \frac{1}{\epsilon} [S_{-t_{\frac{1}{2}}}(12) - 1] f^{(0)}(1) f^{(0)}(2), \qquad (34)$$

where

$$S_{-t_{\frac{1}{2}}}(12) = \exp [-t_{\frac{1}{2}} H(12)].$$

In spite of the factor $1/\epsilon$, $g^{(1)}$ is of order unity. This is best seen by carrying out the Taylor expansion of $f^{(0)}(2)$ in Eq. (33). Then the solution of $g^{(1)}$ is found as follows:

$$g^{(1)}(\mathbf{x}, \mathbf{v}_1, \mathbf{w}_{12}, t_{\frac{1}{2}}) = [S_{-t_{\frac{1}{2}}}(12) - 1] \tfrac{1}{4} \mathbf{w}_{12} \mathbf{w}_{12} : \mathbf{A}, \qquad (35)$$

where

$$\mathbf{A} = f^{(0)}(1) \nabla \nabla f^{(0)}(1) - \nabla f^{(0)}(1) \nabla f^{(0)}(1).$$

For very large $\mathbf{w}_{12}$, we have

$$S_{-t_{\frac{1}{2}}}(12) \mathbf{w}_{12} \mathbf{w}_{12} = \left( \mathbf{w}_{12} + \frac{2}{m} \frac{d\phi}{d\mathbf{x}} t_{\frac{1}{2}} \right) \left( \mathbf{w}_{12} + \frac{2}{m} \frac{d\phi}{d\mathbf{x}} t_{\frac{1}{2}} \right),$$

and

$$g^{(1)}(\mathbf{x}, \mathbf{v}_1, \mathbf{w}_{12}) = \frac{1}{m} \mathbf{w}_{12} \frac{d\phi}{d\mathbf{x}} : \mathbf{A} t_{\frac{1}{2}} = \frac{1}{m} \mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}} t_0 : \mathbf{A}$$

which is exactly the value of $g^{(1)}_{\text{out}}(\mathbf{x}, \mathbf{v}_1, \mathbf{v}_{12}, t_0)$ as $|\mathbf{v}_{12}| \to 0$. We therefore have shown that the solutions for $g^{(1)}$ in the two regions go into each other at the domain of their common validity. Similar analysis can be carried to higher orders. However, we shall not go any further here. The interested reader should refer to Ref. 9.

## 5. KINETIC EQUATIONS OF ONE-PARTICLE FUNCTIONS

We have obtained, in the previous two sections, solutions for correlation functions in two distinct regions of the relative velocity space. Hereafter we shall call the solution in the region of finite relative velocities the outer solution and the solution where the relative velocity is of the order $\epsilon^{\frac{1}{2}}$ the inner solution. In this section, we shall find the matching of these two solutions.

### 5.1 Rate of Relaxation

We have mentioned briefly the rate of relaxation of a system toward the state governed by the Fokker–Planck equation based on the outer solution for $g^{(1)}$ alone. Although $g^{(1)}_{\text{out}}(\mathbf{x}, 12, t_0)$ is well behaved in the limit $t_0 \to \infty$ for $\mathbf{v}_{12}$ small, it does not represent the true solution when $\mathbf{v}_{12}$ becomes small. The solution in this inner region, as we know from the analysis in the last section, is properly approximated by the inner solution $g^{(1)}_{\text{in}}(\mathbf{x}, 12, t_{\frac{1}{2}})$. Using $g^{(1)}_{\text{in}}(\mathbf{x}, 12, t_{\frac{1}{2}})$ instead of $g^{(1)}_{\text{out}}(\mathbf{x}, 12, t_0)$, the integral which gives the rate of relaxation becomes (where $t_{\frac{1}{2}} = \epsilon^{\frac{1}{2}} t$)

$$\frac{\partial}{\partial \mathbf{v}_1} \cdot \int d\mathbf{x}\, d\mathbf{v}_{12} \frac{\partial \phi}{\partial \mathbf{x}} \{ g^{(1)}_{\text{in}}(\mathbf{x}, 12, t_{\frac{1}{2}}) - g^{(1)}_{\text{in}}(\mathbf{x}, 12, \infty) \}$$

$$= \frac{1}{\epsilon} \frac{\partial}{\partial \mathbf{v}_1} \cdot \int d\mathbf{x}\, d\mathbf{v}_{12} \frac{\partial \phi}{\partial \mathbf{x}} (S_{-t_{\frac{1}{2}}} - S_{-\infty}) f^{(0)}(1) f^{(0)}(1). \qquad (36)$$

The pair potential has been assumed to have a finite range. At time $t_{\frac{1}{2}}$, because of the factor $d\phi/d\mathbf{x}$ in Eq. (36), the particle 2 must be within the range of particle 1 in order for the whole expression to be different from zero. For sufficiently large $t_{\frac{1}{2}}$, the streaming operator $S_{-t_{1/2}}$ will definitely bring particle 2 outside the range of particle 1 under the assumption of repulsive force. If the pair potential has an exponential-like tail, the expression in Eq. (36) will decay exponentially as $t_{\frac{1}{2}} \to \infty$. Therefore, in that case, the approach to the kinetic stage is exponentially fast in the time scale of $t_{\frac{1}{2}}$.

### 5.2 One-Particle Equation

We shall now consider the higher order one-particle kinetic equation. The hierarchy equation

for the one-particle distribution function is

$$\frac{\partial f}{\partial t} = \epsilon \frac{n}{m} \int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}} \cdot \frac{\partial g}{\partial \mathbf{v}_1}. \tag{5}$$

We shall neglect the transient behavior of the higher order one-particle distribution function on the fast time scales $t_0$ and $t_{\frac{1}{2}}$. Thus the value of $g$ in Eq. (5) will be the asymptotic value only, i.e., $g(t_0 \to \infty, t_{\frac{1}{2}} \to \infty)$. Since $g$ has two distinct representations in two different regions of $|\mathbf{v}_{12}|$, the integral

$$\int_0^\infty v_{12}^2\, dv_{12} g(\mathbf{x}, \mathbf{v}_1, \mathbf{v}_{12}, t_2) \qquad [\text{with } v_{12} = |\mathbf{v}_{12}|]$$

is split up into two parts, i.e.,

$$\int_0^\infty v_{12}^2\, dv_{12} g(\mathbf{x}, 12) = \int_0^\delta v_{12}^2\, dv_{12} g_{\text{in}}(\mathbf{x}, 12) + \int_\delta^\infty v_{12}^2\, dv_{12} g_{\text{out}}(\mathbf{x}, 12), \tag{37}$$

where $\delta$ is chosen in such a way that

$$\epsilon^{\frac{1}{2}} \ll \delta \ll 1. \tag{38}$$

We can then use the outer solution $g_{\text{out}}$ in the second integral of Eq. (37), and the inner solution $g_{\text{in}}$ in the first one. The order, in terms of expansion parameter $\epsilon$, of the right side of Eq. (37) is determined mainly by the behavior of $g_{\text{out}}(\mathbf{x}, 12)$ near $\mathbf{v}_{12} = 0$. Suppose $g(\mathbf{x}, \mathbf{v}_1, \mathbf{v}_{12})$ goes like $1/v_{12}^4$ as $|\mathbf{v}_{12}| \to 0$, the inner solution must be of order $(1/\epsilon^{\frac{1}{2}})^4 = 1/\epsilon^2$ so that the two solutions can be matched together around $|\mathbf{v}_{12}| \sim \epsilon^{\frac{1}{2}}$. Therefore,

$$\int_0^\delta v_{12}^2\, dv_{12} g(\mathbf{x}, 12) + \int_\delta^\infty v_{12}^2\, dv_{12} g(\mathbf{x}, 12)$$

$$= \frac{\epsilon^{\frac{1}{2}}}{\epsilon^2} \int_0^{\delta/\epsilon^{\frac{1}{2}}} w_{12}^2\, dw_{12} g_{\text{in}}(\mathbf{x}, \mathbf{w}_{12}) + \int_0^\infty v_{12}^2\, dv_{12} g_{\text{out}}(\mathbf{x}, \mathbf{v}_{12}) - \int_0^\delta v_{12}^2\, dv_{12} g_{\text{out}}(\mathbf{x}, \mathbf{v}_{12})$$

$$= \frac{1}{\epsilon^{\frac{1}{2}}} \int_0^\infty w_{12}^2\, dw_{12} g_{\text{in}}(\mathbf{x}, \mathbf{w}_{12}) - \frac{1}{\epsilon^{\frac{1}{2}}} \int_{\delta/\epsilon^{\frac{1}{2}}}^\infty w_{12}^2\, dw_{12} g_{\text{in}}(\mathbf{x}, \mathbf{w}_{12}) + \int_0^\infty v_{12}^2\, dv_{12} g_{\text{out}}(\mathbf{x}, \mathbf{v}_{12}) - \int_0^\delta v_{12}^2\, dv_{12} g_{\text{out}}(\mathbf{x}, \mathbf{v}_{12}).$$

In the second and fourth integrals, the integrand can be expanded in the following ways:

$$g_{\text{in}}(\mathbf{x}, \mathbf{w}_{12}) = \frac{F(0)}{w_{12}^4} + \epsilon^{\frac{1}{2}} \frac{F'(0)}{w_{12}^3} + \frac{\epsilon}{2} \frac{F''(0)}{w_{12}^2} + \cdots,$$

$$g_{\text{out}}(\mathbf{x}, \mathbf{v}_{12}) = \frac{F(0)}{v_{12}^4} + \frac{F'(0)}{v_{12}^3} + \frac{1}{2} \frac{F''(0)}{v_{12}^2} + \cdots.$$

Then

$$-\frac{1}{\epsilon^{\frac{1}{2}}} \int_{\delta/\epsilon^{\frac{1}{2}}}^\infty w_{12}^2\, dw_{12} g_{\text{in}}(\mathbf{x}, \mathbf{w}_{12}) = -\frac{F(0)}{\delta} - F'(0) \int_{\delta/\epsilon^{\frac{1}{2}}}^\infty \frac{dw_{12}}{w_{12}} - \frac{\epsilon^{\frac{1}{2}}}{2} F''(0) \int_0^\infty dw_{12} + \tfrac{1}{2} F''(0)\, \delta + \cdots$$

and

$$-\int_0^\delta v_{12}^2\, dv_{12} g_{\text{out}}(\mathbf{x}, \mathbf{v}_{12}) = \frac{F(0)}{\delta} - \int_0^\infty \frac{F(0)}{v_{12}^2}\, dv_{12} - F'(0) \int_0^\infty \frac{dv_{12}}{v_{12}} - \tfrac{1}{2} F''(0)\, \delta + \cdots.$$

Thus

$$\int_0^\infty v_{12}^2\, dv_{12} g(\mathbf{x}, 12) = \frac{1}{\epsilon^{\frac{1}{2}}} \int_0^\infty [w_{12}^2 g_{\text{in}}(\mathbf{x}, \mathbf{w}_{12}) - \tfrac{1}{2}\epsilon F''(0)]\, dw_{12} + \int_0^\infty \left[ v_{12}^2 g_{\text{out}}(\mathbf{x}, \mathbf{v}_{12}) - F(0) \frac{1}{v_{12}^2} \right] dv_{12} - F'(0) \left[ \int_0^\delta \frac{dv_{12}}{v_{12}} + \int_{\delta/\epsilon^{\frac{1}{2}}}^\infty \frac{dw_{12}}{w_{12}} \right]$$

or

$$\int_0^\infty v_{12}^2\, dv_{12} g(\mathbf{x}, 12) = \frac{1}{\epsilon^{\frac{1}{2}}} \int_0^\infty [w_{12}^2 g_{\text{in}}(\mathbf{x}, \mathbf{w}_{12}) - \tfrac{1}{2}\epsilon F''(0)]\, dw_{12} + \tfrac{1}{2} F'(0) \ln \epsilon + \int_0^\infty \left[ v_{12}^2 g_{\text{out}}(\mathbf{x}, \mathbf{v}_{12}) - \frac{F'(0)}{v_{12}} - \frac{F(0)}{v_{12}^2} \right] dv_{12}.$$

The $F(0)$, $F'(0)$, $F''(0)$ terms in the square brackets were introduced such that the infinite integrals become well defined. We may introduce the symbol $\text{C}\int_0^\infty$ standing for an infinite integral with a proper balancing term at the divergent end of the limit of the integration such as the first and third integral in the last formula. Then

$$\int_0^\infty v_{12}^2\, dv_{12} g = \frac{1}{\epsilon^{\frac{1}{2}}} \text{C}\int_0^\infty w_{12}^2\, dw_{12} g_{\text{in}} + \tfrac{1}{2} F'(0) \ln \epsilon + \text{C}\int_0^\infty v_{12}^2\, dv_{12} g_{\text{out}} \tag{39}$$

where

$$g_{\text{out}} \to F(\mathbf{v}_{12})/v_{12}^4 \quad \text{as} \quad \mathbf{v}_{12} \to 0.$$

Similarly we can work out the expression for the integral when $g_{\text{out}}$ has different behavior at $\mathbf{v}_{12} = 0$. For example

$$
\begin{aligned}
g_{\text{out}} \to \frac{F(\mathbf{v}_{12})}{v_{12}^3} &: \int_0^\infty v_{12}^2\, dv_{12} g \\
&= \tfrac{1}{2}F(0)\ln\epsilon + \int_{\mathrm{C}}\int_0^\infty w_{12}^2\, dw_{12} g_{\text{in}} + \int_{\mathrm{C}}\int_0^\infty v_{12}^2\, dv_{12} g_{\text{out}} \\
g_{\text{out}} \to \frac{F(\mathbf{v}_{12})}{v_{12}^2} &: \int_0^\infty v_{12}^2\, dv_{12} g \\
&= \int_0^\infty v_{12}^2\, dv_{12} g_{\text{out}} + \epsilon^{\frac{1}{2}} \int_{\mathrm{C}}\int_0^\infty w_{12}^2\, dw_{12} g_{\text{in}} \\
g_{\text{out}} \to \frac{F(\mathbf{v}_{12})}{v_{12}} &: \int_0^\infty v_{12}^2\, dv_{12} g \\
&= \int_0^\infty v_{12}^2\, dv_{12} g_{\text{out}} + \epsilon \int_{\mathrm{C}}\int_0^\infty w_{12}^2\, dw_{12} g_{\text{in}} \\
g_{\text{out}} \to F(\mathbf{v}_{12}) &: \int_0^\infty v_{12}^2\, dv_{12} g \\
&= \int_0^\infty v_{12}^2\, dv_{12} g_{\text{out}} + \epsilon^{\frac{1}{2}} \int_{\mathrm{C}}\int_0^\infty w_{12}^2\, dw_{12} g_{\text{in}}.
\end{aligned}
\tag{40}
$$

Using (39) and (40) in the integral of (5), we obtain the coefficients of the various orders of $\epsilon$ as follows:

$$
\begin{aligned}
\epsilon^2 &: \frac{n}{m}\int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g_{\text{out}}^{(1)}}{\partial \mathbf{v}_1} \\
\epsilon^3 &: \frac{n}{m}\int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g_{\text{out}}^{(2)}}{\partial \mathbf{v}_1} \\
\epsilon^{7/2} &: \frac{n}{m}\int_{\mathrm{C}} d\mathbf{x}\, d\mathbf{w}_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g_{\text{in}}^{(1)}}{\partial \mathbf{v}_1} \\
\epsilon^4 \ln\epsilon &: \frac{n}{2m}\int d\mathbf{x}\, d\Omega_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial}{\partial \mathbf{v}_1}\left[F_{35} + F'_{34}\right]_{\mathbf{v}_{12}=0} \\
\epsilon^4 &: \frac{n}{m}\int_{\mathrm{C}} d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g_{\text{out}}^{(3)}}{\partial \mathbf{v}_1},
\end{aligned}
\tag{41}
$$

where $d\Omega_{12}$ is the angular integration of $\mathbf{v}_{12}$, and $F_{35}$, $F_{34}$ are the coefficients of $1/|\mathbf{v}_{12}|^3$, $1/|\mathbf{v}_{12}|^4$ in $g_{\text{out}}^{(3)}$. Note that the coefficient of $\epsilon^{7/2}$ actually includes terms of order $\epsilon^{\frac{1}{2}}$. This is because $g_{\text{in}}^{(1)}$, as defined,[9] is of order unity plus terms of order $\epsilon^{\frac{1}{2}}$.

Using the first term of (41) in (5), we recover on the $t_2$ time scale, the Fokker–Planck equation

$$
\frac{\partial f^{(0)}}{\partial t_2} = \frac{n}{m}\int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g_{\text{out}}^{(1)}}{\partial \mathbf{v}_1} \quad (t_0 \to \infty). \tag{42}
$$

For the higher-order one-particle distribution function the form of the series in terms of $\epsilon$ is dictated by (41); we have

$$
f(\mathbf{v}_1, t_2) = f^{(0)} + \epsilon f^{(1)} + \epsilon^{\frac{1}{2}} f^{(\frac{1}{2})} + \epsilon^2 \ln\epsilon f_\alpha^{(2)} + \epsilon^2 f^{(2)} + \cdots.
$$

Using (5), we obtain the correction equations to the Fokker–Planck equation on the $t_2$ scale as follows:

$$
\begin{aligned}
\frac{\partial f^{(1)}}{\partial t_2} &= \frac{n}{m}\int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g_{\text{out}}^{(2)}}{\partial \mathbf{v}_1}, \\
\frac{\partial f^{(\frac{1}{2})}}{\partial t_2} &= \frac{n}{m}\int_{\mathrm{C}} d\mathbf{x}\, d\mathbf{w}_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g_{\text{in}}^{(1)}}{\partial \mathbf{v}_1}, \\
\frac{\partial f_\alpha^{(2)}}{\partial t_2} &= \frac{n}{2m}\int d\mathbf{x}\, d\Omega_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial}{\partial \mathbf{v}_1}\left[F'_{34} + F_{35}\right], \\
\frac{\partial f^{(2)}}{\partial t_2} &= \frac{n}{m}\int_{\mathrm{C}} d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g_{\text{out}}^{(3)}}{\partial \mathbf{v}_1}.
\end{aligned}
\tag{43}
$$

In principle, one can go on to higher orders successively. However, since the logarithmic terms will appear also in the correlation functions, the nature of the series in terms of $\epsilon$ becomes very complicated and the lengthy algebra is also formidable.

### 5.3 Termination of Time Scales

Since we are dealing with a system with spatial homogeneity, we know, by the $H$-theorem for the Fokker–Planck equation, that the system reaches thermal equilibrium on the time scale $t_2$. The dependence of all distribution functions on slower time scales, $t_3$, $t_4$, $\cdots$ should be trivial from a purely physical point of view. We recall that in the example given in the Appendix, the termination was automatic through the criterion of "nonsecularity." In the kinetic problem this sort of demonstration is more involved, since the solution of the lowest order equation is not so easy to obtain as in the case of the simple examples. To see this, we shall consider only the first correction to the Fokker–Planck equation, i.e., the equation for $f^{(1)}$. Instead of terminating the time scale at $t_2$, we use

$$
\frac{\partial}{\partial t_2} \to \frac{\partial}{\partial t_2} + \epsilon\frac{\partial}{\partial t_3}.
$$

The first equation in (43) then becomes

$$
\frac{\partial f^{(1)}}{\partial t_2} + \frac{\partial f^{(0)}}{\partial t_3} = \frac{n}{m}\int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g_{\text{out}}^{(2)}}{\partial \mathbf{v}_1}\{f^{(0)}, f^{(1)}\} \tag{44}
$$

where $g_{\text{out}}^{(2)}\{f^{(0)}, f^{(1)}\}$ implies a functional dependence of $g_{\text{out}}^{(2)}$ on $f^{(0)}$ and $f^{(1)}$. If we want to show that $f^{(0)}$ does not vary on the $t_3$ scale, we must demonstrate that the second term on the left side of Eq. (44) is secular in the solution for $f^{(1)}$ on the time scale $t_2$. For this purpose we must first obtain the $t_2$ dependence of $f^{(0)}$, which amounts to the solution of the Fokker–Planck equation. In general, this is quite difficult. However, in examining the secularity of the solutions for $f^{(1)}$ we are interested

only in the long-time behavior of each term in the solution $f^{(1)}$ on the $t_2$ scale. The asymptotic solution of the Fokker–Planck equation (42) is simply the $t_2$-time-independent Maxwellian function. We first note that the integral term in Eq. (53) can be written in two parts, i.e.,

$$C_1[f^{(1)}, f^{(0)}] + C_2[f^{(0)}(1), f^{(0)}(2), f^{(0)}(3)].$$

The first term is exactly as in the Fokker–Planck collision integral except that in one place $f^{(0)}$ is replaced by $f^{(1)}$. The second term depends on $f^{(0)}$ only and it vanishes when $f^{(0)}$ is Maxwellian. It is then seen that in the limit of $t_2 \to \infty$, the only secular term in Eq. (44) is

$$(\partial/\partial t_3)f^{(0)}(\mathbf{v}_1, t_2, t_3).$$

Therefore $f^{(0)}$ is required to be independent of $t_3$.

This sort of argument presumably can be carried on further. The termination of the time scale at $t_2$ for a homogeneous system is thus rational within the framework of the multiple-time-scale method.

### 6. THE EFFECT OF INITIAL CORRELATIONS

Up to now we have been investigating only the part of the correlation that arises because of the interaction between particles. It is seen that such correlations become functionals of the one-particle distribution in a time of order of the duration of a collision (except for the type of local secularity discussed in Sec. 3). This functional property is a necessary condition for the system to be describable by a one-particle kinetic equation on a time scale long compared with the duration of a collision. To ensure the one-particle kineticity, we must, in addition, examine the effect of initial correlations. If we accept the kinetic equation obtained in the absence of initial correlations as the right kinetic equation, which is logical, we must choose the initial correlations in such a way that the effect of these initial data on the one-particle function vanishes in a time of order of the duration of a collision. The necessary and sufficient conditions on the initial correlations such that the foregoing requirement is met is of primary importance in kinetic theory. However, since the initial correlations are functions of many variables (for example, the two-particle correlation is a function of three independent vectorial variables $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_{12}$) the search for a necessary and sufficient condition is quite difficult. In what follows we shall restrict ourselves to a sufficiency condition only.

The lower-order equations for the one-particle functions are

$$\partial f^{(0)}/\partial t_0 = 0, \tag{45}$$

$$\frac{\partial f^{(1)}}{\partial t_0} + \frac{\partial f^{(0)}}{\partial t_1} = \frac{n}{m}\int d\mathbf{x}\, d\mathbf{v}_{12}\, \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g^{(0)}}{\partial \mathbf{v}_1}. \tag{46}$$

These equations differ from those without initial correlation by the integral term in (46). The zero-order correlation $g^{(0)}$ is governed by the following equations[14]:

(1) Outer region $|\mathbf{v}_{12}| \gg \epsilon^{\frac{1}{3}}$

$$\frac{\partial g^{(0)}}{\partial t_0} + \mathbf{v}_{12}\cdot\frac{\partial g^{(0)}}{\partial \mathbf{x}} = 0. \tag{47}$$

(2) Inner region $|\mathbf{v}_{12}| \sim \epsilon^{\frac{1}{3}}$

$$\frac{\partial g^{(0)}}{\partial t_{\frac{1}{3}}} + \mathbf{w}_{12}\cdot\frac{\partial g^{(0)}}{\partial \mathbf{x}} - \frac{1}{m}\frac{\partial \phi}{\partial \mathbf{x}}\cdot\left(\epsilon^{\frac{1}{3}}\frac{\partial}{\partial \mathbf{v}_1} + 2\frac{\partial}{\partial \mathbf{w}_{12}}\right)g^{(0)} = 0. \tag{48}$$

These equations are homogeneous. The solutions of them are

$$g^{(0)}_{\text{out}}(\mathbf{x}, 12, t_0) = g^{(0)}(\mathbf{x} - \mathbf{v}_{12}t_0, 12, t_0 = 0), \tag{49}$$

$$g^{(0)}_{\text{in}}(\mathbf{x}, 12, t_{\frac{1}{3}}) = S_{-t_{\frac{1}{3}}}(12)g(\mathbf{x}, 12, t_{\frac{1}{3}} = 0). \tag{50}$$

We may also say simply that the zeroth-order correlation $g^{(0)}$ remains constant following the characteristic lines of the $g^{(0)}$ equations, respectively. Although the correlation persists as the time goes on, its effect on the one-particle function is greatly deemphasized by the nature of the integral in (46). Similar integrals (over velocity space $\mathbf{v}_{12}$ only) with the $g^{(0)}$ given by the free-streaming equation (47) have been analyzed by Van Kampen[15] in the problem of Landau damping of plasma oscillations. The process of this damping is usually referred to as "phase mixing." We note that there are two new features in the integral of (46). First, there is an additional integration in $\mathbf{x}$ space with a finite-range potential factor $d\phi/d\mathbf{x}$. Second, the integration over $|\mathbf{v}_{12}|$ has to be split into two parts, i.e.,

$$\int d\mathbf{x}\, d\mathbf{v}_{12}\, \frac{d\phi}{d\mathbf{x}}\cdot\frac{\partial g^{(0)}}{\partial \mathbf{v}_1} = \frac{\partial}{\partial \mathbf{v}_1}\cdot\int d\mathbf{x}\, d\Omega_{12}\, \frac{d\phi}{d\mathbf{x}}\left\{\int_0^{\delta} g^{(0)}_{\text{in}} + \int_{\delta}^{\infty} g^{(0)}_{\text{out}}\right\}v_{12}^2\, dv_{12} \tag{51}$$

where $\epsilon^{\frac{1}{3}} \ll \delta \ll 1$. Let us first consider the second integral in (51):

$$\int d\mathbf{x}\, d\Omega_{12}\, \frac{d\phi}{d\mathbf{x}}\int_{\delta}^{\infty} v_{12}^2\, dv_{12} g^{(0)}_{\text{out}}. \tag{52}$$

[14] We neglect the three-particle correlation $h^{(0)}$ here for simplicity. The effect of $h^{(0)}$ will be examined at the end of this section.

[15] N. Van Kampen, Physica **21**, 949 (1955).

Let us assume that the initial correlation has a finite range in $\mathbf{x}$ space. The integrand of (52) is different from zero only when the two particles under consideration are within the range of the force at time $t_0$ (because of $d\phi/d\mathbf{x}$) and also within the range of correlation at the initial time $-t_0$ [because of $g^{(0)}(\mathbf{x} - v_{12}t_0, t_0 = 0)$]. Therefore, for sufficiently large $t_0$ (say, $1/\epsilon$) the integral is always zero provided $|\mathbf{v}_{12}|$ is not too small. We have required, however, that the lower limit of the integration be much greater than $\epsilon^{\frac{1}{2}}$, thus $|\mathbf{v}_{12}|\, t_0 \gg \epsilon^{-\frac{1}{2}}$ and the integral (52) is zero in the limit as $t_0 \to \infty$ for all permissible values of $|\mathbf{v}_{12}|$. It is readily seen that the decay is exponential if the assumed initial correlation has an exponential tail. We therefore have from (46) and (51) that

$$\frac{\partial f^{(1)}}{\partial t_0} + \frac{\partial f^{(\frac{1}{2})}}{\partial t_{\frac{1}{2}}} + \frac{\partial f^{(0)}}{\partial t_1} = \frac{n}{m}\frac{\partial}{\partial \mathbf{v}_1}\cdot\int d\mathbf{x}\, d\Omega_{12}\frac{d\phi}{d\mathbf{x}}\int_0^\delta v_{12}^2\, dv_{12} g_{\text{in}}^{(0)} \qquad (53)$$

$$+ \text{exponential decay term in } t_0,$$

or

$$\partial f^{(1)}/\partial t_0 \quad \text{decays exponentially in } t_0,$$

and

$$\frac{\partial f^{(\frac{1}{2})}}{\partial t_{\frac{1}{2}}} + \frac{\partial f^{(0)}}{\partial t_1} = \frac{n}{m}\frac{\partial}{\partial \mathbf{v}_1}\cdot\int d\mathbf{x}\, d\Omega_{12}\frac{d\phi}{d\mathbf{x}}\int_0^\delta v_{12}^2\, dv_{12} g_{\text{in}}^{(0)}. \qquad (54)$$

We now ask the question whether two particles can stay within their range of correlation (we assume that the range of correlation is the same order of magnitude compared as the range of the force) forever. Starting with $|\mathbf{x}| < r_0$, which is required by $d\phi/d\mathbf{x}$, we have seen that the time to travel across the range of force for particles having relative velocities of order $\epsilon^{\frac{1}{2}}$ is of order $\epsilon^{-\frac{1}{2}}$. Thus, in the limit as $t_{\frac{1}{2}} \to \infty$, the particles initially within the range of force will be definitely outside the range of force. The $g_{\text{in}}^{(0)}$ in (54) goes to zero exponentially if the initial $g^{(0)}(\mathbf{x}, \mathbf{v}_1, \mathbf{v}_{12}, t_0 = 0)$ has an exponential tail in its $\mathbf{x}$ dependence. Therefore,

$$\partial f^{(\frac{1}{2})}/\partial t_{\frac{1}{2}} \quad \text{decays exponentially in } t_{\frac{1}{2}},$$

and

$$\partial f^{(0)}/\partial t_1 = 0.$$

We have so far disregarded the relative velocity dependence of the initial correlation $g^{(0)}(\mathbf{x}, \mathbf{v}_1, \mathbf{v}_{12}, t_0 = 0)$ (its $\mathbf{v}_1$ dependence is not of importance in the discussion of the approach to the kinetic regime). It is seen from (52) that a strong singularity of $g^{(0)}(\mathbf{x}, \mathbf{v}_1, \mathbf{v}_{12}, t_0 = 0)$ at $\mathbf{v}_{12} = 0$ might cause trouble in the integral. Supposing that the spatial dependence of $g^{(0)}(\mathbf{x} - \mathbf{v}_{12}\, t_0, 0)$ at large $t_0$ has the exponential form $\epsilon^{-\mathbf{v}_{12}t_0}$, it is easy to see that any algebraic singularity in $\mathbf{v}_{12} = 0$ will not change the conclusion we have drawn before. Therefore the effect of the initial correlation on the one-particle equation vanishes as long as the following conditions are satisfied:

(1) The interaction potential is repulsive and has a finite range.

(2) The initial correlation has a finite range and drops off exponentially as $\mathbf{x} \to \infty$.

(3) The relative velocity dependence of $g^{(0)}(\mathbf{x}, \mathbf{v}_1, \mathbf{v}_{12}, t_0 = 0)$ is not exponentially large at $\mathbf{v}_{12} = 0$.

These conditions are only sufficient ones. One can certainly relax condition (2) with some strengthening of condition (3). After exhausting all the possible combinations, perhaps one could obtain a necessary condition on the initial correlation such that the long-time behavior of the system is independent of the initial data imposed on the system.

The consideration of both inner and outer solutions for $g^{(0)}$ in (51) is essential. Since the outer solution describes the free motion of the particles, the velocity and thus the relative velocity of the particles remains constant in the course of time. The mixing process provided by the velocity integral in (51) becomes more and more ineffective as the relative velocity of the particles goes to zero. However, when the relative velocity between two particles becomes small, the change of the momentum due to their interaction (even though it is weak) can no longer be neglected. It is such interactions which keep the relative velocity from vanishing after each encounter of the particles. This velocity provides enough mixing of the initial data on the time scale $t_{\frac{1}{2}}$. It is easily seen that had we considered the outer solution alone, the condition on the initial correlation would have been much more stringent.[8]

The full equation of the zeroth-order $g^{(0)}$ in the inner region is

$$\frac{\partial g^{(0)}}{\partial t_{\frac{1}{2}}} + \mathbf{w}_{12}\cdot\frac{\partial g^{(0)}}{\partial \mathbf{x}} - \frac{2}{m}\frac{\partial \phi}{\partial \mathbf{x}}\cdot\frac{\partial g^{(0)}}{\partial \mathbf{w}_{12}} = \frac{n}{m}\int d\mathbf{x}\, d\mathbf{v}_3\left[\frac{\partial \phi(\mathbf{x}_1 - \mathbf{x}_3)}{\partial \mathbf{x}_1}\cdot\frac{\partial}{\partial \mathbf{v}_1} + \frac{\partial \phi(\mathbf{x}_2 - \mathbf{x}_3)}{\partial \mathbf{x}_2}\cdot\frac{\partial}{\partial \mathbf{v}_2}\right]h^{(0)}(123).$$

If we impose the same sufficient conditions on

$h^{(0)}(\mathbf{x}, \xi, 123, t_0 = 0)$, it is obvious that the integral terms on the right side of the equation vanish in the limit as $t_0 \to \infty$, $t_{\frac{1}{2}} \to \infty$.

The analysis for the higher orders can be carried out in a similar fashion. As long as the sufficiency conditions are met, the effect of the initial correlations will wash away on the fast time scales.

### 7. QUESTION OF IRREVERSIBILITY

In the previous section we have seen how the effect of the initial correlations on the one-particle function are mixed out during a time interval of the order of the collision time. At first glance, such mixing of the initial data is analogous to usual relaxation processes. One therefore tends to attribute to it the irreversible character of the resulting Fokker–Planck equation or, more generally, the irreversible nature of the kinetic equation. This, however, is not true. Let us first note the following two facts:

(1) The mixing process does not have a preferred direction in time. The equations for the correlations (47) and (48) are time reversible, i.e., they are invariant under the transformation $t \to -t$ and $\mathbf{v} \to -\mathbf{v}$. It is thus not difficult to see that, with the sufficient conditions specified before, the integral in the equation

$$\frac{\partial f^{(1)}}{\partial t_0} + \frac{\partial f^{(\frac{1}{2})}}{\partial t_{\frac{1}{2}}} + \frac{\partial f^{(0)}}{\partial t_1} = \frac{n}{m} \int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}} \cdot \frac{\partial g^{(0)}}{\partial \mathbf{v}_1}$$

also vanishes in the limit as $t_0 \to -\infty$. This argument apparently applies to the higher-order equations too. In fact, the equations of the hierarchy up to this stage are still time-reversible.

(2) The initial correlations themselves do not phase-mix, but only their effect on the one-particle function does phase-mixing and dies out in the long-time limit. This seems to correspond to the Ehrenfest's coarse-graining, which, by itself, is not enough to ensure that the system evolves with a preferred direction of time.[16]

Although the mixing mechanism, as we have just pointed out, is not responsible for the irreversibility of the kinetic equation, it does give one of its important properties, independence of the initial information on the correlation functions.

To single out exactly where the irreversible nature of the kinetic equation has been introduced, we shall concentrate our attention on the equation

$$\frac{\partial f^{(2)}}{\partial t_0} + \frac{\partial f^{(\frac{1}{2})}}{\partial t_{\frac{1}{2}}} + \frac{\partial f^{(1)}}{\partial t_1} + \frac{\partial f^{(0)}}{\partial t_2} = \frac{n}{m} \int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}} \cdot \frac{\partial g^{(1)}}{\partial \mathbf{v}_1}.$$

[16] See Ref. 10, p. 85, for example.

For our present purpose we shall ignore the terms $\partial f^{(\frac{1}{2})}/\partial t_{\frac{1}{2}}$ and $\partial f^{(1)}/\partial t_1$. Using the creation part of $g^{(1)}_{\text{out}}$ only, we obtain

$$\frac{\partial f^{(2)}}{\partial t_0} + \frac{\partial f^{(0)}}{\partial t_2} = \frac{n}{m^2} \frac{\partial}{\partial \mathbf{v}_1} \times \int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}} \left( \frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2} \right) f^{(0)} f^{(0)} \times \int_0^{t_0} d\tau \frac{d\phi}{d\mathbf{x}'} [\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}\tau]. \tag{55}$$

This equation is still invariant under the time-reversal transformation $t \to -t$ (with $t_2 \to -t_2$) and $\mathbf{v} \to -\mathbf{v}$. The kinetic equation (of the lowest order) is obtained by requiring that the solution of Eq. (55) be nonsecular in the limit of $t_0 \to +\infty$, i.e.,

$$\frac{\partial f^{(2)}}{\partial t_0} = -\frac{n}{m^2} \frac{\partial}{\partial \mathbf{v}_1} \cdot \int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}} \left( \frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2} \right) f^{(0)} f^{(0)} \times \int_{t_0}^{\infty} d\tau \frac{d\phi}{d\mathbf{x}'} [\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}\tau], \tag{56}$$

$$\frac{\partial f^{(0)}}{\partial t_2} = \frac{n}{m^2} \frac{\partial}{\partial \mathbf{v}_1} \cdot \int d\mathbf{x}\, d\mathbf{v}_{12} \frac{d\phi}{d\mathbf{x}} \cdot \left( \frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2} \right) f^{(0)} f^{(0)} \times \int_0^{\infty} d\tau \frac{d\phi}{d\mathbf{x}'} [\mathbf{x}' = \mathbf{x} - \mathbf{v}_{12}\tau]. \tag{57}$$

Neither of these equations is invariant under the time-reversal transformation. Thus it is seen that the requirement of nonsecularity of the solution of Eq. (55) in the limit $t_0 \to \infty$ picks out a special class of solutions which is governed by the irreversible equations (56) and (57). Our special interest is, of course, centered on Eq. (56), which carried the system to thermal equilibrium in the time scale of order of the mean free flight time of the particles. Mathematically, one has an equal right to seek a solution of Eq. (55) which is nonsecular in the limit of $t_0 \to -\infty$. The corresponding time-irreversible equation obtained in this way is exactly the same as Eq. (56) except for a minus sign. This equation does not approach thermal equilibrium in the direction of positive $t_2$. From the physical point of view, one can exclude the latter equation by arguing that the system is prepared at $t = 0$, and one is interested only in the evolution of the system in the positive direction of $t$. Nevertheless, the philosophical subtlety of distinguishing the direction of time still remains. We leave this point to our readers who are more philosophically inclined.

### ACKNOWLEDGMENTS

The author is deeply indebted to Professor E. A.

Frieman for the suggestion of the problem and the discussions with him. It is also a great pleasure to acknowledge the discussions with Professor M. D. Kruskal.

The work was performed as a partial fulfillment of the requirement for Ph.D. degree of Princeton University. It was supported under Contract AT(30-1)-1238 with the U. S. Atomic Energy Commission.

## APPENDIX I: METHOD OF MULTIPLE TIME SCALES

We shall consider some simple examples. The method as well as the nature of the expansion in our work in kinetic theory are well demonstrated by the first example

$$\text{(a)} \qquad df/dt = -\epsilon f \,. \tag{A1}$$

The exact solution is trivial (i.e., $f = Ae^{-\epsilon t}$). Supposing that we want to solve (1) approximately by a straightforward series representation, we get

$$f(t) = A[1 - \epsilon t + \epsilon^2 t^2/2 + \cdots + (-1)^n \epsilon^n t^n/n! + \cdots], \tag{A2}$$

where $A$ is an arbitrary constant of integration. Even though the series is convergent, *it is not asymptotic* for large $t$. To get an asymptotic representation, we replace the single time coordinate $t$ in (A1) by a sequence of time coordinates $t_0$, $t_1$, $t_2$, $\cdots$ with $dt_n/dt = \epsilon^n$ (a "space" of times), i.e.,

$$f = f(t_0, t_1, t_2, \cdots). \tag{A3}$$

We treat $t_0$, $t_1$, $t_2$, $\cdots$ as independent variables. Formally it is equivalent to

$$\frac{d}{dt} = \frac{d}{dt_0} + \epsilon \frac{d}{dt_1} + \epsilon^2 \frac{d}{dt_1} + \cdots. \tag{A4}$$

Using this, we obtain a set of equations for successive powers of $\epsilon$. The first equation yields $f^{(0)} = f^{(0)}(t_1, t_2, \cdots)$. We have then from the second equation

$$f^{(1)} = -[df^{(0)}/dt_1 + f^{(0)}]t_0 + B(t_1, t_2, \cdots). \tag{A5}$$

To eliminate the secularity we must require the coefficient of $t_0$ to be zero, i.e.,

$$df^{(0)}/dt_1 + f^{(0)} = 0, \text{ or } f^{(0)} = A(t_2, \cdots)e^{-t_1}. \tag{A6}$$

Using this same procedure we obtain the following solution:

$$f = e^{-t_1}[A_0 + \epsilon B_0 + \cdots], \tag{A7}$$

where $A_0$, $B_0$, $\cdots$ are pure constants which are determined by the initial values of $f^{(0)}$, $f^{(1)}$, $f^{(2)}$, $\cdots$. Series (A7) is a uniformly valid asymptotic expansion.

The integration "constants" $B$ and $A$ in (A5) and (A6) were taken as functions of the appropriate time coordinates instead of pure constants. In other words, consistently with the imbedding of the time coordinate, the initial value, which is a pure constant, is replaced by a set of unknown functions which are to be determined by the condition that the resulting series solution be asymptotic. It is easily checked that this procedure is not essential for the present example. One obtains the same result by taking $A$, $B$, $\cdots$ to be pure constants only. However, for some other problems the replacement proves to be essential. This can be demonstrated by (A8)

$$\text{(b)} \qquad df/dt = -(\epsilon + \epsilon^2)f. \tag{A8}$$

We shall not include the simple algebra here. It suffices to say that in order to obtain an asymptotic representation for (A8), the allowance for the unknown functions is necessary.

While the foregoing examples suggest that it is logical to extend the initial values as well as the time variable itself, we see that the requirement that the series solution be asymptotic gives enough constraints on the extended problem and we obtain a unique meaningful approximation (exact solution in the given examples). This is by no means true in general. One can convince himself by our last example.

$$\text{(c)} \qquad df/dt = -\epsilon f^2.$$

It is not difficult to show that the constraint we have used so far is not strong enough to make the problem well defined, if unknown functions are included.

There is possible confusion between the multiple-time-scales method and the PLK method. The latter perturbation technique has been known to aerodynamicists through Lighthill[17] and Kuo.[18] They applied the method to solve problems of compressible flows and viscous boundary layer, respectively. The idea was first invented by Poincaré in dealing with the perturbation of orbits of celestial bodies. The basic idea of the method is that, in addition to expanding the dependent variable, one expands the independent variables also, i.e.,

$$f = f^{(0)}(z) + \epsilon f^{(1)}(z) + \cdots, \tag{A9}$$

$$t = z + \epsilon T^{(1)}(z) = \epsilon^2 T^{(2)}(z) + \cdots, \tag{A10}$$

[17] M. J. Lighthill, Phil. Mag. [7] **40**, 1179(1949).
[18] H. S. Tsien, *Advances in Applied Mechanics* (Academic Press Inc., New York, 1956).

where $z$ is a parametric variable. $T^{(1)}(z)$, $T^{(2)}(z)$, $\cdots$ are unknown functions. They are determined in such a way that the singularities in the straightforward expansion, if there are any, are eliminated as much as possible. In this regard the two methods are similar, but the similarity ends here. To see this, we invert the series (A10) and substitute it in (A9),

$$f = f^{(0)}[t - \epsilon T^{(1)}(t) \cdots] + \epsilon f^{(1)}[t - \epsilon T^{(1)}(t) \cdots] + \cdots .$$

Compare this with the multiple-time-scale method:

$$f = f^{(0)}(t_0, t_1, \cdots) + \epsilon f^{(1)}(t_0, t_1, \cdots) + \cdots .$$

We see they are quite different in nature. Up to the $n$th order, there are $2n + 1$ unknown functions of of single variables in the PLK method, whereas there are $n + 1$ unknown functions in $n + 1$ variables in the multiple-time-scale method. The extra freedom introduced in the PLK method up to $n$th order has a number of $n$, while in the multiple-time-scale method, it is of $n^2$. It is not difficult to see that the procedure of (A9) and (A10) is not applicable to our first example, as far as the asymptotic series is concerned.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# The Uniqueness of Weak Solutions of the One-Dimensional Scalar Analog to the Navier–Stokes Equation

EDGAR A. KRAUT
*Institute of Geophysics and Planetary Physics, and Department of Physics, University of California Los Angeles, California*
(Received 31 March 1964)

A uniqueness theorem for solutions of the one-dimensional scalar analog to the Navier–Stokes equation is stated and rigorously established.

## 1. INTRODUCTION

IT is well known[1] that weak solutions of partial differential equations are not always unique. In certain cases this may be remedied by introducing higher-order dissipative or dispersive terms. An example where this is the case leads to the one-dimensional scalar analog to the Navier–Stokes equation. This one-dimensional analog was introduced by Burgers[2,3] and has been studied from different points of view by Hopf[4] and Kraichnan.[5]

We consider a mixed initial and boundary-value problem for the nonlinear, inhomogeneous, one-dimensional scalar analog to the Navier–Stokes equation; and following the method of Ladyzhenskaya,[6] we prove that the weak solution is unique.

It is known that the problem under consideration can be linearized by a change of the dependent variable. The uniqueness proof given here does not depend on this fact.

## 2. FORMULATION

Let $Q_T$ represent the closed rectangular region

$$0 \leq x \leq L, \tag{2.1}$$

$$0 \leq t \leq T. \tag{2.2}$$

Consider the solution of

$$u_t + uu_x = \nu u_{xx} + f(x, t), \tag{2.3}$$

subject to the boundary condition

$$u(0, t) = u(L, t) = 0, \tag{2.4}$$

and the initial condition

$$u(x, 0) = a(x), \tag{2.5}$$

where

$$a(0) = a(L) = 0. \tag{2.6}$$

[1] P. Lax, *Nonlinear Problems* (The University of Wisconsin Press, Madison, 1963), pp. 3–12.
[2] J. M. Burgers, Proc. Acad. Sci. Amsterdam **43**, 2 (1940).
[3] J. M. Burgers, Advan. Appl. Mech. **1**, 171 (1948).
[4] E. Hopf, Commun. Pure Appl. Math. **3**, 201 (1950).
[5] R. H. Kraichnan, J. Math. Phys. **2**, 124 (1961).
[6] O. A. Ladyzhenskaya, *The Mathematical Theory of Viscous Incompressible Flow* (Gordon and Breach, Science Publishers, New York, 1963), Chap. 6.

where $z$ is a parametric variable. $T^{(1)}(z)$, $T^{(2)}(z)$, $\cdots$ are unknown functions. They are determined in such a way that the singularities in the straightforward expansion, if there are any, are eliminated as much as possible. In this regard the two methods are similar, but the similarity ends here. To see this, we invert the series (A10) and substitute it in (A9),

$$f = f^{(0)}[t - \epsilon T^{(1)}(t) \cdots] + \epsilon f^{(1)}[t - \epsilon T^{(1)}(t) \cdots] + \cdots .$$

Compare this with the multiple-time-scale method:

$$f = f^{(0)}(t_0, t_1, \cdots) + \epsilon f^{(1)}(t_0, t_1, \cdots) + \cdots .$$

We see they are quite different in nature. Up to the $n$th order, there are $2n + 1$ unknown functions of of single variables in the PLK method, whereas there are $n + 1$ unknown functions in $n + 1$ variables in the multiple-time-scale method. The extra freedom introduced in the PLK method up to $n$th order has a number of $n$, while in the multiple-time-scale method, it is of $n^2$. It is not difficult to see that the procedure of (A9) and (A10) is not applicable to our first example, as far as the asymptotic series is concerned.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# The Uniqueness of Weak Solutions of the One-Dimensional Scalar Analog to the Navier–Stokes Equation

EDGAR A. KRAUT
*Institute of Geophysics and Planetary Physics, and Department of Physics, University of California Los Angeles, California*
(Received 31 March 1964)

A uniqueness theorem for solutions of the one-dimensional scalar analog to the Navier–Stokes equation is stated and rigorously established.

## 1. INTRODUCTION

IT is well known[1] that weak solutions of partial differential equations are not always unique. In certain cases this may be remedied by introducing higher-order dissipative or dispersive terms. An example where this is the case leads to the one-dimensional scalar analog to the Navier–Stokes equation. This one-dimensional analog was introduced by Burgers[2,3] and has been studied from different points of view by Hopf[4] and Kraichnan.[5]

We consider a mixed initial and boundary-value problem for the nonlinear, inhomogeneous, one-dimensional scalar analog to the Navier–Stokes equation; and following the method of Ladyzhenskaya,[6] we prove that the weak solution is unique.

It is known that the problem under consideration can be linearized by a change of the dependent variable. The uniqueness proof given here does not depend on this fact.

## 2. FORMULATION

Let $Q_T$ represent the closed rectangular region

$$0 \leq x \leq L, \tag{2.1}$$

$$0 \leq t \leq T. \tag{2.2}$$

Consider the solution of

$$u_t + uu_x = \nu u_{xx} + f(x, t), \tag{2.3}$$

subject to the boundary condition

$$u(0, t) = u(L, t) = 0, \tag{2.4}$$

and the initial condition

$$u(x, 0) = a(x), \tag{2.5}$$

where

$$a(0) = a(L) = 0. \tag{2.6}$$

[1] P. Lax, *Nonlinear Problems* (The University of Wisconsin Press, Madison, 1963), pp. 3–12.
[2] J. M. Burgers, Proc. Acad. Sci. Amsterdam **43**, 2 (1940).
[3] J. M. Burgers, Advan. Appl. Mech. **1**, 171 (1948).
[4] E. Hopf, Commun. Pure Appl. Math. **3**, 201 (1950).
[5] R. H. Kraichnan, J. Math. Phys. **2**, 124 (1961).
[6] O. A. Ladyzhenskaya, *The Mathematical Theory of Viscous Incompressible Flow* (Gordon and Breach, Science Publishers, New York, 1963), Chap. 6.

### 3. DEFINITION OF WEAK SOLUTIONS

We define a weak solution of the problem (2.3)–(2.5) in the domain $Q_T$ to be a function $u(x, t)$ for which

$$\int_0^L u^4(x, t)\, dx < C_T, \tag{3.1}$$

for all $t \in [0, T]$, where $C_T$ is a constant; and for which the derivatives $u_x$, $u_t$ exist and are square-summable over $Q_T$. The weak solution $u(x, t)$ satisfies the boundary and initial conditions (2.4)–(2.5) and the identity

$$\int_0^T \int_0^L \{u_t\varphi + \nu u_x\varphi_x - \tfrac{1}{2}u^2\varphi_x - f\varphi\}\, dx\, dt = 0, \tag{3.2}$$

for all possible functions $\varphi(x, t)$ such that

$$\varphi(x, t) \in L_2(Q_T), \tag{3.3}$$

$$\varphi_x(x, t) \in L_2(Q_T), \tag{3.4}$$

$$\varphi_t(x, t) \in L_2(Q_T), \tag{3.5}$$

$$\varphi(0, t) = \varphi(L, t) = 0, \tag{3.6}$$

where $L_2(Q_T)$ denotes the Hilbert space of functions which are square summable over $Q_T$.

The force function $f$ in (3.2) is such that the integral involving it is well defined. Thus $f$ can be an ordinary integrable function or a generalized function such as a Dirac delta function.

### 4. RELATION BETWEEN CLASSICAL AND WEAK SOLUTIONS

*Theorem I.* If the problem (2.3)–(2.5) has a classically differentiable solution in $L_2(Q_T)$ with derivatives in $L_2(Q_T)$, then this solution is a weak solution.

*Proof:* We obtain the identity

$$\int_0^T \int_0^L \{u_t + uu_x - \nu u_{xx} - f\}\varphi\, dx\, dt = 0 \tag{4.1}$$

from (2.3). Integration of the second and third terms in (4.1) by parts making use of

$$\varphi u_{xx} = (u_x\varphi)_x - u_x\varphi_x, \tag{4.2}$$

$$\tfrac{1}{2}\varphi(u^2)_x = \tfrac{1}{2}(\varphi u^2)_x - \tfrac{1}{2}\varphi_x u^2, \tag{4.3}$$

and (3.6), yields (3.2) and the conclusion follows.

*Theorem II.* If the problem (2.3)–(2.5) has a weak solution and this weak solution has a square-summable second derivative $u_{xx}$ in the subdomain $Q_T' \subset Q_T$, then $u(x, t)$ satisfies the system (2.3)–(2.5) almost everywhere in $Q_T$.

*Proof:* Let each function $\varphi$ have compact support interior to the domain $Q_T'$ throughout which $u_{xx}$ exists. Then integrating the second and third terms of (3.2) by parts, using (4.2) and (4.3), yields

$$\int_0^T \int_0^L \{u_t + uu_x - \nu u_{xx} - f\}\varphi\, dx\, dt = 0, \tag{4.4}$$

from which follows

$$u_t + uu_x - \nu u_{xx} = f, \tag{4.5}$$

almost everywhere in $Q_T$.

### 5. UNIQUENESS OF WEAK SOLUTIONS

*Theorem III.* The problem (2.3)–(2.5) has no more than one weak solution.

*Proof:* Let $u$ and $v$ be weak solutions of (2.3)–(2.5). Then

$$\int_0^T \int_0^L \{u_t\varphi + \nu u_x\varphi_x - \tfrac{1}{2}u^2\varphi_x - f\varphi\}\, dx\, dt = 0, \tag{5.1}$$

and

$$\int_0^T \int_0^L \{v_t\varphi + \nu v_x\varphi_x - \tfrac{1}{2}v^2\varphi_x - f\varphi\}\, dx\, dt = 0. \tag{5.2}$$

Subtracting (5.2) from (5.1) gives

$$\int_0^T \int_0^L \{(u_t - v_t)\varphi + \nu\varphi_x(u_x - v_x) - \tfrac{1}{2}\varphi_x(u^2 - v^2)\}\, dx\, dt = 0. \tag{5.3}$$

Let

$$\varphi = \begin{cases} (u - v) = w & 0 \le t \le t_1 \\ 0 & t_1 \le t \le T; \end{cases} \tag{5.4}$$

then

$$\int_0^{t_1} \int_0^L \{w_t w + \nu w_x^2 - \tfrac{1}{2}ww_x(w + 2v)\}\, dx\, dt = 0. \tag{5.5}$$

An application of (2.4) reduces (5.5) to

$$\int_0^{t_1} \int_0^L \{ww_t + \nu w_x^2 - \nu ww_x\}\, dx\, dt = 0. \tag{5.6}$$

Define

$$||w(x, t_1)||_2^2 = \int_0^L w^2(x, t_1)\, dx, \tag{5.7}$$

and

$$||w_x(x, t)||_2^2 = \int_0^L w_x^2(x, t)\, dx, \tag{5.8}$$

in terms of which (5.6) becomes

$$\tfrac{1}{2}\,||w(x,\,t_1)||_2^2 + \nu\int_0^{t_1} ||w_x||_2^2\,dt - \int_0^{t_1}\int_0^L www_x\,dx\,dt = 0. \quad (5.9)$$

Two successive applications of Schwarz's inequality to the first integral in the last term of (5.9) gives

$$\int_0^L www_x\,dx \le C_T^{\frac{1}{2}}\,||w_x||_2\,||w||_4, \quad (5.10)$$

where

$$||w||_4^4 = \int_0^L w^4(x,\,t)\,dx. \quad (5.11)$$

It follows from a result obtained by Ladyzhenskaya[7] that

$$||w||_4 \le \epsilon\,||w_x||_2 + C_\epsilon\,||w||_2, \quad (5.12)$$

where $\epsilon$ is any *arbitrary* $\epsilon > 0$; and $C_\epsilon$ is a positive number determined by $\epsilon$, with $\lim_{\epsilon\to 0} C_\epsilon = \infty$. Combining (5.12) and (5.10) gives

$$\int_0^L www_x\,dx \le C_T^{\frac{1}{2}}\{\epsilon\,||w_x||_2^2 + C_\epsilon\,||w_x||_2\,||w||_2\}, \quad (5.13)$$

and completing the square in (5.13) leads to the inequality

$$\int_0^L www_x\,dx \le C_T^{\frac{1}{2}}\left\{\left(2\epsilon\,||w_x||_2^2 + \frac{C_\epsilon^2}{4\epsilon}\,||w||_2^2\right)\right\}. \quad (5.14)$$

Let

$$C_T^{\frac{1}{2}} = C_1, \quad (5.15)$$

and

$$C_2 = C_1(C_\epsilon^2/4\epsilon); \quad (5.16)$$

then (5.14) and (5.9) yield

$$\tfrac{1}{2}\,||w(x,\,t_1)||_2^2 + \nu\int_0^{t_1} ||w_x||_2^2\,dt \le C_1\int_0^{t_1}\left(2\epsilon\,||w_x||_2^2 + \frac{C_\epsilon^2}{4\epsilon}\,||w||_2^2\right)dt. \quad (5.17)$$

[7] Reference 6, pp. 10–12.

Notice that $\epsilon$ is an *arbitrary positive* number, and since $C_1 > 0$, it is permissible to choose $\epsilon$ so that

$$\epsilon = \nu/2C_1, \quad (5.18)$$

then (5.17) reduces to

$$\tfrac{1}{2}\,||w(x,\,t_1)||_2^2 \le C_2\int_0^{t_1} ||w||_2^2\,dt, \quad (5.19)$$

where $C_2 > 0$. Let

$$y(t_1) = \int_0^{t_1} ||w||_2^2\,dt, \quad (5.20)$$

then (5.19) becomes

$$dy(t_1)/dt_1 \le 2C_2y(t_1), \quad (5.21)$$

which can be integrated to give

$$y(t_1) \le y(t_0)\,\exp 2C_2(t_1 - t_0), \quad (5.22)$$

with

$$0 \le t_0 \le t_1 \le T. \quad (5.23)$$

Both $t_0$ and $t_1$ are arbitrary in $[0,\,T]$. Choose $t_0 = 0$, then since $y(0) = 0$ from (5.20) it follows from (5.22) that for arbitrary $t_1$ in $[0,\,T]$

$$y(t_1) \le 0. \quad (5.24)$$

Conversely, (5.20) implies

$$y(t_1) \ge 0; \quad (5.25)$$

hence

$$y(t_1) \equiv 0, \quad (5.26)$$

for arbitrary $t_1$ in $[0,\,T]$, and therefore the solutions $u$ and $v$ must coincide. This proves the theorem.

Note that the validity of the uniqueness proof depends on (5.12) which requires $\epsilon$ to be strictly positive. As a consequence, uniqueness means that the parameter $\nu$ playing the role of a kinematic viscosity and related to $\epsilon$ by (5.18) must be strictly positive.

# Exact Quantization Rules for the One-Dimensional Schrödinger Equation with Turning Points*

PAUL B. BAILEY
*Sandia Laboratory, Albuquerque, New Mexico*
(Received 7 April 1964)

It is pointed out that, for a number of problems, exact quantization rules exist which closely resemble that of Bohr–Wilson–Sommerfeld. In some cases it is shown how these rules may be derived mathematically from the Schrödinger equation.

## 1. INTRODUCTION

THE eigenvalues (point spectrum) of the equation

$$\psi''(z) - (2m/\hbar^2)[V(z) - E]\psi(z) = 0 \qquad (1.1)$$

are frequently estimated by the Bohr–Wilson–Sommerfeld quantization rule

$$\int_{k_1(E)}^{k_2(E)} Q(z)\,dz = (n + \tfrac{1}{2})\pi i \qquad (1.2)$$

in those cases where

$$Q^2(z) = (2m/\hbar^2)[V(z) - E] \qquad (1.3)$$

has two simple zeros (turning points) $k_1(E) < k_2(E)$. Although (1.2) is generally supposed[1] to give good approximations for only the larger eigenvalues, it is very well known that in the case of the harmonic oscillator it happens to give *all* the eigenvalues *exactly*. That is, (1.2) is an exact quantization rule in this one case.

Another curious and very well known fact about (1.2) concerns what we will call the radial wave equation. In this case the function $V(z)$ appearing in (1.1) contains a parameter $L$, an integer, in the form $L(L + 1)$. And the use of (1.2) gives approximating expressions for the eigenvalues $E$ which *become exact* ones upon replacing $L(L + 1)$ by $(L + \frac{1}{2})^2$. In other words, an exact quantization rule for the radial wave equation is obtainable in the form

$$\int_{\tilde{k}_1}^{\tilde{k}_2} \tilde{Q} = (n + \tfrac{1}{2})\pi i, \qquad (1.4)$$

where $\tilde{Q}$ is the function obtained from $Q$ in this case by replacing $L(L + 1)$ by $(L + \frac{1}{2})^2$. $\tilde{k}_1$ and $\tilde{k}_2$ are the zeros of $\tilde{Q}$.

Apparently it has not previously been noticed that exact quantization rules can be had in a number of other cases as well. In each of these cases one can find a suitable function $\tilde{Q}$ for which (1.4) is an exact rule; or, alternatively, one can find a number $\alpha$ and a contour $\Gamma$ such that the rule

$$-\frac{1}{2\pi i}\oint_\Gamma Q = n + \alpha \qquad (1.5)$$

is exact.

Actually one can go just a little further and give a single rule which is exact in all these cases. Namely,

$$\oint \left(Q + \frac{Q'}{2Q} + \frac{\psi'}{\psi}\right) = 0, \qquad (1.6)$$

where the integral is along any simple contour enclosing all the singularities of $Q$ in the finite part[2] of the complex plane.

In Sec. 2 we have listed all the essentially different cases we are aware of for which the differential equation (1.1) has at least two turning points and which can be solved explicitly in terms of known functions. It will be seen that (1.6) is exact in these cases. (We hasten to point out, however, that we do not expect the rule to be exact in *all* cases.) Rule (1.4) is dealt with in Sec. 3.

In Sec. 4 we take up the obvious question of how one could derive (1.6) and (1.4) directly from (1.1) without any prior knowledge of the nature of its solutions. We are able to do this for at least the harmonic oscillator and the radial wave equation by simply determining the asymptotic behavior of the solutions of (1.1). Although the method is a fairly general one it is not obvious how all the other cases could be treated.

Section 5 is a discussion of some of the obvious questions raised by this work.

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

[1] P. M. Morse and H. Feshback, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953). For a mathematical proof, however, see E. C. Titchmarsh, Quart. J. Math. 5, 228 (1954).

[2] In two of the cases we have in mind $Q^2$ is a periodic function, and in such cases the contour is to enclose only the singularities in one period strip.

## 2. THE SPECIAL CASES[3]

(i) The harmonic oscillator:

$$Q^2(z) = (2m/\hbar^2)(z^2 - E), \qquad \psi(-\infty) = 0 = \psi(\infty). \tag{2.1}$$

Eigenvalues

$$E_n = (2n + 1)[\hbar/(2m)^{\frac{1}{2}}]. \tag{2.2}$$

(ii) The radial wave equation:

$$Q^2(z) = \frac{a^2D + L(L+1)}{z^2} - \frac{2a^2D}{z} - a^2E, \qquad \psi(0) = 0 = \psi(+\infty). \tag{2.3}$$

Eigenvalues

$$E_n = -a^2D^2/(n + \alpha)^2, \qquad \alpha = \tfrac{1}{2} + [a^2D + (L + \tfrac{1}{2})^2]^{\frac{1}{2}}. \tag{2.4}$$

(iii) Morse oscillator:

$$Q^2(z) = (2m/\hbar^2)(Ae^{-2az} - 2Ae^{-az} - E), \qquad \psi(-\infty) = 0 = \psi(\infty). \tag{2.5}$$

Eigenvalues

$$E_n = -A\left(1 - \frac{a\hbar}{(2mA)^{\frac{1}{2}}}(n + \tfrac{1}{2})\right)^2. \tag{2.6}$$

(iv) $$Q^2(z) = \frac{2m}{\hbar^2}\left[v\left(\frac{a}{z} - \frac{z}{a}\right)^2 - E\right], \qquad \psi(0) = 0 = \psi(\infty). \tag{2.7}$$

Eigenvalues

$$E_n = \frac{v^{\frac{1}{2}}}{a}\left(\frac{\hbar}{(2m)^{\frac{1}{2}}}(4n + 1 + 2\alpha) - 2av^{\frac{1}{2}}\right), \tag{2.8}$$

where

$$\alpha = \tfrac{1}{2} + (2ma^2v/\hbar^2 + \tfrac{1}{4})^{\frac{1}{2}}.$$

(v) $$Q^2(z) = (2m/\hbar^2)[V\cot^2(\pi z/)a - E], \qquad \psi(0) = 0 = \psi(a). \tag{2.9}$$

Eigenvalues

$$E_n = V + (\hbar^2/2m)(n + \alpha)^2(\pi^2/a^2), \tag{2.10}$$

where

$$\alpha = \tfrac{1}{2} + (2ma^2V/\hbar^2\pi^2 + \tfrac{1}{4})^{\frac{1}{2}}.$$

(vi) $$Q^2(z) = -(2m/\hbar^2)[V\,\mathrm{sech}^2(z/a) - E], \qquad \psi(-\infty) = 0 = \psi(\infty). \tag{2.11}$$

Eigenvalues

$$E_n = -(\hbar^2/2ma^2)(n + \alpha)^2, \tag{2.12}$$

where

$$\alpha = \tfrac{1}{2} + (2ma^2V/\hbar^2 + \tfrac{1}{4})^{\frac{1}{2}}.$$

We do not verify the quantization rules in detail for all these cases, but merely outline the steps needed to check (1.5) and (1.6) in Case (ii). The others are no more difficult.

In Case (ii) the singularities of $Q$ in the finite part of the complex plane are branch points (square roots) at the two turning points, which are on the positive real axis, and a simple pole at the origin with residue $-[a^2D + L(L+1)]^{\frac{1}{2}}$. Our contour, then, is any simple closed curve containing these three points.

The $n$th eigenfunction has $n$ zeros between the two turning points and has the form $z^\alpha \sum_0^\infty a_n z^n$ at the origin, where $\alpha = \tfrac{1}{2} + [a^2D + (L + \tfrac{1}{2})^2]^{\frac{1}{2}}$. Hence

$$\frac{1}{2\pi i}\oint \frac{\psi'}{\psi} = n + \alpha, \quad \text{while} \quad \frac{1}{2\pi i}\oint \frac{Q'}{2Q} = 0,$$

and

$$-\frac{1}{2\pi i}\oint Q = 2\cdot\frac{1}{2\pi}\int_{k_1}^{k_2} |Q| + \text{Residue of } Q \text{ at the origin}$$
$$= \{aD/(-E)^{\frac{1}{2}} - [a^2D + L(L+1)]^{\frac{1}{2}}\} + [a^2D + L(L+1)]^{\frac{1}{2}}$$
$$= aD/(-E)^{\frac{1}{2}}.$$

From these results and (2.4) it is evident that both (1.5) and (1.6) hold.

## 3. THE FUNCTIONS $\tilde{Q}(z)$

Since, in may of the books on quantum mechanics, attempts are made to "justify" (1.2) by means of arguments which use the functions

$$w_\pm(z) = \frac{1}{[Q(z)]^{\frac{1}{2}}}\exp\left(\pm\int^z Q\right) \tag{3.1}$$

as approximants in some sense for the eigenfunctions $\psi(z)$, and since in at least some cases (1.2) can be improved by replacing $Q(z)$ with a suitable $\tilde{Q}(z)$, it seems worthwhile to indicate very briefly why this should be expected.

The functions $w_\pm$ are solutions of the differential equation

$$w'' - (Q^2 + R)w = 0, \tag{3.2}$$

which will resemble our wave equation (1.1) wherever

[3] These equations and their solutions are all to be found in I. Gol'dman and V. Krivchenkov, *Problems in Quantum Mechanics* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1960), as are the evaluations of the more awkward integrals needed for verifying the quantization rules.

$$R = (Q'/2Q)^2 - (Q'/2Q)' \tag{3.3}$$

is small compared to $Q^2$. In the case of the harmonic oscillator, $R$ is small at $\infty$, the only singular point of the differential equation (1.1). However in Case (ii), $R$ is small (relative to $Q^2$) at $\infty$ but not at the origin. [These are the two singular points of (1.1) in this case.] For at the origin, $R$ and $Q^2$ are both $O(1/z^2)$.

If, on the other hand, we compare (1.1) with

$$\tilde{w}'' - (\tilde{Q}^2 + \tilde{R})\tilde{w} = 0, \tag{3.4}$$

where

$$\tilde{Q}^2(z) = \frac{a^2D + (L + \frac{1}{2})^2}{z^2} - \frac{2a^2D}{z} - a^2E, \tag{3.5}$$

and $\tilde{w}$, $\tilde{R}$ are defined by (3.1) and (3.3) with $\tilde{Q}$ in place of $Q$, it will be seen that $\tilde{Q}^2 + \tilde{R}$ is nearly equal to $Q^2$ both at $\infty$ and at 0. Consequently there is good reason for expecting the solutions of (3.4) to approximate the solutions of (1.1). Note that this amounts, formally, to replacing $L(L + 1)$ by $(L + \frac{1}{2})^2$.

Exactly the same considerations apply to the other cases, leading to the following choices of $\tilde{Q}$.

Case (iii): $\tilde{Q}^2(z) = Q^2(z)$. In other words, (1.2) is exact in this case.

Case (iv):

$$\tilde{Q}^2(z) = \left(\frac{2m}{\hbar^2}Va^2 + \frac{1}{4}\right)z^{-2} + \frac{2m}{\hbar^2}\left(\frac{V}{a^2}z^2 - 2V - E\right).$$

Case (v):

$$\tilde{Q}^2(z) = \left(\frac{2m}{\hbar^2}V + \frac{a^2}{4\pi^2}\right)\csc^2\frac{\pi z}{a} - \frac{2m}{\hbar^2}(V + E);$$

the singular points of (1.1) are 0, $a$ (mod $a$), and $\infty$.

Case (vi):

$$\tilde{Q}^2(z) = -\left(\frac{2mV}{\hbar^2} + \frac{1}{4a^2}\right)\operatorname{sech}^2\frac{z}{a} - \frac{2m}{\hbar^2}E;$$

the singular points are $\pm a\frac{1}{2}\pi i$ (mod $a\pi i$) and $\infty$.

Verification of (1.4) in these cases is straightforward and is omitted.

## 4. DERIVATION OF (1.6) FROM (1.1)

First we deal with the harmonic oscillator.

Recall the discussion in Sec. 3 concerning the functions $w_\pm$ and their differential equation (3.2), with now

$$Q^2(z) = (2m/\hbar^2)(z^2 - E), \tag{4.1}$$

and hence

$$R(z) = (3z^2 + 2E)/4(z^2 - E)^2. \tag{4.2}$$

Rewriting (1.1) in the form

$$\psi'' - (Q^2 + R)\psi = -R\psi, \tag{4.3}$$

we can expect that the solution to the integral equation

$$\psi(z) = w(z) - \frac{1}{2}\int_z^b [w_-(z)w_+(y) - w_+(z)w_-(y)]R(y)\psi(y)\,dy \tag{4.4}$$

will satisfy (4.3) if $w$ is any solution of (3.2). Of course the integral here (and elsewhere) must be understood to be a contour integral and the contour must be suitably chosen. We will use (4.4), with suitable choices for $w$ and $b$, to obtain information about the solutions of (1.1) which will allow us to deduce (1.6).

First define a sequence $\psi_n$ by

$$\psi_0 = w_-, \tag{4.5}$$

and

$$\psi_n(z) = w_-(z) - \frac{1}{2}\int_z^\infty [w_-(z)w_+(y) - w_+(z)w_-(y)]R(y)\psi_{n-1}(y)\,dy \tag{4.6}$$

for $n \geq 1$. (It will be seen that these definitions make sense.) Rewrite (4.6) as

$$\frac{\psi_n}{w_-}(z) - 1 = -\frac{1}{2}\int_z^\infty \frac{R}{Q}(y) \times \left[1 - \exp\left(-2\int_z^y Q\right)\right]\frac{\psi_{n-1}}{w_-}(y)\,dy, \tag{4.7}$$

and note that

$$\frac{\psi_{n+1}}{w_-}(z) - \frac{\psi_n}{w_-}(z) = -\frac{1}{2}\int_z^\infty \frac{R}{Q}(y)\left[1 - \exp\left(-2\int_z^y Q\right)\right] \times \left[\frac{\psi_n}{w_-}(y) - \frac{\psi_{n-1}}{w_-}(y)\right]dy \tag{4.8}$$

for $n \geq 1$.

If, now, we can choose a contour $C_z$ running from $z$ to $+\infty$ in such a way that

$$\left|\exp\left(-2\int_z^y Q\right)\right| \leq 1 \quad \text{for} \quad y \in C_z, \tag{4.9}$$

we will have the inequality

$$\left|\frac{\psi_{n+1}}{w_-}(z) - \frac{\psi_n}{w_-}(z)\right|$$

$$\leq \int_{C_z} \left|\frac{R}{Q}(y)\right| \left|\frac{\psi_n}{w_-}(y) - \frac{\psi_{n-1}}{w_-}(y)\right| |dy|$$

$$\leq M(z) \sup_{y \in C_z} \left|\frac{\psi_n}{w_-}(y) - \frac{\psi_{n-1}}{w_-}(y)\right|, \tag{4.10}$$

where

$$M(z) = \int_{C_z} \left|\frac{R}{Q}(y)\right| |dy|$$

$$= O(|z|^{-2}) \quad \text{for large} \quad |z|.$$

In particular, if we can find any region $D$ such that for each $z \in D$ there is a contour $C_z$ lying in $D$ for which (4.9) is valid and $M(z) \leq 1$, then it follows from (4.10) that

$$(\psi_{n+1}/w_-)(z) - (\psi_n/w_-)(z) \to 0$$

uniformly in $D$ as $n \to \infty$,

and that

$$\psi_n/w_- \to \psi/w_- \quad \text{uniformly in } D$$

with $\psi/w_-$ uniformly bounded there. Letting $n \to \infty$ in (4.7) shows that $\psi$ is a solution of (4.4), and using the upper bound of $\psi/w_-$ on the right side of (4.4), we obtain

$$\frac{\psi}{w_-}(z) = 1 + O(|z|^{-2}) \quad \text{uniformly in } D. \tag{4.11}$$

However we must show that $D$ exists (nonempty). It is a simple matter to verify that

$$\exp\left(-2\int_z^\nu Q\right) = \exp\{-[yQ(y) - zQ(z)]\} \times \left(\frac{y + (y^2 - E)^{\frac{1}{2}}}{z + (z^2 - E)^{\frac{1}{2}}}\right)^{E[(2m)^{\frac{1}{2}}/h]},$$

and that $|z + (z^2 - E)^{\frac{1}{2}}| = \text{constant} > E^{\frac{1}{2}}$ is an ellipse in the complex $z$ plane with foci at $\pm E^{\frac{1}{2}}$. Consequently if $R[z] \geq 0$ and $|z + (z^2 - E)^{\frac{1}{2}}| = A > E^{\frac{1}{2}}$, we can define $C_z$ as that contour running from $z$ along the ellipse $|z + (z^2 - E)^{\frac{1}{2}}| = A$ to the positive real axis and thence to $+\infty$. For $y \in C_z$ it turns out that $|\text{angle } [yQ(y) - zQ(z)]| \leq \frac{1}{2}\pi$ with strict inequality at least whenever $y$ is on the real axis with $y \neq z$. Hence $D$ exists and contains all large $z$ for which $R[z] \geq 0$.

To sum up, we have shown that (1.1) has in $D$ a solution $\psi$ with asymptotic behavior given by (4.11). In particular, $\psi(+\infty) = 0$.

In an exactly similar fashion, by taking $b = \infty e^{\pi i}$ and $w = w_-$, we can show the existence of a solution $\varphi$ of (1.1) in a region $D'$ which contains all large $z$ with $R[z] \leq 0$ (or, rather, $\frac{1}{2}\pi \leq \text{angle } z \leq \frac{3}{2}\pi$), and which has the asymptotic behavior

$$(\varphi/w_-)(z) = 1 + O(|z|^{-2}) \quad \text{uniformly in } D'. \tag{4.12}$$

By taking $b = \infty e^{\frac{1}{2}\pi i}$ and $w = w_+$ we find a solution $\theta$ of (1.1) in a region $D''$ which contains all large $z$ with $0 \leq \text{angle } z \leq \pi$ and has asymptotic behavior

$$(\theta/w_+)(z) = 1 + O(|z|^{-2}) \quad \text{uniformly in } D''. \tag{4.13}$$

In particular, $\theta$ does *not* vanish at either $+\infty$ or $\infty e^{\pi i}$.

An important consequence of this last remark is that every solution of (1.1) which vanishes at $+\infty$ is some constant multiple of the function $\psi$ obtained above, and any solution which vanishes at $\infty e^{\pi i}$ is some multiple of the function $\varphi$.

Now, suppose that $E$ is an eigenvalue of (1.1) with $\psi$ the corresponding eigenfunction. In other words, suppose that $\psi$ vanishes not only at $+\infty$ but also at $\infty e^{\pi i}$. Then $\psi = K\varphi$ for some constant $K$. Hence, for any large $z$ on the upper imaginary axis (such $z \in D$ and $D'$ both), we have both

$$(\psi/w_-)(z) = 1 + O(|z|^{-2}),$$

and

$$(\psi/Kw_-)(z) = 1 + O(|z|^{-2}),$$

which implies $K = 1$ and so (4.11) holds uniformly in $D \cup D'$.

But we may assume any eigenfunction of (1.1) is either even or odd (an easy thing to check), so that

$$\psi(xe^{\pi i}) = \pm\psi(x) \tag{4.14}$$

for any large real $x$. And at the same time,

$$w_-(xe^{\pi i}) = \frac{1}{[Q(xe^{\pi i})]^{\frac{1}{4}}} \exp\left(-\int^{xe^{\pi i}} Q\right)$$

$$= \frac{1}{[Q(x)]^{\frac{1}{4}}} \exp(-\tfrac{1}{2}\pi i) \exp\left(-\int^{x} Q\right) \exp\left(-\int^{xe^{\pi i}} Q\right)$$

$$= \exp\left(-\frac{\pi}{2}i - \frac{1}{2}\int_\Gamma Q\right) w_-(x), \tag{4.15}$$

where $\Gamma$ is any simple closed curve containing the two turning points $\pm E^{\frac{1}{2}}$. Using (4.14) and (4.15) in (4.11) with $z = xe^{\pi i}$ gives

$$-\frac{1}{2\pi i}\int_\Gamma Q = n + \frac{1}{2}$$

for some nonnegative integer $n$.

For the radial wave equation the argument is essentially the same as in the previous case, although naturally some of the details differ. By considering the integral equation (4.4) with $b = +\infty$ and $w = w_-$, we show the existence of a solution $\psi$ of (1.1) with asymptotic behavior

$$(\psi/w_-)(z) = 1 + O(|z|^{-2}) \quad \text{in a region } D \tag{4.17}$$

which includes all large $z$ with $0 \leq \text{angle } z \leq \pi$. Taking $b = \infty e^{2\pi i}$ and $w = w_-$ we find a solution $\varphi$ of (1.1) with asymptotic behavior

$$(\varphi/w_-)(z) = 1 + O(|z|^{-2}) \quad \text{in a region } D' \tag{4.18}$$

which includes all large $z$ with $\pi \leq \text{angle } z \leq 2\pi$. And by considering $b = \infty e^{\pi i}$ and $w = w_+$, we conclude that any solution of (1.1) which vanishes at $+\infty$ (or $\infty e^{2\pi i}$) is some multiple of $\psi$ (or $\varphi$).

But of course any solution of (1.1) which vanishes at the origin is of the form $z^\alpha \sum_0^\infty a_n z^n$, $\alpha = \frac{1}{2} + [a^2 D + (L + \frac{1}{2})^2]^{\frac{1}{2}}$. Consequently, if $E$ is an eigenvalue, so that $\psi(0) = 0$, then

$$\psi(xe^{2\pi i}) = e^{2\alpha\pi i}\psi(x) \tag{4.19}$$

for large real $x$, and also

$$\psi(z) = K\varphi(z) \tag{4.20}$$

for some constant $K$. As before we deduce $K = 1$, since $D$ and $D'$ overlap, and so (4.17) holds in $D \cup D'$. Hence, using (4.19) and

$$w_-(xe^{2\pi i}) = \exp\left(\int_\Gamma Q\right) w_-(x) \tag{4.21}$$

where $\Gamma$ is any simple closed curve containing the origin and the two turning points, we deduce the quantization rule

$$-\frac{1}{2\pi i}\int_\Gamma Q = \alpha + n \quad \text{for some integer } n. \tag{4.22}$$

(It turns out that $n$ cannot be negative because of the sign of $Q$.)

Obviously the same procedure can be used to derive (1.4) for these two cases. In fact we could obtain (1.6) for fairly general $Q$, provided only that it bear a sufficiently close resemblance to either (i) or (ii). We omit the details.

## 5. DISCUSSION

Since in the case of the anharmonic oscillator $Q^2(z) = (2m/\hbar^2)(z^4 - E)$ our rule (1.6) is completely equivalent to the Bohr–Wilson–Sommerfeld rule (1.2), which is only approximately correct[4] in this case, we cannot expect it to be exact in all cases. The most one could hope for is that (1.6) be always at least as good as (1.2).

There is one example we know[5] of for which solutions can be obtained only numerically and for which our rule (1.6) differs from (1.2); namely the case

$$Q^2(z) = (2000e^{-3.4(x-1.3)} - 4000e^{-1.7(x-1.3)} + 2/z^2 - E).$$

The ground state $E_0$ has been computed[4] to be $-1923.5296551$. We evaluated the left side of (1.2) using this number of $E$ and obtained 1.57079634. This is remarkably close to (1.2) and distressingly far from (1.6), which is equivalent in this case to (1.2) with $(2 - 2^{\frac{1}{2}})\pi$ in place of $\frac{1}{2}\pi$. Perhaps it would be worthwhile to recompute $E_0$.

In any event, it would be of interest to know for which problems rules like (1.2) or (1.6) hold. Obviously the treatment in Sec. 4 could be applied to fairly general functions $Q^2$ which closely resemble either the harmonic oscillator or the radial wave equation, always obtaining exact rules.

*Note added in proof*: We have now verified that the stated value of $E_0$ is correct, which implies that (1.2) is better in this case than (1.4) or (1.6). That (1.2) gives such a good approximation is probably due to the large coefficients of the exponentials.

---

[4] D. Secrest, K. Cashion, and J. O. Hirschfelder, J. Chem. Phys. **37**, 830 (1962).

[5] The author is indebted to Dr. C. J. MacCallum for bringing a number of papers to his attention.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# Coupled Magnetomechanical Equations for Magnetically Saturated Insulators

H. F. TIERSTEN
*Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey*
(Received 24 March 1964)

The differential equations and boundary conditions governing the macroscopic behavior of nonconducting magnetically saturated media undergoing large deformations, are derived by means of a systematic and consistent application of the laws of continuum physics to a model consisting of an electronic spin continuum coupled to a lattice continuum. The macroscopic effect of the quantum mechanical exchange interaction is included as are dissipation and the associated thermodynamics. The resulting nonlinear equations are specialized to the important case of a small dynamic field superposed on a large static biasing field. Only the linear approximation in the small-field variables is obtained. This final system of linear equations permits the solution of a variety of magnetomechanical boundary-value problems.

## 1. INTRODUCTION

ON the one hand there exists the macroscopic theory of elasticity,[1,2] and on the other what may be called the macroscopic theory of the magnetodynamics of saturated magnetic media.[3-5] The former governs, among other things, the propagation of waves in and the vibrations of an elastic solid and the latter such phenomena as the frequency dependence of the magnetic susceptibility, ferromagnetic resonance, and the propagation of spin waves. In recent years much interest has centered on the magnetostrictive coupling of these two systems[6-8]; but all of these papers use results from classical elasticity theory although it is clear that certain assumptions which are made in classical elasticity theory are not satisfied in such a coupled system. None of the papers go back to the fundamental laws from which classical elasticity theory is derived to see what basic changes result. In addition, all of these papers assume infinitesimal strain at the outset and none of them formulate boundary-value problems.

Now, it is well known that any theory of the mechanical behavior of deformable continuous media (e.g., elasticity) may be obtained by applying the laws of the conservation of mass, linear momentum, angular momentum and energy to an arbitrary element of matter and making suitable constitutive assumptions (e.g., Hooke's law). It is also known that the macroscopic theory of the magnetodynamics of saturated magnetic media may be obtained by applying the law of conservation of angular momentum to an arbitrary region containing electron spins and using the appropriate electromagnetic equations. The application of the law of conservation of angular momentum separately to the two systems should make clear the need to state the fundamental laws and apply them to the coupled system in a consistent manner.

All of the aforementioned considerations are of a macroscopic nature and all discrete microscopic detail has been expressly avoided. Such an approach is readily justifiable since a detailed, consistent microscopic theory leading to the macroscopic equations is not presently available nor does it appear that one will be in the near future. Furthermore, it is well known that a knowledge of the detailed behavior of the particles is, in such instances, not necessary for the determination of the macroscopic field variables. However, useful conceptual macroscopic information may be obtained from a knowledge of microscopic phenomena (e.g., the electronic angular momentum character of the magnetic moment), and this macroscopic information may be incorporated in a continuum theory in a manner consistent with the basic laws of continuum physics. This is the point of view adopted in this paper.

In this paper the field equations for the coupled system are derived by means of a systematic

[1] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Cambridge University Press, Cambridge, England, 1927) 4th ed. (also Dover Publications, Inc., New York, 1944).
[2] C. Truesdell and R. A. Toupin, "The Classical Field Theories" in *Encyclopedia of Physics*, edited by S. Flügge (Springer-Verlag, Berlin, 1960), Vol. III.
[3] J. Smit and H. P. J. Wijn, *Ferrites* (John Wiley & Sons, Inc., New York, 1959).
[4] R. F. Soohoo, *Theory and Application of Ferrites* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1960).
[5] A. G. Gurevich, *Ferrites at Microwave Frequencies*, (translated from the Russian by A. Tybulewicz) (Consultants Bureau, New York, 1963).
[6] C. Kittel, Phys. Rev. **110**, 836 (1958).
[7] A. I. Akhiezer, V. G. Bar'iakhtar, and S. V. Peletminskii, Zh. Eksperim. i Teor. Fiz. **35**, 228 (1958) [English transl.: Soviet Physics—JETP **8**, 157 (1959)].
[8] K. B. Vlasov, Phys. Metals Metal Res. **5**, 385 (1957); Izv. Akad. Nauk. SSSR, Ser. Fiz., **22**, 1159 (1958) (Columbia Tech. Transl. 1149–1157).

application of the laws of continuum physics. The medium being considered is carefully described and the macroscopic field vectors involved in the analysis are carefully defined. No assumption of infinitesimal strain is imposed at the outset. Consequently, the resulting nonlinear equations are valid for finite deformation and large values of the dynamic magnetization. The equations are also specialized to the important case of a small dynamic field superposed on a large static biasing field. Only the linear approximation in the small-field variables is considered. The jump (boundary) conditions at a moving surface of discontinuity are obtained. It is assumed, however, that the frequency–wavelength combinations with which we are concerned are far outside the range associated with electromagnetic propagation, so that the electromagnetic equations may be replaced by the quasimagnetostatic equations. Moreover, as already implied, the material is assumed to be magnetically saturated, so that the direction of the magnetic moment changes but the magnitude does not.

## 2. DESCRIPTION OF THE CONTINUUM

It is well known from microscopic particle physics that a magnetic moment $\mathfrak{M}$ possesses angular momentum $\mathbf{J} = \mathfrak{M}/\gamma$, where the coefficient $\gamma$ depends on the particular material. It is also well known that a magnetic field $\mathbf{H}$ acting on a magnetic moment $\mathfrak{M}$ produces a couple equal to $\mathfrak{M} \times \mathbf{H}$. Furthermore, the application of the conservation of angular momentum to the system gives the magnetodynamic equation $d\mathfrak{M}/dt = \gamma \mathfrak{M} \times \mathbf{H}$. In addition it is known from quantum theory that there is an interaction energy in the microscopic Hamiltonian[9] which has no classical analogue and which is referred to as exchange energy. It has further been established, by Herring and Kittel,[10] assuming only nearest-neighbor interactions between adjacent spins, that the macroscopic effect of this exchange interaction may be represented by an effective magnetic field $\mathbf{H}^e$ which produces a couple $\mathfrak{M} \times \mathbf{H}^e$ acting on the magnetic moment $\mathfrak{M}$.

In any continuum theory of the mechanical behavior of deformable media (e.g., elasticity theory), two distinct types of forces act, body forces $\mathbf{f}$ and contact forces $\mathbf{t}$.[2] The body forces arise as a result of some distant action from within or outside the body. They are long-range forces. The contact forces arise as a result of the contact of adjacent elements of a body. Macroscopically speaking, they are surface forces. Microscopically, they are caused by very short-range near-neighbor interactions between adjacent microscopic elements. As is well known, these contact forces result in the mechanical stress tensor. Note the microscopic similarity between this surface force $\mathbf{t}$ and the effective magnetic field $\mathbf{H}^e$ obtained by Herring and Kittel. In many materials there can act body couples,[2,11,12] and contact couples too.[2,13–15] The contact couples will be expressly excluded from this treatment since they are deemed to be small.[16] The body couples, however, are known to be important in the media with which this paper is concerned. Indeed, as has already been mentioned, such couples ($\mathfrak{M} \times \mathbf{H}$) are precisely the quantities which appear in the aforementioned magnetodynamic equation. Such body couples are taken to be zero in classical elasticity theory.[1] This assumption, upon application of the conservation of angular momentum, results in the symmetry of the stress tensor in that theory. Thus, it should be clear that in magnetoelastic media the stress tensor will not be symmetric; and consequently, the use of results from classical elasticity theory and magnetodynamic theory in coupling the two fields could readily lead to inconsistencies. That is to say, it seems to this author that a theory governing such a combined magnetomechanical system should be obtained by means of a consistent application of fundamental principles to a well-defined macroscopic model. The description of such a model follows.

In view of the previous statements, it seems natural to define two distinct interacting continua. One of these continua will be called the electronic spin continuum and the other the lattice continuum. The electronic spin continuum possesses at each point a macroscopic magnetic moment per unit volume $\mathbf{M}$ and angular momentum $\mathbf{M}/\gamma$ ($\gamma$ a negative number), and it cannot move (translate) with respect to the lattice continuum at that point. It interacts with the local[17] lattice continuum by means of an effective local magnetic field $\mathbf{H}^L$ which exerts a couple per unit volume on the magnetization vector $\mathbf{M}$ by means of the recipe $\mathbf{M} \times \mathbf{H}^L$. Note that by

[9] J. H. Van Vleck, Rev. Mod. Phys. **17**, 27 (1945).
[10] C. Herring and C. Kittel, Phys. Rev. **81**, 869 (1951).
[11] R. A. Toupin, J. Ratl. Mech. Anal. **5**, 849 (1956).
[12] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, Inc., New York, 1960), p. 144.
[13] E. Cosserat and F. Cosserat, *Théorie des Corps Déformable* (Hermann & Cie., Paris, 1909).
[14] E. L. Aero and E. V. Kuvshinskii, Fizika Tverd. Tela **2**, 1399 (1960) [English transl. Soviet Physics—Solid State **2**, 1272 (1961)].
[15] E. S. Rajagopal, Ann. Physik **6**, 192 (1960).
[16] R. D. Mindlin and H. F. Tiersten, Arch. Ratl. Mech. Anal. **11**, 415 (1962).
[17] The word local is used throughout to mean something occupying the same region of space.

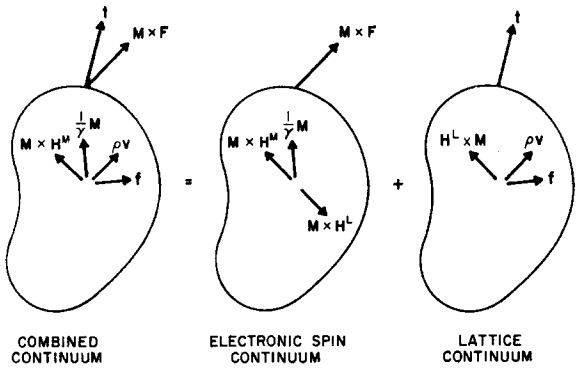


FIG. 1. Schematic diagram showing the linear and angular momentum and force and couple vectors acting in the continua.

virtue of the definition of $\mathbf{H}^L$, without loss of generality we may take $\mathbf{M}\cdot\mathbf{H}^L = 0$. Since $\mathbf{M}\times\mathbf{H}^L$ is a couple exerted by the local lattice continuum on the spin continuum, and angular momentum is conserved, an equal and opposite couple $\mathbf{H}^L\times\mathbf{M}$ must be exerted by the spin continuum on the local lattice continuum. In addition to the couple caused by the lattice, the electronic spin continuum experiences couples caused by the ordinary Maxwellian magnetic field $\mathbf{H}^M$ and couples caused by an effective exchange field $\mathbf{F}$, which will be assumed to act on an element of surface area since it is due to a near-neighbor microscopic interaction. This will be discussed thoroughly in a later section. The lattice continuum is an ordinary mechanical continuum which experiences the aforementioned body couple $\mathbf{H}^L\times\mathbf{M}$ from its interaction with the spin continuum as well as the body forces experienced by the spin continuum, which forces are transferred directly from the spin continuum to the lattice continuum because of the assumption of no relative motion of the two continua. By virtue of the same assumption any interaction force between the two continua caused by $\mathbf{H}^L$ is counterbalanced. The lattice continuum reacts to the forces and couples by means of stresses and deformations in the usual manner. A schematic diagram of the model showing forces and couples is shown in Fig. 1.

### 3. THE LATTICE CONTINUUM

Figure 2 shows a surface $s'$ which separates a portion of the volume $V'$ of the lattice continuum from the remainder. At each point of $s'$, let $\mathbf{n}$ be the unit vector normal to $s'$ pointing out of the volume $V'$. The material on the side of $s'$ toward which $\mathbf{n}$ is directed exerts contact forces, on the material on the other side, which are assumed to consist, at each point of $s'$, of a force per unit area $\mathbf{t}(\mathbf{n})$.[18] Also, at each point in $V'$, let $\mathbf{f}$ be the field of extrinsic and mutual forces per unit volume and $\mathbf{c}$ the field of extrinsic and mutual couples per unit volume. The word extrinsic refers to force fields generated outside the body, and mutual to force fields generated within the body far from the point where they act. The traction vector $\mathbf{t}(\mathbf{n})$ and the body force vector $\mathbf{f}$ are polar vectors, whereas the body couple vector $\mathbf{c}$ is an axial vector. As usual axial vectors are taken as positive in the direction of advance of a right-handed screw.

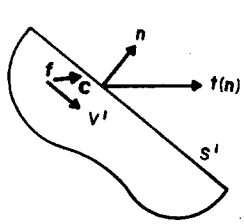


FIG. 2. An arbitrary portion $V'$, of the lattice continuum separated from the remainder by a surface $S'$.

Let us consider the motion of a portion $V$, of a material volume, bounded by a surface $S$ with unit outward normal $\mathbf{n}$. Across $S$ there act traction vectors $\mathbf{t}(\mathbf{n})$, and within $V$ there act body force and body couple vectors $\mathbf{f}$ and $\mathbf{c}$. The equations of the conservation of mass, linear momentum and angular momentum are taken to be, respectively,

$$\frac{d}{dt}\int_V \rho\, dV = 0, \tag{3.1}$$

$$\frac{d}{dt}\int_V \rho\mathbf{v}\, dV = \int_S \mathbf{t}(\mathbf{n})\, dS + \int_V \mathbf{f}\, dV, \tag{3.2}$$

$$\frac{d}{dt}\int_V \mathbf{y}\times\rho\mathbf{v}\, dV = \int_S \mathbf{y}\times\mathbf{t}(\mathbf{n})\, dS + \int_V (\mathbf{y}\times\mathbf{f} + \mathbf{c})\, dV, \tag{3.3}$$

where $d/dt$ denotes the material time derivative, $\rho$ is the mass density, $\mathbf{y}$ is the spatial position vector shown in Fig. 3, and $\mathbf{v}$ is the material velocity vector $d\mathbf{y}/dt$.

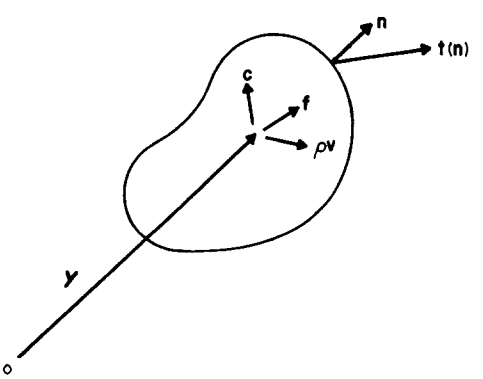


FIG. 3. An arbitrary region of the lattice continuum in motion.

[18] Couples per unit area are assumed to be zero.

Application of Eq. (3.2) to an elementary tetrahedron and taking the limit as the volume of the tetrahedron shrinks to zero yields the definition of the usual stress tensor (dyadic[19]) $\tau$:

$$\mathbf{t}(\mathbf{n}) = \mathbf{n}\cdot\tau = n_i\mathbf{e}_i\cdot\mathbf{e}_j\tau_{jk}\mathbf{e}_k = n_i\tau_{ik}\mathbf{e}_k, \qquad \mathbf{e}_i\cdot\mathbf{e}_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}, \tag{3.4}$$

where $\mathbf{e}_i$ denotes a unit vector in the $i$th direction, and the summation convention for repeated indices is employed. Both dyadic and Cartesian tensor notation will be used in this paper. The one to be used in a specific instance will be determined by convenience and common practice. For instance, the vectors y and v, which have already been introduced, and the vector operator $\nabla_y$ may be written in Cartesian coordinates, respectively, as

$$\mathbf{y} = \mathbf{e}_i y_i, \quad \mathbf{v} = \mathbf{e}_i(d/dt)y_i, \quad \nabla_y = \mathbf{e}_i(\partial/\partial y_i).$$

Thus, $y_i$ represents the rectangular Cartesian coordinates of a material particle at a time $t$, i.e., the Eulerian coordinates.

Now, with (3.4), (3.1), and the divergence theorem, (3.2) becomes

$$\int_V \left(\nabla_y\cdot\tau + \mathbf{f} - \rho\frac{d}{dt}\mathbf{v}\right) dV = 0,$$

from which, since $V$ is arbitrary, we obtain

$$\nabla_y\cdot\tau + \mathbf{f} = \rho(d/dt)\mathbf{v}, \tag{3.5a}$$

or in component form

$$(\partial/\partial y_i)\tau_{ij} + f_j = \rho(d/dt)v_j, \tag{3.5b}$$

which are the usual stress equations of motion.

After substitution of (3.4), Eq. (3.3) takes the component form

$$\frac{d}{dt}\int_V e_{ijk}y_j\rho v_k\, dV = \int_S e_{ijk}y_j n_l\tau_{lk}\, dS + \int_V (e_{ijk}y_j f_k + c_i)\, dV, \tag{3.6}$$

in which we have introduced the skew-symmetric axial tensor $e_{ijk}$, which is defined by

$$e_{ijk} = \begin{cases} +1 & \text{if } ijk \text{ cyclic (123, 231, 312)}, \\ 0 & \text{if any two indices are equal}, \\ -1 & \text{if } ijk \text{ anticyclic (132, 213, 321)}, \end{cases}$$

and which is related to the unit base vectors $\mathbf{e}_i$ by

$$e_{ijk} = \mathbf{e}_i\cdot\mathbf{e}_j\times\mathbf{e}_k.$$

[19] For those interested in more detail see Ref. 2, Sec. 203 or Ref. 1, Sec. 47.

Applying the divergence theorem to (3.6) and using (3.1) as well as the fact that $\partial y_j/\partial y_l = \delta_{jl}$, we obtain

$$\int_V e_{ijk}y_j\left(\frac{\partial\tau_{lk}}{\partial y_l} + f_k - \rho\frac{d}{dt}v_k\right) dV + \int_V (e_{ijk}\tau_{jk} + c_i)\, dV = 0,$$

from which, with (3.5b), we obtain

$$e_{ijk}\tau_{jk} + c_i = 0, \tag{3.7}$$

which yields the symmetry of the stress tensor in classical elasticity theory, since $\mathbf{c} = 0$ in that theory. For the medium described in Sec. 2 we know that the couple c exerted on the lattice continuum by the spin continuum is $\mathbf{H}^L\times\mathbf{M}$. Consequently, we have

$$c_i = e_{ijk}H_j^L M_k,$$

the substitution of which in (3.7) yields

$$e_{ijk}(\tau_{jk} + H_j^L M_k) = 0. \tag{3.8}$$

Equation (3.8) is an axial vector equation, from which the corresponding second-rank antisymmetric polar tensor equation can readily be determined in the usual manner, i.e., by operating on (3.8) with $e_{ilm}$ and using the well-known tensor identity

$$e_{ilm}e_{ijk} = \delta_{lj}\delta_{mk} - \delta_{lk}\delta_{mj},$$

to obtain

$$\tau_{lm} - \tau_{ml} + H_l^L M_m - H_m^L M_l = 0. \tag{3.9}$$

Since $2\tau_{lm}^A = \tau_{lm} - \tau_{ml}$, Eq. (3.9) may be written

$$\tau_{lm}^A = \tfrac{1}{2}(M_l H_m^L - H_l^L M_m), \tag{3.10a}$$

or in vector (dyadic) notation,

$$\tau^A = \tfrac{1}{2}(\mathbf{M}\mathbf{H}^L - \mathbf{H}^L\mathbf{M}), \tag{3.10b}$$

which gives the antisymmetric portion of the mechanical stress tensor in terms of the magnetization and the local magnetic field. Since $\tau = \tau^S + \tau^A$, we may substitute from (3.10b) into (3.5a) to obtain

$$\nabla_y\cdot\tau^S + \tfrac{1}{2}\nabla_y\cdot(\mathbf{M}\mathbf{H}^L - \mathbf{H}^L\mathbf{M}) + \mathbf{f} = \rho(d/dt)\mathbf{v}, \tag{3.11}$$

which is a useful form of the stress equations of motion for the medium being considered.

It should be remembered that Eq. (3.11) is written in Eulerian (spatial) coordinates $y_i$, and that in those coordinates

$$(d/dt)\mathbf{v} = (\partial/\partial t)\mathbf{v} + \mathbf{v}\cdot\nabla_y\mathbf{v}.$$

Equation (3.11) can also be transformed to La-

grangian (material) coordinates, which will turn out to be more useful in this paper. In addition, material as well as spatial coordinates will have to be employed in formulating constitutive equations. Thus, we introduce the material position vector $\mathbf{x}$, which refers to the position of a material particle at some reference time and may be written in the Cartesian form

$$\mathbf{x} = \mathbf{e}_i x_i,$$

where it is understood that[20]

$$x_i = x_i(y_j, t) \quad \text{and} \quad y_j = y_j(x_i, t). \tag{3.12}$$

The transformations shown in (3.12) are usually referred to as the mapping functions of the deformation or simply the deformation. They are assumed to be one-to-one and twice continuously differentiable. We also introduce the material gradient $\nabla_x \equiv \partial/\partial\mathbf{x}$, which, in Cartesian coordinates, takes the form

$$\nabla_x = \mathbf{e}_i \, \partial/\partial x_i.$$

Since the mappings (3.12) are assumed to be one-to-one, the Jacobians of the transformations (3.12) are always different from zero, and by continuity always have the same sign. Thus, we may write

$$J = \det ||\partial y_i/\partial x_j|| > 0. \tag{3.13}$$

We also have the well-known geometric equation

$$dV = J \, dV_0, \tag{3.14}$$

which relates an element of volume $dV$ in the deformed state to the corresponding element $dV_0$ in the reference state. From (3.1) and (3.14), we obtain

$$\rho J = \rho_0, \tag{3.15}$$

wherein $\rho_0$ is the mass density in the reference configuration. Equation (3.15) is one form of the continuity equation.

From the transformations (3.12) and the chain rule of differentiation we have

$$\frac{\partial}{\partial y_j} = \frac{\partial x_m}{\partial y_j} \frac{\partial}{\partial x_m}, \tag{3.16a}$$

which may be written vectorially as

$$\nabla_y = \nabla_y \mathbf{x} \cdot \nabla_x. \tag{3.16b}$$

Using (3.16b), and regarding all variables in (3.11) as functions of the Lagrangian coordinates $x_i$ and the time $t$, we may write (3.11) in the form

$$\nabla_y \mathbf{x} \cdot \nabla_x \cdot \overset{\text{S}}{\tau} + \tfrac{1}{2} \nabla_y \mathbf{x} \cdot \nabla_x \cdot (\mathbf{M}\mathbf{H}^{\text{L}} - \mathbf{H}^{\text{L}}\mathbf{M}) + \mathbf{f} = \rho(\partial \mathbf{v}/\partial t), \tag{3.17}$$

which is the stress equation of motion in Lagrangian coordinates.

## 4. THE ELECTRONIC SPIN CONTINUUM

Now it has been assumed that the electronic spin continuum possesses no linear momentum and that no point can translate with respect to the corresponding point of the lattice continuum. Thus it is clear that the spin continuum expands and contracts with the lattice continuum and must occupy exactly the same volume as the lattice continuum, so that volumetric behavior is governed by Eqs. (3.1), (3.14), and (3.15). Similarly, the conservation of linear momentum simply says that whatever force[21] (of magnetic origin) is applied to a point of the spin continuum, is transferred directly to the lattice continuum at that point. That force has been labeled $\mathbf{f}$ in Sec. 3, and its expression in terms of magnetic quantities is given later on in this section.

Inasmuch as the macroscopic effects of exchange can be adequately treated by considering only nearest-neighbor interactions between adjacent spins,[10] it seems reasonable to represent exchange macroscopically by means of a surface interaction in much the same manner as the stress vector which was discussed in the preceding section. More specifically, we introduce an effective surface exchange field $\mathbf{F}$ which acts across a surface $S$ and produces a couple per unit of surface area equal to $\mathbf{M} \times \mathbf{F}$. The exchange field vector $\mathbf{F}$ is an axial vector which originates in that portion of the spin continuum which is just on the other side of the surface being considered. Note that $\mathbf{F}$ has the dimensions of magnetic field times length, and that its value at a given point will depend on the direction of the normal to the surface at that point. Since only that portion of $\mathbf{F}(\mathbf{n})$ which is perpendicular to $\mathbf{M}$, for any surface through the point, has effectively been defined, we may introduce the condition

$$[\mathbf{F}(\mathbf{n})] \cdot \mathbf{M} = 0, \tag{4.1}$$

for any surface, without loss of generality. We now consider the angular momentum of a volume $V$, of the electronic spin continuum, bounded by a surface $S$ with unit outward normal $\mathbf{n}$. Across $S$ there act the effective magnetic exchange vectors $\mathbf{F}(\mathbf{n})$ discussed above, and within $V$ there act the

[20] For a clear discussion of deformation theory see Ref. 2, Secs. 13, 15, and 16.

[21] Only magnetic forces are being considered in this paper.

ordinary Maxwellian magnetic field $\mathbf{H}^{M}$ and the aforementioned local magnetic field $\mathbf{H}^{L}$. Naturally $\mathbf{F}$ and $\mathbf{H}^{L}$ are axial vectors, as are $\mathbf{M}$ and $\mathbf{H}^{M}$. The equation of the conservation of angular momentum takes the form

$$\int_S \mathbf{M}\times\mathbf{F}(\mathbf{n})\,dS + \int_V \mathbf{M}\times(\mathbf{H}^{M}+\mathbf{H}^{L})\,dV = \frac{d}{dt}\int_V \frac{\mathbf{M}}{\gamma}\,dV, \qquad (4.2)$$

where again, $d/dt$ denotes the material derivative.

The dependence of the exchange field vector $\mathbf{F}(\mathbf{n})$ on the normal $\mathbf{n}$ may be obtained from (4.2) in the following manner. Application of Eq. (4.2) to the tetrahedron shown in Fig. 4 and taking the limit as the volume of the tetrahedron shrinks to zero yields

$$\mathbf{M}\times\left[\mathbf{F}(\mathbf{n})\Delta S + \sum_{i=1}^{3}\mathbf{F}(-\mathbf{e}_i)\Delta S_i\right] = 0, \qquad (4.3)$$

since $h \to 0$, while $\mathbf{M}$, $\mathbf{H}^{M}$, $\mathbf{H}^{L}$, and $(d/dt)(\mathbf{M}/\gamma)$ remain bounded. Since we have introduced the condition (4.1) and $\mathbf{M}$ is assumed nonzero, Eq. (4.3) is satisfied if, and only if

$$\mathbf{F}(\mathbf{n})\Delta S + \sum_{i=1}^{3}\mathbf{F}(-\mathbf{e}_i)\Delta S_i = 0. \qquad (4.4)$$

From Fig. 4 it is clear that

$$\Delta S_i = \Delta S_{-i} = n_i \Delta S,$$

which, with (4.4), gives

$$\mathbf{F}(\mathbf{n}) = -\sum_{i=1}^{3} n_i \mathbf{F}(-\mathbf{e}_i). \qquad (4.5)$$

Consider the surface $\mathbf{n} = (1, 0, 0,)$; then from (4.5) we have

$$\mathbf{F}(\mathbf{e}_1) = -\mathbf{F}(-\mathbf{e}_1).$$

A parallel application of (4.5) to the remaining two perpendicular surfaces shows that

$$\mathbf{F}(\mathbf{e}_i) = -\mathbf{F}(-\mathbf{e}_i), \qquad (4.6)$$

which gives us the important result that the effective magnetic exchange vectors acting upon opposite sides of the same surface are equal in magnitude and opposite in direction. From (4.6) and (4.5), upon eliminating the $\sum$ and reintroducing the summation convention, we have

$$\mathbf{F}(\mathbf{n}) = n_i \mathbf{F}(\mathbf{e}_i),$$

which may be written

$$\mathbf{F}(\mathbf{n}) = n_j \mathbf{e}_j\cdot\mathbf{e}_i\mathbf{F}(\mathbf{e}_i) = \mathbf{n}\cdot\mathbf{e}_i\mathbf{F}(\mathbf{e}_i). \qquad (4.7)$$

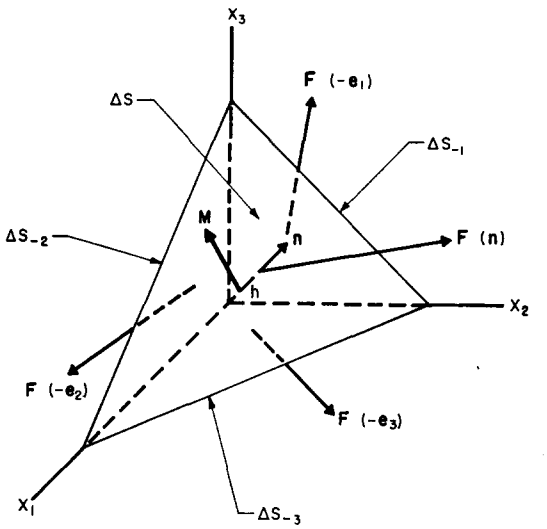


FIG. 4. Elementary tetrahedron of volume $\frac{1}{3}h\Delta S$.

The quantity $\mathbf{e}_i\mathbf{F}(\mathbf{e}_i)$ is a dyadic and hereafter will be denoted by the symbol $-\mathbf{A}$. With this convention, Eq. (4.7) takes the vector form

$$\mathbf{F}(\mathbf{n}) = -\mathbf{n}\cdot\mathbf{A}, \qquad (4.8a)$$

or the Cartesian component form

$$F_i(\mathbf{n}) = -n_j A_{ij}, \qquad (4.8b)$$

in which the first index on the magnetic exchange tensor $\mathbf{A}$ refers to the surface on which the magnetic exchange field acts and the second to the components of the field vector acting on that surface. Since (4.1) holds for arbitrary $\mathbf{n}$, we have, from (4.8) and (4.1), the important condition

$$\mathbf{A}\cdot\mathbf{M} = 0, \qquad (4.9)$$

which reduces from 9 to 6 the number of possible components of $\mathbf{A}$.

Before proceeding we must discuss the saturation condition in some detail. When a material is magnetically saturated, the magnitude of the total magnetic moment is conserved. However, when the volume is varying, the magnitude of the magnetic moment per unit volume $\mathbf{M}$ is not conserved; nevertheless, since the mass is conserved, the magnitude of the magnetic moment per unit mass is conserved. In view of this it is convenient to work with the magnetic moment per unit mass $\boldsymbol{\mu}$, which is given by

$$\boldsymbol{\mu} = \mathbf{M}/\rho. \qquad (4.10)$$

Then the saturation condition can be written in the convenient form

$$\boldsymbol{\mu}\cdot\boldsymbol{\mu} = \mu_S^2, \qquad (4.11)$$

where $\mu_s$ is constant in a homogeneous material. Taking the material time derivative of (4.11) and of the material gradient of (4.11), respectively, we obtain the important conditions

$$\mathfrak{y}\cdot(d/dt)\mathfrak{y} = 0, \tag{4.12}$$

$$(\nabla_x\mathfrak{y})\cdot(d/dt)\mathfrak{y} + \mathfrak{y}\cdot(d/dt)(\mathfrak{y}\nabla_x) = 0, \tag{4.13}$$

which must be satisfied by $\mathfrak{y}$.

After substitution from (4.8) and (4.10) and application of the divergence theorem, Eq. (4.2) takes the component form

$$\int_V e_{ijk}\rho\mu_j\left(-\frac{\partial A_{lk}}{\partial y_l} - \frac{A_{lk}}{\rho}\frac{\partial\rho}{\partial y_l} + H_k^M + H_k^L\right)dV - \int_V e_{ijk}\rho A_{lk}\frac{\partial\mu_j}{\partial y_l}\,dV = \frac{1}{\gamma}\frac{d}{dt}\int_V \rho\mu_i\,dV,$$

from which, with (3.1), we obtain

$$e_{ijk}\rho\mu_j\left(H_k^M - \frac{\partial A_{lk}}{\partial y_l} - \frac{A_{lk}}{\rho}\frac{\partial\rho}{\partial y_l} + H_k^L\right) - e_{ijk}\rho A_{lk}\frac{\partial\mu_j}{\partial y_l} = \frac{1}{\gamma}\rho\frac{d}{dt}\mu_i. \tag{4.14}$$

We must now introduce a further restriction on $\mathbf{A}$, i.e., in addition to (4.9), so that (4.12) be satisfied identically. We must require $A_{lk}\partial\mu_j/\partial y_l$ to be symmetric, i.e.,[22]

$$A_{lk}(\partial\mu_j/\partial y_l) = A_{lj}(\partial\mu_k/\partial y_l), \tag{4.15}$$

which assures us that the last term on the lhs of (4.14) vanishes. Thus, (4.14) becomes

$$e_{ijk}\rho\mu_j\left(H_k^M - \frac{\partial A_{lk}}{\partial y_l} - \frac{A_{lk}}{\rho}\frac{\partial\rho}{\partial y_l} + H_k^L\right) = \frac{1}{\gamma}\rho\frac{d}{dt}\mu_i, \tag{4.16a}$$

which may be written in the vector form

$$\mathfrak{y}\times\left(\mathbf{H}^M - \nabla_y\cdot\mathbf{A} - \frac{1}{\rho}\nabla_y\rho\cdot\mathbf{A} + \mathbf{H}^L\right) = \frac{1}{\gamma}\frac{d}{dt}\mathfrak{y}, \tag{4.16b}$$

which is obviously consistent with (4.12).

We will now determine the expression for the magnetic body force $\mathbf{f}$. Since any force generated by $\mathbf{H}^L$ is local and it has been assumed that there is no relative motion of the two continua, such a force is automatically equilibrated by an interaction between the continua which need not be specified here. On the other hand it is well known, from magnetostatics, that a spatially varying magnetic field $\mathbf{H}^M$ exerts a force on a magnetic dipole; but there appears to be some disagreement in the existing literature[23–25] as to the precise form for $\mathbf{f}$. However, for definiteness we will use the form

$$\mathbf{f} = \mathbf{M}\cdot\nabla_y\mathbf{H}^M, \tag{4.17}$$

for the body force. Other forms could have been chosen equally readily, with only minor changes resulting. As has already been discussed, this $\mathbf{f}$ is applied directly to the lattice continuum.

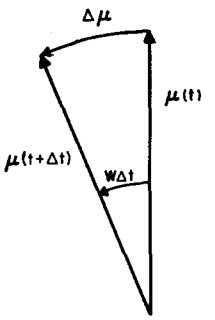


FIG. 5. Incremental motion of magnetization vector in time interval $\Delta t$.

In order for us to write the equation of conservation of energy either for the spin continuum or for the total continuum, we must know the rate at which each couple acting on the spin continuum does work. To this end we consider a typical couple $\mathbf{M}\times\mathbf{H}^P$. The rate at which work is done by this couple in time $\Delta t$ is

$$\Delta W^P/\Delta t = \mathbf{M}\times\mathbf{H}^P\cdot\mathbf{w} = \rho\mathfrak{y}\times\mathbf{H}^P\cdot\mathbf{w}, \tag{4.18}$$

where the angle $\mathbf{w}\Delta t$ through which $\mathfrak{y}$ turns in time $\Delta t$ is shown in Fig. 5. From Fig. 5 it is clear that

$$\mathbf{w}\Delta t = \frac{|\Delta\mathfrak{y}|}{|\mathfrak{y}|}\frac{\mathfrak{y}}{|\mathfrak{y}|}\times\frac{\Delta\mathfrak{y}}{|\Delta\mathfrak{y}|},$$

from which we have

$$\mathbf{w} = (1/\mu_s^2)\mathfrak{y}\times(d/dt)\mathfrak{y}. \tag{4.19}$$

From (4.18) and (4.19) we obtain

$$\frac{dW^P}{dt} = (\rho\mathfrak{y}\times\mathbf{H}^P)\cdot\left(\frac{1}{\mu_s^2}\mathfrak{y}\times\frac{d}{dt}\mathfrak{y}\right), \tag{4.20}$$

for the rate at which work is done by a couple $\mathbf{M}\times\mathbf{H}^P$ acting on $\mathbf{M}$. Using some well-known vector identities, we see that Eq. (4.20) may be written

[22] It can be shown that this condition is equivalent to the invariance of the exchange energy in a rigid rotation of the entire spin continuum with respect to the lattice continuum.

[23] M. Mason and W. Weaver, *The Electromagnetic Field* (The University of Chicago Press, Chicago, 1929) (also Dover Publications, Inc., New York), p. 218.

[24] C. Møller, *The Theory of Relativity* (Clarendon Press, Oxford, England, 1952), Sec. 75.

[25] W. F. Brown, Jr., *Magnetostatic Principles in Ferromagnetism* (North-Holland Publishing Company, Amsterdam, 1962), Chap. 4, Sec. 2.

in the form

$$\frac{dW^{\mathrm{P}}}{dt} = \frac{1}{\mu_S^2}\rho\left(\mathfrak{y}\cdot\mathfrak{y}\mathbf{H}^{\mathrm{P}}\cdot\frac{d}{dt}\mathfrak{y} - \mathfrak{y}\cdot\mathbf{H}^{\mathrm{P}}\mathfrak{y}\cdot\frac{d}{dt}\mathfrak{y}\right),$$

from which, with (4.11) and (4.12), we obtain

$$dW^{\mathrm{P}}/dt = \mathbf{H}^{\mathrm{P}}\cdot\rho(d/dt)\mathfrak{y}. \tag{4.21}$$

Thus it is clear that dotting the rhs of (4.19) into (4.16b) gives us the important energy relation

$$\mathbf{H}^{M}\cdot\frac{d}{dt}\mathfrak{y} - \nabla_y\cdot\mathbf{A}\cdot\frac{d}{dt}\mathfrak{y} - \frac{1}{\rho}\nabla_y\rho\cdot\mathbf{A}\cdot\frac{d}{dt}\mathfrak{y} + \mathbf{H}^{\mathrm{L}}\cdot\frac{d}{dt}\mathfrak{y} = 0, \tag{4.22a}$$

which may be written in the tensor form

$$H_k^{\mathrm{M}}\frac{d}{dt}\mu_k - \frac{\partial A_{lk}}{\partial y_l}\frac{d}{dt}\mu_k - \frac{1}{\rho}\frac{\partial\rho}{\partial y_l}A_{lk}\frac{d}{dt}\mu_k + H_k^{\mathrm{L}}\frac{d}{dt}\mu_k = 0. \tag{4.22b}$$

In general the magnetic field must also satisfy Maxwell's equations

$$\nabla_y\times\mathbf{H}^{M} = \frac{1}{c}\frac{\partial\mathbf{D}}{\partial t} + \frac{4\pi}{c}\mathbf{j}, \tag{4.23}$$

$$\nabla_y\times\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}, \tag{4.24}$$

$$\mathbf{B} = \mathbf{H}^{\mathrm{M}} + 4\pi\mathbf{M}, \tag{4.25}$$

where the vectors $\mathbf{D}$, $\mathbf{j}$, and $\mathbf{E}$ represent the electric displacement, conduction current and electric field, respectively, and $c$ denotes the speed of light. Since we are concerned with nonconducting magnetic media at frequency wavelength combinations far outside the range associated with electromagnetic propagation, the conduction current $\mathbf{j}$ and the time rate of change of electric displacement $\partial\mathbf{D}/\partial t$ may be neglected, and Eqs. (4.23) and (4.24) reduce to the equations of the quasistationary magnetic field,

$$\mathbf{H}^{\mathrm{M}} = -\nabla_y\varphi, \tag{4.26}$$

$$\nabla_y\cdot\mathbf{B} = 0, \tag{4.27}$$

where $\varphi$ is the axial magnetic scalar potential. For our purposes we shall also need the integral relations from which (4.26) and (4.27) may be deduced when suitable continuity conditions are assumed. These integral relations are, respectively,

$$\oint_C \mathbf{H}^{M}\cdot d\mathbf{y} = 0, \tag{4.28}$$

$$\int_S \mathbf{n}\cdot\mathbf{B}\,dS = 0, \tag{4.29}$$

where $C$ denotes an arbitrary circuit and $S$ an arbitrary closed surface, both of which are *stationary* with respect to an inertial reference frame.

## 5. THERMODYNAMIC CONSIDERATIONS

The principle of conservation of energy for the material medium—consisting of both the lattice and electronic spin continua—states that in any volume $V$ of a body bounded by a surface $S$ with unit outward normal $\mathbf{n}$, the rate of increase of energy (kinetic plus internal) is equal to the rate at which work is done by the surface tractions and magnetic exchange torques acting across $S$, less the flux of thermal energy outward across $S$, plus the rate at which energy is supplied to the material from the quasistatic magnetic field. Thus,

$$\frac{d}{dt}\int_V (\tfrac{1}{2}\rho\mathbf{v}\cdot\mathbf{v} + \rho U)\,dV = \int_S \left(\mathbf{t}\cdot\mathbf{v} + \mathbf{F}\cdot\rho\frac{d}{dt}\mathfrak{y} - \mathbf{n}\cdot\mathbf{q}\right)dS + \int_V \epsilon\,dV, \tag{5.1}$$

where $d/dt$ is the material time derivative, $\frac{1}{2}\mathbf{v}\cdot\mathbf{v}$ is the kinetic energy per unit mass, $U$ is the internal energy per unit mass, $\mathbf{t}\cdot\mathbf{v}$ is the rate at which work is done by the surface tractions, $\mathbf{F}\cdot\rho\,d\mathfrak{y}/dt$ is the rate at which work is done by the surface exchange torques, $\mathbf{q}$ is the heat flux vector and $\epsilon$ is the rate of supply of energy per unit volume.[26] The time rate of change of spin kinetic energy vanishes by virtue of (4.11). In view of Eqs. (4.17) and (4.21), the rate at which energy is supplied to the material takes the form

$$\epsilon = \mathbf{H}^{\mathrm{M}}\cdot\rho(d/dt)\mathfrak{y} + (\mathbf{M}\cdot\nabla_y\mathbf{H}^{\mathrm{M}})\cdot\mathbf{v}. \tag{5.2}$$

The first term in (5.2) represents the rate at which the magnetic couples do work and the second term represents the rate at which the magnetic body force does work.

In view of (5.2) and Poyntings theorem, Eq. (5.1) can be written in another form which is particularly enlightening. Poyntings theorem for this quasistatic magnetic field takes the degenerate form

$$-\int_S \frac{1}{4\pi}\mathbf{n}\cdot\left(\varphi\frac{\partial\mathbf{B}}{\partial t}\right)dS - \int_V \mathbf{H}^{M}\cdot\frac{\partial\mathbf{M}}{\partial t}\,dV = \int_V \frac{\partial}{\partial t}\left(\frac{1}{8\pi}\mathbf{H}^{M}\cdot\mathbf{H}^{M}\right)dV, \tag{5.3}$$

where $(4\pi)^{-1}\varphi\partial\mathbf{B}/\partial t$ is the form taken by the Poynting vector $\mathbf{E}\times\mathbf{H}$ in a quasistatic magnetic

[26] For a general discussion of the procedure employed here in coupling the magnetic field to the material medium see Ref. 2, Secs. 284–286.

field. From (4.10) and (3.15), we obtain

$$\rho \frac{d\mathbf{v}}{dt} = \frac{d}{dt}\mathbf{M} + \mathbf{M}(\nabla_v \cdot \mathbf{v}), \tag{5.4}$$

where we have introduced the well-known identity[27] $J^{-1}dJ/dt = \nabla_v \cdot \mathbf{v}$. Equation (5.3) is written in terms of regions and surfaces which are stationary with respect to the reference frame, whereas all other equations in this section are written in terms of material regions and surfaces which are moving with respect to the reference frame. Equation (5.3) may be written in terms of material regions and surfaces by employing the relations

$$\partial/\partial t = d/dt - \mathbf{v}\cdot\nabla_v, \tag{5.5}$$

$$(d/dt)\, dV = \nabla_v \cdot \mathbf{v}\, dV. \tag{5.6}$$

When written in terms of material regions and surfaces, Eq. (5.3) takes the form

$$\begin{aligned} -\int_S \frac{1}{4\pi}\mathbf{n}\cdot\left(\varphi\frac{\partial \mathbf{B}}{\partial t}\right) dS - \int_V \mathbf{H}^{\mathrm{M}}\cdot\frac{d}{dt}\mathbf{M}\, dV \\ + \int_V \mathbf{H}^{\mathrm{M}}\cdot(\mathbf{v}\cdot\nabla_v\mathbf{M})\, dV \\ + \int_S \mathbf{n}\cdot\mathbf{v}\,\frac{1}{8\pi}\mathbf{H}^{\mathrm{M}}\cdot\mathbf{H}^{\mathrm{M}}\, dS \\ = \frac{d}{dt}\int_V \left(\frac{1}{8\pi}\mathbf{H}^{\mathrm{M}}\cdot\mathbf{H}^{\mathrm{M}}\right) dV. \end{aligned} \tag{5.7}$$

Substituting from (5.4) into (5.7), then from (5.7) into (5.2), then from (5.2) into (5.1), recombining terms, and employing the divergence theorem, we obtain

$$\begin{aligned} &\frac{d}{dt}\int_V \left(\tfrac{1}{2}\rho v_k v_k + \rho U + \frac{1}{8\pi}H_k^{\mathrm{M}}H_k^{\mathrm{M}}\right) dV \\ &= \int_S \left(t_i v_i + F_i \rho \frac{d}{dt}\mu_i - n_i q_i - \frac{1}{4\pi}n_i\varphi\frac{\partial B_i}{\partial t}\right. \\ &\quad \left. + n_i \frac{1}{8\pi}H_k^{\mathrm{M}}H_k^{\mathrm{M}}v_i + n_i H_k^{\mathrm{M}}M_k v_i\right) dS, \end{aligned} \tag{5.8}$$

which is a particularly interesting form of the equation of conservation of energy. Equation (5.8) says that the time rate of change of kinetic plus internal plus magnetic field energy is equal to the rate at which work is done by the surface tractions and magnetic exchange torques acting across $S$ less the flux of thermal and magnetic energy outward across $S$ plus a convective flux of magnetic field energy and magnetic dipole energy. Obviously, Eqs. (5.1) and (5.8) are completely equivalent and we may proceed equally readily from either one.

Now, with (3.1), (3.4), (4.8), (4.25), (4.26), (4.27), and the divergence theorem, (5.8) becomes

$$\begin{aligned} \int_V \Bigg[ &\rho\frac{dU}{dt} + \left(\rho\frac{dv_i}{dt} - \frac{\partial \tau_{ij}}{\partial y_j} - M_k\frac{\partial H_i^{\mathrm{M}}}{\partial y_k}\right)v_i \\ &+ \frac{1}{4\pi}H_k^M\frac{dH_k^{\mathrm{M}}}{dt} - \frac{1}{4\pi}H_k^{\mathrm{M}}\left(\frac{\partial H_k^{\mathrm{M}}}{\partial t} + v_i\frac{\partial H_k^{\mathrm{M}}}{\partial y_i}\right. \\ &\left. + 4\pi\frac{\partial M_k}{\partial t} + 4\pi v_i\frac{\partial M_k}{\partial y_i}\right) - M_k H_k^{\mathrm{M}}\frac{\partial v_i}{\partial y_i} \\ &+ \frac{\partial A_{ij}}{\partial y_i}\rho\frac{d}{dt}\mu_j + A_{ij}\frac{\partial \rho}{\partial y_i}\frac{d}{dt}\mu_j \\ &+ A_{ij}\rho\frac{\partial}{\partial y_i}\left(\frac{d}{dt}\mu_j\right) + \frac{\partial q_i}{\partial y_i} - \tau_{ij}\frac{\partial v_j}{\partial y_i}\Bigg] dV = 0, \end{aligned}$$

which, with (5.4), (5.5), (4.17), and (3.5), becomes

$$\begin{aligned} \int_V \Bigg[ &\rho\frac{dU}{dt} - \tau_{ij}\frac{\partial v_j}{\partial y_i} - H_k^M\rho\frac{d\mu_k}{dt} + \frac{\partial A_{ij}}{\partial y_i}\rho\frac{d}{dt}\mu_j \\ &+ A_{ij}\frac{\partial \rho}{\partial y_i}\frac{d}{dt}\mu_j + A_{ij}\rho\frac{\partial}{\partial y_i}\left(\frac{d}{dt}\mu_j\right) + \frac{\partial q_i}{\partial y_i}\Bigg] dV = 0, \end{aligned}$$

from which, since $V$ is arbitrary, we obtain

$$\begin{aligned} \rho\frac{dU}{dt} = \tau_{ij}\frac{\partial v_j}{\partial y_i} + \rho\Bigg(&H_i^M\frac{d\mu_j}{dt} - \frac{\partial A_{ij}}{\partial y_i}\frac{d\mu_j}{dt} \\ &- \frac{1}{\rho}\frac{\partial \rho}{\partial y_i}A_{ij}\frac{d}{dt}\mu_j\Bigg) - A_{ij}\rho\frac{\partial}{\partial y_i}\left(\frac{d}{dt}\mu_j\right) - \frac{\partial q_i}{\partial y_i}. \end{aligned} \tag{5.9}$$

Substituting from (4.22) into (5.9), we obtain

$$\begin{aligned} \rho\frac{dU}{dt} = \tau_{ij}\frac{\partial v_j}{\partial y_i} - \rho H_i^{\mathrm{L}}\frac{d\mu_i}{dt} \\ - \rho A_{ij}\frac{\partial}{\partial y_i}\left(\frac{d}{dt}\mu_j\right) - \frac{\partial q_i}{\partial y_i}, \end{aligned} \tag{5.10}$$

which is a form of the first law of thermodynamics for this combined continuum.

We may now introduce dissipation by assuming that the symmetric part of the stress $\boldsymbol{\tau}^{\mathrm{S}}$ and the local magnetic field $\mathbf{H}^{\mathrm{L}}$ may be written as the sum of a dissipative and a nondissipative part. This is a restrictive assumption, but it is believed to be adequate for the medium being considered. Consequently, we write[28]

$$\boldsymbol{\tau}^{\mathrm{S}} = {}^{\mathrm{R}}\boldsymbol{\tau}^{\mathrm{S}} + {}^{\mathrm{D}}\boldsymbol{\tau}^{\mathrm{S}}, \quad \mathbf{H}^{\mathrm{L}} = {}^{\mathrm{R}}\mathbf{H}^{\mathrm{L}} + {}^{\mathrm{D}}\mathbf{H}^{\mathrm{L}}, \tag{5.11a, b}$$

where

$${}^{\mathrm{D}}\mathbf{H}^{\mathrm{L}}\cdot\mathbf{v} = 0 \tag{5.12}$$

[27] For a derivation, see A. C. Eringen, *Nonlinear Theory of Continuous Media* (McGraw-Hill Book Company, Inc., New York, 1962), Sec. 19.

[28] It is not actually believed that the dissipative portion of the local magnetic field ${}^{\mathrm{D}}\mathbf{H}^{\mathrm{L}}$ exists. However, it is being kept for purposes of comparison with the well-known phenomenological magnetodynamic theory for the rigid solid.

and in each case the superscript R stands for the nondissipative (stored energy) portion and the superscript D for the dissipative portion. Substituting from (5.11b) into (3.10) and employing (4.10), we obtain

$$\boldsymbol{\tau}^{A} = {}^{R}\boldsymbol{\tau}^{A} + {}^{D}\boldsymbol{\tau}^{A}, \tag{5.13}$$

where

$$\begin{aligned} {}^{R}\boldsymbol{\tau}^{A} &= \tfrac{1}{2}\rho(\mathfrak{y}\,{}^{R}\mathbf{H}^{L} - {}^{R}\mathbf{H}^{L}\mathfrak{y}), \\ {}^{D}\boldsymbol{\tau}^{A} &= \tfrac{1}{2}\rho(\mathfrak{y}\,{}^{D}\mathbf{H}^{L} - {}^{D}\mathbf{H}^{L}\mathfrak{y}). \end{aligned} \tag{5.14}$$

Before proceeding further, we must recall a few well-known relations concerning vorticity and spin.[29] These are

$$\begin{aligned} (\boldsymbol{\Omega})_i &= \tfrac{1}{2}(\nabla_y \times \mathbf{v})_i = \tfrac{1}{2}e_{ijk}(\partial v_k/\partial y_j), \\ (\boldsymbol{\omega})_{mn} &= \omega_{mn} = -e_{imn}\Omega_i \\ &= \tfrac{1}{2}(\partial v_m/\partial y_n - \partial v_n/\partial y_m), \\ \Omega_i &= -\tfrac{1}{2}e_{imn}\omega_{mn}, \end{aligned} \tag{5.15}$$

where $\boldsymbol{\Omega}$ is an axial vector called the vorticity which is the local angular velocity of the lattice continuum and $\boldsymbol{\omega}$ is the polar tensor of the axial vector $\boldsymbol{\Omega}$, and is called the spin tensor. Note that $\boldsymbol{\omega}$ is the antisymmetric part of the spatial velocity gradient. The symmetric part of the spatial velocity gradient is given by

$$d_{mn} = \tfrac{1}{2}(\partial v_m/\partial y_n + \partial v_n/\partial y_m), \tag{5.16}$$

and is called the rate-of-deformation tensor. Obviously, we have

$$\partial v_m/\partial y_n = d_{mn} + \omega_{mn}. \tag{5.17}$$

Since $\boldsymbol{\tau} = \boldsymbol{\tau}^{S} + \boldsymbol{\tau}^{A}$, we have from (5.11a) and (5.13)

$$\boldsymbol{\tau} = {}^{R}\boldsymbol{\tau}^{S} + {}^{D}\boldsymbol{\tau}^{S} + {}^{R}\boldsymbol{\tau}^{A} + {}^{D}\boldsymbol{\tau}^{A} \tag{5.18a}$$

$$= {}^{R}\boldsymbol{\tau} + {}^{D}\boldsymbol{\tau}^{S} + {}^{D}\boldsymbol{\tau}^{A}. \tag{5.18b}$$

Substituting from (5.11b) and (5.18b) into (5.10) and employing (5.14b) and (5.15)–(5.17), we obtain

$$\begin{aligned} \rho\frac{dU}{dt} &= {}^{R}\tau_{ij}\frac{\partial v_j}{\partial y_i} - \rho\,{}^{R}H_i^{L}\frac{d\mu_j}{dt} - \rho A_{ij}\frac{\partial}{\partial y_i}\left(\frac{d}{dt}\mu_j\right) \\ &+ {}^{D}\tau_{ij}^{S}d_{ij} - \rho\,{}^{D}H_i^{L}\left(\frac{d}{dt}\mu_i - \omega_{ij}\mu_j\right) - \frac{\partial q_i}{\partial y_i}. \end{aligned} \tag{5.19}$$

All current forms of the theory of irreversible thermodynamics[30–32] indicate that for this case the mathematical expression of the second law takes the form

$$\begin{aligned} \rho\frac{dU}{dt} - {}^{R}\tau_{ij}\frac{\partial v_j}{\partial y_i} + \rho\,{}^{R}H_i^{L}\frac{d}{dt}\mu_i \\ + \rho A_{ij}\frac{\partial}{\partial y_i}\left(\frac{d}{dt}\mu_j\right) = \rho\theta\frac{d\eta}{dt}, \end{aligned} \tag{5.20}$$

where $\theta$ is the positive absolute temperature and $\eta$ is the entropy per unit mass. From (5.19) and (5.20) we also have

$$\begin{aligned} \theta\rho\frac{d\eta}{dt} &= {}^{D}\tau_{ij}^{S}d_{ij} \\ &- \rho\,{}^{D}H_i^{L}\left(\frac{d}{dt}\mu_i - \omega_{ij}\mu_j\right) - \frac{\partial q_i}{\partial y_i}. \end{aligned} \tag{5.21}$$

Now, since

$$\frac{\partial}{\partial y_i}\left(\frac{q_i}{\theta}\right) = \frac{1}{\theta}\frac{\partial q_i}{\partial y_i} - \frac{1}{\theta^2}q_i\frac{\partial\theta}{\partial y_i},$$

we have

$$\begin{aligned} \rho\frac{d\eta}{dt} + \frac{\partial}{\partial y_i}\left(\frac{q_i}{\theta}\right) &= \frac{1}{\theta}\,{}^{D}\tau_{ij}^{S}d_{ij} \\ &- \frac{1}{\theta}\rho\,{}^{D}H_k^{L}\left(\frac{d\mu_k}{dt} - \omega_{kj}\mu_j\right) - \frac{1}{\theta^2}q_i\frac{\partial\theta}{\partial y_i}. \end{aligned} \tag{5.22}$$

The quantity $\rho\, d\eta/dt + \partial/\partial y_i\,(q_i/\theta)$ is called the rate of entropy production and will be represented by the symbol $\Delta$. Then the important postulate of thermodynamic irreversibiltity takes the form

$$\Delta \geq 0. \tag{5.23}$$

## 6. CONSTITUTIVE EQUATIONS

Let us record Eqs. (5.20) and (5.22) along with the condition (5.23) below for convenience:

$$\begin{aligned} \rho\frac{dU}{dt} &= {}^{R}\tau_{ij}\frac{\partial v_j}{\partial y_i} - \rho\,{}^{R}H_i^{L}\frac{d}{dt}\mu_i \\ &- \rho A_{ij}\frac{\partial}{\partial y_i}\left(\frac{d}{dt}\mu_j\right) + \rho\theta\frac{d\eta}{dt}, \end{aligned} \tag{6.1a}$$

$$\begin{aligned} \Delta = \frac{1}{\theta}\,{}^{D}\tau_{ij}^{S}d_{ij} - \frac{1}{\theta}\rho\,{}^{D}H_k^{L}\left(\frac{d\mu_k}{dt} - \omega_{kj}\mu_j\right) \\ - \frac{1}{\theta^2}q_i\frac{\partial\theta}{\partial y_i}, \end{aligned} \tag{6.1b}$$

$$\Delta \geq 0. \tag{6.1c}$$

Equation (6.1a) is concerned with recoverable (stored) energy whereas (6.1b) and (6.1c) are concerned with irrecoverable (dissipated) energy. The portions of the constitutive equations which are derivable from a stored energy function ($U$) and the portions which are associated with the dissipa-

[29] For a discussion of vorticity and spin see Ref. 27, Sec. 21.
[30] See Ref. 2, Secs. 245–247, 256–258.
[31] B. A. Boley and J. H. Weiner, *Theory of Thermal Stresses* (John Wiley & Sons, Inc., New York, 1960); Secs. 1.8–1.11.
[32] S. R. De Groot, *Thermodynamics of Irreversible Processes* (North-Holland Publishing Company, Amsterdam, 1952).

tion are independent and are obtained in completely different manners. Consequently, we may determine either one first. We will determine the dissipative portions of the constitutive equations first.

Although the portions of the constitutive equations which are derivable from a stored energy function will be general and nonlinear, the portions which are of a dissipative nature will be linear, i.e., only linear dissipative processes are considered. Moreover, since we are not interested in the most general possible theory of linear dissipation but only in a theory which is deemed to be adequate for the materials being considered, we assume that

$$ {}^{D}\tau^{S}_{ij} = \zeta_{ijkl} d_{kl}, \qquad q_i = -k_{ij}(\partial\theta/\partial y_j), \tag{6.2a, b} $$

$$ {}^{D}H^{L}_{k} = -\sigma\rho(d\mu_k/dt - \omega_{kj}\mu_j), \tag{6.2c} $$

where a careful analysis shows that[33,34]

$$ \zeta_{ijkl} = Z_{pqrs} \frac{\partial y_i}{\partial x_p} \frac{\partial y_j}{\partial x_q} \frac{\partial y_k}{\partial x_r} \frac{\partial y_l}{\partial x_s}, $$

$$ k_{ij} = \kappa_{rs} \frac{\partial y_i}{\partial x_r} \frac{\partial y_j}{\partial x_s}, \tag{6.3} $$

and $\mathbf{Z}$ and $\boldsymbol{\kappa}$ are referred to the undeformed configuration, and where $\sigma$ is a scalar no matter how anisotropic the media by virtue of (4.12), (5.12), (5.15) and the fact that we are expressly excluding nondissipative gyroscopic terms which contribute nothing to $\Delta$. It should be noted that all the constitutive equations in (6.2) satisfy Noll's principle of material objectivity,[35,36] which requires that any equation describing the constitutive behavior of the material shall be independent of the frame of reference of the observer. The explicit mathematical requirement resulting from the application of the general principle to the case treated here simply states that all variables appearing in (6.2), which are linearly related by the coefficients $\zeta_{ijkl}$, $k_{ij}$, and $\sigma$, shall transform according to the appropriate tensor transformation law under *time-dependent* orthogonal coordinate transformations. The important point to realize is that $(d\mu_k/dt - \omega_{kj}\mu_j)$ is objective whereas $d\mu_k/dt$ is not. Moreover, it should also be noted that if the vector $\boldsymbol{\mu}$ be rigidly attached to the lattice continuum, ${}^{D}\mathbf{H}^{L} = 0$. Substituting from (6.3) into (6.2), employing the well-known relation[37]

$$ \frac{d}{dt} E_{rs} = \frac{\partial y_i}{\partial x_r} \frac{\partial y_j}{\partial x_s} d_{ij}, \tag{6.4} $$

in which the material strain tensor $E_{rs}$ is defined by

$$ E_{rs} = \frac{1}{2}\left(\frac{\partial y_k}{\partial x_r} \frac{\partial y_k}{\partial x_s} - \delta_{rs}\right), $$

and using the chain rule of differentiation, we obtain

$$ {}^{D}\tau^{S}_{ij} = \frac{\partial y_i}{\partial x_p} \frac{\partial y_j}{\partial x_q} Z_{pqrs} \frac{d}{dt} E_{rs}, $$

$$ q_i = -\frac{\partial y_i}{\partial x_r} \kappa_{rm} \frac{\partial\theta}{\partial x_m}. \tag{6.5} $$

The scalar $\sigma$ cannot be negative and the matrices $\kappa_{rs}$ and $Z_{(pq)(rs)}$[38] must be matrices of nonnegative quadratic forms, by virtue of (6.1c). The actual form taken by the matrices $\kappa_{rs}$ and $Z_{(pq)(rs)}$ is determined by the point-group symmetry of the lattice in the undeformed state. In a triclinic crystal there are 21 independent $Z_{(pq)(rs)}$ and 6 independent $\kappa_{rs}$. The $Z_{(pq)(rs)}$ are coefficients of mechanical viscosity, the $\kappa_{rs}$ are the usual coefficients of thermal conductivity, and $\sigma$ is a coefficient of magnetic dissipation which is directly related to the Gilbert damping factor. Note that no crossinteractions of a dissipative nature have been considered. It is, of course, quite conceivable that they exist.

We now turn to the determination of the portions of the constitutive equations which are derivable from a stored energy function. We begin by defining the free energy $F$ in the usual manner, i.e., by

$$ F = U - \eta\theta, \tag{6.6} $$

and substituting the material time derivative of (6.6) in (6.1a) to obtain

$$ \rho \frac{dF}{dt} = {}^{R}\tau_{ij} \frac{\partial v_i}{\partial y_j} - \rho\, {}^{R}H^{L}_{i} \frac{d}{dt}\mu_i - \rho A_{ij} \frac{\partial}{\partial y_j}\left(\frac{d}{dt}\mu_i\right) - \rho\eta \frac{d\theta}{dt}. \tag{6.7} $$

In view of the relations,

$$ \frac{\partial v_i}{\partial y_j} = \frac{\partial x_m}{\partial y_j} \frac{d}{dt} \frac{\partial y_i}{\partial x_m}, \qquad \frac{\partial}{\partial y_j}\left(\frac{d}{dt}\mu_i\right) = \frac{\partial x_m}{\partial y_j} \frac{d}{dt} \frac{\partial \mu_i}{\partial x_m}, $$

Eq. (6.7) can be written in the form

$$ \rho \frac{dF}{dt} = {}^{R}\tau_{ij} \frac{\partial x_m}{\partial y_j} \frac{d}{dt}\left(\frac{\partial y_i}{\partial x_m}\right) - \rho\, {}^{R}H^{L}_{i} \frac{d}{dt}\mu_i - \rho A_{ij} \frac{\partial x_m}{\partial y_j} \frac{d}{dt}\left(\frac{\partial \mu_i}{\partial x_m}\right) - \rho\eta \frac{d\theta}{dt}. \tag{6.8} $$

[33] B. D. Coleman and W. Noll, Arch. Ratl. Mech. Anal. 13, 167 (1963).

[34] Equations (6.2a) and (6.2b) along with (6.3) are special cases of Coleman and Nolls Eqs. $(3.9)_2$ and $(3.9)_4$, respectively.

[35] For thorough discussions of the principle of material objectivity see Ref. 2, Secs. 293, 296; Ref. 27, Secs. 27, 44; and Ref. 36.

[36] W. Noll, J. Ratl. Mech. Anal. 4, 3 (1955), Sec. 4.

[37] For a derivation, see Ref. 27, Sec. 19.

[38] It is to be understood that $Z_{(pq)(rs)}$ is here considered as a $6 \times 6$, 2-index matrix.

Motivated by (6.8), we assume

$$F = F(\partial y_i/\partial x_m,\ \mu_i,\ \partial\mu_i/\partial x_m,\ \theta), \tag{6.9}$$

whence

$$\frac{dF}{dt} = \frac{\partial F}{\partial(\partial y_i/\partial x_m)}\frac{d}{dt}\left(\frac{\partial y_i}{\partial x_m}\right) + \frac{\partial F}{\partial \mu_i}\frac{d}{dt}\mu_i + \frac{\partial F}{\partial(\partial\mu_i/\partial x_m)}\frac{d}{dt}\left(\frac{\partial\mu_i}{\partial x_m}\right) + \frac{\partial F}{d\theta}\frac{d\theta}{dt}. \tag{6.10}$$

At this point we must remember that 12 of the 22 time derivatives appearing on the rhs of (6.8) and (6.10) are not independent, but are connected by the four relations given in (4.12) and (4.13). Consequently, we must introduce four Lagrangian undetermined multipliers, $\lambda$ and $L_m$, then multiply (4.12) by $\lambda$, (4.13) by $L_m$, and add the sum to the rhs of (6.8) while substituting from (6.10) to obtain

$$\left[{}^{\mathrm{R}}\tau_{ij}\frac{\partial x_m}{\partial y_i} - \rho\frac{\partial F}{\partial(\partial y_i/\partial x_m)}\right]\frac{d}{dt}\left(\frac{\partial y_i}{\partial x_m}\right) - \rho\left[{}^{\mathrm{R}}H_i^{\mathrm{L}} - \lambda\mu_i - L_m\frac{\partial\mu_i}{\partial x_m} + \frac{\partial F}{\partial\mu_i}\right]\frac{d}{dt}\mu_i - \rho\left[A_{ij}\frac{\partial x_m}{\partial y_i} - L_m\mu_j + \frac{\partial F}{\partial(\partial\mu_j/\partial x_m)}\right]\frac{d}{dt}\left(\frac{\partial\mu_j}{\partial x_m}\right) - \rho\left[\eta + \frac{\partial F}{\partial\theta}\right]\frac{d\theta}{dt} = 0. \tag{6.11}$$

Since we have introduced the proper number of undetermined multipliers in the usual Lagrangian manner, we may treat all 22 time derivatives appearing in (6.11) as if they are independent. Moreover, we assume that ${}^{\mathrm{R}}\tau$, ${}^{\mathrm{R}}\mathbf{H}^{\mathrm{L}}$, $\mathbf{A}$, $\eta$, $\lambda$, and $\mathbf{L}$ are independent of $d(\partial y_i/\partial x_m)/dt$, $d\mu_i/dt$, $d(\partial\mu_i/\partial x_m)/dt$, and $d\theta/dt$. Consequently, from (6.11), we must have

$${}^{R}\tau_{ij}\frac{\partial x_m}{\partial y_i} = \rho\frac{\partial F}{\partial(\partial y_i/\partial x_m)}, \tag{6.12a}$$

$${}^{R}H_i^{L} = -\frac{\partial F}{\partial\mu_i} + \lambda\mu_i + L_m\frac{\partial\mu_i}{\partial x_m}, \tag{6.12b}$$

$$A_{ij}\frac{\partial x_m}{\partial y_i} = -\frac{\partial F}{\partial(\partial\mu_j/\partial x_m)} + L_m\mu_j, \tag{6.12c}$$

$$\eta = -\partial F/\partial\theta. \tag{6.12d}$$

The Lagrangian multipliers may be determined by substituting from (6.12b) and (6.12c), respectively, into the conditions ${}^{\mathrm{R}}\mathbf{H}^{\mathrm{L}}\cdot\mathbf{\mu} = 0$ and $\mathbf{A}\cdot\mathbf{\mu} = 0$. The results are

$$\lambda = \frac{1}{\mu_s^2}\frac{\partial F}{\partial\mu_k}\mu_k, \quad L_m = \frac{1}{\mu_s^2}\frac{\partial F}{\partial(\partial\mu_k/\partial x_m)}\mu_k, \tag{6.13a, b}$$

where we have made use of (4.11) and the fact that the material gradient of (4.11) vanishes. Substituting from (6.13) into (6.12) and solving the resulting equations for ${}^{\mathrm{R}}\tau$ and $\mathbf{A}$, we find

$${}^{\mathrm{R}}\tau_{ij} = \rho\frac{\partial y_i}{\partial x_m}\frac{\partial F}{\partial(\partial y_j/\partial x_m)}, \tag{6.14a}$$

$${}^{\mathrm{R}}H_i^{\mathrm{L}} = -\frac{\partial F}{\partial\mu_i} + \frac{1}{\mu_s^2}\left[\frac{\partial F}{\partial\mu_k}\mu_k\mu_i + \frac{\partial F}{\partial(\partial\mu_k/\partial x_m)}\mu_k\frac{\partial\mu_i}{\partial x_m}\right], \tag{6.14b}$$

$$A_{ij} = -\frac{\partial y_i}{\partial x_m}\left[\frac{\partial F}{\partial(\partial\mu_j/\partial x_m)} - \frac{1}{\mu_s^2}\frac{\partial F}{\partial(\partial\mu_k/\partial x_m)}\mu_k\mu_j\right], \tag{6.14c}$$

$$\eta = -\frac{\partial F}{\partial\theta}, \tag{6.14d}$$

in which we have made use of the well-known relation

$$(\partial y_l/\partial x_m)(\partial x_m/\partial y_i) = \delta_{li}. \tag{6.15}$$

We must note clearly that $F$ cannot be *any* function of $\partial y_k/\partial x_l$, $\mu_k$, $\partial\mu_k/\partial x_l$, and $\theta$ because $F$ must be invariant in a rigid rotation of the deformed and magnetized body, and any arbitrary function of the 22 assumed variables (7 vectors at the point $y_k$ and a scalar) will not be so invariant.[39] Now, there is a theorem on invariant functions of several vectors due to Cauchy[40], and used by Toupin[11] in a similar connection which states that if $f(V_i^1, V_i^2, \cdots, V_i^n)$ is a single-valued function of the components of $n$ vectors which is invariant in a rigid rotation of the system of vectors, $f$ must reduce at most to a function of their lengths and scalar products, $\pi^{AB} = V_i^A V_i^B$, and the determinants of their components taken three at a time, $\Delta^{ABC} = e_{ijk}V_i^A V_j^B V_k^C$.[41] Thus, in our case, the theorem asserts that the stored energy function $F$ must reduce at most to a function of $\theta$ and

$$C_{ij} = (\partial y_k/\partial x_i)(\partial y_k/\partial x_j), \tag{6.16a}$$

$$\mu_S^2 = \mu_k\mu_k, \tag{6.16b}$$

$$G_{lm} = (\partial\mu_k/\partial x_l)(\partial\mu_k/\partial x_m), \tag{6.16c}$$

$$N_l = (\partial y_k/\partial x_l)\mu_k, \tag{6.16d}$$

$$D_{lm} = (\partial\mu_k/\partial x_l)(\partial y_k/\partial x_m), \tag{6.16e}$$

$$H_m = \mu_k(\partial\mu_k/\partial x_m), \tag{6.16f}$$

$$J = \tfrac{1}{6}e_{ijk}e_{lmn}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\frac{\partial y_k}{\partial x_n}, \tag{6.16g}$$

[39] Toupin gives an excellent discussion of this point in Ref. 11, pp. 887–888; 901–904.

[40] A. L. Cauchy, "Mémoire sur les Systèmes Isotropes de Points Matériels," Mem. Acad. Sci. XXII, 615 (1850) [Oevres (1) 2, 351.]

[41] A proof may be found in H. Weyl, *The Classical Groups, Their Invariants and Representations* (Princeton University Press, Princeton, New Jersey, 1946), Chap. 2, Sec. 9.

$$L = \tfrac{1}{6}e_{ijk}e_{lmn}\frac{\partial\mu_i}{\partial x_l}\frac{\partial\mu_j}{\partial x_m}\frac{\partial\mu_k}{\partial x_n}, \tag{6.16h}$$

$$P_{rn} = \tfrac{1}{2}e_{ijk}e_{lmr}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\frac{\partial\mu_k}{\partial x_n}, \tag{6.16i}$$

$$Q_{rn} = \tfrac{1}{2}e_{ijk}e_{lmr}\frac{\partial y_i}{\partial x_l}\frac{\partial\mu_j}{\partial x_m}\frac{\partial\mu_k}{\partial x_n}, \tag{6.16j}$$

$$W_{lm} = e_{ijk}\frac{\partial y_i}{\partial x_l}\frac{\partial\mu_j}{\partial x_m}\mu_k, \tag{6.16k}$$

$$K_n = \tfrac{1}{2}e_{ijk}e_{lmn}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\mu_k, \tag{6.16l}$$

$$R_s = \tfrac{1}{2}e_{ijk}e_{lms}\frac{\partial\mu_i}{\partial x_l}\frac{\partial\mu_j}{\partial x_m}\mu_k. \tag{6.16m}$$

Consequently, we find that the original list of variables shown in (6.9) (excluding $\theta$) can occur only in the 63 combinations listed in (6.16). We will now show that only the 18 quantities $C_{lm}$, $N_l$, $D_{lm}$ need be considered without any loss in generality, since all of the remaining 45 quantities are expressible in terms of the aforementioned 18. To this end, we first solve (6.16d) and (6.16e) for $\mu_k$ and $\partial\mu_k/\partial x_m$ in terms of $N_l$, $\partial x_l/\partial y_m$, and $D_{mn}$ and obtain

$$\mu_k = N_l(\partial x_l/\partial y_k), \tag{6.17}$$

$$\partial\mu_k/\partial x_m = D_{mn}(\partial x_n/\partial y_k), \tag{6.18}$$

and then note that

$$C^{-1}_{ij} = (\partial x_l/\partial y_k)(\partial x_j/\partial y_k), \tag{6.19}$$

where $\mathbf{C}^{-1}$ denotes the reciprocal of $\mathbf{C}$. We then substitute from (6.17) and (6.18) into (6.16) and with the aid of (6.19) obtain

$$\mu_S^2 = N_l C^{-1}_{lr} N_r, \qquad G_{lm} = D_{mn}C^{-1}_{nr}D_{lr}, \tag{6.20a,b}$$

$$H_m = N_l C^{-1}_{ln} D_{mn}. \tag{6.20c}$$

Now, it is well known[42] that

$$\det C_{mn} = J^2. \tag{6.21}$$

In view of (6.21), we evaluate $\det D_{mn}$ and find[43]

$$L = (1/J)\det D_{mn}. \tag{6.22}$$

We now substitute from (6.17) and (6.18) into the remaining relations in (6.16) and obtain

$$\begin{aligned} P_{rn} &= JC^{-1}_{rm}D_{nm}, \\ Q_{rn} &= (1/2J)e_{tks}e_{lmr}C_{lt}D_{mk}D_{ns}, \\ W_{lm} &= (1/J)e_{snr}C_{ls}D_{mn}N_r, \\ K_r &= JC^{-1}_{rs}N_s, \\ R_k &= (1/2J)e_{rst}e_{lmk}D_{lr}D_{ms}N_t. \end{aligned} \tag{6.23}$$

We have now demonstrated that the quantities appearing on the left-hand sides of Eqs. (6.20)–(6.23) are directly dependent on the $C_{lm}$, $N_l$, and $D_{lm}$, and consequently may be eliminated from the list of variables in (6.16) without any loss of generality. Thus, we have reduced $F$ to the form

$$F = F(C_{rs}, N_r, D_{rs}, \theta), \tag{6.24}$$

in place of the form shown in (6.9).

Now, we must realize that any arbitrary function $F$ of $\mathbf{C}$, $\mathbf{N}$, $\mathbf{D}$, and $\theta$ will not necessarily satisfy (4.15) by virtue of (6.14c). To find the additional restrictions on $F$ which are engendered by (4.15) we first note that

$$\frac{\partial F}{\partial(\partial\mu_l/\partial x_m)} = \frac{\partial F}{\partial D_{rs}}\frac{\partial D_{rs}}{\partial(\partial\mu_l/\partial x_m)} = \frac{\partial F}{\partial D_{ms}}\frac{\partial y_l}{\partial x_s}, \tag{6.25}$$

and then substitute from (6.25) into (6.14c) to obtain

$$A_{ij} = -\frac{\partial y_i}{\partial x_m}\frac{\partial F}{\partial D_{mr}}\left[\frac{\partial y_j}{\partial x_r} - \frac{1}{\mu_s^2}\frac{\partial y_k}{\partial x_r}\mu_k\mu_j\right]. \tag{6.26}$$

We now substitute from (6.26) into (4.15), and employ (6.17)–(6.19) with the result

$$\begin{aligned} &-\frac{\partial F}{\partial D_{mp}}D_{mr}C^{-1}_{rq} + \frac{1}{\mu_s^2}\frac{\partial F}{\partial D_{ms}}D_{mr}C^{-1}_{rq}\frac{\partial y_k}{\partial x_s}\frac{\partial x_p}{\partial y_j}\mu_k\mu_j \\ &\quad = -\frac{\partial F}{\partial D_{mq}}D_{mr}C^{-1}_{rp} + \frac{1}{\mu_s^2}\frac{\partial F}{\partial D_{ms}}D_{mr}C^{-1}_{rp}\frac{\partial y_k}{\partial x_s}\frac{\partial x_q}{\partial y_j}\mu_k\mu_j. \end{aligned} \tag{6.27}$$

These comprise a system of three independent differential equations in the 19 variables $\mathbf{C}$, $\mathbf{N}$, $\mathbf{D}$, and $\theta$, which must be satisfied by $F$. Consequently, $F$ must reduce to an arbitrary function of any $19 - 3 = 16$ functionally independent solutions of (6.27), which must be composed of $\mathbf{C}$, $\mathbf{N}$, $\mathbf{D}$, and $\theta$. It is trivially obvious that $\mathbf{C}$, $\mathbf{N}$, and $\theta$ constitute 10 such solutions. Six additional solutions are given by $D_{ia}C^{-1}_{ab}D_{jb} \equiv \Gamma_{ij}$, as may be verified by substituting any function $f$ of $\boldsymbol{\Gamma}$ into (6.27) to obtain

$$\begin{aligned} \frac{\partial f}{\partial\Gamma_{ij}}\Bigg[&-D_{ia}C^{-1}_{ap}D_{jr}C^{-1}_{rq} + \frac{1}{\mu_s^2}D_{ia}C^{-1}_{as}D_{jr}C^{-1}_{rq} \\ &\times\frac{\partial y_k}{\partial x_s}\frac{\partial x_p}{\partial y_l}\mu_k\mu_l + D_{ia}C^{-1}_{aq}D_{jr}C^{-1}_{rp} \\ &- \frac{1}{\mu_s^2}D_{ia}C^{-1}_{as}D_{jr}C^{-1}_{rp}\frac{\partial y_k}{\partial x_s}\frac{\partial x_q}{\partial y_l}\mu_k\mu_l\Bigg] = 0, \end{aligned} \tag{6.28}$$

in which we have introduced the convention $\partial f/\partial\Gamma_{ij} = \partial f/\partial\Gamma_{ji}$ and employed the relations

[42] See Ref. 2, Sec. 30 for a proof.

[43] See Appendix A for detailed derivations of Eqs. (6.22) and (6.23).

$$\frac{\partial f}{\partial D_{mq}} = \frac{\partial f}{\partial \Gamma_{ij}} \frac{\partial \Gamma_{ij}}{\partial D_{mq}} = 2 \frac{\partial f}{\partial \Gamma_{im}} D_{ia} C^{-1}_{aq}. \tag{6.29}$$

In (6.28), the second and fourth terms inside the brackets vanish identically since

$$D_{ia} C^{-1}_{as} (\partial y_k / \partial x_s) \mu_k \equiv \mu_k (\partial \mu_k / \partial x_i) = 0,$$

and the diagonal components ($i = j$) of the remaining two terms naturally cancel one another, while the off-diagonal components ($i \neq j$) cancel one another in pairs upon contraction with $\partial f / \partial \Gamma_{ij}$, and the $\Gamma_{ij}$ are indeed six solutions of (6.27). From (6.20b) we see that $\mathbf{\Gamma} \equiv \mathbf{G}$, and thus that $F$ may be reduced to the form

$$F = F(\mathbf{C}, \mathbf{N}, \mathbf{G}, \theta), \tag{6.30}$$

in place of the form shown in (6.24). It is interesting to note that $\mathbf{G}$ is invariant in a rigid rotation of the entire spin continuum with respect to the lattice continuum. Thus it is clear that condition (4.15) has served to make the exchange energy invariant in a rigid rotation of the entire spin system as it is in the quantum mechanical description.[9,10]

It should be noted that $C_{lm}$ does not vanish when the material is in its natural undeformed state. Consequently, it will be convenient for our purposes to replace Greens' deformation tensor $\mathbf{C}$ in the energy function $F$ by the material strain tensor $\mathbf{E}$, where

$$E_{lm} = \tfrac{1}{2}(C_{lm} - \delta_{lm}), \tag{6.31}$$

and does vanish when the material is in the undeformed state. Of course, since $\mathbf{C}$ is uniquely determined by $\mathbf{E}$, this replacement is always permissible and we may replace (6.30) by

$$F = F(\mathbf{E}, \mathbf{N}, \mathbf{G}, \theta). \tag{6.32}$$

From (6.32), (6.31), and (6.16), we obtain the relations

$$\frac{\partial F}{\partial (\partial y_i / \partial x_m)} = \frac{\partial F}{\partial E_{rm}} \frac{\partial y_i}{\partial x_r} + \frac{\partial F}{\partial N_m} \mu_i, \tag{6.33a}$$

$$\frac{\partial F}{\partial \mu_i} = \frac{\partial F}{\partial N_l} \frac{\partial y_i}{\partial x_l}, \tag{6.33b}$$

$$\frac{\partial F}{\partial (\partial \mu_j / \partial x_m)} = 2 \frac{\partial F}{\partial G_{rm}} \frac{\partial \mu_j}{\partial x_r}, \tag{6.33c}$$

in which we have introduced the conventions $\partial F / \partial E_{rs} = \partial F / \partial E_{sr}$, and $\partial F / \partial G_{rm} = \partial F / \partial G_{mr}$. Substituting from (6.33c) into (6.13b), we find

$$L_m = \frac{1}{\mu_s^2} \frac{\partial F}{\partial (\partial \mu_k / \partial x_m)} \mu_k = \frac{2}{\mu_s^2} \frac{\partial F}{\partial G_{rm}} \frac{\partial \mu_k}{\partial x_r} \mu_k = 0. \tag{6.34}$$

Substituting from (6.33) into (6.14) and employing (6.34), we obtain

$${}^{\mathrm{R}}\tau_{ij} = \rho \left[ \frac{\partial y_i}{\partial x_r} \frac{\partial F}{\partial E_{rm}} \frac{\partial y_j}{\partial x_m} + \frac{\partial y_i}{\partial x_m} \frac{\partial F}{\partial N_m} \mu_j \right], \tag{6.35a}$$

$${}^{\mathrm{R}}H^{\mathrm{L}}_i = -\frac{\partial F}{\partial N_l} \left[ \frac{\partial y_i}{\partial x_l} - \frac{1}{\mu_s^2} \frac{\partial y_k}{\partial x_l} \mu_k \mu_i \right], \tag{6.35b}$$

$$A_{ij} = -2 \frac{\partial y_i}{\partial x_m} \frac{\partial F}{\partial G_{mr}} \frac{\partial \mu_j}{\partial x_r}, \quad \eta = -\frac{\partial F}{\partial \theta}, \tag{6.35c}$$

in which it is to be assumed that $\partial E_{rm} / \partial E_{mr} = 0$ and $\partial G_{mr} / \partial G_{rm} = 0$ in differentiating $F$. Remembering that ${}^{\mathrm{R}}\boldsymbol{\tau} = {}^{\mathrm{R}}\boldsymbol{\tau}^{\mathrm{S}} + {}^{\mathrm{R}}\boldsymbol{\tau}^{\mathrm{A}}$, and substituting first from (5.14a), and then from (6.35b) into (6.35a) we find

$${}^{R}\tau^{S}_{ij} = \rho \left[ \frac{\partial y_i}{\partial x_r} \frac{\partial F}{\partial E_{rs}} \frac{\partial y_j}{\partial x_s} + \frac{1}{2} \frac{\partial F}{\partial N_l} \left( \frac{\partial y_i}{\partial x_l} \mu_j + \frac{\partial y_j}{\partial x_l} \mu_i \right) \right]. \tag{6.36}$$

Thus, it is clear that the antisymmetric portion of the nondissipative part of the stress tensor is derivable from an energy function and has just the value required by the conservation of angular momentum.

From (5.11), (6.2c), (6.5), (6.35b)–(6.35d), and (6.36) we may write the constitutive equations in the invariant vector form

$$\begin{aligned}
\boldsymbol{\tau}^{s} &= \rho \mathbf{y} \nabla_x \cdot \frac{\partial F}{\partial \mathbf{E}} \cdot \nabla_x \mathbf{y} + \tfrac{1}{2} \rho \left( \frac{\partial F}{\partial \mathbf{N}} \cdot \nabla_x \mathbf{y} \boldsymbol{\mu} \right)^{s} \\
&\quad + \mathbf{y} \nabla_x \cdot \left( \mathbf{Z} : \frac{d}{dt} \mathbf{E} \right) \cdot \nabla_x \mathbf{y}, \\
\mathbf{H}^{\mathrm{L}} &= -\frac{\partial F}{\partial \mathbf{N}} \cdot \nabla_x \mathbf{y} + \frac{1}{\mu_s^2} \frac{\partial F}{\partial \mathbf{N}} \cdot \nabla_x \mathbf{y} \cdot \boldsymbol{\mu} \boldsymbol{\mu} \\
&\quad - \sigma \rho \left( \frac{d}{dt} \boldsymbol{\mu} - \boldsymbol{\omega} \cdot \boldsymbol{\mu} \right), \\
\mathbf{A} &= -2 \mathbf{y} \nabla_x \cdot \partial F / \partial \mathbf{G} \cdot \nabla_x \boldsymbol{\mu}, \\
\eta &= -\partial F / \partial \theta, \quad \mathbf{q} = -\mathbf{y} \nabla_x \cdot \boldsymbol{\kappa} \cdot \nabla_x \theta.
\end{aligned} \tag{6.37}$$

Thus, all that remains in the determination of explicit constitutive relations is the selection of a form for $F$.

## 7. SURFACES OF DISCONTINUITY

In this section we determine the jump conditions at moving surfaces of discontinuity which may or may not be material. To this end we begin with (3.1), (3.2), (4.2), (4.28), (4.29), and (5.22), respectively, in the form

$$\frac{d}{dt} \int_V \rho \, dV = 0, \tag{7.1}$$

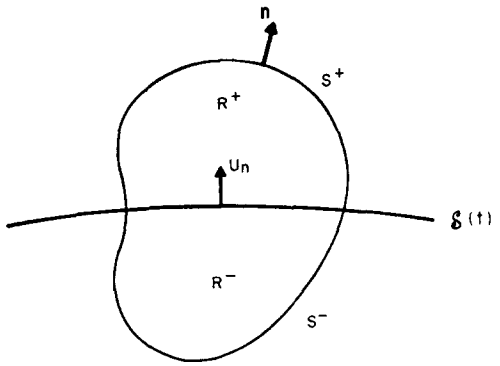


FIG. 6. Diagram showing a surface of discontinuity moving through an arbitrary region.

$$\frac{d}{dt}\int_V \rho \mathbf{v}\, dV = \int_S \mathbf{n}\cdot\boldsymbol{\tau}\, dS + \int_S \mathbf{n}\cdot\mathbf{T}^{\mathrm{M}}\, dS, \tag{7.2}$$

$$\int_S \mathbf{n}\cdot\mathbf{A}\times\rho\mathfrak{y}\, dS + \int_V \rho\mathfrak{y}\times(\mathbf{H}^{\mathrm{M}}+\mathbf{H}^{\mathrm{L}})\, dV = \frac{d}{dt}\int_V \frac{1}{\gamma}\rho\mathfrak{y}\, dV, \tag{7.3}$$

$$\oint_C \mathbf{H}^{\mathrm{M}}\cdot d\mathbf{y} = 0, \tag{7.4}$$

$$\int_S \mathbf{n}\cdot\mathbf{B}\, dS = 0, \tag{7.5}$$

$$\frac{d}{dt}\int_V \rho\eta\, dV + \int_S \mathbf{n}\cdot\frac{1}{\theta}\mathbf{q}\, dS = \int_V \Delta\, dV, \tag{7.6}$$

where $\mathbf{T}^{\mathrm{M}}$ is the Maxwell tensor defined by

$$\mathbf{T}^{\mathrm{M}} = \frac{1}{4\pi}(\mathbf{B}\mathbf{H}^{\mathrm{M}} - \tfrac{1}{2}\mathbf{H}^{\mathrm{M}}\cdot\mathbf{H}^{\mathrm{M}}\mathbf{I}_v), \tag{7.7}$$

and $\mathbf{I}_v$ is the idemfactor $\mathbf{e}_i\mathbf{e}_i$. Note that

$$\nabla_v\cdot\mathbf{T}^{\mathrm{M}} = \mathbf{M}\cdot\nabla_v\mathbf{H}^{\mathrm{M}} = \mathbf{f},$$

so that

$$\int_S \mathbf{n}\cdot\mathbf{T}^{\mathrm{M}}\, dS = \int_V \mathbf{f}\, dV,$$

and (3.2) may be obtained from (7.2) when suitable continuity conditions are assumed. Equation (3.3) is not being employed since it gives the same jump conditions as (3.2).

The jump conditions are obtained by applying (7.1)–(7.6) to a region $R$ containing a moving surface of discontinuity $\mathcal{S}$ shown in Fig. 6 and then allowing the region $R$ to shrink down to $\mathcal{S}$ in such a way that the volume of $R$ vanishes while the area of $\mathcal{S}$ remains finite as shown in Fig. 7. Equations (7.1)–(7.3) and (7.6) are presently written for a moving material region. The procedure for determining the jump conditions, consists of writing these equations for a fixed region $R$ which coincides instantaneously with the moving region while considering the fact that $\mathcal{S}$ is a surface of discontinuity moving with a speed $u_n$, which divides $R$ into two parts $R^+$ and $R^-$, in each of which the pertinent functions are continuously differentiable.[44] Thus, Eqs. (7.1)–(7.3) and (7.6) must first be converted to the proper form and then applied as stated. On the other hand, Eqs. (7.4) and (7.5) are already written for such a fixed region. We will give a detailed derivation of the jump conditions resulting from (7.3), but will simply write down the jump conditions resulting from the other equations, since they are obtained in a similar manner and are generally well known.

Equations (5.5) and (5.6) along with the divergence theorem enable us to write (7.3) in the form

$$\int_S \mathbf{n}\cdot\mathbf{A}\times\rho\mathfrak{y}\, dS - \int_S \frac{1}{\gamma}\mathbf{n}\cdot\mathbf{v}\rho\mathfrak{y}\, dS + \int_V \rho\mathfrak{y}\times(\mathbf{H}^{\mathrm{M}}+\mathbf{H}^{\mathrm{L}})\, dV = \frac{\partial}{\partial t}\int_V \frac{1}{\gamma}\rho\mathfrak{y}\, dV, \tag{7.8}$$

in which $S$ encloses a region $V$ in which $\mathbf{v}$ and $\rho\mathfrak{y}$ are continuously differentiable. Application of a generalized form of Eq. (7.8) to the region shown in Fig. 6 yields

$$\begin{aligned}&\int_{S^+}\mathbf{n}\cdot\mathbf{A}\times\rho\mathfrak{y}\, dS + \int_{S^-}\mathbf{n}\cdot\mathbf{A}\times\rho\mathfrak{y}\, dS\\ &-\int_{S^+}\frac{1}{\gamma}\mathbf{n}\cdot\mathbf{v}\rho\mathfrak{y}\, dS + \int_{\mathcal{S}}\frac{1}{\gamma^+}u_n\rho^+\mathfrak{y}^+\, dS\\ &-\int_{S^-}\frac{1}{\gamma}\mathbf{n}\cdot\mathbf{v}\rho\mathfrak{y}\, dS - \int_{\mathcal{S}}\frac{1}{\gamma^-}u_n\rho^-\mathfrak{y}^-\, dS\\ &+\int_{R^+}\rho\mathfrak{y}\times(\mathbf{H}^{\mathrm{M}}+\mathbf{H}^{\mathrm{L}})\, dV + \int_{R^-}\rho\mathfrak{y}\times(\mathbf{H}^{\mathrm{M}}+\mathbf{H}^{\mathrm{L}})\, dV\\ &-\int_{R^+}\frac{1}{\gamma}\frac{\partial}{\partial t}(\rho\mathfrak{y})\, dV - \int_{R^-}\frac{1}{\gamma}\frac{\partial}{\partial t}(\rho\mathfrak{y})\, dV = 0.\end{aligned} \tag{7.9}$$

Taking the limit of (7.9) as the volume $R^+ + R^-$ approaches zero while the area of $S$ remains finite

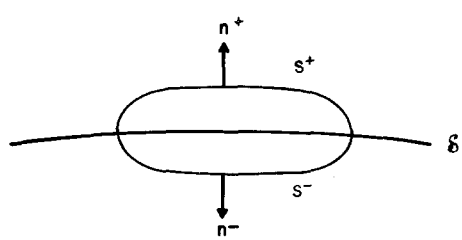


FIG. 7. Diagram showing the limiting region considered in obtaining jump conditions.

[44] For an excellent discussion of the procedure see Ref. 2, Secs. 192 and 193.

and assuming that $\rho$, $\mathfrak{y}$, $\mathbf{H}^{\mathrm{M}}$, $\mathbf{H}^{\mathrm{L}}$ and $\partial(\rho\mathfrak{y})/\partial t$ remain bounded, we find the jump condition

$$\mathbf{n}^{+}\cdot[\mathbf{A}\times\rho\mathfrak{y}]+u_n\left[\frac{1}{\gamma}\,\rho\mathfrak{y}\right]-\mathbf{n}^{+}\cdot\left[\mathbf{v}\,\frac{1}{\gamma}\,\rho\mathfrak{y}\right]=0, \tag{7.10}$$

where we have introduced the conventional notation $[\mathbf{C}]$ for $\mathbf{C}^{+}-\mathbf{C}^{-}$. Equation (7.10) must be satisfied at any possible surface of discontinuity, be it material or not. If the surface of discontinuity is material, $u_n=\mathbf{n}^{+}\cdot\mathbf{v}^{+}=\mathbf{n}^{+}\cdot\mathbf{v}^{-}$ and (7.10) reduces to

$$\mathbf{n}^{+}\cdot[\mathbf{A}\times\rho\mathfrak{y}]=0. \tag{7.11}$$

The remaining jump conditions—which may be obtained in essentially the same manner and correspond, respectively, to (7.1), (7.2), (7.6), and (7.5)—are given by

$$u_n[\rho]-\mathbf{n}^{+}\cdot[\mathbf{v}\rho]=0, \tag{7.12}$$

$$\mathbf{n}^{+}\cdot[\boldsymbol{\tau}+\mathbf{T}^{\mathrm{M}}]+u_n[\rho\mathbf{v}]-\mathbf{n}^{+}\cdot[\mathbf{v}\rho\mathbf{v}]=0, \tag{7.13}$$

$$\mathbf{n}^{+}\cdot\left[\frac{1}{\theta}\,\mathbf{q}\right]-u_n[\rho\eta]+\mathbf{n}^{+}\cdot[\mathbf{v}\rho\eta]=0, \tag{7.14}$$

$$\mathbf{n}^{+}\cdot[\mathbf{B}]=0. \tag{7.15}$$

The tangential jump condition corresponding to (7.8) is obtained in the same manner as in magnetostatics and is given by

$$\mathbf{n}^{+}\times[\mathbf{H}^{\mathrm{M}}]=0. \tag{7.16}$$

If the surface of discontinuity is material, (7.12) degenerates to nothing and (7.13) and (7.14), respectively, reduce to

$$\mathbf{n}^{+}\cdot[\boldsymbol{\tau}+\mathbf{T}^{\mathrm{M}}]=0, \tag{7.17}$$

$$\mathbf{n}^{+}\cdot\left[\frac{1}{\theta}\,\mathbf{q}\right]=0, \tag{7.18}$$

whereas (7.15) and (7.16) remain unchanged.

## 8. RECAPITULATION OF THE NONLINEAR THEORY

In this section we set forth in one place the complete system of differential equations and boundary conditions at material surfaces of discontinuity of the nonlinear magnetomechanical theory which has been developed. These consist of the following:

*The dynamic equations*

$$\nabla_y\cdot\boldsymbol{\tau}+\mathbf{M}\cdot\nabla_y\mathbf{H}^{\mathrm{M}}=\rho(d/dt)\mathbf{v},$$

$$\boldsymbol{\tau}^{\mathrm{A}}=\tfrac{1}{2}(\mathbf{M}\mathbf{H}^{\mathrm{L}}-\mathbf{H}^{\mathrm{L}}\mathbf{M}),\qquad \rho J=\rho_0,$$

$$\mathfrak{y}\times\left(\mathbf{H}^{\mathrm{M}}-\nabla_y\cdot\mathbf{A}-\frac{1}{\rho}\,\nabla_y\rho\cdot\mathbf{A}+\mathbf{H}^{\mathrm{L}}\right)=\frac{1}{\gamma}\,\frac{d}{dt}\,\mathfrak{y},$$

$$\mathbf{H}^{\mathrm{M}}=-\nabla_y\varphi,\qquad \nabla_y\cdot\mathbf{B}=0,$$

$$\rho\,\frac{d\eta}{dt}+\nabla_y\cdot\left(\frac{1}{\theta}\,\mathbf{q}\right)=\frac{1}{\theta}\,\frac{d\mathbf{E}}{dt}:\mathbf{Z}:\frac{d\mathbf{E}}{dt}+\rho^2\sigma\left(\frac{d\mathfrak{y}}{dt}-\boldsymbol{\omega}\cdot\mathfrak{y}\right)\cdot\left(\frac{d\mathfrak{y}}{dt}-\boldsymbol{\omega}\cdot\mathfrak{y}\right)+\nabla_x\theta\cdot\boldsymbol{\kappa}\cdot\nabla_x\theta. \tag{8.1a–g}$$

*The constitutive equations*

$$\boldsymbol{\tau}^{\mathrm{S}}=\rho\mathbf{y}\nabla_x\cdot\frac{\partial F}{\partial\mathbf{E}}\cdot\nabla_x\mathbf{y}+\tfrac{1}{2}\rho\left(\frac{\partial F}{\partial\mathbf{N}}\cdot\nabla_x\mathbf{y}\mathfrak{y}\right)^{\mathrm{S}}+\mathbf{y}\nabla_x\cdot\left(\mathbf{Z}:\frac{d\mathbf{E}}{dt}\right)\cdot\nabla_x\mathbf{y},$$

$$\mathbf{H}^{\mathrm{L}}=-\frac{\partial F}{\partial\mathbf{N}}\cdot\nabla_x\mathbf{y}+\frac{1}{\mu_{\mathrm{S}}^2}\,\frac{\partial F}{\partial\mathbf{N}}\cdot\nabla_x\mathbf{y}\cdot\mathfrak{y}\mathfrak{y}-\sigma\rho\left(\frac{d\mathfrak{y}}{dt}-\boldsymbol{\omega}\cdot\mathfrak{y}\right), \tag{8.2}$$

$$\mathbf{A}=-2\mathbf{y}\nabla_x\cdot\frac{\partial F}{\partial\mathbf{G}}\cdot\nabla_x\mathfrak{y},$$

$$\eta=-\frac{\partial F}{\partial\theta},\qquad \mathbf{q}=-\mathbf{y}\nabla_x\cdot\boldsymbol{\kappa}\cdot\nabla_x\theta.$$

*The relations*

$$\boldsymbol{\tau}=\boldsymbol{\tau}^{\mathrm{S}}+\boldsymbol{\tau}^{\mathrm{A}},\qquad \mathbf{v}=\frac{d}{dt}\,\mathbf{y},$$

$$\frac{d\mathbf{E}}{dt}=\tfrac{1}{2}\nabla_x\mathbf{y}\cdot(\nabla_y\mathbf{v}+\mathbf{v}\nabla_y)\cdot\mathbf{y}\nabla_x,$$

$$\boldsymbol{\omega}=\tfrac{1}{2}(\nabla_y\mathbf{v}-\mathbf{v}\nabla_y),\quad J=\det\nabla_x\mathbf{y},\quad \mathbf{E}=\tfrac{1}{2}(\nabla_x\mathbf{y}\cdot\mathbf{y}\nabla_x-\mathbf{I}_x), \tag{8.3a–j}$$

$$\mathfrak{y}=\frac{\mathbf{M}}{\rho},\qquad \mathbf{B}=\mathbf{H}^{\mathrm{M}}+4\pi\mathbf{M},$$

$$\mathbf{N}=\nabla_x\mathbf{y}\cdot\mathfrak{y},\qquad \mathbf{G}=\nabla_x\mathfrak{y}\cdot\mathfrak{y}\nabla_x.$$

*The auxiliary condition*

$$\mathfrak{y}\cdot\mathfrak{y}=\mu_{\mathrm{S}}^2, \tag{8.4}$$

and the identity

$$\nabla_y=\nabla_y\mathbf{x}\cdot\nabla_x. \tag{8.5}$$

*The boundary conditions*

$$\mathbf{n}^{+}\cdot[\boldsymbol{\tau}+(1/4\pi)(\mathbf{B}\mathbf{H}^{\mathrm{M}}-\tfrac{1}{2}\mathbf{H}^{\mathrm{M}}\cdot\mathbf{H}^{\mathrm{M}}\mathbf{I}_y)]=0,$$

$$\mathbf{n}^{+}\cdot[\mathbf{A}\times\rho\mathfrak{y}]=0,\qquad \mathbf{n}^{+}\times[\mathbf{H}^{\mathrm{M}}]=0, \tag{8.6}$$

$$\mathbf{n}^{+}\cdot[\mathbf{B}]=0,\qquad \mathbf{n}^{+}\cdot[(1/\theta)\mathbf{q}]=0.$$

Equations (8.1)–(8.3) may readily be reduced to seven equations in seven variables by employing (8.4), (8.5), and making the appropriate straightforward substitutions. The seven equations correspond to the three of (8.1a), any two of (8.1d),

(8.1f) and (8.1g). The variables are the three components of $\mathbf{y}$, and two of the three components of $\mathfrak{y}$, $\varphi$ and $\theta$. The seven variables must also satisfy the nine jump conditions in (8.6), possibly along with some statements about $[\![y]\!]$, $[\![\theta]\!]$ and $[\![(1/\mu_S^2)\mathfrak{y} \times d\mathfrak{y}/dt]\!]$ at every material surface of discontinuity. However, these equations are hopelessly nonlinear and impossible to solve in their present form, and consequently must be modified before anything can be done with them.

## 9. SUPERPOSITION OF A SMALL DYNAMIC FIELD ON A LARGE STATIC FIELD

The linear equations for a small dynamic field superimposed on a large static field will now be obtained from the system of nonlinear equations recorded in Sec. 8. However, before proceeding we must introduce the intermediate configuration $\boldsymbol{\xi}$ in addition to the natural (material) and present (spatial) configurations $\mathbf{x}$ and $\mathbf{y}$. The intermediate configuration is the present configuration at some time $t = {}_0t$. Consequently, the brief discussion of deformation theory given in Sec. 3 and Eq. (3.12) imply that

$$\begin{aligned} \xi_i &= \xi_i(x_k), \qquad x_i = x_i(\xi_k), \\ y_i &= y_i(\xi_k, t), \qquad \xi_i = \xi_i(y_k, t). \end{aligned} \tag{9.1}$$

Since we are concerned with a small deformation superposed on a large, we define the mechanical displacement $\mathbf{u}$ by

$$y_i(\boldsymbol{\xi}, t) = \xi_i + u_i(\boldsymbol{\xi}, t), \tag{9.2}$$

with the restriction

$$|\partial u_i/\partial \xi_k| \ll 1. \tag{9.3}$$

Also, let ${}_0\mathfrak{y}$, ${}_0\varphi$ and ${}_0\theta$ denote the remaining portion of a static solution of the nonlinear equations referred to the intermediate configuration $\boldsymbol{\xi}$. Moreover, only situations in which ${}_0\theta$ is constant are being considered. Then we may define the dynamic portions of the magnetic moment per unit mass $\tilde{\mathfrak{y}}$, the magnetic potential $\tilde{\varphi}$ and the temperature $\tilde{\theta}$ by

$$\mathfrak{y}(\boldsymbol{\xi}, t) = {}_0\mathfrak{y}(\boldsymbol{\xi}) + \tilde{\mathfrak{y}}(\boldsymbol{\xi}, t), \tag{9.4a}$$

$$\varphi(\boldsymbol{\xi}, t) = {}_0\varphi(\boldsymbol{\xi}) + \tilde{\varphi}(\boldsymbol{\xi}, t), \tag{9.4b}$$

$$\theta(\boldsymbol{\xi}, t) = {}_0\theta + \tilde{\theta}(\boldsymbol{\xi}, t), \tag{9.4c}$$

with the restrictions

$$\frac{|\tilde{\mathfrak{y}}|}{|{}_0\mathfrak{y}|} \ll 1, \qquad \frac{\tilde{\varphi}}{{}_0\varphi} \ll 1, \qquad \frac{\tilde{\theta}}{{}_0\theta} \ll 1. \tag{9.5}$$

We now wish to express the constitutive equations (8.2) in terms of the small dynamic variables $\mathbf{u}$, $\tilde{\mathfrak{y}}$, $\tilde{\varphi}$, and $\tilde{\theta}$. However, before we can do this we must first express all quantities appearing in (8.2) in terms of the small dynamic variables and then write the partial derivatives of $F$ with respect to $\mathbf{E}$, $\mathbf{N}$, $\mathbf{G}$, and $\theta$ as a power series in these newly defined variables, retaining, of course, only linear dynamic terms.

Now, from (9.1) and the chain rule of differentiation, we may write

$$\frac{\partial y_k}{\partial x_l} = \frac{\partial \xi_m}{\partial x_l}\frac{\partial y_k}{\partial \xi_m}, \tag{9.6}$$

and from (9.2) we obtain

$$\partial y_k/\partial \xi_m = \delta_{km} + \partial u_k/\partial \xi_m. \tag{9.7}$$

Substituting from (9.4a), (9.6), and (9.7) into (8.3f), (8.3i), and (8.3j), respectively, and retaining only linear dynamic terms, we find

$$\begin{aligned} E_{pq} &= {}_0E_{pq} + \frac{1}{2}\frac{\partial \xi_j}{\partial x_p}\frac{\partial \xi_n}{\partial x_q}\left(\frac{\partial u_j}{\partial \xi_n} + \frac{\partial u_n}{\partial \xi_j}\right), \\ N_p &= {}_0N_p + \frac{\partial \xi_k}{\partial x_p}\tilde{\mu}_k + \frac{\partial \xi_m}{\partial x_p}{}_0\mu_k\frac{\partial u_k}{\partial \xi_m}, \\ G_{pq} &= {}_0G_{pq} + \left(\frac{\partial\, {}_0\mu_k}{\partial x_p}\frac{\partial \xi_r}{\partial x_q} + \frac{\partial\, {}_0\mu_k}{\partial x_q}\frac{\partial \xi_r}{\partial x_p}\right)\frac{\partial \tilde{\mu}_k}{\partial \xi_r}, \end{aligned} \tag{9.8}$$

in which the static quantities have been defined by

$$\begin{aligned} {}_0E_{pq} &= \frac{1}{2}\left(\frac{\partial \xi_k}{\partial x_p}\frac{\partial \xi_k}{\partial x_q} - \delta_{pq}\right), \\ {}_0N_l &= \frac{\partial \xi_k}{\partial x_l}{}_0\mu_k, \qquad {}_0G_{pq} = \frac{\partial\, {}_0\mu_k}{\partial x_p}\frac{\partial\, {}_0\mu_k}{\partial x_q}. \end{aligned} \tag{9.9}$$

Expansion of $\partial F/\partial E_{rs}$ in a Taylor series about the $\boldsymbol{\xi}$ configuration yields

$$\begin{aligned} \frac{\partial F}{\partial E_{rs}} &= \frac{\partial F}{\partial E_{rs}}\Big)_0 + \frac{\partial^2 F}{\partial E_{pq}\,\partial E_{rs}}\Big)_0 (E_{pq} - {}_0E_{pq}) \\ &+ \frac{\partial^2 F}{\partial N_p\,\partial E_{rs}}\Big)_0 (N_p - {}_0N_p) + \frac{\partial^2 F}{\partial G_{pq}\,\partial E_{rs}}\Big)_0 \\ &\times (G_{pq} - {}_0G_{pq}) + \frac{\partial^2 F}{\partial \theta\,\partial E_{rs}}\Big)_0 (\theta - {}_0\theta) + \cdots. \end{aligned} \tag{9.10}$$

Since only the equations which are linear in the dynamic variables are considered here, terms of higher order than the first may be ignored. Equations analogous to (9.10) may be obtained for the quantities $\partial F/\partial N_l$, $\partial F/\partial G_{rs}$, and $\partial F/\partial \theta$ simply by replacing $E_{rs}$ in (9.10), respectively, by $N_l$, $G_{rs}$, and $\theta$. Substituting from (9.8) and (9.4c) into (9.10) yields

$$\frac{\partial F}{\partial E_{rs}} = \frac{\partial F}{\partial\, {}_0E_{rs}} + \frac{1}{2}\frac{\partial^2 F}{\partial\, {}_0E_{pq}\partial\, {}_0E_{rs}}\frac{\partial \xi_j}{\partial x_p}\frac{\partial \xi_n}{\partial x_q}\left(\frac{\partial u_j}{\partial \xi_n} + \frac{\partial u_n}{\partial \xi_j}\right)$$

$$+\frac{\partial^2 F}{\partial\,{}_0N_p\partial\,{}_0E_{rs}}\left(\frac{\partial\xi_k}{\partial x_p}\tilde{\mu}_k+\frac{\partial\xi_m}{\partial x_p}\,{}_0\mu_k\frac{\partial u_k}{\partial\xi_m}\right)$$

$$+\frac{\partial^2 F}{\partial\,{}_0G_{pq}\partial\,{}_0E_{rs}}\left(\frac{\partial\,{}_0\mu_k}{\partial x_p}\frac{\partial\xi_r}{\partial x_q}\right.$$

$$\left.+\frac{\partial\,{}_0\mu_k}{\partial x_q}\frac{\partial\xi_r}{\partial x_p}\right)\frac{\partial\tilde{\mu}_k}{\partial\xi_r}+\frac{\partial^2 F}{\partial\,{}_0\theta\partial\,{}_0E_{rs}}\tilde{\theta}, \tag{9.11}$$

with analogous expressions available for $\partial F/\partial N_l$, $\partial F/\partial G_{rs}$, and $\partial F/\partial\theta$. From (8.1c), (8.3e), (9.6), and the fact that the determinant of a matrix product is equal to the product of the determinants of the two matrices, we obtain

$$\rho=\rho_0(\det\partial\xi_m/\partial x_l)^{-1}(\det\partial y_k/\partial\xi_m)^{-1}, \tag{9.12}$$

in which $\rho_0\,(\det\partial\xi_m/\partial x_l)^{-1}\equiv\tilde{\rho}$ represents the mass density in the intermediate configuration. Now, from (9.7) and (9.3), in the linear approximation we obtain

$$(\det\partial y_k/\partial\xi_m)^{-1}=1-\partial u_k/\partial\xi_k. \tag{9.13}$$

Thus, we may write

$$\rho=\tilde{\rho}(1-\partial u_k/\partial\xi_k). \tag{9.14}$$

From (9.2) we see that (8.3b) may be written

$$v_i=\partial u_i/\partial t. \tag{9.15}$$

From (9.2) and (9.1d) we obtain

$$\delta_{km}=\partial\xi_k/\partial y_m+\partial u_k/\partial y_m. \tag{9.16}$$

From the chain rule of differentiation, (9.7), (9.15), and (9.16), we find that in the linear approximation (8.3c) and (8.3d), respectively, may be written

$$\frac{d}{dt}E_{rs}=\frac{1}{2}\frac{\partial\xi_k}{\partial x_r}\frac{\partial\xi_l}{\partial x_s}\frac{\partial}{\partial t}\left(\frac{\partial u_k}{\partial\xi_l}+\frac{\partial u_l}{\partial\xi_k}\right),$$
$$\omega_{kl}=\frac{1}{2}\frac{\partial}{\partial t}\left(\frac{\partial u_k}{\partial\xi_l}-\frac{\partial u_l}{\partial\xi_k}\right). \tag{9.17}$$

We may now substitute from (9.4), (9.6), (9.7), (9.11), (9.14), and (9.17) into (8.2) and after some tedious manipulation in which only static terms and terms linear in the dynamic variables are retained, we obtain

$$\tau_{ij}^{\mathrm{S}}={}_0\tau_{ij}^{\mathrm{S}}+\tilde{\tau}_{ij}^{\mathrm{S}},\qquad H_i^{\mathrm{L}}={}_0H_i^{\mathrm{L}}+\tilde{H}_i^{\mathrm{L}},$$
$$A_{ij}={}_0A_{ij}+\tilde{A}_{ij},\qquad \eta=\tilde{\eta},\qquad q_i=\tilde{q}_i, \tag{9.18}$$

in which the static terms are given by

$${}_0\tau_{ij}^{\mathrm{S}}=\tilde{\rho}\left[\frac{\partial F}{\partial\,{}_0E_{rs}}\frac{\partial\xi_i}{\partial x_r}\frac{\partial\xi_j}{\partial x_s}\right.$$
$$\left.+\frac{1}{2}\frac{\partial F}{\partial\,{}_0N_l}\left(\frac{\partial\xi_i}{\partial x_l}\,{}_0\mu_j+\frac{\partial\xi_j}{\partial x_l}\,{}_0\mu_i\right)\right],$$
$${}_0H_i^{\mathrm{L}}=-\frac{\partial F}{\partial\,{}_0N_l}\left(\frac{\partial\xi_i}{\partial x_l}-\frac{1}{\mu_s^2}\frac{\partial\xi_k}{\partial x_l}\,{}_0\mu_k\,{}_0\mu_i\right), \tag{9.19}$$
$${}_0A_{ij}=-2\frac{\partial\xi_i}{\partial x_m}\frac{\partial F}{\partial\,{}_0G_{mn}}\frac{\partial\,{}_0\mu_j}{\partial x_n},$$

and the dynamic terms are given by

$$\tilde{\tau}_{ij}^{\mathrm{S}}=\tilde{c}_{ijkm}\frac{\partial u_k}{\partial\xi_m}+\tilde{e}_{kij}\tilde{\mu}_k+\tilde{\gamma}_{ijkm}\frac{\partial\tilde{\mu}_k}{\partial\xi_m}$$
$$+a_{ij}\tilde{\theta}+\tfrac{1}{2}\tilde{Z}_{ijkl}\frac{\partial}{\partial t}\left(\frac{\partial u_k}{\partial\xi_l}+\frac{\partial u_l}{\partial\xi_k}\right),$$
$$\tilde{H}_i^{\mathrm{L}}=\tilde{\chi}_{ik}\tilde{\mu}_k+\tilde{g}_{ikm}\frac{\partial u_k}{\partial\xi_m}+\tilde{\eta}_{irj}\frac{\partial\tilde{\mu}_r}{\partial\xi_j}$$
$$+b_i\tilde{\theta}-\sigma\tilde{\rho}\left[\frac{\partial\tilde{\mu}_i}{\partial t}-\frac{\partial}{\partial t}\left(\frac{\partial u_i}{\partial\xi_j}-\frac{\partial u_j}{\partial\xi_i}\right){}_0\mu_j\right],$$
$$\tilde{A}_{ij}=\tilde{\beta}_{ijkl}\frac{\partial\tilde{\mu}_l}{\partial\xi_k}+\tilde{\nu}_{kij}\tilde{\mu}_k$$
$$+\tilde{h}_{ijkl}\frac{\partial u_k}{\partial\xi_l}+f_{ij}\tilde{\theta}, \tag{9.20a–e}$$
$$\eta=C\tilde{\theta}+\pi_{kl}\frac{\partial u_k}{\partial\xi_l}+\tilde{\psi}_k\tilde{\mu}_k+p_{kl}\frac{\partial\tilde{\mu}_k}{\partial\xi_l},$$
$$\tilde{q}_i=-\tilde{\kappa}_{ij}\frac{\partial\tilde{\theta}}{\partial\xi_j},$$

where the unfamiliar coefficients in (9.20) are given by

---

$$\tilde{c}_{ijkm}=\tilde{\rho}\left[\frac{\partial F}{\partial\,{}_0E_{rs}}\left(-\frac{\partial\xi_i}{\partial x_r}\frac{\partial\xi_j}{\partial x_s}\delta_{km}+\frac{\partial\xi_i}{\partial x_r}\frac{\partial\xi_m}{\partial x_s}\delta_{jk}+\frac{\partial\xi_j}{\partial x_s}\frac{\partial\xi_m}{\partial x_r}\delta_{ik}\right)\right.$$

$$-\frac{1}{2}\frac{\partial F}{\partial\,{}_0N_l}\left(\frac{\partial\xi_i}{\partial x_l}\,{}_0\mu_j\delta_{km}+\frac{\partial\xi_j}{\partial x_l}\,{}_0\mu_i\delta_{km}-\frac{\partial\xi_m}{\partial x_l}\,{}_0\mu_j\delta_{ik}-\frac{\partial\xi_m}{\partial x_l}\,{}_0\mu_i\delta_{jk}\right)+\frac{\partial^2 F}{\partial\,{}_0E_{pq}\partial\,{}_0E_{rs}}\frac{\partial\xi_i}{\partial x_r}\frac{\partial\xi_j}{\partial x_s}\frac{\partial\xi_k}{\partial x_p}\frac{\partial\xi_m}{\partial x_q}+\frac{\partial^2 F}{\partial\,{}_0E_{rs}\partial\,{}_0N_l}$$

$$\left.\times\left(\frac{1}{2}\frac{\partial\xi_k}{\partial x_r}\frac{\partial\xi_m}{\partial x_s}\frac{\partial\xi_i}{\partial x_l}\,{}_0\mu_j+\frac{1}{2}\frac{\partial\xi_k}{\partial x_r}\frac{\partial\xi_m}{\partial x_s}\frac{\partial\xi_j}{\partial x_l}\,{}_0\mu_i+\frac{\partial\xi_i}{\partial x_r}\frac{\partial\xi_j}{\partial x_s}\frac{\partial\xi_m}{\partial x_l}\,{}_0\mu_k\right)+\frac{1}{2}\frac{\partial^2 F}{\partial\,{}_0N_p\partial\,{}_0N_l}\frac{\partial\xi_m}{\partial x_p}\,{}_0\mu_k\left(\frac{\partial\xi_i}{\partial x_l}\,{}_0\mu_j+\frac{\partial\xi_j}{\partial x_l}\,{}_0\mu_i\right)\right],$$

$$\tilde{e}_{kij}=\tilde{\rho}\left[\frac{1}{2}\frac{\partial F}{\partial\,{}_0N_l}\left(\frac{\partial\xi_i}{\partial x_l}\delta_{jk}+\frac{\partial\xi_j}{\partial x_l}\delta_{ik}\right)+\frac{\partial^2 F}{\partial\,{}_0N_p\partial\,{}_0E_{rs}}\frac{\partial\xi_i}{\partial x_r}\frac{\partial\xi_j}{\partial x_s}\frac{\partial\xi_k}{\partial x_p}+\frac{1}{2}\frac{\partial^2 F}{\partial\,{}_0N_p\partial\,{}_0N_l}\frac{\partial\xi_k}{\partial x_p}\left(\frac{\partial\xi_i}{\partial x_l}\,{}_0\mu_j+\frac{\partial\xi_j}{\partial x_l}\,{}_0\mu_i\right)\right],$$

$$\tilde{\gamma}_{ijkm} = \tilde{\rho}\left[2\frac{\partial^2 F}{\partial\,{}_0G_{pq}\partial\,{}_0E_{rs}}\frac{\partial\xi_i}{\partial x_r}\frac{\partial\xi_j}{\partial x_s}\frac{\partial\,{}_0\mu_k}{\partial x_p}\frac{\partial\xi_m}{\partial x_q} + \frac{\partial^2 F}{\partial\,{}_0G_{pq}\partial\,{}_0N_l}\frac{\partial\xi_m}{\partial x_q}\frac{\partial\,{}_0\mu_k}{\partial x_p}\left(\frac{\partial\xi_i}{\partial x_l}\,{}_0\mu_j + \frac{\partial\xi_j}{\partial x_l}\,{}_0\mu_i\right)\right],$$

$$a_{ij} = \tilde{\rho}\left[\frac{\partial^2 F}{\partial\,{}_0\theta\partial\,{}_0E_{rs}}\frac{\partial\xi_i}{\partial x_r}\frac{\partial\xi_j}{\partial x_s} + \frac{1}{2}\frac{\partial^2 F}{\partial\,{}_0\theta\partial\,{}_0N_l}\left(\frac{\partial\xi_i}{\partial x_l}\,{}_0\mu_j + \frac{\partial\xi_j}{\partial x_l}\,{}_0\mu_i\right)\right],$$

$$\tilde{Z}_{ijkl} = Z_{pqrs}\frac{\partial\xi_i}{\partial x_p}\frac{\partial\xi_j}{\partial x_q}\frac{\partial\xi_k}{\partial x_r}\frac{\partial\xi_l}{\partial x_s},$$

$$\tilde{\chi}_{ik} = \left[\frac{1}{\mu_S^2}\frac{\partial F}{\partial\,{}_0N_l}\frac{\partial\xi_n}{\partial x_l}({}_0\mu_n\delta_{ik} + {}_0\mu_i\delta_{nk}) - \frac{\partial^2 F}{\partial\,{}_0N_p\partial\,{}_0N_l}\frac{\partial\xi_k}{\partial x_p}\left(\frac{\partial\xi_i}{\partial x_l} - \frac{1}{\mu_s^2}\frac{\partial\xi_n}{\partial x_l}\,{}_0\mu_n\,{}_0\mu_i\right)\right],$$

$$\tilde{g}_{ikm} = \left[-\frac{\partial F}{\partial\,{}_0N_l}\frac{\partial\xi_m}{\partial x_l}\left(\delta_{ki} - \frac{1}{\mu_s^2}\,{}_0\mu_k\,{}_0\mu_i\right) - \frac{\partial^2 F}{\partial\,{}_0E_{pq}\partial\,{}_0N_l}\frac{\partial\xi_k}{\partial x_p}\frac{\partial\xi_m}{\partial x_q}\left(\frac{\partial\xi_i}{\partial x_l} - \frac{1}{\mu_s^2}\frac{\partial\xi_n}{\partial x_l}\,{}_0\mu_n\,{}_0\mu_i\right)\right.$$
$$\left. - \frac{\partial^2 F}{\partial\,{}_0N_p\partial\,{}_0N_l}\frac{\partial\xi_m}{\partial x_p}\left(\frac{\partial\xi_i}{\partial x_l}\,{}_0\mu_k - \frac{1}{\mu_s^2}\frac{\partial\xi_n}{\partial x_l}\,{}_0\mu_n\,{}_0\mu_i\,{}_0\mu_k\right)\right],$$

$$\tilde{\eta}_{irj} = -2\frac{\partial^2 F}{\partial\,{}_0G_{pq}\partial\,{}_0N_l}\frac{\partial\xi_j}{\partial x_q}\frac{\partial\,{}_0\mu_r}{\partial x_p}\left(\frac{\partial\xi_i}{\partial x_l} - \frac{1}{\mu_s^2}\frac{\partial\xi_k}{\partial x_l}\,{}_0\mu_k\,{}_0\mu_i\right),$$

$$b_i = -\frac{\partial^2 F}{\partial\,{}_0\theta\partial\,{}_0N_l}\left(\frac{\partial\xi_i}{\partial x_l} - \frac{1}{\mu_s^2}\frac{\partial\xi_k}{\partial x_l}\,{}_0\mu_k\,{}_0\mu_i\right),$$

$$\tilde{\beta}_{ijkl} = -2\left[\frac{\partial F}{\partial\,{}_0G_{mr}}\frac{\partial\xi_i}{\partial x_m}\frac{\partial\xi_k}{\partial x_r}\delta_{jl} + 2\frac{\partial^2 F}{\partial\,{}_0G_{pq}\partial\,{}_0G_{mr}}\frac{\partial\xi_i}{\partial x_m}\frac{\partial\xi_k}{\partial x_q}\frac{\partial\,{}_0\mu_j}{\partial x_r}\frac{\partial\,{}_0\mu_l}{\partial x_p}\right],$$

$$\tilde{\nu}_{kij} = -2\frac{\partial^2 F}{\partial\,{}_0N_p\partial\,{}_0G_{mr}}\frac{\partial\xi_i}{\partial x_m}\frac{\partial\xi_k}{\partial x_p}\frac{\partial\,{}_0\mu_j}{\partial x_r},$$

$$h_{ijkl} = -2\left[\frac{\partial F}{\partial\,{}_0G_{mr}}\frac{\partial\xi_l}{\partial x_m}\frac{\partial\,{}_0\mu_j}{\partial x_r}\delta_{ki} + \frac{\partial^2 F}{\partial\,{}_0E_{pq}\partial\,{}_0G_{mr}}\frac{\partial\xi_i}{\partial x_m}\frac{\partial\xi_k}{\partial x_p}\frac{\partial\xi_l}{\partial x_q}\frac{\partial\,{}_0\mu_j}{\partial x_r} + \frac{\partial^2 F}{\partial\,{}_0N_p\partial\,{}_0G_{mr}}\frac{\partial\xi_i}{\partial x_m}\frac{\partial\xi_l}{\partial x_p}\frac{\partial\,{}_0\mu_j}{\partial x_r}\,{}_0\mu_k\right],$$

$$f_{ij} = -2\frac{\partial^2 F}{\partial\,{}_0\theta\partial\,{}_0G_{mr}}\frac{\partial\xi_i}{\partial x_m}\frac{\partial\,{}_0\mu_j}{\partial x_r},$$

$$\pi_{kl} = -\left[\frac{\partial^2 F}{\partial\,{}_0E_{pq}\partial\,{}_0\theta}\frac{\partial\xi_k}{\partial x_p}\frac{\partial\xi_l}{\partial x_q} + \frac{\partial^2 F}{\partial\,{}_0N_p\partial\,{}_0\theta}\frac{\partial\xi_l}{\partial x_p}\,{}_0\mu_k\right],$$

$$C = -\frac{\partial^2 F}{\partial\,{}_0\theta^2}, \qquad \tilde{\psi}_k = -\frac{\partial^2 F}{\partial\,{}_0N_l\partial\,{}_0\theta}\frac{\partial\xi_k}{\partial x_l}, \qquad p_{kl} = -2\frac{\partial^2 F}{\partial\,{}_0G_{pq}\partial\,{}_0\theta}\frac{\partial\xi_l}{\partial x_q}\frac{\partial\,{}_0\mu_k}{\partial x_p},$$

$$\tilde{\kappa}_{ij} = \kappa_{rm}\frac{\partial\xi_i}{\partial x_r}\frac{\partial\xi_j}{\partial x_m}. \tag{9.21}$$

Utilizing the chain rule of differentiation, substituting from (9.18), (9.4), (9.14)–(9.16) into (8.1a), (8.1b), (8.1d), (8.1e), (8.1f) and (8.1g), respectively, employing the static solution of the nonlinear equations referred to the intermediate configuration $\xi$ and retaining only linear dynamic terms, we obtain

$$\frac{\partial\tilde{\tau}_{ij}}{\partial\xi_i} + \tilde{\rho}\,{}_0\mu_i\frac{\partial\tilde{H}_j^{\mathrm{M}}}{\partial\xi_i} - \left(\frac{\partial\,{}_0\tau_{ij}}{\partial\xi_m} + \tilde{\rho}\,{}_0\mu_k\frac{\partial\,{}_0H_j^{\mathrm{M}}}{\partial\xi_k}\delta_{im} + \tilde{\rho}\,{}_0\mu_i\frac{\partial\,{}_0H_j^{\mathrm{M}}}{\partial\xi_m}\right)\frac{\partial u_m}{\partial\xi_i} + \tilde{\rho}\frac{\partial\,{}_0H_j^{\mathrm{M}}}{\partial\xi_i}\tilde{\mu}_i = \tilde{\rho}\frac{\partial^2 u_j}{\partial t^2},$$

$$\tilde{\tau}_{ij}^{\mathrm{A}} = \tfrac{1}{2}\tilde{\rho}\left[\tilde{\mu}_i\,{}_0H_j^{\mathrm{L}} - {}_0\mu_i\tilde{H}_j^{\mathrm{L}} - {}_0H_i^{\mathrm{L}}\tilde{\mu}_j - {}_0\mu_j\tilde{H}_i^{\mathrm{L}} - ({}_0\mu_i\,{}_0H_j^{\mathrm{L}} - {}_0H_i^{\mathrm{L}}\,{}_0\mu_j)\frac{\partial u_k}{\partial\xi_k}\right],$$

$$e_{ijk}\,{}_0\mu_j\left[\tilde{H}_k^{\mathrm{M}} - \frac{\partial\tilde{A}_{lk}}{\partial\xi_l} + \tilde{H}_k^{\mathrm{L}} - {}_0A_{lk}\frac{\partial^2 u_m}{\partial\xi_l\,\partial\xi_m} + \left(\frac{\partial\,{}_0A_{lk}}{\partial\xi_m} - \frac{{}_0A_{lk}}{\tilde{\rho}}\frac{\partial\tilde{\rho}}{\partial\xi_m}\right)\frac{\partial u_m}{\partial\xi_l} - \frac{\tilde{A}_{lk}}{\tilde{\rho}}\frac{\partial\tilde{\rho}}{\partial\xi_l}\right]$$
$$+ \frac{1}{\mu_s^2}e_{ijk}\tilde{\mu}_j\,{}_0\mu_k\,{}_0\mu_r\left[{}_0H_r^{M} - \frac{\partial\,{}_0A_{lr}}{\partial\xi_l}\right] = \frac{1}{\gamma}\frac{\partial\tilde{\mu}_i}{\partial t},$$

$$\tilde{H}_i^{\mathrm{M}} = -\frac{\partial\tilde{\varphi}}{\partial\xi_i} - {}_0H_k^{M}\frac{\partial u_k}{\partial\xi_i},$$

$$\frac{\partial \tilde{H}_i^M}{\partial \xi_i} + 4\pi \frac{\partial}{\partial \xi_i}(\tilde{\rho}\tilde{\mu}_i) - \frac{\partial}{\partial \xi_m}({}_0H_i^M + 4\pi\tilde{\rho}\ {}_0\mu_i)\frac{\partial u_m}{\partial \xi_i} - 4\pi \frac{\partial}{\partial \xi_k}\left(\tilde{\rho}\ {}_0\mu_k \frac{\partial u_i}{\partial \xi_i}\right) = 0,$$

$${}_0\theta\tilde{\rho}\frac{\partial \eta}{\partial t} = Z_{ijkl}\frac{d}{dt}E_{ij}\frac{d}{dt}E_{kl} + \sigma\tilde{\rho}^2\left(\frac{\partial}{\partial t}\tilde{\mu}_k - \omega_{kl}\ {}_0\mu_l\right)\left(\frac{\partial}{\partial t}\tilde{\mu}_k - \omega_{kj}\ {}_0\mu_j\right) - \frac{\partial \tilde{q}_k}{\partial \xi_k}, \qquad (9.22a\text{–}f)$$

where $d\mathbf{E}/dt$ and $\omega$ are given by (9.17) and $\tilde{\tau}^*$, $\tilde{\mathbf{H}}^L$, $\tilde{\mathbf{A}}$, $\tilde{\eta}$, and $\tilde{\mathbf{q}}$ are given by the constitutive equations (9.20). Thus we see that if the dissipation equation (9.22f) is coupled to the others, the system is still nonlinear even in this simple case. However, if either

$$\frac{\partial^2 F}{\partial\ {}_0\theta \partial\ {}_0E_{rs}} = \frac{\partial^2 F}{\partial\ {}_0\theta \partial\ {}_0N_t} = \frac{\partial^2 F}{\partial\ {}_0\theta \partial\ {}_0G_{rs}} = 0, \qquad (9.23)$$

or

$$Z_{ijkl} = \sigma = 0, \qquad (9.24)$$

the system can be linearized. If all of (9.23) are satisfied then (9.22f) and (9.20d, e) uncouple from the remaining equations, none of which then contain $\tilde{\theta}$. The remaining equations can then be solved independently of (9.22f) and (9.20d, e), after which these latter equations, which are linear in $\tilde{\theta}$, may be satisfied to find $\tilde{\theta}$. If on the other hand all of (9.24) are satisfied but some of (9.23) are not the system remains coupled, but (9.22f) becomes linear.

The boundary conditions are obtained by substituting from (9.4), (9.14), and (9.18) into (8.6), employing the static boundary conditions while assuming that the configuration $\mathbf{y}$ is approximately the configuration $\boldsymbol{\xi}$ and retaining only linear dynamic terms with the result

$$\begin{aligned}
&\mathbf{n}^+\cdot[\tilde{\tau} + (1/4\pi)(\tilde{\mathbf{B}}\ {}_0\mathbf{H}^M + {}_0\mathbf{B}\tilde{\mathbf{H}}^M \\
&\qquad - {}_0\mathbf{H}^M\cdot\tilde{\mathbf{H}}^M\mathbf{I}_\xi)] = 0, \\
&\mathbf{n}^+\cdot[{}_0\mathbf{A}\times\tilde{\rho}\tilde{\mathbf{u}} + \tilde{\mathbf{A}}\times\tilde{\rho}\ {}_0\mathbf{u} - {}_0\mathbf{A}\times\tilde{\rho}\ {}_0\mathbf{u}\nabla_\xi\cdot\mathbf{u}] = 0, \qquad (9.25) \\
&\mathbf{n}^+\times[\tilde{\mathbf{H}}^M] = 0, \quad \mathbf{n}\cdot[\tilde{\mathbf{B}}] = 0, \quad \mathbf{n}^+\cdot[\tilde{\mathbf{q}}] = 0,
\end{aligned}$$

where

$$\begin{aligned}
{}_0\mathbf{B} &= {}_0\mathbf{H}^M + 4\pi\tilde{\rho}\ {}_0\mathbf{u}, \\
\tilde{\mathbf{B}} &= \tilde{\mathbf{H}}^M + 4\pi\tilde{\rho}\tilde{\mathbf{u}} - 4\pi\tilde{\rho}\ {}_0\mathbf{u}\nabla_\xi\cdot\mathbf{u}, \qquad (9.26) \\
\nabla_\xi &= \mathbf{e}_i\,\partial/\partial\xi_i.
\end{aligned}$$

Along with (9.25) there may be some statements about $[\mathbf{u}]$, $[\tilde{\theta}]$ and $[1/\mu_0^2\ {}_0\mathbf{u}\times\partial\tilde{\mathbf{u}}/\partial t]$.

We now take a polynomial approximation for $F$ in the form

$$\begin{aligned}
F = {}&\frac{1}{2\rho_0}c_{ijkl}E_{ij}E_{kl} + \tfrac{1}{2}\rho_0\chi_{ij}N_iN_j + \tfrac{1}{2}\rho_0\alpha_{ij}G_{ij} \\
&+ \tfrac{1}{2}c\theta^2 + \epsilon_{kij}E_{ij}N_k + \rho_0 b_{ijkl}E_{ij}N_kN_l \\
&+ \rho_0^2 f_{kij}G_{ij}N_k + \rho_0\gamma_{ijkm}E_{ij}G_{km} + \nu_{ij}E_{ij}\theta \\
&+ \text{higher-order terms in } \mathbf{N}, \qquad (9.27)
\end{aligned}$$

where the constants $c_{ijkl}$, $\chi_{ij}$, $\alpha_{ij}$, $c$, $\epsilon_{kij}$, $b_{ijkl}$, $f_{kij}$, $\gamma_{ijkm}$, and $\nu_{ij}$ will be referred to as the elastic, anisotropy, exchange, thermal, piezomagnetic, magnetostrictive, magnetoexchange, exchangestrictive and thermoelastic constants, respectively, Note that no thermomagnetic coupling term has been included in (9.27). We are now in a position to formulate and solve boundary-value problems, which we will do in a forthcoming paper.

It is interesting to note that the material constants in (9.20) are composed of combinations of the material constants in (9.27). However, on the basis of existing experimental evidence for yttrium iron garnet, one term in each combination is dominant and all other terms in that combination are negligibly small compared with it. On the basis of the same experimental information the antisymmetric portion of the stress tensor turns out to be negligibly small compared with the symmetric portion. On the other hand, it does not seem evident *a priori* that such terms will always be negligible. These matters will be discussed in detail in a forthcoming paper.


## ACKNOWLEDGMENTS

The author wishes to thank J. H. Rowen and H. Matthews for motivating and encouraging this investigation throughout and for stimulating discussions. The author also would like to acknowledge valuable discussions with R. N. Thurston and G. A. Coquin, and thank M. Lax and J. A. Lewis for reading and commenting on much of the text.


## APPENDIX A

To derive (6.22) consider the scalar

$$\det D_{mn} = \tfrac{1}{6}e_{ijk}e_{lmn}D_{il}D_{jm}D_{kn},$$

from which, with (6.16e), we obtain

$$\det D_{mn} = \tfrac{1}{6}e_{ijk}\frac{\partial \mu_r}{\partial x_i}\frac{\partial \mu_s}{\partial x_j}\frac{\partial \mu_t}{\partial x_k}e_{lmn}\frac{\partial y_r}{\partial x_l}\frac{\partial y_s}{\partial x_m}\frac{\partial y_t}{\partial x_n},$$

whence, with (6.16g) and (6.16h) we find

$$\det D_{mn} = \tfrac{1}{6}e_{rst}Le_{rst}J = LJ. \qquad (A1)$$

The relations in (6.23) are derived with the aid of some of (6.16), (6.17)–(6.19), and the well-known tensor identity $e_{lms}e_{lmr} = 2\delta_{rs}$ as follows:

$$P_{rn} = \tfrac{1}{2}e_{ijk}e_{lmr}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}D_{np}\frac{\partial x_p}{\partial y_k},$$

$$= \tfrac{1}{2}e_{ijk}e_{lmr}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\frac{\partial y_k}{\partial x_s}C^{-1}_{sp}D_{np},$$

$$= \tfrac{1}{2}e_{lms}Je_{lmr}C^{-1}_{sp}D_{np} = JC^{-1}_{rp}D_{np}; \quad \text{(A2)}$$

$$Q_{rn} = \tfrac{1}{2}e_{ijk}e_{lmr}\frac{\partial y_i}{\partial x_l}D_{mp}\frac{\partial x_p}{\partial y_j}D_{ns}\frac{\partial x_s}{\partial y_k},$$

$$= \tfrac{1}{2}e_{ijk}e_{lmr}\frac{\partial x_t}{\partial y_i}\frac{\partial x_p}{\partial y_j}\frac{\partial x_s}{\partial y_k}C_{lt}D_{mp}D_{ns},$$

$$= \frac{1}{2J}e_{tps}e_{lmr}C_{lt}D_{mp}D_{ns}; \quad \text{(A3)}$$

$$W_{lm} = e_{ijk}\frac{\partial y_i}{\partial x_l}D_{mn}\frac{\partial x_n}{\partial y_j}N_r\frac{\partial x_r}{\partial y_k},$$

$$= e_{ijk}\frac{\partial x_s}{\partial y_i}\frac{\partial x_n}{\partial y_j}\frac{\partial x_r}{\partial y_k}C_{ls}D_{mn}N_r,$$

$$= \frac{1}{J}e_{snr}C_{ls}D_{mn}N_r; \quad \text{(A4)}$$

$$K_r = \tfrac{1}{2}e_{ijk}e_{lmr}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}N_p\frac{\partial x_p}{\partial y_k},$$

$$= \tfrac{1}{2}e_{ijk}e_{lmr}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\frac{\partial y_k}{\partial x_s}C^{-1}_{sp}N_p,$$

$$= \tfrac{1}{2}e_{lms}Je_{lmr}C^{-1}_{sp}N_p = JC^{-1}_{rp}N_p; \quad \text{(A5)}$$

$$R_k = \tfrac{1}{2}e_{ijp}e_{lmk}\frac{\partial x_r}{\partial y_i}\frac{\partial x_s}{\partial y_j}\frac{\partial x_t}{\partial y_p}D_{lr}D_{ms}N_t,$$

$$= \frac{1}{2J}e_{rst}e_{lmk}D_{lr}D_{ms}N_t. \quad \text{(A6)}$$

# Quantization of Electrodynamics in the Axial Gauge

YORK-PENG YAO*
*Lyman Laboratory of Physics, Harvard University, Cambridge, Massachussetts*
(Received 20 February 1964)

The so-called axial gauge condition $A_3(x) = 0$ is shown to be inconsistent with the condition of Lorentz invariance. This inconsistency is resolved herein.

## INTRODUCTION

RECENTLY, it was pointed out by Schwinger[1] that the description of non-Abelian gauge fields in the axial gauge cannot be complete. In fact, such a gauge condition, namely $A_3(x) = 0$, cannot lead to a complete and consistent theory in electrodynamics either. We here explore the necessary modifications, which must be made, so as to have an acceptable formalism.

The inconsistency is resolved once we are able to unite the assumption that all local field quantities must vanish at spatial infinities with the requirement that the electromagnetic field has two and only two independent degrees of freedom. We shall also show that these conditions are enough to eliminate the two-dimensional gauge arbitrariness inherent in the condition $A_3(x) = 0$.[1] These considerations can be extended to non-Abelian gauge fields.

## GAUGE CONDITIONS

We shall use the Hermitian Dirac field $\Psi$. The Lagrangian density is

$$L = -\tfrac{1}{2}F^{\mu\nu}(\partial_\mu A_\nu - \partial_\nu A_\mu) + \tfrac{1}{4}F^{\mu\nu}F_{\mu\nu} + \tfrac{1}{2}\Psi\beta\gamma^\mu\cdot\partial_\mu\Psi + (\tfrac{1}{2}i)m\Psi\beta\Psi + j^\mu A_\mu,$$

where the current is

$$j^\mu = -(\tfrac{1}{2}i)e\Psi\beta\gamma^\mu q\Psi,$$

and

$$q = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad q^2 = 1$$

is the antisymmetric imaginary charge matrix. $\beta = i\gamma^0$ (and $\beta\gamma^\mu$) are real (imaginary) and antisymmetric (symmetric), and

$$\{\gamma^\mu, \gamma^\nu\} = -2g^{\mu\nu} = -2\begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}.$$

In this paper, all Latin indices run from 1 to 2; all Greek, 0 to 3. Repeated indices are to be summed over properly, unless specified otherwise.

We tentatively assume that $A_3(x) = 0$ is a complete and consistent gauge condition. The Euler's equations, which result from the above Lagrangian density, are divided into two sets: equations of motion and equations of constraint. The former set includes all those which contain explicit time derivatives, i.e.,

$$\left[\gamma^\mu\left(\frac{1}{i}\partial_\mu - eqA_\mu\right) + m\right]\Psi = 0,$$

$$\partial_0 A_i = \partial_i A_0 + F_{0i},$$

$$\partial_0 F^{i0} = j^i - \partial_3 F^{i3} - \partial_i F^{ij}, \tag{1}$$

except one,

$$\partial_0 F^{30} = j^3 - \partial_i F^{3i}, \tag{2}$$

which can be shown to follow from (1) and the continuity equation

$$\partial_\mu j^\mu = 0.$$

Equation (2) is therefore a member of the latter set, which, in addition, consists of

$$F_{03} = -\partial_3 A_0, \tag{3}$$

$$F_{ij} = \partial_i A_j - \partial_j A_i, \tag{4}$$

$$F_{3j} = \partial_3 A_j, \tag{5}$$

and

$$\partial_i F^{0i} + \partial_3 F^{03} = j^0. \tag{6}$$

The significance of this separation is that all field quantities in (1)—$A^i$, $F^{0i}$, and $\psi$—are independent dynamical variables. The rest of the field quantities are dependent; they can be expressed in terms of $A^i$, $F^{0i}$, and $\psi$.

The integrated form of (6) is

$$F^{03}(x) = \tfrac{1}{2}\int d\mathbf{x}'\epsilon(x_3 - x_3')\delta^2(x - x') \times [j^0(x') - \partial_i' F^{0i}(x')], \tag{7}$$

where

$$\delta^2(x - x') = \delta(x_1 - x_1')\delta(x_2 - x_2')$$

* John Parker Fellow. Present address: Institute for Advanced Study, Princeton, New Jersey.

[1] J. Schwinger, Phys. Rev. **130**, 402 (1963); R. L. Arnowitt and S. I. Fickler, Phys. Rev. **127**, 1821 (1962).

and

$$\epsilon(x) = \begin{cases} 1 & x > 0, \\ -1 & x < 0, \end{cases}$$

while its Fourier transform is

$$ik_3 F^{03}(\mathbf{k}) = j^0(\mathbf{k}) - ik_i F^{0i}(\mathbf{k}), \tag{8}$$

wherein we have also defined

$$O(x) = \sum_{\mathbf{k}} \frac{1}{V^{\frac{1}{2}}} e^{i\mathbf{k}\cdot\mathbf{x}} O(\mathbf{k}, t) = \sum_{\mathbf{k}} \frac{1}{V^{\frac{1}{2}}} e^{i\mathbf{k}\cdot\mathbf{x}} O(\mathbf{k})$$

for any field quantity $O(x)$, $V = L^3$ being the volume within which we enclose our system for quantization.

When $k_3 = 0$, if $F^{03}(\mathbf{k})$ is finite, (8) becomes

$$k_3 = 0: \qquad j^0(\mathbf{k}) - ik_i F^{0i}(\mathbf{k}) = 0. \tag{9}$$

This is equivalent to the assumption that in (7)

$$F^{03}(x) = \pm\tfrac{1}{2} \int d\mathbf{x}' \delta^2(x - x') \times [j^0(x') - \partial_i' F^{0i}(x')] = 0 \qquad x_3 = \pm\infty. \tag{10}$$

In turn, if we accept (10), we obtain (9). These two equivalent conditions assert that the longitudinal components of $F^{0i}$ ($k_3 = 0$) [denoted by $F^{0iL}(k_3 = 0)$] are not independent degrees of freedom. We here see a contradiction, because (1) expresses that all Fourier components of $F^{0i}(x)$ are independent.

This alarm is, however, not completely unpleasant, for (2) leads to

$$A_0(x) = \tfrac{1}{2} \lim_{L\to\infty} \int_{-\frac{1}{2}L}^{\frac{1}{2}L} d\mathbf{x}'(|x_3 - x_3'| - L/2)\delta^2(x - x') \times [j^0(x') - \partial_i' F^{0i}(x')].$$

If (9), or its equivalent (10), does not hold, then $A_0(x) \to \infty$. To have a convergent theory in the axial gauge, it is mandatory that (9) and (10) should be a part of the gauge conditions. On the other hand, if this is the situation, we have yet to overcome the contradiction we just mentioned. There is one way out of this dilemma. We shall first of all show that $A_3(\mathbf{k}) = 0$ cannot be maintained when $k_3 = 0$ for the class of potentials $A_\mu(x)$ which have finite Fourier transforms around $k_3 = 0$ and over a certain domain of $k_1$ and $k_2$.[2] Under an infinitesimal Lorentz transformation,

$$x \to \bar{x} = lx = x + \epsilon x,$$

and

$$A_\mu(x) \to \bar{A}_\mu(\bar{x}) = A_\mu(x) + \epsilon_\mu^\nu A_\nu(x) + \partial_\mu \Lambda(x),$$

where $\Lambda(x)$ is an infinitesimal gauge transformation, which must accompany the Lorentz transformation $l$ in order to guarantee that the same gauge condition holds in the other Lorentz frame. To have an invariant gauge condition, we must have

$$\bar{A}_3(\bar{x}) = 0 = A_3(x) + \epsilon_3^\nu A_\nu(x) + \partial_3 \Lambda(x)$$

or, in Fourier components,

$$\epsilon_{3i} A^i(\mathbf{k}) + \epsilon_{30} A^0(\mathbf{k}) + ik_3 \Lambda(\mathbf{k}) = 0.$$

For the class of potentials we are considering, when $k_3 = 0$, $\Lambda(\mathbf{k})$ must be finite in order that

$$\bar{A}_i(\bar{k}) = A_i(\mathbf{k}) + \epsilon_i^\nu A_\nu(\mathbf{k}) + ik_i \Lambda(\mathbf{k})$$

be consistent. Therefore, we have

$$k_3 = 0; \qquad \epsilon_{3i} A^i(\mathbf{k}) + \epsilon_{30} A^0(\mathbf{k}) = 0,$$

$\epsilon_{3i}$ and $\epsilon_{30}$ are arbitrary; the above condition can be matched only if

$$k_3 = 0; \qquad A^i(\mathbf{k}) = A^0(\mathbf{k}) = 0. \tag{11}$$

But this is too stringent; therefore, necessarily,

$$k_3 = 0; \qquad A_3(\mathbf{k}) \neq 0$$

although

$$k_3 \neq 0; \qquad A_3(\mathbf{k}) = 0. \tag{12}$$

A new gauge condition is now needed, which must be compatible with (9), when $k_3 = 0$. In fact, (9) already suggests what this new condition may be, namely the radiation gauge condition

$$k_3 = 0; \qquad k_i A^i(\mathbf{k}) = 0. \tag{13}$$

There are still other advantages in choosing (13):

(1) We can have a Lorentz-invariant theory; e.g., it can be shown that[3]

$$-i[T^0(x), T^0(x')]_{x_0 = x_0'} = -[T^k(x) + T^k(x')]\partial_k \delta(\mathbf{x} - \mathbf{x}') \qquad (k = 1, 2, 3)$$

in the modified axial gauge which we propose here, but

$$-i[T^0(x), T^0(x')]_{x_0 = x_0'} = -[T^k(x) + T^k(x')]\partial_k \delta(\mathbf{x} - \mathbf{x}') - [(i/2L)F^{03}(x)\Psi(x')\beta\gamma^3 eq\Psi(x')\delta^2(x - x') + x \leftrightarrow x'] + \text{additional terms.}$$

[2] The requirement here is more easily matched when we consider quantization in a finite volume.

[3] To be presented as a part of the Ph.D. thesis (Physics Department, Harvard University).

in the unmodified axial gauge. $T^0$ and $T^k$ are, respectively, the energy and momentum densities of the electromagnetic together with the matter fields. The extra terms in the latter case make it impossible to have a Lorentz-invariant theory in a finite volume.[4]

(2) The gauge arbitrariness mentioned by Schwinger[1] is eliminated. That is, if $A_3(x) = 0$ were the only gauge condition, we would have the freedom of making any arbitrary gauge transformation $\Lambda(x_1, x_2, x_0)$, independent of $x_3$, and the gauge condition would still be maintained, i.e.,

$$A_3(x) \rightarrow A_3(x) + \partial_3\Lambda(x_1, x_2, x_0) = 0.$$

However, if we accept (13), under such a gauge transformation,

$$k_3 = 0; \qquad k^iA_i(\mathbf{k}) \rightarrow k^iA_i(\mathbf{k}) + ik_ik^i\Lambda(\mathbf{k}) = 0$$

or

$$\partial_i\partial^i\Lambda(x_1, x_2, x_0) = 0. \tag{14}$$

The only gauge arbitrariness now is $\Lambda(x_1, x_2, x_0)$, which is a solution of (14). Just as in the conventional radiation gauge, we choose as the standard gauge $\Lambda(\mathbf{k})$, which makes

$$k_3 = 0; \qquad A_i^L(\mathbf{k}) = 0.$$

In this way, the two-dimensional gauge freedom is completely eliminated.

(3) It will be shown in another context[2] that the spectral density of the one-photon Green's function, $\langle \text{vac} \mid A_\mu(\mathbf{k})\, A_\mu(-\mathbf{k}) \mid \text{vac} \rangle$ (no summation), is nonnegative for $k_3 \neq 0$. (13) is naturally the logical choice, among all the gauges we know of, to continue this property to $k_3 = 0$. To phrase it differently, the vectors in the Hilbert space, supplemented with conditions (12) and (13), have finite positive-definite norm and thus allow for probability interpretation.

In closing, it may be remarked that here is an example, which stresses once again that the electromagnetic field has two and only two independent dynamical degrees of freedom. Any violation of this aspect will lead to difficulty in formulating a consistent theory.

The results will be used in a separate note[2,5] to give a proof of the connection between spin and statistics with an electromagnetic field.

ACKNOWLEDGMENT

I would like to thank Professor J. Schwinger for his suggestion of this investigation and his encouragements.

[4] J. Schwinger, Phys. Rev. 127, 324 (1962).

[5] Y. P. Yao, J. Math Phys. 5, 1322 (1964) (following paper)

# Spin and Statistics with an Electromagnetic Field

YORK-PENG YAO*
*Lyman Laboratory, Harvard University, Cambridge, Massachusetts*
(Received 28 February 1964; final manuscript received 9 June 1964)

Because of the impossibility of simultaneously satisfying the requirements of manifest Lorentz covariance and positive-definite (finite) norm in the Hilbert space, no simple proof of the connection between spin and statistics with an electromagnetic field has been given. This note is to point out that it is indeed not necessary to have manifest Lorentz covariance in the full 3 + 1 space to show such a connection. Using the axial gauge $A_3 = 0$, we have succeeded in constructing a simple straightforward proof.

## INTRODUCTION

IN the proofs of the connection between spin and statistics, it is generally assumed that[1,2]

(1) There is a unique invariant vacuum, the lowest energy state.

(2) The spectral density is semipositive-definite, so that probability interpretation exists in the theory.

(3) The theory is manifestly Lorentz covariant.

These assumptions are not entirely met by the electromagnetic field; we have the annoying situation that, in any manifestly Lorentz-covariant gauges, (2) is violated; while, in the radiation guage, (2) is restored at the expense of (3). The relationship between spin and statistics is known to hold, however. It is therefore urgent to re-examine whether all these three assumptions are necessary to construct a proof.

In the axial gauge, $A_3 = 0$,[3,4] Assumption (3) is given up. We shall use this fact to show that (3) is actually not needed to show the connection between spin and statistics.

The proof given here may be considered as an extension of the work of Brown and Schwinger.

There has been a paper by D. Boulware,[5] which deals with the same problem we are considering here. However, he had to work in both the radiation guage and the Lorentz gauge in order to make use of all the three requirements stated above. We are of the opinion that our proof is more straightforward and can be easily generalized.

## SPIN AND STATISTICS

Let $\chi$ be a Hermitian field of finite multiplicity. In the proof of the connection between spin and statistics by Brown and Schwinger, it is essential that the spectral density $m(p, \kappa^2)$ (see later for definition) depends on $p$ only algebraically. For this reason, their proof cannot be carried through to include electrodynamics in the radiation gauge,

$$\nabla \cdot \mathbf{A} = 0.$$

Similarly, in the axial gauge, $A_3 = 0$, we fail to establish that $m(p, \kappa^2)$, depends on $p_3$ only algebraically, as the second-order mass operator indicates.[4] It is then necessary to modify the proof given in I somewhat.

We use the notations

$$\bar{p}^\mu = (p^0, p^1, p^2, 0), \qquad \bar{x}^\mu = (x^0, x^1, x^2, 0),$$

and

$$\bar{p} \cdot \bar{x} = p^1 x^1 + p^2 x^2 - p^0 x^0,$$

i.e., the Lorentz metric is $(-1, 1, 1, 1)$.

Using the conditions that the vacuum is invariant under translations in space and time and that only positive frequency timelike energy momentum vectors $|p>$ can exist in the physical quantum vector space, we write

$$\langle \chi(x)\chi(x') \rangle = \langle \chi e^{iP(x-x')} \chi \rangle$$
$$= \int_0^\infty d\kappa^2 \frac{1}{2L} \left( \sum_{p_3 \neq 0} + \sum_{p_3 = 0} e^{ip_3(x_3 - x_3')} \right)$$
$$\times \int \frac{d\bar{p}}{(2\pi)^2} e^{i\bar{p}\cdot(\bar{x}-\bar{x}')} \eta_+(p)\ \delta(p^2 + \kappa^2) m(p, \kappa^2), \quad (1)$$

where $P$ are the translation (energy–momentum)

* John Parker Fellow. Present address: Institute for Advanced Study, Princeton, New Jersey.

[1] N. Burgoyne, Nuovo Cimento 8, 607, (1958); G. Luders and B. Zumino, Phys. Rev. 110, 1450 (1958).

[2] L. Brown and J. Schwinger, Progr. Theoret. Phys., (Kyoto) 26, 917 (1961). This paper will be referred to as I. It is suggested that the readers go through this paper first, since we shall follow the notations and the basic concepts presented there.

[3] Y. P. Yao, J. Math. Phys. 5, 1319 (1964) (previous paper).

[4] Y. P. Yao, Ph. D. thesis (1964), Department of Physics, Harvard University. Consistency, Lorentz invariance, etc. of quantization in a finite volume have been verified here.

[5] D. Boulware, J. Math. Phys. 3, 50 (1962).

operators, and

$$\eta_+(p) = \eta_-(-p) = \begin{cases} 1 & p^0 > 0, \\ 0 & p^0 < 0. \end{cases}$$

The Hermitian spectral density

$$m(p, \kappa^2) = m^{*T}(p, \kappa^2) = (2\pi)^2(2L)\langle \chi M(p)\chi\rangle \qquad (2)$$

must be a positive matrix over some region of the momentum space; at least over $p^0 \geq 0$, in order to have probability interpretation in the theory. $2L$ is the extension along the third direction in which we enclose the system and $M(p)$ is a nonnegative Hermitian projection operator. (With proper normalization, it is $|p><p|$.)

The complex conjugate of (1) is

$$\langle \chi(x')\chi(x)\rangle = \langle \chi(x)\chi(x')\rangle^*$$

$$= \int_0^\infty d\kappa^2 \frac{1}{2L}\left(\sum_{p_3\neq 0} + \sum_{p_3=0} e^{ip_3(x_3-x_3')}\right)$$

$$\times \int \frac{d\bar{p}}{(2\pi)^2} e^{i\bar{p}\cdot(\bar{x}-\bar{x}')}\eta_-(p)\ \delta(p^2+\kappa^2)m^*(-p,\kappa^2). \qquad (3)$$

Using

$$\eta_\pm(p) = \tfrac{1}{2}(1 \pm \epsilon(p)),$$

we combine (1) and (3) to yield the commutator $[\ ]_-$ or the anticommutator $[\ ]_+$

$$\langle [\chi(x), \chi(x')]_\pm\rangle = \left(\sum_{p_3=0} + \sum_{p_3\neq 0}\frac{1}{2L} e^{ip(x_3-x_3')}\right)$$

$$\times \int \frac{d\bar{p}}{(2\pi)^2}\, d\kappa^2\, e^{i\bar{p}\cdot(\bar{x}-\bar{x}')}\ \delta(p^2+\kappa^2)\tfrac{1}{2}[m(p,\kappa^2)$$

$$\pm m^*(-p,\kappa^2) + \epsilon(p)(m(p,\kappa^2) \pm m^*(-p,\kappa^2))]. \qquad (4)$$

We shall examine the structure of $m(p, \kappa^2)$.

To each homogeneous, proper, orthochronous Lorentz transformation

$$x' = lx, \qquad \det l = 1, \qquad l_0^0 \geq 1,$$

there correspond a unitary operator $U(l)$ and a finite real matrix $L(l)$, such that

$$\bar{\chi}(x') = L(l)\chi(x)$$

and

$$\bar{\chi}(x) = U^{-1}(l)\chi(x)U(l). \qquad (5)$$

For an infinitesimal transformation

$$l_\mu^\nu = \delta_\mu^\nu + \delta\omega_\mu^\nu, \qquad \delta\omega_{\mu\nu} = -\delta\omega_{\nu\mu},$$

we have

$$U(\delta\omega) = 1 + \tfrac{1}{2}i\delta\omega^{\mu\nu}J_{\mu\nu}$$

and

$$L(\delta\omega) = 1 + \tfrac{1}{2}i\delta\omega^{\mu\nu}S_{\mu\nu} + \delta\omega^{\mu\nu}O_{\mu\nu}(\Lambda),$$

$$O_{\mu\nu}(\Lambda) = 0, \qquad \mu, \nu = 0, 1, 2,$$

where $O_{\mu\nu}(\Lambda)$ is a function of the gauge transformation $\Lambda$, which must accompany the Lorentz transformation $l$ to maintain the gauge condition in the new coordinate system. For example, in electrodynamics, expanding

$$A(x) = \int \frac{dp}{(2\pi)^4} e^{ip\cdot x}A(p)$$

and

$$\Psi(x) = \int \frac{dp}{(2\pi)^4} e^{ip\cdot x}\Psi(p),$$

we have[4]

$$(LA(p))_\mu = l_\mu^\nu A_\nu(p) + ip_\mu\lambda(p)$$

and

$$L\Psi(p) = (1 + \tfrac{1}{2}i\delta\omega_{\mu\nu}\sigma^{\mu\nu})\Psi(p)$$

$$+ ieq\int \frac{dk}{(2\pi)^4}\lambda(p-k)\Psi(k), \qquad (6)$$

where

$$\lambda(p) = \frac{-1}{ip_3}\delta\omega^{3\kappa}A_\kappa(p)$$

and

$$\sigma^{\mu\nu} = i\gamma^\mu\gamma^\nu, \qquad \mu \neq \nu.$$

Equation (6) can be integrated to a finite rotation. Let us consider a specific case when we make a rotation in the 1–3 plane through an angle $\alpha$. It can be shown that[6]

$$\bar{A}^0(x') = A^0(x) + \partial^0\lambda,$$

$$\bar{A}^1(x') = (A^1(x) + \partial^1\lambda)\cos\alpha + (A^3(x) + \partial^3\lambda)\sin\alpha,$$

$$\bar{A}^2(x') = A^2(x) + \partial^2\lambda,$$

$$\bar{A}^3(x') = -(A^1(x) + \partial^1\lambda)\sin\alpha + (A^3(x) + \partial^3\lambda)\cos\alpha,$$

and

$$\bar{\Psi}(x') = e^{\frac{1}{2}i\alpha\sigma^{13}}e^{ieq\lambda}\Psi(x), \qquad (7)$$

where the accompanying gauge transformation is

$$\lambda = -A^1(x)\sin\alpha/\partial_3' \qquad (8)$$

and

$$x^{1\prime} = x^1\cos\alpha + x^3\sin\alpha,$$

$$x^{3\prime} = -x^1\sin\alpha + x^3\cos\alpha.$$

(We have tacitly avoided mentioning the complica-

[6] D. Boulware, Ph.D. thesis (1962), Department of Physics, Harvard University.

tion caused by $p_3 = 0$. This does not, however, cause any limitation on the following proof.[7])

The differential characterization of (5) is (See I, p. 920)

$$\frac{1}{i}\left(p_\mu \frac{\partial}{\partial p^\nu} - p_\nu \frac{\partial}{\partial p^\mu}\right) m(p, \kappa^2) + S_{\mu\nu} m(p, \kappa^2) + m(p, \kappa^2) S^T_{\mu\nu} = 0, \qquad \mu, \nu = 0, 1, 2, \tag{9}$$

since no gauge transformation is needed if we do not disturb the third axis. After we introduce $p_4 = ip_0$ and $S_{k4} = iS_{0k}$, as far as the dependence on $\bar{p}$ is concerned, (9) is a statement of the rotational invariance of a system in the 0–1–2 Euclidean subspace. $m(p, \kappa^2)$ is therefore a sum of products of some arbitrary functions of $\kappa^2$ and $p_3$, finite numerical matrices, and three-dimensional spherical harmonics. These spherical functions are homogeneous algebraic functions of $\bar{p}$, of degrees no greater than $2S$, $S$ being the highest spin value associated with $\chi$, and irreducible with respect to the contraction $-p^2 = \kappa^2$.

When $\alpha = \pi$, (8) gives $\lambda = 0$. This result is very important, because it means that no gauge transformation is needed if we rotate the coordinate axes in the 1–3 plane by $\pi$,[8] (although this result is proved for spin-$\frac{1}{2}$ electrodynamics, we believe it to be true in all other cases.) If, after this operation, we further rotate the coordinate system in the 3–4 plane by $\pi$, the combined effect is the total reflection $x^\mu \to -x^\mu$ and the corresponding matrix $L(l)$ is[9]

$$R_{st} = e^{\pi i S_{13}} e^{\pi i S_{24}}.$$

Using the properties

$$R^2_{st} = 1,$$

$iS_{13}$ being real, and $iS_{24} = -S_{02}$ being imaginary, we have

$$R^*_{st} = (-1)^{2S} R_{st}.$$

This last relation and (5) lead to

$$m(-p, \kappa^2) = R_{st} m(p, \kappa^2) R^T_{st} = (-1)^{2S} R_{st} m(p, \kappa^2) R^{*T}_{st}.$$

On the other hand, $m(p, \kappa^2) \geq 0$ for $p^0 \geq 0$, and therefore we have analytically continued $m(p, \kappa^2)$ to negative value of $p^0$, supplying the property

$$p^0 > 0 : (-1)^{2S} m(p, \kappa^2) \geq 0. \tag{10}$$

[7] As we shall see later, if the system is enclosed in a finite volume, we shall avoid using the spectral density around $p_3 = 0$. When the volume becomes infinite, then points with $p_3 = 0$ have measure zero and thus should not complicate matters.

[8] We believe that this result is also true, if we have a guage condition which distinguishes two of the spatial axes from the other one (but does not involve the time component.)

[9] J. Schwinger, Lecture Notes, Summer Institute in Theoretical Physics, Brandeis University (1959).

We return to (4). We have just shown that $m(p, \kappa^2)$ depends on $\bar{p}$ algebraically. For this reason, it is legitimate to write (4) as [in other words, we replace $\bar{p}$ by $1/i\ \bar{\partial}$ to emphasize the algebraic structure]

$$\begin{aligned}
&\langle [\chi(x), \chi(x')]_\pm \rangle \\
&= \int_0^\infty d\kappa^2 \frac{1}{2L} \sum_{p_3 \neq 0} e^{ip_3(x_3 - x_3')} \frac{1}{2}\left[ m\left(\frac{1}{i}\bar{\partial}, p_3, \kappa^2\right) \pm m^*\left(\frac{1}{i}\bar{\partial}, -p_3, \kappa^2\right)\right] \int \frac{d\bar{p}}{(2\pi)^2} e^{i\bar{p}\cdot(\bar{x} - \bar{x}')}\, \delta(p^2 + \kappa^2) \\
&+ \int_0^\infty d\kappa^2 \frac{1}{2L} \sum_{p_3 \neq 0} e^{ip_3(x_3 - x_3')} \frac{1}{2}\left[ m\left(\frac{1}{i}\bar{\partial}, p_3, \kappa^2\right) \mp m^*\left(\frac{1}{i}\bar{\partial}, -p_3, \kappa^2\right)\right] \int \frac{d\bar{p}}{(2\pi)^2} e^{i\bar{p}\cdot(\bar{x} - \bar{x}')}\, \delta(p^2 + \kappa^2)\epsilon(p) \\
&+ \frac{1}{2L} \int \frac{d\bar{p}}{(2\pi)^2}\, d\kappa^2\, e^{i\bar{p}\cdot(\bar{x} - \bar{x}')}\, \delta(p^2 + \kappa^2) \tfrac{1}{2}[m(p, \kappa^2) \\
&\pm m^*(-p, \kappa^2) + \epsilon(p)(m(p, \kappa^2) \mp m^*(-p, \kappa^2))]_{p_3} = 0.
\end{aligned} \tag{11}$$

The function, designated as

$$\Delta(\bar{x} - \bar{x}', \kappa^2 + p_3^2) = \int \frac{d\bar{p}}{(2\pi)^2} e^{i\bar{p}\cdot(\bar{x} - \bar{x}')}\, \delta(p^2 + \kappa^2)\epsilon(p)$$

is odd with respect to $x_0 - x_0'$, because of $\epsilon(p)$, i.e.,

$$\Delta(-(x_0 - x_0'), \mathbf{x} - \mathbf{x}', \kappa^2) = -\Delta(\bar{x} - \bar{x}', \kappa^2).$$

This results in

$$\Delta(\bar{x} - \bar{x}', \kappa^2) = 0 \quad \text{when} \quad x^0 = x^{0\prime}.$$

But $\Delta(\bar{x} - \bar{x}', \kappa^2)$ is an invariant function in the 0–1–2 subspace; the invariant statement is then

$$\Delta(\bar{x} - \bar{x}', \kappa^2) = 0, \qquad (\bar{x} - \bar{x}')^2 > 0.$$

It is at this kind of spatial separation from each other, namely,

$$(x - x')^2 \geq (\bar{x} - \bar{x}')^2 = (x_1 - x_1')^2 + (x_2 - x_2')^2 - (x_0 - x_0')^2 > a^2,$$

that we assign the two $\chi$ fields in what follows. $a$ is some arbitrary but sufficiently large real number.

The vanishing of the commutator or the anticommutator for spatial separations is now assumed. Because $x_3 - x_3'$ can run over all values, the uniqueness of Fourier expansion ensures that each individual component of the Fourier transform in (11) with respect to $p_3$ is null. Thus, when $p_3 \neq 0$,

$$\int_0^\infty d\kappa^2 \left[ m\left(\frac{1}{i}\bar{\partial}, p_3, \kappa^2\right) \pm m^*\left(\frac{1}{i}\bar{\partial}, -p_3, \kappa^2\right)\right] \times \int \frac{d\bar{p}}{(2\pi)^2} e^{i\bar{p}\cdot(\bar{x} - \bar{x}')}\, \delta(p^2 + \kappa^2) = 0, \ (\bar{x} - \bar{x}')^2 > a^2. \tag{12}$$

We have the representation, for $(\bar{x} - \bar{x}')^2 > a^2$

$$\Delta^1(\bar{x} - \bar{x}', \kappa^2 + p_3^2) \equiv \int \frac{d\bar{p}}{(2\pi)^2} e^{i\bar{p}\cdot(\bar{x}-\bar{x}')}\, \delta(p^2 + \kappa^2)$$

$$= \frac{1}{(\pi)^{\frac{1}{2}}} \int_0^\infty \frac{d\lambda}{(\lambda)^{\frac{1}{2}}} \exp\left(-\lambda\bar{x}^2 - \frac{\kappa^2 + p_3^2}{4\lambda}\right),$$

and (12) becomes

$$\int_0^\infty d\kappa^2 \left[ m\left(\frac{1}{i}\bar{\partial}, p_3, \kappa^2\right) \pm m^*\left(\frac{1}{i}\bar{\partial}, -p_3, \kappa^2\right)\right]$$

$$\times \int_0^\infty \frac{d\lambda}{(\lambda)^{\frac{1}{2}}} \exp\left(-\lambda\bar{x}^2 - \frac{\kappa^2 + p_3^2}{4\lambda}\right) = 0. \qquad (13)$$

The proof from here on duplicates that of I. Because of what we said about the structure of $m[(1/i)\bar{\partial}, p_3, \kappa^2]$ after Eq. (9) and because $\Delta^1(\bar{x} - \bar{x}', \kappa^2 + p_3^2)$ is a function of $(\bar{x} - \bar{x}')^2$ only, we expand

$$\left[ m\left(\frac{1}{i}\bar{\partial}, p_3, \kappa^2\right) \pm m^*\left(\frac{1}{i}\bar{\partial}, -p_3, \kappa^2\right)\right]$$

$$\times \Delta^1(\bar{x} - \bar{x}', \kappa^2 + p_3^2) = \sum_{l=1}^{2S} m_{l\pm}(p_3, \kappa^2) Y_l(\bar{x} - \bar{x}')$$

$$\times \left(\frac{\partial}{\partial(\bar{x} - \bar{x}')^2}\right)^l \Delta^1(\bar{x} - \bar{x}', \kappa^2 + p_3^2) \qquad (14)$$

into harmonic series where $Y_l(\bar{x} - \bar{x}')$ are three-dimensional spherical harmonics, and $m_{l\pm}(p_3, \kappa^2)$ are the associated expansion coefficients, being scalar functions in $p_3$ and $\kappa^2$.

Because $(\partial/\partial(\bar{x} - \bar{x}')^2)^l \Delta^1(\bar{x} - \bar{x}', \kappa^2 + p_3^2) \not\equiv 0$, we must have for each $l = 0, 2S$

$$\int_0^\infty d\lambda\, e^{-\lambda z^2} \int_0^\infty d\kappa^2$$

$$\times \exp\left[-(\kappa^2 + p_3^2)/4\lambda\right](\lambda)^{-\frac{1}{2}} m_{l\pm}(p_3, \kappa^2) = 0.$$

We restrict ourselves to the class of functions $m_{l\pm}(p_3, \kappa^2)$ which can grow at most algebraically with $\kappa^2$.[10] We now invoke Laplace's lemma,[2] which states that if

$$\int_0^\infty (dx)\, e^{-tx} f(x) = 0 \qquad (t \geq 1)$$

for $f(x)$ such that

$$\int_0^\infty (dx)\, e^{-x} |f(x)|^2 < \infty,$$

then

$$f(x) = 0.$$

We apply this lemma first with respect to the $\lambda$ transform and then the $\kappa^2$ transform. We come up with the conclusion

$$m_{l\pm}(p_3, \kappa^2) \equiv 0, \qquad l = 0, 2S.$$

Therefore it follows from (4) that the vanishing of the commutator $(-)$ or the anticommutator $(+)$ for sufficient spacelike separation requires

$$m(p, \kappa^2) \mp m^*(-p, \kappa^2) = 0, \quad \text{respectively.}$$

Comparing this with (2) and (4), we see that $m(p, \kappa^2) \equiv 0$, if we assume the wrong connection between spin and statistics. This cannot be so, for $m(p, \kappa^2) \equiv 0$ implies $\chi \equiv 0$. Thus, we must demand the commutator condition (Bose–Einstein statistics) to go with integral spin fields, and the anticommutator conditions (Fermi–Dirac statistics) to go with half-integral spin fields.

We also see that the same proof can be carried out, if we have a theory which is manifestly Lorentz-covariant in a plane containing the 0 axis and one of the spatial axes.


## ACKNOWLEDGMENTS

It is my pleasure to thank Professor Schwinger for continuous guidance and encouragement. I would also like to thank Dr. D. Boulware for discussions and a critical reading of the manuscript.


## APPENDIX I[4]

The one-photon Green's function has the spectral representation

$$\langle A_\mu(x) A_\nu(x')\rangle = \int \frac{d^3\bar{p}}{(2\pi)^3} \frac{1}{2L} \sum_{p_4} \int_0^\infty d\kappa^2\, e^{ip\cdot(x-x')}$$

$$\times \eta_+(p)\, \delta(p^2 + \kappa^2) A_{\mu\nu}(p),$$

where

$$p_3 = 0 : A_{\mu\nu}(p)$$

$$= \left(g_{\mu\nu} - \frac{p_\mu p_\nu - (n\cdot p)(n_\mu p_\nu + n_\nu p_\mu)}{(p + n(n\cdot p))^2}\right) A_1(\kappa^2),$$

$n = (1, 0, 0, 0)$ a timelike unit vector;

$$p_3 \not\equiv 0 : A_{\mu\nu}(p)$$

$$= \left(g_{\mu\nu} + \frac{p_\mu p_\nu}{2}\left(\frac{1}{(m\cdot p - i\epsilon)^2} + \frac{1}{(m\cdot p + i\epsilon)^2}\right)\right.$$

$$\left. - \frac{m_\mu p_\nu + m_\nu p_\mu}{2}\left(\frac{1}{(m\cdot p - i\epsilon)} + \frac{1}{(m\cdot p + i\epsilon)}\right)\right) A_2(\kappa^2)$$

[10] This condition can always be satisfied, if we have some finite equal-time commutation relations of the $\chi$ [i. e., finite coefficients multiplied by $\delta(\mathbf{x} - \mathbf{x}')$], because then we can show that there are finite sum rules (integrals with respect to $\kappa^2$) that $m(p, \kappa^2)$ have to obey. See, for instance, the Appendix.

($\lim \epsilon \to 0$). $m = (0, 0, 0, 1)$ a spacelike unit vector, with

$$A_1(\kappa^2) \geq 0, \qquad A_2(\kappa^2) \geq 0,$$

$$\int_0^\infty A_1(\kappa^2)\, d\kappa^2 = \int_0^\infty A_2(\kappa^2)\, d\kappa^2 = 1;$$

it is apparent that

$$A_{\mu\mu}(p) \geq 0 \quad \begin{matrix}\text{(semipositive-definite)} \\ \text{(no sum).}\end{matrix}$$

The one-electron Green's function has the spectral representation

$$\langle \psi(x)\psi(x') \rangle = -\int \frac{d^3\bar{p}}{(2\pi)^2} \int_0^\infty d\kappa^2 \frac{1}{2L} \left\{ \sum_{p_3} e^{ip\cdot(x-x')} \times \eta_+(p)\, \delta(p^2 + \kappa^2) i\beta[(\gamma^0 p_0 + \gamma^1 p_1 + \gamma^2 p_2) \times A(p_3^2, \kappa^2) + \gamma^3 p_3 B(p_3^2, \kappa^2) + mC(p_3^2, \kappa^2)].$$

We have the properties

$$A(p_3^2, \kappa^2) \geq 0 \quad \text{(semipositive-definite)}$$

and

$$\int_0^\infty A(p_3^2, \kappa^2)\, d\kappa^2 = 1.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# The Wave Equation and the Green's Dyadic for Bounded Magnetoplasmas

Y. J. SETO AND ARWIN A. DOUGAL
*Department of Electrical Engineering, The University of Texas, Austin, Texas*
(Received 4 February 1964; final manuscript received 24 April 1964)

In studies of electromagnetic wave propagation and radiation in magnetoplasmas, the wave equation takes the form of a dyadic–vector Helmholtz equation. The investigation here shows that the dyadic–vector Helmholtz equation is solvable by the separation method in four cylindrical coordinate systems. Solutions in the form of complete sets of eigenfunctions are possible when boundary surfaces are present. For problems involving current sources in the plasma, the Green's dyadics for finite or semifinite domains can be constructed from the complete sets of eigenfunctions which are solutions to the homogeneous equation. The Green's dyadic for infinite domain is also shown to be obtained from that for a semifinite domain through a limiting process.

## INTRODUCTION

THE presence of a static magnetic field in a plasma region results in an effective electric conductivity which is of dyadic form. Assuming monochromatic waves, the equation describing the waves, generated by a source, $\mathbf{J}_s$, in such an anisotropic medium may be written as

$$\nabla \times \nabla \times \mathbf{E} - \mathbf{k}\cdot\mathbf{E} = \mathbf{J}_s. \tag{1}$$

Written in matrix form, the dyadic $\mathbf{k}$ is

$$\mathbf{k} = \begin{bmatrix} k_\perp & k_T & 0 \\ -k_T & k_\perp & 0 \\ 0 & 0 & k_\parallel \end{bmatrix}. \tag{2}$$

Assuming spatial homogeneity, the parameters $k_\perp$, $k_T$, and $k_\parallel$ are constants with respect to time and space coordinates.

Solutions for Eq. (1) in terms of auxiliary Green's function for infinite domain have been discussed by Bunkin[1] and subsequently extended by Chow[2] and Brandstater.[3] However, the solutions of Eq. (1) for a bounded region have proved to be more difficult to obtain. The studies dealt with here reveal that, in order to solve for a finite-domain or semifinite-domain Green's function, a better understanding of the free wave equation, $\mathbf{J}_s = 0$ in Eq. (1), is needed, and that the Green's function may be constructed from the solutions of the homogeneous equation.

## THE HOMOGENEOUS EQUATION

The homogeneous equation describing free wave propagation is

[1] F. V. Bunkin, Zh. Eksperim. i Teor. Fiz. **32**, 338 (1957) [English transl.: Soviet Phys.—JETP **5**, 277 (1957)].
[2] Y. Chow, Trans. IRE trans. Antennas Propagation **10**, 464 (1962).
[3] J. J. Brandstater, *An Introduction to Waves, Rays, and Radiation in Plasma* (McGraw-Hill Book Company, Inc., New York, 1963) Chap. 9.

($\lim \epsilon \to 0$). $m = (0, 0, 0, 1)$ a spacelike unit vector, with

$$A_1(\kappa^2) \geq 0, \qquad A_2(\kappa^2) \geq 0,$$

$$\int_0^\infty A_1(\kappa^2)\, d\kappa^2 = \int_0^\infty A_2(\kappa^2)\, d\kappa^2 = 1;$$

it is apparent that

$$A_{\mu\mu}(p) \geq 0 \quad \begin{matrix}\text{(semipositive-definite)} \\ \text{(no sum).}\end{matrix}$$

The one-electron Green's function has the spectral representation

$$\langle \psi(x)\psi(x') \rangle = -\int \frac{d^3\bar{p}}{(2\pi)^2} \int_0^\infty d\kappa^2 \frac{1}{2L} \left\{ \sum_{p_3} e^{ip\cdot(x-x')} \times \eta_+(p)\, \delta(p^2 + \kappa^2) i\beta[(\gamma^0 p_0 + \gamma^1 p_1 + \gamma^2 p_2) \times A(p_3^2, \kappa^2) + \gamma^3 p_3 B(p_3^2, \kappa^2) + mC(p_3^2, \kappa^2)].$$

We have the properties

$$A(p_3^2, \kappa^2) \geq 0 \quad \text{(semipositive-definite)}$$

and

$$\int_0^\infty A(p_3^2, \kappa^2)\, d\kappa^2 = 1.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# The Wave Equation and the Green's Dyadic for Bounded Magnetoplasmas

Y. J. SETO AND ARWIN A. DOUGAL
*Department of Electrical Engineering, The University of Texas, Austin, Texas*
(Received 4 February 1964; final manuscript received 24 April 1964)

In studies of electromagnetic wave propagation and radiation in magnetoplasmas, the wave equation takes the form of a dyadic–vector Helmholtz equation. The investigation here shows that the dyadic–vector Helmholtz equation is solvable by the separation method in four cylindrical coordinate systems. Solutions in the form of complete sets of eigenfunctions are possible when boundary surfaces are present. For problems involving current sources in the plasma, the Green's dyadics for finite or semifinite domains can be constructed from the complete sets of eigenfunctions which are solutions to the homogeneous equation. The Green's dyadic for infinite domain is also shown to be obtained from that for a semifinite domain through a limiting process.

## INTRODUCTION

THE presence of a static magnetic field in a plasma region results in an effective electric conductivity which is of dyadic form. Assuming monochromatic waves, the equation describing the waves, generated by a source, $\mathbf{J}_s$, in such an anisotropic medium may be written as

$$\nabla \times \nabla \times \mathbf{E} - \mathbf{k}\cdot\mathbf{E} = \mathbf{J}_s. \tag{1}$$

Written in matrix form, the dyadic $\mathbf{k}$ is

$$\mathbf{k} = \begin{bmatrix} k_\perp & k_T & 0 \\ -k_T & k_\perp & 0 \\ 0 & 0 & k_\parallel \end{bmatrix}. \tag{2}$$

Assuming spatial homogeneity, the parameters $k_\perp$, $k_T$, and $k_\parallel$ are constants with respect to time and space coordinates.

Solutions for Eq. (1) in terms of auxiliary Green's function for infinite domain have been discussed by Bunkin[1] and subsequently extended by Chow[2] and Brandstater.[3] However, the solutions of Eq. (1) for a bounded region have proved to be more difficult to obtain. The studies dealt with here reveal that, in order to solve for a finite-domain or semifinite-domain Green's function, a better understanding of the free wave equation, $\mathbf{J}_s = 0$ in Eq. (1), is needed, and that the Green's function may be constructed from the solutions of the homogeneous equation.

## THE HOMOGENEOUS EQUATION

The homogeneous equation describing free wave propagation is

[1] F. V. Bunkin, Zh. Eksperim. i Teor. Fiz. **32**, 338 (1957) [English transl.: Soviet Phys.—JETP **5**, 277 (1957)].
[2] Y. Chow, Trans. IRE trans. Antennas Propagation **10**, 464 (1962).
[3] J. J. Brandstater, *An Introduction to Waves, Rays, and Radiation in Plasma* (McGraw-Hill Book Company, Inc., New York, 1963) Chap. 9.

$$\nabla \times \nabla \times \mathbf{E} - \mathbf{k} \cdot \mathbf{E} = 0. \tag{3}$$

Equation (3) is seen to resemble a vector Helmholtz equation except that $\mathbf{k}$ is a dyadic. It is well known that the scalar Helmholtz equation is separable in eleven coordinate systems, and that the vector Helmholtz equation is separable in only six coordinate systems.[4] Despite the fact that the dyadic–vector Helmholtz equation has been frequently encountered in connection with the studies of crystal materials and plasma fields, and that its solutions have been obtained and used extensively for problems involving boundaries in the rectangular coordinate systems and the circular cylindrical coordinate systems,[5,6,7] additional investigation into the separability of the dyadic–vector Helmholtz equation is desirable. The separability of Eq. (3) will be studied here, since by determining the coordinate systems in which this equation is separable one not only gains the knowledge of exactly in what coordinate systems the equation is solvable by a separation method, but one also hopefully attempts solutions in the form of eigenfunctions when boundaries are involved. The eigenfunction solutions will be of great help in constructing the finite-domain or semifinite-domain Green's dyadics.

In the application of boundary-value problems, separation into the form that conveniences the fitting of boundary surfaces is most desirable. Hence, it is advisable to separate this dyadic–vector Helmholtz equation in terms of longitudinal $\mathbf{L}$, and transverse $\mathbf{M}$ and $\mathbf{N}$ vector components.

The first term in Eq. (3) is a vector operating term, $\nabla \times \nabla \times \mathbf{E}$. A review of the separability of a vector Helmholtz equation shows that the coordinate system in which this vector operating term facilitates separation must be a coordinate system in which one of the scale factors is unity, and that the ratio of the remaining two scale factors must be independent of the coordinate corresponding to the unity scale factor. The six coordinate systems which meet these requirements are the spherical, the conical, and the four cylindrical coordinate systems.

Pertaining to magnetoactive plasma, Eq. (2) implies that the static magnetic field is in the direction parallel or antiparallel to the coordinate corresponding to the unity scale factor. Without losing generality, this coordinate is denoted $\xi_3$, and its unit vector, $\mathbf{a}_3$. A close examination shows that only four out of the six coordinate systems are physically realizable for such alignment of the static magnetic field; namely, the four cylindrical coordinate systems including the rectangular, the circular cylindrical, the elliptical cylindrical, and the parabolic cylindrical coordinate systems. In each system, $\xi_3$ corresponds to the $z$ axis.

It may first seem to be pessimistic that the number of permissible coordinate systems has been reduced to only four from eleven right at the onset. Fortunately, however, it turns out that no other restriction will be imposed that further reduces the number of permissible coordinate systems.

In attempting the solution of Eq. (3), the difficulty lies in the fact that each term in the equation is a purely transverse vector, while due to the dyadic $\mathbf{k}$, the vector field $\mathbf{E}$, in general, is not entirely transverse. Since it is desirable to separate the equation in terms of transverse and longitudinal components, $\mathbf{E}$ must be expressed in terms of all three vector components $\mathbf{L}$, $\mathbf{M}$, and $\mathbf{N}$, i.e.,

$$\begin{aligned} \mathbf{E} &= \mathbf{L} + \mathbf{M} + \mathbf{N}, \\ \mathbf{L} &= -\nabla_\perp \Phi - \nabla_\parallel \Phi, \\ \mathbf{M} &= \nabla_\perp \Psi \times \mathbf{a}_3, \\ \mathbf{N} &= \nabla_\perp (\nabla_\parallel \cdot \mathbf{a}_3 \mathrm{X}) - (\nabla_\perp^2 \mathrm{X}) \mathbf{a}_3; \end{aligned} \tag{4}$$

where $\Phi$, $\Psi$, and X are three scalar functions to be determined. The subscript $\perp$ indicates the components of operator or vector which are perpendicular to $\mathbf{a}_3$, whereas $\parallel$ indicates parallel to $\mathbf{a}_3$.

Expanding $\mathbf{k} \cdot \mathbf{E}$ into vector form, Eq. (3) may be broken into two equations, one contains the $\perp$ vectors, the other contains the $\parallel$ vectors. It is also recognized that Eq. (3) implies

$$\nabla \cdot (\mathbf{k} \cdot \mathbf{E}) = 0, \tag{5}$$

which yields a third equation. After some manipulation, the three basic equations become

$$\nabla_\perp^2 (\nabla_\perp^2 \Psi) + \nabla_\parallel^2 (\nabla_\perp^2 \Psi) + k_T \nabla_\parallel \cdot \mathbf{a}_3 (\nabla_\perp^2 \mathrm{X}) - k_T \nabla_\perp^2 \Phi + k_\perp (\nabla_\perp^2 \Psi) = 0, \tag{6}$$

$$\nabla_\perp^2 (\nabla_\perp^2 \mathrm{X}) + \nabla_\parallel^2 (\nabla_\perp^2 \mathrm{X}) + k_\parallel \nabla_\perp^2 \mathrm{X} + k_\parallel \nabla_\parallel \cdot \mathbf{a}_3 \Phi = 0, \tag{7}$$

$$k_T (\nabla_\perp^2 \Psi) + (k_\parallel - k_\perp) \nabla_\parallel \cdot \mathbf{a}_3 (\nabla_\perp^2 \mathrm{X}) + k_\perp \nabla_\perp^2 \Phi + k_\parallel \nabla_\parallel^2 \Phi = 0. \tag{8}$$

Close examination of Eqs. (6)–(8) shows that

[4] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953) Vol. II, Chap. 13.
[5] W. P. Allis, S. J. Buchsbaum, and A. Bers, *Waves in Anisotropic Plasmas* (Technology Press, Cambridge, Massachusetts, 1963) Part II.
[6] H. Suhl and L. R. Walker, Bell System Tech. J. **33**, 579, 939, 1133 (1954).
[7] A. A. Th.M. van Trier, Appl. Sci. Res. (Netherlands) **B3**, 305 (1953).

solutions may be obtained if the three scalar functions each satisfies

$$\nabla^2(\Phi, \Psi, \mathrm{X}) + T_m^2(\Phi, \Psi, \mathrm{X}) = 0, \tag{9}$$

or

$$\nabla_\perp^2(\Phi, \Psi, \mathrm{X}) + \kappa_m^2(\Phi, \Psi, \mathrm{X}) = 0 \tag{10a}$$

and

$$\nabla_\parallel^2(\Phi, \Psi, \mathrm{X}) + k_m^2(\Phi, \Psi, \mathrm{X}) = 0, \tag{10b}$$

with

$$\kappa_m^2 = T_m^2 - k_m^2, \tag{10c}$$

where $k_m^2$ is the separation constant for separation of $\xi_3$. $T_m^2$ in Eqs. (9) or (10c) must satisfy an eighth-order determinant equation

$$\begin{vmatrix} (T_m^2-k_\perp) & -k_T & k_T \\ 0 & (T_m^2-k_m^2)(T_m^2-k_\parallel) & -k_\parallel k_m^2 \\ k_T(T_m^2-k_m^2) & (k_\parallel-k_\perp)(T_m^2-k_m^2) & k_\perp(T_m^2-k_m^2)+k_\parallel k_m^2 \end{vmatrix} = 0. \tag{11}$$

The order of Eq. (11) appears to be too high to be readily solved at first, but it turns out that the resulting secular equation is only of fourth order in $T_m$, since the other four roots, $T_m^2 = k_m^2$, and $T_m^2 = 0$, are trivial and may be discarded. The secular equation yield by Eq. (11) is

$$(T_m^2 - k_m^2 - k_\parallel)[k_T^2 - k_\perp(T_m^2 - k_\perp)] - k_\parallel k_m^2(T_m^2 - k_\perp) = 0. \tag{12}$$

Equation (12) may be readily solved for $T_m^2$ in terms of $k_m^2$, or for $k_m^2$ in terms of $T_m^2$, depending upon the manner of the boundaries set up in the problem. Let the solutions of the scalar functions be

$$\Psi = A\Psi_\perp(\xi_1, \xi_2, \kappa_m)\Psi_\parallel(\xi_3, k_m), \tag{13}$$

$$\mathrm{X} = B\mathrm{X}_\perp(\xi_1, \xi_2, \kappa_m)\mathrm{X}_\parallel(\xi_3, k_m), \tag{14}$$

$$\Phi = C\Phi_\perp(\xi_1, \xi_2, \kappa_m)\Phi_\parallel(\xi_3, k_m). \tag{15}$$

If the boundaries are parallel to the $\xi_3$ = constant surfaces, $\Psi_\parallel$, $\mathrm{X}_\parallel$, and $\Phi_\parallel$ are sets of eigenfunctions and $k_m$ are the eigenvalues with index $m$; then $\kappa_m$, obtained from Eqs. (12) and (10c), will describe the dispersion relation for propagation in the $(\xi_1, \xi_2)$ space. Conversely, if the boundary surfaces are perpendicular to $\xi_3$ = constant surfaces, $\Psi_\perp$, $\mathrm{X}_\perp$, and $\Phi_\perp$ will consist of sets of eigenfunctions with $\kappa_m$ consisting of the eigenvalues. $\Psi_\parallel$, $\mathrm{X}_\parallel$, and $\Phi_\parallel$ describe the propagation in $\mathbf{a}_3$ direction with $k_m$ being the parameter describing the dispersion relation. In either case there will be another eigenvalue with index $n$, resulting from the separation of Eq. (10b), which is not apparent in Eqs. (13)–(15). Of course, when the boundary is a completely self-enclosed one, there are three sets of eigenvalues with indices $m$, $n$, and $l$. The solutions $\Psi$, X, and $\Phi$ are not entirely independent. By substitution of Eqs. (13)–(15) into Eqs. (6)–(8), it is possible to obtain a functional relation between the constants $A$, $B$, and $C$, thus reducing the number of arbitrary constants to one.

Without restricting the generality of the two succeeding sections on the inhomogeneous equation and the Green's function, and on the infinite-domain Green's dyadic, a readily understandable illustration is that of a plasma region bounded by two parallel, infinitely large, conducting plates of finite separation $d$, with a static magnetic field imposed upon the plasma in the direction normal to the boundary plates. The solution for outgoing waves can be found in a circular cylindrical coordinate system. Assuming the origin of the coordinate system is located midway between the plates

$$\Psi = \sum_{m,n} A_{mn}H_n^{(2)}(\kappa_m r)e^{in\theta}\cos k_m z, \tag{16}$$

$$\mathrm{X} = \sum_{m,n} B_{mn}H_n^{(2)}(\kappa_m r)e^{in\theta}\sin k_m z, \tag{17}$$

$$\Phi = \sum_{m,n} C_{mn}H_n^{(2)}(\kappa_m r)e^{in\theta}\cos k_m z; \tag{18}$$

with two sets of eigenvalues, i.e.,

$$n = 0, \pm 1, \pm 2, \cdots,$$

$$k_m = m\pi/d; \quad m = 0, 1, 2, \cdots. \tag{19}$$

As stated above, functional relations between $A_{mn}$, $B_{mn}$, and $C_{mn}$ may be obtained by substituting Eqs. (16)–(18) into Eqs. (6)–(8). Since $\cos k_m z$ or $\sin k_m z$ when summed on $m$ constitutes a complete set of eigenfunctions, this complete set is a complete solution of Eq. (10b). Also, the functions $e^{in\theta}$ when summed on $n$ yield a complete set of eigenfunctions which satisfies an equation resulting from separation of $\theta$ from Eq. (10a). Therefore, by virtue of the completeness theorem for several variables,[8] the functions $\Psi$, X, and $\Phi$ as shown in Eqs. (16)–(18) are complete sets which satisfy Eq. (9). Consequently, the wave field $\mathbf{E}$, obtained from Eq. (4), having three orthogonal components, namely,

$$\nabla_\perp\Psi \times \mathbf{a}_3, \quad \nabla_\perp(\nabla_\parallel \cdot \mathbf{a}_3\mathrm{X} - \Phi) \text{ and } \mathbf{a}_3(\nabla_\perp^2\mathrm{X} + \nabla_\parallel \cdot \mathbf{a}_3\Phi),$$

with each component consisting of complete sets,

[8] See for example, R. Courant, and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), Vol. I, p. 56.

must by itself be complete. Hence, $\mathbf{E}$ so obtained is a complete solution to Eq. (3). In addition, it can also be shown that for a physically realizable problem, the boundary conditions require that either the tangential component of electric field or the tangential component of the magnetic field vanishes at the boundary; from this, the solution obtained by Eq. (4) can be proved to be unique.[9]

### INHOMOGENEOUS EQUATION AND THE GREEN'S FUNCTION

When a source $\mathbf{J}_s(\mathbf{r})$ is presented in the bounded region, it can be shown that Eq. (1) is solvable in terms of an integral representation

$$\mathbf{E}(\mathbf{r}) = \int_{v_0} \mathbf{G}(\mathbf{r} \mid \mathbf{r}_0)\cdot\mathbf{J}_s(\mathbf{r}_0)\, dv_0, \tag{20}$$

where the kernel $\mathbf{G}(\mathbf{r} \mid \mathbf{r}_0)$ is the usual Green's dyadic function except that instead of satisfying Eq. (1) with a dyadic impulse source, it satisfies the following:

$$\nabla\times\nabla\times\mathbf{G} - \tilde{\mathbf{k}}\cdot\mathbf{G} = \mathbf{I}\delta(\mathbf{r} - \mathbf{r}_0), \tag{21}$$

where $\mathbf{I}$ is the idemfactor and $\tilde{\mathbf{k}}$ is the conjugate of $\mathbf{k}$. The use of the conjugate of $\mathbf{k}$ in Eq. (21) is necessary if it is desired to include the cases where $\mathbf{k}$ is not Hermitian, (see Appendix). In addition to satisfying Eq. (21), the Green's function must also satisfy the same boundary condition that the field satisfies.

The derivation of a Green's function to be discussed here depends upon whether there are boundary surfaces parallel to the $\xi_3$ = const surface. For brevity, only the case with boundary surface parallel to the $\xi_3$ = const surface will be derived here. It is assured that the Green's dyadic for the case of no boundary surface parallel to $\xi_3$ = const surface may also be derived with the same technique, except for some minor modifications.

In view of the form of the solutions to the homogeneous equation, and in view of the fact that the three scalar functions are not independent functions, it is proposed that the Green's dyadic takes the form

$$\mathbf{G} = \mathbf{G}_M + \mathbf{G}_N + \mathbf{G}_L, \tag{22}$$

where

$$\mathbf{G}_M = \sum_{m,n} \{(\nabla_\perp\varphi_{mn}\times\mathbf{a}_3)f_m\}\mathbf{F}_{mn}(\xi_1^0, \xi_2^0, \xi_3^0), \tag{23}$$

$$\mathbf{G}_N = \sum_{m,n} \{(\nabla_\perp\varphi_{mn})(\nabla_\parallel\cdot\mathbf{a}_3 g_m) - (\nabla_\perp^2\varphi_{mn})g_m\mathbf{a}_3\} \times \mathbf{G}_{mn}(\xi_1^0, \xi_2^0, \xi_3^0), \tag{24}$$

$$\mathbf{G}_L = \sum_{m,n} \{\nabla_\perp\varphi_{mn}f_m + \varphi_{mn}(\nabla_\parallel f_m)\}\mathbf{H}_{mn}(\xi_1^0, \xi_2^0, \xi_3^0), \tag{25}$$

and where $\mathbf{F}_{mn}$, $\mathbf{G}_{mn}$, and $\mathbf{H}_{mn}$ are functions of source coordinates only.

$\varphi_{mn}$ is a two-variable function of variables $\xi_1$ and $\xi_2$, satisfying Eq. (10a). The $\xi_3$ dependent functions $f_m$ and $g_m$ are the two independent solutions of Eq. (10b); their relation is dictated by whether there is a closed or open boundary in $\xi_3$. For the case of closed boundary in $\xi_3$, the relation is

$$\nabla_\parallel\cdot\mathbf{a}_3 g_m = \mp k_m f_m, \qquad \nabla_\parallel\cdot\mathbf{a}_3 f_m = \pm k_m g_m. \tag{26}$$

The choice of upper or lower sign in Eq. (26) depends on the type of boundary condition and the choice of coordinate origin in the problem. To be general, both signs will be kept throughout this derivation. Finally, the index $n$ in Eqs. (23)–(25) may be a single index or a double index, depending upon whether the boundary perpendicular to $\xi_3$ is open or closed.

The vectors $\mathbf{M}$, $\mathbf{N}$, and $\mathbf{L}$ are not necessarily orthogonal in space, but $\nabla_\perp\varphi_{mn}\times\mathbf{a}_3$, $\nabla_\perp\varphi_{mn}$, and $\mathbf{a}_3$ are three orthogonal vectors. If a unit vector $\mathbf{b}$, and a two-variable-dependent function $\rho_{mn}(\xi_1, \xi_2)$ are defined such that

$$\nabla_\perp\varphi_{mn}(\xi_1, \xi_2) = \rho_{mn}(\xi_1, \xi_2)\mathbf{b}, \tag{27}$$

then the unit vectors $\mathbf{b}$, $\mathbf{b}\times\mathbf{a}_3$, and $\mathbf{a}_3$ are mutually orthogonal in space. Multiplication of these unit vectors $\mathbf{b}$, $\mathbf{b}\times\mathbf{a}_3$, and $\mathbf{a}_3$ in turn into Eq. (20) yields a set of three mutually orthogonal equations:

$$\sum_{m,n} \{-(\nabla^2 + k_\perp)(\rho_{mn}f_m)\mathbf{F}_{mn} - k_T(\rho_{mn}f_m)(k_m\mathbf{G}_{mn}) + k_T(\rho_{mn}f_m)\mathbf{H}_{mn}\} = \mathbf{b}\times\mathbf{a}_3\delta(\mathbf{r} - \mathbf{r}_0), \tag{28}$$

$$\sum_{m,n} \{k_T(\rho_{mn}f_m)\mathbf{F}_{mn} - (\nabla^2 + k_\perp)(\rho_{mn}f_m)(k_m\mathbf{G}_{mn}) + k_\perp(\rho_{mn}f_m)\mathbf{H}_{mn}\} = \mathbf{b}\delta(\mathbf{r} - \mathbf{r}_0), \tag{29}$$

$$\sum_{m,n} \{(\nabla^2 + k_\parallel)(\nabla^2 + k_m^2)(\varphi_{mn}g_m)(k_m\mathbf{G}_{mn}) - k_\parallel k_m^2(\varphi_{mn}g_m)\mathbf{H}_{mn}\} = k_m\mathbf{a}_3\delta(\mathbf{r} - \mathbf{r}_0). \tag{30}$$

In Eqs. (28)–(30), the relation given by Eqs. (22)–(26) has been substituted. The operator $\nabla^2$ is a three-dimensional operator operating on the observer coordinate functions only, i.e., $(\rho_{mn}f_n)$ or $(\varphi_{mn}g_m)$. When operation on the source coordinate is needed, the operators will be distinguished by a superscript o.

In order to express the $\xi_3^o$ dependent functions explicitly, and to express the source coordinate functions in component form, $\mathbf{F}_{mn}$, $\mathbf{G}_{mn}$, and $\mathbf{H}_{mn}$ may be assumed

[9] See for example, J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941) pp. 486–488.

$$\mathbf{F}_{mn} = F^{\times}_{mn}(\xi^{\circ}_1, \xi^{\circ}_2)\mathfrak{f}^{\times}_m(\xi^{\circ}_3)\mathbf{b}\times\mathbf{a}_3 + F^{\perp}_{mn}(\xi^{\circ}_1, \xi^{\circ}_2)\mathfrak{f}^{\perp}_m(\xi^{\circ}_3)\mathbf{b} + F^{\parallel}_{mn}(\xi^{\circ}_1, \xi^{\circ}_2)\mathfrak{f}^{\parallel}_m(\xi^{\circ}_3)\mathbf{a}_3, \tag{31a}$$

$$\mathbf{G}_{mn} = G^{\times}_{mn}(\xi^{\circ}_1, \xi^{\circ}_2)\mathfrak{g}^{\times}_m(\xi^{\circ}_3)\mathbf{b}\times\mathbf{a}_3 + G^{\perp}_{mn}(\xi^{\circ}_1, \xi^{\circ}_2)\mathfrak{g}^{\perp}_m(\xi^{\circ}_3)\mathbf{b} + G^{\parallel}_{mn}(\xi^{\circ}_1, \xi^{\circ}_2)\mathfrak{g}^{\parallel}_m(\xi^{\circ}_3)\mathbf{a}_3. \tag{31b}$$

$$\mathbf{H}_{mn} = H^{\times}_{mn}(\xi^{\circ}_1, \xi^{\circ}_2)\mathfrak{h}^{\times}_m(\xi^{\circ}_3)\mathbf{b}\times\mathbf{a}_3 + H^{\perp}_{mn}(\xi^{\circ}_1, \xi^{\circ}_2)\mathfrak{h}^{\perp}_m(\xi^{\circ}_3)\mathbf{b} + H^{\parallel}_{mn}(\xi^{\circ}_1, \xi^{\circ}_2)\mathfrak{h}^{\parallel}_m(\xi^{\circ}_3)\mathbf{a}_3. \tag{31c}$$

After some vector manipulations, Eqs. (28)–(30) are broken into a set of nine sixth-order equations

$$\sum_{m,n}\mathfrak{M}(\rho_{mn}f_m)^2(\varphi_{mn}g_m)F^{\times}_{mn}\mathfrak{f}^{\times}_m = \sum_{m,n}\{k_{\parallel}k_m^2\nabla^2 - k_{\perp}\nabla^2(\nabla^2 + k_{\parallel} + k_m^2)\} \times (\rho_{mn}f_m)(\varphi_{mn}g_m)\delta(\mathbf{r} - \mathbf{r}_0), \tag{32}$$

$$\sum_{m,n}\mathfrak{M}(\rho_{mn}f_m)^2(\varphi_{mn}g_m)F^{\perp}_{mn}\mathfrak{f}^{\perp}_m = \sum_{m,n}\{-k_T\nabla^2(\nabla^2 + k_{\parallel} + k_m^2)\} \times (\rho_{mn}f_m)(\varphi_{mn}g_m)\delta(\mathbf{r} - \mathbf{r}_0), \tag{33}$$

$$\sum_{m,n}\mathfrak{M}(\rho_{mn}f_m)^2(\varphi_{mn}g_m)F^{\parallel}_{mn}\mathfrak{f}^{\parallel}_m = \sum_{m,n}\pm k_T k_m\nabla^2(\rho_{mn}f_m)^2\delta(\mathbf{r} - \mathbf{r}_0), \tag{34}$$

$$\sum_{m,n}\mathfrak{M}(\rho_{mn}f_m)(\varphi_{mn}g_m)G^{\times}_{mn}\mathfrak{g}^{\times}_m = \sum_{m,n}\pm k_T k_{\parallel}k_m(\varphi_{mn}g_m)\delta(\mathbf{r} - \mathbf{r}_0), \tag{35}$$

$$\sum_{m,n}\mathfrak{M}(\rho_{mn}f_m)(\varphi_{mn}g_m)G^{\perp}_{mn}\mathfrak{g}^{\perp}_m = \sum_{m,n}\mp k_{\parallel}k_m(\nabla^2 + k_{\perp})(\varphi_{mn}g_m)\delta(\mathbf{r} - \mathbf{r}_0), \tag{36}$$

$$\sum_{m,n}\mathfrak{M}(\rho_{mn}f_m)(\varphi_{mn}g_m)G^{\parallel}_{mn}\mathfrak{g}^{\parallel}_m = \sum_{m,n}\{k_T^2 + k_{\perp}(\nabla^2 + k_{\perp})\}(\rho_{mn}f_m)\delta(\mathbf{r} - \mathbf{r}_0), \tag{37}$$

$$\sum_{m,n}\mathfrak{M}(\rho_{mn}f_m)(\varphi_{mn}g_m)H^{\times}_{mn}\mathfrak{h}^{\times}_m = \sum_{m,n}k_T(\nabla^2 + k_{\parallel})(\nabla^2 + k_m^2)(\varphi_{mn}g_m)\delta(\mathbf{r} - \mathbf{r}_0), \tag{38}$$

$$\sum_{m,n}\mathfrak{M}(\rho_{mn}f_m)(\varphi_{mn}g_m)H^{\perp}_{mn}\mathfrak{h}^{\perp}_m = \sum_{m,n} -(\nabla^2 + k_{\parallel}) \times (\nabla^2 + k_{\perp})(\nabla^2 + k_m^2)(\varphi_{mn}g_m)\delta(\mathbf{r} - \mathbf{r}_0), \tag{39}$$

$$\sum_{m,n}\mathfrak{M}(\rho_{mn}f_m)(\varphi_{mn}g_m)H^{\parallel}_{mn}\mathfrak{h}^{\parallel}_m = \sum_{m,n}\pm k_m\{k_T^2 + (\nabla^2 + k_{\perp})^2\}(\rho_{mn}f_m)\delta(\mathbf{r} - \mathbf{r}_0), \tag{40}$$

where the operator $\mathfrak{M}$ is an observer coordinate operator, it can be considered to be operating on any one of the two or three observer coordinate functions immediately to its right,

$$\mathfrak{M} = \begin{cases} k_{\perp}T_m^2(T_m^2 - {}_2T_m^2)(\nabla^2 - T_m^2), & \text{if } T_m^2 \to {}_1T_m^2; \\ k_{\perp}T_m^2(T_m^2 - {}_1T_m^2)(\nabla^2 - T_m^2), & \text{if } T_m^2 \to {}_2T_m^2; \end{cases} \tag{41}$$

where ${}_1T_m^2$ and ${}_2T_m^2$ are the two nontrivial roots of Eq. (12). In view of Eq. (9), all operators $\nabla^2$ to the right of the equality sign in Eqs. (32)–(40) are replaced by $(-T_m^2)$. Substitute Eq. (41) into Eqs. (32)–(40) and drop out the functions common to both sides of the equality sign. Multiply both sides by $f_m^*$ or $g_m^*$, whichever one is appropriate. Then integrate over the entire bounded $\xi_3$ space, utilizing the orthogonal properties of the eigenfunctions $f_m$ and $g_m$,

$$\int f_m f_m^*\, dv_{\xi_3} = \Lambda_m^2, \qquad \int g_m g_m^*\, dv_{\xi_3} = \Lambda_m^2, \tag{42}$$

where $\Lambda_m^2$ is the normalization factor. The asterisk (*) indicates the complex conjugate. The integration yields distinct solutions for the $\xi_3^{\circ}$-dependent functions,

$$\mathfrak{f}^{\parallel}_m(\xi^{\circ}_3) = \mathfrak{h}^{\parallel}_m(\xi^{\circ}_3) = \mathfrak{g}^{\parallel}_m(\xi^{\circ}_3) = (1/\Lambda_m^2)g_m^*(\xi^{\circ}_3), \qquad \mathfrak{f}^{\times}_m(\xi^{\circ}_3) = \mathfrak{f}^{\perp}_m(\xi^{\circ}_3) = \mathfrak{g}^{\times}_m(\xi^{\circ}_3) = \mathfrak{g}^{\perp}_m(\xi^{\circ}_3) = \mathfrak{h}^{\times}_m(\xi^{\circ}_3) = \mathfrak{h}^{\perp}_m(\xi^{\circ}_3) = (1/\Lambda_m^2)f_m^*(\xi^{\circ}_3). \tag{43}$$

Equation (42) together with Eq. (43) indicate that the functions $f_m(\xi_3)f_m^*(\xi_3^{\circ})$ and $g_m(\xi_3)g_m^*(\xi_3^{\circ})$ are one-dimensional scalar Green's function, both satisfying a one-dimensional equation:

$$\nabla^2_{\parallel}(f_m f_m^*,\ g_m g_m^*) + k^2(f_m f_m^*,\ g_m g_m^*) = \delta(\xi_3 - \xi_3^0). \tag{44}$$

After integrating out the $\xi_3$- and $\xi_3^{\circ}$-dependent functions, Eqs. (32)–(40) become convenient two-dimensional simultaneous equations involving $(\xi_1, \xi_2)$ and $(\xi_1^{\circ}, \xi_2^{\circ})$ only. It may be demanded that

$$\{\nabla_{\perp}^2 + \kappa_m^2\}(\rho_{mn}\mathfrak{F}^{\times}_{mn})(\mathbf{b}\times\mathbf{a}_3)(\mathbf{b}\times\mathbf{a}_3) = -\kappa_m^2(\nabla_{\perp}^2 + \kappa_m^2)\,||\varphi_{mn}\mathfrak{G}^{\times}_{mn}||\,(\mathbf{b}\times\mathbf{a}_3)(\mathbf{b}\times\mathbf{a}_3), \tag{45}$$

and

$$\{\nabla_{\perp}^2 + \kappa_m^2\}(\rho_{mn}\mathfrak{F}^{\perp}_{mn})\mathbf{bb} = \kappa_m^2(\nabla_{\perp}^2 + \kappa_m^2)\,||\varphi_{mn}\mathfrak{G}^{\perp}_{mn}||\,\mathbf{bb}, \tag{46}$$

where the two double bars bracketing a function indicates only the scalar is being considered. $\mathfrak{F}_{mn}$ and $\mathfrak{G}_{mn}$ represent anyone of the $(\xi_1^{\circ}, \xi_2^{\circ})$ functions corresponding to $\rho_{mn}$ and $\varphi_{mn}$, respectively. With Eqs. (45) and (46), all nine equations in Eqs.

(32)–(40) may be represented by one symbolic equation,

$$\{\nabla_\perp^2 + T_m^2 - k_m^2\} g_{mn}(\xi_1, \xi_2 \mid \xi_1^0, \xi_2^0) = (\text{constant})\delta(\xi_1 - \xi_1^0)\delta(\xi_2 - \xi_2^0). \quad (47)$$

One thus has reduced the problem of solving the Green's dyadic to one of searching for an appropriate two-dimensional scalar Green's function. Exact solutions of Eq. (47) depend upon the coordinate systems employed and the type of boundary perpendicular to $\xi_3$ = const surfaces considered. In general, it can be written symbolically

$$g_{mn} = \frac{(\text{constant})}{\nu^2}\varphi_{mnl}(\xi_1, \xi_2)\tilde{\varphi}_{mnl}(\xi_1^0, \xi_2^0), \quad \text{for closed boundary in } \perp; \quad (48)$$

$$\frac{(\text{constant})}{\nu^2}\varphi_{mn}(\xi_1, \xi_2)\tilde{\varphi}_{mn}(\xi_1^0, \xi_2^0), \quad \text{for open boundary in } \perp.$$

In the case of a closed boundary in $\perp$, $\tilde{\varphi}_{mnl} = \varphi_{mnl}^*$ is the complex conjugate of $\varphi_{mnl}$; and $\nu^2 = \Lambda_n^2\Lambda_l^2$, $\Lambda_n^2$ and $\Lambda_l^2$ are the two normalization factors. In the case of an open boundary, $\varphi_{mn}$ and $\tilde{\varphi}_{mn}$ are the two independent solutions of Eq. (10a) and $\nu^2$ is a constant involving the Wronskian of the two independent solutions. Let

$$t_{mn}(\xi_1, \xi_2 \mid \xi_1^0, \xi_2^0) = \varphi_{mnl}(\xi_1, \xi_2)\tilde{\varphi}_{mnl}(\xi_1^0, \xi_2^0), \quad \text{for closed boundary in } \perp; \quad (49)$$

$$\varphi_{mn}(\xi_1, \xi_2)\tilde{\varphi}_{mn}(\xi_1^0, \xi_2^0), \quad \text{for open boundary in } \perp.$$

The complete Green's dyadic is found to be

$$\mathbf{G}(\mathbf{r} \mid \mathbf{r}_0) = \sum_{m,n,l}\frac{1}{\tilde{\eta}_m\nu^2\Lambda_m^2}\left\{\frac{\kappa_m^2 - k_\parallel + \frac{k_\parallel}{k_\perp}k_m^2}{\kappa_m^2}(\nabla_\perp \times \mathbf{a}_3)(\nabla_\perp^0 \times \mathbf{a}_3)\right.$$
$$+ \frac{T_m^2 - k_\perp}{k_\perp\kappa_m^2}(\kappa_m^2 - k_\parallel)(\nabla_\perp)(\nabla_\perp^0)$$
$$+ \frac{1}{k_\perp k_m^2}[k_T^2 - (k_\perp - k_m^2)(T_m^2 - k_\perp)](\nabla_\parallel)(\nabla_\parallel^0)$$
$$- \frac{k_T}{k_\perp}\left(\frac{\kappa_m^2 - k_\parallel}{\kappa_m^2}\right)[(\nabla_\perp \times \mathbf{a}_3)(\nabla_\perp^0) + (\nabla_\perp)(\nabla_\perp^0 \times \mathbf{a}_3)]$$
$$- \frac{k_T}{k_\perp}[(\nabla_\perp \times \mathbf{a}_3)(\nabla_\parallel^0) + (\nabla_\parallel)(\nabla_\perp^0 \times \mathbf{a}_3)]$$
$$\left. + \frac{T_m^2 - k_\perp}{k_\perp}[(\nabla_\perp)(\nabla_\parallel^0) + (\nabla_\parallel)(\nabla_\perp^0)]\right\}$$
$$\times t_{mn}(\xi_1, \xi_2 \mid \xi_1^0, \xi_2^0)f_m(\xi_3)f_m^*(\xi_3^0); \quad (50)$$

where

$$\tilde{\eta}_m = \begin{cases} {}_1T_m^2 - {}_2T_m^2, & \text{if } T_m^2 \to {}_1T_m^2; \\ {}_2T_m^2 - {}_1T_m^2, & \text{if } T_m^2 \to {}_2T_m^2. \end{cases} \quad (50a)$$

Symbolically, Eq. (50) may be written as

$$\mathbf{G} = \sum_m\sum_n\sum_l\sum_i\sum_j\{S^{i,j}(k_m^2, \kappa_m^2)(\mathfrak{L}^i) \times (\mathfrak{L}_0^j)t_{mn}f_mf_m^*\}, \quad (51)$$

where $\mathfrak{L}^i$ are the space coordinate differential operators, i.e., $\nabla_\perp \times \mathbf{a}_3$, $\nabla_\parallel$, and $\nabla_\perp$; and $S^{i,j}(k_m^2, \kappa_m^2)$ are the algebraic functions shown in Eq. (50); the indices $i$ and $j$ run on the terms corresponding to the vectors $\mathbf{a}_3$, $\mathbf{b}$, and $\mathbf{b} \times \mathbf{a}_3$.

In the circular cylindrical coordinate system with open boundary in $r$ and $\theta$, the solution for Eq. (47) is

$$g_{mn}(\mathbf{r}_\perp \mid \mathbf{r}_{0\perp}) = \frac{(\text{constant})}{\nu^2}e^{in(\theta-\theta_0)} \times \begin{cases} J_n(\kappa_m r_0)H_n^{(2)}(\kappa_m r), & r \geq r_0; \\ H_n^{(2)}(\kappa_m r_0)J_n(\kappa_m r), & r \leq r_0. \end{cases} \quad (52)$$

Identifying $f_m(\xi_3)$ with $\cos k_m z$ and $g_m(\xi_3)$ with $\sin k_m z$, in line with Eqs. (16)–(18), the Green's dyadic valid for the case of two parallel conducting plates is (for $r \geq r_0$ only)

$$\mathbf{G}(\mathbf{r} \mid \mathbf{r}_0) = \frac{j}{4}\sum_{m,n}\frac{\kappa_m^2 - k_\parallel + k_\parallel\frac{k_m^2}{k_\perp}}{\tilde{\eta}_m\kappa_m^2 d}\{(\nabla_\perp H_n^{(2)} \times \mathbf{a}_3)(\nabla_\perp^0 J_n \times \mathbf{a}_3)\cos k_m z \cos k_m z_0\}$$
$$+ \frac{j}{4}\sum_{m,n}\frac{(\kappa_m^2 - k_\parallel)(T_m^2 - k_\perp)}{\tilde{\eta}_m k_\perp\kappa_m^2 d}(\nabla_\perp H_n^{(2)})(\nabla_\perp^0 J_n)\cos k_m z \cos k_m z_0$$
$$+ \frac{j}{4}\sum_{m,n}\frac{k_T^2 - (k_\perp - k_m^2)(T_m^2 - k_\perp)}{\tilde{\eta}_m k_\perp d}H_n^{(2)}J_n \sin k_m z \sin k_m z_0\mathbf{a}_3\mathbf{a}_3$$
$$- \frac{j}{4}\sum_{m,n}\frac{k_T(\kappa_m^2 - k_\parallel)}{\tilde{\eta}_m k_\perp\kappa_m^2 d}\{(\nabla_\perp H_n^{(2)} \times \mathbf{a}_3)(\nabla_\perp^0 J_n)\cos k_m z \cos k_m z_0 + (\nabla_\perp H_n^{(2)})(\nabla_\perp^0 J_n \times \mathbf{a}_3)\cos k_m z \cos k_m z_0\}$$
$$+ \frac{j}{4}\sum_{m,n}\frac{k_T k_m}{\tilde{\eta}_m k_\perp d}\{(\nabla_\perp H_n^{(2)} \times \mathbf{a}_3)(J_n\mathbf{a}_3)\cos k_m z \sin k_m z_0 + (H_n^{(2)}\mathbf{a}_3)(\nabla_\perp^0 J_n \times \mathbf{a}_3)\sin k_m z \cos k_m z_0\}$$
$$- \frac{j}{4}\sum_{m,n}\frac{k_m(T_m^2 - k_\perp)}{\tilde{\eta}_m k_\perp d}\{(\nabla_\perp H_n^{(2)})(J_n\mathbf{a}_3)\cos k_m z \sin k_m z_0 + (H_n^{(2)}\mathbf{a}_3)(\nabla_\perp^0 J_n)\sin k_m z \cos k_m z_0\}. \quad (53)$$

### THE INFINITE-DOMAIN GREEN'S DYADIC

The transition of finite- or semifinite-domain Green's dyadic to the infinite-domain Green's dyadic may be obtained through a limiting process. Since the infinite-domain Green's function has been derived by a number of authors and employed extensively,[1,2,3] no attempt will be made here to derive the infinite-domain Green's dyadic into its final form. The main purpose here is to show that such transition is possible.

As an example, take the Green's dyadic of the two-parallel-plates case given in Eq. (53). As the plates recede to infinity, i.e., $d \to \infty$, the summation on $m$ goes over to an integral. Written in the symbolic form of Eq. (51), the transformed infinite-domain Green's dyadic $\mathbf{g}(\mathbf{r} \mid \mathbf{r}_0)$ is

$$\mathbf{g}(\mathbf{r} \mid \mathbf{r}_0) = \sum_{i,j} \sum_{n} (\mathfrak{F}^i)(\mathfrak{F}_0^j) \int_{0}^{\infty} S^{i,j}(k) t_n f f^* \, dk. \quad (54)$$

In Eq. (54) the order of integration and differentiation operations has been interchanged and the subscript $m$ has been dropped. Evidently, the Green's dyadic for infinite domain can be obtained through a set of auxiliary scalar functions, $I^{i,j}$, as represented by the integral in Eq. (54). In view of Eq. (52), Eq. (44), and the fact that $ff^*$ is an even function, after using the addition theorem to perform the summation on $n$, $I^{i,j}$ may be written

$$I^{i,j} = \frac{1}{2} \int_{-\infty}^{\infty} S^{i,j} H_0^{(2)}(\kappa r'_\perp) e^{-jks'} \, dk, \quad (55)$$

where

$$r'_\perp = |\mathbf{r}_\perp - \mathbf{r}_{0\perp}|, \qquad z' = |z - z_0|.$$

It should be noted that the new coordinate system has its origin at the source point. This choice of a new origin may require subsequent transformation back to the original origin.

An exact solution of Eq. (55) is tedious and is not easily attainable; however, an asymptotic solution which is valid for waves at large distance from the source may be obtained through the method of steepest descent.

Assuming that interest is in the accuracy of the solution only to the order of $1/r'_\perp$, the zero-order Hankel function may be expanded into its asymptotic form. Retaining only the first term, Eq. (55) then becomes

$$I^{i,j} = \frac{e^{-j\frac{1}{4}\pi}}{2} \int_{-\infty}^{\infty} \frac{S^{i,j} e^{-jks'} e^{-j\kappa r'_\perp}}{(\pi\kappa r'_\perp)^{\frac{1}{2}}} \, dk. \quad (56)$$

At this point, it would deem more convenient to change the coordinate system from that of circular cylindrical coordinates to that of spherical coordinates $(R, \varphi, \alpha)$; where

$$r'_\perp = R \sin \varphi,$$
$$z' = R \cos \varphi.$$

Under the new coordinate system, Eq. (56) becomes

$$I^{i,j}(\mathrm{R}) = \frac{e^{-j(\pi/4)}}{2} \int_{-\infty}^{\infty} \frac{S^{i,j}}{(\pi\kappa R \sin \varphi)^{\frac{1}{2}}} \times \exp\left[-jR(\kappa \sin \varphi + k \cos \varphi)\right] dk. \quad (57)$$

It is recalled that

$$T^2 = \kappa^2 + k^2. \quad (58)$$

$T$ is therefore the total propagation factor. Now for the sake of convenience, instead of $k$, a new integration parameter, $\tau$, may be employed, such that

$$\kappa = T \sin \tau, \qquad k = T \cos \tau. \quad (59)$$

The parameter $\tau$ has the same significance as the angle which measures the wave normal if $T$ is a constant; however, in the present case $T$ is not a constant. In fact, combining Eq. (12), Eq. (59) and Eq. (58) yields an expression for $T$ in terms of $\tau$,

$$T_{1,2}^2 = \frac{-[k_\perp(k_\perp - k_\parallel) - k_T^2] \sin^2 \tau + 2k_\perp k_\parallel \pm \{[k_\perp(k_\perp - k_\parallel) - k_T^2]^2 \sin^4 \tau + 4k_\parallel^2 k_T^2 \cos^2 \tau\}^{\frac{1}{2}}}{2[k_\perp \sin^2 \tau + k_\parallel \cos^2 \tau]}, \quad (60)$$

where, the subscript 1 and 2 on $T^2$ represents the choice of plus or minus sign in Eq. (60). For simplicity, the subscripts on $T^2$ are dropped, assuming that it is permissible to work with one wave at a time. The integral for $I^{i,j}$ becomes

$$I^{i,j} = \frac{1}{(j4R \sin \varphi)^{\frac{1}{2}}} \int_C \frac{S^{i,j}}{[\pi T(\tau) \sin \tau]^{\frac{1}{2}}} e^{-RU(\tau)} \, d\tau, \quad (61)$$

where

$$U(\tau) = jT(\tau) \cos(\tau - \varphi). \quad (62)$$

Examination of the exponent shows that the real part of $U(\tau)$ approaches $+\infty$ as $k$ approaches $\pm\infty$. The saddle point of the integration is determined by

$$dU/d\tau = 0, \quad (63)$$

which yields

$$1T(\tau_0)/[(d/d\tau)T(\tau)]_{\tau=\tau_0} = \tan(\tau_0 - \varphi). \quad (64)$$

The contour of integration, $C$, is then chosen such that the path goes through the saddle point, $\tau_0$, and that the imaginary part of $U$ is constant. Following the method of steepest descent, the solution for Eq. (61) is therefore

$$I^{i,j} = \frac{S^{i,j}(\tau_0)}{2n(\tau_0)} \frac{e^{-RU(\tau_0)}}{R(\sin\varphi)^{\frac{1}{2}}}, \quad (65)$$

where

$$\eta(\tau_0) = \left\{2T \sin\tau\left[\left(\frac{dT}{d\tau} - T\right)\cos(\tau - \varphi) - 2\frac{dT}{d\tau}\sin(\tau - \varphi)\right]\right\}^{\frac{1}{2}}_{\tau=\tau_0}. \quad (66)$$

The electric field intensity $\mathbf{E}$ in infinite domain may, therefore, be obtained from

$$\mathbf{E} = \sum_{i,j}\int_{v_0}\left[(\mathfrak{L}^i)(\mathfrak{L}_0^j)\frac{S^{i,j}}{2n(\tau_0)}\frac{e^{-RU(\tau_0)}}{(R^2\sin\varphi)^{\frac{1}{2}}}\right]\cdot\mathbf{J}_s\, dv_0, \quad (67)$$

providing that all parameters, including the differential operators are properly transformed to the correct observer and source coordinates in the spherical coordinate system.

## CONCLUSION

The results of separability studies in this work indicate that the dyadic–vector Helmholtz equation is solvable by the separation technique in four cylindrical coordinate systems. It may be noted that the solutions obtained by the separation technique are uncoupled, i.e., it is possible to solve for one field of the waves without explicit knowledge of the other field. Such simplicity may be contrasted to the coupled field solution that often prevailed in the past. In the past, free wave solutions in a bounded anisotropic plasma often has been obtained by direct manipulations of Maxwell's equations and the generalized Ohm's law. Such manipulations often led to second order differential equations such that the fields are coupled, i.e., the electric field is solvable in terms of the magnetic field and vice versa. Except for some special cases, to uncouple the fields, the order of the differential equations must be raised beyond two and thereby increases the difficulty in obtaining solutions in simple form.

The Green's dyadic constructed through sets of eigenfunctions for finite or semifinite domain problems is expressed in terms of differential operators which have the advantage of ease of operation over integral operators.

The form of solutions for infinite domain problems as shown in Eq. (67) is not exactly of the same form obtained by Bunkin.[1] The most noticeable difference lies in the manner of operation. Bunkin's solution requires two second-order differential operations, while Eq. (67) requires only two first-order differential operations. However, the result of Eq. (67) compares favorably with that obtained by Bunkin.

## APPENDIX A

The inhomogeneous equation is

$$\nabla\times\nabla\times\mathbf{E} - \mathsf{k}\cdot\mathbf{E} = \mathbf{J}_s. \quad (A1)$$

A Green's equation is assumed

$$\nabla\times\nabla\times\mathsf{G} - \mathsf{K}\cdot\mathsf{G} = \mathsf{I}\delta(\mathbf{r} - \mathbf{r}_0). \quad (A2)$$

Multiply $\mathsf{G}$ from the right into Eq. (A1) and multiply $\mathbf{E}$ from the left into Eq. (A2), subtract and integrate over the entire space on the source coordinate yielding

$$\mathbf{E}(\mathbf{r}) = \int \mathsf{G}\cdot\mathbf{J}_s\, dv_0 + \int \{\mathbf{E}\cdot\nabla\times\nabla\times\mathsf{G} - (\nabla\times\nabla\times\mathbf{E})\cdot\mathsf{G}\}\, dv_0 - \int \{\mathbf{E}\cdot\mathsf{K}\cdot\mathsf{G} - \mathsf{k}\cdot\mathbf{E}\cdot\mathsf{G}\}\, dv_0. \quad (A3)$$

Using Green's theorem the second integral can be transformed into a surface integral. If the Green's dyadic satisfies the same boundary conditions the $\mathbf{E}$ field satisfies, the surface integral vanishes.

The dyadics $\mathsf{k}$ and $\mathsf{K}$ in the third integral may be expressed in terms of their symmetrical components (subscript $s$) and antisymmetrical components (subscript $a$)

$$\mathsf{k} = \mathsf{k}_s + \mathsf{k}_a, \quad (A4)$$

$$\mathsf{K} = \mathsf{K}_s + \mathsf{K}_a. \quad (A5)$$

Substituting into the integral, assuming $\mathsf{G}$ being symmetric and reciprocal with respect to $\mathbf{r}$ and $\mathbf{r}_0$, it is found that

$$\mathbf{E}\cdot\mathsf{K}_s\cdot\mathsf{G} - \mathsf{k}_s\cdot\mathbf{E}\cdot\mathsf{G} = 0, \quad \text{if} \quad \mathsf{K}_s = \mathsf{k}_s;$$

and

$$\mathbf{E}\cdot\mathsf{K}_a\cdot\mathsf{G} - \mathsf{k}_a\cdot\mathbf{E}\cdot\mathsf{G} = 0, \quad \text{if} \quad \mathsf{K}_a = -\mathsf{k}_a.$$

Thus, it is shown that for Eq. (20) to hold, the third integral must also vanish, or

$$\mathsf{K} = \tilde{\mathsf{k}}. \quad (A6)$$

## APPENDIX B

Equations (45) and (46) in essence demand that the Green's dyadic be symmetrical. They also imply a condition for the $(\xi_3, \xi_3^0)$ functions such that the solution for Eq. (44) must be chosen to satisfy

$$\nabla_{\parallel}\nabla_{\parallel}^{0}(f_m f_m^*, g_m g_m^*) = k_m^2(f_m f_m^*, g_m g_m^*)\mathbf{a}_3\mathbf{a}_3 \ . \quad \text{(B1)}$$

Equations (45) and (46) along with Eq. (B1) imply

$$\nabla_{\perp}^{0} \times \mathbf{a}_3 = -\nabla_{\perp} \times \mathbf{a}_3, \quad \text{(B2)}$$

$$\nabla_{\perp}^{0} = -\nabla_{\perp}, \quad \text{(B3)}$$

and

$$\nabla_{\parallel}^{0} = -\nabla_{\parallel}. \quad \text{(B4)}$$

The symbolic form of the Green's dyadic, as is given in Eq. (51), cannot be symmetric unless the source coordinates operator $\mathfrak{L}_0^i$ and the observer coordinates operator $\mathfrak{L}^i$ can be interchanged. Of course, Eqs. (B2) through (B4) are not the only possible conditions that may force the Green's dyadic to be symmetric.

The Green's dyadic is symmetric only in the coordinate system for which the Green's dyadic is constructed. Using variational technique, a given Green's dyadic may be transformed to one that is valid for a problem of different boundary configuration in a different coordinate system. But the symmetrical property of the original Green's dyadic is not necessarily retained in the transformation. This is especially true in the case of the Green's dyadic for the infinite domain.

## ACKNOWLEDGMENT

The work reported here was supported in part by the National Aeronautics and Space Administration.

# Triality Type and its Generalization in Unitary Symmetry Theories*

C. R. HAGEN AND A. J. MACFARLANE
*Department of Physics and Astronomy, University of Rochester, Rochester, New York*
(Received 24 March 1964)

Within the context of an extension of the $SU_3$-symmetry theory recently suggested by Gell-Mann and further developed by the authors, certain aspects of the theory of the special unitary groups are examined. The plurality type of a given representation is introduced as the generalization of the triality concept to $SU_{n+1}$ and is shown to be associated with a multiplicative conservation law. Theorems for the reduction of representations of $SU_{n+1}$ with respect to $SU_n \otimes U_1^{(n)}$ are derived which are subsequently used to relate plurality type to the existence of fractional eigenvalues for the generator $Y_1^{(n)}$ of $U_1^{(n)}$.

## I. INTRODUCTION

IN a recent paper, Gell-Mann[1] has raised the question of the possible occurrence in nature of particles (quarks) which can be associated with the states of the triplet representations of $SU_3$. Such particles must necessarily have nonintegral eigenvalues of $Q/e$ if their quantum numbers are to satisfy the Gell-Mann–Nishijima formula

$$Q/e = I_z + \tfrac{1}{2}Y. \tag{1.1}$$

In the language of Biedenharn and Fowler,[2] we may say that the triplet representations of $SU_3$ have triality type[3] $t \neq 0$, and that the nonintegral charge associated with the basic quark is a characteristic of all states belonging to representations of $SU_3$ with $t \neq 0$. Of course, all particles known at present have been accommodated within representations of $SU_3$ with $t = 0$ for which integral eigenvalues of $Q/e$ follow from Eq. (1.1).

More recently, the authors[4] have suggested a theory of strong interaction symmetries, wherein representations of $SU_3$ with $t \neq 0$ can be used for particles whose charges are integral multiples of $e$. It is the purpose of the present paper to provide the mathematical background necessary for this theory. Although we believe that much of the mathematical work to be exhibited is of intrinsic value, we wish to present it in its physical context. Accordingly we begin with a review of the content of our previous paper.[4]

If new particles are discovered whose isospin and hypercharge suggest that they belong to representations of $SU_3$ with $t \neq 0$, then one can postulate that there exists a new conserved quantum number for the strongly interacting particles. We refer to it as $Y^{(3)}$, to distinguish it from hypercharge $Y \equiv Y^{(2)}$, and specify the following assignments:

(a) $Y^{(3)} = 0$ for representations of $SU_3$ with $t=0$,
(b) $Y^{(3)} \neq 0$ for representations with $t \neq 0$.

From (a), we see that $Y^{(3)} = 0$ for all particles known at present. From (b), it follows that one can arrange that states of representations of $SU_3$ with $t \neq 0$ correspond to integral eigenvalues of $Q/e$ by replacing Eq. (1.1) by

$$Q/e = I_z + \tfrac{1}{2}Y^{(2)} + \tfrac{1}{3}Y^{(3)}. \tag{1.2}$$

In order to provide a definite framework within which our postulates can be realized, we propose the extension of the symmetry group of the strong interactions from $SU_3 \otimes U_1^{(3)}$, where $U_1^{(3)}$ is the gauge group generated by $Y^{(3)}$, to $SU_4$. This extension, of course, is the analog of the previous extension of the strong interaction symmetry group from $SU_2 \otimes U_1^{(2)}$, i.e., isospin and hypercharge symmetry, to $SU_3$. The realization of the above ideas follows if one assumes that the $SU_4$ theory uses only representations of $SU_4$ with quadrality $k = 0$. We see this by noting that the representations of $SU_3 \otimes U_1^{(3)}$ contained within such representations of $SU_4$ always associate $Y^{(3)} = 0$ with $t = 0$, and $Y^{(3)} \neq 0$ with $t_3 \neq 0$ in such a way that (1.2) leads to integral eigenvalues of $Q/e$. The $SU_4$ theory thus developed resembles $SU_3$ theory as it now stands. The question of the occurrence of representations of $SU_4$ with $k \neq 0$ arises, and the same sequence of ideas repeats. Equation (1.3) leads to nonintegral eigenvalues of $Q/e$ for states of representations of $SU_4$ with $k \neq 0$. So one can postulate again a new quantum number $Y^{(4)}$, whose assignments, made in analogy with those for $Y^{(3)}$, are such that replacement of (1.2) by

* This research was supported in part by the U. S. Atomic Energy Commission.
[1] M. Gell-Mann, Phys. Letters 8, 214 (1964).
[2] L. C. Biedenharn and Earle C. Fowler (preprint, 1963). See also G. E. Baird and L. C. Biedenharn, paper presented at the Conference on Symmetry Principles at High Energy, University of Miami, January, 1964.
[3] The definition of triality type for a representation of $SU_3$ is given as the $n = 2$ case of the definition (2.6), below.
[4] C. R. Hagen and A. J. Macfarlane, Phys. Rev. 135, B 432 (1964).

$$Q/e = I_z + \tfrac{1}{2}Y^{(2)} + \tfrac{1}{3}Y^{(3)} + \tfrac{1}{4}Y^{(4)} \qquad (1.3)$$

obviates the occurrence of nonintegral eigenvalues of $Q/e$. Then as before we can realize this situation by extending the strong interaction symmetry group from $SU_4 \otimes U_1^{(4)}$, where $U_1^{(4)}$ is generated by $Y^{(4)}$, to $SU_5$, agreeing to use only representations of $SU_5$ of quintality $q = 0$. In this way a finite hierarchy of unitary symmetry theories can be generated, each with its appropriate definition of $Q$.

Of central importance in the theory outlined above is the concept of triality type for representations of $SU_3$ and its generalization to $SU_{n+1}$, which, for want of a better alternative, has been given the generic name, plurality type.[5] In Sec. II, after a preliminary discussion of notation, we define plurality type and obtain its most important property, namely its conservation modulo $(n + 1)$. We also define a multiplicative quantum number closely related to plurality type. In Sec. III, theorems are derived for the reduction of representations of $SU_{n+1}$ with respect to the subgroup $SU_n \otimes U_1^{(n)}$. These are used to relate the occurrence of fractional eigenvalues of $Y_1^{(n)}$, canonically defined by such a subgroup inclusion, to the plurality type of representations of $SU_{n+1}$. We then prove that the successive definitions of $Q/e$ have the properties claimed above.

## II. DEFINITION AND PROPERTIES OF PLURALITY TYPE

We begin by discussing the relationship[6] of irreducible representations (IR's) of unitary groups to Young diagrams. It is well known that IR's of $U_{n+1}$ may be uniquely characterized by a set of integers

$$[f]_{n+1} = [f_1, f_2, \cdots f_{n+1}] \qquad (2.1)$$

such that $f_i \geq f_{i+1}$ for $i = 1, \cdots, n$. It is sufficient for present purposes to note that those IR's with $f_{n+1} \geq 0$ can be placed in one-to-one correspondence with Young diagrams of $(n + 1)$ rows, $f_i$ being the number of boxes in the $i$th row. On restriction from $U_{n+1}$ to $SU_{n+1}$, IR's of $U_{n+1}$ remain irreducible, but no longer do inequivalent IR's remain inequivalent. Indeed, the representations $[f]_{n+1}$ and $[f']_{n+1}$ of $U_{n+1}$ become equivalent representations of $SU_{n+1}$ if

$$f_i' - f_i = e \qquad (2.2)$$

for all $i$ and some integer $e$. Hence we can and always do arrange to have $f_{n+1} = 0$ for all IR's of $SU_{n+1}$. In other words, the IR's of $SU_{n+1}$ can be put into one-to-one correspondence with Young diagrams of $n$ rows and can thus be characterized by the more appropriate notation

$$\{l\}_n = \{l_1, \cdots l_n\}, \qquad (2.3)$$

where

$$l_i \geq l_{i+1} \quad \text{for} \quad i = 1, \cdots n - 1,$$

and

$$l_n \geq 0.$$

In view of the common usage of the highest weight notation[7] for IR's of $SU_{n+1}$, it is desirable to make explicit its relation to the notation (2.3). An IR of $SU_{n+1}$ can be uniquely characterized by the values $\lambda_i \geq 0$, $i = 1, \cdots, n$ of the $n$ components of its highest weight

$$(\lambda)_n = (\lambda_1, \lambda_2, \cdots, \lambda_n). \qquad (2.4)$$

If $\{l\}_n$ and $(\lambda)_n$ are the respective notations for the same representation of $SU_{n+1}$, the $l_i$ and $\lambda_i$ are related by

$$\lambda_i = l_i - l_{i+1}, \qquad i = 1, 2, \cdots, n - 1 \qquad (2.5a)$$

and

$$\lambda_n = l_n. \qquad (2.5b)$$

One can regard the $\lambda_i$ as giving the number of boxes by which the $i$th row of the Young diagram of $\{l\}_n$ projects beyond the $(i + 1)$th row.

We may now give the definition of the plurality type $p\{l\}_n$ of the IR $\{l\}_n$ of $SU_{n+1}$ in the form

$$p\{l\}_n = \sum_{i=1}^{n} l_i, \qquad \text{modulo } (n + 1), \qquad (2.6)$$

which for the special case of $SU_3$ reduces to the definition of triality type given by Biedenharn and Fowler.[2] The importance of the concept of plurality type in the representation theory of $SU_{n+1}$ stems from the following theorem.

*Theorem.* For all $\{k\}_n$ that occur in the direct product

$$\{l\}_n \otimes \{m\}_n, \qquad (2.7)$$

one has

$$p\{k\}_n = p\{l\}_n + p\{m\}_n, \quad \text{modulo } (n + 1). \qquad (2.8)$$

[5] The concept of plurality type is implicit in the work of H. Weyl, Lecture Notes, Princeton University, Princeton, New Jersey, (1935). See also E. Stiefel, Commun. Math. Helv. 14, 350 (1943). We thank the referee for having drawn our attention to these sources.

[6] For proof of the statements made in this paragraph, see M. Hamermesh, *Group Theory* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1962), Chap. 10.

[7] See G. Racah, "Group Theory and Spectroscopy," Lecture Notes, Princeton University, Princeton, New Jersey (1951).

The proof follows directly from the Littlewood method[8] of performing the explicit reduction of direct products, a technique which has been applied by Edmonds[9] in the $n = 2$ case ($SU_3$). This method involves the addition, in as many different ways as are allowed, of a total number $m = \sum_{i=1}^{n} m_i$ of boxes to the Young diagram of $\{l\}_n$ to form a Young diagram with $(l + m) = \sum_{i=1}^{n} (l_i + m_i)$ boxes and at most $(n + 1)$ rows. Allowed ways are those which are in accordance with a clearly stated set of rules, whose precise nature need not concern us here. To each one of the distinct allowed ways of obtaining a suitable Young diagram, there corresponds one and only one irreducible constituent of the direct product (2.7). If the Young diagram of any such irreducible constituent $\{k\}_n$ has $n$ rows or less, we have $k = \sum_{i=1}^{n} k_i = l + m$, so that the result (2.8) follows directly from the definition (2.6). However, if the Young diagram has $(n + 1)$ rows with $x$ ($x > 0$) boxes in the last row, we can replace it by an equivalent Young diagram with $x$ less boxes in each row, and in particular, no boxes in the last row. Calling the IR of $SU_{n+1}$ to which the latter Young diagram corresponds $\{k\}_n$, we get $k = l + m - (n + 1)x$, which again implies (2.8), thus completing the proof of the theorem.

Instead of introducing plurality type as a quantum number which is conserved modulo $(n + 1)$ for IR's of $SU_{n+1}$, we may introduce the conserved multiplicative quantum number

$$P\{l\}_n = \exp [2\pi i p\{l\}_n/(n + 1)]. \tag{2.9}$$

Clearly the result

$$P\{k\}_n = P\{l\}_n P\{m\}_n$$

follows from the theorem.

Finally, in preparation for the next section, we fix a suitable notation for the IR's of the group $U_1$ to which the remarks made at the start of this section do not apply. The group $U_1$ is a one parameter Abelian gauge group, and, in general, the quantity $e^{iy\phi}$ for any real number $y$ is an IR of $U_1$ corresponding to the element $e^{i\phi}$ of the group. We refer to the IR as $y$ without any brackets. In the case of the hypercharge gauge group discussed in the next section, the allowed numbers $y$ are the eigenvalues of the generator $Y$ of the gauge group.

[8] See D. E. Littlewood, *Theory of Group Characters and Matrix Representation of Groups*, (Oxford University Press, Oxford, England, 1940), pp. 91–98.

[9] A. R. Edmonds, Proc. Roy. Soc. (London) **A268**, 567 (1962).

## III. REDUCTION THEOREMS

We now turn to the derivation of an important theorem for the reduction of a representation of $SU_{n+1}$ with respect to the subgroup $SU_n \otimes U_1^{(n)}$ of $SU_{n+1}$. It will be noted that the intended theorem contains a statement of the allowed values of $Y_1^{(n)}$, the generator of $U_1^{(n)}$, within any representation of $SU_{n+1}$. It is also worth pointing out that the subgroup $SU_n \otimes U_1^{(n)}$ is precisely that used by Biedenharn[10] to provide a canonical method of labeling basis states of representations of $SU_{n+1}$.

It is sufficient to give a detailed treatment of the special case of the reduction of a representation of $SU_3$ with respect to $SU_2 \otimes U_1^{(2)}$, $Y^{(2)} = Y$ (hypercharge) being the generator of $U_1^{(2)}$. The generalization of the result to the case of $SU_{n+1}$ and the method of obtaining it, can in this way be made obvious. The problem for $SU_3$ has, of course, already been solved by many people, some of whom[11] employ methods related to ours, while others[12] utilize completely different approaches.

The essential tool is the Weyl formula for the character of an IR $[l]_3 = [l_1, l_2, l_3]$ of $U_3$ corresponding to an element of $U_3$ with eigenvalues $\epsilon_i$ ($|\epsilon_i| = 1$, $i = 1, 2$ and 3). We use only the $l_3 = 0$ case of this formula, so that the IR $[l_1, l_2, 0] \to \{l_1, l_2\} = \{l\}_2$ on restriction to $SU_3$. In terms of the $\epsilon_i$ this restriction is expressed by

$$\epsilon_1\epsilon_2\epsilon_3 = 1. \tag{3.1}$$

The formula in question is obtained from[13]

$$\chi([l]_3, \epsilon_1, \epsilon_2, \epsilon_3) = \begin{vmatrix} \epsilon_1^{l_1+2} & \epsilon_1^{l_2+1} & \epsilon_1^{l_3} \\ \epsilon_2^{l_1+2} & \epsilon_2^{l_2+1} & \epsilon_2^{l_3} \\ \epsilon_3^{l_1+2} & \epsilon_3^{l_2+1} & \epsilon_3^{l_3} \end{vmatrix} \div \begin{vmatrix} \epsilon_1^2 & \epsilon_1 & 1 \\ \epsilon_2^2 & \epsilon_2 & 1 \\ \epsilon_3^2 & \epsilon_3 & 1 \end{vmatrix} \tag{3.2}$$

for general $[l]_3$ by setting $l_3 = 0$. A simple manipulation on the rows of the determinant in the numerator casts the right side of (3.2), with $l_3 = 0$, into the form

[10] L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963). See also L. C. Biedenharn, "Group Theoretical Approaches to Nuclear Spectroscopy" in *Lectures in Theoretical Physics*, edited by W. E. Brittin *et al.* (Interscience Publishers, Inc., New York, 1963), Vol. 5.

[11] J. E. Wess, Nuovo Cimento **15**, 52 (1960); M. Ikeda, S. Ogawa, and Y. Ohnuki, Progr. Theoret. Phys. (Kyoto) **22**, 715 (1959); S. Okubo, Progr. Theoret. Phys. (Kyoto) **27**, 949 (1962); A. J. Macfarlane, E. C. G. Sudarshan, and C. Dullemond, Nuovo Cimento **30**, 845 (1963).

[12] L. C. Biedenharn, Phys. Letters **3**, 254 (1963) and lectures cited in Footnote 10; D. L. Pursey, Proc. Roy. Soc. (London) **A275**, 284 (1963); N. Mukunda, (private communication).

[13] H. Weyl, *Theory of Groups and Quantum Mechanics* (Dover Publications, Inc., New York, 1931), p. 381.

$$(\epsilon_1 - \epsilon_2)^{-1} \sum_{r=0}^{l_1+1} \sum_{s=0}^{l_2} (\epsilon_1^r \epsilon_2^s - \epsilon_1^s \epsilon_2^r) \epsilon_3^{l_1+l_2-1-r-s}, \tag{3.3}$$

formulas of the type

$$a^{m+1} - b^{m+1} = (a - b) \sum_{i=1}^{m} a^i b^{m-i}$$

having been used to cancel the factors $(\epsilon_1 - \epsilon_3)$ and $(\epsilon_2 - \epsilon_3)$ of the determinant in the denominator. We can replace the summation in (3.3) by a summation $\sum_{r=l_2+1}^{l+1} \sum_{s=0}^{l_2}$ since the part $\sum_{r=0}^{l_2} \sum_{s=0}^{l_2}$ of the summation in (3.3) is a symmetric sum over an antisymmetric summand and thus vanishes. Then setting $m_1 = r_1 - 1$ and $m_2 = s$ we get

$$\chi([l_1, l_2, 0], \epsilon_1, \epsilon_2, \epsilon_3) = \sum_{m_1=l_2}^{l_1} \sum_{m_2=0}^{l_2} \chi([m_1, m_2], \epsilon_1, \epsilon_2) \epsilon_3^{l_1+l_2-m_1-m_2}, \tag{3.4}$$

where

$$\chi([m_1, m_2], \epsilon_1, \epsilon_2) = \begin{vmatrix} \epsilon_1^{m_1+1} & \epsilon_1^{m_2} \\ \epsilon_2^{m_1+1} & \epsilon_2^{m_2} \end{vmatrix} \div \begin{vmatrix} \epsilon_1 & 1 \\ \epsilon_2 & 1 \end{vmatrix} \tag{3.5}$$

is the character of the IR $[m_1, m_2]$ of $U_2$ corresponding to an element of $U_2$ with eigenvalues $\epsilon_1$ and $\epsilon_2$. We use (3.4) only for elements of $SU_3$ and write

$$\chi(\{l_1, l_2\}, \epsilon_1, \epsilon_2, \epsilon_3),$$

the restriction (3.1) being implied by the fact that the IR argument of $\chi$ is an IR of $SU_3$.

We now go on to consider the restriction of the element of $SU_3$ to an element of the subgroup $SU_2 \otimes U_1^{(2)}$ of $SU_3$. In physical language we restrict attention from the general $SU_3$-symmetry transformation to the product of isospin and hypercharge gauge transformations. Isospin and hypercharge transformations are described by matrices which when diagonalized have the respective forms

$$\begin{pmatrix} \epsilon^{\frac{1}{2}} & 0 & 0 \\ 0 & \epsilon^{-\frac{1}{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta^{\frac{1}{2}} & 0 & 0 \\ 0 & \eta^{\frac{1}{2}} & 0 \\ 0 & 0 & \eta^{-\frac{2}{3}} \end{pmatrix} \tag{3.6}$$

with $|\epsilon| = |\eta| = 1$. The corresponding product matrix is

$$\begin{pmatrix} \epsilon^{\frac{1}{2}}\eta^{\frac{1}{2}} & 0 & 0 \\ 0 & \epsilon^{-\frac{1}{2}}\eta^{\frac{1}{2}} & 0 \\ 0 & 0 & \eta^{-\frac{2}{3}} \end{pmatrix}, \tag{3.7}$$

so that we must examine the element of $SU_3$ with eigenvalues

$$\epsilon_1 = \epsilon^{\frac{1}{2}}\eta^{\frac{1}{2}}, \qquad \epsilon_2 = \epsilon^{-\frac{1}{2}}\eta^{\frac{1}{2}}, \qquad \epsilon_3 = \eta^{-\frac{2}{3}}, \tag{3.8}$$

restriction (3.1) being automatically satisfied. We thus obtain the character of the IR $\{l_1, l_2\}$ of $SU_3$, corresponding to the element of $SU_3$ specified by (3.8), as a compound character of $SU_2 \otimes U_1^{(2)}$, and hence discover that IR's of $SU_2 \otimes U_1^{(2)}$ are contained in the IR $\{l_1, l_2\}$ of $SU_3$. Equations (3.4)–(3.8) give

$$\chi(\{l_1, l_2\}, \epsilon^{\frac{1}{2}}\eta^{\frac{1}{2}}, \epsilon^{-\frac{1}{2}}\eta^{\frac{1}{2}}, \eta^{-\frac{2}{3}}) = \sum_{m_1=l_2}^{l_1} \sum_{m_2=0}^{l_2} \chi(\{m_1 - m_2\}, \epsilon^{\frac{1}{2}}, \epsilon^{-\frac{1}{2}}) \eta^{m_1+m_2-\frac{2}{3}(l_1+l_2)}.$$

We conclude that the IR $\{l_1, l_2\}$ of $SU_3$ contains within it the IR $\{m_1 - m_2\} \otimes Y^{(2)}$ of $SU_2 \otimes U_1^{(2)}$, where

$$y^{(2)} = m_1 + m_2 - \tfrac{2}{3}(l_1 + l_2) \tag{3.9}$$

once and only once for each pair of integers $m_1$ and $m_2$ allowed by the inequalities

$$l_1 \geq m_1 \geq l_2 \geq m_2 \geq 0. \tag{3.10}$$

This is to be recognized as a nontrivial refinement of a result given by Weyl.[14]

Since in particle physics, it is customary to use the highest weight notation for IR's of $SU_3$, we use Eq. (2.5) to translate this last result into the statement that the IR $(\lambda_1, \lambda_2)$ of $SU_3$ contains within it an $(I, Y)$ multiplet with $I$ and $Y$ values given by

$$2I = m_1 - m_2,$$

$$Y = m_1 + m_2 - \tfrac{2}{3}(\lambda_1 + 2\lambda_2) \tag{3.11}$$

for each pair of values of $m_1$ and $m_2$ allowed by

$$\lambda_1 + \lambda_2 \geq m_1 \geq \lambda_2 \geq m_2 \geq 0. \tag{3.12}$$

We note that Eqs. (3.11) and (3.12) exactly reproduce the result stated without proof as Eqs. (2.8)–(2.10) of the last-named paper of Footnote 11.

The derivation of a reduction theorem for $SU_{n+1}$ analogous to that contained in Eqs. (3.9) and (3.10) for $SU_3$ proceeds along the same lines and we therefore confine ourselves to a statement of the result.[15]

The IR $\{l\}_n$ of $SU_{n+1}$ contains within it the IR $\{m_1 - m_n, \cdots, m_{n-1} - m_n\} \otimes y^{(n)}$ of $SU_n \otimes U_1^{(n)}$, where $y^{(n)}$ is given by

$$y^{(n)} = m - nl/(n+1),$$

$$l = \sum_{i=1}^{n} l_i, \qquad m = \sum_{i=1}^{n} m_i$$

[14] H. Weyl, Ref. 13, p. 391.

[15] Such a result has been obtained in a different manner by I. M. Gelfand and M. L. Cetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950). See also G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963). We thank the referee for having drawn our attention to these sources.

once and only once for each distinct (ordered) set of integers $m_1, m_2 \cdots m_n$ allowed by the inequalities

$$l_1 \geq m_1 \geq l_2 \cdots \geq l_n \geq m_n \geq 0.$$

This also is to be recognized as a nontrivial generalization of a result given by Weyl.[13]

We can now show how the occurrence of fractional eigenvalues of $Y_1^{(n)}$ in any IR $\{l\}_n$ of $SU_{n+1}$ is related to the plurality type $p\{l\}_n$ of the representation. In the case of $SU_3$, the values of $Y^{(2)} = Y$ that occur in $\{l_1, l_2\}$ follow from (3.9) and (3.10) and are

$$\tfrac{1}{3}(l_1 + l_2) - l_1, \quad \tfrac{1}{3}(l_1 + l_2) - l_1 + 1 \cdots \tfrac{1}{3}(l_1 + l_2) - 1, \quad \tfrac{1}{3}(l_1 + l_2).$$

Herein the spacing in unity and each value $y^{(2)}$ of $Y^{(2)}$ is of the form

$$y^{(2)} = \tfrac{1}{3}p\{l\}_2 + h \tag{3.13}$$

for some integer $h$. Thus, according to whether the triality type of an IR of $SU_3$ is 0, 1, 2, its allowed hypercharge eigenvalues are of the form "integer", "integer plus one-third", "integer plus two-thirds". This can be summarized in the operator relationship

$$P\{l\}_2 = \exp[2\pi i Y^{(2)}],$$

which follows from Eqs. (2.9) and (3.13). Similarly, we see that the allowed values of $Y^{(n)}$ in the IR $\{l\}_n$ of $SU_{n+1}$ are

$$l/(n+1) - l_1, \quad l/(n+1) - l_1 + 1, \cdots, l/(n+1) - 1, \quad l/(n+1),$$

where $l = \sum_{i=1}^{n} l_i$. As expected, the spacing is unity and each $y^{(n)}$ is of the form

$$y^{(n)} = p\{l\}_n/(n+1) + h \tag{3.14}$$

for some integer $h$. The equivalent operator relationship is

$$P\{l\}_n = \exp[2\pi i Y^{(n)}]. \tag{3.15}$$

Either Eq. (3.14) or Eq. (3.15) may be regarded as containing the desired relationship between the plurality type of the IR of $SU_{n+1}$ and the occurrence in the representation of fractional eigenvalues of $Y_1^{(n)}$.

Finally, we show that the successive definitions of $Q/e$, Eqs. (1.2), (1.3), etc., each lead to integral eigenvalues in their appropriate contexts. It is sufficient to consider Eq. (1.2) within the $SU_4$-symmetry theory which uses only IR's $\{l\}_3$ of $SU_4$ with $k \equiv p\{l\}_3 = 0$. Consider any representation $\{m_1 - m_3, m_2 - m_3\} \otimes y^{(3)}$ of $SU_3 \otimes U_1^{(3)}$, where $y^{(3)} = m_1 + m_2 + m_3 - \frac{3}{4}(l_1 + l_2 + l_3)$, contained within such an IR $\{l\}_3$ of $SU_4$. We note that $\frac{1}{3}y^{(3)}$ is of the form $\frac{1}{3}t$ plus an integer where $t = m_1 + m_2 - 2m_3 = m_1 + m_2 + m_3$, modulo 3, is the triality of $\{m_1 - m_3, m_2 - m_3\}$. Also from (3.11) it follows that all the values of $(I_z + \frac{1}{2}Y)$ contained in $\{m_1 - m_3, m_2 - m_3\}$ are of the form integer minus $\frac{1}{3}t$, so that the same is true for all values of $(I_z + \frac{1}{2}Y)$. Hence we see that $Q/e$, as given by Eq. (1.3), has integral eigenvalues.

# Solution of a Singular Integral Equation from Scattering Theory*

R. P. KENSCHAFT† AND R. D. AMADO
*Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania*
(Received 11 February 1964)

The solution of a singular integral equation of the form

$$\psi(\omega, \omega_0) = -\frac{1}{\omega} - \frac{1}{\pi}\int_\mu^\infty \frac{\rho(\omega_1, \omega_0)\psi(\omega_1, \omega_0)\, d\omega_1}{\omega_1 + \omega - \omega_0 - i\epsilon}$$

is obtained. The equation first appeared in the attempt of Källén and Pauli to solve $V - \theta$ scattering in the Lee model. The methods developed by one of us to solve that problem are turned to solving the integral equation itself.

## I. INTRODUCTION

IN discussing the Lee model,[1] Källén and Pauli[2] encountered an integral equation for the $V - \theta$ scattering states which they were unable to solve; this equation was not of the usual singular type often found in particle physics.[3,4] Since then, one of us has solved $V - \theta$ scattering in the Lee model[5] by a different approach, which is extended in the present paper to yield a solution to the integral equation of KP.

In obtaining the solution, we use physical arguments (such as the relation of the Lippmann–Schwinger equation to the Schrödinger equation) as a guide, but the result, checked by direct substitution, is quite rigorous. It is not however the most general solution. Rather than obscure the origin of our arguments we rely heavily on the previous work in the Lee model[5]; we do not wish to imply by this that our method or results are restricted to that model.

In Sec. II we present the integral equation and our solution. In Sec. III, using a method developed by one of us,[6] we derive an integral equation for the scattering amplitude, simply related to the equation of KP; we then derive an expression for the scattering amplitude by the methods of VCLM. In Sec. IV we show by direct substitution the validity of our solution.

## II

The integral equation encountered by KP for the $V - \theta$ scattering states in the Lee model can be written in the form[7]

$$\psi(\omega, \omega_0) = -\frac{1}{\omega} - \frac{1}{\pi}\int_\mu^\infty \frac{\operatorname{Im} h(\omega')\, d\omega'}{\omega' + \omega - \omega_0 - i\epsilon}\frac{\psi(\omega', \omega_0)}{h(\omega_0 - \omega')}, \tag{1}$$

where we define

$$h(\omega) = \omega + \frac{\omega^2}{\pi}\int_\mu^\infty \frac{U(\omega_1)\, d\omega_1}{\omega_1(\omega_1 - \omega)}. \tag{2}$$

$U(\omega_1)$ is an arbitrary real function bounded on the interval of integration and vanishing at both ends thereof. $U$ must be further restricted so that the only zero of $h(\omega)$ occurs at $\omega = 0$, and so that the quantity $Z$, defined by

$$Z = 1 - \frac{1}{\pi}\int_\mu^\infty \frac{U(\omega_1)\, d\omega_1}{\omega_1} = \lim_{\omega\to\infty}\frac{h(\omega)}{\omega},$$

exists and satisfies $0 < Z < 1$. The lower limit $\mu$ is a positive real number. We find that $h(\omega)$ is analytic in the complex plane cut along the line $\mu \leq \omega < \infty$, with

$$\operatorname{Im} h(\omega + i0) \equiv \operatorname{Im} h(\omega) = \omega U(\omega),$$

and that $h(\omega)$ is real for $\omega \leq \mu$.[8]

Equation (1) defines a function $\psi$ of $\omega$ with a pole at $\omega = 0$ and a branch cut on the line $-\infty < \omega \leq \omega_0 - \mu$, with the discontinuity across the

* Supported in part by the National Science Foundation.
† National Science Foundation Cooperative Fellow.
[1] T. D. Lee, Phys. Rev. 95, 1329 (1954).
[2] G. Källén and W. Pauli, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. 30, No. 7 (1955). This paper will be referred to as KP.
[3] R. Omnes, Nuovo Cimento 8, 316 (1958).
[4] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff, Ltd., Groningen, The Netherland, 1953).
[5] R. D. Amado, Phys. Rev. 122, 696 (1961), referred to as VCLM.
[6] R. D. Amado, Phys. Rev. 132, 485 (1963).
[7] In the notation of KP, we have defined $\psi$ by

$$\phi_1(\mathbf{k}, \mathbf{k}_0) = \delta_{\mathbf{k}\mathbf{k}_0} + \frac{g^2}{2V}\frac{f(\omega)f(\omega_0)}{(\omega\omega_0)^{\frac{1}{2}}}\frac{\psi(\omega, \omega_0)}{h(\omega_0 - \omega)}.$$

An over-all correction of sign in Eq. (57) of KP has been included in (1).
[8] In the notation of VCLM, $U(\omega)$ is $g^2u^2(\omega)(\omega^2 - \mu^2)^{1/2}/4\pi\omega$ and $h(\omega)$ is $\omega[1 - \beta(\omega)]$.

branch cut at the point $\omega$ related to $\psi(\omega_0 - \omega, \omega_0)$. This equation therefore differs from the standard singular equation encountered in particle physics[3] and elasticity theory[4] in that the discontinuity of the function at a point is not related to the function at that point. For that reason KP were unable to solve the equation. In this paper we present a particular solution of Eq. (1) obtained by a means different from that of Omnes and Muskhelishvili. That solution is given by

$$\psi(\omega, \omega_0) = \frac{h(\omega_0 - \omega)}{\omega - \omega_0}\left[\frac{\omega_0}{\omega h(\omega_0)} + \frac{2A(\omega, \omega_0)}{1 - h(\omega_0)A(\omega_0, \omega_0)}\right], \tag{3}$$

in which we define

$$A(\omega, \omega_0) = \frac{1}{\pi}\int_{\mu}^{\infty} \frac{\omega_1\, d\omega_1}{(\omega_1 + \omega - \omega_0 - i\epsilon)h(\omega_0 - \omega_1)} \times \operatorname{Im}\frac{1}{h(\omega_1)}. \tag{4}$$

In the complex $\omega$ plane, $A$ is analytic except for a left-hand cut $-\infty < \omega \leq \omega_0 - \mu$; the discontinuity across the cut is purely imaginary if and only if $\omega_0 \leq 2\mu$.[9] The analytic properties possessed by $\psi(\omega, \omega_0)$ of Eq. (3) are in agreement with the analytic properties ascribed by the integral equation (1) to $\psi$, and we show in Sec. IV by direct substitution that it satisfies the equation.

## III

The problem in obtaining a solution to (1) is to relate the equation to the solution for the $V - \theta$ scattering amplitude obtained by a different method.[5] Since (1) was encountered by KP in solving for the state vector—that is, in solving the integral form of the "Schrödinger equation"—we attempt to find a corresponding "Lippmann–Schwinger equation" for the $t$ matrix. This can be derived by methods recently presented by one of us.[6] Adapting these methods to the Hamiltonian for the $VN\theta$ system given in VCLM, we may derive an integral equation for the $V - \theta$ scattering amplitude of the form

$$t(\omega, \omega_0) = b(\omega, \omega_0) + \sum_{k'} b(\omega, \omega')G(\omega')t(\omega', \omega_0), \tag{5}$$

where the Born term is given, in the notation of VCLM, by

$$b(\omega, \omega_0) = \frac{g^2}{2\Omega}\frac{u(\omega)u(\omega_0)}{(\omega\omega_0)^{\frac{1}{2}}}\frac{1}{E - m - \omega - \omega_0}, \tag{6}$$

and the $V$ propagator by

$$G(\omega') = 1/h(E - m - \omega'). \tag{7}$$

Letting $E = m + \omega_0 + i\epsilon$, defining

$$\psi(\omega, \omega_0) = \frac{2\Omega}{g^2}\frac{(\omega\omega_0)^{\frac{1}{2}}}{u(\omega)u(\omega_0)}\, t(\omega, \omega_0), \tag{8}$$

and writing the summation as an integral, we find that we have generated Eq. (1).

The dispersion-theoretic methods utilized in VCLM lead to an expression for the transition amplitude off the energy shell, which should be simply related to $t$ of (5). Contracting first the $V$ particle from the left, and then the $\theta$ from the left in $\langle V\theta_{\omega}^{(-)} \mid V\theta_{\omega_0}^{(+)}\rangle$, we obtain, in the notation of that paper, an expression analogous to $T(\omega)$ of VCLM:

$$T(\omega, \omega_0) = \sum_{S}\frac{(2\Omega\omega_0)^{\frac{1}{2}}}{u(\omega_0)}\frac{\langle 0 \mid f \mid S\rangle\langle S \mid j \mid V\theta_{\omega_0}^{(+)}\rangle}{S + \omega - \omega_0 - m - i\epsilon}. \tag{9}$$

Using the definitions of $K$ and $F$ as given in VCLM, we obtain

$$T(\omega, \omega_0) = \sum_{k'}\frac{u^2(\omega')}{\omega'}\frac{K(\omega')F(\omega', \omega_0)}{\omega' + \omega - \omega_0 - i\epsilon}. \tag{10}$$

We note that $T(\omega_0, \omega_0)$ is equal to $T(\omega_0)$, giving the correct result on the energy shell.

Results derived or quoted in VCLM lead to an expression for $T$ in the form

$$\begin{aligned} T(\omega, \omega_0) &= (g^2/\omega)[\omega_0/h(\omega_0)] + g^2N(\omega_0)A(\omega, \omega_0) \\ &\equiv T_1 + T_2, \end{aligned} \tag{11}$$

where $A(\omega, \omega_0)$ is given by Eq. (4) and $N(\omega_0)$ by

$$N(\omega_0) = 1 + g^{-2}h(\omega_0)T(\omega_0, \omega_0) \tag{12}$$

$$= 2[1 - h(\omega_0)A(\omega_0, \omega_0)]^{-1}. \tag{13}$$

The last expression results from Eqs. (11) and (12).

As a function of the complex variable $\omega$, $T(\omega, \omega_0)$ has a simple pole at $\omega = 0$ and a branch cut along the line $-\infty < \omega \leq \omega_0 - \mu$; the integral equation (1) assigns the same analytic properties to $\psi(\omega, \omega_0)$. Attempting to match residues at the pole, and recognizing that $1 - \beta$ represents[8] the effects of taking the $V$ particle off the energy shell, we try a solution of the form

$$\psi(\omega, \omega_0) = (1/g^2)[h(\omega_0 - \omega)/(\omega - \omega_0)]T(\omega, \omega_0), \tag{14}$$

giving the result quoted in (3).

## IV

The validity of the *ansatz* (14) is proved by direct substitution in the integral equation. Defining $\phi_1$

[9] This condition corresponds to no production of $\theta$ particles.

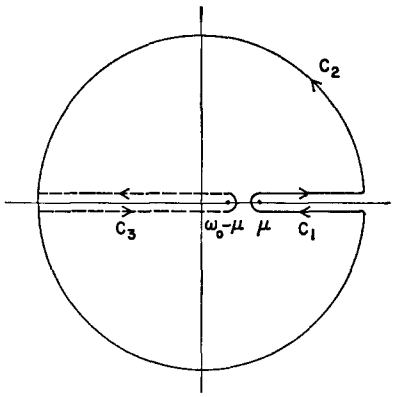


FIG. 1. Contours of integration.

and $N(\omega_0)\phi_2$ as the terms generated, respectively, by $T_1$ and $T_2$ of (11) upon substitution in the integral of (1), we seek to show that

$$\psi(\omega, \omega_0) = \phi_1 + N\phi_2 - \omega^{-1}.$$

To calculate $\phi_1$, given by

$$\phi_1 = \frac{-\omega_0}{h(\omega_0)} \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} h(\omega')\, d\omega'}{\omega'(\omega' + \omega - \omega_0 - i\epsilon)(\omega' - \omega_0 - i\epsilon)},$$

we use the property $h(\omega - i0) = h^*(\omega + i0)$ to write

$$\phi_1 = \frac{-\omega_0}{h(\omega_0)} \frac{1}{2\pi i} \int_{C_1} \frac{h(\omega')\, d\omega'}{\omega'(\omega' + \omega - \omega_0 - i\epsilon)(\omega' - \omega_0 - i\epsilon)},$$

where the contour $C_1$ comes from $\infty$ to $\mu$ just below the real axis and returns to $\infty$ just above, as in Fig. 1. Closing the contour by a large circle $C_2$, for which the contribution to the integral vanishes, we find by the calculus of residues that

$$\phi_1 = \frac{1}{\omega} \frac{h(\omega_0 - \omega)}{\omega_0 - \omega} \frac{\omega_0}{h(\omega_0)} - \frac{1}{\omega}. \tag{15}$$

To calculate $\phi_2$, given by

$$\phi_2 = \frac{1}{\pi} \int_\mu^\infty \frac{\operatorname{Im} h(\omega_1)\, d\omega_1}{(\omega_1 + \omega - \omega_0 - i\epsilon)(\omega_1 - \omega_0 - i\epsilon)} \times \frac{1}{\pi} \int_\mu^\infty \frac{\omega'\, d\omega'}{(\omega' + \omega_1 - \omega_0 - i\epsilon)h(\omega_0 - \omega')} \operatorname{Im} \frac{-1}{h(\omega')},$$

we can justifiably reverse the order of integration. Denoting by $I$ the result of the integration over $\omega_1$, we obtain, by the above method,

$$I = \frac{h(\omega_0 - \omega')}{\omega'(\omega' - \omega)} + \frac{h(\omega_0)}{\omega\omega'} - \frac{h(\omega_0 - \omega)}{\omega(\omega' - \omega)} \equiv I_A + I_B + I_C.$$

We then define $\phi_2 = \phi_{2A} + \phi_{2B} + \phi_{2C}$ in a corresponding manner, finding immediately that

$$\phi_{2A} = \omega^{-1} - [h(\omega)]^{-1} \tag{16}$$

and

$$\phi_{2B} = -h(\omega_0)A(\omega_0, \omega_0)\omega^{-1}. \tag{17}$$

To obtain a recognizable expression for $\phi_{2C}$, given by

$$\phi_{2C} = \frac{h(\omega_0 - \omega)}{\omega\pi} \int_\mu^\infty \frac{\omega'\, d\omega'}{(\omega' - \omega)h(\omega_0 - \omega')} \operatorname{Im} \frac{1}{h(\omega')},$$

we again attempt to form a closed contour of integration. Since the integral now has a left-hand branch cut from $\omega_0 - \mu$ to $-\infty$, we must introduce the contour $C_3$ of Fig. 1, obtaining

$$\phi_{2C} = \frac{h(\omega_0 - \omega)}{\omega 2\pi i} \oint_{C_1 + C_3 - C_3} \frac{\omega'\, d\omega'}{(\omega' - \omega)h(\omega_0 - \omega')h(\omega')} + \frac{h(\omega_0 - \omega)}{\omega 2\pi i} \int_{C_3} \frac{\omega'\, d\omega'}{(\omega' - \omega)h(\omega_0 - \omega')h(\omega')}.$$

The first integral may now be evaluated by residues, yielding

$$\phi_{2C} = \frac{1}{h(\omega)} - \frac{1}{\omega} \frac{h(\omega_0 - \omega)}{\omega_0 - \omega} \frac{\omega_0}{h(\omega_0)} + \phi_{C3}, \tag{18}$$

where $\phi_{C3}$ is the integral over $C_3$ that, after reconverting to a line integral and changing the variable of integration, reduces to

$$\phi_{C3} = \frac{h(\omega_0 - \omega)}{\omega\pi} \int_\mu^\infty \frac{(\omega' - \omega_0)\, d\omega'}{(\omega' + \omega - \omega_0 - i\epsilon)h(\omega_0 - \omega')} \times \operatorname{Im} \frac{1}{h(\omega')} \tag{19}$$

$$= \frac{h(\omega_0 - \omega)}{\omega_0 - \omega} \left[ \frac{\omega_0}{\omega} A(\omega_0, \omega_0) - A(\omega, \omega_0) \right].$$

Collecting terms, and using (13), we find that

$$\phi_1 + N(\omega_0)\phi_2 - \omega^{-1} = \psi,$$

as desired.

Thus we have shown that (3) satisfies the integral equation (1); it is, of course, a particular solution of (1). Arbitrary amounts of the solution of the homogeneous equation corresponding to (1) can be added; however, we do not know the form of these homogeneous solutions. Nevertheless, it seems clear that (3) is the particular solution appropriate to a unitary $V - \theta$ scattering amplitude, for reasons analogous to those given in VCLM.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 9 SEPTEMBER 1964

# The Clebsch–Gordan Series for $SU(3)$*

SIDNEY COLEMAN
*Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts*
(Received 31 March 1964)

A method is explained for the calculation of the reduction of direct products of irreducible representations of $SU(3)$. The method is believed to be simpler and more efficient than the usual algorithms.

THIS note explains a method for the calculation of the reduction of direct products of irreducible representations of $SU(3)$. This method is both simpler in its foundations and more efficient in its application than the usual methods, based on Young tableaux or weight diagrams. No attempt has been made to search the mathematical literature, but I believe the method to be new.

The method proceeds in two steps: First we decompose the direct product of two irreducible representations into a direct sum of certain special reducible representations, which will be defined below. Then we decompose the special reducible representations into a direct sum of irreducible representations.

We shall use the familiar characterization[1] of irreducible representations of $SU(3)$ as the transformations induced on irreducible tensorial sets by unitary unimodular transformations of a three-dimensional complex vector space. In particular, we shall denote by $D^{(n,m)}$—or, for brevity, simply by $(n, m)$—that irreducible representation which has as its basis the set of all tensors with $n$ upper indices and $m$ lower indices, that are completely symmetric among the upper indices, completely symmetric among the lower indices, and traceless (the contraction of any upper index with any lower index gives zero). The dimension of $(n, m)$ is the number of linearly independent tensors of the specified type. It is a simple exercise in combinatorics to calculate that

$$\dim (n, m) = \tfrac{1}{2}(n + 1)(m + 1)(n + m + 2). \quad (1)$$

We shall also need certain special reducible representations, which we shall denote by $D^{(n,n';m,m')}$—or, for brevity, simply by $(n, n'; m, m')$. The representation $(n, n'; m, m')$ is defined as that representation which has for its basis the set of all tensors with $n + n'$ upper indices and $m + m'$ lower indices, that are completely symmetric among the first $n$ upper indices, completely symmetric among the last $n'$ upper indices, completely symmetric among the first $m$ lower indices, completely symmetric among the last $m'$ lower indices, and traceless. Roughly speaking, $(n, n'; m, m')$ may be thought of as the direct product of $(n, m)$ and $(n', m')$ with all traces removed, but without any symmetrization.

It is a simple matter to decompose the direct product of irreducible representations into our special reducible representations. We merely separate out all tensors that can be obtained by contracting, in all possible ways, indices from the set of $n$ with indices from the set of $m'$, and indices from the set of $n'$ with indices from the set of $m$. That is to say,

$$\begin{aligned}(n, m) \otimes (n', m') &= (n, n'; m, m') \oplus (n - 1, n'; m, m' - 1) \\ &\quad \oplus (n, n' - 1; m - 1, m') \\ &\quad \oplus (n - 1, n' - 1; m - 1, m' - 1) \oplus \cdots.\end{aligned}$$

The process terminates whenever we run out of indices to contract; that is, whenever a zero appears in the series on the right. In more compact form,

$$(n, m) \otimes (n', m') = \sum_{i=0}^{\min(n,m')} \sum_{j=0}^{\min(n',m)} (n - i, m - j; n' - j, m' - i), \quad (2)$$

where the summation sign indicates a direct sum.

We now wish to decompose one of our special reducible representations into direct sums of irreducible representations. In the language of tensors, we want to decompose an arbitrary tensor from the basis of $(n, m; n', m')$ into a sum of linear combinations of completely symmetric traceless tensors. Let us begin with the upper indices. Let

$$T^{i_1 \cdots i_n i_{n+1} \cdots i_{n+n'}}_{j_1 \cdots j_m j_{m+1} \cdots j_{m+m'}},$$

* Work supported in part by the Air Force Office of Scientific Research, under Contract No. AF 49(638)589.

[1] See, for example, R. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. 34, 1 (1962); J. J. de Swart, Rev. Mod. Phys. 35, 916 (1963). These articles contain extensive references to the earlier literature.

be an arbitrary tensor of the type under discussion. Let us choose a pair of upper indices; with no loss of generality they may be $i_1$ and $i_{n+1}$. We may write the tensor as the sum of two tensors, one of which is symmetric under interchange of these indices, and the other of which is antisymmetric. Using the $\epsilon$ tensor, we may write the antisymmetric part in terms of a tensor of lower rank,

$$S_{ki_1\cdots i_m i_{m+1}\cdots j_{m+m'}}^{i_2\cdots i_n i_{n+2}\cdots i_{n+n'}} = \epsilon_{ki_1 i_{n+1}} T_{i_1\cdots j_{m+m'}}^{i_1\cdots i_{n+n'}}.$$

The surprising fact, which enormously simplifies the whole reduction, is that this tensor is *already* completely symmetric in its lower indices.

*Proof:* For example, let us take the indices $j_1$ and $j_{m+1}$. We prove the tensor is symmetric under interchange of these indices by showing that their contraction with the $\epsilon$-tensor vanishes,

$$\begin{aligned}\epsilon^{ri_1 i_{m+1}} S_{k\cdots j_{m+m'}}^{i_2\cdots i_{n+n'}} &= \epsilon^{ri_1 i_{m+1}} \epsilon_{ki_1 i_{n+1}} T_{i_1\cdots j_{m+m'}}^{i_1\cdots i_{n+n'}} \\ &= (\delta_k^r \delta_{i_1}^{j_1} \delta_{i_{n+1}}^{j_{m+1}} - \delta_k^{j_1} \delta_{i_1}^r \delta_{i_{n+1}}^{j_{m+1}} + \text{cyclic perms.}) \\ &\quad \times T_{j_1\cdots j_{m+m'}}^{i_1\cdots i_{n+n'}}.\end{aligned}$$

But, by the tracelessness of $T$, the right-hand side of this equation is zero. Similar arguments work for any pair of indices. Q.E.D.

Thus, the symmetrization is very simple. We may remove pairs of upper indices, adding a lower index whenever we do so; or, alternatively, we may remove pairs of lower indices, adding an upper index whenever we do so—but we can never remove both a pair of upper indices and a pair of lower indices, for once we have removed a pair of upper (lower) indices, the tensor is already completely symmetric in its lower (upper) indices. The process terminates when we run out of indices. Returning from the basis space to the representation, we may write the decomposition in compact form:

$$\begin{aligned}(n, n'; m, m') &= (n + n', m + m') \\ &\quad \oplus \sum_{i=1}^{\min(n,n')} (n + n' - 2i, m + m' + i) \\ &\quad \oplus \sum_{j=1}^{\min(m,m')} (n + n' + j, m + m' - 2j), \qquad (3)\end{aligned}$$

where the summation sign again represents the direct sum.

To demonstrate the efficiency of this method, we conclude with two examples. All arithmetic is shown.

*Example 1:* $(1, 1) \otimes (1, 1)$
By Eq. (2),

$$\begin{aligned}(1, 1) \otimes (1, 1) &= (1, 1; 1, 1) \oplus (1, 0; 0, 1) \\ &\quad \oplus (0, 1; 1, 0) \oplus (0, 0; 0, 0).\end{aligned}$$

By Eq. (3),

$$(1, 1; 1, 1) = (2, 2) \oplus (0, 3) \oplus (3, 0),$$

$$(1, 0; 0, 1) = (1, 1),$$

$$(0, 1; 1, 0) = (1, 1),$$

and

$$(0, 0; 0, 0) = (0, 0).$$

The desired decomposition is the sum of all the terms on the right. If we use Eq. (1) to write this in terms of the notation in which representations are labeled by their dimensions, we find the familiar result

$$8 \otimes 8 = 27 \oplus 10 \oplus \overline{10} \oplus 8 \oplus 8 \oplus 1.$$

*Example 2:* $(2, 2) \otimes (3, 0)$
By Eq. (2),

$$\begin{aligned}(2, 2) \otimes (3, 0) &= (2, 3; 2, 0) \oplus (2, 2; 1, 0) \oplus (2, 1; 0, 0).\end{aligned}$$

By Eq. (3),

$$(2, 3; 2, 0) = (5, 2) \oplus (3, 3) \oplus (1, 4),$$

$$(2, 2; 1, 0) = (4, 1) \oplus (2, 2) \oplus (0, 3),$$

and

$$(2, 1; 0, 0) = (3, 0) \oplus (1, 1).$$

In the alternative notation,

$$\begin{aligned}27 \otimes 10 &= 81 \oplus 64 \oplus 35 \oplus \overline{35} \oplus 27 \oplus \overline{10} \\ &\quad \oplus 10 \oplus 8.\end{aligned}$$