JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 6, NUMBER 12 DECEMBER 1965

# On the Representations of the Semisimple Lie Groups. V. Some Explicit Wigner Operators for $SU_3$*

G. E. BAIRD
*Northrop Space Laboratories, Huntsville, Alabama*
AND
L. C. BIEDENHARN
*Department of Physics, Duke University, Durham, North Carolina*
(Received 1 February 1965)

A general method for the calculation of the [21 ··· 10] Wigner operators is presented and used to obtain specific results for the $SU_3$ group. General expressions for the $SU_3$ reduced Wigner operators are given for tensor operators transforming like the representations [100], [110], and [210].

## I. INTRODUCTION AND SUMMARY

THE present series of papers[1] is concerned with the explicit constructive determination of the representations of the semisimple Lie groups by an extension of the Racah–Wigner angular momentum calculus which was developed for $SU_2$. As discussed earlier, the program to be followed has been laid out in detail by the work of Wigner[2] and Racah[3] and consists of essentially three problems: (a) the determination of invariant operators ("Casimir invariants") that uniquely specify the irreducible representations, (b) the determination of sufficient "labeling operators" to uniquely specify the states of an irreducible representation, and (c) the determination of explicit Wigner coefficients by solution of the problem of simple reducibility.[4,5] In I a more complete discussion of these problems is given and solution for general Casimir operators ($I_n$) for the unitary groups is constructed.[6] II contains a solution of the labeling for the unitary groups—based on Weyl's branching theorem—which is used there to explicitly determine the matrices of the generators for the unitary groups, $SU_n$. In III the analog of the $SU_2$ 1-$j$ symbol is defined for $SU_n$ and an appropriate generalization to $SU_n$ of the Condon–Shortley phase convention is given. In IV it is shown that the multiplicity of the general $SU_n$ tensor operator may be put into a one-to-one correspondence with the multiplicity structure of the corresponding states, and a definition for the reduced Wigner operators is introduced.

It is the purpose of the present paper to derive detailed results for the $SU_3$ [210] Wigner operators (octet operator matrix elements) which couple an

* Supported in part by the U. S. Army Research Office (Durham) and the National Science Foundation.

[1] L. C. Biedenharn, J. Math. Phys. 4, 436 (1963); G. E. Baird and L. C. Biedenharn, *ibid.* 4, 1449 (1963); 5, 1723 (1964); 5, 1730 (1964); these are referred to as I, II, III, and IV, respectively.

[2] E. P. Wigner, lecture notes, Princeton University, Princeton, New Jersey, 1955 (unpublished); "On the Matrices which Reduce the Kronecker Product of Representations of Simply Reducible Groups," in *Selected Papers on the Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic Press Inc., New York, 1965).

[3] G. Racah, Ergeb. d. exacten Naturw. 37, 28 (1965).

[4] E. P. Wigner, Am. J. Math. 63, 57 (1941).

[5] Actually problems (b) and (c) are not distinct, for in both the problem is to uniquely specify the states of the irreducible representations of a sub-group contained in the uniquely labeled states of the irreducible representation of the larger group. Conditions under which such a problem is soluble have been discussed by E. P. Wigner, unpublished.

[6] For other recent work on this problem see: M. Umezawa, Nuclear Phys. 57, 65 (1964); M. Micu (unpublished).

arbitrary representation $[pq]$ with the representation [210] to give a final representation $[p'q']$, where $[pq] \neq [p'q']$. The method employed is outlined in IV and depends upon results contained in II, III, and IV for the explicit calculation of the formulas. The method is completely general, as one may calculate both the nondiagonal [21 $\cdots$ 10] Wigner operator matrix elements for the general unitary group $SU_n$ by its use as well as the diagonal operator matrix elements (from the generators and the symmetric coupling coefficient). However, detailed results are given only for the $SU_3$ case.

For completeness we also give (as Wigner coefficients) matrix elements of the two octet operators $\begin{pmatrix} 1 & & \\ 1 & 1 & \\ 2 & 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & & \\ 2 & 0 & \\ 2 & 1 & 0 \end{pmatrix}$, which couple the representation $[pq]$ to itself. For the $\begin{pmatrix} 1 & & \\ 2 & 1 & 1_0 \end{pmatrix}$ operator this requires only a phase change from results given already in II. Similarly $\begin{pmatrix} 1 & & \\ 2 & 0 & \\ 2 & 1 & 0 \end{pmatrix}$ is completely defined in I–III as the (orthogonal) operator

$$X_A^{[2]} - (I_3/I_2)X_A$$

upon normalizing. Numerical and algebraic tables of these operators have appeared earlier.[7,8]

We feel that these results are of interest for several reasons: First the method used is a general one, and it may be employed to calculate [21 $\cdots$ 10] Wigner operator matrices for the general unitary group. The presentation of these results is a logical extension of the prior papers in this series[1] and depends extensively upon results contained in these papers; a detailed example is very helpful in illustrating these basic results. An adequate and economical notation for the general $SU_n$ *reduced Wigner operator*, introduced in IV, is also employed. Thus the presentation of a general scheme for the calculation of the [21 $\cdots$ 10] Wigner operators for $SU_n$ with phase conventions included, gives a synoptic view of the entire problem. Matrix elements of the complete "octet" tensor operators in $SU_3$ are of considerable current interest in elementary-particle physics; it is this interest which accounts for the large number of more or less complete numerical and algebraic tabulations.[9] To a large extent these earlier results satisfy the particular needs of elementary-particle physics; the major reason for the current work lies rather in its emphasis on the general structure of the tensor operator classification problem and its systematic resolution. The key to this resolution lies in the fundamental Wigner coefficients which, in a logical sense, imply the structure of the general tensor operator. These fundamental Wigner coefficients have been given already for the general unitary groups in II.

However, it is essential to note that any explicit Wigner operators obtained from these general results must properly take into account the canonical phase conventions discussed in III. It is therefore useful to specialize to $U_3$ and give explicit results to illustrate both the method and the necessarily complicated notational conventions.

## II. THE $SU_3$ ("SPINOR") WIGNER COEFFICIENTS

The results to follow are all dependent upon the explicit matrices[10] for the $SU_n$ generators given by II, Eqs. (60)–(62). It was noted that the matrix element of $E_{n,k}$ factored into a reduced matrix element multiplied by an $SU_{n-1}$ Wigner coefficient. Thus we obtain the $SU_3$ spinor and conjugate spinor (the [100] and [110] $SU_3$ representations) Wigner coefficients from the matrices of the generators $E_{i,4}$ and $E_{4,i}$. We may further factor these $SU_3$ "spinor" Wigner coefficients into a reduced Wigner coefficient multiplied by an $SU_2$ Wigner coefficient.

The reduced Wigner coefficient is a function of the $U_3$ and $U_2$ representation labels but not of the $U_1$ label (the $SU_2$ "magnetic" quantum number). All of the dependence upon the $U_1$ label is contained in the $SU_2$ Wigner coefficient.[11] The changes in the $U_2$ and $U_3$ representation labels can be indicated

[7] Numerical tabulations of the operators have been given by: J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1963); K. T. Hecht, Bull. Am. Phys. Soc. **8**, 57 (1963); P. McNamee, S. J. Chilton, and Frank Chilton, Rev. Mod. Phys. **36**, 1005 (1964).

[8] Algebraic results for these operators include: L. C. Biedenharn, Phys. Letters **3**, 69 (1962); **3**, 254 (1963); D. L. Pursey, Proc. Roy. Soc. (London) **A275**, 284 (1963).

[9] J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1963), gave six tables of numerical values; P. McNamee, S. J. Chilton, and Frank Chilton, Rev. Mod. Phys. **36**, 1005 (1964), gave five numerical tables; M. A. Rashid, Nuovo Cimento **26**, 118 (1962), gave six numerical tables; A. R. Edmonds, Proc. Roy. Soc. (London) **A268**, 567 (1962), gave five numerical tables. For algebraic results see: L. C. Biedenharn, Phys. Letters **3**, 69 (1962); **3**, 254 (1963); D. L. Pursey, Proc. Roy. Soc. (London) **A275**, 284 (1963); D. Lurie and A. J. Macfarlane, J. Math Phys. **5**, 565 (1964). (These results for matrix elements of the octet operator differ from our nondiagonal matrices in the normalizing factors of certain of the octet matrix elements. Our [210] matrix elements correspond to Wigner coefficients for these cases.) See also J. G. Kuriyan, D. Lurie, and A. J. Macfarlane, J. Math. Phys. **6**, 722 (1965).

[10] Note the two misprints in II, Eq. (60): (a) the factor $(m_{j,n} - m_{i,n-1} - j \pm i + 1)$ should read:

$$(m_{j,n} - m_{i,n} - j + i + 1);$$

(b) in the final Gelfand pattern "$m$" should be replaced by $m_{n,n}$.

[11] The transformation from the $U_2$ Gelfand labels to the more familiar notation for $SU_2$ is accomplished by use of the relations: $2J = m_{1,2} - m_{2,2}$; $M = m_{1,1} - \frac{1}{2}(m_{1,2} + m_{2,2})$.

by specifying two indices $i_1$ and $i_2$. For the spinor reduced Wigner coefficients an arbitrary $U_3$ representation $[m_{1,3}m_{2,3}m_{3,3}]$ is coupled to the $U_3$ representation [100]; the final Gelfand state label will then be identical to the initial Gelfand state label *except* that

$$m'_{i_j,4-j} = m_{i_j,4-j} + 1, \tag{1}$$

where $m$ is the initial state and $m'$ is the final state. An arbitrary state $(m)$ is given by

$$(m) = \begin{bmatrix} m_{1,3} & m_{2,3} & m_{3,3} \\ & m_{1,2} & m_{2,2} \\ & & m_{1,1} \end{bmatrix}. \tag{2}$$

In the interest of convenience to the reader let us note the translation of the Gelfand pattern into the usual elementary particle symbols. The state $(m)$ given in Eq. (2) above would read

$$(m) = \begin{bmatrix} p & & q & & 0 \\ & I + \frac{Y}{2} + \frac{p+q}{3} & & -I + \frac{Y}{2} + \frac{p+q}{3} & \\ & & I_z + \frac{Y}{2} + \frac{p+q}{3} & & \end{bmatrix}, \tag{3}$$

where $I$, $I_z$, and $Y$ are the isospin, $z$ component of isospin, and the hypercharge, respectively; $p$ and $q$ designate the representation by a Young pattern, but often one sees in the literature the notation $\lambda = p - q$, $\mu = q$.

An alternate specification of the reduced Wigner operator is given in IV and is summarized here. The reduced canonical tensor operator is designated by the notation

$$\begin{bmatrix} (\alpha) : & n \\ (\beta) : & n-1 \end{bmatrix} \equiv \left\langle \begin{matrix} \alpha \\ A \\ a \end{matrix} \right\rangle \cdot \left\langle \begin{matrix} \beta \\ B \\ b \end{matrix} \right\rangle, \tag{4}$$

where the operator $\langle A \rangle$ is in $SU_n$, the operator $\langle B \rangle$ is in $SU_{n-1}$, and the dot product implies a conjugation of $\langle B \rangle$.

Several things about this notation should be noted. The notation "$(\alpha) : n$" designates the upper $(SU_n)$ Gelfand pattern of $\langle A \rangle$, including the labels of $[A]$. Similarly "$(\beta) : n - 1$" designates the upper $(SU_{n-1})$ Gelfand pattern of $\langle B \rangle$, including the labels of $[B]$. Thus the labels "$(a)$" and "$(b)$" on the right hand side of Eq. (4) are suppressed; however, the labels $a_{i,n-1}$ and $b_{i,n-1}$ must agree identically for the dot product to be meaningful. The remaining labels $a_{i,j}$ and $b_{i,j}$ $(j \leq n - 2)$ are summed over in the dot product. Thus the reduced tensor operator notation is *a complete specification of the changes in the Gelfand pattern.*

As is shown in IV the *upper* Gelfand pattern specifies the changes in the $SU_n$ representation labels of a Gelfand state by the relations

$$\begin{aligned} m'_{i,n} &= m_{i,n} + \Delta_i, \\ \Delta_i &= -(i-1)m_{i-1} + \sum_{k=i}^{n-2} m_i + nm_{n-1}, \end{aligned} \tag{5}$$

where the $m_i$ are the fixed diagonal quantum numbers of the upper Gelfand pattern and are given in III, Eq. (5). [As these $\Delta$-values are for $SU_n$, $m_{n,n}$ is zero as is $\Delta_n$.]

By comparing this notation to that used in II, Eqs. (60)–(62), one sees that the patterns $(\alpha)$ and $(\beta)$ furnish a complete and unambiguous designation of the *changes* in the representation labels $[m_{i,n}] \to [m'_{i,n}]$ and $[m_{i,n-1}] \to [m'_{i,n-1}]$ induced by the tensor operator. Thus the notation of Eq. (4) is a direct generalization of the notation of II, which concerned only the special changes ($\pm 1$ in one index only) associated with the fundamental ("spinor") Wigner operators.

It should be noted, for completeness, that the explicit decomposition of the canonical $SU_n$ tensor operator is thus given by iterating the basic decomposition

$$\left\langle \begin{matrix} \alpha \\ A_n \\ a \end{matrix} \right\rangle = \sum_{(\beta)} \begin{bmatrix} (\alpha) : & n \\ (\beta) : & n-1 \end{bmatrix} \left\langle \begin{matrix} \beta \\ B \\ a \end{matrix} \right\rangle, \tag{6}$$

where the Gelfand pattern given by $B$ and $a$ on the right specifies the $U_{n-1}$ pattern $(a)$.

To illustrate the utility of the notation for an elementary case, the reduced tensor operators for the [100] representation of $SU_3$ are given in Table I. The *initial* $SU_3$ representation used in Table I is $[m_{1,3}\ m_{2,3}\ m_{3,3} = 0]$. Table I is constructed so that it leads to the *final* representation labels in $U_2$ of $[m_{1,2} - \delta_{i,3}\ m_{2,2} - \delta_{i,3}]$ so that coefficients constructed from this table are explicitly orthogonal. [This choice for the $U_2$ labels—which stems from Condon–Shortley—is not standard (for example, the last entry in Ref. 9 uses a different convention).] Table I and following tables give reduced Wigner coefficients between *lexical* Gelfand states. Needless to say, any Wigner coefficient involving the coupling of a *nonlexical* Gelfand state is defined to vanish.

The index $i$ used in Table I corresponds to the index $i_1$ defined in II. [It indicated that $m_{i_1,3} \to$

TABLE I. $SU_3$ [100] reduced Wigner coefficients. (Note that $i = 1 \Leftrightarrow a = b = 1$; $i = 2 \Leftrightarrow a = 1$, $b = 0$; $i = 3 \Leftrightarrow a = b = 0$. $S(x) \equiv$ sign of $(x)$; $s(0) = +1$.)

| $(\beta) =$ | $i_2 =$ | $\left\langle \begin{pmatrix} b \\ a\ 0 \\ 1\ 0\ 0 \end{pmatrix} : 3 \atop (\beta) \quad : 2 \right\rangle =$ |
|---|---|---|
| $\begin{pmatrix} 0 \\ 0\ 0 \end{pmatrix}$ | - | $\left[\dfrac{\prod_{j=1}^{2}(m_{j,2} - m_{i,3} - j + i - 1)}{\prod_{\substack{j=1 \\ j\neq i}}^{3}(m_{j,3} - m_{i,3} - j + i)}\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 1 \\ 1\ 0 \end{pmatrix}$ | 1 | $S(1-i)\left[\dfrac{(m_{i,3} - m_{2,2} - i + \upsilon)}{(m_{1,2} - m_{2,2} + 1)} \times \prod_{\substack{j=1 \\ j\neq i}}^{3}\left(\dfrac{m_{j,3} - m_{1,2} - j + 2}{m_{j,3} - m_{i,3} - j + i}\right)\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 0 \\ 1\ 0 \end{pmatrix}$ | 2 | $S(2-i)\left[\dfrac{(m_{1,2} - m_{i,3} - 2 + i)}{(m_{1,2} - m_{2,2} + 1)} \times \prod_{\substack{j=1 \\ j\neq i}}^{3}\left(\dfrac{m_{j,3} - m_{2,2} - j + 3}{m_{j,3} - m_{i,3} - j + i}\right)\right]^{\frac{1}{2}}$ |

$m_{i_1,3} + 1$.] Knowledge of this raising action is contained in the $SU_3$ reduced tensor operator. For convenience in the writing of Table I, let us make the following definitions (kronecker $\delta$'s):

$$a = \delta_{i,1} + \delta_{i,2}, \qquad b = \delta_{i,1}. \tag{7}$$

Using these definitions of $a$ and $b$, we give in Table I the $SU_3$ [100] ($SU_3$ spinor) reduced Wigner coefficients. The complete $SU_3$ [100] Wigner coefficients are obtained by multiplying the reduced Wigner coefficients of Table I by the appropriate $SU_2$ spin-$\frac{1}{2}$ Wigner coefficients. In the first column of Table I the upper Gelfand pattern $(\beta)$ is specified where $(\beta)$ belongs to the tensor operator

$$\left\langle \begin{pmatrix} b \\ a\ 0 \\ 1\ 0\ 0 \end{pmatrix} : 3 \atop (\beta) \quad : 2 \right\rangle.$$

In the second column the value of $i_2$ is given so that the correspondence between the two alternate

TABLE II. $SU_3$ [110] reduced Wigner coefficients. (Note that $i = 1 \Leftrightarrow c = d = 0$; $i = 2 \Leftrightarrow c = 0$, $d = 1$; $i = 3 \Leftrightarrow c = d = 1$. $S(x) \equiv$ sign of $(x)$, $s(0) = +1$.)

| $(\beta) =$ | $i_2 =$ | $\left\langle \begin{pmatrix} d \\ 1\ c \\ 1\ 1\ 0 \end{pmatrix} : 3 \atop (\beta) \quad : 2 \right\rangle =$ |
|---|---|---|
| $\begin{pmatrix} 1 \\ 1\ 1 \end{pmatrix}$ | - | $\left[\dfrac{\prod_{j=1}^{2}(m_{j,2} - m_{i,3} - j + i)}{\prod_{\substack{j=1 \\ j\neq i}}^{3}(m_{j,3} - m_{i,3} - j + i)}\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 0 \\ 1\ 0 \end{pmatrix}$ | 1 | $-S(1-i)\left[\dfrac{(m_{i,3} - m_{2,2} - i + 2)}{(m_{1,2} - m_{2,2} + 1)} \times \prod_{\substack{j=1 \\ j\neq i}}^{3}\left(\dfrac{m_{j,3} - m_{1,2} - j + 1}{m_{j,3} - m_{i,3} - j + i}\right)\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 1 \\ 1\ 0 \end{pmatrix}$ | 2 | $S(2-i)\left[\dfrac{(m_{1,2} - m_{i,3} - 1 + i)}{(m_{1,2} - m_{2,2} + 1)} \times \prod_{\substack{j=1 \\ j\neq i}}^{3}\left(\dfrac{m_{j,3} - m_{2,2} - j + 2}{m_{j,3} - m_{i,3} - j + i}\right)\right]^{\frac{1}{2}}$ |

TABLE III. $SU_3$ [210] reduced Wigner coefficients. [See Eq. (12) for definition of symbols.]

$$(\beta) = \left\langle \begin{bmatrix} c \\ a & b \\ 2 & 1 & 0 \end{bmatrix} : 3 \\ (\beta) \quad : 2 \right\rangle =$$

$\begin{bmatrix} 0 \\ 1 & 0 \end{bmatrix}$

$$S(1-j)\left[\frac{(m_{j,3}-m_{2,2}-j+2)(m_{i,3}-m_{1,2}-i+2)(m_{i,3}-m_{2,2}-i+3)}{A(m_{1,2}-m_{2,2}+1)} \times \prod_{\substack{k=1\\k\neq j}}^{3}(m_{k,3}-m_{1,2}-k+1)\right]^{\frac{1}{2}}$$

$\begin{bmatrix} 1 \\ 1 & 0 \end{bmatrix}$

$$-S(2-j)\left[\frac{(m_{1,2}-m_{j,3}-1+j)(m_{i,3}-m_{1,2}-i+2)(m_{i,3}-m_{2,2}-i+3)}{A(m_{1,2}-m_{2,2}+1)} \times \prod_{\substack{k=1\\k\neq j}}^{3}(m_{k,3}-m_{2,2}-k+2)\right]^{\frac{1}{2}}$$

$\begin{bmatrix} 2 \\ 2 & 1 \end{bmatrix}$

$$S(1-i)\left[\frac{(m_{i,3}-m_{2,2}-i+3)(m_{j,3}-m_{1,2}-j+1)(m_{j,3}-m_{2,2}-j+2)}{A(m_{1,2}-m_{2,2}+1)} \times \prod_{\substack{k=1\\k\neq i}}^{3}(m_{k,3}-m_{1,2}-k+2)\right]^{\frac{1}{2}}$$

$\begin{bmatrix} 1 \\ 2 & 1 \end{bmatrix}$

$$S(2-i)\left[\frac{(m_{j,3}-m_{2,2}-j+2)(m_{j,3}-m_{1,2}-j+1)(m_{1,2}-m_{i,3}-2+i)}{A(m_{1,2}-m_{2,2}+1)} \times \prod_{\substack{k=1\\k\neq i}}^{3}(m_{k,3}-m_{2,2}-k+3)\right]^{\frac{1}{2}}$$

$\begin{bmatrix} 0 \\ 2 & 0 \end{bmatrix}$

$$S(2-i)S(1-j)\left[\frac{(m_{1,2}-m_{i,3}-2+i)(m_{j,3}-m_{2,2}-j+2)}{A(m_{1,2}-m_{2,2}+1)(m_{1,2}-m_{2,2}+2)} \times \prod_{\substack{k=1\\k\neq i}}^{3}(m_{k,3}-m_{2,2}-k+3) \times \prod_{\substack{k=1\\k\neq j}}^{3}(m_{k,3}-m_{1,2}-k+1)\right]^{\frac{1}{2}}$$

$\begin{bmatrix} 1 \\ 2 & 0 \end{bmatrix}$

$$\left[\frac{(m_{i,3}-m_{1,2}-i+2)(m_{i,3}-m_{2,2}-i+3)(m_{j,3}-m_{1,2}-j+1)(m_{j,3}-m_{2,2}-j+2)}{2A(m_{1,2}-m_{2,2})(m_{1,2}-m_{2,2}+2)} \times \prod_{\substack{k=1\\k\neq i\\k\neq j}}^{3}(2m_{k,3}-m_{1,2}-m_{2,2}-2k+4)\right]^{\frac{1}{2}}$$

$\begin{bmatrix} 2 \\ 2 & 0 \end{bmatrix}$

$$-S(1-i)S(2-j)\left[\frac{(m_{i,3}-m_{2,2}-i+3)(m_{1,2}-m_{j,3}-1+j)}{A(m_{1,2}-m_{2,2})(m_{1,2}-m_{2,2}+1)} \times \prod_{\substack{k=1\\k\neq i}}^{3}(m_{k,3}-m_{1,2}-k+2) \times \prod_{\substack{k=1\\k\neq j}}^{3}(m_{k,3}-m_{2,2}-k+2)\right]^{\frac{1}{2}}$$

$\begin{bmatrix} 1 \\ 1 & 1 \end{bmatrix}$

$$\left[\frac{3(m_{i,3}-m_{1,2}-i+2)(m_{i,3}-m_{2,2}-i+3)(m_{j,3}-m_{1,2}-j+1)(m_{j,3}-m_{2,2}-j+2)}{2A}\right]^{\frac{1}{2}}$$

descriptions of the reduced tensor operator may be easily made. [The index $i_2$ indicates that the label $m_{i_2,2} \rightarrow m_{i_2,2} + 1$.]

In Table II the $SU_3$ [110] (conjugate spinor) reduced Wigner coefficients are given. The conjugate-spinor Wigner coefficients factor into the given reduced Wigner coefficient multiplied by an $SU_2$ Wigner coefficient.

For convenience in writing Table II let us define the parameters

$$c = 1 - \delta_{i,1} - \delta_{i,2}, \qquad d = 1 - \delta_{i,1}. \tag{8}$$

Using the above definitions of $c$ and $d$, the [110] $SU_3$ reduced Wigner coefficients are given. Table II is arranged in the same manner as Table I; however, $i_2$ now specifies the lowering of an $SU_2$ label [that is, $m_{i_2,2} \rightarrow m_{i_2,2} - 1$]. Again in Table II the *initial* $SU_3$ representation is $[m_{1,3}\, m_{2,3}\, m_{3,3} = 0]$ and the reduced Wigner coefficients couple to the *final* $U_2$ labels $[m_{1,2} + \delta_{i,3}\, m_{2,2} + \delta_{i,3}]$, leading once again to explicitly orthonormal $su_3$ states.

## III. DETERMINATION OF THE "VECTOR" COEFFICIENTS

It is well known that the generators of the unitary group $SU_n$ play a dual role: (1) as operators and (2) the carrier space of the adjoint representation. It is the latter role of which we make use: It establishes a mapping from the generators onto the basis states of the adjoint representation $X_A \rightarrow |X_A\rangle$.

Thus one has the result that the matrices of the generators (when properly phased and normalized) are themselves Wigner coefficients. (For $SU_2$ these are the Wigner coefficients $C^{J1J}_{MmM'}$.) These coefficients are, however, necessarily diagonal between the initial and final representations. To obtain the *nondiagonal* parts of the Wigner coefficients for $SU_n$, one may make use of a mapping of the generators of $SU_n$ into $SU_{n+1}$ systems. This mapping is achieved by using the anticommutator

$$E_{i,j} \rightarrow [E_{i,4}, E_{4,j}]_+ = E_{i,4}E_{4,j} + E_{4,j}E_{i,4}. \tag{9}$$

Thus by mapping into operators of the next higher group ($SU_{n+1}$), we may "raise" and "lower" the representation labels in $SU_n$ and obtain nondiagonal Wigner coefficients via the mapping of generators onto states

$$E_{i,j} \rightarrow |E_{i,j}\rangle. \tag{10}$$

As is discussed above, a phase is involved in this mapping and must be included in a calculation of the Wigner coefficients by this technique. The phase convention used here is a generalization of the Condon–Shortley phase convention for $SU_2$ and results in a consistent phase convention for any generator mapped onto an equivalent state. The results of this convention for $SU_n$ are

$$E_{i,j} \rightarrow (-)^{i+1} |E_{i,j}\rangle, \qquad H_i \rightarrow (-)^{i+1} |H_i\rangle, \tag{11}$$

where the $E_{i,j}$ are the nondiagonal (in the $SU_{n-1}$ subgroup) generators and the $H_i$ are the $n - 1$ diagonal generators.

By use of the Wigner coefficients in II, Eqs. (60)–(62) and of the two mappings (1) $SU_n$ generators into $SU_{n+1}$ operators and (2) generators onto states, the nondiagonal matrix elements of the $SU_n$ "vector" operators may be determined. These matrix elements factor into a product of a reduced matrix element multiplied by an $SU_{n-1}$ Wigner coefficient (as did the spinor Wigner coefficients).

Specific results for the $SU_3$ [210] ($SU_3$ "vector") Wigner coefficients have been obtained by use of the techniques outlined above. These results are presented in Table III using the reduced canonical tensor operator notation given in IV and summarized in Sec. II above.

For an $SU_3$ representation $[m_{1,3}\, m_{2,3}\, m_{3,3}=0]$, the reduced tensor operator notation gives the changes in these $SU_3$ labels [as is indicated by Eq. (5)]. Table III is further constructed so that the final $SU_2$ representation labels are $[m_{1,2} + d\ \ m_{2,2} + d]$, where $d$ is given in Eq. (12) below.

Let us define the parameters $a$, $b$, $c$, $d$, and $A$ for use in Table III by

$$a = 1 + \delta_{i,1} + \delta_{i,2}, \qquad b = \delta_{i,3},$$
$$c = \delta_{i,1} + \delta_{j,2} + \delta_{i,3}, \qquad d = -\delta_{i,3} + \delta_{j,3},$$
$$A = \left\{ \prod_{\substack{k=1 \\ k\neq i \\ k\neq j}}^{3} [(m_{k,3} - m_{i,3} - k + i) \times (m_{k,3} - m_{j,3} - k + j)] \right\} \times (m_{j,3} - m_{i,3} - j + i)(m_{i,3} - m_{j,3} - i + j + 1), \tag{12}$$

where $i$ and $j$ are indices in the formulas of Table III and obey the relations $1 \leq i, j \leq 3$, $i \neq j$.

In Table III the upper Gelfand pattern corresponding to the pattern ($\beta$) of Eq. (4) gives the $SU_2$ character of the reduced Wigner operators. Thus the upper Gelfand patterns $\binom{2}{2}\ \binom{0}{0}$, $\binom{1}{2}\ \binom{0}{0}$, and $\binom{0}{2}\ \binom{0}{0}$ give the $SU_2$ $\Delta_1$-values of 2, 0, and $-2$. From the relationship $J = \frac{1}{2}(m_{1,2} - m_{2,2})$, one sees that these

TABLE IV. Generator reduced matrix elements (*phased*).

| $(\beta) =$ | $\left(\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 2 & 1 & 0 \end{bmatrix} : 3 \right) = \begin{bmatrix} & & \\ & (\beta) & \\ & & \end{bmatrix} : 2$ |
|---|---|
| $\begin{bmatrix} 0 & \\ 1 & 0 \end{bmatrix}$ | $\left[\frac{(m_{1,3}-m_{1,2})(m_{1,2}-m_{2,3}+1)(m_{1,2}-m_{3,3}+2)}{6(m_{1,2}-m_{2,2}+1)I_2}\right]^{\frac{1}{2}}$ |
| $\begin{bmatrix} 1 & \\ 1 & 0 \end{bmatrix}$ | $-\left[\frac{(m_{1,3}-m_{2,2}+1)(m_{2,3}-m_{2,2})(m_{2,2}-m_{3,3}+1)}{6(m_{1,2}-m_{2,2}+1)I_2}\right]^{\frac{1}{2}}$ |
| $\begin{bmatrix} 2 & \\ 2 & 1 \end{bmatrix}$ | $\left[\frac{(m_{1,3}-m_{1,2}+1)(m_{1,2}-m_{2,3})(m_{1,2}-m_{3,3}+1)}{6(m_{1,2}-m_{2,2}+1)I_2}\right]^{\frac{1}{2}}$ |
| $\begin{bmatrix} 1 & \\ 2 & 1 \end{bmatrix}$ | $\left[\frac{(m_{1,3}-m_{2,2}+2)(m_{2,3}-m_{2,2}+1)(m_{2,2}-m_{3,3})}{6(m_{1,2}-m_{2,2}+1)I_2}\right]^{\frac{1}{2}}$ |
| $\begin{bmatrix} 0 & \\ 2 & 0 \end{bmatrix}$ | $0$ |
| $\begin{bmatrix} 1 & \\ 2 & 0 \end{bmatrix}$ | $\left[\frac{(m_{1,2}-m_{2,2})(m_{1,2}-m_{2,2}+2)}{12I_2}\right]^{\frac{1}{2}}$ |
| $\begin{bmatrix} 2 & \\ 2 & 0 \end{bmatrix}$ | $0$ |
| $\begin{bmatrix} 1 & \\ 1 & 1 \end{bmatrix}$ | $\frac{-[\frac{1}{2}(m_{1,2}+m_{2,2})-\frac{1}{3}(m_{1,3}+m_{2,3}+m_{3,3})]}{(I_2)^{\frac{1}{2}}}$ |

$SU_2$ $\Delta_1$-values imply that $\Delta J = +1$, 0, and $-1$, and thus this set of operators acts like an $SU_2$ representation with $J = 1$, that is, an $SU_2$ "vector." Similarly, the $SU_2$ character of the other terms may be calculated.

The upper pattern $(\beta)$ also gives the $SU_3$ character of the reduced Wigner operator. By application of the generator $H_2$ [see III, Eq. (8)] one may find the *change* $\Delta m_2$ in the $m_2$ eigenvalue. Thus,

$$\begin{aligned}\Delta m_2 &= \tfrac{1}{2}(m_{1,2}+m_{2,2}) - \tfrac{1}{3}(m_{1,3}+m_{2,3}+m_{3,3}),\\ &= \tfrac{1}{2}(\beta_{1,2}+\beta_{2,2}) - \tfrac{1}{3}(2+1+0),\\ &= \tfrac{1}{2}(\beta_{1,2}+\beta_{2,2}) - 1.\end{aligned} \tag{13}$$

Table IV contains the $\langle 2\ {}^{1}_{1}\ {}^{1}_{1}\ 0\rangle$ reduced Wigner coefficients. [These are just the generator matrix elements suitably phased by use of the phase convention given in Eq. (11).]

Table V contains reduced matrix elements of the $\langle 2\ {}^{2}_{1}\ {}^{1}_{1}\ 0\rangle$ operator which is just the operator $[X^{[2]} - (I_3/I_2)X]$. Both of the Tables IV and V are diagonal in their $SU_3$ representation labels coupling the representation $[m_{1,3}m_{2,3}m_{3,3}]$ to itself; the reduced operators of these tables lead to the final $SU_2$ representation $[m_{1,2}m_{2,2}]$. For convenience in writing Table V let us make the following definitions:

$$\begin{aligned}\lambda &= \tfrac{1}{2}(m_{1,2}-m_{2,2}),\\ m &= \tfrac{1}{2}(m_{1,2}+m_{2,2}) - \tfrac{1}{3}(m_{1,3}+m_{2,3}+m_{3,3}),\\ I_2 &= \tfrac{1}{9}[(m_{1,3}-m_{3,3})^2 + (m_{2,3}-m_{3,3})^2\\ &\qquad - (m_{1,3}-m_{3,3})(m_{2,3}-m_{3,3})\\ &\qquad + 3(m_{1,3}-m_{3,3})],\\ I_3 &= 2^{-1}\cdot 3^{-4}(m_{1,3}-2m_{2,3}+m_{3,3})\\ &\qquad \times (2m_{1,3}-m_{2,3}-m_{3,3}+3)\\ &\qquad \times (m_{1,3}+m_{2,3}-2m_{3,3}+3),\\ B &= 6(\tfrac{1}{9}I_2^3 + \tfrac{1}{36}I_2^2 - I_3^2)/I_2\end{aligned} \tag{14}$$

where $I_2$ and $I_3$ are the two invariants (defined in I)

of $SU_3$. (Note that when $B$ is zero, the operator is defined to vanish—this occurs for 'triangle' representations.)

Thus Table III gives a complete specification of the $SU_3$ [210] Wigner operators which are non-diagonal in representation labels while Table IV and Table V give the [210] operator matrix elements which are diagonal in representation labels.

TABLE V. $\langle \begin{smallmatrix} 1 \\ 2\ 0 \end{smallmatrix} \rangle$ operator reduced matrix elements: $[X^{(2)} - (I_3/I_2)X]$. [Cf. Eq. (14) for symbols.]

| $(\beta) =$ | $\left(\begin{array}{c} 1 \\ 2\ 0 \\ 2\ 1\ 0 \\ (\beta) \end{array} 3: \right) =$ |
|---|---|
| $\begin{pmatrix} 0 \\ 1\ 0 \end{pmatrix}$ | $\frac{1}{3}[\lambda - m + \frac{1}{2} - 3(I_3/I_2)]\left[\frac{(m_{1,3} - m_{1,2})(m_{1,2} - m_{2,3} + 1)(m_{1,2} - m_{3,3} + 2)}{B(m_{1,2} - m_{2,2} + 1)}\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 1 \\ 1\ 0 \end{pmatrix}$ | $\frac{1}{3}[\lambda + m + \frac{1}{2} + 3(I_3/I_2)]\left[\frac{(m_{1,3} - m_{2,2} + 1)(m_{2,3} - m_{2,2})(m_{2,2} - m_{3,3} + 1)}{B(m_{1,2} - m_{2,2} + 1)}\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 2 \\ 2\ 1 \end{pmatrix}$ | $\frac{1}{3}[\lambda - m + \frac{1}{2} - 3(I_3/I_2)]\left[\frac{(m_{1,3} - m_{1,2} + 1)(m_{1,2} - m_{2,3})(m_{1,2} - m_{3,3} + 1)}{B(m_{1,2} - m_{2,2} + 1)}\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 1 \\ 2\ 1 \end{pmatrix}$ | $-\frac{1}{3}[\lambda + m + \frac{1}{2} + 3(I_3/I_2)]\left[\frac{(m_{1,3} - m_{2,2} + 2)(m_{2,3} - m_{2,2} + 1)(m_{2,2} - m_{3,3})}{B(m_{1,2} - m_{2,2} + 1)}\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 0 \\ 2\ 0 \end{pmatrix}$ | $\left[\frac{(m_{1,3} - m_{1,2})(m_{1,2} - m_{2,3} + 1)(m_{1,2} - m_{3,3} + 2)(m_{1,3} - m_{2,2} + 2)(m_{2,3} - m_{2,2} + 1)(m_{2,2} - m_{3,3})}{9B(m_{1,2} - m_{2,2} + 2)(m_{1,2} - m_{2,2} + 1)}\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 1 \\ 2\ 0 \end{pmatrix}$ | $\{\frac{1}{6}m + [\lambda(\lambda + 1)]^{-1}[I_3 + \frac{1}{2}mI_2 + \frac{1}{6}m(1 - m^2)] - (I_3/I_2)\}\left[\frac{(m_{1,2} - m_{2,2})(m_{1,2} - m_{2,2} + 2)}{2B}\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 2 \\ 2\ 0 \end{pmatrix}$ | $-\left[\frac{(m_{1,3} - m_{1,2} + 1)(m_{1,2} - m_{2,3})(m_{1,2} - m_{3,3} + 1)(m_{1,3} - m_{2,2} + 1)(m_{2,3} - m_{2,2})(m_{2,2} - m_{3,3} + 1)}{9B(m_{1,2} - m_{2,2} + 1)(m_{1,2} - m_{2,2})}\right]^{\frac{1}{2}}$ |
| $\begin{pmatrix} 1 \\ 1\ 1 \end{pmatrix}$ | $-[6/B]^{\frac{1}{2}}\{\frac{1}{6}[\lambda(\lambda + 1) - m^2 - I_2] - (I_3/I_2)m\}$ |

# A Relation between the Hydrogen Atom and Multidimensional Harmonic Oscillators

D. BERGMANN AND Y. FRISHMAN
*The Weizmann Institute of Science, Rehovoth, Israel*
(Received 3 May 1965)

A simple connection between the radial Schrödinger equation for the bound states of a hydrogen atom with angular momentum $l$ and that of an isotropic harmonic oscillator of even dimension from 2 to $4l + 4$ is noted. The case of highest dimensionality $4l + 4$ is shown to lead in a very simple way to the energy levels and degeneracies of the hydrogen atom, once the energy levels of a one-dimensional harmonic oscillator are known.

## INTRODUCTION

THE purpose of this paper is to point out an apparently unnoticed connection between the radial Schrödinger equations of the hydrogen atom and isotropic harmonic oscillators of various dimensions. The connection is made by performing simple transformations on the equation and the wavefunctions. As a byproduct, we get what is perhaps the easiest way to obtain the energy levels of the hydrogen atom (nonrelativistic) once one has solved the problem of the one-dimensional harmonic oscillator.

## TRANSFORMATION OF THE HYDROGEN ATOM EQUATION

We start from the radial Schrödinger equation for levels of the hydrogen atom with angular momentum $l$ and negative energies $E_H$

$$\left\{-\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + \frac{l(l+1)\hbar^2}{2mr^2} - \frac{ze^2}{r} - E_H\right\}y(r) = 0.$$

We define a dimensionless argument $\rho$

$$r = \left(\frac{\hbar^2}{-8mE_H}\right)^{\frac{1}{2}}\rho = l_0\rho$$

in terms of which the equation is

$$\left\{\frac{d^2}{d\rho^2} - \frac{l(l+1)}{\rho^2} - \frac{ze^2}{4E_H\rho l_0} - \frac{1}{4}\right\}y(l_0\rho) = 0.$$

Substituting $\rho = x^2$ we get

$$\left\{\frac{d^2}{dx^2} - \frac{1}{x}\frac{d}{dx} - \frac{4l(l+1)}{x^2} - \frac{ze^2}{E_H l_0} - x^2\right\}y(l_0x^2) = 0.$$

Now we substitute for the wavefunction $y(l_0x^2) = x^\lambda u(x)$, where $\lambda$ remains to be fixed

$$\left\{\frac{d^2}{dx^2} + \frac{2\lambda - 1}{x}\frac{d}{dx} + \frac{\lambda(\lambda-2) - 4l(l+1)}{x^2} - x^2 - \frac{ze^2}{l_0E_H}\right\}u(x) = 0.$$

By choosing $\lambda = 2l + 2 - L$, where $L$ is any integer between 0 and $2l + 1$ inclusive, this equation becomes

$$\left\{\frac{d^2}{dx^2} + \frac{N-1}{x}\frac{d}{dx} - \frac{L(L+N-2)}{x^2} - x^2 + \frac{2ze^2}{\hbar}\left(\frac{2m}{-E_H}\right)^{\frac{1}{2}}\right\}u(x) = 0,$$

where we have denoted $N \equiv 2\lambda = 4l + 4 - 2L$. The first two terms in the equation are the radial part of the Laplacian operator in $N$ dimensions, while the third term is the result of the $N$-dimensional Laplacian acting on an $N$-dimensional spherical harmonic of order $L$.

The equation we have obtained may be compared with the radial equation for an $N$-dimensional isotropic harmonic oscillator of mass $m$, frequency $\omega$, and "angular momentum" $L$, written in terms of the dimensionless argument $x = (m\omega/\hbar)^{\frac{1}{2}}r$

$$\left\{\frac{d^2}{dx^2} + \frac{N-1}{x}\frac{d}{dx} - \frac{L(L+N-2)}{x^2} - x^2 + \frac{2E_{\text{osc}}}{\hbar\omega}\right\}u(x) = 0.$$

The lowest dimensionality which can thus be obtained is $N = 2$, for which we must take $L = 2l + 1$. Any other even dimensionality up to $N = 4l + 4$ can be obtained by choosing successively lower $L$ until for $L = 0$ the highest $N$ is obtained.

## THE ENERGY LEVELS OF HYDROGEN (NONRELATIVISTIC)

The simplest way to obtain these is by considering the highest $N$, which goes together with $L = 0$. This means that the levels of hydrogen correspond in this case to those levels of the $(4l + 4)$-dimensional harmonic oscillator which are invariant under rotations in $(4l + 4)$-dimensional space. The ground state of this oscillator is clearly such a state. Any ex-

cited state of an oscillator can be obtained by acting on the ground state with suitable functions of the creation operators $a_i^+$ $(i = 1, \cdots N)$ of the $N$ different cartesian oscillators into which it decomposes. In order to make only scalar states from the ground state, we must act on it only with scalar functions of $a_i^+$. Any such operator must be a function of $A^+ \equiv \sum_{i=1}^{N} (a_i^+)^2$, so we can generate all the desired states of the oscillator by acting on the ground state with $A^+$ any number of times. In this way it is clear that we get an infinite multiplet of nondegenerate oscillator states with energies

$$E_{\text{osc.}n} = \hbar\omega(2l + 2 + 2n) \qquad n = 0, 1, 2, \cdots .$$

This means that the hydrogen atom energies for a for a certain $l$ are going to be given by

$$\frac{2ze^2}{\hbar}\left(\frac{2m}{-E_{\text{H}}}\right)^{\frac{1}{2}} = \frac{2E_{\text{osc}\,n}}{\hbar\omega} = 4(l + n + 1),$$

$$E_{\text{H}}(l, n) = -\frac{z^2e^4m}{2\hbar^2(l + n + 1)^2}.$$

Thus, the quantum number $n$ defined to characterize the harmonic oscillator levels is just the radial quantum number of the hydrogen atom.

## ACKNOWLEDGMENT

We would like to thank Prof. H. J. Lipkin for suggesting a related problem to us, out of which this work has grown.

# Quantum Electrodynamics of the Stimulated Emission of Radiation. II

R. M. BEVENSEE
*University of California, Berkeley, California**
(Received 25 January 1965)

The equations of motion for the resonant modes in a lossy cavity, which couples them through the nonresonant thermal loss mechanism described by Senitzky, are examined. The second-order perturbation treatment of dissipation is justified for long time intervals by reference to an intermittent similarity transformation. Self-consistent expressions for commutators such as $\langle [P_n(t_1), P_m(t_2)] \rangle_d$, $P_n$ being the operator-amplitude of the electric field in the $n$th mode, as averaged over the loss mechanism $d$ with respect to its *diagonal* density matrix, are obtained from Laplace transform expressions for these operators. These averaged commutators describe correlations of fluctuations in different modes, in a way compatible with the ordinary commutation relations for lossless modes. The self-consistent anticommutator expressions, averaged over the loss mechanism, imply the Bose–Einstein formula for thermal energy in each mode at time $t_2 = t_1$. This is shown from direct examination of anticommutators and also from an integration of the correlation function. Commutators for the total electric and magnetic fields at two space-time points in the cavity have the proper causal properties. The dissipation terms in the field equations of motion cannot be derived from an equivalent Hamiltonian unless a modified analysis such as that of Kemeny is undertaken. Finally, the approximate equations of motion for the mode operator-amplitudes of a parametric amplifier are derived in the creation-annihilation operator framework and the spectral density of noise obtained at the signal frequency. The equilibrium noise present due to coupling with the idle modes is proportional to the number of photons in the idle frequency mode instead of that number plus one, in contrast to the conclusion of Wagner and Hellwarth. In this system the loss-averaged operators evolve from initial thermal equilibrium values to parametric equilibrium values according to nonlinear equations of motion containing time dependent commutators.

## 1. INTRODUCTION

In Part I of this series[1] we developed the equations of motion for the electric and magnetic field operator-amplitudes of one resonant cavity mode interacting with a thermal wall loss mechanism and molecules. The nonlinear equations for the time-average field energy and molecular power flow were accurate through second-order perturbation theory, i.e., through terms quadratic in the field-wall loss and field-molecular coupling coefficients. The perturbation expressions for these operators as derived for a "short" time interval (long compared to the signal period but short compared to the decay time) were extended over long time intervals by application of an intermittent similarity (IS) transformation which restored the perturbed molecular operator matrices to their unperturbed forms at the end of each short time interval. This transformation also acted upon the density matrix of the system in the Heisenberg representation. Fortunately the description of the expectation value of time-average field energy and its dispersion over long intervals did not involve the individual elements of the density matrix.

In this description of the stimulated emission process we tacitly assumed along with Senitzky[2] that the field-loss mechanism interaction depended upon the field amplitude times a zero-order loss mechanism expectation value. We discarded the detailed behavior of the loss mechanism operators[3] in the equation of motion for the field-loss coupling operator $\Gamma$ ($F = \Gamma/4\pi c$ in this paper) as driven by the electric field, and we took the loss mechanism density matrix to be purely diagonal.

In order to justify the use of zero-order loss mechanism operators through successive short-time intervals with many resonant cavity modes present, we deduce in Sec. 2 an appropriate IS transformation. This same transformation applied to the global density matrix as always a direct product of field, loss mechanism, and perhaps molecular density[1] matrices makes the expectation values of all operators depend on the diagonal portion of the loss mechanism matrix as well as on a transformed field expectation

* Present address: Lawrence Radiation Laboratory, University of California, Livermore, California.

[1] R. M. Bevensee, J. Math. Phys. (N. Y.) 5, 308 (1964).

[2] I. R. Senitzky, Phys. Rev. 119, 670 (1960).

[3] This means replacement of an expectation value like

$$\langle \epsilon^{i\mathcal{H}_d^{(0)}(t_1-t_2)/\hbar} [\Gamma(t_2), [\Gamma(t_3), \mathcal{H}_d(t_3)]P(t_3)] \times \epsilon^{-i\mathcal{H}_d^{(0)}(t_1-t_2)/\hbar} \rangle_d$$

by

$$\langle [\Gamma^{(0)}(t_2), [\Gamma^{(0)}(t_3), \mathcal{H}_d^{(0)}]] \rangle_d P(t_3)$$

where superscript (0) denotes a zero-order unperturbed operator of the loss mechanism with Hamiltonian $\mathcal{H}_d$ and $\langle\ \rangle_d$ implies an expectation value over loss mechanism space only. In this paper the exponentials are retained and only the zero-order operators are involved in expectation values during short time intervals.

which is just the initial expectation within errors quadratic in the (short) time interval. In this way the effects of the interaction appear only in the field portion of the density matrix.

As a result we are able to derive the self-consistent steady-state commutators and anticommutators between modes of frequency $\simeq\omega$, as averaged over the loss mechanism space, with fractional error on the order of $\beta/\omega$, $\beta$ being the dissipation constant for energy. These expectations imply a subsequent averaging to be performed over field space. It seems to us, therefore, that Senitzky's formalism is entirely adequate for treating the dissipation phenomena during long time intervals.

We now clarify our analysis by comparison with the treatment of Willis and Bergmann[4] of the equation of motion for the density matrix of an arbitrary physical system in contact with an infinite thermal reservoir of fermion (or boson) free particles. These authors treated weak but sustained interaction between system and reservoir by generalizing the Wigner–Weisskopf treatment of line broadening in such a way that only a single reservoir particle was present at a time. The result was a self-consistent solution for the matrix elements of a time-independent kernel $K$ in the equation of motion $d\bar{\rho}/dt = K\bar{\rho}$, $\bar{\rho}$ being the density matrix of the system as the global density matrix $\bar{\mu}$ (called $\rho_g$ herein) averaged over the reservoir coordinates: $\bar{\rho} = \langle\bar{\mu}\rangle_d$. Subscript d denotes a loss mechanism ensemble average. As a result they avoid the inconsistent rerandomization of phase procedure at the end of each "short" time interval (short in the sense defined earlier) but at the expense of overly restrictive assumptions for the sake of self-consistency. Perhaps the most serious one is the restriction of the matrix elements of $K\bar{\rho}(t)$ to certain system energy states[5] to guarantee the time independence of $K$. These relations could not reasonably be applied to the resonant modes of a cavity in contact with a fermion reservoir. They take the reservoir to be diagonal at all times and obtain equations of motion for matrix elements like $\langle m|\bar{\mu}|n\rangle$ between reservoir states which "move" according to the elements of $\bar{\mu}$ which are diagonal in reservoir coordinates. We, too, find that, even though the first-order perturbation of the reservoir by the system leads to dissipation, only the diagonal elements of $\bar{\mu}$ in loss mechanism space appear in expectation values of operators averaged over this space during a succession of short-time intervals, in each of which the expectations of field operators can be followed. We do not make their assumption of one reservoir particle at a time; rather, the important matrix elements of the loss mechanism states represent transitions between all energy levels $i$ and $j$. The equation of motion which Willis and Bergmann obtain is valid over a time interval such that $\lambda^2 t \ll 1$, $\lambda$ measuring the strength of interaction, and our solution is valid over a succession of similar intervals.

In this paper we assume throughout, except in Sec. 8, that no energy sources are present in the cavity. We first derive the equations of motion for the electric and magnetic field operators $P_n(t)$ and $Q_n(t)$ of a typical $n$th mode of a lossless cavity, coupled to the other modes through the agency of the wall loss mechanism. The philosophical question then arises: do we know the commutation relations between the various mode amplitudes, consistent with the loss mechanism coupling? The work in Sec. 8 demonstrates the plausible fact that if the loss mechanism were gradually turned on we would have to study the system with equations of motion containing time-dependent commutators which would evolve, in a loss-averaged sense, to thermal equilibrium values. About all we can safely say about a cavity in which the fields have come to equilibrium with the loss mechanism is that the density matrix of the latter is diagonal or nearly so. It is sensible, therefore, to consider a subensemble of systems all with the same field density matrix initially and derive the equations of motion for the field operators by employing *initially unknown commutators* between all pairs of mode operators, averaged over loss mechanism space.[6] These averaged commutators represent time correlations between mode operators caused by the driving loss mechanism, as distinct from the coupling between modes created by the first-order perturbation in loss mechanism responsible for dissipation. Correlation between modes is strong if their resonant frequency separation is on the order of the decay constant of either. The

[4] C. R. Willis and P. G. Bergmann, Phys. Rev. **128**, 391 (1962).

[5] For example, $E_\alpha - E_\epsilon + E_\gamma - E_\beta = 0$, the Greek letters denoting states of the system.

[6] One might tend to believe that the loss-averaged commutators, and anticommutators such as $\langle[P_n(t_1), P_m(t_2)]\rangle_d$ should be obtained self-consistently from the equations of motion for the field operators according to the Heisenberg prescriptions using *unaveraged* commutators. Because these commutators imply correlations in fluctuations between mode operators due to the zero-order thermal loss mechanism the commutators should be interpreted as averages over loss mechanism space. Although dissipation terms appear in the equations of motion for the field operators from averaging of the perturbed portion of the loss mechanism over its space, the significant terms in the similarly averaged commutation relations result from averaging the products of the operators, not from the products of the averaged operators.

averaged commutators (Sec. 4) and anticommutators (Sec. 5) between mode amplitudes are nonzero because of the peculiar stochastic properties of the zero-order uncoupled loss mechanism operator $F^{(0)}$ which connects the electric field with the loss mechanism Hamiltonian. However, we show in Sec. 6 that the self-consistent loss-averaged commutators for electric and magnetic field components at two space–time points in the cavity have the causal properties expected of them.

Senitzky has shown[7] that the net commutator $\langle[P_n(t), Q_n(t)]\rangle_d$ of a single mode interacting with the loss mechanism remains exactly $-i\hbar$ in the second-order perturbation treatment. One may verify the fact that this commutator remains $-i\hbar$ even if an assemblage of molecules is present[8] with a Gaussian spread in transition frequencies about the cavity mode resonant frequency. Now, such a commutator does not represent uncertainties due to the measurement process but rather correlations in fluctuations due to the thermal loss mechanism which determines the minimum uncertainties of measurement, according to the particular relations $[P_n(t), Q_n(t)] = -i\hbar$ and $\Delta P_n \Delta Q_n \cong \hbar$ for a given $n$th mode. This statement is in accord with Bohr's philosophy[9] but it seems somewhat of a coincidental yet consistent fact that the thermal dissipation mechanism has the property of maintaining precisely this uncertainty product which is ordinarily inferred from measurement limitations neglecting dissipation.

With respect to a general Hamiltonian we can define and use an equivalent Hamiltonian (Sec. 7) from which the equations of motion for the loss-averaged field operators follow from the ordinary commutation relations for a lossless cavity, with two exceptions. First, the dissipation terms must be obtained by averaging certain loss mechanism terms in the correct equations of motion for the field operators if the procedure of this paper is followed. Second, the terms in the equations of motion due to nonthermal energy sources and sinks must be obtained from the new-equilibrium, loss-averaged commutation relations. A loss term in the equivalent Hamiltonian quadratic in field variables would not yield the correct dissipation terms in the equations of motion for the field operators unless that term involved both physical and nonphysical or adjoint field operators and a variational or action principle were employed to obtain the equations of motion. For example, Kemeny[10] employed Schwinger's action principle as applied to the system Lagrangian written in terms of adjoint operators defined on the reverse time track of the system.

According to the results of Sec. 5 the spectral density of energy in the cavity is composed of contributions from the modes with resonant frequencies near the frequency of interest, each mode having a Lorentzian spectrum about its own resonant frequency. The total energy in each of these spectrums is just the Bose–Einstein value. When the parametric coupling described by Wagner and Hellwarth[11] is introduced the equations of motion for the field operators in the creation–annihilation operator framework are altered and a fundamental result emerges. That portion of the equilibrium thermal noise spectrum in the range of signal frequencies due to the parametric coupling is proportional to the number of photons $n_i$ in the relevant idler mode, not $n_i + 1$. This comes about because the spectrum of energy must be measured by the full averaged anticommutator, not half of it as Wagner and Hellwarth claim, because of the unusual stochastic nature of the loss mechanism. The symmetrical form of the anticommutator implies the same dependence of spectral intensity upon $n_s$, the number of photons in the relevant signal frequency mode, as upon $n_i$.

## 2. PROPERTIES OF THE LOSS MECHANISM

According to Ref. 2 the electric and magnetic fields of the $n$th cavity mode, defined without loss, are

$$E_n(r, t) = -4\pi c P_n(t) u_n(r), \tag{1a}$$

$$H_n(r, t) = Q_n \nabla \times u_n \tag{1b}$$

with the vector patterns conveniently normalized as

$$\int u_n \cdot u_m \, dv = \delta_{nm}, \quad \int \nabla \times u_n \cdot \nabla \times u_m \, dv = \delta_{nm} k_n^2 \tag{2a}$$

$$u_n \times n = 0 \text{ on walls}, \qquad k_n = \omega_n / c, \tag{2b}$$

$n$ being the unit normal pointing out of the cavity. Classically the equation of motion for $\dot{Q}_n(t)$ is obtained by substituting mode expansions for $E$ and $H$ into one of Maxwell's equations,

$$\nabla \times E = -\int [n \times E] \delta(r - r_{\text{wall}}) \, dv - \dot{H}/c, \tag{3}$$

[7] I. R. Senitzky, Phys. Rev. 115, 227 (1959).
[8] I. R. Senitzky, Phys. Rev. 119, 1807 (1960).
[9] D. Bohm and Y. Aharonov, Phys. Rev. 122, 1649 (1961).
[10] G. Kemeny, Phys. Rev. 133, A69 (1964).
[11] W. G. Wagner and R. W. Hellwarth, Phys. Rev. 133, A915 (1964).

whereupon dot-multiplication by $\nabla \times u_n$ and integration over the cavity volume yields an equation for $\dot{Q}_n$ in terms of $P_n$ and an additional integral over the walls of the form

$$\int ds \sum_m Q_m(t) \nabla \times u_m \cdot \nabla \times u_n \tag{4}$$

because $E \propto J \propto n \times H$ on the walls. Quantum mechanically the dissipation term may be obtained by including a field-loss mechanism energy term in the Hamiltonian, of the form

$$\mathcal{H}_{fd} = 4\pi c \sum_k \alpha_k P_k(t) F, \qquad F = F^{(0)} + F^{(1)} \tag{5}$$

(superscript 1 denotes a perturbation due to the coupling) whereupon it turns out that after averaging the operators $P_n$ and $Q_n$ over loss mechanism space the equation of motion for $\dot{P}_n$ and not $\dot{Q}_n$ contains a term of the form

$$\sum_m P_m(t) \alpha_m \alpha_n. \tag{6}$$

This corresponds to (4) if $\alpha_m \alpha_n$ has the significance of[12]

$$\alpha_m \alpha_n = \left( \int ds \, \nabla \times u_m \cdot \nabla \times u_n \right) \times \left( \int (\nabla \times u_n)^2 \, dv \right)^{-1}. \tag{7}$$

For the mode coupling of interest $k_n \cong k_m$.

The zero-order loss mechanism operator $F^{(0)}$ has only off-diagonal matrix elements $F^{(0)}(t)_{ij} = F_{ij} \epsilon^{i\omega_{ij} t}$, where it is convenient to take each $F_{ij}$ to have phase zero. The stochastic properties of $F^{(0)}$ are summarized by the following loss mechanism ensemble expectations[2]:

$$\langle F^{(0)}(t) \rangle_d = 0 \tag{8a}$$

$$(4\pi c)^2 \langle F^{(0)}(t_1) F^{(0)}(t_2) \rangle_d = (\hbar A/2) \left[ \int_0^\infty d\omega' \, B(\omega') \epsilon^{-\gamma' + i\omega'(t_1 - t_2)} + \int_0^\infty d\omega' \, B(\omega') \epsilon^{-i\omega'(t_1 - t_2)} \right], \quad \gamma' = \hbar\omega'/kT. \tag{8b}$$

The $\langle \mathcal{O} \rangle_d$ refers throughout this paper to an ensemble average of an operator over the loss mechanism states, with respect to its density operator $\rho_d$. $A$ is a normalizing factor in its unperturbed density matrix representation

$$\rho_{mn} = \delta_{mn} A \epsilon^{-E_m/kT}, \tag{9a}$$

where

$$A^{-1} = \sum_i \epsilon^{-E_i/kT} \cong \int_0^\infty \rho(E) \epsilon^{-E/kT} \, dE. \tag{9b}$$

In (8b) $B(\omega')$ [Eq. (32) of Ref. 2] is an integral over energy of a function of $\rho(E)$ as well as of $E$ and the magnitudes of the matrix elements of $F^{(0)}$ connecting various loss dipole states. Eq. (8b) implies a subsequent integration over times $t_1$ and $t_2$ when calculating the expectation value of any operator over the loss mechanism space.

It does not seem possible to derive self-consistent commutation relations by the methods of this paper unless there is a strong weighting factor in favor of one frequency $\omega$ in (8b), $\omega$ being the resonant frequency of that mode for which we wish commutation relations with neighboring modes. In fact, these relations for a given mode with its neighbors are not unduly affected until the frequency separations of the neighbors from the given mode become comparable to the decay constant, and at high frequencies (usually above the microwave range) where this occurs the strong weighting factor is present. For example, in the He–Ne gas laser described by Haken and Sauermann[13] the decay constant is considerably less than the frequency spacing of the two active cavity modes so the considerations of our paper are unimportant.

For strong weight over a narrow range of frequencies about $\omega$ (8) are written in the forms

$$\langle F^{(0)}(t) \rangle_d = 0, \tag{10a}$$

$$(4\pi c)^2 \langle F^{(0)}(t_1) F^{(0)}(t_2) \rangle_d = (4c^2 \hbar \beta_0/\omega)(1 - \epsilon^{-\gamma})^{-1} [\epsilon^{-\gamma} \times \int_0^\infty d\omega' \, \epsilon^{i\omega'(t_1 - t_2)} + \int_0^\infty d\omega' \, \epsilon^{-i\omega'(t_1 - t_2)}], \tag{10b}$$

with

$$\beta_0 = (\omega/8c^2) AB(\omega)(1 - \epsilon^{-\gamma}), \qquad \gamma = \hbar\omega/kT. \tag{10c}$$

The decay constant for the energy in the $n$th mode turns out to be $\beta_0 \alpha_n \alpha_n$.

This description of the interaction between the modes and loss mechanism is founded upon an essential continuum of energy states of the latter

[12] Actually $\alpha_k F$ stands for $\alpha_{kx} F_x + \ldots + \alpha_{kz} F_z$ in (5) and $\alpha_n \alpha_m = \alpha_{nx} \alpha_{mx} + \ldots + \alpha_{nz} \alpha_{mz}$ in (6) and (7). $F^{(1)}$ in (11) and (13) and $\mathcal{H}_d^{(1)}$ of (14), as well as the zero-order operators of (12), can be broken into component form whence the IS transformation of (15) should be considered as three separate transformations applied to the $x$, $y$, and $z$ operators separately. Any term appearing in the form $\alpha_n \alpha_m$ or $\alpha_n F$ should be interpreted in a dot-product sense.

[13] H. Haken and H. Sauermann, Z. Physik **173**, 261, Sec. 7 (1963); also *ibid.* **176**, 47 (1963).

and will become suspect at high frequencies above the optical range where the separation in energy between consecutive modes becomes comparable to the level separation of the loss mechanism.

We now show that an IS transformation exists at the end of a short time interval which will restore the $F$- and $\mathcal{H}_d$-matrices ($\mathcal{H}_d$ is the loss mechanism Hamiltonian) to their unperturbed off-and on-diagonal forms, respectively. These simple forms enable a repetition of the same field-loss mechanism description through successive (short) time intervals, with a changing field density matrix (see Appendix A). The general element of $F^{(1)}$ is, from the Heisenberg equations of motion for the $F$ and $\mathcal{H}_d$ operators, which both commute (self-consistently it turns out) with the field in the interaction term (5),

$$F^{(1)}(t)_{ik} = \frac{4\pi c}{\hbar^2}\int_0^t dt_1 \int_0^{t_1} dt_2 \langle i|\, \epsilon^{i\mathcal{H}_d^{(0)}(t-t_1)/\hbar}[F^{(0)}(t_1), \times [F^{(0)}(t_2), \mathcal{H}_d^{(0)}]]\epsilon^{-i\mathcal{H}_d^{(0)}(t-t_1)/\hbar}\,|k\rangle_d \sum_n \alpha_n P_n(t_2). \tag{11}$$

With

$$[F^{(0)}(t_1), [F^{(0)}(t_2), \mathcal{H}_d^{(0)}]]_{ik} = -\hbar \sum_j \omega_{jk}F_{ij}F_{jk}\epsilon^{i\omega_{ij}t_1} \times \epsilon^{i\omega_{jk}t_2} + \hbar\sum_j \omega_{ij}F_{ij}F_{jk}\epsilon^{i\omega_{jk}t_1}\epsilon^{i\omega_{ij}t_2}, \tag{12}$$

where

$$\omega_{ij} = \omega_i - \omega_j,$$

substitution of (12) into (11), with the exponentials diagonal, yields, after changing

$$\int_0^t dt_1 \int_0^{t_1} dt_2 \quad \text{to} \quad \int_0^t dt_2 \int_{t_2}^t dt_1$$

and performing the $t_1$-integration,

$$F^{(1)}(t)_{ii} = \frac{8\pi c}{\hbar}\int_0^t dt_2 \sum_j F_{ij}F_{ji} \times \sin\omega_{ij}(t-t_2)\sum_n \alpha_n P_n(t_2), \tag{13a}$$

$$F^{(1)}(t)_{ik} = \frac{4\pi c}{\hbar}\epsilon^{i\omega_{ik}t} \times \int_0^t dt_2\, S(t,t_2)\sum_n \alpha_n P_n(t_2), \tag{13b}$$

where

$$S(t,t_2) = -i\sum_j F_{ij}F_{jk}(\epsilon^{i\omega_{kj}(t-t_2)} - \epsilon^{i\omega_{ji}(t-t_2)}). \tag{13c}$$

$\mathcal{H}_d^{(1)}$ is strictly off-diagonal,

$$\mathcal{H}_d^{(1)}(t)_{ik} = \frac{4\pi c}{i\hbar}\hbar\omega_{ik}F_{ik}\int_0^t dt_1\, \epsilon^{i\omega_{ik}t_1}\sum_n \alpha_n P_n(t_1) = (\mathcal{H}_d^{(1)})^*_{ki}. \tag{14}$$

The asterisk denotes a complex conjugate.

The $T$-matrix which restores $F$ to unperturbed form, $TF(t)T^\dagger = F^{(0)}(t)$, and $\mathcal{H}_d$ to the diagonal form $\mathcal{H}_d^{(0)}$ within terms proportional to the $\alpha_n$ has the elements

$$T_{ii} = 1, \qquad T_{ij} = \frac{4\pi c}{i\hbar}F_{ij}\int_0^t dt_1\, \epsilon^{i\omega_{ij}t_1}\sum_n \alpha_n P_n(t_1) = -T^*_{ji}. \tag{15}$$

Here $T^\dagger T = I$, diagonal, with error terms proportional to the $(\alpha_n t)^2$. One can verify the fact that the first-order elements of $TF^{(0)}T^\dagger$ cancel those of $TF^{(1)}T^\dagger \cong F^{(1)}$, and the first-order element $(T\mathcal{H}_d^{(0)}T^\dagger)_{ik}$ cancels $(T\mathcal{H}_d^{(1)}T^\dagger)_{ik}$.

### 3. EQUATIONS OF MOTION FOR THE CAVITY MODE OPERATORS

The Hamiltonian for the complete system, including the interaction energy (5), is

$$\mathcal{H}_{tot} = \int dv \sum_{mn}\left(2\pi c^2 P_m u_m \cdot P_n u_n + \frac{1}{8\pi}Q_m\nabla\times u_m \cdot Q_n \nabla \times u_n\right) + \mathcal{H}_{fd} + \mathcal{H}_d. \tag{16a}$$

By orthogonality of the mode patterns this is

$$\mathcal{H}_{tot} = \sum_n\left[2\pi c^2 P_n^2 + \frac{\omega_n^2}{8\pi c^2}Q_n^2\right] + 4\pi c\sum_n \alpha_n P_n F + \mathcal{H}_d. \tag{16b}$$

The loss-averaged commutators to be derived self-consistently are written

$$\langle[P_n(t), P_k(t)]\rangle_d = i\hbar X_{nk}, \qquad X_{nn} = 0, \tag{17a}$$

$$\langle[P_n(t), Q_k(t)]\rangle_d = -i\hbar Y_{nk}, \qquad Y_{nn} = 1, \tag{17b}$$

$$\langle[Q_n(t), Q_k(t)]\rangle_d = i\hbar Z_{nk}, \qquad Z_{nn} = 0, \tag{17c}$$

The Heisenberg equations of motion yield, for the field operators,

$$\dot P_n(t) = (i\hbar)^{-1}[P_n, \mathcal{H}_{tot}] = -(\omega_n^2/4\pi c^2)Q_n + \sum_{k\neq n}[4\pi c^2 X_{nk}P_k - (\omega_k^2/4\pi c^2)Y_{nk}Q_k] + 4\pi c\sum_k X_{nk}\alpha_k F(t), \tag{18}$$

with a similar equation for $\dot Q_n$.

In integral form this is

$$P_n(t) = P_n^{(0)}(t) + 4\pi c \sum_k \alpha_k X_{nk} \times \int_0^t dt_1 \, [F^{(0)} + F^{(1)}]_{t_1} \cos \omega_n(t - t_1) - \frac{\omega_n}{c} \sum_k \alpha_k Y_{nk} \int_0^t dt_1 \, [F^{(0)} + F^{(1)}]_{t_1} \times \sin \omega_n(t - t_1) + \int_0^t dt_1 \, V_{1n}(t_1) \times \cos \omega_n(t - t_1) - \frac{\omega_n}{4\pi c^2} \times \int_0^t dt_1 \, V_{2n}(t_1) \sin \omega_n(t - t_1), \tag{19a}$$

with a similar equation for $Q_n$ in which

$$\dot{P}_n^{(0)} = -(\omega_n^2/4\pi c^2) Q_n^{(0)} \qquad \dot{Q}_n^{(0)} = 4\pi c^2 P_n^{(0)}, \tag{19b}$$

$$V_{1n} = \sum_{k \neq n} \left[ 4\pi c^2 X_{nk} P_k - \frac{\omega_k^2}{4\pi c^2} Y_{nk} Q_k \right], \tag{19c}$$

$$V_{2n} = \sum_{k \neq n} \left[ 4\pi c^2 Y_{nk} P_k + \frac{\omega_k^2}{4\pi c^2} Z_{nk} Q_k \right]. \tag{19d}$$

Now we average $P_n$ and $Q_n$ of direct physical interest over the loss mechanism ensemble; we examine first the term

$$\int_0^t dt_1 \, \langle F^{(1)}(t_1) \rangle_d \sin \omega_n(t - t_1) = \int_0^t dt_1 \sum_k \langle k | \, \rho_d F^{(1)} \, | k \rangle \sin \omega_n(t - t_1) = \int_0^t dt_1 \sum_k (\rho_d)_{kk} F^{(1)}(t_1)_{kk} \sin \omega_n(t - t_1). \tag{20}$$

$F^{(1)}$ by (11) and (13) is the special case of an operator in $\Theta_f \Theta_d$ form (f—field; d—dissipation mechanism). We show in Appendix A, Eq. (A5), that only the diagonal $\rho_d^{(0)}$ enters into the global expectation value of such an operator. Therefore (20) yields, with reference to (13a),

$$\langle F^{(1)}(t_1) \rangle_d = \frac{8\pi c}{\hbar} A \sum_i \epsilon^{-E_i/kT} \sum_j F_{ij} F_{ji} \times \int_0^t dt_2 \sin \omega_{ij}(t_1 - t_2) \sum_n \alpha_n P_n(t_2) \tag{21}$$

and this yields, in a straightforward fashion[2] for (20),

$$\frac{\omega_n}{c} \int_0^t dt_1 \, \langle F^{(1)}(t_1) \rangle_d \sin \omega_n(t - t_1) = \beta_0 \sum_l \alpha_l \int_0^t dt_1 \, P_l(t_1) \cos \omega_n(t - t_1). \tag{22}$$

Similarly, the other $F^{(1)}$ term appearing in the equation of motion for $Q_n$ analogous to (19) evaluates as

$$-\frac{\omega_n}{c} \int_0^t dt_1 \, \langle F^{(1)}(t_1) \rangle_d \cos \omega_n(t - t_1) = \beta_0 \sum_l \alpha_l \int_0^t dt_1 \, P_l(t_1) \sin \omega_n(t - t_1) \tag{23}$$

Upon substituting (22) and (23) into (19a) and its mate for $Q_n$ and differentiating twice so as to form second-order differential equations we obtain, with $\beta_n = \beta_0 \alpha_n \alpha_n$, $\sum_k \alpha_k X_{nk} = 0$ and $\sum_k \alpha_k Y_{nk} = \alpha_n$ as discussed in Appendix C, and the relation

$$(\omega_k^2/4\pi c^2) Z_{nk} = 4\pi c^2 X_{nk}$$

from symmetry considerations,

$$\ddot{P}_n(t) + \beta_n \dot{P}_n + \omega_n^2 P_n = -(\omega_n^2/c)\alpha_n F^{(0)}(t) - 4\pi c^2 \beta_0 \sum_k \alpha_k^2 X_{nk} P_k - \alpha_n \beta_0 \sum_{l \neq n} \alpha_l \dot{P}_l - \beta_0 \sum_{k \neq n} \alpha_k^2 Y_{nk} \dot{P}_k - (\omega_n^2/4\pi c^2) V_{2n}(t) + \dot{V}_{1n}(t), \tag{24}$$

$$\ddot{Q}_n(t) + \beta_n \dot{Q}_n + \omega_n^2 Q_n = 4\pi c \alpha_n (\dot{F}^{(0)} + \beta_n F^{(0)}) + \beta_0 (4\pi c^2/\omega_n)^2 \sum_k \alpha_k^2 X_{nk} \cdot \times (\dot{P}_k + \beta_n P_k) - 4\pi c^2 \alpha_n \beta_0 \sum_{l \neq n} \alpha_l P_l - 4\pi c^2 \beta_0 \sum_{k \neq n} \alpha_k^2 Y_{nk} P_k + 4\pi c^2 V_{1n} + \dot{V}_{2n}. \tag{25}$$

The Laplace transforms of these equations enable a self-consistent evaluation of the $X$- and $Y$-commutators in the next section.

## 4. THE SELF-CONSISTENT COMMUTATORS

In this section the results and the methods of obtaining them are merely summarized. The Laplace transforms of (24) and (25), with

$$P_n(s) = \int_0^\infty P_n(t) \epsilon^{-st} \, dt,$$

for example, yield four groups of terms in each equation, consisting of (a) initial-value terms labeled by superscript (a) below, (b) terms proportional to $F^{(0)}(s)$, (c) terms proportional to $\beta_0$, and (d) terms involving $V_{1n}(s)$ and $V_{2n}(s)$. A typical commutator

in the time domain can then be formed, and has the appearance of

$$[P_n(t_1), P_m(t_2)] = \left(\frac{1}{2\pi i}\right)^2 \int_{\epsilon-i\infty}^{\epsilon+i\infty} ds \int ds' \times \frac{[N_n(s), N_m(s')]\epsilon^{st_1+s't_2}}{(s-s_n)(s-s_n^*)(s'-s_m)(s'-s_m^*)}, \quad (26)$$

where $N$ denotes an involved collection of the above groups of terms. Upon taking an average over the loss mechanism the $N_n^{(a)}N_m^{(a)}$ product of initial-value terms is retained, the $N_n^{(a)}N_m^{(b)} + N_n^{(b)}N_m^{(a)}$ terms average to zero, the $N^{(a)}N^{(c)}$ terms are quadratic in $\beta_0$ and therefore neglected, the $N^{(b)}N^{(b)}$ expectation is nonzero and will be evaluated below along with the $N^{(b)}N^{(c)}$ contribution, and the $N_n^{(c)}N_m^{(c)}$ terms are quadratic in $\beta_0$. The loss-mechanism expectations of terms involving $V_{1n}$ and $V_{2n}$ in (26) are zero for reasons presented in Sec. 8 following (56).

We now show that the $N^{(b)}N^{(c)}$ terms in (26), the ones proportional to both $\beta_0$ and $F^{(0)}$, are negligible. Examination of (19), (20), (22), and (23) shows that the $\beta_0$-terms in $P$ arise from averaging the $F^{(1)}$ terms. So we examine averages like

$$\left\langle \int_0^t dt_1 \int_0^{t'} dt_2\, F^{(0)}(t_1) \begin{matrix}\sin\\ \cos\end{matrix} \omega_n(t-t_1) F^{(1)}(t_2) \times \begin{matrix}\sin\\ \cos\end{matrix} \omega_m(t'-t_2) \right\rangle_d$$

in $\langle [P_n(t), P_m(t')]\rangle_d$. This average depends on

$$\int_0^{t'} dt_2 \langle F^{(0)}(t_1)F^{(1)}(t_2)\rangle_d \begin{matrix}\sin\\ \cos\end{matrix} \omega_m(t'-t_2) = \sum_{ik} \int_0^{t'} dt_2\, (\rho_d)_{kk} F_{ki} \epsilon^{i\omega_k t_1} \times F^{(1)}(t_2)_{ik} \begin{matrix}\sin\\ \cos\end{matrix} \omega_m(t'-t_2), \quad (27)$$

because only the diagonal elements of $\rho_d^{(0)}$ are involved. Each $ik$ term is zero for the following reason. If (13b) is substituted into (27) and the integration carried out with respect to the argument of $F_{ik}^{(1)}$ after inverting the order of integration four sums arise, over the $j$ index in $S(t, t_2)$. One of the sums has the form

$$\epsilon^{-i\omega_m t} \sum_j F_{ij}F_{jk} \frac{\epsilon^{i(\omega_{kj}+\omega_n)t} - \epsilon^{i(\omega_{kj}+\omega_n)t_2}}{i(\omega_{kj}+\omega_n)},$$

in which the predominant terms are those for which $\omega_{kj} + \omega_n \cong 0$. This group of terms will cancel a similar group and the other two sums will also cancel *provided* the loss mechanism is homogeneous to the extent that

$$F(\omega_i, \omega_k + \omega_n)F(\omega_j, \omega_j - \omega_n) \cong F(\omega_i - \omega_n, \omega_k)F(\omega_j + \omega_n, \omega_j). \quad (28)$$

So that $N^{(b)}N^{(c)}$ terms are zero in (26).

We have made the assumption that the field operators commute with the coupling operator $F = F^{(0)} + F^{(1)}$ on a loss mechanism ensemble basis. This will be so, self-consistently, if we can show in (19) that

$$\int_0^t dt_1 \langle [F^{(0)}(t_1) + F^{(1)}(t_1), F^{(0)}(t) + F^{(1)}(t)]\rangle_d \begin{matrix}\sin\\ \cos\end{matrix} \omega_n(t-t_1) \cong 0, \quad (29)$$

with $F^{(1)}$ given by (13). Neglecting higher-order field and loss mechanism correlations this will be true according to detailed evaluation of the zero-order portion and for the same reason each first-order term of (27) was zero.

The conclusion that the $N^{(b)}N^{(c)}$ terms are zero in (26) is supported by substituting the predominant term for each $P_k(s)$ into a typical term in the integrand, as obtained from the Laplace transform of (24), converting the summation into a complex plane integration, whereupon the result vanishes as $s$ or $s'$ approaches the relevant pole in (26).

The predominant terms in commutator (26), then, are the $N^{(a)}N^{(a)}$ initial-value terms and the $N^{(b)}N^{(b)}$ terms proportional to $\langle [F^{(0)}(s), F^{(0)}(s')]\rangle_d$. Straightforward evaluation according to the work in Appendix B yields, for the net commutator

$$\langle [P_n(t_1), P_m(t_2)]\rangle_d = -\frac{i\hbar\beta_0\alpha_n\alpha_m\omega}{4\pi c^2} \epsilon^{-\beta_n\tau/2} \frac{\beta_{nm}\sin\omega_n'\tau + \omega_{nm}'\cos\omega_n'\tau}{D_{nm}}, \quad \text{if } \tau = t_1 - t_2 \geq 0, \quad (30a)$$

$$= -\frac{i\hbar\beta_0\alpha_n\alpha_m\omega}{4\pi c^2} \epsilon^{\beta_m\tau/2} \frac{\beta_{nm}\sin\omega_m'\tau + \omega_{nm}'\cos\omega_m'\tau}{D_{nm}}, \quad \text{if } \tau \leq 0. \quad (30b)$$

with

$$\beta_{nm} = (\beta_n + \beta_m)/2$$
$$\omega_{nm}' = \omega_n' - \omega_m', \qquad \omega_n' = \mathrm{Im}\,(s_n) = \omega_n[1 - (\beta_n/2\omega_n)^2]^{\frac{1}{2}}, \quad (30c)$$
$$D_{nm} = \beta_{nm}^2 + (\omega_{nm}')^2.$$

These expressions were derived self-consistently in the following way: the $N^{(b)}N^{(b)}$ portion of the commutator contains exponentially decaying terms; the $N^{(a)}N^{(a)}$ portion contains initial-value terms like

$\langle[P_n(0),\ P_m(0)]\rangle_d$ which decay as $\epsilon^{-(\beta_n t_1+\beta_m t_2)/2}$. By choosing these latter commutators to have the thermal values specified by (30) for $t_1 = t_2 = 0$ all the transient terms cancel in $\langle[P_n(t_1),\ P_m(t_2)]\rangle_d$ and this commutator for $t_1 = t_2 = t$ does have the value specified at $t = 0$.

Evaluation of the $[P_n(t_1),\ Q_m(t_2)]$ commutator averaged over loss mechanism space is carried out analogously; the net value is

$$\langle[P_n(t_1),\ Q_m(t_2)]\rangle_d = i\hbar\beta_0\alpha_n\alpha_m\epsilon^{-\beta_n\tau/2}\frac{-\beta_{nm}\cos\omega_n'\tau + \omega_{nm}'\sin\omega_n'\tau}{D_{nm}},$$

$$\text{if}\quad \tau = t_1 - t_2 \geq 0, \tag{31a}$$

$$= i\hbar\beta_0\alpha_n\alpha_m\epsilon^{\beta_m\tau/2}\frac{-\beta_{nm}\cos\omega_m'\tau + \omega_{nm}'\sin\omega_m'\tau}{D_{nm}},$$

$$\text{if}\quad \tau \leq 0. \tag{31b}$$

The commutator $\langle[Q_n(t_1),\ Q_m(t_2)]\rangle_d$ is given by (30) multiplied by the factor $(4\pi c^2/\omega)^2$, $\omega_n \cong \omega_m \cong \omega$.

Equations (30) and (31) illustrate correlations in fluctuations caused by the thermal loss mechanism in the following way. If $\overline{\Delta P_n(t_1)}$ is the average thermal fluctuation during a brief time interval $\Delta t \ll 2\pi/\omega_n$ and $\Delta P_m(t_2)$ is a later thermal fluctuation then the minimum product of these two is given by the magnitude of (30b); here $\overline{\Delta P_n(t_1)}\Delta Q_n(t_1) \cong \hbar$ at the end of the brief interval, consistent with the magnitude of (31) at $\tau = 0$.

Equations (30) and (31) evaluated at $t_1 = t_2 = t$ give the values of $X$, $Y$, and $Z$ for (17) as

$$X_{nk} = -(\omega/4\pi c^2)\beta_0\alpha_n\alpha_k\omega_{nk}'/D_{nk}, \tag{32a}$$

$$Y_{nk} = \beta_0\alpha_n\alpha_k\beta_{nk}/D_{nk}, \qquad \omega \cong \omega_n \cong \omega_k, \tag{32b}$$

$$Z_{nk} = (4\pi c^2/\omega)^2 X_{nk}. \tag{32c}$$

It seems to us these are in error by fractional parts on the order of $\beta_n/\omega_n \cong \beta_k/\omega_k$.

## 5. ANTICOMMUTATORS; THERMAL ENERGY SPECTRUM

The anticommutators such as $\langle[P_n(t_1),\ P_m(t_2)]_+\rangle_d$ averaged over loss mechanism space, where $P_nP_m + P_mP_n = [P_n,\ P_m]_+$, are evaluated according to exactly the same considerations as applied to the commutators. If the anticommutator associated with (26) is formed the contribution of the $V_{1n}$- and $V_{2n}$-terms is zero for the same reasons in Sec. 4 that these terms were negligible in the commutator. The effect of the sign difference between commutator and anticommutator is a removal of the factor $i$ and evaluation of energy terms. The slowly varying transient terms which arise from $\langle[F^{(0)}(t_1),\ F^{(0)}(t_2)]_+\rangle_d$ exactly cancel the thermal portion of the initial-value anticommutators, the net result being

$$\pi c^2\langle[P_n(t_1,\ P_m(t_2)]_+\rangle_d = \tfrac{1}{2}W_{th,nm}(t_1 - t_2) + \epsilon^{-(\beta_n t_1+\beta_m t_2)/2} \times \Big\{\pi c^2[\bar{P}_n(0),\ \bar{P}_m(0)]_+\cos(\omega_n' t_1 - \omega_m' t_2) + \frac{\omega}{4}[\bar{P}_n(0),\ \bar{Q}_m(0)]_+\sin(\omega_n' t_1 - \omega_m' t_2)\Big\} \tag{33a}$$

with

$$W_{th,nm}(t_1 - t_2) = \frac{\hbar\omega}{2}\frac{1+\epsilon^{-\gamma}}{1-\epsilon^{-\gamma}}\beta_0\alpha_n\alpha_m \times \begin{cases}\epsilon^{-\beta_n\tau/2}(\beta_{nm}\cos\omega_n'\tau - \omega_{nm}'\sin\omega_n'\tau)/D_{nm}, \\ \qquad\qquad \tau = t_1 - t_2 \geq 0, \\ \epsilon^{+\beta_m\tau/2}(\beta_{nm}\cos\omega_m'\tau - \omega_{nm}'\sin\omega_m'\tau)/D_{nm}, \quad \tau \leq 0.\end{cases} \tag{33b}$$

The bar denotes the nonthermal component of an operator and $\gamma = \hbar\omega/kT$. The expression for $(\omega^2/16\pi c^2)\langle[Q_n(t_1),\ Q_m(t_2)]_+\rangle_d$, which is the magnetic energy in the $n$th mode for $m = n$ and $t_1 = t_2$, has the same thermal portion and a slightly different nonthermal portion. For reference we also quote the mixed anticommutator

$$\omega\langle[P_n(t),\ Q_m(t)]_+\rangle_{d(\text{thermal})} = -\hbar\omega\frac{1+\epsilon^{-\gamma}}{1-\epsilon^{-\gamma}}\beta_0\alpha_n\alpha_m\omega_{nm}'/D_{nm}. \tag{34}$$

By (16b) the total thermal field energy in the cavity is the sum of the energies in the various modes, where the energy in the $m$th mode is obtained by substituting $m = n$ in (33). That energy is

$$W_{th,mm} = (\hbar\omega_m/2)[1 + 2(\epsilon^{\gamma} - 1)^{-1}] = \hbar\omega_m n \tag{35}$$

the energy of a Bose–Einstein distribution of $n$ photons among quantum states of the mode.

It is difficult to judge the effect of the coupling terms in (24) and (25) on the frequency spectrum of energy without referring to the real correlation function for the electric (or magnetic) field in the time domain. According to Cummings[14] the real correlation function for the electric field is

$$\phi_E(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} E(r,\ t+\tau)\cdot E(r,\ t)\,dt = \tfrac{1}{2}\langle[E(r,\ t+\tau),\ E(r,\ t)]_+\rangle_d \tag{36}$$

[14] F. W. Cummings, Am. J. Phys. 30, 898 (1962).

by the ergodic properties of the loss mechanism. The spectral density of electric energy at point $r$ is

$$I_E(\omega) = (2/\pi)\int_0^\infty \phi_E(\tau)\cos\omega\tau\, d\tau. \quad (37)$$

Evaluation of the $n$th and $m$th mode contributions to $\phi(\tau)$ is

$$[\phi_E(\tau)]_{nm} = \pi c^2\langle[P_n(t+\tau), P_m(t)]_+\rangle_d u_n(r)\cdot u_m(r), \quad (38)$$

and this, together with the thermal anticommutator given by (33), yields the $nm$th mode component of the spectral density,

$$[I_E(\omega)]_{nm} = \frac{\hbar\omega}{4\pi}\frac{1+\epsilon^{-\gamma}}{1-\epsilon^{-\gamma}}\beta_0\alpha_n\alpha_m[\beta_{nm}\beta_n/2 + \omega_{nm}(\omega-\omega_n')]D_{nm}^{-1}[(\beta_n/2)^2+(\omega-\omega_n)^2]^{-1}u_n\cdot u_m, \quad (39)$$

where terms which are smaller by a factor $(\beta/\omega)^2$ have been neglected. Summation over $m$, with $\sum_m \alpha_m\beta_{nm}u_m/D_{nm} \cong \alpha_n u_n\beta_n$, $\sum_m \alpha_m\omega_{nm}\cdot u_m/D_{nm} \cong 0$ according to the properties of these terms described in Appendix C yields, for the spectral electric energy intensity associated with mode $n$ at point $r$,

$$[I_E(\omega)]_n = \frac{\hbar\omega}{8\pi}\frac{1+\epsilon^{-\gamma}}{1-\epsilon^{-\gamma}}\beta_n\left[\left(\frac{\beta_n}{2}\right)^2+(\omega-\omega_n')^2\right]^{-1}u_n^2(r). \quad (40)$$

The total electric intensity at $\omega$ is thus due to the superposition of Lorentzian spectra, each centered on one of the mode resonant frequencies and associated with a mode amplitude of time dependence $\epsilon^{-\beta_n t/2}\cos(\omega_n' t+\varphi_n)$. Integration of (40) over frequency yields precisely the total electric energy implied by (35).

Regarding the symmetry properties of $\phi$, it is easy to see from (33) that a thermal correlation function such as

$$\phi_{r,r';s,s}(\tau) = \tfrac{1}{2}\langle[E_s(r, t+\tau), E_s(r', t)]_+\rangle_d \quad (41)$$

between the same Cartesian components $s$ at both points satisfies the relation

$$\phi_{r,r';s,s'}(\tau) = \phi_{r,r';s,s'}(-\tau), \quad (42)$$

for $s' = s$ because the $nm$th term in the mode summation for $\tau > 0$ is exactly the $mn$th term in the summation for $\tau < 0$. Even if $s' \neq s$, (42) is still true, verified by summing over the appropriate index in (33) according to the rules in Appendix C.

Mehta and Wolf[15] have derived complex correlation tensors for the thermal field in a lossless cavity; the relationship between these tensors and (42) is established from the relation

$$\langle E_s^r(r, t+\tau)E_{s'}^r(r', t)\rangle_d = \mathrm{Re}\,\tfrac{1}{2}\langle E_s(r, t+\tau)E_{s'}^*(r', t)\rangle_d, \quad (43)$$

where $E_s = E_s^r + iE_s^i$, $E_s^r$ being the real $s$ component of the field used in this paper and $E_s^i$ its Hilbert transform. These authors define a complex correlation tensor $\mathcal{E}_{ss'}(r, r', \tau) \propto \langle E_s(r, t+\tau).E_{s'}^*(r', t)\rangle_d$ and describe some of its properties. They also refer to authors who have studied correlation functions in lossless cavities.

## 6. CAUSAL BEHAVIOR OF THE FIELDS; CORRELATION

Consider the commutator $\langle[E_s(r, t), E_{s'}(r', t')]\rangle_d$ for two Cartesian components $s$ and $s'$ of the complete electric field according to (1). From (30b) we obtain, for $t' \geq t$ say,

$$\langle[E_s(r, t), E_{s'}(r', t')]\rangle_d = -4\pi i\hbar\sum_{nm}\omega_m\beta_0\alpha_m\alpha_n\epsilon^{\beta_m\tau/2}[\beta_{nm}\sin\omega_m'\tau + \omega_{nm}'\cos\omega_m'\tau]D_{nm}^{-1}u_{ns}(r)u_{ms'}(r'), \quad (44)$$

with $\beta_{nm}$, $\omega_{nm}'$, and $D_{nm}$ defined in (30c) and $\tau = t - t' \leq 0$. The functional dependence on the $n$th mode is slowly varying except in $D_{nm} = \beta_{nm}^2 + (\omega_n' - \omega_m')^2$. Hence we can change $\alpha_n$, $\beta_n$, and $u_{ns}$ into continuous functions of a complex variable $\Omega$ and evaluate the sum over $n$ by a contour integral in the $\Omega$-plane, according to the residues at the poles $\omega_m' \pm i\beta(\omega_m')$. By the considerations in Appendix C we obtain

$$\langle[E_s(r, t), E_{s'}(r', t')]\rangle_d = -4\pi i\hbar\sum_m \epsilon^{\beta_m\tau/2}\omega_m'\sin\omega_m'\tau u_{ms}(r)u_{ms'}(r'). \quad (45)$$

The sum is closely related to the Green's function in terms of lossless modes for the three-dimensional wave equation,

$$\nabla\times\nabla\times G(r, t; r_0, t_0) + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}G = a_s\delta(r-r_0)\delta(t-t_0), \quad (46)$$

$a_s$ being a unit vector in the $s$ direction. $G$, obtained by a procedure outlined by Morse and Feshbach[16], is

$$G(r, t; r_0, t_0) = \pm c^2\sum_m \frac{\sin\omega_m(t-t_0)}{\omega_m}u_{ms}(r_0)u_m(r), \begin{cases}+, & t \geq t_0,\\ -, & t \leq t_0.\end{cases} \quad (47)$$

[15] C. L. Mehta and E. Wolf, Phys. Rev. 134, A1143 (1964).

[16] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw Hill Book Company, Inc., New York, 1953), p. 849.

If we relabel $r$, $t$, $r_0$, and $t_0$ as $r'$, $t'$, $r$, and $t$, respectively, we can write (45), with $\beta_m = 0$, as

$$\langle [E_s(r, t), E_{s'}(r', t')]_{\beta=0} \rangle_d = \frac{4\pi i\hbar}{c^2} \frac{\partial^2}{\partial(t')^2} G_{s'}(r', t'; r, t). \quad (48)$$

Since $G$ represents a propagation of a disturbance from $r$ to $r'$ at light velocity or less, the electric field components commute at two space–time points outside the light cone.

One may easily verify that $\langle [H_s(r, t), H_{s'}(r', t')] \rangle_d$ is exactly equal to (48) and that the third commutator for $\beta_n = 0$ is

$$\langle [E_s(r, t), H_{s'}(r', t')]_{\beta=0} \rangle_d = \frac{4\pi i\hbar}{c} \frac{\partial}{\partial t'} \nabla' \times G(r', t'; r, t)|_{s'-\text{comp}} \quad (49)$$

## 7. EQUIVALENT HAMILTONIANS

If no molecules or other energy sources are present in the cavity the equation of motion (18) for $P_n(t)$ and its mate for $Q_n(t)$ may be derived from the equivalent Hamiltonian

$$\mathcal{H}_{eq} = \sum_{k,n} \left( 4\pi c^2 X_{nk} P_n Q_k + \frac{\omega_n \omega_k}{8\pi c^2} Y_{nk} Q_n Q_k + 2\pi c^2 Y_{nk} P_n P_k \right) + 4\pi c \sum_n \alpha_n P_n F^{(0)} \quad (50)$$

by the Heisenberg prescription employing the lossless-cavity commutation relations $[P_n, P_m] = 0$, etc., *provided* the dissipation terms proportional to $\beta_0$ are subsequently introduced into the equations of motion. A dissipation term cannot appear in (50) because it would have to be of the form $\beta_0 \sum_{nk} \alpha_n \alpha_k Q_n P_k$ so as to yield the correct dissipation term in the $P_n$ equation obtained by substituting (22) and (23) into (19), but then an incorrect dissipation term would appear in the $\dot{Q}_n$ equation.

The equation (24) for $\ddot{P}_n$ without the terms involving the other modes corresponds to Kemeny's[10] equation of motion (5.8) for an operator $q_-(t) + q_+(t)$ in his notation, $q_\pm$ being the two amplitudes of the vector magnetic potential of one mode on the forward and reverse time tracks, respectively, of a closed cycle $t_2 \to t_1 \to t_2$. Generally the equations of motion will differ and involve different dynamical variables, denoted by the subscripts $\pm$, in the forward as opposed to the backward time direction. $\frac{1}{2}(q_- + q_+)$ is sort of an average mode operator on the forward time segment of the path. Kemeny introduced dissipation terms into the effective Lagrangian and then derived equations of motion for $q_\pm$ by employing Schwinger's action principle, whereby the closed-cycle action between two points $t_1$ and $t_2$ of the path is made an extremum.

An equivalent Hamiltonian in the sense of our paper must be used cautiously if other energy terms appear in the original Hamiltonian because these latter terms yield components in the equations of motion according to the *loss-averaged* commutators, not the lossless-cavity ones. What is more, these equilibrium commutators may have values different from the thermal ones if a perturbing Hamiltonian causes thermal noise power to transfer between modes of widely different resonant frequencies. This happens in the parametric amplifier of the next Section.

## 8. THERMAL NOISE OF A PARAMETRIC AMPLIFIER

First we convert the first-order differential equations for $P_n$ and $Q_n$, obtained by substituting (22) and (23) into (19) and its mate and differentiating once, into creation–annihilation operator form. With the definitions

$$p_n = (4\pi c^2)^{\frac{1}{2}} P_n, \qquad q_n = Q_n/(4\pi c^2)^{\frac{1}{2}}, \quad (51a)$$

$$J = (4\pi)^{\frac{1}{2}} F^{(0)} \quad (51b)$$

such that

$$[p_n, q_n] = -i\hbar, \quad (51c)$$

the result is, from considerations in Appendix C,

$$\dot{p}_n = -\omega_n^2 q_n - \beta_0 \alpha_n \sum_l \alpha_l p_l - \beta_0 \sum_{k \neq n} \alpha_k^2 Y_{nk} p_k + (4\pi c)^{\frac{1}{2}} V_{1n}, \quad (52a)$$

$$\dot{q}_n = p_n + \alpha_n J + \beta_0 \frac{4\pi c^2}{\omega_n^2} \sum_k \alpha_k^2 X_{nk} p_k + V_{2n}/(4\pi c^2)^{\frac{1}{2}}. \quad (52b)$$

$V_{1n}$ and $V_{2n}$ are defined in (19). Upon defining the creation and annihilation operators in the usual way,

$$a_n^\dagger = \frac{p_n + i\omega_n q_n}{(2\hbar\omega_n)^{\frac{1}{2}}}, \qquad a_n = \frac{p_n - i\omega_n q_n}{(2\hbar\omega_n)^{\frac{1}{2}}}, \quad (53)$$

the commutation relations from (17), which we show are somewhat inconsistent, are

$$\langle [a_n, a_m] \rangle_d = -\langle [a_n^\dagger, a_m^\dagger] \rangle_d = \frac{1}{2(\omega_n \omega_m)^{\frac{1}{2}}} \beta_0 \alpha_n \alpha_m \frac{\beta_{nm} \omega'_{nm}}{D_{nm}}, \quad (54a)$$

$$\langle [a_n, a_m^\dagger] \rangle_d = -i\beta_0 \alpha_n \alpha_m / (\omega'_{nm} - i\beta_{nm}). \quad (54b)$$

Here the magnitude of (54a) is always far less than that of (54b).

With these relations we introduce the simple parametric energy term of Wagner and Hellwarth[11]

$$\mathcal{H}_{\text{par}} = \hbar\beta_c \sum_{n,m} (a_{1n}^\dagger a_{2m}^\dagger \epsilon^{-i\omega_p t} + a_{1n} a_{2m} \epsilon^{i\omega_p t}) \qquad (55)$$

where 1 denotes a mode of resonant frequency near the signal frequency $\omega_1$ of interest and 2 denotes a mode of resonant frequency near the idle frequency $\omega_2$. The pump frequency is $\omega_p = \omega_1 + \omega_2$. Upon converting (52) into $a_n - a_n^\dagger$ form, in which the $\beta_0$-terms (representing dissipation through perturbation of the loss mechanism) are introduced symmetrically, and with consistent neglect of commutators of the form (54a) we obtain

$$\dot{a}_{1n} + \tfrac{1}{2}\beta_{1n} a_{1n} + i\omega_{1n} a_{1n} = -i(\omega_1/2\hbar)^{\frac{1}{2}} \alpha_{1n} J + \sum_{k \neq n} (4\pi c^2 X_{1,nk} - i\omega_{1k} Y_{1,nk}) \times a_{1k} - i\beta_c \sum_k \frac{-i\beta_0 \alpha_{1n}\alpha_{1k}}{\omega'_{1,nk} - i\beta_{1,nk}} (\sum_m a_{2m}^\dagger) \epsilon^{-i\omega_p t}. \qquad (56a)$$

Several $\beta_0$-terms have been neglected by the approximation $\beta_n/\omega_n \ll 1$. Also

$$\dot{a}_{1n}^\dagger + \tfrac{1}{2}\beta_{1n} a_{1n}^\dagger - i\omega_{1n} a_{1n}^\dagger = i(\omega_1/2\hbar)^{\frac{1}{2}} \alpha_{1n} J + \sum_{k \neq n} (4\pi c^2 X_{1,nk} + i\omega_{1k} Y_{1,nk}) \times a_{1k}^\dagger - i\beta_c \sum_k \frac{i\beta_0 \alpha_{1n}\alpha_{1k}}{-\omega'_{1,nk} - i\beta_{1,nk}} (\sum_m a_{2m}) \epsilon^{i\omega_p t}. \qquad (56b)$$

The $X$- and $Y$-terms in (56) correspond to the $V_{1n}$- and $V_{2n}$-terms in (24)–(26); we shall now show that the former terms do not contribute to commutators and anticommutators of the form $\langle [a_{1n}(t_1), a_{1m}(t_2)]_\mp \rangle_d$, etc., with or without parametric coupling. It follows that these terms will not contribute to the spectral intensity of energy. A Laplace transform of (56a) yields, for any $k$,

$$a_{1k}(s) = \frac{K(s)}{s + (\beta_k - \beta_p)/2 + i\omega_k}, \qquad (57)$$

where we have anticipated a self-consistent $\beta_p > 0$ term which will be proportional to $\beta_c^2$ of the parametric coupling, and $K$ turns out to be analytic. This goes into the first summation of (56a) written as

$$-\sum_{k \neq n} \frac{\omega_{1n}\beta_0 \alpha_{1n}\alpha_{1k}}{\omega'_n - \omega'_k - i(\beta_n + \beta_k)/2} a_{1k}(s). \qquad (58)$$

After converting this sum into an integral in the complex $\Omega$ plane, with $\omega_k$ changed to $\Omega$ and the denominator of (58) appearing as

$$[\Omega + i\beta_\Omega/2 - (\omega'_n - i\beta_n/2)][\Omega - i(\beta_\Omega - \beta_p)/2 - is],$$

with the poles at $\Omega = \omega'_k$ guaranteed by the factor $\partial G(\Omega)/\partial\Omega / G(\Omega)$, $G$ having zeros at the $\omega'_k$, it turns out that after integrating over $\Omega$ two of the residue terms cancel at the pole $s = -(3\beta_n - \beta_p)/2 - i\omega'_n$ in the subsequent integration over $s$ and the other two residue terms cancel at the pole $s = -(\beta_n - \beta_p)/2 - i\omega'_n$. Here the approximation $\omega'_n = \omega_n$ is made; presumably a more accurate analysis would account for the higher-order difference between these two (see (30c)). As a result the $X$- and $Y$-terms in (56) do not contribute to $a_{1n}(t)$ and $a_{1n}^\dagger(t)$, thereby justifying the first two statements of this paragraph.

Returning now to (56) we take the Laplace transform of both equations, neglecting the first sum in each, and easily evaluate the last sums by introducing the density of modes per unit frequency interval, $g_{1n}$, in the signal frequency range. We obtain

$$(s + \tfrac{1}{2}\beta_{1n} + i\omega_{1n}) a_{1n}(s) = a_{1n}(0) - i(\omega_1/2\hbar)^{\frac{1}{2}} \alpha_{1n} J(s) - i\pi\beta_c \beta_{1n} g_{1n} \sum_m a_{2m}^\dagger (s + i\omega_p). \qquad (59)$$

The corresponding equation for $a_{1n}^\dagger(s)$ is obtained from this by changing $i$ to $-i$ everywhere and $a_{2m}^\dagger$ to $a_{2m}$. Notice that, since $J(s^*) = J^*(s)$, we have $a_{1n}(s^*) = [a_{1n}^\dagger(s)]^*$ and similarly for $a_{2m}(s)$ for any $s$.

If we take the Wagner–Hellwarth Eq. (13) without their $c_{sj0}$ and $c_{ij0}^\dagger$ sums, move their $\beta^2 a_s$ term to the right side and then sum over the idle modes we have essentially our (59), in which $a_{2m}^\dagger(s + i\omega_p)$ is replaced by its equivalent to (59). We now discuss the properties of the general solution by treating the case of one signal mode at frequency $\omega_1$ and one idle mode at frequency $\omega_2$, with $\omega_p = \omega_1 + \omega_2$. The $\sum_{k \neq n}$ sums in (56) are noncontributory; the second sums reduce to $\pm 1$, respectively, and so a Laplace transformation yields

$$a_1(s) = \frac{N(s)}{D(s)} [a_1(0) - i(\omega_1/2\hbar)^{\frac{1}{2}} \alpha_1 J(s)] - \frac{i\beta_c}{D(s)} [a_2^\dagger(0) + i(\omega_2/2\hbar)^{\frac{1}{2}} \alpha_2 J(s + i\omega_p)] \qquad (60a)$$

with

$$N(s) = s + \tfrac{1}{2}\beta_2 + i\omega_1, \qquad D(s) = (s + \tfrac{1}{2}\beta_1 + i\omega_1) N(s) - \beta_c^2, \qquad (60b)$$

and with a similar expression for $a_1^\dagger(s^*)$ obtained from the changes $a_1$, $a_2^\dagger$, $s$, and $i$ to $a_1^\dagger$, $a_2$, $s^*$, and $-i$ everywhere, respectively.

From (60a) and its mate the thermal energy at signal frequency $\omega_1$ is obtained, for $s = i\omega$, as

$$W_{th}(\omega_1, t) = \tfrac{1}{2}\hbar\omega_1\left(\frac{1}{2\pi i}\right)^2 \int_{\epsilon-i\infty}^{\epsilon+i\infty} ds \int ds^* \, \epsilon^{(s+s^*)t} \times \langle [a_1(s), a_1^\dagger(s^*)]_+ \rangle_d. \quad (61)$$

Routine evaluation shows this expression does reduce to $W_{th}(\omega_1, 0)$ at $t = 0$ and, with consistent neglect of the $\langle [a_1(0), a_2(0)]_+ \rangle$ and $\langle [a_1^\dagger, a_2^\dagger]_+ \rangle$ anticommutators, approaches as $t \to \infty$

$$W_{th}(\omega_1, t = \infty) = \hbar\omega_1[\bar\beta_1\bar\beta_2(\bar\beta_1 + \bar\beta_2)]^{-1} \times [\beta_1(\bar\beta_1\bar\beta_2 + \beta_2^2)n_1 + (2\beta_c)^2\beta_2 n_2] \quad (62a)$$

where

$$\bar\beta_{1,2} = \tfrac{1}{2}(\beta_1 + \beta_2) \pm 2[(\beta_1 - \beta_2)^2/4 + \beta_c^2]^{\frac{1}{2}}, \quad (62b)$$

$$n_{1,2} = \tfrac{1}{2}[1 + \epsilon^{-\gamma(\omega_{1,2})}]/[1 - \epsilon^{-\gamma(\omega_{1,2})}],$$

$n$ being the number of photons in the field at either $\omega_1$ or $\omega_2$. The commutator $\langle [a_1(t), a_1^\dagger(t)] \rangle_d$ at $t = \infty$ is just (62a) but without the $\hbar\omega_1/2$ and $n_{1,2}$ factors.

With regard to the behavior of $W_{th}$ and the commutator during the interim $0 < t < \infty$ it is not hard to show that transient fluctuations can be considerable if, for example, the decay constant $\bar\beta_2$ in the presence of parametric coupling is $\cong 0$ and the system is highly oscillatory. The mathematical implication here that the transient fluctuations occur even if the system starts from parametric equilibrium must be interpreted operationally. That is, in order to determine the initial thermal energy in a short time we must measure the system, thereby perturbing the initial state so that in almost all cases the system does not start exactly from the state of parametric equilibrium.

This simplification of the general case indicates that the transient noise behavior of the parametric amplifier is more complicated than first supposed, and somewhat inconsistent, for we assumed time-independent commutation relations (54) but derived a time-dependent one with the equilibrium value mentioned above. For small parametric coupling we have a good first-order answer. But as a matter of principle the transient behavior of the parametric amplifier, as averaged over the loss mechanism ensemble, depends on a solution to the operator equations of motion which depend on time-dependent commutators obtained self-consistently from the time-dependent operators themselves. We have not carried out such a solution even for the above-described simple model, in which the slow variation of the commutators could be utilized. Saying it another way, the evolution of an amplifier from thermal equilibrium to parametric equilibrium proceeds according to nonlinear differential equations with time-varying commutator coefficients.

We have implied all along that the loss-averaged system described is a statistical representation of a sub-ensemble of "global" systems, and the former must yet be averaged over field coordinates with respect to the field density matrix.

## APPENDIX A. EXPECTATIONS OF VARIOUS OPERATORS

We show that restoration of the loss mechanism operators $F$ and $\mathcal{H}_d$ to their unperturbed off- and on-diagonal forms at the end of a short-time interval by the appropriate IS transformation (which may have vector nature according to the footnote 12) causes a transformation of the density matrix such that $\rho_d$ remains diagonal in loss mechanism space and $\rho_f$ of field space changes only by terms *quadratic* in the time.

Consider $\langle \rho_g \mathcal{O}_d \rangle = \mathrm{Tr}\,(\rho_g \mathcal{O}_d)$, $\rho_g$ being the global density matrix of direct product form and $\mathcal{O}_d$ being an operator which is explicitly a function of loss mechanism coordinates. Upon applying the IS transformation $\mathcal{O}_d^{(0)} = T_d \mathcal{O}_d T_d^\dagger$, $T_d$ being a function of the field according to (15), we can write for the global expectation after any short time interval

$$\langle \rho_g \mathcal{O}_d \rangle = \langle \rho_f \times \rho_d^{(0)} \mathcal{O}_d \rangle = \langle T_d \rho_f \times \rho_d^{(0)} T_d^\dagger \mathcal{O}_d^{(0)} \rangle = \langle T_d \rho_f T_d^\dagger \rangle_f \langle \rho_d^{(0)} \mathcal{O}_d^{(0)} \rangle_d. \quad (A1)$$

The last equality is justified by relating the matrix elements of $\rho_{g,f,d}$ as

$$\langle f'd| \, \rho_g \, |f''d'\rangle = \langle f'| \, \rho_f \, |f''\rangle \langle d| \, \rho_d \, |d'\rangle. \quad (A2)$$

From (A1) it follows that the loss mechanism expectation alone is

$$\langle \rho_g \mathcal{O}_d \rangle_d = T_d \rho_f T_d^\dagger \langle \rho_d^{(0)} \mathcal{O}_d^{(0)} \rangle_d. \quad (A3)$$

The field average in (A1) is nearly $\langle \rho_f \rangle$ since $T_d$ is unitary within terms quadratic in the time.

Consider now $\langle \rho_g \mathcal{O}_f \rangle$, a global expectation of an explicit field operator. There is no advantage in transforming by $T_d$ so we write

$$\langle \rho_g \mathcal{O}_f \rangle = \langle \rho_f \times \rho_d \mathcal{O}_f \rangle = \langle \rho_f \mathcal{O}_f \rangle_f \quad (A4)$$

because of $\mathrm{Tr}\,(\rho_d) = 1$ and (A2) for $d' = d$.

By (A2) we may verify that the global expectation value of a mixed operator is, after an IS transformation,

$$\langle \rho_g \mathcal{O}_f \mathcal{O}_d \rangle = \langle T_d \rho_g \mathcal{O}_f T_d^\dagger \mathcal{O}_d^{(0)} \rangle = \langle T_d \rho_f \mathcal{O}_f T_d^\dagger \rangle_f \langle \rho_d^{(0)} \mathcal{O}_d^{(0)} \rangle_d. \quad (A5)$$

The field average in (A5) in $\langle \rho_f \mathcal{O}_f \rangle_f$ with error quadratic in the (short) time.

If molecules were present the above relations would be valid except for (A2) because all the elements of $\rho_g$ would be defined with respect to $\rho_f \times \rho_d \times \rho_m$ as

$$\sum_m \langle f'md|\ \rho_g\ |f''md'\rangle = \langle f'|\ \rho_f\ |f''\rangle\langle d|\ \rho_d\ |d'\rangle \quad \text{(A6)}$$

in place of (A2). In addition, a new IS transformation of molecular variables by $T_m$, also a function of the field, back to unperturbed form implies

$$\langle \rho_g \mathcal{O}_m\rangle = \langle T_m \rho_f \times \rho_d \times \rho_m T_m^\dagger \mathcal{O}_m^{(0)}\rangle = \langle T_m \rho_f T_m^\dagger\rangle_f \langle \rho_m^{(0)} \mathcal{O}_m^{(0)}\rangle_m \quad \text{(A7)}$$

provided

$$\sum_d \langle f'md|\ \rho_g\ |f''m'd\rangle = \langle f'|\ \rho_f\ |f''\rangle\langle m|\ \rho_m\ |m'\rangle. \quad \text{(A8)}$$

## APPENDIX B. EVALUATION FOR TERMS OF $\langle [P_n(t_1),\ P_m(t_2)]\rangle_d$

Consider the integral

$$I_1 = \left(\frac{1}{2\pi i}\right)^2 \int_{\epsilon-i\infty}^{\epsilon+i\infty} ds \int_{\epsilon-i\infty}^{\epsilon+i\infty} ds' \times \frac{\epsilon^{st_1+s't_2}}{(s-s_n)(s-s_n^*)(s'-s_m)(s'-s_m^*)} \times \langle [F^{(0)}(s),\ F^{(0)}(s')]\rangle_d \quad \text{(B1)}$$

in which

$$\langle [F^{(0)}(s),\ F^{(0)}(s')]\rangle_d = \int_0^\infty dt_3 \int_0^\infty dt_4\ \epsilon^{-st_3-s't_4}\langle [F^{(0)}(t_3),\ F^{(0)}(t_4)]\rangle_d. \quad \text{(B2)}$$

The last expectation value is given by (10) with respect to $\rho_d^{(0)}$. Integration over $t_4$ and then $t_3$ of half this commutator yields

$$I_2 = \int_0^\infty dt_3 \int_0^\infty dt_4\ \epsilon^{-st_3-s't_4}\langle F^{(0)}(t_3)F^{(0)}(t_4)\rangle = (4c^2\hbar\beta_0/\omega)(4\pi c)^{-2}(1-\epsilon^{-\gamma})^{-1}\int_0^\infty d\omega'' \times \left[\frac{\epsilon^{-\gamma}}{(s-i\omega'')(s'+i\omega'')} + \frac{1}{(s+i\omega'')(s'-i\omega'')}\right]. \quad \text{(B3)}$$

Integration of this over $s'$ by means of the residues to the left of the $s'$ contour yields

$$I_3 = \int_{\epsilon-i\infty}^{\epsilon+i\infty} ds' \frac{\epsilon^{s't_2}}{(s'-s_m)(s'-s_m^*)} I_2(s, s') = (4c^2\hbar\beta_0/\omega)(4\pi c)^{-2}(1-\epsilon^{-\gamma})^{-1} \times 2\pi i \int_0^\infty d\omega'' \left[\frac{\epsilon^{-\gamma}}{s-i\omega''}\left(\frac{\epsilon^{s_m t_2}}{(s_m-s_m^*)(s_m+i\omega'')} + \frac{\epsilon^{s_m^* t_2}}{(s_m^*-s_m)(s_m^*+i\omega'')} + \frac{\epsilon^{-i\omega'' t_2}}{(i\omega''+s_m)(i\omega''+s_m^*)}\right)\right.$$
$$\left. + \frac{1}{s+i\omega''}\left(\frac{\epsilon^{s_m t_2}}{(s_m-s_m^*)(s_m-i\omega'')} + \frac{\epsilon^{s_m^* t_2}}{(s_m^*-s_m)(s_m^*-i\omega'')} + \frac{\epsilon^{i\omega'' t_2}}{(i\omega''-s_m)(i\omega''-s_m^*)}\right)\right]. \quad \text{(B4)}$$

We now integrate this with respect to $s$, ignore the small (and rapidly-varying) terms, cancel some terms, neglect terms by the approximation $\beta \ll \omega$, and take $\int d\omega''$ whenever the difference $(t_1 - t_2)$ does not appear. There obtains

$$I_4 = \int_{\epsilon-i\infty}^{\epsilon+i\infty} ds \frac{\epsilon^{st_1}}{(s-s_n)(s-s_n^*)} I_3(s, \omega'') = (c^2\hbar\beta_0/\omega^3)(4\pi c)^{-2}(1-\epsilon^{-\gamma})^{-1}(2\pi i)^3 \times \left[\epsilon^{-\gamma}\left(-\frac{i\epsilon^{s_n t_1 + s_m^* t_2}}{s_n+s_m^*} + \frac{1}{2\pi i}\int_0^\infty \frac{d\omega''\ \epsilon^{i\omega''(t_1-t_2)}}{(i\omega''-s_n)(i\omega''+s_m^*)}\right) - \left(\frac{i\epsilon^{s_n^* t_1 + s_m t_2}}{s_n^*+s_m} + \frac{1}{2\pi i}\int_0^\infty \frac{d\omega''\ \epsilon^{-i\omega''(t_1-t_2)}}{(i\omega''+s_n^*)(i\omega''-s_m)}\right)\right]. \quad \text{(B5)}$$

Evaluation of $I_4$ according to the sign of $t_1 - t_2$ furnishes half of $I_1$; the other half follows readily. When $I_1$ is used to evaluate the $N^{(b)}N^{(b)}$ term of (26) the transient portion which does not depend on $(t_1 - t_2)$ cancels the $N^{(a)}N^{(a)}$ portion, where the initial-value commutators $\langle [P_n(0),\ P_m(0)]\rangle_d$, etc., have the values specified by (32); the net result is precisely (30).

In the equation for $Q_n(s)$ from (25) $F^{(0)}(0)$ will appear; integrals of the form of $I_1$ but involving $\langle [F^{(0)}(s),\ F^{(0)}(0)]\rangle_d$ and $\langle [F^{(0)}(0),\ F^{(0)}(0)]\rangle_d$ evaluate to zero.

## APPENDIX C. EVALUATION OF SUMS OVER THE CAVITY MODES

Consider a sum over the modes of the form

$$S_1 = \sum_n \frac{f(\omega_j) - f(\omega_n)}{D_{jn}} \quad \text{(C1)}$$

where the numerator might represent $X_{jn}$ of (32), for example, and $D_{jn}$ is given by (30c). $S_1$ can be represented by a contour integral in the complex $\Omega$ plane:

$$S_1 = \frac{1}{2\pi i}\oint_c d\Omega \times \frac{f(\omega_j) - f(\Omega)}{[\Omega - \omega_j - i\beta_j(\Omega)][\Omega - \omega_j + i\beta_j(\Omega)]} \frac{G'(\Omega)}{G(\Omega)}, \quad \text{(C2)}$$

where $\beta_j(\Omega) = [\beta_j + \beta(\Omega)]/2$, $G(\Omega)$ has poles simple or multiple at all the $\Omega = \omega_n$, and contour $C$ is shown in Fig. 1. The prime denotes $d/d\Omega$. If $f(\Omega) \to 0$ sufficiently rapidly for large $|\Omega|$ we can continue contour $C$ into contour $C'$ and enclose the poles labled $x$ in the figure. Evaluation of $S_1$ by the residues at these two poles yields

$$S_1 = -\frac{1}{2i\beta_{jj}}\left[(f(\omega_j) - f(\omega_j + i\beta_{jj}))\frac{G'}{G}\bigg|_{\omega_j + i\beta_{jj}} - (f(\omega_j) - f(\omega_j - i\beta_{jj}))\frac{G'}{G}\bigg|_{\omega_j - i\beta_{jj}}\right] \quad \text{(C3)}$$

where $\beta_{jj} = \beta_j$. Upon expanding the terms about $\omega_j$ we get

$$S_1 = \left[\frac{\partial f}{\partial \Omega}\frac{G''}{G'}\right]_{\omega_j}. \quad \text{(C4)}$$

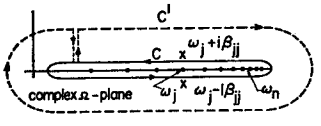


FIG. 1. Distribution of cavity mode resonant frequencies in the complex $\Omega$ plane. Integration around contour $C$ is extended to integration around $C + C'$ and evaluation made in terms of residues at $\omega_j \pm i\beta_{jj}$.

Here

$$\frac{G''}{G'}\bigg|_{\omega_j} = 2\left(\cdots + \frac{m_{j-1}}{\omega_j - \omega_{j-1}} + \frac{m_{j+1}}{\omega_j - \omega_{j+1}} + \cdots\right), \quad \text{(C5)}$$

$m_i$ being the multiplicity of modes at $\omega_i$. We *assume* that the distribution of mode resonant frequencies is quite homogeneous so that $G''/G' \cong 0$ and hence that $S_1 \cong 0$.

Now consider the other type of sum,

$$S_2 = \sum_n \frac{f_e(\omega_j) + f_e(\omega_n)}{D_{jn}}, \quad \text{(C6)}$$

where the numerator might represent $Y_{jn}$ of (32), for example. Evaluation of this by a contour integration yields

$$S_2 = 2f_e(\omega_j)/\beta_{jj}^2 + \frac{1}{i\beta_{jj}}\frac{\partial f_e}{\partial \Omega}\bigg|_{\omega_j}. \quad \text{(C7)}$$

The first term is just the $n = j$ term in (C6). It is reasonable to define $f_e(\Omega)$ to be an impulsive function which is even about each $\omega_n$, in which case the second term in (C7) is zero.

# Cartesian Tensors in Spaces with an Indefinite Metric with Applications to Relativity*

LESLIE L. FOLDY
*Case Institute of Technology, Clevelend, Ohio*
(Received 22 April 1965)

A new formalism for dealing with Cartesian tensor analysis in flat spaces with an indefinite metric with the same ease as in Euclidean spaces is introduced. It avoids the necessity of distinguishing covariant and contravariant indices and the consequent use of the metric tensor in raising and lowering indices; neither does it require the introduction of imaginary coordinates and components of tensors. It is based on the use of a modified Einstein summation convention combined with a modified differentiation with respect to tensor components in such a way that all formal manipulations are essentially identical with those employed in Cartesian tensor analysis in Euclidean spaces. The further introduction of a modified matrix multiplication and a modified definition of a determinant serves to round out the formal analogy with the Euclidean space. The convenience, simplicity, and typographical economy of the new formalism is illustrated by examples drawn from special relativity. The formalism can be trivially generalized to complex linear vector spaces with an indefinite metric.

## INTRODUCTION

CARTESIAN tensor analysis in Euclidean spaces is particularly simple since it is unnecessary to distinguish between covariant and contravariant indices. In a flat space with an indefinite metric, like the Minkowski space of special relativity, the difference between covariant and contravariant components of a vector (in a "Cartesian" coordinate system) involves only differences in sign between some of the components, with associated sign differences for components of higher rank tensors. To keep account of these sign changes one can employ the usual apparatus of general tensor analysis with upper and lower indices distinguishing contravariance and covariance, respectively, and the metric tensor as the medium for raising or lowering indices as necessary. While this is simple in concept it is cumbersome in practice and seems an unnecessarily complicated procedure for doing the required "bookkeeping" on plus and minus signs.

One way of restoring the simplicity of Euclidean space to spaces with indefinite metrics is to introduce imaginary components. The distinction between covariant and contravariant indices then disappears but at the price of dealing with tensors with both real and imaginary components. This also has its annoyances, particularly in quantum mechanics where complex conjugations are frequently necessary, and one must keep in mind which symbols represent real and which imaginary quantities.

I have recently discovered that there exists a simple way of dealing with flat spaces with an indefinite metric with the same ease as in the case of Euclidean space and without the intervention o imaginary quantities.[1] No care need be exercised with respect to the distinction between covariant and contravariant components—one uses one set in a consistent way. Furthermore all manipulations are formally the same as in Cartesian tensor analysis in a Euclidean space, and only the final evaluation of expressions is made according to (slightly) modified rules.

The new scheme rests on two simple modifications of the usual operations of Cartesian tensor analysis:

(i) A modification of the Einstein summation convention for repeated (dummy) indices.

(ii) A modification of the usual partial differentiation with respect to components of a vector or tensor, which in combination with the modified summation convention satisfies all the formal manipulative properties of ordinary partial differentiation.

In conjunction with the above it proves convenient to introduce also matrices and determinants modified to "mesh" with the above changes. We refer to these modified operations and quantities as $M$-summation, $M$-differentiation, $M$-matrices, and $M$-determinants, respectively, where the "$M$" may

* This work was supported in part by the U. S. Atomic Energy Commission.

[1] I have been able to trace the basic idea involved in this formalism to the notation introduced by Feynman [R. P. Feynman, Phys. Rev. 76, 749 (1949), Sec. 3.] and the whole formalism can be considered to be a simple and natural extension of Feynman's notational devices. It is not unlikely that an identical or similar formalism has been published previously and was not revealed in what must of necessity be a relatively superficial literature search; nor were any of the author's colleagues who were queried on this question aware of the possibility of the generalization here presented. In the event that an exposition of this formalism does exist elsewhere, I apologize to its author (or authors) for this oversight and hope that the present publication may serve to bring it into wider acquaintanceship among physicists and mathematicians who may find it useful.

be considered to stand for "metric" or "modified" as desired.

We emphasize that there is nothing profound or mathematically novel in the formalism we describe. We justify its publication purely on the grounds that it is convenient for both computational and didactic purposes and has typographical advantages. We shall present our formalism in rather general form though we have in mind primarily its application to special relativity. We limit our discussion to real spaces, but there appears to be no problem about enlarging its scope to complex linear vector spaces with an indefinite metric.

## CONVENTIONS AND DEFINITIONS

We consider a real linear vector space with $n$ dimensions. A point in the space referred to a particular Cartesian coordinate system is designated $x \equiv x_\mu$ ($\mu = 1, 2, \cdots n$). The metric tensor is then diagonal: $g_{\mu\nu} = 0$ for $\mu \neq \nu$, and its diagonal elements are either $+1$ or $-1$: $g_{\mu\nu} = g_\nu$ for $\mu = \nu$ with $g_\mu = \pm 1$. In Minkowski space, which we shall sometimes use for examples, we shall take[2] $g_1 = g_2 = g_3 = -g_4 = 1$.

### Modified Summation Convention ($M$-summation)

A repeated Greek index is to be summed over all values 1, 2, $\cdots$, $n$, with a coefficient for each term which is $+1$ or $-1$ according as the corresponding $g_\mu$ is $+1$ or $-1$.

As examples for the Minkowski space, we quote

$$a_\mu b_\mu = a_1 b_1 + a_2 b_2 + a_3 b_3 - a_4 b_4,$$

$$\begin{aligned} a_{\mu\nu} b_{\mu\nu} = {} & a_{11} b_{11} + a_{12} b_{12} + a_{13} b_{13} - a_{14} b_{14} \\ & + a_{21} b_{21} + a_{22} b_{22} + a_{23} b_{23} - a_{24} b_{24} \\ & + a_{31} b_{31} + a_{32} b_{32} + a_{33} b_{33} - a_{34} b_{34} \\ & - a_{41} b_{41} - a_{42} b_{42} - a_{43} b_{43} + a_{44} b_{44}. \end{aligned}$$

We shall sometimes write $a \cdot b$ for $a_\mu b_\mu$ and $a^2$ for $a \cdot a = a_\mu a_\mu$.

It follows from our convention on summation that

$$\begin{aligned} g_{\mu\nu} a_\nu &= a_\mu, \\ g_{\mu\nu} a_\mu b_\nu &= a_\mu b_\mu = \mathbf{a} \cdot \mathbf{b}, \\ g_{\mu\nu} g_{\sigma\nu} &= g_{\mu\nu}, \end{aligned} \tag{1}$$

so that $g_{\mu\nu}$ behaves under $M$-summation as the Kronecker symbol $\delta_{\mu\nu}$ does under the ordinary summation, that is, as an "identity".

### $M$-differentiation

We use the usual symbol $\partial/\partial a_\mu$ to represent ordinary partial differentiation, but we now introduce a modified differentiation, $M$-differentiation. Let $\phi(a)$ be a function of the "vector" $a_\mu$. We define

$$\partial_M \phi / \partial a_\mu = \delta_{\mu\nu} (\partial \phi / \partial a_\nu), \tag{2}$$

where the $M$-summation convention is to be employed on the repeated index on the right. Here $\delta_{\mu\nu}$ is the ordinary Kronecker symbol defined by

$$\delta_{\mu\nu} = \begin{cases} 1 & \text{for } \mu = \nu = 1, 2, \cdots n, \\ 0 & \text{otherwise.} \end{cases}$$

This means

$$\partial_M \phi / \partial a_\mu = \begin{cases} \partial \phi / \partial a_\mu & \text{if } g_\mu = +1, \\ -\partial \phi / \partial a_\mu & \text{if } g_\mu = -1. \end{cases} \tag{3}$$

More generally, if we have a function $\phi$ of a "tensor" $t_{\mu\nu} \cdots$, we define the $M$-derivative by

$$\partial_M \phi / \partial t_{\mu\nu} \ldots = \delta_{\mu\alpha} \delta_{\nu\beta} \cdots (\partial \phi / \partial t_{\alpha\beta} \ldots). \tag{4}$$

With the $M$-derivative, the usual rules apply for differentiating sums and products. Furthermore, and this has motivated its definition,

$$\phi(a + da) = \phi(a) + [\partial_M \phi(a) / \partial a_\mu]\, da_\mu, \tag{5}$$

$$\phi(t + dt) = \phi(t) + [\partial_M \phi(t) / \partial t_{\mu\nu} \ldots]\, dt_{\mu\nu} \ldots, \tag{6}$$

and also, if $\phi$ is a function of a vector $\mathbf{b}$ which in turn is a function of a vector $\mathbf{a}$, then

$$\partial_M \phi / \partial a_\mu = (\partial_M \phi / \partial b_\nu)(\partial_M b_\nu / \partial a_\mu). \tag{7}$$

These formulas illustrate the neat "meshing" of $M$-summation with $M$-differentiation to yield results which are formally identical with the corresponding formulas in Euclidean spaces. One has also the particular relations

$$\begin{aligned} \partial_M (a_\nu b_\nu) / \partial a_\mu &= b_\mu, \\ \partial_M (a^2) / \partial a_\mu &= 2 a_\mu, \\ \partial_M a_\mu / \partial a_\nu &= g_{\mu\nu}, \end{aligned}$$

of which the last again exhibits tbe close kinship of the metric tensor in the present formalism with the Kronecker symbol of the conventional formalism.

## TRANSFORMATION OF COORDINATES

We consider now the transformation from one set of "Cartesian" axes in which the coordinates of

[2] Somewhat later we go over to the currently more popular convention of designating the time coordinate by $x_0$ with the metric: $g_0 = -g_1 = -g_2 = -g_3 = 1$.

a point are $x \equiv x_\mu$ to another in which the coordinates of the same point are $x' \equiv x'_\mu$. Such a transformation is a linear transformation

$$x'_\mu = \Lambda_{\mu\nu} x_\nu, \tag{8}$$

($M$-summation implied on the right!) satisfying the condition

$$x'^2 = x'_\mu x'_\mu = \Lambda_{\mu\nu}\Lambda_{\mu\sigma} x_\nu x_\sigma = x_\nu x_\nu = x^2.$$

This defines a rotation of axes and requires the coefficients to satisfy the condition

$$\Lambda_{\mu\nu}\Lambda_{\mu\sigma} = g_{\nu\sigma}, \tag{9}$$

or, defining

$$\begin{aligned} \Lambda^t_{\mu\nu} &= \Lambda_{\nu\mu}, \\ \Lambda^t_{\nu\mu}\Lambda_{\mu\sigma} &= g_{\nu\sigma} = \Lambda_{\nu\mu}\Lambda^t_{\mu\sigma}. \end{aligned} \tag{10}$$

(The interpretation of this last equation as a "matrix" equation will be taken up in the following section.) It follows from these results that

$$x_\nu = x'_\mu \Lambda_{\mu\nu} = \Lambda^t_{\nu\mu} x'_\mu, \tag{11}$$

and

$$\begin{aligned} \partial_\nu x'_\mu &= \Lambda_{\mu\nu} = \Lambda^t_{\nu\mu}, \\ \partial'_\mu x_\nu &= \Lambda^t_{\nu\mu} = \Lambda_{\mu\nu}, \end{aligned} \tag{12}$$

where we have abbreviated

$$\partial_\nu = \partial_M / \partial x_\nu, \qquad \partial'_\mu = \partial_M / \partial x'_\mu.$$

It should be noted that the identity transformation is represented by $\Lambda_{\mu\mu} = g_{\mu\nu}$:

$$x'_\mu = g_{\mu\nu} x_\nu = x_\mu.$$

An infinitesimal transformation (rotation) then has the form

$$\Lambda_{\mu\nu} = g_{\mu\nu} + \epsilon\lambda_{\mu\nu}, \tag{13}$$

where $\epsilon$ is an infinitesimal and condition (9) implies that $\lambda_{\mu\nu}$ is antisymmetric on its indices:

$$\lambda_{\nu\mu} = -\lambda_{\mu\nu} = -\lambda^t_{\nu\mu}. \tag{14}$$

In the special case of Minkowski space, the various inversion transformations are represented by

$$\begin{aligned} \Lambda_{\mu\nu} &= -\delta_{\mu\nu} \quad \text{(space inversion)}, \\ \Lambda_{\mu\nu} &= \delta_{\mu\nu} \quad \text{(time inversion)}, \\ \Lambda_{\mu\nu} &= -g_{\mu\nu} \quad \text{(space–time inversion)}. \end{aligned} \tag{15}$$

If under a change of axes (reference frame) a set of quantities $a_\mu$ transforms according to

$$a'_\mu = \Lambda_{\mu\nu} a_\nu,$$

we define it to be a *vector*; more generally a *tensor* transforms as

$$t'_{\mu\nu\ldots} = \Lambda_{\mu\alpha}\Lambda_{\nu\beta} \cdots t_{\alpha\beta\ldots}, \tag{16}$$

while a *scalar* $\phi$ is left invariant. In particular, let $\phi(x)$ be a scalar function of position so that $\phi'(x') = \phi(x)$; then one readily verifies that $\partial_\mu\phi(x)$ is a vector,[3] i.e.

$$\partial'_\mu \phi'(x') = \Lambda_{\mu\nu}\partial_\nu\phi(x). \tag{17}$$

That $g_{\mu\nu}$, the metric tensor, is indeed a tensor with the same components in all frames follows from

$$g'_{\mu\nu} = \Lambda_{\mu\alpha}\Lambda_{\nu\beta} g_{\alpha\beta} = \Lambda_{\mu\alpha}\Lambda_{\nu\alpha} = g_{\mu\nu}.$$

## $M$-MATRICES AND $M$-DETERMINANTS

We have noted above that Eq. (10) expressing the condition for a rotation of reference frame is similar to a matrix equation; the difference is only that the sum on the repeated index is an $M$-summation rather than an ordinary summation. It is therefore convenient to define *$M$-matrices* (that is, arrays) whose row-by-column multiplication is carried out by $M$-summation instead of ordinary summation. The unit $M$-matrix will in this case be represented by the metric tensor $g_{\mu\nu}$. Successive rotations of reference frame will be associated with $M$-matrix multiplication of the associated $M$- matrices for the individual rotations. Equation (10) asserts that $\Lambda_{\mu\nu}$ is an *orthogonal* $M$-matrix while the totality of orthogonal $M$-matrices will constitute an *$M$-matrix representation* of the group of rotations in the space. The generators of infinitesimal rotations, $-i\lambda_{\mu\nu}$, are then *Hermitian* $M$-matrices.

If we were to write out the unit $M$-matrix as an array we would note that its *ordinary* determinant would not necessarily be $+1$; in fact for the Minkowski space, in particular, it would be $-1$. Hence, the familiar rule that the determinant of a product of matrices is the product of the determinants of the factors would not be valid. To rectify this it is convenient to introduce an *$M$-determinant* of an $M$-matrix such that this rule is regained. This is most simply accomplished by defining the $M$-determinant of an $M$-matrix as the product of its ordinary determinant with the ordinary determinant of the unit $M$-matrix. Thus, if we designate the ordinary determinant of a square array $A_{\mu\nu}$ by $\Delta(A_{\mu\nu})$ and its $M$-determinant by $\Delta_M(A_{\mu\nu})$, then[4]

[3] More generally, the $M$-derivative of a tensor yields a tensor of rank increased by unity.

[4] One finds readily that $\Delta_M(\delta_{\mu\nu}) = \Delta(g_{\mu\nu})$, each being plus or minus one according as the number of negative $g_\mu$ is even or odd.

$$\Delta_M(A_{\mu\nu}) = \Delta(g_{\mu\nu})\Delta(A_{\mu\nu}). \tag{18}$$

In the following section we shall consider the Levi-Civita symbol $\epsilon_{\mu\nu\sigma}\cdots$ in our formalism. One can then easily show that the $M$-determinant of an $M$-matrix $A_{\mu\nu}$ can be defined with its use by the same formula as the ordinary determinant is defined in the conventional formalism:

$$\Delta_M(A_{\mu\nu}) = \epsilon_{\mu\nu\sigma\ldots}A_{1\mu}A_{2\nu}A_{3\sigma}\cdots, \tag{19}$$

where $M$-summation is now to be used on the repeated indices on the right.

With the above definition of the $M$-determinant, all restricted (i.e., proper, orthochronous) Lorentz transformations in Minkowski space have $M$-determinants equal to $+1$ with $\Lambda_{44} \geq +1$; improper, antichronous transformations have $M$-determinant $+1$ but $\Lambda_{44} \leq -1$; improper, orthochronous and proper, antichronous transformations have $M$-determinant $-1$ and $\Lambda_{44} \geq 1$ and $\Lambda_{44} \leq -1$, respectively.

We remark finally that the solution of a set of linear equations of the form

$$A_{\mu\nu}x_\nu = y_\mu, \tag{20}$$

with $M$-summation implied on the left, can be expressed in the familiar determinantal form, that is, $x_\nu$ is the usual ratio of two determinants (either both ordinary or both $M$-determinants) but multiplied by 1 or $-1$ according as $g_\mu$ is plus or minus one.[5] If the equations are homogeneous, that is, the $y_\mu$ are all zero, then the condition for a non-trivial solution of the equations is that $\Delta_M(A_{\mu\nu}) \neq 0$, or equivalently, $\Delta(A_{\mu\nu}) \neq 0$.

There is one point in which a little care is required in dealing with $M$-determinants. The usual definition of the cofactor of an element $A_{\mu\nu}$ of a determinant involves striking out the row and column in which $A_{\mu\nu}$ occurs; this leaves one unclear as to how to evaluate the resultant determinant, which has $n - 1$ rows and columns, in our formalism. A better rule for constructing the cofactor is the following: The cofactor $a_{\mu\nu}$ of an element $A_{\mu\nu}$ of a determinant is obtained by replacing all elements in the $\mu$th row and $\nu$th column by zeros except for the element $A_{\mu\nu}$ itself which is replaced by unity. This is equivalent to the ordinary rule for ordinary determinants but works equally well for $M$-determinants. One then has as usual

$$A_{\mu\nu}a_{\mu\sigma} = g_{\mu\sigma}\Delta_M(A_{\mu\nu}),$$

which allows one to define the reciprocal matrix $A^{-1}_{\mu\nu}$ to the matrix $A_{\mu\nu}$ as

$$A^{-1}_{\mu\nu} = a_{\nu\mu}/\Delta_M(A_{\mu\nu}), \tag{21}$$

(assuming $\Delta_M(A_{\mu\nu}) \neq 0$) with the property

$$A^{-1}_{\mu\nu}A_{\nu\sigma} = A_{\mu\nu}A^{-1}_{\nu\sigma} = g_{\mu\sigma}. \tag{22}$$

## THE LEVI-CIVITA SYMBOL

In view of its prominent role in relativity and in other situations, it is now necessary to determine whether any modification of the Levi-Civita symbol is required to bring it into consistency with our formalism employing $M$-summation. One finds that the usual definition

$$\epsilon_{\mu\nu\ldots} = \begin{cases} +1 \text{ if } \mu\nu\cdots \text{ is an even permutation of } 12\cdots n, \\ -1 \text{ if } \mu\nu\cdots \text{ is an odd permutation of } 12\cdots n, \\ 0 \text{ otherwise,} \end{cases} \tag{23}$$

suffices; it would be very awkward for us if it did not since its usefulness hinges on its being completely antisymmetric on its indices. One can establish its consistency with our definition of the $M$-determinant as exemplified in Eq. (19). Furthermore, considered as a symbol with the same components in every reference frame it transforms as a relative tensor (or tensor density) under coordinate transformations:

$$\epsilon'_{\mu\nu\ldots} = \Delta_M(\Lambda_{\mu\nu})\epsilon_{\mu\nu\ldots}. \tag{24}$$

In the particular instance of Minkowski space it therefore transforms as a tensor under restricted Lorentz transformations, and hence, can be used to form pseudoscalars and pseudovectors from completely antisymmetric tensors of the fourth and third rank, respectively, just as in the conventional formalism.

One important caution with respect to its use must be noted, however, which is a consequence of the fact that it is not a true tensor. Some of the familiar identities which it satisfies in a strict Euclidean geometry are valid only in a slightly modified form. We quote here the correct form of three of these identities from which the character of the others is readily inferred:

$$\begin{aligned} \epsilon_{\mu\nu\sigma\ldots}\epsilon_{\mu\nu\sigma\ldots} &= n!\,\Delta_M(\delta_{\mu\nu}), \\ \epsilon_{\alpha\mu\nu\ldots}\epsilon_{\beta\mu\nu\ldots} &= (n-1)!\,\Delta_M(\delta_{\mu\nu})g_{\alpha\beta}, \\ \epsilon_{\alpha\beta\mu\nu\ldots}\epsilon_{\gamma\delta\mu\nu\ldots} &= (n-2)!\,\Delta_M(\delta_{\mu\nu})[g_{\alpha\gamma}g_{\beta\delta} - g_{\alpha\delta}g_{\beta\gamma}]. \end{aligned} \tag{25}$$

It will be noted that these reduce to the familiar

[5] This follows immediately on noting that Eq. (20) is equivalent to the same set of equations with ordinary summation on the left if each $x_\nu$ for which $g_\nu = -1$ is replaced by its negative.

identities for Euclidean space where $\Delta_M(\delta_{\mu\nu}) = 1$ and $g_{\mu\nu} = \delta_{\mu\nu}$.

We shall terminate our discussion of the $M$-summation formalism at this point. There are other matters worthy of investigation, integral theorems, for example, but by this point it should be clear that the formalism is a coherent one and further elaboration would only consist in paralleling the conventional treatment in the new notation.

## EXAMPLES

We shall conclude our exposition of the $M$-summation formalism by illustrating its application to two familiar situations in special relativity, the vector meson (Proca–Maxwell) field and the Dirac equation, in order to demonstrate its basic simplicity and typographic economy[6] over the usual tensor formalism. In this section we depart from our previous notation for Minkowski space and adhere instead to the current preference of using $x_0$ to represent the time coordinate: $x_\mu = (x_0, x_i, x_2, x_3) = (t, \mathbf{x})$; also we take the metric tensor to have diagonal elements: $g_0 = -g_1 = -g_2 = -g_3 = 1$. Thus,

$$x^2 = x_\mu x_\mu = x_0^2 - x_1^2 - x_2^2 - x_3^2.$$

### Proca–Maxwell Field

We shall use this case as an example of the derivation of field equations from a Lagrangian. We write the four-vector potential as $A_\mu = (\phi, \mathbf{A})$ where $\phi$ is the usual scalar potential and $\mathbf{A}$ the usual vector potential. The field tensor $F_{\mu\nu}$ is then given by

$$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu, \tag{26}$$

with $\partial_\mu$ representing $M$-differentiation; thus,

$$\begin{aligned} \mathbf{E} &= -\text{grad}\,\phi - \partial\mathbf{A}/\partial t = (F_{01}, F_{02}, F_{03}), \\ \mathbf{B} &= \text{curl}\,\mathbf{A} = (F_{23}, F_{31}, F_{12}), \end{aligned} \tag{27}$$

and $F_{\mu\nu}$ is antisymmetric on its indices. We shall derive the field equations in the presence of an external four-vector current source $j_\mu = (\rho, \mathbf{j})$ where $\rho$ is the charge density and $\mathbf{j}$ the current density.

The appropriate Lagrangian density is then

$$\begin{aligned} \mathfrak{L} &= \tfrac{1}{2}(\mathbf{E}^2 - \mathbf{B}^2) + \tfrac{1}{2}\kappa^2(\phi^2 - \mathbf{A}^2) + \mathbf{j}\cdot\mathbf{A} - \rho\phi, \\ &= -\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} + \tfrac{1}{2}\kappa^2 A_\mu A_\mu - j_\mu A_\mu. \end{aligned} \tag{28}$$

The usual variation of the action expressed in terms of $M$-derivatives and with the $M$-summation convention is then

$$\begin{aligned} \delta\int d^4x\mathfrak{L} &= \int d^4x\{[\partial_M\mathfrak{L}/\partial A_\nu]\,\delta A_\nu \\ &\quad + [\partial_M\mathfrak{L}/\partial(\partial_\nu A_\mu)]\delta(\partial_\nu A_\mu)\}, \\ &= \int d^4x\{[\partial_M\mathfrak{L}/\partial A_\mu] \\ &\quad - \partial_\nu[\partial_M\mathfrak{L}/\partial(\partial_\nu A_\mu)]\}\,\delta A_\mu, \end{aligned} \tag{29}$$

whose vanishing yields the Euler–Lagrange equations:

$$\partial_M\mathfrak{L}/\partial A_\mu - \partial_\nu[\partial_M\mathfrak{L}/\partial(\partial_\nu A_\mu)] = 0. \tag{30}$$

With

$$\begin{aligned} \partial_M\mathfrak{L}/\partial A_\mu &= \kappa^2 A_\mu - j_\mu, \\ \partial_M\mathfrak{L}/\partial(\partial_\nu A_\mu) &= -F_{\mu\nu}, \end{aligned} \tag{31}$$

this becomes

$$\partial_\nu F_{\mu\nu} + \kappa^2 A_\mu = j_\mu, \tag{32}$$

or

$$\begin{aligned} \text{div}\,\mathbf{E} + \kappa^2\phi &= \rho, \\ \text{curl}\,\mathbf{B} - \partial\mathbf{E}/\partial t + \kappa^2\mathbf{A} &= \mathbf{j}. \end{aligned} \tag{33}$$

To obtain the canonical energy–momentum density tensor $T'_{\mu\nu}$ in the absence of sources, $j_\mu = 0$, we write the condition that the Lagrangian density involves space–time coordinates only implicitly through the fields by

$$\begin{aligned} \partial_\mu\mathfrak{L} &- [\partial_M\mathfrak{L}/\partial A_\nu]\,\partial_\mu A_\nu \\ &- [\partial_M\mathfrak{L}/\partial(\partial_\nu A_\sigma)]\,\partial_\mu\,\partial_\nu A_\sigma = 0. \end{aligned} \tag{34}$$

Using the Euler–Lagrange equations this may be rewritten as

$$\partial_\nu T'_{\mu\nu} = 0. \tag{35}$$

with

$$T'_{\mu\nu} = [\partial_M\mathfrak{L}/\partial(\partial_\nu A_\sigma)]\,\partial_\mu A_\sigma - g_{\mu\nu}\mathfrak{L}. \tag{36}$$

The symmetric energy–momentum density tensor is obtained from this by adding the tensor

$$T''_{\mu\nu} = \partial_\sigma(F_{\sigma\nu}A_\mu), \tag{37}$$

whose divergence, $\partial_\nu T''_{\mu\nu}$ vanishes identically because of the antisymmetry of $F_{\sigma\mu}$. Straightforward calculation using the field Eqs. (32) with $j_\mu = 0$ and (31) then yields

$$\begin{aligned} T_{\mu\nu} = T'_{\mu\nu} + T''_{\mu\nu} &= -F_{\sigma\mu}F_{\sigma\nu} \\ &+ \kappa^2 A_\mu A_\nu + \tfrac{1}{2}g_{\mu\nu}[\tfrac{1}{2}F_{\alpha\beta}F_{\alpha\beta} - \kappa^2 A_\alpha A_\alpha]. \end{aligned} \tag{38}$$

[6] Even greater typographical economy could be achieved in what follows by using the familiar "comma notation" to represent an $M$-derivative, e.g., $\partial_M A_\mu/x_\nu = A_{\mu,\nu}$.

In the presence of conserved currents

$$\partial_\mu j_\mu = \partial\rho/\partial t + \operatorname{div} \mathbf{j} = 0, \tag{39}$$

so that $\partial_\mu A_\mu = 0$, one finds that, with $T_{\mu\nu}$ given by (38),

$$\partial_\nu T_{\mu\nu} = F_{\mu\nu} j_\nu. \tag{40}$$

### The Dirac Equation

Our discussion of the Dirac equation in the $M$-summation formalism will be limited to the statement of some of the pertinent relations and the proof of its covariance. We shall suppress all spinor indices.[7]

We first write the Dirac equation in its familiar noncovariant form:

$$i\partial\psi/\partial t = (\beta m - i\boldsymbol{\alpha}\cdot\boldsymbol{\nabla})\psi. \tag{41}$$

If we multiply the equation by $-i\beta$ and let

$$\gamma_\mu = (\beta, \beta\boldsymbol{\alpha}), \tag{42}$$

it can be rewritten in the $M$-summation formalism as

$$\gamma_\mu \partial_\mu \psi + im\psi = 0. \tag{43}$$

The $\gamma_\mu$ satisfy the anticommutation relations

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2g_{\mu\nu}. \tag{44}$$

Of these matrices, $\gamma_0$ is Hermitian and $\gamma_1$, $\gamma_2$, $\gamma_3$ are anti-Hermitian; thus

$$\gamma_\mu^+ = \gamma_0\gamma_\mu\gamma_0 = \delta_{\mu\nu}\gamma_\nu. \tag{45}$$

The adjoint spinor is defined as usual by

$$\bar{\psi} = \psi^+\gamma_0. \tag{46}$$

We now consider the covariance of the equation under Lorentz transformations,

$$x'_\mu = \Lambda_{\mu\nu} x_\nu, \tag{47}$$

assuming the Dirac wavefunction to transform locally as usual:

$$\psi'(x') = S(\Lambda)\psi(x), \tag{48}$$

$$\psi(x) = S^{-1}(\Lambda)\psi'(x'). \tag{49}$$

[7] Spinor indices, if explicitly included, would be summed according to the usual convention, not by an $M$-summation.

Then

$$\partial_\mu\psi(x) = S^{-1}\,\partial'_\nu\psi'(x')\,\partial_\mu x'_\nu = \Lambda_{\nu\mu}S^{-1}\,\partial'_\nu\psi'(x'). \tag{50}$$

Substituting (49) and (50) into (43) and multiplying by $S$, one obtains as the condition for covariance

$$S(\Lambda)\gamma_\mu S^{-1}(\Lambda) = \Lambda_{\mu\nu}\gamma_\nu. \tag{51}$$

Under an infinitesimal Lorentz transformation

$$\Lambda_{\mu\nu} = g_{\mu\nu} + \epsilon\lambda_{\mu\nu},$$

with

$$S(\Lambda) = 1 + \epsilon T(\lambda), \tag{52}$$

one easily shows that

$$T(\lambda) = \tfrac{1}{4}\lambda_{\sigma\tau}\gamma_\sigma\gamma_\tau, \tag{53}$$

yields a solution of (51). For the space-inversion transformation where $\Lambda_{\mu\nu} = \delta_{\mu\nu}$, Eq. (51) takes the form

$$S\gamma_\mu S^{-1} = \delta_{\mu\nu}\gamma_\nu, \tag{54}$$

which by virtue of (45) is satisfied by $S = \gamma_0 = S^{-1}$.

## CONCLUSION

We believe that we have demonstrated the utility and convenience of the $M$-summation formalism in dealing with Cartesian tensor analysis in spaces with an indefinite metric. The use of lower indices only not only increases the transparency of many relations, but when dealing with nonsymmetric tensors avoids the need for careful positioning of superscripts and subscripts. The formalism has an advantage over the use of imaginary coordinates in that one need not keep track of the real or imaginary character of vector and tensor components. While facility in the use of $M$-summation and $M$-differentiation requires a little practice, it appears to the author to be well repaid in the convenience of calculation in this formalism.

## ACKNOWLEDGMENT

Finally, I would like to express my appreciation to my class in Field Theory for their patience with errors in plus and minus signs while this formalism was in the process of development.

# Thermodynamic Evolution Equation for a Quantum Statistical Gas

S. FUJITA*
*Department of Physics, Pennsylvania State University, University Park, Pennsylvania*
(Received 22 April 1965)

It is shown by a perturbation diagram method that partial Green's functions satisfy closed evolution equations exact in the thermodynamic limit except for the neglect of the contribution of initial particle correlations. This contribution is negligible for a system at a time $t \gg \tau_0$ (average collision duration) when the main interaction processes are localized in space and time. In particular this makes unnecessary the often imposed assumption of the absence of initial particle correlations for the derivation of the usual Boltzmann equation or its generalizations.

## 1. INTRODUCTION

IN an earlier paper[1] the author demonstrated that for a quantum imperfect gas of *distinguishable* particles a closed evolution equation for the one-body density matrix or Wigner's distribution function can be obtained when the contribution arising from the initial particle correlations is negligible. The present article deals with a quantum gas of *indistinguishable* particles.

The essential difference between the classical- and quantum-statistical systems lies in the symmetry property of a many-body state under the permutation of particle indices. Thus, a quantum statistical interaction process is a process in which particles participate not only with their intermolecular potentials but also with the principle of indistinguishability (Pauli's exclusion principle for fermions). The symmetry property of a many-body state may be most simply taken into account by means of second quantization operators. The momentum–space annihilation and creation operators are introduced in the text though the theory could be developed essentially independent of the momentum representation.

It is found that partial Green's functions $g^>$ and $g^<$, satisfy simultaneous equations exact in the thermodynamic limit, Kadanoff–Baym's equations[2] supplemented by the terms typical of representing contributions arising from the initial particle correlations. It is found that (1) the contribution of the initial particle correlations should die out in a time of the order of the average collision duration $\tau_0$, and that (2) for a system at a time $t \gg \tau_0$, the forementioned equations should reduce to closed equations for $g^>$ and $g^<$, and thus give a good starting point for the discussion of various problems. The finding (1) implies in particular that the assumption of the absence of initial particle correlations[3] (or the factorizability of the initial density into one-body densities[4]) is not required for the derivation of the usual Boltzmann equation and its generalizations. This point was previously argued by Prigogine and his collaborators[4–6] from the study of the $N$-body distribution function.

In the Appendix, M. Fitelson, A. St. Pierre, and the present author discuss the effect of initial particle correlations.

## 2. ONE-BODY DENSITY MATRIX AND ITS MECHANICAL EVOLUTION

Let us consider an imperfect gas characterized by the Hamiltonian[7]

$$H \equiv \sum_i h_0^{(i)} + \lambda \sum_{i>j} v^{(ij)} \tag{2.1}$$
$$= \int_{\mathbf{p}_1} \epsilon_1 a_1^\dagger a_1 + \tfrac{1}{4}\lambda \int_{\mathbf{p}_1}\int_{\mathbf{p}_2}\int_{\mathbf{p}_3}\int_{\mathbf{p}_4} v(12,34) a_1^\dagger a_2^\dagger a_4 a_3,$$

$$v(12,34) \equiv [v(\mathbf{p}_1 - \mathbf{p}_3) \pm v(\mathbf{p}_1 - \mathbf{p}_4)] \times \delta_\Omega^{(3)}(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{p}_3 - \mathbf{p}_4), \tag{2.2}$$

$$\int_{\mathbf{p}} \equiv \frac{(2\pi)^3}{\Omega} \sum_{\mathbf{p}} \to \int d^3p, \quad \text{as} \quad \Omega \to \infty, \tag{2.3}$$

$$\delta_\Omega^{(3)}(\mathbf{p}) \equiv \frac{\Omega}{(2\pi)^3}\delta_{\mathbf{p},0}^{(3)} \to \delta^{(3)}(\mathbf{p}), \quad \text{as} \quad \Omega \to \infty,$$

where the conventional notations are used: the second quantization operators $\{a_1^\dagger, a_2\}$ satisfy the commutation or anticommutation rules according to whether the particles are bosons or fermions; the

* On Leave from l'Universite Libre de Bruxelles, Brussels, Belgium.

[1] S. Fujita, J. Math. Phys. (N. Y.) **6**, 1004 (1965).

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

[3] H. Fröhlich and A. W. B. Taylor, Proc. Phys. Soc. (London) **A83**, 739 (1964).

[4] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962).

[5] R. Brout, Physica **22**, 509 (1956); R. Brout and I. Prigogine, Physica **22**, 621 (1956).

[6] P. Résibois, Physica **27**, 541 (1961).

[7] The following notations are due to Hugenholtz, N. M. Hugenholtz, *The Many-Body Problem*, edited by Dunod (Dunod Cie., Paris, 1959), p. 1.

upper signs are for bosons and the lower signs for fermions.

The momentum density matrix is defined by

$$n(\mathbf{p}_1\mathbf{p}_2, t) \equiv \frac{(2\pi)^3}{\Omega} \mathrm{Tr}\,\{\hat{n}_{12}\rho(t)\}, \tag{2.4}$$

$$\hat{n}_{12} \equiv a_2^\dagger a_1, \tag{2.5}$$

where $\rho(t)$ is the density operator satisfying the Liouville equation

$$i\frac{\partial}{\partial t}\langle n|\ \rho(t)\ |n'\rangle = \sum_{n''}[\langle n|\ H\ |n''\rangle\langle n''|\ \rho(t)\ |n'\rangle - \langle n|\ \rho(t)\ |n''\rangle\langle n''|\ H\ |n'\rangle], \tag{2.6}$$

$$|n\rangle \equiv |n_1 n_2 n_3 \cdots\rangle. \tag{2.7}$$

Following Prigogine and Résibois[6,8] we may introduce a new specification of $n$–$n'$ matrix elements of an operator $A$ by

$$\langle n|\ A\ |n'\rangle = A_{n-n'}\left(\frac{n+n'}{2}\right). \tag{2.8}$$

If one defines a set of numbers $(N, \nu)$ such that

$$n - n' \equiv \nu, \qquad \tfrac{1}{2}(n + n') \equiv N, \tag{2.9}$$

then (2.8) is equivalent to

$$\langle N + \tfrac{1}{2}\nu|\ A\ |N - \tfrac{1}{2}\nu\rangle \equiv A_\nu(N). \tag{2.10}$$

In this specification one can write (2.6) as

$$i\frac{\partial}{\partial t}\rho_\nu(N, t) = \sum_{\nu'}(\nu|\ \mathfrak{K}(N)\ |\nu')\rho_{\nu'}(N, t), \tag{2.11}$$

where $(\nu|\mathfrak{K}(N)|\nu')$ is defined by

$$(\nu|\ \mathfrak{K}(N)\ |\nu') \equiv \eta^{+\nu'}H_{\nu-\nu'}(N)\eta^{-\nu} - \eta^{-\nu'}H_{\nu-\nu'}(N)\eta^{\nu}, \tag{2.12}$$

with $\eta^{\nu'}$ denoting a displacement operator such that

$$\eta^{\pm\nu'}f(N_i) = f(N_i \pm \tfrac{1}{2}\nu_i)\eta^{\pm\nu'} \tag{2.13}$$

for a function $f$ of $N_i$.

Direct calculation yields that

$$(\nu|\ \mathfrak{K}_0(N)\ |\nu') = \sum_i \epsilon_i\nu_i\ \delta_{\nu,\nu'} \equiv \int_{\mathbf{p}_1}\int_{\mathbf{p}_2}\int_{\mathbf{p}_3}\int_{\mathbf{p}_4}(\nu|\ \mathfrak{V}_{12,34}(N)\ |\nu'),$$

$$(\nu|\ \mathfrak{V}(N)\ |\nu') = \int_{\mathbf{p}_1}\int_{\mathbf{p}_2}\int_{\mathbf{p}_3}\int_{\mathbf{p}_4} v(12, 34)\ \delta_{\nu_1',\nu_1-1} \times \delta_{\nu_2',\nu_2-1}\ \delta_{\nu_3',\nu_3+1}\ \delta_{\nu_4',\nu_4+1}\prod_{i\geq 5}\delta_{\nu_i',\nu_i}, \tag{2.14}$$

$$v(12, 34) \equiv v(12, 34)\Big\{\eta^{+\nu_1'+\nu_2'+\nu_3'+\nu_4'}\prod_{k=1}^{4}\delta_{N_k,\frac{1}{2}}\eta^{-\nu_1-\nu_2-\nu_3-\nu_4} - \eta^{-\nu_1'-\nu_2'-\nu_3'-\nu_4'}\prod_{k=1}^{4}\delta_{N,\frac{1}{2}}\eta^{\nu_1+\nu_2+\nu_3+\nu_4}\Big\}. \tag{2.15}$$

Here we see that the elements $(\nu|\mathfrak{K}_0|\nu')$ and $(\nu|\mathfrak{V}|\nu')$ are subject to the same sort of selection rules as $\langle n|\ H_0\ |n'\rangle$ and $\langle n|\ V\ |n'\rangle$ are, i.e. $\mathfrak{K}_0$ is diagonal and $\mathfrak{V}$ has nontrivial elements between those $\nu$ and $\nu'$ which can be reached by the transposition of two particles.

Using (2.4), (2.11), and (2.14) one finds that

$$\left[\frac{\partial}{\partial t} + i(\epsilon_2 - \epsilon_1)\right]n(\mathbf{p}_1\mathbf{p}_2, t) = -i\frac{(2\pi)^3}{\Omega}\sum_{N\nu'\nu}\hat{n}_{12\nu}(N) \times (\nu|\ \mathfrak{V}e^{-i\mathfrak{K}(N)t}\ |\nu')\rho_{\nu'}(N, 0). \tag{2.16}$$

If it is so desired, the term on the right-hand side can be expressed in terms of an integral of the product of the potential and two-body density matrix (Bogoliubov–Born–Green–Kirkwood–Yvon equation[9]). Therefore, the mechanical evolution equation is *not* a *closed* equation for the one-body density matrix $n(\mathbf{p}_1\mathbf{p}_2, t)$. This situation is the same as in the case of classical statistics. In this case it was shown that the foremention integral can be expressed in terms of one-body density matrix in the thermodynamic limit if the effect of initial particle correlation is negligible. Does a similar simplification occur in the quantum statistical case? This question is examined in the following sections.

## 3. DIAGRAM ANALYSIS IN ONE-RESOLVENT FORMALISM

From (2.5) and (2.10)

$$\hat{n}_{12\nu}(N) = \langle N + \tfrac{1}{2}\nu|\ a_2^\dagger a_1\ |N - \tfrac{1}{2}\nu\rangle = [(N_1 + \tfrac{1}{2})(N_2 - \tfrac{1}{2})]^{\frac{1}{2}}\ \delta_{\nu_1,-1}\ \delta_{\nu_2,1}\prod_{i\geq 3}\delta_{\nu_i,0}. \tag{3.1}$$

[8] I. Prigogine and P. Résibois, *Superfluidité et équation de transport quantique* (Inst. interuniversitaire Sc. Nucl. Belg., Brussels, 1960).

[9] J. Yvon, *La théorie statistique des fluides et l'équation d'état* (Hermann & Cie., Paris, 1935); N. Bogoliubov, *Problemi Dynamitchesky Theorie v Statistitcheskey Phisike* (OGIS, Moscow, 1946) [English transl., *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland Publishing Company, Amsterdam, 1962), Vol. 1, p. 1]; J. G. Kirkwood, J. Chem. Phys. 14, 180 (1946); 15, 72 (1947); M. Born and H. S. Green, *A General Kinetic Theory of Fluids* (Cambridge University Press, Cambridge, England, 1949).

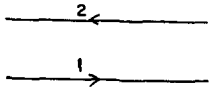


FIG. 1. An $N - \nu$ diagram representing a state in $\nu$.

Following Résibois[6] we may represent the state $\nu \equiv \{-1, 1, 0, 0 \cdots\}$ by the diagram shown in Fig. 1, where the horizontal particle line with $\mathbf{p}_1$ running to the right represents $\nu_1 = -1$ and the line with $\mathbf{p}_2$ running to the left $\nu_2 = 1$. This diagram representation may be generalized to an arbitrary state $\nu$ by drawing several directed particle lines.

In (2.14) we see that the perturbation $\mathcal{V}$ brings out a change in four momentum states. We may represent $(0|\mathcal{V}_{12,34}|-1, -1, 1, 1, 0, 0, \cdots)$ and $(1, 0, 0\cdots|\mathcal{V}_{12,34}|0, -1, 1, 1, 0, 0, \cdots)$ by the diagrams a and b, respectively, in Fig. 2.

We may expand $e^{-i\mathcal{H}t}$ in the perturbation series:

$$e^{-i\mathcal{H}t} = e^{-i\mathcal{H}_0 t}\left[1 + \sum_1^\infty (-i\lambda)^k \int_0^t d\tau_1 \times \int_0^{\tau_1} d\tau_2 \cdots \int_0^{\tau_{k-1}} d\tau_k \mathcal{V}(\tau_1)\mathcal{V}(\tau_2)\cdots\mathcal{V}(\tau_k)\right], \quad (3.2)$$

$$\mathcal{V}(\tau) \equiv e^{+i\mathcal{H}_0\tau}\mathcal{V}e^{-i\mathcal{H}_0\tau} = \int_{\mathbf{p}_1}\int_{\mathbf{p}_2}\int_{\mathbf{p}_3}\int_{\mathbf{p}_4} v(12, 34)e^{i(\epsilon_1+\epsilon_2-\epsilon_3-\epsilon_4)\tau}. \quad (3.3)$$

We wish to represent by diagrams the perturbation expansion of the density matrix

$$n(\mathbf{p}_1\mathbf{p}_2, t) = \frac{(2\pi)^3}{\Omega}\sum_{N,\nu,\nu'} \hat{n}_{12\nu}(N)(\nu|\, e^{-i\mathcal{H}(N)t}\,|\nu')\rho_{\nu'}(N, 0). \quad (3.4)$$

Components $\rho_\nu(N, 0)$ with nonzero $\nu$ describe particle correlations due to interparticle potential and hydrodynamic inhomogeneities. They may be represented by a horizontal broken line with interaction vertices and inhomogeneity marks $X$. It is seen from (2.2) that the sum of the momenta of the incoming particle lines at a vertex should be equal to the sum of the momenta of the outgoing lines. In contrast the algebraic sum of the momenta of an incoming and an outgoing line at an inhomogeneity mark $X$ is not zero.

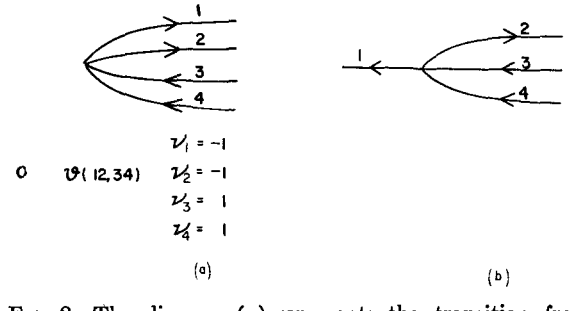


FIG. 2. The diagram (a) represents the transition from $(-1, -1, 1, 1, 0, 0, \ldots)$ to $(0, 0, \ldots)$ due to $\mathcal{V}_{12,34}$, while the diagram (b) represents the transition from $(0, -1, 1, 1, 0, 0, \ldots)$ to $(1, 0, 0, \ldots)$.

Some typical diagrams are drawn in Fig. 3, where we leave out momentum values except for those having the fixed momenta $\mathbf{p}_1$ and $\mathbf{p}_2$. By such an (partially) *unlabeled* diagram we may imply a collection of momentum-labeled diagrams of the same diagrammatic structure.

A diagram is said to be *connected* if any two points on it can be reached from one to the other without leaving it. Otherwise, the diagram will be called *disconnected*. For example, the diagram (a) in Fig. 3 is connected while the diagram (b) is disconnected.

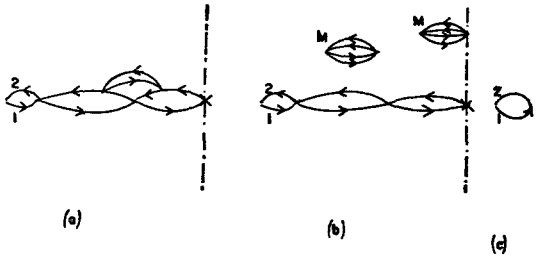


FIG. 3. Diagrams representing components of the density matrix in (3.4). A connected diagram (a) and a disconnected diagram (b) which contains vertices of the type $M$. A proper diagram (c).

Certain diagrams have an interaction vertex of the type $M$ like those two in (b) which have all the incoming and outgoing lines on their right. Such diagrams contribute nothing as in the case of classical statistics. This, in the quantum statistical case, is known as the theorem on left-multidentate structures.[10]

As all the disconnected diagrams have $M$-type interaction vertices, we have only to deal with connected diagrams containing the lines with $\mathbf{p}_1$ and $\mathbf{p}_2$.

We may analyze these diagrams in the following manner:

A linked diagram may or may not consist of two parts which can be separated out by cutting a pair of unlabeled particle lines. In the first case we shall say that it is an *improper* diagram; and in the latter case it is *proper*. For example, the diagram (a) in Fig. 3 is improper and the diagram (c) proper. A part which is suspended by such a pair of lines will be called a *self-energy part*, following the terminology used in the quantum field theory. We shall further say that we *reduce* an improper diagram into a sim-

[10] S. Fujita, J. Math. Phys. 3, 1246 (1962).

pler one when we replace with a single line a pair of lines suspending a self-energy part. If we repeatedly make such reduction for an arbitrary improper diagram, we shall finally come down to a proper diagram. In the present case of diagrams representing the density matrix in (3.4), all the improper diagrams are reducible to the only proper diagram (c) in Fig. 3.

We can extend our diagram representation to the term on the right-hand side of (2.16). Some typical diagrams are drawn in Fig. 4, where the vertex at $t$ (time) $= t$ is shown by a vertical dotted line. The improper diagram a is reducible to the proper diagram b.

It is now verified that any proper diagram consists of the two particle lines with the fixed momenta $\mathbf{p}_1$ and $\mathbf{p}_2$ and a proper self-energy part. The diagram c is the simplest proper diagram.

Thus far we could proceed in a line similar to that of the classical statistical case. Now we shall discuss an important characteristic of the quantum statistical diagrams.

A proper diagram is composed of several particle lines running between vertices. Such a line may be classified according to whether it runs from a vertex to the same or between two vertices. In the first case the line can be dressed with self-energy parts and inhomogeneity marks in the same way as for the only proper diagram (line) in Fig. 3 (c) representing a density matrix. For a line of the second type a similar dressing is possible but the line runs between two vertices of different times. This makes it impossible to express a dressed line of the second type simply in terms of a density matrix.

It is seen that except for the simplest proper diagram c in Fig. 4 every proper diagram contains lines of the second type. This implies eventually that within the present diagram analysis we may not obtain a closed evolution equation for the density matrix even if the contribution of the initial particle correlations are omitted. However, it is possible to derive a closed set of equations for partial Green's functions as was first conceived by Kadanoff and Baym. This will be indicated in the next section.

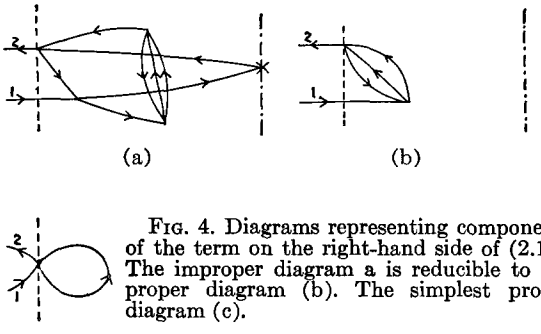


FIG. 4. Diagrams representing components of the term on the right-hand side of (2.16). The improper diagram a is reducible to the proper diagram (b). The simplest proper diagram (c).

## 4. EVOLUTION EQUATIONS FOR PARTIAL GREEN'S FUNCTION

Partial Green's functions are defined by

$$\begin{aligned} g^{>}(1,2) &\equiv g^{>}(\mathbf{p}_1 t_1, \mathbf{p}_2 t_2) \\ &\equiv -i\frac{(2\pi)^3}{\Omega}\,\mathrm{Tr}\,\{a(1)a^{\dagger}(2)\rho\}, \\ g^{<}(1,2) &\equiv \mp i\frac{(2\pi)^3}{\Omega}\,\mathrm{Tr}\,\{a^{\dagger}(2)a(1)\rho\}, \end{aligned} \tag{4.1}$$

where $a(1)$ and $a^{\dagger}(2)$ are annihilation and creation operators in the Heisenberg picture:

$$\begin{aligned} a(1) &\equiv U^{\dagger}(t_1)a_1 U(t_1), \\ a^{\dagger}(1) &\equiv U^{\dagger}(t_1)a_1 U(t_1). \end{aligned} \tag{4.2}$$

The evolution operator $U(t)$ is defined by

$$i(\partial/\partial t)U(t) = H(t)U(t) \equiv [H_0 + \lambda V(t)]U(t), \tag{4.3}$$

$$\lim_{t\to t_0=0} U(t_1, t_0) \equiv \lim_{t\to 0} U(t) = 1, \tag{4.4}$$

with the initial time $t_0$ being chosen to be zero for simplicity.

In the present case of the time-independent Hamiltonian $H$ in (2.1), the evolution operator $U(t)$ can be written as

$$U(t) \equiv e^{-itH}. \tag{4.5}$$

The function $g^{<}$ is related to the momentum density matrix in (2.4) by

$$\lim_{t_2\to t_1} \pm ig^{<}(1,2) = n(t_1)_{12}. \tag{4.6}$$

The mechanical evolution equations for $g^{>}$ and $g^{<}$ can be derived from the Heisenberg equation of motion for $a(1)$ and $a^{\dagger}(2)$. In the presence of pairwise interactions these equations are hierachy equations of the lowest order similar to the Bogoliubov–Born–Green–Kirkwood–Yvon equations. Kadanoff and Baym argued that under certain conditions these equations can be, in the thermodynamic limit, transformed into closed equations for $g^{>}$ and $g^{<}$.[2] In an earlier paper the present author investigated the underlying conditions by means of a perturbation theory.[11] Here a system of noninteracting electrons in an impurity field is treated with the initial con-

[11] S. Fujita, Physica **30**, 848 (1964).

dition of no particle correlations and no spatial inhomogeneities. The adaptation of this perturbation theory to an imperfect gas is immediate. The inclusion of the initial particle correlations and inhomogeneities will introduce the same modifications as those discussed in the preceding section.

In the following we shall briefly outline the derivation of the corrected Kadanoff–Baym equations as the arguments proceed similarly to those in Ref. 11.

As $g^<$ contains three evolution operators $\{(\exp(it_2H), \exp[-i(t_2 - t_1)H], \exp(-it_1H)\}$ the technique of $N - \nu$ representation in Sec. 3 does not help very much. We may analyze the perturbation expansion of $g^<$ in the ordinary $n$-space (momentum occupation number space).

The three evolution operators may be expanded in perturbation series with the use of

$$e^{-itH} \equiv e^{-itH_0}S(t), \tag{4.7}$$

$$S(t) = 1 + \sum_1^\infty (-i\lambda)^k \int_0^t d\tau_1 \int_0^{\tau_1} d\tau_2 \cdots \int_0^{\tau_{k-1}} d\tau_k \times V(\tau_1)V(\tau_2)\cdots V(\tau_k), \tag{4.8}$$

$$V(\tau) \equiv e^{i\tau H_0}Ve^{-i\tau H_0} = \frac{1}{4}\int_{p_1}\int_{p_2}\int_{p_3}\int_{p_4} v(12, 34)e^{i(\epsilon_1+\epsilon_2-\epsilon_3-\epsilon_4)\tau}a_1^\dagger a_2^\dagger a_4 a_3. \tag{4.9}$$

From (4.1), (4.2), (4.5), and (4.7), one has

$$g^<(1, 2) = \pm i\frac{(2\pi)^3}{\Omega} \times \mathrm{Tr}\{S^\dagger(t_2)a_2^\dagger(t_2)S'(t_2, t_1)a_1(t_1)S(t_1)\rho\}, \tag{4.10}$$

$$a_1(t) \equiv e^{itH_0}a_1e^{-itH_0}, \tag{4.11}$$

$$S'(t_2, t_1) \equiv e^{it_2H_0}S(t_2 - t_1)e^{-it_2H_0} = 1 + \sum_1^\infty (-i\lambda)^k \int_{t_1}^{t_2} d\tau_1 \int_{t_1}^{\tau_1} d\tau_2 \cdots \int_{t_1}^{\tau_{k-1}} d\tau_k \times V(\tau_1)V(\tau_2)\cdots V(\tau_k) = 1 + \sum_1^\infty (i\lambda)^k \int_{t_2}^{t_1} d\tau_1 \int_{t_2}^{\tau_1} d\tau_2 \cdots \int_{t_2}^{\tau_{k-1}} d\tau_k \times V(\tau_k)\cdots V(\tau_2)V(\tau_1). \tag{4.12}$$

In general the density operator $\rho$ may be expanded as

$$\rho \equiv \sum_n \sum_{n'} \rho(n', n)\,|n'\rangle\langle n| = \sum_n |n\rangle\langle n|\,\rho(n) + \sum_n \int_{p_1}\int_{p_2} a_1^\dagger a_2\,|n\rangle\langle n|\,\rho(\mathbf{p}_1\mathbf{p}_2, n) + \sum_n \int_{p_1}\int_{p_2}\int_{p_3}\int_{p_4} a_1^\dagger a_3^\dagger a_4 a_2\,|n\rangle\langle n|\,\rho(\mathbf{p}_1\mathbf{p}_3\mathbf{p}_2\mathbf{p}_4, n) + \cdots, \tag{4.13}$$

$$\rho(n', n) \equiv \langle n'|\,\rho\,|n\rangle, \tag{4.14}$$

$$\rho(\mathbf{p}_1\mathbf{p}_2, n) \equiv \langle n|\,a_2^\dagger a_1\rho\,|n\rangle, \qquad \rho(\mathbf{p}_1\mathbf{p}_3, \mathbf{p}_2\mathbf{p}_4, n) \equiv \langle n|\,a_2^\dagger a_4^\dagger a_3 a_1\rho\,|n\rangle. \tag{4.15}$$

Substituting the expansion (4.13) of $\rho$ into (4.10) one sees that one has to deal with $n$–$n$ elements of the form

$$\langle n|\,V(\tau)\cdots V(\tau')a^\dagger\cdots a\,|n\rangle. \tag{4.16}$$

Such an element will vanish unless the set of momenta associated with the creation operators is equal to the set of momenta associated with the annihilation operators. Thus the pairings of creation and annihilation operators with common momenta are possible and will be indicated by directed particle lines in the following diagram representation of the expanded terms of $g^<$.

Draw a horizontal boundary line. The $a^\dagger(2)$ is denoted by a point at $t = t_2$ at which a particle line is to start and $a(1)$ by a point at $t = t_1$ at which a particle is to arrive. All the interactions $V(\tau)$ arising from the three $S$ can be ordered according to their time arguments: Those $V(\tau)$ from $S^\dagger(t_2)$ are represented by vertices below the boundary line, and those $V(\tau)$ from $S(t_1)$ by vertices above the boundary. Those $V(\tau)$ from $S'(t_2, t_1)$ are represented by vertices above or below according to whether $t_2 > t_1$ or $t_2 < t_1$. Each of these vertices will have two particle lines arriving and two particle lines leaving.

The operators $a^\dagger \cdots a$ in (4.16) associated with the initial inhomogeneities and/or particle correlations will be represented by marks $X$ and/or vertices on the vertical line at $t = 0$ above the boundary.

Every particle line being labeled with a momentum, a diagram could be thought to represent a schematic change in state by means of interactions $V(\tau)$ just as an $N$–$\nu$ diagram in the preceding section.[7] In fact according to the previous analysis the diagrams in the $n$–$n$ representation have a close correspondence with those in the $N$–$\nu$ representation.[12] If the horizontal boundary line were suppressed, the diagrammatical structures in both representations would be the same. We can make diagram analysis of the same sort. The notions of, labeled and unlabeled diagrams, connected and disconnected diagrams, $M$-type vertices, connected-diagram expansion, self-energy diagrams, proper and

[12] S. Fujita, Physica **27**, 940 (1961).

improper diagrams, reduction of diagrams, and so on, can be used.

In the present diagram analysis, we purposely avoid the use of Feynman diagrams representing contractions which would arise by the application of the contraction theorem.[13] However, as a connected diagram will not in general (except for the cases of long range forces and those of degenerate bosons) have any thermodynamic singularities, the difference between the present analysis and the analysis by means of Feynman diagrams will be negligible in the thermodynamic limit. In particular a system of bosons in nondegenerate phase may be included in the present analysis although this system does not allow the usual form of the contraction theorem with respect to the $n$–$n$ expectation value.

Differentiating $g^<(1, 2)$ with respect to $t_1$, one obtains

$$\begin{aligned} &i\frac{\partial}{\partial t_1} g^<(1, 2) \\ &= \pm i \frac{(2\pi)^3}{\Omega} \mathrm{Tr}\, \{a^\dagger(2)U^\dagger(t_1)[H, a_1]U(t_1)\rho\} \\ &= \epsilon_1 g^<(1, 2) \mp \frac{i}{2}\frac{(2\pi)^3}{\Omega}\int_{\mathbf{p}_3}\int_{\mathbf{p}_4}\int_{\mathbf{p}_5} v(13, 45) \\ &\quad \times \mathrm{Tr}\, \{S^\dagger(t_2)a_2^\dagger(t_2)S'(t_2, t_1)a_3^\dagger(t_1)a_5(t_1)a_4(t_1)S(t_1)\rho\}. \end{aligned} \tag{4.17}$$

We can diagram-analyze the last term in the last member in the same way. This time however, the point at $t = t_1$ will have one leaving and two arriving lines.

It is not difficult to see that in any proper diagram the particle line starting from the point at $t = t_2$ ends either at the point at $t = t_1$ or at a unique interaction vertex 3 which may arise from the three $S$ and $\rho$ in (4.17).

In the first alternative the contribution of such diagrams and their associated improper ones will be included in

$$u(1)g^<(1, 2), \tag{4.18}$$

$$u(1) \equiv \pm i\lambda \iint d^3p_3\, d^3p_4 v(13, 14)g^<(\mathbf{p}_4 t_1, \mathbf{p}_3 t_1). \tag{4.19}$$

In the second alternative if the vertex 3 originates in one of the three $S$, this case will be accounted for by

$$\begin{aligned} &\int_0^{t_1} dt_3 \int d^3p_3 \Sigma^>(1, 3)g^<(3, 2) \\ &\quad - \int_0^{t_2} dt_3 \int d^3p_3 \Sigma^<(1, 3)g^>(3, 2) \\ &\quad - \int_{t_1}^{t_2} dt_3 \int d^3p_3 \Sigma^<(1, 3)g^<(3, 2), \end{aligned} \tag{4.20}$$

$$\begin{aligned} \Sigma^>(1, 3) &\equiv -i\lambda^2 \int\cdots\int \Big(\prod_{i=4}^{9} d^3p_i\Big)[v(19, 78)v(54, 63) \\ &\quad \times \mathrm{Tr}\, \{S^\dagger(t_1)a_9^\dagger(t_1)a_8(t_1)a_7(t_1) \\ &\quad \times S'(t_1, t_3)a_4^\dagger(t_3)a_5^\dagger(t_3)a_6(t_3)S(t_3)\rho\}]_\sigma \\ &= -\Sigma^>(3, 1)^*, \end{aligned}$$

$$\begin{aligned} \Sigma^<(1, 3) &\equiv i\lambda^2 \int\cdots\int \Big(\prod_{i=4}^{9} d^3p_i\Big)[v(16, 45)v(78, 39) \\ &\quad \times \mathrm{Tr}\, \beta S^\dagger(t_3)a_7^\dagger(t_3)a_8^\dagger(t_3)a_9(t_3) \\ &\quad \times S'(t_3, t_1)a_6^\dagger(t_1)a_5(t_1)a_4(t_1)S(t_1)\rho\}]_\sigma \\ &= -\Sigma^<(3, 1)^*, \end{aligned} \tag{4.21}$$

where the subscript $\sigma$ means that contribution corresponding to the whole set of proper self-energy parts defined with dressed-particle lines. A dressed particle is the sum of an undressed line and those lines which upon reduction give rise to the undressed line, and represents a $g^>$ or $g^<$. If the vertex 3 is located on the vertical line at $t = 0$, the contribution of such diagrams and their associated improper diagrams cannot be expressed simply in terms of $g^>$ and $g^<$. If we denote this contribution by $D$, we may write the evolution equation for $g^<(1, 2)$ in the thermodynamic limit as

$$\begin{aligned} &[i(\partial/\partial t_1) - \epsilon_1 - u(1)]g^<(1, 2) \\ &= \int_0^{t_1} dt_3 \int d^3p_3[\Sigma^>(1, 3) - \Sigma^<(1, 3)]g^<(3, 2) \\ &\quad - \int_0^{t_2} dt_3 \int d^3p_3 \Sigma^<(1, 3)[g^>(3, 2) - g^<(3, 2)] + D^{(a)}. \end{aligned} \tag{4.22a}$$

This is one of *corrected* Kadanoff–Baym equations[2] describing the evolution of partial Green's functions $g^>$ and $g^<$. The other three equations can be obtained by transforming the $t_1$- and $t_2$-derivatives of $g^>(1, 2)$ and $g^<(1, 2)$, e.g.,

$$\begin{aligned} &[-i(\partial/\partial t_2) - \epsilon_2 - u(2)]g^<(1, 2) \\ &= \int_0^{t_1} dt_3 \int d^3p_3[g^>(1, 3) - g^<(1, 3)]\Sigma^<(3, 2)] \\ &\quad - \int_0^{t_2} dt_3 \int d^3p_3 g^<(1, 3)[\Sigma^>(3, 2) - \Sigma^<(3, 2)] + D^{(b)}. \end{aligned} \tag{4.22b}$$

[13] G. C. Wick, Phys. Rev. **80**, 268 (1950); K. Nishikawa, J. Phys. Soc. (Japan) **15**, 78 (1960); C. Bloch and C. De Dominicis, Nucl. Phys. **7**, 459 (1958).

The terms $D^{(a)}$, $D^{(b)}$ $\cdots$ may be written down from the diagrams. In particular if the initial density operator $\rho$ is chosen to be the grand canonical operator

$$\rho = e^{\alpha N-\beta H}/\mathrm{Tr}\,\{e^{\alpha N-\beta H}\}, \tag{4.23}$$

$D^{(a)}$ and $D^{(b)}$ can be expressed as

$$D^{(a)} = \int_0^\beta d\beta_3 \int d^3p_3 \Sigma'^{>}(1, 3')g'^{<}(3', 2),$$
$$D^{(b)} = \int_0^\beta d\beta_3 \int d^3p_3 g'^{>}(1, 3')\Sigma'^{<}(3', 2), \tag{4.24}$$

$$g'^{<}(1', 2) \equiv \mp i\frac{(2\pi)^3}{\Omega}\mathrm{Tr}\,\{a^\dagger(2)a_1'(\beta_1)\rho\},$$
$$g'^{>}(1, 2') \equiv \mp i\frac{(2\pi)^3}{\Omega}\mathrm{Tr}\,\{a(1)a_2'^*(\beta_2)\rho\}, \tag{4.25}$$

$$\Sigma'^{>}(1, 2') \equiv \lambda^2 \int \cdots \int \left(\prod_{i=4}^{9} d^3p_i\right)[v(16, 45)v(78, 29)$$
$$\times \mathrm{Tr}\,\{S^\dagger(t_1)a_6^\dagger(t_1)a_5(t_1)a_4(t_1)S(t_1)$$
$$\times a_7'^*(\beta_2)a_8'^*(\beta_2)a_9'(\beta_2)\rho\}]_\sigma, \tag{4.26}$$

$$\Sigma'^{<}(1', 2) \equiv \lambda^2 \int \cdots \int \left(\prod_{i=4}^{9} d^3p_i\right)[v(78, 29)v(16, 45)$$
$$\times \mathrm{Tr}\,\{S^\dagger(t_2)a_7^\dagger(t_2)a_8^\dagger(t_2)a_9(t_2)$$
$$\times S(t_2)a_6'^*(\beta_1)a_5(\beta_1)a_4(\beta_1)\rho\}]_\sigma,$$

$$a_1'(\beta_1) \equiv e^{\beta_1 H_0}a_1e^{-\beta_1 H_0} = e^{-\beta_1\epsilon_1}a_1,$$
$$a_1'^*(\beta_1) \equiv e^{\beta_1 H_0}a_1^\dagger e^{-\beta_1 H_0} = e^{\beta\epsilon_1}a_1. \tag{4.27}$$

## 5. REMARKS

A. In analogy with the classical statistical case, partial Green's functions $g^>(\mathbf{p}_1t_1, \mathbf{p}_2t_2)$ and $g^<$ obey a set of closed simultaneous equations except for the neglect of the contribution $D$ of the initial particle correlations. It is known[14] that for the purpose of calculating transport coefficients the Markoffian approximation to the evolution equations can be used. In this stage of approximation the density matrix $n$ could be shown to satisfy a closed equation as in the case of the Uehling–Uhlenbeck equation[15] (a very dilute gas) and in a more general equation previously discussed by the author (homogeneous system).[16]

B. Corrected Kadanoff–Baym's equations (4.22) are derived for Green's functions describing a system of interacting particles, obeying the Bose–Einstein or the Fermi–Dirac statistics, with an initial density operator $\rho$ corresponding to a state of *arbitrary particle correlations and inhomogeneities.* The underlying restrictions are that (1) Green's functions should be finite and (2) the interaction processes involving a finite fraction of the total number of particles should be negligible. The restriction (2) excludes the case of interacting bosons in degenerate phase. The notion of the self-energy parts $\Sigma^<$ and $\Sigma^>$ is essentially independent of representation and could be introduced without referring to the momentum representation (see Ref. 2). The equations (4.22) are valid also for time-dependent interactions $v(t)$, ($H_0$ = time independent), for which, it is noted, the perturbation expansion similar to (4.8) is still available.

C. The term $D$ in (4.22) represents contribution arising from the initial particle correlations. This term $D$ would die out in the time of the order of the average collision duration $\tau_c$. A particular case is explicitly worked out in the Appendix. A self-energy part $\Sigma^>(1, 2)$ or $\Sigma^<$ will in general consist of two parts: one part expressible in terms of $g^>$ and $g^<$, and the remainder. This latter part is found to be always connected with the initial particle correlations, and therefore, could be left out if one is concerned with a system at a time $t \gg \tau_c$. Thus, apart from the contribution of the initial particle correlations, the Eqs. (4.22) can be considered as closed equations for $g^>$ and $g^<$.

D. Kadanoff and Baym treated partial Green's functions $g^>$ and $g^<$ defined with the grand canonical density operator, which is a main difference from the present treatment. By choosing that the initial time tends to $-\infty$, they implicitly *assumed* the vanishing of the effect of the initial particle correlations, which of course depends on the given system.

## APPENDIX: INITIAL PARTICLE CORRELATIONS*

In this appendix we show that the contribution arising from the initial particle correlations will die out in a time of the order of the average collision time $\tau_c$.

Subtracting (4.22a) from (4.22b) side by side and setting that $\mathbf{p}_2 = \mathbf{p}_1$ and $t_2 = t_1$, one obtains

$$(\partial/\partial t_1)n(\mathbf{p}_1t_1) = \pm i(\partial/\partial t_1)g^<(1, 1)$$
$$= \mp\int_0^{t_1} dt_3 \int d^3p_3[\Sigma^<(1, 3)g^>(3, 1) + \text{c.c.}$$

(complex conjugate)]

[14] R. Balescu, Physica **27**, 693 (1961).
[15] E. A. Uehling and G. E. Uhlenbeck, Phys. Rev. **43**, 552 (1933).
[16] S. Fujita, Physica **29**, 1087 (1963).

* By S. Fujita, Michael Fitelson, and Anthony St. Pierre, Pennsylvania State University, University Park, Pennsylvania.

$$\pm \int_0^{t_1} dt_3 \int d^3p_3[\Sigma^>(1,3)g^<(3,1) + \text{c.c.}] \mp (D^{(b)} - D^{(a)}). \quad \text{(A.1)}$$

From this equation we may derive the Boltzmann equation for a dilute homogeneous gas. Such a derivation may be seen in the monograph by one of the authors (S. F.).[17] Here, we discuss the behavior of the term $\mp(D^{(b)} - D^{(a)})$ due to the initial particle correlations. To illustrate the main points, let us consider the simplest possible case.

We assume (a) the grand canonical density operator for $\rho$, (b) weak and short-range forces, and (c) classical statistics. With these conditions the term $\mp(D^{(b)} - D^{(a)})$ will be written down from (4.24)–(4.27)

$$D(t_1) \equiv \int_0^\beta d\lambda \iiint d^3p_4\, d^3p_5\, d^3p_6\, |v(\mathbf{p}_1 - \mathbf{p}_4)|^2 f_1 f_6 \times e^{\lambda[\epsilon_1+\epsilon_6-\epsilon_4-\epsilon_5]} e^{-it(\epsilon_1+\epsilon_6-\epsilon_4-\epsilon_5)} \times \delta^{(3)}(\mathbf{p}_1 + \mathbf{p}_6 - \mathbf{p}_4 - \mathbf{p}_5) + \text{c.c.}, \quad \text{(A.2)}$$

$$f_1 \equiv e^{\alpha - \beta p_1^2}. \quad \text{(A.3)}$$

Introducing the variable

$$q \equiv \mathbf{p}_1 - \mathbf{p}_4,$$

one may rewrite (A.2) as

$$D(t_1) = f_1 \iint d^3p_5\, d^3q\, |v(\mathbf{q})|^2 f_{p_5-q} \times \exp\left[+\frac{it_1}{M}\mathbf{q}\cdot(\mathbf{p}_5 - \mathbf{p}_1)\right] \times \int_0^\beta d\lambda \exp\left[-\frac{\lambda}{M}\mathbf{q}\cdot(\mathbf{p}_5 - \mathbf{p}_1)\right] + \text{c.c.} \quad \text{(A.4)}$$

When $t_1$ is large, the value of the integrand near $(q/M)\cdot(p_5 - p_1) = 0$ is most important. In this condition the $\lambda$-integral may be approximated by $\beta$. This approximation becomes exact anyway in the high temperature limit: $\beta \to 0$. Thus,

$$D(t_1) \cong -\beta f_1 \iint d^3p_5\, d^3q\, |v(\mathbf{q})|^2 \times \exp\left[\frac{\alpha - \beta(\mathbf{p}_5 - \mathbf{q})^2}{2M}\right] \times \exp\left[i\frac{t_1}{M}\mathbf{q}\cdot(\mathbf{p}_5 - \mathbf{p}_1)\right] + \text{c.c.} \quad \text{(A.5)}$$

[17] S. Fujita, *Non-Equilibrium Quantum Statistical Mechanics* (Saunders, Philadelphia, to be published in 1966).

Let us assume the form of a potential $v$ to be Gaussian with a range $R$:

$$\lambda v(r) = \lambda e^{-r^2/R^2}$$

$$v(q) \equiv \frac{1}{(2\pi)^3}\int d^3r\, e^{-iq\cdot r} e^{-r^2/R^2} = \frac{1}{(2\pi)^3}(R\pi^{\frac{1}{2}})^3 e^{-\frac{1}{4}R^2q^2}. \quad \text{(A.6)}$$

If we use this, we can easily evaluate the integral (A.5) exactly. The calculation involves Fourier–Gauss integrations twice like that in (A.6). We thus obtain

$$D(t_1) = \lambda^2\beta\left(\frac{2\pi M}{\beta}\right)^{\frac{3}{2}} \frac{R^3}{64\pi^3} e^{\alpha} f_1 \left[\frac{R^2}{2} + \frac{t_1^2}{2M\beta} - i\frac{t_1}{M}\right]^{-\frac{3}{2}} \times \exp\left[-p_1^2t_1^2/4M^2\left(\frac{R^2}{2} + \frac{t_1^2}{2M\beta} - i\frac{t_1}{M}\right)\right] + \text{c.c.} \quad \text{(A.7)}$$

The absolute magnitude of this quantity as a function of $t_1$ has a peak at the origin and decreases with oscillations as $t_1$ becomes large. The value of $t_1$ for which the real part of the exponential function become one-half the value at the origin may be estimated by

$$p_1^2t_1^2/4M^2\left(\frac{R^2}{2} + \frac{t_1^2}{2M\beta}\right) = 1, \quad \text{(A.8)}$$

which yields

$$t_1^2 = \frac{2R^2}{(p_1/M)^2 - \frac{2}{3}v_\beta^2} \sim \frac{R^2}{v_\beta^2} \equiv \tau_c^2, \quad \text{(A.9)}$$

$$\tfrac{1}{2}Mv_\beta^2 \equiv \tfrac{3}{2}kT \equiv \tfrac{3}{2}\frac{1}{\beta}. \quad \text{(A.10)}$$

The equation (A.9) shows that $|D(\mathbf{p}_1 t_1)|$ decreases in the average in a time of the order of $\tau_c$ as we wished to show.

We may qualitatively understand this result in the following way: In the oscillatory exponential in (A.2) $\epsilon \equiv \epsilon_1 + \epsilon_6 - \epsilon_4 - \epsilon_5$ is the transferred energy before and after the two-body interaction process. This exponential is weighted by the matrix elements of the potential and others. If the potential has a range $R$, the weight as a function of $\epsilon$ may be shown to decrease rapidly for $|\epsilon| > \langle v_1\rangle/R \equiv \tau_c^{-1}$. Therefore, if $t_1$ is much greater than $\tau_c$, the rapidly oscillatory exponential makes the contributions negligibly small.

Such qualitative argument can be applied to more

general cases: It may be verified, say by a perturbation treatment, that the contribution of initial particle correlations are in general composed of integrals of the form

$$\int \cdots \int e^{i\epsilon t_1}\varphi(\epsilon, t_1). \qquad \text{(A.11)}$$

We may then apply the same argument to each integral. This argument relies on the smooth behavior of $\varphi(\epsilon, t_1)$ as function of $\epsilon$. The thermodynamic singularities are eliminated by the connected-diagram treatment. Singularities due to the possible bound-state formation, long-range forces, etc., must be examined in practical cases.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 12 DECEMBER 1965

# Momentum-Space Analyticity Properties of the Bergman-Weil Integral for the Three Point Function*

G. KÄLLÉN
*Department of Theoretical Physics, University of Lund, Lund, Sweden*
AND
J. S. TOLL†
*Department of Physics and Astronomy, University of Maryland, College Park, Maryland*
(Received 10 May 1965)

The momentum-space analyticity domain of the Bergman–Weil integral representation of the vertex function or three-point function is investigated using the assumption that thresholds can be introduced simply as lower limits of integration for the mass variables in the representation. It is shown that this assumption leads to a regularity domain which is larger than the domain following from general physical assumptions of Lorentz invariance, local commutativity and reasonable mass spectrum. To simplify the discussion we also assume that the vertex function is regular when all three complex variables lie in the same half-plane. Standard techniques for the evaluation of Feynman diagrams have proved to be inconvenient for this investigation and we have developed new methods taking explicit advantage of the fact that we only have two external vectors and, hence, can work in a two-dimensional Lorentz space. Further, the vanishing of the masses for certain internal lines has also been exploited. The techniques we have used here might be of interest also in other connections.

## 1. INTRODUCTION

ONE of the unsolved problems in a systematic approach to quantum field theory is to determine the exact analyticity domain in momentum space of the three-point function in theories fulfilling the three standard assumptions of[1]

(i) Lorentz invariance,

(ii) Reasonable mass spectrum, i.e., the assumption that every energy momentum vector of a physical state is either identically zero (for the vacuum) or timelike with positive energy above a given minimum mass,

(iii) Local commutativity, i.e., the assumption that two field operators commute for spacelike separations.

Several years ago, the corresponding problem was solved for the analyticity domain $\mathfrak{U}$ in configuration space of the three-point function and it was shown that the vacuum expectation value of the ordinary product of three (scalar) fields is the boundary value of an analytic function regular in a certain domain bounded by pieces of analytic hypersurfaces (viz. cuts, $F'_{k1}$ and $\mathfrak{F}$).[2] Afterwards, a representation of the most general function analytic in this domain and sufficiently bounded at infinity was written down[3] with the aid of a generalization of the Cauchy

* Assisted in part by the U. S. Air Force Office of Scientific Research through contract AFOSR 500-64 and the National Science Foundation through grant NSF GP 3221.

† Present address: Department of Physics, State University of New York at Stony Brook, Stony Brook, New York.

[1] These assumptions have been implicitly used by many authors since very long ago, e.g., W. Pauli, Prog. Theoret. Phys. (Kyoto) 5, 526 (1950). They were explicitly stated and used, e.g., by H. Umezawa and S. Kamefuchi, Prog. Theoret. Phys. (Kyoto) 6, 543 (1951); G. Källén, Helv. Phys. Acta 25, 417 (1952); J. Schwinger, Proc. Nat. Acad. Sci. 37, 455 (1951). Later, these assumptions have also been used by H. Lehman, K. Symanzik, W. Zimmermann, Nuovo Cimento 1, 205 (1955) and by A. Wightman, Phys. Rev. 101, 860 (1956).

[2] G. Källén and A. Wightman, Kgl. Danske Videnskab. Selskab. Mat.-Fys. Skrifter 1, No. 6 (1958).

[3] G. Källén, J. Toll, Helv. Phys. Acta 33, 753 (1960).

general cases: It may be verified, say by a perturbation treatment, that the contribution of initial particle correlations are in general composed of integrals of the form

$$\int \cdots \int e^{i\epsilon t_1}\varphi(\epsilon, t_1). \qquad \text{(A.11)}$$

We may then apply the same argument to each integral. This argument relies on the smooth behavior of $\varphi(\epsilon, t_1)$ as function of $\epsilon$. The thermodynamic singularities are eliminated by the connected-diagram treatment. Singularities due to the possible bound-state formation, long-range forces, etc., must be examined in practical cases.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 12 DECEMBER 1965

# Momentum-Space Analyticity Properties of the Bergman-Weil Integral for the Three Point Function*

G. KÄLLÉN
*Department of Theoretical Physics, University of Lund, Lund, Sweden*
AND
J. S. TOLL†
*Department of Physics and Astronomy, University of Maryland, College Park, Maryland*
(Received 10 May 1965)

The momentum-space analyticity domain of the Bergman–Weil integral representation of the vertex function or three-point function is investigated using the assumption that thresholds can be introduced simply as lower limits of integration for the mass variables in the representation. It is shown that this assumption leads to a regularity domain which is larger than the domain following from general physical assumptions of Lorentz invariance, local commutativity and reasonable mass spectrum. To simplify the discussion we also assume that the vertex function is regular when all three complex variables lie in the same half-plane. Standard techniques for the evaluation of Feynman diagrams have proved to be inconvenient for this investigation and we have developed new methods taking explicit advantage of the fact that we only have two external vectors and, hence, can work in a two-dimensional Lorentz space. Further, the vanishing of the masses for certain internal lines has also been exploited. The techniques we have used here might be of interest also in other connections.

## 1. INTRODUCTION

ONE of the unsolved problems in a systematic approach to quantum field theory is to determine the exact analyticity domain in momentum space of the three-point function in theories fulfilling the three standard assumptions of[1]

(i) Lorentz invariance,

(ii) Reasonable mass spectrum, i.e., the assumption that every energy momentum vector of a physical state is either identically zero (for the vacuum) or timelike with positive energy above a given minimum mass,

(iii) Local commutativity, i.e., the assumption that two field operators commute for spacelike separations.

Several years ago, the corresponding problem was solved for the analyticity domain $\mathfrak{U}$ in configuration space of the three-point function and it was shown that the vacuum expectation value of the ordinary product of three (scalar) fields is the boundary value of an analytic function regular in a certain domain bounded by pieces of analytic hypersurfaces (viz. cuts, $F'_{k1}$ and $\mathfrak{F}$).[2] Afterwards, a representation of the most general function analytic in this domain and sufficiently bounded at infinity was written down[3] with the aid of a generalization of the Cauchy

* Assisted in part by the U. S. Air Force Office of Scientific Research through contract AFOSR 500-64 and the National Science Foundation through grant NSF GP 3221.

† Present address: Department of Physics, State University of New York at Stony Brook, Stony Brook, New York.

[1] These assumptions have been implicitly used by many authors since very long ago, e.g., W. Pauli, Prog. Theoret. Phys. (Kyoto) 5, 526 (1950). They were explicitly stated and used, e.g., by H. Umezawa and S. Kamefuchi, Prog. Theoret. Phys. (Kyoto) 6, 543 (1951); G. Källén, Helv. Phys. Acta 25, 417 (1952); J. Schwinger, Proc. Nat. Acad. Sci. 37, 455 (1951). Later, these assumptions have also been used by H. Lehman, K. Symanzik, W. Zimmermann, Nuovo Cimento 1, 205 (1955) and by A. Wightman, Phys. Rev. 101, 860 (1956).

[2] G. Källén and A. Wightman, Kgl. Danske Videnskab. Selskab. Mat.-Fys. Skrifter 1, No. 6 (1958).

[3] G. Källén, J. Toll, Helv. Phys. Acta 33, 753 (1960).

integral given by S. Bergman and A. Weil. It turned out that this representation, after a suitable number of modifications, could be written as a sum of four terms, three of which could be recognized as being contributions from the comparatively simple Feynman diagram shown in Fig. 1. This diagram has been called the "Mercedes diagram."[4] The fourth term in the Bergman–Weil integral is associated with the boundary curve $\mathfrak{F}$ for the configuration-space function. It can be shown[6] that every finite order in perturbation theory with an interaction Hamiltonian which is a multinomial in the fields to a three-point function equal to a superposition of Mercedes diagrams and simpler triangle diagrams, i.e., the term corresponding to $\mathfrak{F}$ is replaced by the contribution from three cuts. In this paper we are concentrating our attention on functions analytic in this domain $\mathfrak{U}'$ characteristic of perturbation theory and bounded by the cuts and the analytic hypersurfaces $F'_{kl}$. The Bergman–Weil integral for functions regular in $\mathfrak{U}'$ can be written[3] in the form

$$F(\tilde{z}) = \iiint_0^\infty da\,db\,dc \int_0^\infty dv \sum_{i=1}^{3} F_M(\tilde{z}; a; v)^{(i)}\varphi^{(i)}(a; v) + \iiint_0^\infty da\,db\,dc\, F_\Delta(\tilde{z}; a)\varphi_\Delta(a). \quad (1)$$

Here, $\tilde{z}$ is a shorthand notation for the three complex variables $\tilde{z}_1$, $\tilde{z}_2$, and $\tilde{z}_3$ defined by

$$\tilde{z}_1 = -(x - x')^2, \quad (2a)$$

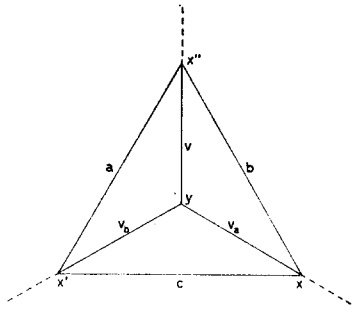


FIG. 1. The Mercedes diagram.

$$\tilde{z}_2 = -(x - x'')^2, \quad (2b)$$

$$\tilde{z}_3 = -(x' - x'')^2. \quad (2c)$$

(We use the space favored metric $x^2 = \bar{x}^2 - x_0^2$.) Further, $a$ on the right-hand side of Eq. (1) is a short notation for the three mass squares $a$, $b$, and $c$ in Fig. 1. The function $F_M(\tilde{z}; a; v)$ is the contribution from the Mercedes diagram in Fig. 1. It is given by

$$F_M(\tilde{z}; a; v) = \int dy\, \Delta_F(x - y, v_a)\, \Delta_F(x' - y, v_b)\, \Delta_F(x'' - y, v_c) \times \Delta_F(x - x', c)\, \Delta_F(x - x'', b)\, \Delta_F(x' - x'', a). \quad (3)$$

(The integration $dy$ in Eq. (3) goes over all four components of the vector $y$.) The functions $\Delta_F(x, a)$ in Eq. (3) are the conventional Feynman singular functions defined by

$$\Delta_F(x, a) = \Delta_F[\tilde{z}, a] = \frac{2}{i}\frac{1}{(2\pi)^4}\int dp\, \frac{e^{ipx}}{p^2 + m^2 - i\epsilon} = -(1/8\pi)(a/\tilde{z})^{\frac{1}{4}}H_1^{(1)}((a\tilde{z})^{\frac{1}{2}}), \quad (4)$$

$$\tilde{z} = -x^2 - i\epsilon. \quad (4a)$$

The function $H_1^{(1)}$ in (4) is the usual Hankel function of first order and exponentially damped in the upper half-plane. The square root in Eq. (4) is defined to have a positive imaginary part. Further, Eqs. (3) and (4) are valid for time-ordered products of the field operators. However, it can be proved that the ordinary vacuum expectation value is another boundary value of the same analytic function[2] and we can here study either one of them.

When the index $i$ in the sum on the right hand of Eq. (1) is equal to one, we put $v_c = v$, $v_a = v_b = 0$; when $i = 2$, we have $v_b = v$, $v_a = v_c = 0$ and for $i = 3$ we put $v_a = v$, $v_b = v_c = 0$. The functions $\varphi^{(i)}(a; v)$ are weight functions which are arbitrary except that they should decrease so rapidly at infinity that the integrals are convergent. The function $F_\Delta(\tilde{z}, a)$ is the contribution from the triangle diagram. It is given by

$$F_\Delta(\tilde{z}, a) = \Delta_F[\tilde{z}_1, c]\, \Delta_F[\tilde{z}_2, b]\, \Delta_F[\tilde{z}_3, a]. \quad (5)$$

The function $\varphi_\Delta(a)$ is a weight with properties similar to the other weights.

The representation (1) above is correct also when the minimum mass in assumption (ii) is zero. In many cases of physical interest it is desirable to make explicit use of a positive value for this minimum mass. This implies that the mass spectrum consists of a set of sharp mass values $m_i$ corre-

[4] Cf., e.g., Ref. 5 below. The name "Mercedes" appears naturally if one draws the outer triangle as a circle.

[5] W. S. Brown, J. Math. Phys. 3, 321 (1962).

[6] This statement can be found in the literature for the momentum-space function [cf. K. Symanzik, Progr. Theoret. Phys. 20, 690 (1958)] in the form that there are no singularities when all of the momentum space $z$-variables (as defined below) lie in the same half-plane. Writing down the Bergman–Weil integral for such a momentum space function and transforming it in the way described in Ref. 3, one finds a result which corresponds to a configuration space function regular analytic when all the configuration space variables $\tilde{z}_j$ lie in the same half-plane.

sponding to single-particle states and continua starting from the mass values $m_i + m_j$ corresponding to scattering states. Under those circumstances, the Fourier transforms of the ordinary vacuum expectation values vanish below certain thresholds. If we assume that we are only considering operators where the matrix elements between the vacuum and one-particle states vanish (this is the case, e.g., for current operators), these thresholds are determined by the scattering states and given by $M = m_i + m_j$. Because of various selection rules these thresholds are different for different operators. An inclusion of these thresholds does not enlarge the analyticity domain of the $x$-space function $F(\xi)$ in Eq. (1). This is seen from the simple remark that one can find a large class of three-point functions fulfilling also the sharper version of assumption (ii) by restricting the domain of integration for the masses $a$, $b$, and $c$ in Eq. (1) by relations of the form $a^2 + b^2 \geq M_1$, $a^2 + c^2 \geq M_2$, and $b^2 + c^2 \geq M_3$, where $M_i$ denotes the three thresholds for the operators under consideration. The function $F(\xi)$ obtained in this way has singularities in the same places as when all three thresholds vanish.[7]

However, the momentum-space analytic function $H(z)$ of the variables $z_i$ defined by

$$z_1 = -p^2, \tag{6a}$$

$$z_2 = -p'^2, \tag{6b}$$

$$z_3 = -(p - p')^2 \tag{6c}$$

does have a larger domain of analyticity as a consequence of the introduction of thresholds. Here, $p$ and $p'$ are the momentum-space vectors obtained by taking the Fourier transform of the time-ordered vacuum expectation value as follows:

$$H(z) = \frac{1}{(2\pi)^8} \iint dx'\, dx'' \times \exp\left[ip'(x' - x) + ip(x - x'')\right] F(\xi). \tag{7}$$

If the function $F(\xi)$ considered as function of $\xi$ [cf. Eqs. (2)] is regular analytic in $\mathfrak{U}'$ and properly bounded at infinity, the function $H(z)$ considered as function of the variables $z$ is also regular in the corresponding domain $\mathfrak{U}'$ of $z$.

The duality indicated above concerning the analyticity domains for the two functions $F(\xi)$ and $H(z)$ ceases when detailed mass spectrum properties are assumed. We have already mentioned that the function $F(\xi)$ may still be singular at any boundary point of $\mathfrak{U}'(\xi)$. However, in momentum space, the mass thresholds give new lower limits on the three cuts, and the analyticity domain of $H(z)$ can be extended through the corners exposed by the withdrawal of the cuts. The resulting holomorphy envelope $\mathfrak{U}''$ has not yet been determined.

The investigation described in this paper was motivated by the desire to locate the domain $\mathfrak{U}''$. A general superposition of the Mercedes and triangle diagrams such that each term satisfies the threshold conditions separately is analytic in a domain $\mathfrak{U}'''(z)$ which is at least as large as $\mathfrak{U}''(z)$. A natural conjecture is that $\mathfrak{U}'''(z)$ might be equal to $\mathfrak{U}''(z)$.

In this paper we show that this conjecture is false, i.e., that $\mathfrak{U}'''(z)$ is actually larger than $\mathfrak{U}''(z)$. Thus the most general three-point function satisfying nonzero threshold conditions can be obtained as a superposition of Mercedes and triangle diagrams only by including terms which violate the threshold conditions separately. (The unallowed singularities of these terms must, of course, cancel in the superposition.) We prove this result by considering a special example in some detail and showing that, for this example, the representation (1) is not able to produce the necessary momentum-space singularities if the thresholds are introduced in the way indicated above. The special case in question is one which, among many other topics, was studied by Brown.[5] It is obtained when only one threshold is different from zero. If we suppose that it is the quantity $M_3$ which is nonzero while $M_1 = M_2 = 0$, we have several qualitatively different configurations of the variables $z_i$ to consider. When $z_1$ and $z_2$ lie in opposite half planes, Brown was able to find the corresponding piece of the boundary of $\mathfrak{U}''$ by studying the ordinary triangle diagram. He was also able to show that the singularities produced by the triangle diagram when $z_1$ and $z_2$ lie in the same half-plane are sometimes a finite distance away from the boundary of an upper bound of $\mathfrak{U}''$. This upper bound is obtained by displacing the boundary curve $F'_{12}$ for the zero threshold case[2] by the amount $M_3^2$ in the variable $z_3$. The curve obtained in this way, the displaced $F'_{12}$ or the $\hat{F}'_{12}$, is an analytic hypersurface which intersects the cut above the threshold. Therefore, it defines, together with the cut, a natural domain of holomorphy which is an upper bound for $\mathfrak{U}''$. As the slope of the asymptote in the lower half-plane for $z_3$ is the same for $F'_{12}$ and $\hat{F}'_{12}$ it follows that the boundary of $\mathfrak{U}''$ also has this slope. (Note that the domain $\mathfrak{U}$ which has $F'_{12}$ as a boundary is a lower bound for $\mathfrak{U}''$.) By explicitly studying the singularity manifold of each term in

[7] Cf. Ref. 3.

(1) separately, we show that its contribution to the function $H(z)$ has a regularity domain with a boundary which, on every Riemann sheet, has an asymptotic slope different from the slope of $F'_{12}$ and $\hat{F}'_{12}$. The only exceptional case is a curve which turns out to be identically the same as $F'_{12}$ and it has been shown by Brown[5] that no piece of this curve can be part of the boundary of $\mathfrak{U}''$ for a nonzero value of $M_3$. This general technique allows us to reach our conclusion without any discussion of the Riemann sheet to which the singularities obtained from formal pinching arguments belong.

## 2. THE LEADING MOMENTUM-SPACE SINGULARITY OF THE MERCEDES DIAGRAM

Apart from a constant factor, the momentum-space function of the Mercedes diagram in Fig. 1 is given by

$$H_M(z;a;v) = \iiint \frac{dq\,dq'\,dQ}{(Q^2+a)[(Q-p-q)^2+b][(Q-p'-q')^2+c][q^2+v_c][q'^2+v_b][(q-q')^2+v_a]}. \quad (8)$$

Here, $p$ and $p'$ are the two external energy momentum vectors of the diagram. The complex variables $z$ are related to these vectors by Eqs. (6). The singularities of the denominators in Eq. (8) are interpreted with conventional, infinitesimal negative imaginary parts of all the mass squares $a$, $b$, $c$, etc. The standard technique to evaluate the integral (8) is to write the product of the denominators as an integral over one denominator depending on six Feynman parameters. Afterwards, the integrations over the vectors $q$, $q'$, and $Q$ can be explicitly performed and yields a characteristic denominator raised to a certain power. This power itself depends on the number of dimensions in the space of the vectors. Apart from a numerical factor this is the only way in which the dimensionality of the Lorentz space enters. As singularities in the function $H(z;a;v)$ are produced by pinches of poles in the integrand, it follows that the position of these singularities is independent of the number of dimensions in the Lorentz space. (The exact form of the singularity does depend on the number of dimensions but does not concern us here.) The only complication which could possibly occur would be if the dimensionality of the Lorentz space is chosen so low that the external vectors are linearly dependent. As we here have two independent vectors ($p$ and $p'$) it is enough to consider the function $H$ in Eq. (8) in a two-dimensional Lorentz space with one time dimension and one space dimension. For this case, the original integral (8) is itself six-dimensional and it is not particularly convenient to use the standard technique to recast one six-dimensional integral into another equally complicated form. Indeed we find it easier to discuss the pinching of the denominators in Eq. (8) directly in the space of the two-dimensional vectors of integration.

The leading singularity of the function $H$ in Eq. (8), i.e., the singularity which is characteristic of the Mercedes diagram itself and not present in contracted diagrams with a smaller number of internal lines, is obtained at the "big pinch" where all the denominators vanish simultaneously. We are going to call this singularity the "Landau" singularity.[8] Actually, the vanishing of all the denominators in (8) imposes six algebraic relations between the six components of the three vectors of integration. Consequently, all values of the external vectors $p$ and $p'$ normally yield one or more sets of vectors $q$, $q'$, and $Q$ such that all the denominators in (8) vanish at the same time. To be more specific, let us first determine the vector $q$ from the simultaneous vanishing of the denominators involving $b$ and $v_c$. In general, there are two vectors $q$ of this kind, viz.

$$q = [2(Q-p)^2]^{-1}\{((Q-p)^2 + b - v_c)(Q-p) \pm (\hat{Q}-\hat{p})[\lambda(-(Q-p)^2, b, v_c)]^{\frac{1}{2}}\}, \quad (9)$$

where the symbol $\hat{p}$ indicates the vector which is obtained from $p$ by the interchange of space and time components. Further, the quantity $\lambda(a, b, c)$ is the quadratic expression

$$\lambda(a,b,c) = a^2 + b^2 + c^2 - 2ab - 2ac - 2bc. \quad (10)$$

Note that the "dual pseudovector" $\hat{p}$ has the following properties

$$p\hat{p} = 0, \quad (11a)$$

$$\hat{p}^2 = -p^2, \quad (11b)$$

$$p\hat{q} = -q\hat{p}, \quad (11c)$$

$$\hat{\hat{p}} = p. \quad (11d)$$

From the vanishing of the two denominators involving $c$ and $v_b$ we get an expression analogous to (9) but with $q$ replaced by $q'$, $p$ by $p'$, $c$ by $b$, and

[8] L. D. Landau, Nucl. Phys. 13, 181 (1959).

$v_c$ by $v_b$. When these two formulas are introduced in the denominator involving $v_a$, we get an algebraic expression relating the vectors $Q$, $p$, and $p'$. After appropriate squarings to remove radicals, one obtains one functional relation which we denote as

$$F(Q, p, p', b, c, v_a, v_b, v_c) = 0. \tag{12}$$

The zeros of this function describe the possible singularities coming from the total pinch of the five denominators in the integrations over $q$ and $q'$. In the following integration over $Q$, these singularities are going to pinch with each other and with the remaining denominator $Q^2 + a$, thereby producing the singularities in the final result. Whether or not the final function is really singular at a point determined in this way depends on which Riemann sheet we are considering, i.e., whether or not the path of integration is actually trapped between the singularities of the integrands in the successive integrations. For the moment we are not concerned about these problems but only want to determine all possible $p$ and $p'$ which give a singularity on some Riemann sheet.

A pinch in the integrations over the first of the two components of the vector $Q$ and involving both the denominators $F$ and $Q^2 + a = 0$ can be used to determine and to eliminate this component of $Q$. In the next and last integration we must then ask for a double root of the remaining algebraic denominator. Another way to state this situation is to say that we ask for a double root in, e.g., $F$ considered as function of one of the components of $Q$ when the other component is determined from the vanishing of the denominator $Q^2 + a$. Most elegantly, this is formulated with the aid of a Lagrangian multiplier $L$ and the following equations:

$$G(Q) = F(Q, \cdots) + L(Q^2 + a) = 0, \tag{13a}$$

$$\nabla_Q G(Q) = \nabla_Q F + 2LQ = 0, \tag{13b}$$

$$Q^2 + a = 0. \tag{13c}$$

The multiplier $L$ can now be eliminated to give

$$F(Q, \cdots) = 0, \tag{14a}$$

$$\hat{Q} \nabla_Q F(Q, \cdots) = 0, \tag{14b}$$

$$Q^2 + a = 0. \tag{14c}$$

After the components of the vector $Q$ are eliminated between the three equations (14), the resulting algebraic expression $M(z; a; v) = 0$ gives the position of the Landau singularity for the Mercedes diagram. In practice, such an elimination leads to a very complicated algebraic expression and we prefer to use the parametric form (14) or expressions equivalent to it directly in our discussion below. However before we discuss the detailed properties of these singularities from the big pinch, we want to give a list of all other singularities of the Mercedes diagram obtained from lower-order pinches and show that, on the physical sheet, they are all inside the singularity domain obtained from the simple triangle diagram.

## 3. MOMENTUM-SPACE SINGULARITIES FROM LOWER-ORDER PINCHES IN THE MERCEDES DIAGRAM

In the last section we have discussed the singularities obtained when all the six denominators in the integral expression (8) vanished simultaneously and pinched. Other, lower order, singularities are obtained when at least one of the denominators is not involved. We have several cases to consider:

### 3.1. The Denominator $(q - q')^2 + v_a$ Does Not Participate in the Pinch

The position of the singularities obtained in this way is independent of the denominator containing $v_a$. Consequently, we can investigate a simpler integral obtained by omitting this denominator. The integration over the vector $q$ is now elementary and yields an expression $L(-(Q-p)^2, b, v_c)$ involving a logarithm.[9] This function has singularities at the points

$$(Q - p)^2 + (b^{\frac{1}{2}} + \epsilon v_c^{\frac{1}{2}})^2 = 0, \tag{15}$$

where $\epsilon^2 = 1$. The root corresponding to $\epsilon = -1$ does not occur on the physical sheet. The integration over the vector $q'$ is entirely analogous and yields the singularities

$$(Q - p')^2 + (c^{\frac{1}{2}} + \epsilon' v_b^{\frac{1}{2}})^2 = 0. \tag{16}$$

The integration over the vector $Q$ now involves exactly the same singularities as the simple triangle diagram with the mass-squares a, $(b^{\frac{1}{2}} + v_c^{\frac{1}{2}})^2$, and $(c^{\frac{1}{2}} + v_b^{\frac{1}{2}})^2$. To be more explicit, we can represent the logarithms yielding the singularities (15) and (16) by the following formula[9]

$$L(z, b, v) = 2i\pi \int_{M_{b,v}^2}^{\infty} \frac{du}{[\lambda(u, b, v)]^{\frac{1}{2}}} \frac{1}{u - z}, \tag{17}$$

$$M_{b,v} = b^{\frac{1}{2}} + v^{\frac{1}{2}}. \tag{17a}$$

Consequently, the simplified integral $H_M^{(s)}$ under discussion is given by[9]

[9] G. Källén and J. Toll, J. Math. Phys. 6, 299 (1965), especially Eqs. (7) and (25).

$$H_M^{(\bullet)}(z; a; v)$$

$$= -4\pi^2 \int_{M_{b\bullet c}^2}^{\infty} \int_{M_{c\bullet b}^2}^{\infty} \frac{du\, du'}{[\lambda(u, b, v_c)]^{\frac{1}{2}}[\lambda(u', c, v_b)]^{\frac{1}{2}}}$$

$$\times \int \frac{dQ}{(Q^2 + a)[(Q - p)^2 + u][(Q - p')^2 + u']}$$

$$= -4\pi^2 \int_{M_{b\bullet c}^2}^{\infty} \int_{M_{c\bullet b}^2}^{\infty} \frac{du\, du'}{[\lambda(u, b, v_c)]^{\frac{1}{2}}[\lambda(u', c, v_b)]^{\frac{1}{2}}}$$

$$\times \frac{1}{\Phi(z_1, z_2, z_3; u', u, a)} \sum_k P_k L(z_k, u_1, u_m)$$

$$\equiv -4\pi^2 \int_{M_{b\bullet c}^2}^{\infty} \int_{M_{c\bullet b}^2}^{\infty} \frac{du\, du'}{[\lambda(u, b, v_c)]^{\frac{1}{2}}[\lambda(u', c, v_b)]^{\frac{1}{2}}}$$

$$\times H_\Delta(z; u), \tag{18}$$

where

$$u_1 = u', \tag{18a}$$

$$u_2 = u, \tag{18b}$$

$$u_3 = a, \tag{18c}$$

$$P_k = \partial\Phi/\partial u_k, \tag{18d}$$

$$\Phi = \Phi(z_1, z_2, z_3; u_1, u_2, u_3)$$

$$= \sum_k u_k^2 z_k + \sum_k u_k u_1 (z_m - z_1 - z_k)$$

$$+ \sum_k u_k z_k (z_k - z_1 - z_m) + z_1 z_2 z_3. \tag{18e}$$

Here and in the following we will use the notation $k$, 1, $m$ for a cyclic permutation of 1, 2, 3.

Equation (18) shows explicitly how the function $H_M^{(\bullet)}$ can be written as a superposition of functions $H_\Delta$ defined in analogy with $H_M$ in Eq. (8) but corresponding to the triangle diagram instead of the Mercedes diagram. The singularities of this function are well-known.[2] Apart from cuts along the positive real axes in the variables $z_k$ and starting from the thresholds $(u_1^{\frac{1}{2}} + u_m^{\frac{1}{2}})^2$ it has singularities at the following zero of the manifold $\Phi$

$$z_k - u_1 - u_m = (-1/2u_k)$$

$$\times [(z_1 - u_m - u_k)(z_m - u_k - u_1) + R_1^{\frac{1}{2}} R_m^{\frac{1}{2}}], \tag{19}$$

$$R_k = \lambda(z_k, u_1, u_m). \tag{19a}$$

The square roots appearing in Eq. (19) are defined to behave as $R_k^{\frac{1}{2}} \sim z_k - u_1 - u_m$ for $|z_k| \to \infty$. The singularities (19) are relevant when $z_1$ and $z_m$ have the same sign for their imaginary parts and the imaginary part of $z_k$ has the opposite sign. All other solutions of the algebraic equation $\Phi = 0$ correspond to singularities which do not lie on the physical sheet.

The main conclusion to be drawn from the analysis above is that the function $H_M^{(\bullet)}$ has singularities on the principal sheet in a domain which is not larger than the domain covered by the singularities from the triangle diagram function $H_\Delta(z; u)$ where $u_1 \geq M_{b\bullet c}^2$, $u_2 \geq M_{c\bullet b}^2$, and $u_3 = a$. Therefore, the singularities from this lower-order pinch for the Mercedes diagram do not extend into a larger domain than that which has already been obtained from the triangle diagram.

### 3.2. The Denominator $q'^2 + v_b$ Does Not Participate in the Pinch

The integration over $q'$ is very similar to the previous case and yields the singularity (16) with $v_b$ replaced by $v_a$ and the vector $Q$ replaced by the vector $Q' = Q - q$. Introducing $Q'$ in place of $Q$ everywhere, the integration over $q$ involves only the two denominators with the mass-squares $a$ and $v_c$. Consequently, the resulting function has singularities on the manifold

$$Q'^2 + (a^{\frac{1}{2}} + v_c^{\frac{1}{2}})^2 = 0. \tag{20}$$

The modified Eq. (16), Eq. (20), and the original denominator involving $b$ again give an integration yielding "triangle singularities" or

$$\Phi(z_1, z_2, z_3; (c^{\frac{1}{2}} + v_a^{\frac{1}{2}})^2, b, (a^{\frac{1}{2}} + v_c^{\frac{1}{2}})^2) = 0. \tag{21}$$

Only that root of Eq. (21) which is defined in analogy with Eq. (19) above is relevant on the physical sheet.

### 3.3. The Denominator $q^2 + v_c$ Does Not Participate in the Pinch

This case is obtained from case in Sec. 3.2 by interchanging $p$ and $p'$, $b$ and $c$ and $v_b$ and $v_c$ and thus yields

$$\Phi(z_1, z_2, z_3; c, (b^{\frac{1}{2}} + v_a^{\frac{1}{2}})^2, (a^{\frac{1}{2}} + v_b^{\frac{1}{2}})^2) = 0, \tag{22}$$

where, again, only the root analogous to Eq. (19) is on the physical sheet.

### 3.4. The Denominator $(Q - p' - q')^2 + c$ Does Not Participate in the Pinch

In this case, the vector $p'$ does not appear in the singularity manifold which, therefore, depends only on the invariant $z_1$. From general principles it follows that such singularities can only lie on the positive real axis. An explicit argument of the same kind as used in Secs. 3.1, 3.2, and 3.3 above gives the following branch points:

$$z_1 - (a^{\frac{1}{2}} + \epsilon_1 b^{\frac{1}{2}} + \epsilon_2 v_c^{\frac{1}{2}})^2 = 0, \tag{23a}$$

$$z_1 - (a^{\frac{1}{2}} + \epsilon_1 b^{\frac{1}{2}} + \epsilon_3 v_a^{\frac{1}{2}} + \epsilon_4 v_b^{\frac{1}{2}})^2 = 0. \tag{23b}$$

The symbols $\epsilon_i$ are, as before, either equal to $+1$ or to $-1$. On the physical sheet, all the $\epsilon_i$ are equal to $+1$.

### 3.5. The Denominator $(Q - p - q)^2 + b$ Does Not Participate in the Pinch

This case is obtained from case in Sec. 3.4 by interchanging $p$ and $p'$, $b$ and $c$ and $v_b$ and $v_c$ and thus yields

$$z_2 - (a^{\frac{1}{2}} + \epsilon_1' c^{\frac{1}{2}} + \epsilon_2' v_b^{\frac{1}{2}})^2 = 0, \tag{24a}$$

$$z_2 - (a^{\frac{1}{2}} + \epsilon_1' c^{\frac{1}{2}} + \epsilon_3' v_a^{\frac{1}{2}} + \epsilon_4' v_c^{\frac{1}{2}})^2 = 0. \tag{24b}$$

### 3.6. The Denominator $Q^2 + a$ Does Not Participate in the Pinch

By introducing the vectors $Q' = Q - q$, $q'' = q - q'$, and $p'' = p' - p$ one finds that this case is analogous to Secs. 3.4 and 3.5 above and yields

$$z_3 - (b^{\frac{1}{2}} + \epsilon_1'' c^{\frac{1}{2}} + \epsilon_2'' v_a^{\frac{1}{2}})^2 = 0, \tag{25a}$$

$$z_3 - (b^{\frac{1}{2}} + \epsilon_1'' c^{\frac{1}{2}} + \epsilon_3'' v_b^{\frac{1}{2}} + \epsilon_4'' v_c^{\frac{1}{2}})^2 = 0. \tag{25b}$$

### 3.7. Two or More Denominators do not Participate in the Pinch

When two of the denominators do not participate in the pinch, one finds by arguments analogous to those presented above that one gets only the same branch points on the cuts as exhibited in Eqs. (23)–(25). When three or more denominators do not participate, the integral is independent of the external vectors. Therefore, such pinches do not give interesting singularities.

## 4. FURTHER DISCUSSION OF THE BIG PINCH

The argument in the previous section shows that the only possibility of obtaining singularities on the physical sheet from the Mercedes diagram outside the "triangle region" comes from the big pinch discussed in Sec. 2 and exhibited in Eqs. (14). To carry through the details of the corresponding algebra it is convenient to utilize as much as possible the fact that two of the masses $v$ vanish in the case of interest for the representation (1). Because of the inherent symmetry it is enough to study, e.g., the case $i = 3$, where $i$ is the summation index in Eq. (1). The other two terms in Eq. (1) can be obtained from this case by permutations.

To obtain the function $F$ in Eqs. (12) and (14) for this case we have to eliminate the four components of the vectors $q$ and $q'$ between the five equations

$$q^2 = 0, \tag{26a}$$

$$q'^2 = 0, \tag{26b}$$

$$(q - q')^2 + v = 0, \tag{26c}$$

$$(Q - p - q)^2 + b = 0, \tag{26d}$$

$$(Q - p' - q')^2 + c = 0. \tag{26e}$$

The first three of these equations have the solution

$$q_x = \epsilon q_0; \qquad \epsilon = \pm 1, \tag{27a}$$

$$q_x' = -\epsilon q_0' = \epsilon v/4q_0, \tag{27b}$$

$$q_0' = -v/4q_0. \tag{27c}$$

When Eqs. (27) are substituted in (26d) and (26e) and the number $q_0$ eliminated between the resulting expressions, one obtains

$$F = [(Q - p)^2 + b][(Q - p')^2 + c] - v[(Q - p)(Q - p') + \epsilon(Q - p)(\hat{Q} - \hat{p}')]. \tag{28}$$

Actually, it is sufficient to treat the case $\epsilon = +1$ in this equation. Afterwards, the case $\epsilon = -1$ can be obtained by a space reflection. Such a transformation does not change the scalar products of the external vectors which are our basic variables.

Putting $\epsilon = +1$, we next want to eliminate the two components of the vector $Q$ between the three equations (14) with the explicit form of (14a) given by (28). To do this it is convenient to introduce, instead of the components of $Q$, two invariant quantities $\tau$ and $\sigma$ given by the following definitions:

$$\tau = -\xi^{-1}(Qp + Q\hat{p}) = -z_2^{-1}(Qp' + Q\hat{p}'), \tag{29a}$$

$$\sigma = -(\xi/az_1)(Qp - Q\hat{p}) = -a^{-1}(Qp' - Q\hat{p}'), \tag{29b}$$

$$\xi = (pp' + \hat{p}p'). \tag{29c}$$

The two quantities $\tau$ and $\sigma$ are essentially ratios of light cone coordinates[9] of the vectors $Q$, $p$, and $p'$. Equation (14c) says that the product of these two variables is equal to one

$$\tau\sigma = 1. \tag{30}$$

Eliminating one of the variables, say $\sigma$, between Eqs. (14a) and (14c) we find a relation which the remaining variable $\tau$ has to fulfil. After some algebraic manipulations we get

$$\tau^2\xi^2 A + \tau\xi B + C = 0, \tag{31}$$

$$A = \tau^2 z_2 + \tau(c - a - z_2) + a, \tag{31a}$$

$$B = A(b - a - z_1) + vD, \tag{31b}$$

$$C = z_1(aA - vD), \tag{31c}$$

$$D = \tau(a - z_2\tau). \tag{31d}$$

In terms of these new definitions, Eq. (14b) is equivalent to the condition that the expression (31) has a double root when considered as function of $\tau$ for $\xi$ fixed, or, in other words, that the derivative of this expression with respect to $\tau$ vanishes,

$$\xi^2 \frac{d}{d\tau}(\tau^2 A) + \xi \frac{d}{d\tau}(\tau B) + \frac{dC}{d\tau} = 0. \tag{32}$$

An explicit algebraic elimination of the variable $\tau$ between Eqs. (31) and (32) has not proved very practical. It can be performed and leads to a polynomial in $z_1$, $z_2$, and $\xi$ which is of the eighth degree in $\xi$. The variable $z_3$ is related to the variable $\xi$ through

$$z_3 = z_1 + z_2 - \xi - z_1 z_2/\xi. \tag{33}$$

The polynomial for $\xi$ obtained in this way is rather complicated. Here, we only exhibit the eighth degree term in $\xi$ and the term which is independent of $\xi$

$$a^2\lambda(z_1, a, b)\lambda(z_2, a, c)\xi^8 + \cdots + (z_1 z_2)^4(a + v)^2 \times \lambda(z_1, b, a + v)\lambda(z_2, c, a + v) = 0. \tag{34}$$

All the coefficients not written out explicitly in Eq. (34) are polynomials in the variables $z_k$, $a$, $b$, $c$, and $v$. In particular, they involve no denominators. Finally, it should be remarked that the algebraic properties of the surface $M(z; a; v)$ are usually more easily studied with the aid of the parametric representation (31), (32), and (33) than from the explicit and rather involved eighth-degree polynomial (34).

Equations (31), (32), and (33) are arranged in a form where it is convenient to study the variable $z_3$ as a function of the other quantities. If we, instead, want to consider, e.g., $z_1$ as the dependent variable we have to rearrange Eq. (33) and solve for $z_1$. This gives

$$z_1 = \xi + z_3 + z_2 z_3/(\xi - z_2). \tag{35}$$

When this expression is substituted in Eqs. (31) one finds after some algebraic rearrangements

$$\alpha\xi^2 + \beta\xi + \gamma = 0, \tag{36}$$

$$\alpha = A\tau(\tau - 1), \tag{36a}$$

$$\beta = -z_2\alpha + \Lambda[\tau(b - a - z_3) + a] + vD(\tau - 1), \tag{36b}$$

$$\gamma = \Lambda[\tau(a - b)z_2 + a(z_3 - z_2)] - vD(\tau z_2 + z_3 - z_2). \tag{36c}$$

In this case we have to replace Eq. (32) by the following condition

$$\xi^2 \frac{d\alpha}{d\tau} + \xi \frac{d\beta}{d\tau} + \frac{d\gamma}{d\tau} = 0. \tag{37}$$

It turns out that if one eliminates one of the variables $\tau$ or $\xi$ between the two Eqs. (36) and (37) one obtains a polynomial of the twelfth-degree instead of the eighth-degree polynomial obtained from Eqs. (31) and (32). To obtain the manifold $M(z; a; v)$ in this way one then substitutes the solution for $\xi$ in Eq. (35) to get $z_1$. The explicit expression for the twelfth-degree equation for $\xi$ is somewhat complicated. It can be formally simplified if we introduce the quantity $\eta = \xi - z_2$ as variable instead of $\xi$. This is also a convenient variable as $z_1$ becomes infinite when $\eta$ is either infinite or zero. The expression analogous to Eq. (34) now reads

$$\eta^{12}c^2\lambda(z_2, a, c) + \cdots + z_2^6 z_3^6(c - v)^2\lambda(z_2, a, c - v) = 0. \tag{38}$$

Finally, if one wants to consider $z_2$ as dependent variable, one can use Eqs. (36) and (37) but with $z_2$ replaced by $z_1$ and $b$ and $c$ interchanged. The quantity $z_2$ itself is then obtained from (35) with $z_1$ replacing $z_2$ on the right-hand side.

The formalism developed in this section is sufficient to discuss all the algebraic properties of the surface $M(z; a; v)$ which we need. Here, we only mention that in the limit when $v$ goes to zero the surface $M(z; a; 0)$ includes the manifold $\Phi(z_1, z_2, z_3; c, b, a)$. To see this we remark that the expression in Eq. (31) for $v = 0$ simplifies to

$$[\tau^2\xi^2 + \tau\xi(b - a - z_1) + z_1 a] \times [\tau^2 z_2 + \tau(c - a - z_2) + a] = 0. \tag{39a}$$

If the two roots of $\tau$ which are equal both come from the first factor, we must have

$$\lambda(z_1, a, b) = 0, \tag{39b}$$

or

$$z_1 = (a^{\frac{1}{2}} \pm b^{\frac{1}{2}})^2 = M_{ab}^2 = \text{real number}. \tag{39c}$$

These solutions do not appear when we solve $M(z; a; v)$ for $z_3$ or for $z_2$. However, they do appear as limiting cases when we solve for $z_1$. Straight forward calculations show that the limiting values of the numbers $\tau$ and $\xi$ are

$$\tau = (M_{ab}^2 + a - b)/4z_2 M_{ab}^2 \times [M_{ab}^2 + z_2 - z_3 \pm \{\lambda(z_2, z_3, M_{ab}^2)\}^{\frac{1}{2}}], \tag{40a}$$

$$\xi = \tfrac{1}{2}[M_{ab}^2 + z_2 - z_3 \mp \{\lambda(z_2, z_3, M_{ab}^2)\}^{\frac{1}{2}}]. \tag{40b}$$

We note that the result given in Eqs. (40) corresponds to four values of $\tau$ and $\xi$ as the mass $M_{ab}$ can assume the two values indicated in Eq. (39c). Consequently, four of the twelve roots for $z_1$ mentioned after Eq. (37) collapse in pairs to the threshold values (39c) in the limit when $v$ goes to zero. The remaining eight roots as well as the eight roots of Eq. (34) go over into the manifold $\Phi(z_1, z_2, z_3; c, b, a)$ and correspond to the simultaneous vanishing of the two factors in Eq. (39a). The corresponding values of $\tau$ and $\xi$ are given by

$$\tau = (1/2z_2)(z_2 + a - c + \epsilon R_2^{\frac{1}{2}}), \tag{41a}$$

$$\xi = (1/4a)(z_2 + a - c - \epsilon R_2^{\frac{1}{2}})(z_1 + a - b + \epsilon' R_1^{\frac{1}{2}}). \tag{41b}$$

As before, $\epsilon$ and $\epsilon'$ are equal to $\pm 1$ independently. We have assumed that a is different from zero in Eq. (41b). Each one of the four values of $\xi$ given in (41b) are double roots of Eq. (34) in the limit when $v$ goes to zero. For small values of $v$ each double root splits up in two roots which differ by a term proportional to $v^{\frac{1}{2}}$. The corresponding roots of $\tau$ and $z_3$ or $z_1$, respectively, are fourfold in the limit when $v$ goes to zero and differ by terms proportional to $v^{\frac{1}{2}}$ for small values of $v$. The actual values of $z_1$ and $z_3$ for $v = 0$ are given by Eq. (19) with $k = 1$ or 3, respectively, and appropriate assignments of the variables $u_i$ in terms of $a$, $b$, and $c$.

Taken literally, the discussion above is relevant only for one term in the representation (1), viz. the term with $i = 3$. To obtain the term with $i = 1$, we have to permute $z_1$ and $z_3$ as well as the mass squares $a$ and $c$. To obtain the term with $i = 2$ we permute instead, the variables $z_2$ and $z_3$ and the mass squares $a$ and $b$.

## 5. THE MANIFOLD $M(z; a; v)$ FOR $z_1$ AND $z_2$ FIXED AND $z_3$ VERY LARGE

As was mentioned in the introduction, we are particularly interested in the situation when two of the variables $z_k$ have fixed values with the same sign for their imaginary parts and the third variable $z_k$ goes to infinity in the opposite half plane. Because of the inherent asymmetry of the manifold $M(z; a; v)$, we have to discuss two cases, one of which is obtained when the imaginary parts of $z_1$ and $z_2$ have the same sign while the imaginary part of $z_3$ has the opposite sign. The second case is obtained when, e.g., $z_2$ and $z_3$ have the same sign for their imaginary parts while $z_1$ lies in the opposite half-plane. In this section we discuss the first of these two cases. From the work of Brown[5] follows that the boundary of $\mathfrak{U}''$ for this case has an asymptote with a slope parallel to the complex number $-z_1z_2$ also in the case when $M_3 \neq 0$ but $M_1 = M_2 = 0$. Therefore, we are interested in the asymptotic form of the singularity manifold $M(z; a; v)$ when $z_3$ is very large and $z_1$ and $z_2$ are fixed complex numbers in, e.g., the upper half-plane. We want to compare the slope $-z_1z_2$ found by Brown with the slope of the boundary of the regularity domain for all values of $a$, $b$, $c$, and $v$ consistent with the threshold condition $b^{\frac{1}{2}} + c^{\frac{1}{2}} = M_3$.

From Eq. (33) follows that the variable $z_3$ goes to infinity (for fixed values of $z_1$ and $z_2$) when the number $\xi$ is either very large or very small. From Eq. (34) we conclude that $\xi$ can be infinite only when at least one of the mass squares $a$, $b$, $c$, or $v$ is infinite or when the coefficient of $\xi^8$ vanishes. For complex values of $z_1$ and $z_2$ the eighth-degree coefficient can vanish for $a = 0$. Further, $\xi = 0$ is a root of Eq. (34) only when the $\xi$-independent term vanishes or when at least one of the quantities $a$, $b$, $c$, or $v$ becomes infinite. If the imaginary parts of $z_1$ and $z_2$ are different from zero, the $\xi$-independent term in (34) vanishes only for $a = v = 0$.

### 5.1. The Limit When at Least One of the Mass Squares $a$, $b$, or $c$ is Very Large

The solutions of the algebraic equations (34) and (38) all correspond to singularities on some Riemann sheet but all of them are not expected to lie on the "physical sheet." A detailed discussion of the particular Riemann sheet of a given singularity is, in general, very complicated. However, a few special cases can be easily handled without too explicit calculations. Consider the equation $M(z; a; v) = 0$ and pick one particular solution for the variable $z_3$ of the form

$$z_3 = f(z_1, z_2; a, b, c; v). \tag{42}$$

The function $f(z_1, z_2; a, b, c; v)$ is an analytic function of $z_1$ and $z_2$ regular everywhere except for isolated branch points. The value of $z_3$ given by (42) corresponds to a singularity of $H_M(z; a; v)$ only on certain Riemann sheets. If the point (42) is not a singular point of $H_M$ at an initial point on a given sheet, it remains a non-singular point of $H_M$ for a continous variation of $z_1$, $z_2$, and the mass squares, provided that the point (42) does not pass through another singularity of $H_M$ and provided that $z_1$ and $z_2$ do not pass through a branch point of $f$. This property will be exploited to show that, for sufficiently large values of at least one of the quantities $a$, $b$, and $c$, the solution (42) is never relevant

on the physical Riemann sheet. This "physical sheet" is defined by evaluating Eq. (8) with the vectors $p$, $p'$, and $p - p'$ all real and space-like (which implies all $z_k$ real and negative) and continuing analytically from there. To specify the physical sheet completely it is also necessary to indicate certain cuts. First, we have the cuts (23), (24), and (25) along the real axes. In addition, we also have cuts from the branch points (18), (21), and (22). The detailed position of these cuts does not matter for our argument except that they must always lie outside the domain $\mathfrak{U}''$. Evidently, such cuts are always possible.

In the limit when at least one of the quantities $a$, $b$, and $c$ goes to infinity, some of the roots (42) also become infinite. In this case, the leading term of Eq. (42) is of the form

$$z_3 = Kf^{(0)}(z_1, z_2) + \cdots , \tag{43}$$

where $K$ is a real quantity which goes to infinity. The function $f^{(0)}(z_1, z_2)$ stays finite in the limit. In general, it depends on $z_1$ and $z_2$[10] as well as on those mass squares and mass ratios which stay finite. Further, in the limit we are considering, two of the three thresholds $M_k$ go to infinity. At least one of these thresholds is either $M_1$ or $M_2$. For definiteness, let us assume that $M_2$ goes to infinity. We also know that when $z_1$ (or any of the variables $z_k$) is on the negative real axis, the domain $\mathfrak{U}''$ includes the product of the two other $z$-planes cut along the real axes from the thresholds (23), (24), and (25). Starting from a given configuration of $z_1$ and $z_2$ with both of these variables in the upper half-plane, we can now move $z_1$ and $z_2$ together in such a way that $f^{(0)}(z_1, z_2)$ remains fixed while $z_1$ is moving to some point on the negative real axis. We choose such a path for the variation of $z_1$ and $z_2$ that we avoid all branch points of the function $f^{(0)}(z_1, z_2)$ and such that $z_1$ does not cross any cut. Further, the path must be such that $z_2$ remains finite. Because of the algebraic character of the functions $f^{(0)}(z_1, z_2)$ and $M(z; a; v)$ this is always possible. When $z_1$ is on the negative real axis the only singularity in the $z_2$-plane is the cut starting from the very large number $M_2^2$. None of the branch points (18.e), (21), and (22) lie on the physical sheet. When $z_1$ is moving, the branch points in the $z_2$-plane also move. Normally, i.e., except for certain isolated values of $z_1$ which we also avoid, these branch points in the $z_2$-plane either remain at finite points or always stay near infinity when $K$ is large. Those branch points in the $z_2$-plane which remain at finite points never come up on the physical sheet as the branch cut on this sheet starts from a very large threshold. The branch points which are close to infinity are evidently not circled by $z_2$ during its motion. Consequently, the singularity of the function $H_M(z; a; v)$ remains on the sheet where it is when $z_1$ is negative and real, i.e., not on the physical sheet. This proves that all the solutions of the equation $M(z; a; v) = 0$ correspond to singularities which do not lie on the physical sheet when at least one of the variables $a$, $b$, and $c$ is very large.

We may note that the argument presented above is very general and makes very little use of the detailed form of the manifold $M(z; a; v)$. Therefore, it applies to the discussion of most vertex Feynman diagrams. In particular, we remark that it can be used to show that all roots of the equation $\Phi(z; a) = 0$ lie off the physical sheet as soon as at least one of the mass squares $a$, $b$, and $c$ is large enough. Intuitively, these results are very reasonable as the limiting case which we are considering corresponds to intermediate states with particles of very large masses. These particles should have only a very small influence at any finite momentum and thus are not expected to give rise to singularities for finite momenta. Our result actually proves more than this as we have shown that there are no singularities on the physical sheet when two of the three variables $z_k$ are finite even if the third variable $z_k$ goes to infinity.

### 5.2. The Limit When $v$ is Very Large and $a$, $b$, and $c$ finite

The intuitive argument at the end of the last paragraph suggests that also in this case all the singularities of $H_M(z; a; v)$ should lie off the physical sheet. However, the argument actually used in Sec. 5.1 does not cover the situation when only $v$ is very large as the thresholds (23a), (24a), and (25b) stay finite. Therefore, both $M_1$ and $M_2$ are finite. Consequently, we must handle the case of very large $v$ by more explicit calculations.

Equations (31) and (32) can be solved explicitly in the limit of very large $v$. After some elementary algebraic work one finds that the eight solutions in this limit become

$$\xi_{1,2} = (z_1 z_2/a)[1 \pm 2i(bc)^{\frac{1}{2}}/av], \tag{44}$$

$$\xi_{3,4} = -z_1 v/(z_1 - M_{ab}^2), \tag{45a}$$

$$M_{ab}^2 = (a^{\frac{1}{2}} \pm b^{\frac{1}{2}})^2, \tag{45b}$$

[10] The exceptional case when $f^{(0)}(z_1, z_2)$ is actually independent of $z_1$ and $z_2$ is uninteresting for our discussion as the corresponding asymptote in the $z_2$-plane is clearly inadmissable for $z_1$ and $z_2$ negative and is thereby eliminated as a possible boundary of $\mathfrak{U}''$.

$$\xi_{5,6} = -z_2 v/(z_2 - M_{ac}^2), \tag{46a}$$

$$M_{ac}^2 = (a^{\frac{1}{2}} \pm c^{\frac{1}{2}})^2, \tag{46b}$$

$$\xi_{7,8} = -v \pm 2i[v(b-a)(c-a)/a]^{\frac{1}{2}}. \tag{47}$$

The two roots $\xi_{1,2}$ correspond to the variable $z_3$ having the value

$$z_3 = z_1 + z_2 - a - z_1 z_2/a + O(v^{-\frac{1}{2}}). \tag{48}$$

In the limit when $v$ goes to infinity this is exactly the curve $F_{12}'$ mentioned in the introduction. It has been proved by Brown[7] that every point on this curve is inside $\mathfrak{U}''$ for $M_3 \neq 0$. Consequently, the singularities of $H_M(z; a; v)$ corresponding to the roots in Eq. (44) do not lie on the physical sheet. The roots $\xi_{3,4}$ in Eq. (45) and $\xi_{5,6}$ in Eq. (46) correspond to very large values of $z_3$. However, the slope of the asymptote of the curve described by $z_3$ for fixed $z_1$ and $z_2$, when $v$ goes to infinity depends on only one of the variables $z_1$ and $z_2$. Consequently, it is possible to find values of $z_1$ and $z_2$ such that either of these asymptotes lies inside $\mathfrak{U}''$. Therefore, at least for certain cases the corresponding singularity points do not lie on the physical sheet. In principle, it would be possible to argue by continuity from here and discuss whether or not these points ever come up on the physical sheet by passing through one of the cuts (18e), (21), or (22). However, we do not want to enter into these details as it is clear from the formulas given that the slope of the asymptotes under discussion is not the "correct" slope $-z_1 z_2$ which is required for the boundary of $\mathfrak{U}''$. This last observation also holds for the roots $\xi_{7,8}$ where the slope of the asymptotes is essentially horizontal. Consequently, none of the roots of the manifold $M(z; a; v)$ obtained in the limit when $v$ goes to infinity can give the asymptotic boundary of $\mathfrak{U}''$.

### 5.3. The Limit When $a$ Goes to Zero

The discussion in Secs. 5.1 and 5.2 takes care of all cases when the mass squares $a$, $b$, $c$, and $v$ are very large. As has been shown previously, the only remaining case when the quantity $\xi$ becomes very large is obtained in the limit when a goes to zero. Also for this case one finds that it is possible to obtain explicit formulae for the eight roots of Eqs. (31) and (32). An algebraic computation using an expansion technique in powers of a or $a^{\frac{1}{2}}$ and assuming $v \neq 0$ gives

$$\xi_{1,2} = z_2(z_1 - M_{bv}^2)/(z_2 - c), \tag{49a}$$

$$M_{bv}^2 = (b^{\frac{1}{2}} \pm v^{\frac{1}{2}})^2, \tag{49b}$$

$$\xi_{3,4} = z_1(z_2 - M_{cv}^2)/(z_1 - b), \tag{50a}$$

$$M_{cv}^2 = (c^{\frac{1}{2}} \pm v^{\frac{1}{2}})^2, \tag{50b}$$

$$\xi_{5,6} = (z_1 - b)(z_2 - c)/a \pm 2i(bcv/a)^{\frac{1}{2}}, \tag{51}$$

$$\xi_{7,8} = -\frac{vz_1z_2}{(z_1 - b)(z_2 - c)}\left[1 \pm 2i\left(\frac{a}{v}\right)^{\frac{1}{2}} \times \left[\frac{(z_1 - b)^2 - vb}{z_1(z_1 - b)}\right]^{\frac{1}{2}}\left[\frac{(z_2 - c)^2 - vc}{z_2(z_2 - c)}\right]^{\frac{1}{2}}. \tag{52}$$

By inspection one finds that all the roots listed above correspond to a finite $\xi$ and thus also to a finite $z_3$ when $a$ goes to zero with the exception of the two roots $\xi_{5,6}$ in Eq. (51). The value of $z_3$ obtained from $\xi_{5,6}$ approaches asymptotically a parabola with an axis which has the slope $-(z_1 - b)(z_2 - c)$. As the mass squares $b$ and $c$ are coupled by the threshold condition $b^{\frac{1}{2}} + c^{\frac{1}{2}} = M_3 \neq 0$, this slope is never equal to $-z_1 z_2$. Without discussing the Riemann sheet of this singularity we can state that, if it is on the physical sheet and corresponds to a part of the boundary of $\mathfrak{U}'''$ for some value of $b$ and $c$, it cannot be part of the boundary of $\mathfrak{U}''$.

There remains to discuss the case when both $a$ and $v$ go to zero simultaneously with a finite ratio. [The cases when this ratio is either zero or infinite can be obtained as special cases of Eqs. (49)–(52) in one case and Eqs. (41) in the other case.] By calculations similar to those indicated earlier one finds that some of the roots for $\xi$ and $\tau$ can be understood as special cases of the formulas given above. This is the case for the roots given in Eqs. (49) and (50). For the other roots one finds, when $a$ and $v$ are of the same order of magnitude,

$$\xi_{5,6} = \frac{(z_1 - b)(z_2 - c)}{a} - b\frac{z_2 - c}{z_1 - b} - c\frac{z_1 - b}{z_2 - c} \pm 2i(bcv/a)^{\frac{1}{2}}, \tag{53}$$

$$\xi_{7,8} = [z_1 z_2/(z_1 - b)(z_2 - c)](a^{\frac{1}{2}} \pm iv^{\frac{1}{2}})^2. \tag{54}$$

We note in passing that the limit of Eqs. (51) and (52) when $v$ goes to zero can be understood as special cases of Eqs. (53) and (54) when $a$ is much smaller than $v$. In a similar way, we can understand those limits of Eqs. (41) which are obtained when $a$ goes to zero and which do not agree with the limit when $v$ goes to zero in Eqs. (49) and (50), as special cases of Eqs. (53) and (54) with $v$ much smaller than $a$.

One finds that the two roots for $\xi$ which are given by Eq. (53) are very large while the two roots in Eq. (54) are very small. In either case $z_3$ is very large but the slopes of the curves obtained in the $z_3$-plane are given by $-(z_1 - b)(z_2 - c)$ in one case and by $(z_1 - b)(z_2 - c)(a^{\frac{1}{2}} \pm iv^{\frac{1}{2}})^{-2}$ in the other case. The first of these numbers is never equal to

$-z_1z_2$. For fixed values of $b$ and $c$ it is evidently always possible to find positive, real numbers $a$ and $v$ such that $(z_1 - b)(z_2 - c)(a^{\frac{1}{2}} \pm iv^{\frac{1}{2}})^{-2}$ has the same argument as $-z_1z_2$. However, these values of $a$ and $v$ can never correspond to the *boundary* of the singularity domain for $H_M(z; a; v)$. Small changes in $a$ and $v$ bring the point from this solution of $M(z; a; v) = 0$ inside the domain $\mathfrak{U}''$, where the corresponding singularity must not be on the physical sheet. By continuity arguments we then conclude that it has to be off the physical sheet also when the point in question lies outside $\mathfrak{U}''$ but close to its boundary. It can never come up on the physical sheet until it crosses a cut starting from one of the triangle singularities in Eqs. (18e), (21), or (22). However, these branch points are known to lie a large distance away from the boundary of $\mathfrak{U}''$ in the asymptotic region where $z_3$ is very large. Summarizing, we find that the singularities obtained when both $a$ and $v$ are very small never lie on the boundary of $\mathfrak{U}''$.

Our discussion so far has shown that the singularities of the Mercedes diagram always lie in such positions that they, asymptotically, are very far away from the curve $\hat{F}'_{12}$ mentioned in the introduction and known to be the boundary of an upper bound of $\mathfrak{U}''$ as long as we consider the case when $z_1$ and $z_2$ have the same sign for their imaginary parts while the imaginary part of $z_3$ has the opposite sign. The remaining case when, e.g., $z_2$ and $z_3$ lie in the same half-plane and $z_1$ in the opposite half-plane is discussed below.

As an illustration of the algebraic discussion above we show in Figs. 2 and 3 a numerical example of

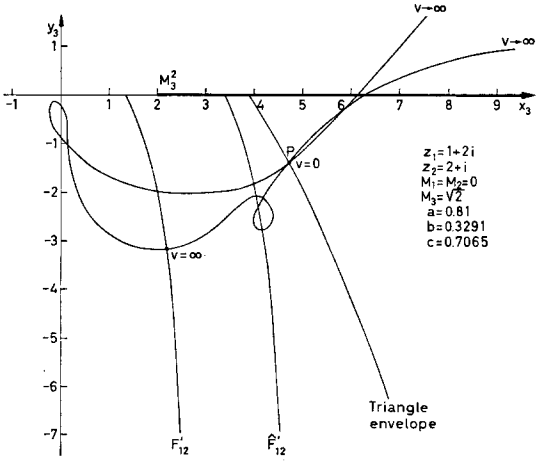


FIG. 2. Singularity manifolds corresponding to Eq. (34) plotted in the $z_3$-plane with $z_1 = 1 + 2i$, $z_2 = 2 + i$, $M_1 = M_2 = 0$, $M_3 = \sqrt{2}$, $a = 0.81$. The mass squares $b$ and $c$ are chosen in such a way that the point $v = 0$ lies on the triangle envelope. This implies $b = 0.3291$ and $c = 0.7065$.

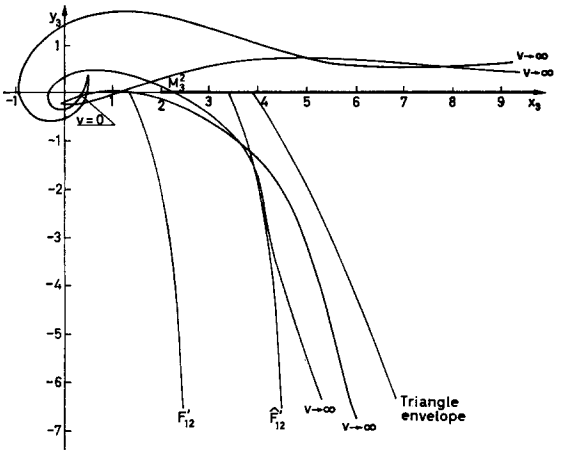


FIG. 3. The remaining roots of Eq. (34) for the same parameter values as in Fig. 2. This figure shows those roots which degenerate to the nonrelevant root of the Φ-manifold in the limit $v = 0$.

the eight solutions of the equation $M(z; a; v) = 0$ obtained for fixed values of $z_1$, $z_2$, $a$, $b$, and $c$. The mass square $v$ is varying from zero to infinity and we have plotted $z_3$ as function of $v$. The actual numerical values of the fixed parameters are shown in the plots. The mass squares $b$ and $c$ have been chosen in such a way that the point obtained for $v = 0$ and corresponding to the relevant root of Eq. (41) lies on the envelope of the singularity domain of the triangle diagram with the same thresholds as the Mercedes diagram. Figure 2 shows those four roots which start at this relevant triangle singularity for $v = 0$ while Fig. 3 shows the other four roots which start at the nonrelevant triangle singularity. In our example it happens that two of the roots in Fig. 2 end up at a point on the curve $F'_{12}$ and corresponding to Eq. (48). The other two roots go to infinity. By inspection one sees that only those two roots which lie above the point $P$ and below the positive real axis in Fig. 2 could be on the physical sheet and correspond to relevant singularities. However, both of these solutions are in a domain where we already have singularities from the triangle diagram and, therefore, they are not interesting for our discussion. By inspection one also sees that all the four roots exhibited in Fig. 3 must be off the physical sheet. For completeness, it should be remarked that the situation shown here may be changed for other values of, e.g., the parameter $a$. If this mass square is increased, it happens that one of the roots in Fig. 3 ends up on the curve $F'_{12}$ while one more of the roots in Fig. 2 goes off to infinity. In our numerical example this happens for a value of $a$ of about 0.88. In this particular case the equation $M(z; a; v) = 0$ has a double root for one particular value of $v$.

## 6. THE MANIFOLD $M(z; a; v)$ FOR $z_2$ AND $z_3$ FIXED AND $z_1$ VERY LARGE

Next, we consider the case when the two variables $z_2$ and $z_3$ are fixed in the upper half-plane and we solve our manifold $M(z; a; v)$ for $z_1$, i.e., use the formulas exhibited in Eqs. (36) and (37). Here, we are mainly interested in the case when $z_1$ is very large. The same technique which was used in Sec. 5 can be applied here too. First, we remark that the argument which was used in Sec. 5.1 to dispose of the case when at least one of the mass squares $a$, $b$, or $c$ becomes infinite can be used again and gives the result that the corresponding singularities of $H_M(z; a; v)$ do not lie on the physical sheet. Consequently, we consider only finite $a$, $b$, and $c$ below. The mass square $v$, however, may go to infinity. When $v$ is finite, we find from Eq. (38) that $\eta$ goes to infinity only for $c = 0$ and $\eta$ goes to zero for $c = v$ or for $\lambda(z_2, a, c - v) = 0$. The last condition can only be fulfilled for $v > c$ as long as $\mathrm{Im}(z_2) \neq 0$. As in Sec. 5, we consider these cases separately.

### 6.1. The Limit When $v$ is Very Large

This case can be discussed with the same method as we used earlier, e.g., in Sec. 5.2. After some algebraic work one finds that the 12 roots for $\xi$ and $\tau$ when $v$ is very large become

$$\xi_{1,2} = \frac{z_2 z_3}{a - z_2}\left[1 \pm \frac{2i}{a - z_2}\left(\frac{abc}{v}\right)^{\frac{1}{2}}\right] + z_2, \tag{55}$$

$$\xi_{3,4} = z_2 - z_3(1 + M_{ab}^2/v), \tag{56}$$

$$\xi_{5,6} = z_2 v/(M_{ac}^2 - z_2), \tag{57}$$

$$\xi_{7,8} = -v \pm [v(c - a)]^{\frac{1}{2}}, \tag{58}$$

$$\xi_{9,10} = (z_2 - a)/cv \pm (2/c)vz_3[(z_2 - a)^2 - ac]^{\frac{1}{2}}, \tag{59}$$

$$\xi_{11,12} = z_2 + \frac{z_2 z_3}{v} \pm \frac{2z_2 z_3}{v}\left(\frac{b}{v}\right)^{\frac{1}{2}}. \tag{60}$$

The quantities $M_{ab}$ in Eq. (56) and $M_{ac}$ in Eq. (57) are the same combinations which appeared in Eqs. (45b) and (46b). From Eq. (35) one finds that $z_1$ does not become infinite unless $\xi$ is either infinite or equal to $z_2$. By inspection one sees from the equations above that this happens for all cases above except the first four ones. However, the values for $z_1$ obtained from the eight interesting roots above are approximately given by

$$z_1 \cong vz_2/(M_{ac}^2 - z_2) \quad \text{from Eq. (57)}, \tag{61}$$

$$z_1 \cong -v \quad \text{from Eq. (58)}, \tag{61b}$$

$$z_1 \cong (v/c)(z_2 - a) \quad \text{from Eq. (59)}, \tag{61c}$$

$$z_1 \cong v \quad \text{from Eq. (60)}. \tag{61d}$$

Evidently, none of these roots corresponds to a possible boundary of $\mathfrak{U}''$. Consequently, even if some of the roots found here are on the physical sheet, they need not be discussed further here.

### 6.2. The Case When the Mass Squares $c$ and $v$ are Equal, Finite, and Nonvanishing

At the beginning of Sec. 6 we found that at least one root $\eta$ in Eq. (38) goes to zero for $c = v$ independently of the values of the other mass squares and of the values of $z_2$ and $z_3$. In this case, $z_1$ becomes infinite and we are interested in the slope of the corresponding asymptote. To find the value of $\tau$ for which this infinity occurs, we return to Eq. (36) and substitute $\xi = z_2$ in this expression. We find

$$(a - z_2\tau)[(a - z_2\tau)(1 - \tau) + \tau(c - v)] = 0. \tag{62}$$

From Eq. (62) follows immediately that we have a double root in $\tau$ given by

$$\tau = a/z_2 \tag{62a}$$

when $c = v$. Next, we use an expansion technique around this point and find with the aid of Eqs. (35), (36), and (37) the following behavior of $z_1$ in a neighborhood of $c = v$:

$$z_1 = \frac{z_2 z_3}{\eta} = -\frac{4c}{(c - v)^2}(z_2 - a)(z_3 - b). \tag{63}$$

Because the two mass squares $a$ and $b$ are coupled by the relation $a^{\frac{1}{2}} + b^{\frac{1}{2}} = M_1$, the asymptotic slope implied by Eq. (63) is always different from the slope $-z_2 z_3$ required for the boundary of $U''$. Consequently, the root of the manifold $M(z; a; v)$ which goes to infinity for $c = v$ is not interesting from our point of view even if it lies on the physical sheet.

A few remarks about the asymptotic solution presented in Eq. (63) may be appropriate. First of all, we note that the expression for $\eta$ in Eq. (63) behaves as $(c - v)^2$ in the limit when $c$ goes to $v$. From Eq. (38) we see that only one root of this kind occurs. Consequently, there is no other possibility for $z_1$ to become infinite than in the case considered here. We further note that the right-hand side of Eq. (63) becomes indeterminate if the common value of $c$ and $v$ is zero. Consequently, the limit when both $c$ and $v$ vanish is not adequately discussed by the formulas given here. We shall return to this question in a later section.

### 6.3. The Case $\lambda(z_2, a, c - v) = 0$

An entirely analogous discussion can be carried through for the second case mentioned at the be-

ginning of Sec. 6, viz. $\lambda(z_2, a, c - v) = 0$. From Eq. (62) we see that this situation corresponds to a double root in $\tau$ of the square bracket. More explicitly, this means

$$\pm z_2^{\frac{1}{2}} = \pm a^{\frac{1}{2}} \pm (c - v)^{\frac{1}{2}}. \tag{64}$$

If $z_2$ is a complex number, Eq. (64) can only be fulfilled if $v$ is larger than $c$ and we get

$$a = [\mathrm{Re}\, z_2^{\frac{1}{2}}]^2 = \tfrac{1}{2}[\mathrm{Re}\,(z_2) + |z_2|], \tag{64a}$$

$$v - c = [\mathrm{Im}\, z_2^{\frac{1}{2}}]^2 = \tfrac{1}{2}[-\mathrm{Re}\,(z_2) + |z_2|]. \tag{64b}$$

The corresponding value of $\tau$ obtained from Eq. (62) is

$$\tau = \tfrac{1}{2}[1 + |z_2|/z_2]. \tag{64c}$$

An expansion technique of exactly the same kind as used above now gives the following expression for $\eta$ in a neighborhood of the point (64)

$$\eta = \frac{z_3}{b - z_3} \frac{1}{2v} \frac{z_2 - |z_2|}{z_2 + |z_2|} [\epsilon_1(z_2 - |z_2|) - \epsilon_2(z_2 + |z_2|)], \tag{65a}$$

$$\epsilon_1 = a - \tfrac{1}{2}(\mathrm{Re}\,(z_2) + |z_2|), \tag{65b}$$

$$\epsilon_2 = v - c + \tfrac{1}{2}(\mathrm{Re}\,(z_2) - |z_2|). \tag{65c}$$

The corresponding value of $z_1$ becomes approximately

$$z_1 = \frac{z_2 z_3}{\eta} = -z_2(z_3 - b) \frac{z_2 + |z_2|}{z_2 - |z_2|} \times \frac{2v}{\epsilon_1(z_2 - |z_2|) - \epsilon_2(z_2 + |z_2|)}. \tag{66}$$

Clearly, the asymptotic slope implied by Eq. (66) is entirely different from the slope of $F'_{23}$ and we find that the corresponding curve, if it lies on the physical sheet, is certainly not the boundary of $\mathfrak{U}''$. The expansions given in Eqs. (65) and (66) are valid as long as $z_2$ and $z_3$ are complex numbers with nonvanishing imaginary parts and as long as the mass square $v$ is different from zero. The last condition is always fulfilled as long as Eqs. (64) are valid.

### 6.4. The Case When the Mass Square $c$ Vanishes

The discussion in the previous sections has disposed of all cases when the mass squares $a$, $b$, $c$, and $v$ either go to infinity or approach finite values in such a way that the complex number $z_1$ becomes infinite. Further, the case when the mass square $v$ alone goes to zero has been disposed of earlier in sec. 4. Consequently, we now have to study the remaining cases when either the mass $c$ alone goes to zero or when the masses $c$ and $v$ both approach zero simultaneously. In this section we discuss the first of these cases.

The special case when the mass square $c$ goes to zero can be discussed using Eqs. (36) and (37) in the same way as we have previously discussed Eqs. (31) and (32). However, the algebraic work here becomes somewhat more involved and it is not possible to give explicit values for all the roots in the same way as we have done before. Putting $c$ exactly equal to zero in Eq. (36) and introducing $\eta$ instead of $\xi$ we get

$$(a - z_2\tau)\{\eta^2\tau(\tau - 1)^2 + \eta[\tau(\tau - 1)(b - v - z_3) - (a - z_2\tau)(\tau - 1)^2] + \tau v z_3 + z_3(a - z_2\tau)(\tau - 1)\} = 0. \tag{67}$$

If the first factor $a - z_2\tau$ pinches with the expression inside the square bracket, one gets a second-degree equation for $\eta$. Actually, each root in that second-degree expression corresponds to a double root of Eq. (38) as one finds after considering also terms which are of first order in $c$ in Eq. (67). Using an expansion technique of the same kind as before, one gets the following expressions for the corresponding four roots:

$$\eta_i = \tfrac{1}{2}[z_2/(a - z_2)][v + z_3 - b + \epsilon\lambda^{\frac{1}{2}} + \epsilon'[2(cv)^{\frac{1}{2}}/\lambda^{\frac{1}{2}}]X^{\frac{1}{2}}] \quad \text{for} \quad i = 1, \cdots, 4, \tag{68}$$

$$X = (v - b - z_3 + \epsilon\lambda^{\frac{1}{2}})^2 + \epsilon a\lambda^{\frac{1}{2}}/(a - z_2)^2 \times [(v - b + \epsilon\lambda^{\frac{1}{2}})^2 - z_3^2], \tag{68a}$$

$$\lambda = \lambda(z_3, b, v), \tag{68b}$$

where $\epsilon$ and $\epsilon'$ as usual assume the values $\pm 1$ independently.

The condition that the big square bracket in Eq. (67) has a double root leads formally to an eighth degree equation in $\eta$. However, it so happens that the two highest-order terms in this eighth-degree expression have zero coefficients and, consequently, we get only six finite values of $\eta$ in this way. The detailed form of the sixth-degree equation is rather complicated. Here, we only give its highest- and lowest-order terms in the same way as we have indicated the corresponding terms in Eq. (38)

$$\eta_i^6\lambda(z_3, b, v) + \cdots + z_2^2 z_3^4\lambda(z_2, a, -v) = 0, \quad \text{for} \quad i = 5, \cdots, 10. \tag{69}$$

The remaining two roots of $\eta$ in Eq. (38) both behave as $1/c$. More explicitly, one finds

$$\eta_{11,12} = -(1/c)(z_2 - a)(z_3 - M_{bv}^2). \tag{70}$$

Clearly, 10 of the 12 roots obtained in the limit when $c$ goes to zero correspond to finite values of $\eta$. Two of these finite values can possibly be equal to zero. This happens if the last factor $\lambda(z_2, a, -v)$ in Eq. (69) vanishes. However, the roots obtained in this way are special cases of the roots discussed earlier in Sec. 6.3. The corresponding asymptotic slope is given by Eq. (66) with $c = 0$ and, consequently, uninteresting. The infinite roots for $\eta$ obtained in Eq. (70) correspond to an asymptotic slope not equal to the interesting slope $-z_2z_3$ except in the special case $a = 0$ and $b = v = M_1^2$. Consequently, we investigate this special case in some more detail. When substituting the mass values just mentioned in the original Eq. (36) but leaving $c$ arbitrary, we find after some simple rearrangements that the whole expression (36) can be written in the following way:

$$\tau^2\{\xi(\xi - z_2)z_2(1 - \tau)^2 + [z_2z_3 - c(\xi - z_2)]\xi(1 - \tau) + M_1^2c(\xi - z_2) - \xi cz_3 + M_1^2z_2z_3\} = 0. \quad (71)$$

The interesting roots for $\xi$ corresponding to Eq. (70) occur when the curly bracket in Eq. (71) has a double root. The exact condition for this to happen is given by

$$[c(\xi - z_2) + z_2z_3][c\xi^2 + \xi z_2(z_3 - c - 4M_1^2) + 4M_1^2z_2^2] = 0. \quad (72)$$

The two roots of $\xi$ which become very large when $c$ goes to zero are given by

$$\xi_{11} = (z_2/2c)\{c + 4M_1^2 - z_3 - [\lambda(z_3, c, 4M_1^2)]^{\frac{1}{2}}\} \approx z_2z_3/(z_3 - 4M_1^2) - z_2(z_3 - 4M_1^2)/c, \quad (73a)$$

$$\xi_{12} = z_2 - z_2z_3/c. \quad (73b)$$

The root given in Eq. (73a) evidently corresponds to the plus sign in $M_{b,v}$ in Eq. (70) and is uninteresting from our point of view as the asymptotic slope is incorrect. The result in Eq. (73b) yields exactly the curve $F_{23}'$ which has been shown by Brown[7] to be outside $\mathfrak{U}''$. It follows that the corresponding root of the manifold $M(z; a; v) = 0$ does not give a singularity on the physical sheet. This is also intuitively reasonable as we expect that only singularities associated with thresholds where the sum of the masses appears should be on the physical sheet (cf. Sec. 3). As a result of this analysis we are able to conclude that the infinities of $z_1$ obtained when $c$ alone goes to zero do not form part of the boundary of $\mathfrak{U}''$.

### 6.5. The Case When the Two Mass Squares $c$ and $v$ Go to Zero Simultaneously

Equation (38) shows that the only remaining possibility which might be interesting occurs when both $c$ and $v$ vanish simultaneously. We can treat this case with the aid of Eqs. (36) and (37). Straightforward calculations show that four of the roots for $\tau$ and $\xi$ are given by Eqs. (40) and correspond to $z_1 = M_{ab}^2$. Therefore, they are the same as four of the roots obtained for $v = 0$ but $c \neq 0$. The remaining eight roots are

$$\eta_{5,6} = -(1/c)(z_2 - a)(z_3 - b) \pm 2[(vb)^{\frac{1}{2}}/c](z_2 - a), \quad (74)$$

$$\eta_{7,8} = -z_3\frac{z_2 - a}{z_3 - b} \pm \frac{2z_3v^{\frac{1}{2}}}{(z_3 - b)^2} \times [b(z_2 - a)^2 - a(z_3 - b)^2]^{\frac{1}{2}}, \quad (75)$$

$$\eta_{9,10} = -z_2\left\{\frac{z_3 - b}{z_2 - a} - \frac{c}{(z_3 - b)(z_2 - a)^3} \times \left[(b(z_2 - a)^2 - a(z_3 - b)^2)^{\frac{1}{2}} \pm (z_2 - a)\left(\frac{bv}{c}\right)^{\frac{1}{2}}\right]^2\right\}, \quad (76)$$

$$\eta_{11,12} = -z_2z_3M_{cv}^2/(z_2 - a)(z_3 - b). \quad (77)$$

The symbol $M_{cv}$ is defined in analogy with, e.g., Eq. (45b). No assumption has been made about the relative magnitude of $c$ and $v$ in Eqs. (74)–(77). In particular, these roots agree with the limiting form of Eqs. (41) for the special case that $v$ goes to zero faster than $c$. Further, when $c$ is much smaller than $v$, Eqs. (70) and (74) agree. The four roots given in Eq. (68) correspond to the four roots in Eqs. (76) and (77) in the same limit.

By inspection one finds that only the roots given in Eqs. (74) and (77) correspond to very large values of $z_1$. In both cases the slope of the asymptote in the $z_1$-plane is given by $-(z_2 - a)(z_3 - b)$. As at least one of the mass squares $a$ and $b$ has to be different from zero because of the inequality $a^{\frac{1}{2}} + b^{\frac{1}{2}} \geq M_1$, this slope is never equal to the desired number $-z_2z_3$. Consequently, none of the roots obtained in this limit can give the boundary of $\mathfrak{U}''$.

As at the end of Sec. 5, we illustrate the algebraic discussion with numerical plots. Figures 4, 5, and 6 show an example analogous to the case illustrated in Figs. 2 and 3. Here, we have solved the manifold $M(z; a; v) = 0$ for $z_1$ but still consider $v$ as an independent variable while $a$, $b$, and $c$ are fixed. As before, the values of these three latter quantities are chosen in such a way that one of the points

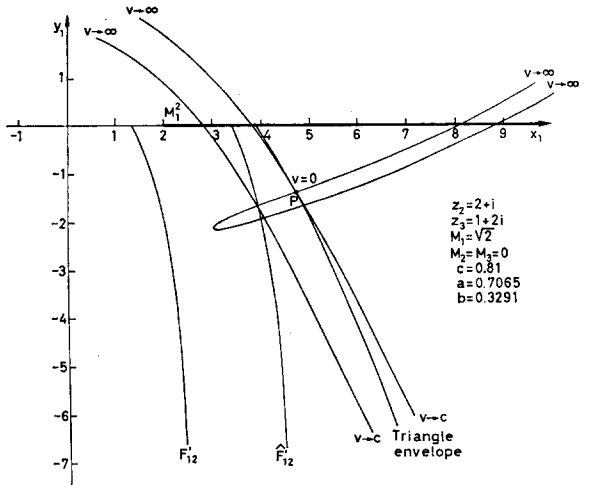


FIG. 4. Singularity manifolds corresponding to Eq. (38) plotted in the $z_1$-plane with $z_2 = 2 + i$, $z_3 = 1 + 2i$, $M_1 = \sqrt{2}$, $M_2 = M_3 = 0$, $c = 0.81$. The mass squares $a$ and $b$ are chosen in such a way that the point $v = 0$ lies on the triangle envelope.

obtained in the limit when $v$ goes to zero lies on the envelope of the triangle singularity domain. In Figs. 4 and 5 we show those roots which correspond to Eq. (41) in the limit when $v$ goes to zero, while Fig. 6 shows those four roots which correspond to Eqs. (39). In this case we note that one of the roots in Fig. 4 goes to infinity also for a finite value of $v$. Actually, this situation corresponds to the case discussed in Sec. 6.2 when $v$ becomes equal to $c$. As before, one finds that none of the roots shown in these figures can correspond to a singularity on the physical sheet and outside the domain where the triangle diagram already gives singularities.

## 7. DISCUSSION

The result proved above can be summarized in the statement that the momentum-space regularity domain $\mathfrak{U}'''$ of the function represented by the Bergman–Weil integral (1) with thresholds introduced in a simple minded way in terms of lower limits on the mass square variables $a$, $b$, and $c$ is larger than the momentum space domain $\mathfrak{U}''$ obtained from the general physical assumptions (i), (ii) and (iii) listed in the introduction together with the assumption of regularity when all variables $z_i$ have the same sign for their imaginary parts. This is a somewhat disappointing result from several points of view. First, it shows that the most general function regular in $\mathfrak{U}''$ cannot be simply represented with the same representation which works so well for the domain $\mathfrak{U}'$ obtained with zero thresholds. Second, our result is purely negative as it does not suggest any alternative technique to obtain the domain $\mathfrak{U}''$ or to represent the most general functions regular in $\mathfrak{U}''$. Of course, a function regular in $\mathfrak{U}''$ is also regular in $\mathfrak{U}'$ and, therefore, can be represented in the form (1) but with integrations over the mass squares $a$, $b$, and $c$ which extend from zero to infinity. The momentum-space support properties of $H(z)$ which follow from the detailed mass spectral conditions must then be taken care of by intricate cancellations between the four terms in (1) and is not fulfilled for each term separately.

An alternative question of some physical interest not discussed in this paper is the exact shape of the domain $\mathfrak{U}'''$. It is remarkable that the slopes of the asymptotes we have studied here and which have

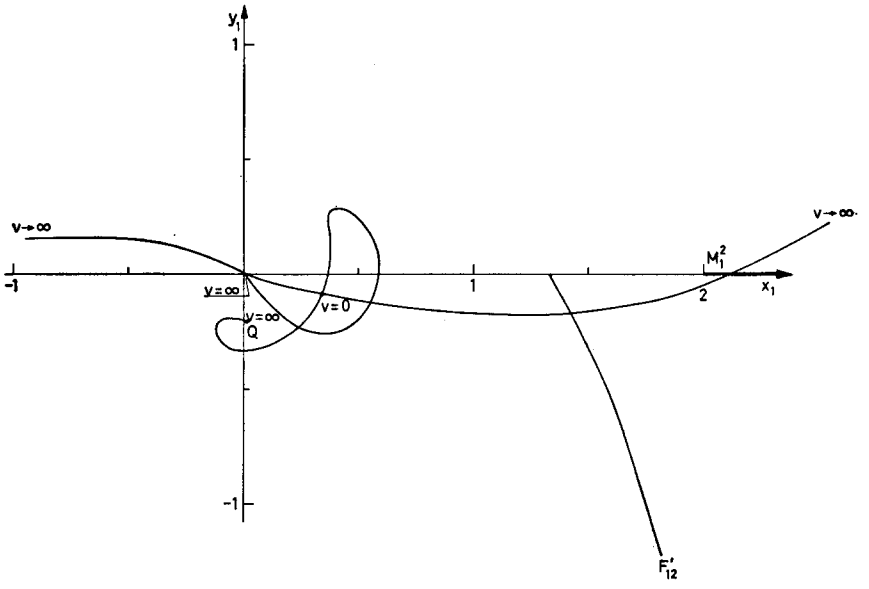


FIG. 5. Four other roots of Eq. (38) for the same parameter values as in Fig. 4. This figure shows those roots which degenerate to the non-relevant root of the Φ-manifold in the limit $v = 0$. Note that different scales are used in Figs. 4 and 5.

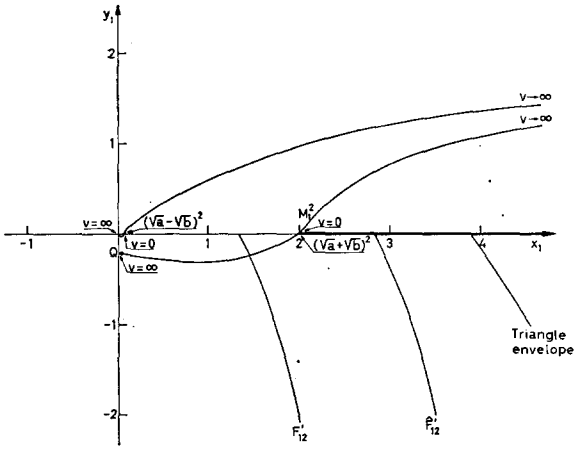


FIG. 6. The remaining roots of Eq. (38) for the same parameter values as in Figs. 4 and 5. This figure shows those roots which degenerate to the two thresholds $(a^{\frac{1}{2}} \pm b^{\frac{1}{2}})^2$ in the limit $v = 0$. Note the scale. The point $Q$ in this figure is the same as the point $Q$ in Fig. 5.

not been proved to be off the physical sheet have all been given by $-(z_1 - b)(z_2 - c)$ with $b^{\frac{1}{2}} + c^{\frac{1}{2}} \geq M_3$ when we consider $z_3$ as the dependent variable and by a symmetric result obtained when $z_1$ is considered as the dependent variable. This slope is exactly the same as the slope obtained in momentum space for the triangle diagram with thresholds introduced in a direct way. Consequently, our result does not contradict the conjecture that the domain $\mathfrak{U}'''$ for the Mercedes diagram with simple minded thresholds is the same as the momentum-space domain for the triangle with thresholds. To prove this conjecture one would have to investigate more carefully the Riemann sheet of the corresponding singularities and also to discuss the position of the singularities for finite values of the dependent variable. This cannot be done without a considerable effort and we do not feel that the present situation justifies such a calculation. The interest in the Mercedes diagram comes from its connection with the Bergman-Weil integral and once the complications associated with nonzero thresholds have been exhibited, we do not think it is worth while to carry the investigation further.

## ACKNOWLEDGMENTS

The research reported above has been carried out during a period of several years, alternatively at the Department of Physics and Astronomy, Maryland University, College Park, Maryland and the Department of Theoretical Physics, University of Lund, Lund, Sweden. It is a pleasure to acknowledge helpful discussion with several members of both these departments. The numerical calculations necessary to produce the graphs have been performed by the Computing Science Center at Maryland and the electronic computer SMIL in Lund. We especially want to thank J. Bohman and A. Ogus for help with the coding of the necessary programs.

# Green's Functions and Superfluid Hydrodynamics*

JOSEPH W. KANE† AND LEO P. KADANOFF‡
*Department of Physics, University of Illinois, Urbana, Illinois*
(Received 19 April 1965)

The two-fluid hydrodynamics of Landau and Tisza is related to the thermodynamic Green's function formulation of the many-body problem. A specific approximation for the self-energies yields directly the superfluid hydrodynamics in the limit of slow space and time variation. The approximation chosen is the simplest possible form for the self energies which includes the effects of collisions and which also satisfies the differential conservation laws for the mass, momentum, and energy. Of course these self energies are not adequate for a realistic description of $^4$He. Nonetheless the structure of this model may provide some insight into the exact theory. For example, in the model employed here the self-energies and Green's functions obey several integral identities which are utilized in the derivation of the two-fluid hydrodynamics, and which yield useful expressions for some thermodynamic quantities. It is speculated that the identities and the consequent expressions for the various thermodynamic quantities may have a wider range of validity than their derivation.

## 1. INTRODUCTION

We examine the nonequilibrium behavior of superfluid systems using the techniques of thermodynamic Green's functions. The standard Green's function methods[1,2] have been extensively employed in the determination of the equilibrium properties of the Boson system, and in the response to disturbances which vary relatively rapidly in space and time.[3-7]

However the calculation of the response of disturbances which vary quite slowly in space and time, i.e., the hydrodynamic limit, presents special complications since perturbation theory converges very slowly in this limit. Chapters 6–10 of Ref. 1 outline a procedure which can be applied to derive the hydrodynamic equations as the response of the single-particle Green's functions to an external field. In the long wavelength limit, the response is completely dominated by the flow of conserved quantities, i.e., mass, momentum, and energy. A qualitatively correct description of this flow can only be obtained if the approximation for the Green's functions is "conserving," i.e., includes the effects of the conservation laws in the sense of Ref. 8.

Hohenberg and Martin have computed the density correlation function[9] using the simplest conserving approximation for the one particle Green's function, that of Girardeau and Arnowitt.[10] This approximation however neglects collisions and they recognize that these collisions are necessary to enforce local thermodynamic equilibrium which implies hydrodynamic behavior. The work of Ref. 10 is then unsuitable for an understanding of hydrodynamic behavior; it is instead a discussion of the collisionless ($\omega\tau \gg 1$) domain. Several recent works[11-13] have discussed the differences between these domains.

We use an approximation which is one step more complex than that of Ref. 9. It is the simplest approximation which includes the two essential features of the hydrodynamic domain: collisions and the conservation laws. It is, of course, far too simple to represent the properties of He II, but we hope that deriving the two-fluid hydrodynamics of Landau[14,15] and Tisza[16] from this approximation

* Supported in part by Office of Naval Research Contract 1834(12).
† National Science Foundation Predoctoral Fellow.
‡ Alfred P. Sloan Foundation Fellow. Presently on leave from the University of Illinois at Cavendish Laboratory, University of Cambridge, Cambridge, England.
[1] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (W. A. Benjamin, Inc., New York, 1963).
[2] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963).
[3] C. De Dominicis and P. C. Martin, J. Math. Phys. (N. Y.) **5**, 14, 31 (1964).
[4] J. Gavoret and P. Nozières, Ann. Phys. **28**, 349 (1964).
[5] N. M. Hugenholtz and D. Pines, Phys. Rev. **116**, 489 (1959).
[6] S. T. Beliaev, Zh. Eksperim. i Teor. Fiz. **34**, 417 (1958) [English transl.: Soviet Phys.—JETP **7**, 289 (1958)].
[7] Chapter 5 of Ref. 2.
[8] G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).
[9] P. C. Hohenberg, thesis, Harvard University (1962) (unpublished).
[10] M. Girardeau and R. Arnowitt, Phys. Rev. **113**, 755 (1959).
[11] A. Miller, D. Pines, and P. Nozières, Phys. Rev. **127**, 1452 (1962).
[12] P. C. Hohenberg and P. C. Martin, Phys. Rev. Letters **12**, 69 (1964).
[13] P. Nozières and D. Pines, *Theory of Quantum Liquids*, to be published (W. A. Benjamin, Inc., New York).
[14] L. Landau, Zh. Eksperim. i. Teor. Fiz. **5**, 71 (1941).
[15] L. Landau and E. Lifshitz, *Fluid Mechanics* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1959), p. 510.
[16] L. Tisza, Nature **141**, 913 (1938).

will provide some indications of the structure of the exact theory.

In a recent technical report, Bogliubov[17] has described the relation of hydrodynamics to the exact Green's function theory. Our work differs from Bogliubov's in that we are restricted to a specific Green's function approximation. Thus we do not prove that the two-fluid hydrodynamic model is true, but instead attempt to provide a more detailed insight into just how the hydrodynamics emerge. In particular we can see explicitly that the superfluid velocity $v_s$ enters the theory as the gradient of the phase of the condensate wavefunction, while the normal fluid velocity $v_n$ enters as a parameter in the solution of the detailed balancing relation.

To obtain the two-fluid equations explicitly we make use of a set of Green's function identities which we believe are new. These "identities" are manifestly true in our approximation but probably have a wider range of validity than implied by our very simple approximation for the Green's function. Further discussion of this point must await further study.

In Sec. 2 of the paper we define the approximation for the condensate wavefunctions $\langle\Psi(1)\rangle$ and for the single-particle Green's function. Included in this section are the exact, real time equations of motion for the system. In Sec. 3 these equations are specialized to slow space and time variation, and are then taken as the generalized Boltzmann equations in the sense of Kadanoff and Baym.[1] In this section we must overcome one technical difficulty. As pointed out by Bogliubov[17] and others[12,18] the phase of the condensate wavefunction is rapidly varying in space and time, and must be treated with some care. A gauge transformation is made so that the rapid variation can be treated in a consistent way. In the course of performing this guage transformation we are led to introduce $v_s$ and the local chemical potential as the space and time derivations of the condensate phase.

In Sec. 4 we show that the conservation laws for mass, momentum, and energy can be correctly obtained from the Boltzmann equations of Sec. 3. Useful expressions for the energy density, energy current, and the stress tensor are obtained by using the Green's function identities mentioned above.

A procedure analogous to that of Chapman and Enskog[19] is used in Sec. 5 to obtain what we term the "local thermodynamic equilibrium" solutions of the Boltzmann equation. These solutions are correct in the hydrodynamic limit of extremely slow space and time variation of all physical quantities, i.e., for $\omega\tau \ll 1$.

The local equilibrium solutions are inserted into the conservation laws in Sec. 6 to obtain the two-fluid hydrodynamic equations. In the course of this analysis we use the identities of Sec. 4 to obtain several new expressions which give microscopic identifications of the quantities appearing in the two-fluid equations. A particularly useful expression for the entropy is thereby obtained.

## 2. GREEN'S FUNCTION APPROXIMATIONS

The superfluid system is conveniently described in terms of the spinor wave field operators[20]:

$$\Psi(1) = \begin{pmatrix} \psi(1) \\ \psi^\dagger(1) \end{pmatrix}, \qquad \Psi^\dagger(1) = \overbrace{\psi^\dagger(1)\psi(1)}, \tag{2.1}$$

which are used to define the correlation functions

$$\begin{aligned} h(1,1') &= -i\langle\Psi(1)\rangle\langle\Psi^\dagger(1')\rangle, \\ g^>(1,1') &= -i\langle\Psi(1)\Psi^\dagger(1')\rangle, \\ g^<(1,1') &= -i\langle\Psi^\dagger(1')\Psi(1)\rangle, \\ \tilde{g}^<(1,1') &= g^<(1,1') - h(1,1'). \end{aligned} \tag{2.2}$$

Here $\Psi(1)$ denotes the spinor wave field operator at the space–time point $(\mathbf{r}_1, t_1)$. The correlation functions are defined for real times as in Chap. 8 of Ref. 1. These functions then obey the equations of motion

$$\begin{aligned} &\int_{-\infty}^{\infty} d\bar{1}\{g_0^{-1}(1,\bar{1}) - \Sigma_{HF}(1,\bar{1})\}\tilde{g}^<(\bar{1},1') \\ &= \int_{-\infty}^{t_1} d\bar{1}\,\Gamma(1,\bar{1})\tilde{g}^<(\bar{1},1') - \int_{-\infty}^{t_{1'}} d\bar{1}\,\Sigma^<(1,\bar{1})a(\bar{1},1'), \\ &\int_{-\infty}^{\infty} d\bar{1}\,\tilde{g}^<(1,\bar{1})\{g_0^{-1}(\bar{1},1') - \Sigma_{HF}(\bar{1},1')\} \\ &= \int_{-\infty}^{t_1} d\bar{1}\,a(1,\bar{1})\Sigma^<(\bar{1},1') - \int_{-\infty}^{t_{1'}} d\bar{1}\,\tilde{g}^<(1,\bar{1})\Gamma(\bar{1},1'), \end{aligned} \tag{2.3}$$

$$\begin{aligned} &\int_{-\infty}^{\infty} d\bar{1}\{g_0^{-1}(1,\bar{1}) - S_{HF}(1,\bar{1})\}h(\bar{1},1') \\ &\qquad = \int_{-\infty}^{t_1} d\bar{1}\,\gamma(1,\bar{1})h(\bar{1},1'), \end{aligned}$$

[17] N. N. Bogoliubov, *The Question of the Hydrodynamics of the Superfluid Liquid* (Dubna, 1963). We wish to thank Dr. Bogoluibov for sending us this copy of his work.

[18] The analogous point for a superconductor was made by V. Ambegaokar and L. P. Kadanoff, Nuovo Cimento **22**, 914 (1961).

[19] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge University Press, London, 1960).

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

$$\int_{-\infty}^{\infty} d\bar{1}\, h(1,\bar{1})\{g_0^{-1}(\bar{1},1') - S_{HF}(\bar{1},1')\} = -\int_{-\infty}^{t_{1'}} d\bar{1}\, h(1,\bar{1})\gamma(\bar{1},1'),$$

where[21]

$$g_0^{-1}(1,1') = \left\{ i\tau^{(3)}\frac{\partial}{\partial t_1} + \frac{\nabla_1^2}{2} + \mu_0 - u(1)\right\}\delta(1,1'). \tag{2.4}$$

The limits on the integrals of (2.3) are those for the time integral. The remaining integrations are over all space. In these equations, $u(1)$ is an external potential coupled to the density which serves to drive the system away from equilibrium, $\mu_0$ is the constant chemical potential for the system, and $\tau^{(3)}$ is the Pauli spin matrix

$$\tau^{(3)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

$\Sigma_{HF}$ and $S_{HF}$ are the Hartree–Fock self energies for $\tilde{g}$ and $h$, respectively. These are shown diagrammatically in Fig. 1 and are given algebraically by

$$\begin{aligned} \Sigma_{HF}(1,1') &= \frac{i}{2}\int d\bar{2}\, v(1,\bar{2})\,\mathrm{tr}\, g(\bar{2},\bar{2})\delta(1,1') \\ &\quad + iv(1,1')g(1,1'), \\ S_{HF}(1,1') &= \frac{i}{2}\int d\bar{2}\, v(1,\bar{2})\,\mathrm{tr}\, g(\bar{2},\bar{2})\delta(1,1') \\ &\quad + iv(1,1')\tilde{g}(1,1'), \end{aligned} \tag{2.5}$$

where tr denotes the matrix trace. We have adopted the convention that at equal times, $(\psi(1)\psi^\dagger(1'))^+ = \psi^\dagger(1')\psi(1)$. The interparticle potential is $v(1,1') = v(|\mathbf{r}_1 - \mathbf{r}_{1'}|)\delta(t_1 - t_{1'})$ so $\Sigma_{HF}$ and $S_{HF}$ are local in time.

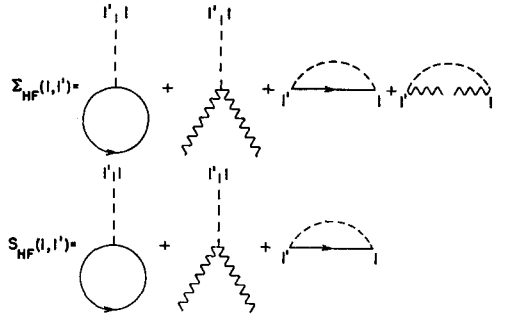


FIG. 1. The Hartree–Fock contributions to the self-energies. $_{1'}$— — —$_1$ denotes the two-body potential $v(1,1')$, $_{1'}$∿∿∿$_1$ denotes the single-particle condensate propagator $h(1,1')$, and $_{1'}$⟶⟶$_1$ the propagator $\tilde{g}(1,1')$.

[21] We use units in which $\hbar = m = 1$.

In Eqs. (2.3) we have defined

$$\begin{aligned} \Gamma(1,1') &= \Sigma^>(1,1') - \Sigma^<(1,1'), \\ \gamma(1,1') &= S^>(1,1') - S^<(1,1'), \\ a(1,1') &= \tilde{g}^>(1,1') - \tilde{g}^<(1,1'), \end{aligned} \tag{2.6}$$

where $\Sigma^\gtrless$ and $S^\gtrless$ are the remaining parts of the self energies of $\tilde{g}$ and $h$, respectively.

In our approximate analysis we choose $\Sigma$ and $S$ to be the simplest expressions which contain collisions between particles and satisfy the laws of conservation of mass, momentum, and energy. We include terms of second order in the interparticle potential but exclude exchange terms. This choice[9] is shown in Fig. 2 and has the analytical form

$$\begin{aligned} \Sigma^\gtrless(1,1') &= -\frac{1}{2}\int d\bar{2}\int d\bar{3}\, v(1,\bar{2})v(1',\bar{3}) \\ &\times \{\tilde{g}^\gtrless(1,1')\,\mathrm{tr}\,[\tilde{g}^\lessgtr(\bar{3},\bar{2})\tilde{g}^\gtrless(\bar{2},\bar{3}) + \tilde{g}^\lessgtr(\bar{3},\bar{2})h(\bar{2},\bar{3}) \\ &+ h(\bar{3},\bar{2})\tilde{g}^\gtrless(\bar{2},\bar{3})] + h(1,1')\,\mathrm{tr}\,[\tilde{g}^\lessgtr(\bar{3},\bar{2})\tilde{g}^\gtrless(\bar{2},\bar{3})]\}, \end{aligned} \tag{2.7}$$

$$\begin{aligned} S^\gtrless(1,1') &= -\frac{1}{2}\int d\bar{2}\int d\bar{3}\, v(1,\bar{2})v(1',\bar{3}) \\ &\times \tilde{g}^\gtrless(1,1')\,\mathrm{tr}\,[\tilde{g}^\lessgtr(\bar{3},\bar{2})\tilde{g}^\gtrless(\bar{2},\bar{3})]. \end{aligned} \tag{2.8}$$

This approximation has two obvious faults. First since it is only a portion of the second-order term in a perturbation expansion, it cannot be expected to be correct at He II densities. Nonetheless, from the structure of our result we hope to obtain some insight into the structure of the He II theory. Secondly it does not satisfy the Hugenholty–Pines[5] theorem, so the equilibrium single-particle function $\tilde{g}$ will not contain the phonon spectrum at low frequencies and wave number. However as pointed out by Hohenberg in Chap. 4 of Ref. 9, a properly gapless response can be obtained in the density-correlation function even from a $\tilde{g}$ which has a gap, if the approximation is conserving.

## 3. THE BOLTZMANN EQUATIONS

We specialize the exact transport equations of the system (2.3) to the case in which the external

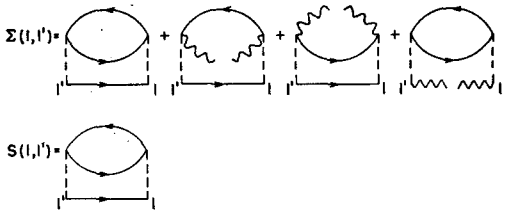


FIG. 2. The remaining parts of the self-energies in the approximation of this paper. The notation is the same as that of Fig. 1.

potential varies slowly in space and time. If the external potential induces a disturbance of wavelength much longer than thermal wavelengths and frequencies much smaller than characteristic particle energies, then in the normal system $g^<(1, 1')$ can be expected to vary slowly as a function of

$$\mathbf{R} = \tfrac{1}{2}(\mathbf{r}_1 + \mathbf{r}_{1'}), \qquad T = \tfrac{1}{2}(t_1 + t_{1'}), \tag{3.1}$$

and be peaked about the zeros of

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_{1'}, \qquad t = t_1 - t_{1'}. \tag{3.2}$$

This type of variation was exploited in Ref. 1. to derive a "generalized Boltzmann equation" for the normal system.

However in the superfluid there is a special difficulty. If $\langle\Psi(1)\rangle$ is rewritten in the form

$$\langle\Psi(1)\rangle = [n_0(1)]^{\frac{1}{2}} e^{i\varphi(1)\tau^{(3)}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{3.3}$$

then

$$h(1, 1') = -i[n_0(1)]^{\frac{1}{2}} e^{i\varphi(1)\tau^{(3)}} \times (1 + \tau^{(1)}) e^{-i\varphi(1')\tau^{(3)}} [n_0(1')]^{\frac{1}{2}} \tag{3.4}$$

where $n_0$ and $\varphi$ are real. In equilibrium the gradient of $\varphi$ is the superfluid velocity, and the time derivative of $\varphi$ is related to the chemical potential[12,17] Extending these identifications to the nonequilibrium sysstem, we identify

$$\begin{aligned} \nabla_1\varphi(1) &= \mathbf{v}_s(1), \\ \partial\varphi(1)/\partial t_1 &= -[\mu(1) - \mu_0 + \tfrac{1}{2}v_s^2(1)], \end{aligned} \tag{3.5}$$

where $\nabla_1\mu_0 = 0$. It can then be seen that $\varphi(\mathbf{R}, T)$ is a rapidly varying function of $(\mathbf{R}, T)$ and induces anomolously strong $(\mathbf{R}, T)$ variations in the off diagonal elements of $h$ and thus in $\tilde{g}$. So $h(1, 1')$ in the form (3.4) and the related $\tilde{g}(1, 1')$ will not satisfy a Boltzmann equation like that of Ref. 1. For this reason we perform a gauge transformation to remove the strong $\mathbf{R}, T$-dependence of $\varphi$.

Consider the gauge transformed correlation functions

$$\begin{aligned} h'(1, 1') &= e^{-i\varphi(1)\tau^{(3)}} h(1, 1') e^{i\varphi(1')\tau^{(3)}}, \\ \tilde{g}'^<(1, 1') &= e^{-i\varphi(1)\tau^{(3)}} \tilde{g}^<(1, 1') e^{i\varphi(1')\tau^{(3)}}. \end{aligned} \tag{3.6}$$

Since we expect the physical quantities $h'$, $\tilde{g}'$, $\mathbf{v}_s$, and $\mu$ to vary slowly as functions of $(\mathbf{R}, T)$ we can write generalized Boltzmann equations for these gauge-transformed correlation functions. Notice that $h'(1, 1')$ has the form

$$h'(1, 1') = -i[n_0(1)]^{\frac{1}{2}}(1 + \tau^{(1)})[n_0(1')]^{\frac{1}{2}} \tag{3.7}$$

and we interpret $\frac{1}{2}\operatorname{tr} h'(1, 1')\,|_{1'\to 1}$ as the probability that a particle, added to the system at the space–time point $(\mathbf{r}_1, t_1)$, enters the condensed state.

The gauge transformation leaves Eqs. (2.3) form invariant if $g_0^{-1}$ is replaced by

$$g_0'^{-1}(1, 1') = \left\{ i\tau^{(3)} \frac{\partial}{\partial t_1} - \frac{\partial\varphi(1)}{\partial t_1} + \tfrac{1}{2}[\nabla_1 + i\nabla_1\varphi(1)\tau^{(3)}]^2 + \mu_0 \right\} \delta(1, 1'). \tag{3.8}$$

Since we shall hereafter use the gauge-transformed functions exclusively, the primes will be dropped to simplify the notation. We have also dropped the external potential in (3.8). Our analysis can then be viewed as a description of the system as it returns slowly to equilibrium after being disturbed by the external potential.

To obtain the generalized Boltzmann equations consider the change of variables $(1, 1') \to (\mathbf{r}, t; \mathbf{R}, T)$ according to

$$\begin{aligned} &g^<(1, 1') \\ &= g^<[\mathbf{r}_1 - \mathbf{r}_{1'}, t_1 - t_{1'}; \tfrac{1}{2}(\mathbf{r}_1 + \mathbf{r}_{1'}), \tfrac{1}{2}(t_1 + t_{1'})] \\ &= g^<(\mathbf{r}, t; \mathbf{R}, T) \\ &= -i \int \frac{d^3p\, d\omega}{(2\pi)^4} \exp(i\mathbf{p}\cdot\mathbf{r} - i\omega t) g^<(\mathbf{p}, \omega; \mathbf{R}, T). \end{aligned} \tag{3.9}$$

Notice that it is again possible to split $g$ as

$$g^< = \tilde{g}^< + h$$

where, if we neglect second derivatives of $n_0(\mathbf{R}, T)$,

$$h(\mathbf{p}, \omega; \mathbf{R}, T) = n_0(\mathbf{R}, T)(1 + \tau^{(1)})(2\pi)^4 \delta(\mathbf{p})\delta(\omega). \tag{3.10}$$

The generalized Boltzmann equations are now obtained as the Fourier transforms of Eqs. (2.3) for the case in which the variation of $g^<(\mathbf{p}, \omega; \mathbf{R}, T)$ in $\mathbf{R}$ and $T$ is very slow. In particular we assume that $g_0^{-1}$ and the self-energies vary very little as $\mathbf{R}$ is changed by a characteristic excitation wavelength or $T$ is changed by an inverse excitation energy. By using a procedure exactly analogous to that of Ref. 1, Chap. 9, when we take proper account of the matrix character of the correlation functions we find

$$\begin{aligned} &g_0^{-1}\tilde{g}^< + \frac{i}{2}[g_0^{-1}, \tilde{g}^<] = \operatorname{Re}\Sigma\tilde{g}^< + \Sigma^<\operatorname{Re}\tilde{g} \\ &- \frac{i}{2}[\Gamma\tilde{g}^< - \Sigma^< a] + \frac{i}{2}\{[\operatorname{Re}\Sigma, \tilde{g}^<] + [\Sigma^<, \operatorname{Re}\tilde{g}]\} \\ &+ \tfrac{1}{4}\{[\Gamma, \tilde{g}^<] - [\Sigma^<, a]\}, \end{aligned} \tag{3.11}$$

$$\tilde{g}^< g_0^{-1} + \frac{i}{2}[\tilde{g}^<, g_0^{-1}] = \tilde{g}^<\operatorname{Re}\Sigma + \operatorname{Re}\tilde{g}\Sigma^<$$

$$-\frac{i}{2}[a\Sigma^{>}-\tilde{g}^{>}\Gamma]+\frac{i}{2}\{[\tilde{g}^{>},\operatorname{Re}\Sigma]+[\operatorname{Re}\tilde{g},\Sigma^{>}]\}$$
$$-\tfrac{1}{4}\{[\tilde{g}^{>},\Gamma]-[a,\Sigma^{>}]\}, \tag{3.12}$$

$$g_0^{-1}h+\frac{i}{2}[g_0^{-1},h]=\operatorname{Re}Sh$$
$$-\frac{i}{2}\gamma h+\frac{i}{2}[\operatorname{Re}S,h]+\tfrac{1}{4}[\gamma,h], \tag{3.13}$$

$$hg_0^{-1}+\frac{i}{2}[h,g_0^{-1}]=h\operatorname{Re}S$$
$$+\frac{i}{2}h\gamma+\frac{i}{2}[h,\operatorname{Re}S]-\tfrac{1}{4}[h,\gamma], \tag{3.14}$$

where all quantities depend on $(\mathbf{p},\omega;\mathbf{R},T)$ and

$$g_0^{-1}(\mathbf{p},\omega;\mathbf{R},T)$$
$$=[\omega-\mathbf{p}\cdot\mathbf{v}_s(\mathbf{R},T)]\tau^{(3)}-\tfrac{1}{2}p^2+\mu(\mathbf{R},T). \tag{3.15}$$

In these equations we have used the generalized Poisson brackets

$$[X,Y]=\frac{\partial X}{\partial\omega}\frac{\partial Y}{\partial T}-\frac{\partial X}{\partial T}\frac{\partial Y}{\partial\omega}$$
$$+\nabla_{\mathbf{R}}X\cdot\nabla_{\mathbf{p}}Y-\nabla_{\mathbf{p}}X\cdot\nabla_{\mathbf{R}}Y \tag{3.16}$$

and have defined

$$\begin{aligned}\operatorname{Re}S&=S_{\mathrm{HF}}(\mathbf{p})+\operatorname{Re}S(\mathbf{p},\omega),\\ \operatorname{Re}\Sigma&=\Sigma_{\mathrm{HF}}(\mathbf{p})+\operatorname{Re}\Sigma(\mathbf{p},\omega),\end{aligned} \tag{3.17}$$

$$\operatorname{Re}\Sigma(\mathbf{p},\omega)=P\int\frac{d\omega'}{2\pi}\frac{\Gamma(\mathbf{p},\omega')}{\omega-\omega'},$$

$$\operatorname{Re}\tilde{g}(\mathbf{p},\omega)=P\int\frac{d\omega'}{2\pi}\frac{a(\mathbf{p},\omega')}{\omega-\omega'},$$

where $P$ denotes the Cauchy principal value.

The Hartree–Fock contributions to the self energies have the form [with the $(\mathbf{R},T)$ dependence suppressed]

$$\Sigma_{\mathrm{HF}}(\mathbf{p},\omega)$$
$$=\frac{i}{2}\int\frac{d^3p'\,d\omega'}{(2\pi)^4}v(\mathbf{p}=0)\operatorname{tr}[(1+\tau^{(3)})g^{<}(\mathbf{p}',\omega')]$$
$$+i\int\frac{d^3p'\,d\omega'}{(2\pi)^4}v(|\mathbf{p}-\mathbf{p}'|)\{\tfrac{1}{2}\operatorname{tr}[(1+\tau^{(3)})g^{<}(\mathbf{p}',\omega')]$$
$$+\tfrac{1}{2}(1+\tau^{(3)})g^{<}(\mathbf{p}',\omega')\tfrac{1}{2}(1-\tau^{(3)})$$
$$+\tfrac{1}{2}(1-\tau^{(3)})g^{<}(\mathbf{p}',\omega')\tfrac{1}{2}(1+\tau^{(3)})\} \tag{3.18}$$

$$S_{\mathrm{HF}}(\mathbf{p},\omega)=\Sigma_{\mathrm{HF}}(\mathbf{p},\omega)$$
$$-i\int\frac{d^3p'\,d\omega'}{(2\pi)^4}v(|\mathbf{p}-\mathbf{p}'|)h(\mathbf{p}',\omega'). \tag{3.19}$$

The cumbersome matrix form of (3.18) and (3.19) is necessary to avoid the constant term in the lower right-hand element of $\tilde{g}^{<}(\Omega,\omega)$ which brings in an unwanted term

$$\int\frac{d^3p'\,d\omega'}{(2\pi)^4}\operatorname{tr}[\tfrac{1}{2}(1-\tau^{(3)})g^{<}(\mathbf{p}'\omega')]$$
$$=\langle\psi(\mathbf{R},T)\psi^{+}(\mathbf{R},T)\rangle$$
$$=\langle\psi^{+}(\mathbf{R},T)\psi(\mathbf{R},T)\rangle+\delta(\mathbf{R}).$$

The remaining terms in our self energy approximation have the form

$$\Sigma^{>}(\mathbf{p},\omega)=-\frac{1}{2}\int\frac{d^3p'\,d\omega'}{(2\pi)^4}\int\frac{d^3p''\,d\omega''}{(2\pi)^4}\int\frac{d^3\bar{p}\,d\bar{\omega}}{(2\pi)^4}$$
$$\times(2\pi)^4\delta(\mathbf{p}-\bar{\mathbf{p}}-\mathbf{p}'+\mathbf{p}'')\delta(\omega-\bar{\omega}-\omega'+\omega'')$$
$$\times[v(|\mathbf{p}''-\bar{\mathbf{p}}|)]^2\{\tilde{g}^{>}(\mathbf{p}',\omega')\operatorname{tr}[\tilde{g}^{<}(\mathbf{p}'',\omega'')\tilde{g}^{>}(\bar{\mathbf{p}},\bar{\omega})$$
$$+\tilde{g}^{<}(\mathbf{p}'',\omega'')h(\bar{\mathbf{p}},\bar{\omega})+h(\mathbf{p}'',\omega'')\tilde{g}^{>}(\bar{\mathbf{p}},\bar{\omega})]$$
$$+h(\mathbf{p}',\omega')\operatorname{tr}[\tilde{g}^{<}(\mathbf{p}'',\omega'')\tilde{g}^{>}(\bar{\mathbf{p}},\bar{\omega})]\}, \tag{3.20}$$

$$S^{>}(\mathbf{p},\omega)=-\frac{1}{2}\int\frac{d^3p'\,d\omega'}{(2\pi)^4}\int\frac{d^3p''\,d\omega''}{(2\pi)^4}\int\frac{d^3\bar{p}\,d\bar{\omega}}{(2\pi)^4}$$
$$\times(2\pi)^4\delta(\mathbf{p}-\bar{\mathbf{p}}-\mathbf{p}'+\mathbf{p}'')\delta(\omega-\bar{\omega}-\omega'+\omega'')$$
$$\times[v(|\mathbf{p}''-\bar{\mathbf{p}}|)]^2\tilde{g}^{>}(\mathbf{p}',\omega')\operatorname{tr}[\tilde{g}^{<}(\mathbf{p}'',\omega'')\tilde{g}^{>}(\bar{\mathbf{p}},\bar{\omega})]. \tag{3.21}$$

## 4. CONSERVATION LAWS AND IDENTITIES

The differential conservation laws for mass or particle number, momentum, and energy can be obtained by various methods and have as many forms. Since the result of each procedure is correct in the same domain as the Boltzmann Eqs. (3.11) to (3.14), the results must be equivalent. We invoke this equivalence in comparing two of these methods, not only to obtain the conservation laws, but to show that several identities which are important for our analysis appear in a natural way.

The conservation laws can be obtained from the exact transport Eqs. (2.3) in the limit of slow space and time variation, or equivalently, from the generalized Boltzmann equations themselves. It is also possible to obtain the conservation expressions in terms of a "conserving" two-particle Green's function $g_2$. A "conserving" $g_2$ is one which leads to a conserving approximation for $g$. This will always be true if $g_2$ satisfies[8]

$$\int d\bar{1}\,g_0^{-1}(1,\bar{1})g(\bar{1},1')=\delta(1,1')$$
$$+\frac{i}{2}\int d\bar{1}\,v(1,\bar{1})g_2(1,\bar{1};1',\bar{1}), \tag{4.1a}$$

$$\int d\bar{1}\, g(1,\bar{1})g_0^{-1}(\bar{1},1') = \delta(1,1') + \frac{i}{2}\int d\bar{1}\, g_2(1,\bar{1};1'\,\bar{1})v(\bar{1},1'),$$

and

$$g_2(1,2;1',2') = g_2(2,1;2',1'). \tag{4.1b}$$

In Eq. (4.1a) the integral over $\bar{1}$ is an abbreviation for a space–time integral plus a trace over the spinor variables of $\bar{1}$. Whenever the time ordering in $g_2$ is ambigous we pick the ordering which gives the density of particles.

Equations (4.1a) are just the exact transport Eqs. (2.3) and condition (4.1b) is satisfied by the two-particle Green's function corresponding to the special self-energy approximation (2.5), (2.7), and (2.8). $g_2$ is shown diagrammatically in Fig. 3. We outline the procedure leading to the momentum conservation law and its associated identities, and merely state the results for the energy and number conservation laws.

Following the approach of Chap. 10, Ref. 1, the momentum conservation law, written in terms of the conserving two-particle Green's function is

$$\frac{\partial}{\partial T}[\rho \mathbf{v}_s + \mathbf{j}_0]_k = -\sum_l \frac{\partial}{\partial R_l}[\rho(\mathbf{v}_s)_k(\mathbf{v}_s)_l + (\mathbf{v}_s)_k(\mathbf{j}_0)_l + (\mathbf{v}_s)_l(\mathbf{j}_0)_k + (\mathbf{T}_0^{(\alpha)})_{lk} + (\mathbf{T}_0^{(\beta)})_{lk}], \tag{4.2}$$

where we have suppressed the $(\mathbf{R}, T)$ dependence. In Eq. (4.2) $\mathbf{j}_0$ and $\mathbf{T}_0 = \mathbf{T}_0^{(\alpha)} + \mathbf{T}_0^{(\beta)}$ are, respectively, the mass current and stress tensor viewed in a coordinate system moving with velocity $\mathbf{v}_s$, and $\rho$ is the total mass density. These are conveniently expressed in terms of the Green's functions by

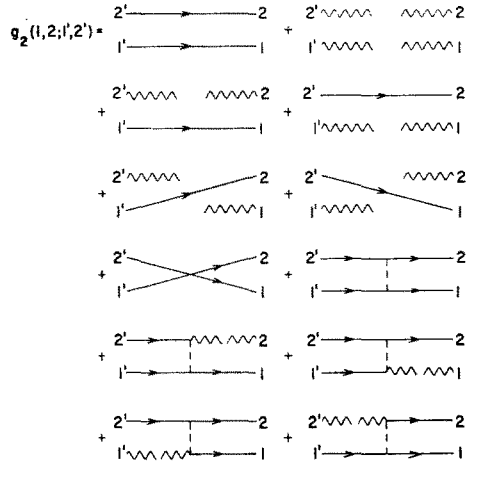


FIG. 3. The diagrams for the $g_2$ which generates our approximations for $\Sigma$ and $S$.

$$\rho = \tfrac{1}{2}\,\mathrm{Tr}\,(1 + \tau^{(3)})g^<(\mathbf{p},\omega) = n_0 + \tfrac{1}{2}\,\mathrm{Tr}\,\tilde{g}^<(\mathbf{p},\omega) + \text{const}, \tag{4.3}$$

$$\mathbf{j}_0 = \tfrac{1}{2}\,\mathrm{Tr}\,\mathbf{p}(1 + \tau^{(3)})g^<(\mathbf{p},\omega) = \tfrac{1}{2}\,\mathrm{Tr}\,\mathbf{p}\tau^{(3)}\tilde{g}^<(\mathbf{p},\omega), \tag{4.4}$$

$$(\mathbf{T}_0^{(\alpha)})_{kl} = \tfrac{1}{2}\,\mathrm{Tr}\,\mathbf{p}_k\mathbf{p}_l(1 + \tau^{(3)})g^<(\mathbf{p},\omega) = \tfrac{1}{2}\,\mathrm{Tr}\,\mathbf{p}_k\mathbf{p}_l\tilde{g}^<(\mathbf{p},\omega) + \text{const}'. \tag{4.5}$$

In these expressions $\mathrm{Tr} = \mathrm{tr}\int d^3p d\omega(2\pi)^{-4}$, and the constants are independent of $\mathbf{R}$ and $T$. Finally the potential term in the stress tensor is given by

$$(\mathbf{T}_0^{(\beta)})_{kl} = \tfrac{1}{8}\,\mathrm{tr}\int d^3 r_2 \frac{(\mathbf{r}_1 - \mathbf{r}_2)_k(\mathbf{r}_1 - \mathbf{r}_2)_l}{|\mathbf{r}_1 - \mathbf{r}_2|} \times \frac{\partial v(|\mathbf{r}_1 - \mathbf{r}_2|)}{\partial|\mathbf{r}_1 - \mathbf{r}_2|} g_2(1,2;1,2)\Big|_{\substack{t_1=t_2\\ 1=\mathbf{R},T}}. \tag{4.6}$$

Alternatively we can obtain the momentum conservation laws from the generalized Boltzmann equations. Take the sum of the second equation in (3.11) and (3.13). Subtract from this the sum of the second equation in (3.12) and (3.14). Multiply the result by $\mathbf{p}$ and apply Tr. This yields a momentum conservation law exactly in the form (4.2) except that we find the alternative expression for $\nabla_{\mathbf{R}}\cdot\mathbf{T}_0^{(\beta)}$

$$\nabla_{\mathbf{R}}\cdot\mathbf{T}_0^{(\beta)} = \tfrac{1}{2}\,\mathrm{Tr}\,\mathbf{p}\{[\mathrm{Re}\,\Sigma, \tilde{g}^<] + [\Sigma^<, \mathrm{Re}\,\tilde{g}] + [\mathrm{Re}\,S, h]\} + \tfrac{1}{2}\,\mathrm{Tr}\,\mathbf{p}\{\Sigma^>\tilde{g}^< - \Sigma^<\tilde{g}^>\}. \tag{4.7}$$

Since the two different paths must lead to the same result, the expression (4.7) must be the identical to the divergence of (4.6). We are thus led to look into the structure of (4.7) to see if this equivalence can be demonstrated.

The left-hand side is of course proportional to a spatial derivative. In our approximation, $\Sigma^>$ and $\Sigma^<$ contain neither space nor time derivatives to first order in $(\nabla_{\mathbf{R}}, \partial/\partial T)$. Hence the final trace on the right-hand side of (4.7) contains no space–time derivatives. By writing

$$[X, Y] = [X, Y]_{\omega,T} + [X, Y]_{\mathbf{p},\mathbf{R}},$$

$$[X, Y]_{\omega,T} = \frac{\partial X}{\partial\omega}\frac{\partial Y}{\partial T} - \frac{\partial X}{\partial T}\frac{\partial Y}{\partial\omega}, \tag{4.8}$$

$$[X, Y]_{\mathbf{p},\mathbf{R}} = -\nabla_{\mathbf{p}}X\cdot\nabla_{\mathbf{R}}Y + \nabla_{\mathbf{R}}X\cdot\nabla_{\mathbf{p}}Y,$$

it is possible to divide the first trace of (4.7) into two parts, one containing only space derivatives and the other only time derivatives. The obvious way that we can satisfy (4.7) would be to have the terms containing no derivatives cancel out,

$$\mathrm{Tr}\,\mathbf{p}[\Sigma^>\tilde{g}^< - \Sigma^<\tilde{g}^>] = 0, \tag{4.9}$$

and those containing only time derivatives also be zero,

$$\mathrm{Tr}\,\mathbf{p}\{[\mathrm{Re}\,\Sigma,\tilde{g}^<]_{\omega,T} + [\Sigma^<,\mathrm{Re}\,\tilde{g}]_{\omega,T} + [\mathrm{Re}\,S,h]_{\omega,T}\} = 0, \quad (4.10)$$

while the remaining terms containing explicit space derivatives add up to a divergence as

$$\nabla_{\mathbf{R}}\cdot\mathbf{T}_0^{(\beta)} = \tfrac{1}{2}\,\mathrm{Tr}\,\mathbf{p}\{[\mathrm{Re}\,\Sigma,g^<]_{\mathbf{p},\mathbf{R}} + [\Sigma^<,\mathrm{Re}\,\tilde{g}]_{\mathbf{p},\mathbf{R}} + [\mathrm{Re}\,S,h]_{\mathbf{p},\mathbf{R}}\}. \quad (4.11)$$

The arguments leading to (4.9)–(4.11) are only speculative at this point, but represent the motivation for writing them down and considering them to be a reflection of the momentum conservation laws in the superfluid system. The next step is to check these results. The approximations (3.18)–(3.21) for the self-energies are inserted into our proposed relations, and after considerable algebra (4.9)–(4.11) are verified. That is, the cancellation indicated in (4.9) and (4.10) does in fact occur, and the right hand side of (4.11) is seen to be a total derivative. This derivative is just what emerges when our approximate $g_2$ is inserted into (4.6).

The identities (4.10) and (4.11) are quite important in our discussion of the hydrodynamics of the two-fluid system. They appear to fail for high temperatures and strong interactions,[22] but we speculate that they may be correct whenever the quasiparticle approximation is valid, i.e., for sufficiently low temperatures or weak interactions.

A rather similar, but more complex argument motivates the appearance of identities connected with the differential energy conservation law. In terms of the two-particle Green's functions we obtain

$$\frac{\partial}{\partial T}\{\tfrac{1}{2}\rho v_s^2 + \mathbf{j}_0\cdot\mathbf{v}_s + K_0 + V\} = -\nabla_{\mathbf{R}}\cdot\{\mathbf{v}_s[\tfrac{1}{2}\rho v_s^2 + \mathbf{j}_0\cdot\mathbf{v}_s + K_0 + V] + \tfrac{1}{2}v_s^2\mathbf{j}_0 + \mathbf{T}_0\cdot\mathbf{v}_s + \mathbf{j}_{E_0}^{(\alpha)} + \mathbf{j}^{(\beta)}_{E_0}\}. \quad (4.12)$$

The left-hand side of (4.12) is the time derivative of the energy density and the right-hand side the divergence of the energy current. $V$ is the potential energy density

$$V = -\tfrac{1}{8}\,\mathrm{Tr}\int d\bar{1}\,v(1,\bar{1})g_2(1,\bar{1};1,\bar{1})\,|_{1=\mathbf{R},T}, \quad (4.13)$$

while $K_0$ is the kinetic energy density in the coordinate frame moving with velocity $\mathbf{v}_s$,

$$K_0 = \tfrac{1}{2}\,\mathrm{Tr}\,\{(1+\tau^{(3)}) \times [(\omega - \mathbf{p}\cdot\mathbf{v}_s) + \mu]g^<(\mathbf{p},\omega)\} - 2V. \quad (4.14)$$

The term $\mathbf{j}_{E_0} = \mathbf{j}_{E_0}^{(\alpha)} + \mathbf{j}_{E0}^{(\beta)}$, where $E_0 = K_0 + V$, is the energy current viewed in the moving frame. The terms are given by

$$\mathbf{j}_{E_0}^{(\alpha)} = \tfrac{1}{2}\,\mathrm{Tr}\,(1+\tau^{(3)})\mathbf{p}(\omega - \mathbf{p}\cdot\mathbf{v}_s + \mu)g^<(\mathbf{p},\omega) \quad (4.15)$$

and

$$\mathbf{j}_{E_0}^{(\beta)} = -\frac{i}{16}\mathrm{Tr}\,\tau^{(3)}\int d\bar{1}\,v(1,\bar{1})\,(\mathbf{r}_1 - \mathbf{r}_{\bar{1}}) \times [\nabla_{\bar{1}}\cdot(\nabla_1 - \nabla_{1'})g_2(1,\bar{1};1'\bar{1}]_{1'=1=\mathbf{R},T}. \quad (4.16)$$

The energy conservation law obtained from the Boltzmann equation does not seem to have the form (4.12). However, after considerable manipulation, it is seen that the form is the same as that of (4.12) if the following identities are obeyed:

$$\mathrm{Tr}\,\omega\{\Sigma^>\tilde{g}^< - \Sigma^<\tilde{g}^>\} = 0, \quad (4.17)$$

$$\frac{\partial V}{\partial T} = \tfrac{1}{2}\,\mathrm{Tr}\,\omega\{[\mathrm{Re}\,\Sigma,\tilde{g}^<]_{\omega,T} + [\Sigma^<,\mathrm{Re}\,\tilde{g}]_{\omega,T} + [\mathrm{Re}\,S,h]_{\omega,T}\}, \quad (4.18)$$

$$\nabla_{\mathbf{R}}\cdot\mathbf{j}_{E_0}^{(\beta)} = -\tfrac{1}{2}\,\mathrm{Tr}\,\omega\{[\mathrm{Re}\,\Sigma,\tilde{g}^<]_{\mathbf{p},\mathbf{R}} + [\Sigma^<,\mathrm{Re}\,\tilde{g}]_{\mathbf{p},\mathbf{R}} + [\mathrm{Re}\,S,h]_{\mathbf{p},\mathbf{R}}\} - \nabla_{\mathbf{R}}\cdot\{\mathbf{T}_0^{(\beta)}\cdot\mathbf{v}_s - V\mathbf{v}_s\}. \quad (4.19)$$

Then making use of our explicit expressions for the self-energies and $g_2$ we discover that (4.17)–(4.19) are indeed true.

Equations (4.18) and (4.19) are two more identities which will prove important in our discussion of the two-fluid equations. They are related to the energy conservation law in precisely the same way as (4.10) and (4.11) are related to the momentum conservation law.

For completeness we note that connected with the number conservation law

$$\frac{\partial\rho}{\partial T} + \nabla_{\mathbf{R}}\cdot(\rho\mathbf{v}_s + \mathbf{j}_0) = 0 \quad (4.20)$$

are the identities

$$\mathrm{Tr}\,\tau^{(3)}\{[\mathrm{Re}\,\Sigma,\tilde{g}^<]_{\omega,T} + [\Sigma^<,\mathrm{Re}\,\tilde{g}]_{\omega,T} + [\mathrm{Re}\,S,h]_{\omega,T}\} = 0, \quad (4.21)$$

$$\mathrm{Tr}\,\tau^{(3)}\{[\mathrm{Re}\,\Sigma,\tilde{g}^<]_{\mathbf{p},\mathbf{R}} + [\Sigma^<,\mathrm{Re}\,\tilde{g}]_{\mathbf{p},\mathbf{R}} + [\mathrm{Re}\,S,h]_{\mathbf{p},\mathbf{R}}\} = 0. \quad (4.22)$$

These identities ensure the gauge invariance of the previous identities, but will not be important in our later work.

[22] We are indebted to Dr. G. Baym for pointing out that these identities may not be exact to all orders in the potential at finite temperature. In fact trouble occurs in the third-order terms. However, according to R. Craig, the identities become increasingly accurate as the temperature approaches zero.

## 5. LOCAL THERMODYNAMIC EQUILIBRIUM SOLUTIONS

The Boltzmann equations of Sec. (3) are valid only if $\tilde{g}$ and $h$ change slowly as $\mathbf{R}$ varies over distances comparable to characteristic particle wave lengths and as $T$ varies over times comparable to the characteristic inverse particle energies. We now investigate the domain in which $\tilde{g}$ and $h$ vary slowly over distances comparable to the mean free path of a typical excitation and times of the order of the mean free time of such an excitation. Mathematically we require that the space and time derivatives in Eqs. (3.11) through (3.14) be small compared to those terms involving no such derivatives, e.g., that

$$\left[g_0^{-1} - \operatorname{Re}\Sigma + \frac{i}{2}\Gamma, \tilde{g}^<\right] \ll \left(g_0^{-1} - \operatorname{Re}\Sigma + \frac{i}{2}\Gamma\right)\tilde{g}^<. \tag{5.1}$$

This hydrodynamic domain ($\omega\tau \ll 1$) usually has a much smaller domain of validity than that of the generalized Boltzmann equations since the excitations are rather sharply defined in energy. However, very near the lamda point, the Boltzmann equations and hydrodynamic description of the system may have the same ranges of validity, i.e.,

$$\nabla_{\mathbf{p}}\cdot\nabla_{\mathbf{R}} g^{\gtrless} \ll g^{\gtrless}, \qquad \frac{\partial}{\partial\omega}\frac{\partial}{\partial T} g^{\gtrless} \ll g^{\gtrless}. \tag{5.2}$$

In the hydrodynamic limit the generalized Poisson bracket terms, containing space and time derivatives, are negligibly small and Eqs. (3.11) and (3.13) reduce to

$$\left[g_0^{-1} - \operatorname{Re} S + \frac{i}{2}\gamma\right]h = 0, \tag{5.3}$$

$$[g_0^{-1} - \operatorname{Re}\Sigma]\tilde{g}^{\gtrless} - \Sigma^{\gtrless}\operatorname{Re}\tilde{g} = -\frac{i}{2}[\Sigma^>\tilde{g}^< - \Sigma^<\tilde{g}^>]. \tag{5.4}$$

Since $h(\mathbf{p},\omega)$ contains delta functions in momentum and energy (see 3.10), Eq. (5.3) becomes

$$[\mu(\mathbf{R}, T) - \operatorname{Re} S(0, 0; \mathbf{R}, T)](1 + \tau^{(1)}) = -\frac{i}{2}\gamma(0, 0; \mathbf{R}, T)(1 + \tau^{(1)}). \tag{5.5}$$

Before considering the consequences of (5.5) further we first turn to Eq. (5.4).

In the usual theory for the normal system, Eqs. (5.4) would have a solution in which the spectral weight function $a(\mathbf{p}, \omega; \mathbf{R}, T)$ would be determined by equations exactly analogous to those of the equilibrium theory, i.e.,

$$a(\mathbf{p}, \omega) = i\left\{\left[g_0^{-1} - \operatorname{Re}\Sigma + \frac{i}{2}\Gamma\right] - \left[g_0^{-1} - \operatorname{Re}\Sigma - \frac{i}{2}\Gamma\right]^{-1}\right\}. \tag{5.6}$$

Then since $a(\mathbf{p}, \omega)$ satisfies a dispersion relation,

$$\operatorname{Re}\tilde{g}(\mathbf{p}, \omega) = P\int\frac{d\omega'}{2\pi}\frac{a(\mathbf{p}, \omega')}{\omega - \omega'} = \frac{1}{2}\left\{\left[g_0^{-1} - \operatorname{Re}\Sigma + \frac{i}{2}\Gamma\right]^{-1} + \left[g_0^{-1} - \operatorname{Re}\Sigma - \frac{i}{2}\Gamma\right]^{-1}\right\}. \tag{5.7}$$

It can be seen that (5.6) and (5.7) explicitly satisfy (5.4) rewritten in the form

$$[g_0^{-1} - \operatorname{Re}\Sigma]a = \Gamma\operatorname{Re}\tilde{g},$$

if we allow, $a$, $g_0^{-1}$, $\operatorname{Re}\Sigma$, $\Gamma$, and $\operatorname{Re}\tilde{g}$ to depend on $\mathbf{R}$ and $T$. Also in the normal system theory the terms on the right-hand side of (5.4) must vanish. This is the mathematical statement of detailed balance. If we consider in the superfluid system that the right-hand side of (5.4) represents the net rate of change of the density of particles with momentum $\mathbf{p}$ and energy $\omega$ at $(\mathbf{R}, T)$ then in local thermodynamic equilibrium

$$\Sigma^>(\mathbf{p}, \omega; \mathbf{R}, T)\tilde{g}^<(\mathbf{p}, \omega; \mathbf{R}, T) - \Sigma^<(\mathbf{p}, \omega; \mathbf{R}, T)\tilde{g}^>(\mathbf{p}, \omega; \mathbf{R}, T) = 0. \tag{5.8}$$

Since we can always write

$$\tilde{g}^>(\mathbf{p}, \omega; \mathbf{R}, T) = f(\mathbf{p}, \omega; \mathbf{R}, T)a(\mathbf{p}, \omega; \mathbf{R}, T),$$
$$\tilde{g}^<(\mathbf{p}, \omega; \mathbf{R}, T) = (1 + f(\mathbf{p}, \omega; \mathbf{R}, T))a(\mathbf{p}, \omega; \mathbf{R}, T), \tag{5.9}$$

we find that (5.8) can only be satisfied if the detailed balancing relations

$$\Sigma^>(\mathbf{p}, \omega; \mathbf{R}, T) = f(\mathbf{p}, \omega; \mathbf{R}, T)\Gamma(\mathbf{p}, \omega; \mathbf{R}, T),$$
$$\Sigma^<(\mathbf{p}, \omega; \mathbf{R}, T) = (1 + f(\mathbf{p}, \omega; \mathbf{R}, T))\Gamma(\mathbf{p}, \omega; \mathbf{R}, T), \tag{5.10}$$

are satisfied. Detailed study of the structure of the approximation for $\Sigma^{\gtrless}$ in Eq. (3.20) indicates that a general expression for $f(\mathbf{p}, \omega; \mathbf{R}, T)$ is

$$f(\mathbf{p}, \omega; \mathbf{R}, T) = \{\exp[\beta(\mathbf{R}, T)(\omega - \mathbf{p}\cdot\mathbf{v}(\mathbf{R}, T))] - 1\}^{-1}. \tag{5.11}$$

By analogy with the equilibrium solution we interpret $\beta(\mathbf{R}, T)$ as the local inverse temperature in energy units, and $\mathbf{v}(\mathbf{R}, T)$ is an arbitrary vector with dimensions of velocity. We shall find in Sec. 6

that $\mathbf{v}(\mathbf{R}, T)$ can be interpreted as the normal fluid velocity.[23] The usual normal system $f(\mathbf{p}, \omega; \mathbf{R}, T)$ contains $\mu(\mathbf{R}, T)$ in the exponential. This term is not present in (5.11) since the number of excitations is not conserved. The macroscopically occupied condensed state acts as a particle sink. Expressions (5.6), (5.7), (5.9)–(5.11) define the solution to the Boltzmann equations for $\tilde{g}^{\gtrless}$ in local thermodynamic equilibrium and allow simplification of (5.5).

Using (5.9) we see from (3.21) that

$$\gamma(0, 0; \mathbf{R}, T) = 0$$

and since Re $S(\mathbf{p}, \omega; \mathbf{R}, T)$ can be shown to have no $\tau^{(2)}$ component

$$\text{Re } S(0, 0; \mathbf{R}, T) = \text{Re } S_0(0, 0; \mathbf{R}, T)\tau^{(0)} + \text{Re } S_1(0, 0; \mathbf{R}, T)\tau^{(1)} \tag{5.12}$$

because Re $S_3(\mathbf{p}, \omega; \mathbf{R}, T)$ is an odd function of $(\mathbf{p}, \omega)$. Then (5.5) reduces to

$$\mu(\mathbf{R}, T) = \text{Re } S_0(0, 0; \mathbf{R}, T) + \text{Re } S_1(0, 0; \mathbf{R}, T). \tag{5.13}$$

We view this equation as determining $\mu(\mathbf{R}, T)$ as a function of the parameters of the system $\beta(\mathbf{R}, T)$, $v_s^2(\mathbf{R}, T)$, $v^2(\mathbf{R}, T)$, and $\rho(\mathbf{R}, T)$.

We now have the conservation laws, identities, and local equilibrium solutions needed to obtain the two-fluid hydrodynamic equations for the superfluid system.

## 6. THE TWO-FLUID THERMODYNAMICS

The two-fluid hydrodynamic equations[15] are the following set of relations:

(A) The energy, momentum, and number conservation laws (4.2), (4.12), and (4.20).

(B) A set of equations obeyed by the superfluid velocity $v_s$. These are a direct consequence of the identifications (3.5) which imply

$$\frac{\partial}{\partial T}\mathbf{v}_s = -\nabla_{\mathbf{R}}(\mu + \tfrac{1}{2}v_s^2), \tag{6.1}$$

$$\text{curl } \mathbf{v}_s = 0, \tag{6.2}$$

in the absence of vortices.

(C) A set of thermodynamic relations obeyed by the densities and currents of conserved quantities.[15] These are, first of all, that $\mathbf{j}_0$, $\mathbf{T}_0$, $\mathbf{j}_{E_0}$, and $E_0$ each depend upon $\mu(\mathbf{R}, T)$, $\beta(\mathbf{R}, T)$, and $\mathbf{v}_n(\mathbf{R}, T) - \mathbf{v}_s(\mathbf{R}, T)$, but not upon $\mathbf{v}_n$ or $\mathbf{v}_s$ individually. In particular, the mass current must be proportional to $(\mathbf{v}_n - \mathbf{v}_s)$ as

$$\mathbf{j}_0 = \rho_n(\mathbf{v}_n - \mathbf{v}_s) \tag{6.3}$$

with $\rho_n$ being a function of $\mu$, $\beta$, and $(\mathbf{v}_n - \mathbf{v}_s)^2$. Also, the stress tensor is required to have the form

$$(\mathbf{T}_0)_{kl} = (\mathbf{T}_0^{(\alpha)} + \mathbf{T}_0^{(\beta)})_{kl} = P\delta_{kl} + \rho_n(\mathbf{v}_n - \mathbf{v}_s)_k(\mathbf{v}_n - \mathbf{v}_s)_l. \tag{6.4}$$

That the $P$ in (6.4) has the physical significance of the pressure is indicated by the thermodynamic relation

$$dP = \rho\, d\mu + \rho s\, d\theta + \mathbf{j}_0 \cdot d(\mathbf{v}_n - \mathbf{v}_s), \tag{6.5}$$

where $\theta$ is the temperature and $s$ is the entropy per unit mass defined by

$$\rho\theta s = E_0 + P - \mu p - \rho_n(\mathbf{v}_n - \mathbf{v}_s)^2. \tag{6.6}$$

Finally the energy current must be

$$\mathbf{j}_{E_0} = [E_0 + P - \mu\rho_s](\mathbf{v}_n - \mathbf{v}_s), \tag{6.7}$$

where $\rho_s = \rho - \rho_n$ is the superfluid density.

To complete the description of the connection between our approximation and the two-fluid hydrodynamics, we must indicate how our equations imply the relations (6.3)–(6.7).

The first question to be considered is the dependence upon $\mathbf{v}_n$ and $\mathbf{v}_s$. Notice that $g^{>}(\mathbf{p}, \omega')$ and $g^{<}(\mathbf{p}, \omega')$, with $\omega' = \omega + \mathbf{p}\cdot\mathbf{v}_s$, are propagators viewed in a system moving with velocity $\mathbf{v}_s$. We should therefore expect that the propagators depend only on $(\mathbf{v}_n - \mathbf{v}_s)$, not on $\mathbf{v}_n$ and $\mathbf{v}_s$ independently. To check this, note that $g_0^{-1}(\mathbf{p}, \omega + \mathbf{p}\cdot\mathbf{v}_s)$ is independent of both $\mathbf{v}$ and $\mathbf{v}_s$. Notice also that the expressions for the self-energies remain form invariant under the change of variables $\omega \to \omega' = \omega + \mathbf{p}\cdot\mathbf{v}_s$. The only function which now contains any dependence on $\mathbf{v}$ or $\mathbf{v}_s$ is $f(\mathbf{p}, \omega; \mathbf{R}, T)$ which becomes

$$f(\mathbf{p}, \omega') = f(\mathbf{p}, \omega + \mathbf{p}\cdot\mathbf{v}_s) = \frac{1}{\exp[\omega' - \mathbf{p}\cdot(\mathbf{v} - \mathbf{v}_s)] - 1}.$$

So $g^{\gtrless}(\mathbf{p}, \omega')$ depends only on $(\mathbf{v} - \mathbf{v}_s)$. We can then identify $\mathbf{v}$ with the normal fluid velocity $\mathbf{v}_n$.

Expressions (4.3)–(4.5), (4.10), (4.11), (4.14), (4.15), (4.18), and (4.19) guarantee that all explicit reference to $\mathbf{v}_s$ in $\rho$, $\mathbf{j}_0$, $\mathbf{T}_0$, $E_0$, and $\mathbf{j}_{E_0}$ disappears when the transformation $\omega \to \omega' = \omega + \mathbf{p}\cdot\mathbf{v}_s$ is made within the Green's functions. $\mathbf{v}_s$ only enters via $f(\mathbf{p}, \omega')$ which depends only on $(\mathbf{v}_n - \mathbf{v}_s)$. Thus we see how the physical quantities lose their dependence on $\mathbf{v}_n$ and $\mathbf{v}_s$ and become functions of the relative velocity. In our evaluation of these physical

[23] J. Bardeen, Rev. Mod. Phys. 34, 667 (1962).

quantities we exploit this dependence by replacing $\mathbf{v}_s$ in $g_0^{-1}$ by zero and letting $\mathbf{v}$ in $f(\mathbf{p}, \omega)$ be $(\mathbf{v}_n - \mathbf{v}_s)$.

Next we show how the thermodynamic results emerge from the identities of Sec. 4. Consider first the momentum density. From (4.4)

$$\frac{\partial \mathbf{j}_0}{\partial T} = \tfrac{1}{2} \operatorname{Tr} \mathbf{p}[\omega\tau^{(3)}, g^<(\mathbf{p}, \omega)]$$

$$= \tfrac{1}{2} \operatorname{Tr} \mathbf{p}\{[g_0^{-1}, \tilde{g}^<]_{\omega,T} + [g_0^{-1}, h]_{\omega,T}\}$$

since $g_0^{-1} = \omega\tau^{(3)} - (p^2/2m) + \mu$. Using the identity (4.10) this can be written as

$$\frac{\partial \mathbf{j}_0}{\partial T} = \tfrac{1}{2} \operatorname{Tr} \mathbf{p}\{[g_0^{-1} - \operatorname{Re} \Sigma, \tilde{g}^<]_{\omega,T}$$

$$- [\Sigma^<, \operatorname{Re} \tilde{g}]_{\omega,T} + [g_0^{-1} - \operatorname{Re} S, h]_{\omega,T}\}.$$

The term involving $h$ vanishes because $h \alpha \delta(\mathbf{p})$. After an integration by parts we obtain

$$\frac{\partial \mathbf{j}_0}{\partial T} = \frac{\partial}{\partial T} \frac{1}{2} \operatorname{Tr} \mathbf{p}\left\{[g_0^{-1} - \operatorname{Re} \Sigma] \frac{\partial \tilde{g}^<}{\partial \omega} + \frac{\partial \Sigma^<}{\partial \omega} \operatorname{Re} \tilde{g}\right\}. \tag{6.8}$$

Since $\partial/\partial T$ refers to changes in $\mathbf{v}_n - \mathbf{v}_s$, $\mu$, and $\beta$, and the trace in (6.8) vanishes at $\mathbf{v}_n - \mathbf{v}_s = 0$, the equation may be integrated with a vanishing constant of integration. Because $\tilde{g}^< = fa$ and $\Sigma^< = f\Gamma$, the momentum density is

$$\mathbf{j}_0 = \frac{1}{2} \operatorname{Tr} \mathbf{p} f\left\{\frac{\partial \operatorname{Re} \tilde{g}^{-1}}{\partial \omega} a - \frac{\partial \Gamma}{\partial \omega} \operatorname{Re} \tilde{g}\right\}$$

$$- \frac{1}{2} \operatorname{Tr} \mathbf{p} \frac{\partial f}{\partial \omega} \Gamma \operatorname{Re} \tilde{g}, \tag{6.9}$$

where we have used the local equilibrium result

$$\operatorname{Re} \tilde{g}^{-1} a = \Gamma \operatorname{Re} \tilde{g}.$$

If we define

$$\tilde{g}^{-1} = g_0^{-1} - \operatorname{Re} \Sigma - \frac{i}{2} \Gamma, \tag{6.10}$$

and notice that $g_0^{-1}$, Re $\Sigma$, and $\Gamma$ are each Hermitian matrices in the $2 \times 2$ $\tau$ space we can rewrite the term inside the curly bracket under the first trace of (6.9) as

$$\operatorname{tr}\left\{\frac{\partial \operatorname{Re} \tilde{g}^{-1}}{\partial \omega} a - \frac{\partial \Gamma}{\partial \omega} \operatorname{Re} \tilde{g}\right\}$$

$$= 2 \operatorname{tr}\{\operatorname{Im} \log[-\tau^{(3)} \tilde{g}^{-1}(\mathbf{p}, \omega_{fl}\}. \tag{6.11}$$

The $-\tau^{(3)}$ has been inserted inside the logarithm in order that this term will vanish at $\omega = -\infty$. This permits an integration by parts in the first term of (6.9) and with the definition

$$X(\mathbf{p}, \omega) = -2 \operatorname{Im} \log[-\tau^{(3)} \tilde{g}^{-1}(\mathbf{p}, \omega)]$$

$$- \Gamma(\mathbf{p}, \omega) \operatorname{Re} \tilde{g}(\mathbf{p}, \omega) \tag{6.12}$$

we can write

$$\mathbf{j}_0 = \frac{1}{2} \operatorname{Tr}\left[\mathbf{p} \frac{\partial f}{\partial \omega} X\right] = \rho_n(\mathbf{v}_n - \mathbf{v}_s). \tag{6.13}$$

[We note in passing that our identification of $\rho_n$ differs from that of Götze[24] who essentially defines $\rho_n = \rho - n_0$. This is incorrect because it completely neglects the depletion effect.[5]]

An almost identical argument applied to $\mathbf{j}_{E_s} = \mathbf{j}_{E_s}^{(\alpha)} + \mathbf{j}_{E_s}^{(\beta)}$ defined by (4.15) and (4.19) indicates that

$$\mathbf{j}_{E_s} = -\tfrac{1}{2} \operatorname{Tr}[\omega(\nabla_p f) X] + \mu \mathbf{j}_0. \tag{6.14}$$

The explicit term containing $\mu$ comes from the $\mu$ in (4.15). Since $\nabla_p f = -(\mathbf{v}_n - \mathbf{v}_s) \partial f/\partial \omega$, Eq. (6.14) implies that

$$\mathbf{j}_{E_s} = \left\{\frac{1}{2} \operatorname{Tr}\left[\omega \frac{\partial f}{\partial \omega} X\right] + \mu \rho_n\right\}(\mathbf{v}_n - \mathbf{v}_s). \tag{6.15}$$

The analysis of the energy density and the stress tensor follow along exactly the same lines. From (4.14) and (4.18),

$$\frac{\partial E_0}{\partial T} = \frac{\partial}{\partial T}\left\{\frac{1}{2} \operatorname{Tr}\left[\omega \frac{\partial f}{\partial \omega} X\right]\right\}$$

$$+ \frac{1}{2} \operatorname{Tr}\left[\frac{\partial f}{\partial \omega} X\right] + \mu \frac{\partial \rho}{\partial T}, \tag{6.16}$$

where $E_0 = K_0 + V$. From (4.5) and (4.11)

$$\frac{\partial}{\partial R_l} (\mathbf{T}_0)_{kl} = -\frac{\partial}{\partial \mathbf{R}_l}\left\{\frac{1}{2} \operatorname{Tr}\left[\mathbf{p}_k \left(\frac{\partial f}{\partial \mathbf{p}_l}\right) X\right]\right\}$$

$$- \delta_{kl} \frac{1}{2} \operatorname{Tr} \frac{\partial f}{\partial \mathbf{R}_l} X + \delta_{kl} \rho \frac{\partial}{\partial \mathbf{R}_l} \mu. \tag{6.17}$$

Since we can interpret $\partial/\partial T$ and $\nabla_{\mathbf{R}}$ as producing changes in the thermodynamic parameters $\mathbf{v}_n - \mathbf{v}_s$, $\beta$, and $\mu$, Eq. (6.16) may be written as a relation between thermodynamic differentials

$$dE_0 = d\left\{\frac{1}{2} \operatorname{Tr}\left[\omega \frac{\partial f}{\partial \omega} X\right]\right\} + \tfrac{1}{2} \operatorname{Tr}[(df) X] + \mu \, d\rho \tag{6.18}$$

and we also see that the stress tensor is of the form (6.4) with

$$dP = -\tfrac{1}{2} \operatorname{Tr}[(df) X] + \rho \, d\mu. \tag{6.19}$$

The differential of $f$ is

$$df = -\frac{\partial f}{\partial \omega}\left\{\mathbf{p} \cdot d(\mathbf{v}_n - \mathbf{v}_s) + \frac{d\theta}{\theta}[\omega - \mathbf{p} \cdot (\mathbf{v}_n - \mathbf{v}_s)]\right\}, \tag{6.20}$$

[24] W. Götze, Z. Physik 117, 300 (1964).

so that (6.19) implies that $dP$ has the form (6.5) with the entropy per unit mass $s$ given by

$$\rho\theta s = \tfrac{1}{2}\,\mathrm{Tr}\left\{\frac{\partial f}{\partial\omega}\,[\omega - \mathbf{p}\cdot(\mathbf{v}_n - \mathbf{v}_s)]X\right\}. \quad (6.21)$$

At this point, only Eqs. (6.6) and (6.7) remain unverified. By adding (6.18) and (6.19), and using the definition of the entropy we find

$$d\{E_0 + P - \mu\rho - \rho\theta s - \rho_n(\mathbf{v}_n - \mathbf{v}_s)^2\} = 0. \quad (6.22)$$

The enthalpy, $E_0 + P$, is then defined by integrating (6.22) and choosing the constant of integration to be zero. We then obtain Eq. (6.6). The last step is to notice that (6.21) and (6.15) imply (6.7). We nally conclude that the identities of Section (4) include all of the thermodynamics of the two fluid model.

## ACKNOWLEDGMENTS

We would like to thank Prof. G. Baym, R. Craig, Dr. P. Hohenberg, Prof. P. C. Martin, and Prof. D. Pines for their extremely useful and constructive critisicm of the work described here.

## APPENDIX: QUASI-PARTICLE EXPRESSION FOR THE ENTROPY

In this appendix we investigate the relationship of Eq. (6.21) with the standard quasi-particle expression for the entropy. To facilitate the discussion, we handle $\mathbf{v}_n - \mathbf{v}_s$ in a slightly different manner than that used in Sec. 6. In Sec. 6 we took $g_0^{-1} = \omega\tau^{(3)} - (p^2/2m) + \mu$ and inserted the $\mathbf{v}_n - \mathbf{v}_s$ dependence via $f$. Here we change the origin of $\omega$ by letting $\omega \to \omega + \mathbf{p}\cdot(\mathbf{v}_n - \mathbf{v}_s)$ so that

$$g_0^{-1} = [\omega + \mathbf{p}\cdot(\mathbf{v}_n - \mathbf{v}_s)]\tau^{(3)} - \frac{p^2}{2m} + \mu, \quad (A.1)$$

and then

$$f(\omega) = 1/(e^{\beta\omega} - 1) \quad (A.2)$$

independent of $\mathbf{v}_n - \mathbf{v}_s$. With this change of variables Eq. (6.21) becomes

$$\rho\theta s = -\mathrm{Tr}\,\omega\frac{\partial f}{\partial\omega}\{\mathrm{Im}\log[-\tau^{(3)}\tilde{g}^{-1}] + \tfrac{1}{2}\Gamma\,\mathrm{Re}\,\tilde{g}\}. \quad (A.3)$$

The quasi-particle approximation is obtained by considering $\Gamma$ to be very small for small values of $\omega$, where $\partial f/\partial\omega$ is large. Then $-\tau^{(3)}\tilde{g}^{-1}$ is essentially real with only an infinitesimal positive imaginary part. To write this in a convenient form, we decompose Re $\Sigma$ into its matrix components. Using the fact that the $\tau^{(2)}$ term may be shown to vanish

$$\Sigma(\mathbf{p},\omega) = [\omega + \mathbf{p}\cdot(\mathbf{v}_n - \mathbf{v}_s)][1 - Z(\mathbf{p},\omega)]\tau^{(3)} + \chi(\mathbf{p},\omega)\tau^{(0)} + \phi(\mathbf{p},\omega)\tau^{(1)}. \quad (A.4)$$

After some manipulation (A.3) reduces to

$$\rho\theta s = -\int\frac{d^3p\,d\omega}{(2\pi)^4}\,\omega\frac{\partial f}{\partial\omega}\,\mathrm{Im}\{\log Z\Lambda_+ + \log Z\Lambda_-\} \quad (A.5)$$

with

$$\Lambda_\pm = -[\omega + \mathbf{p}\cdot(\mathbf{v}_n - \mathbf{v}_s)] \pm\left[\left(\frac{p^2}{2m} - \mu + \chi\right)^2 Z^{-2} - \phi^2 Z^{-2}\right]^{\frac{1}{2}}, \quad (A.6)$$

where in (A.6) the square root is taken to be positive. Since Re $Z(\mathbf{p},\omega)$ is positive for small $\omega$, the imaginary parts of the logarithms in (A.5) are determined by the sign of $\Lambda$. Because of the simple structure of the integrand in (A.5), it is useful to define the quasi-particle energy $E(\mathbf{p})$ (assumed unique) to be the value of the square root in (A.6) which for a given $\mathbf{p}$ makes $\Lambda_+ = 0$; i.e.,

$$E(\mathbf{p}) = \left\{Z^{-2}(\mathbf{p},\omega)\left[\frac{p^2}{2m} - \mu + \chi(\mathbf{p},\omega)\right]^2 - Z^{-2}(\mathbf{p},\omega)\phi^2(\mathbf{p},\omega)\right\}^{\frac{1}{2}}_{\omega = E(\mathbf{p}) - \mathbf{p}\cdot(\mathbf{v}_n - \mathbf{v}_s)}. \quad (A.7)$$

From this definition and from the fact that $\phi$, $Z$, and $\chi$ can be shown to be even under the interchange $(\mathbf{p},\omega) \to (-\mathbf{p},-\omega)$, (A.5) reduces to the simple form

$$\rho\theta s = -\int\frac{d^3p}{(2\pi)^3}\int_{E(\mathbf{p}) - \mathbf{p}\cdot(\mathbf{v}_n - \mathbf{v}_s)}^{\infty} d\omega\,\omega\frac{\partial f}{\partial\omega}, \quad (A.8)$$

so that we obtain the standard result

$$\rho s = -k\int\frac{d^3p}{(2\pi)^3}\{f(\omega)\log f(\omega) - [1 + f(\omega)]\log[1 + f(\omega)]\}_{\omega = E(\mathbf{p}) - \mathbf{p}\cdot(\mathbf{v}_n - \mathbf{v}_s)}, \quad (A.9)$$

where $k$ is the Boltzmann constant.

Precisely the same argument can be applied to the derivation of the other quasi-particle results. For example Eq. (6.13) implies

$$\mathbf{j}_0 = \int\frac{d^3p}{(2\pi)^3}\,\mathbf{p}f[E(\mathbf{p}) - \mathbf{p}\cdot(\mathbf{v}_n - \mathbf{v}_s)]. \quad (A.10)$$

# A New Phase-Space Distribution Function in the Statistical Theory of the Electromagnetic Field*

YUTAKA KANO†
*Department of Physics and Astronomy, University of Rochester, Rochester, New York*
(Received 17 July 1964)

In a previous paper a certain new probability distribution function $q(z)$ relating to *blackbody radiation* was introduced. In the present paper the properties of this function for a *general radiation field* are studied. Unlike the phase-space distribution function of Sudarshan (1963), this function is nonnegative and is an ordinary function. A series expansion for $q(z)$ is given, and it is shown that the series is absolutely convergent for all eigenvalues $z$ of the destruction operator. It is also shown that the density matrix in the Fock representation can be uniquely determined from this probability distribution function, and *vice versa*. The relation between $q(z)$ and the Sudarshan's phase-space distribution function is discussed.

## INTRODUCTION

IT has recently been pointed out by Glauber[1] that a thorough quantum treatment as the basis of the theory of measurement of the electromagnetic field is desirable. Using quantum electrodynamics he has developed the formal properties of electromagnetic field correlation functions by taking an appropriate definition of a coherent state.[2,3] An interesting representation of the density matrix for describing the statistical properties of a radiation field has been introduced by Sudarshan[4,5] and Glauber.[3] Making use of the overcompleteness of the eigenstates $|z$ of the photon destruction operator, Sudarshan has shown that any density matrix can be written in "diagonal" form with the state vector $|z\rangle$ as base (Glauber's $P$ representation). Subsequently interference effects between independent light beams have been discussed using Sudarshan's representation.[6] Also the stationarity and the random phase condition relating to the representation of the density matrix have been investigated.[7,8] However, the validity of the results obtained in these investigations is still open to question, since the convergence of the Sudarshan representation of the phase-space distribution function has so far not been proved. The main purpose of the present paper is to investigate some of the properties of a certain new related probability distribution function $q(z)$, which is free of some of the shortcomings of the Sudarshan representation. This function $q(z)$ is essentially defined as the expectation value of the density matrix with respect to the eigenstate of the photon destruction operator. Unlike the phase-space distribution function of Sudarshan, $q(z)$ is nonnegative. It is shown that the matrix elements of the density matrix in the Fock representation may be uniquely determined from $q(z)$. The relation between $q(z)$ and the phase-space distribution function of Sudarshan is also discussed.

## THE PROBABILITY DISTRIBUTION FUNCTION

Let us consider the electromagnetic field consisting of single momentum-spin state. Following Sudarshan,[4] we begin with an outline of the representation of canonical creation and destruction operators. If $a$ and $a^\dagger$ satisfy the canonical commutation relation,

$$[a, a^\dagger] = 1, \tag{1}$$

every irreducible representation is equivalent to the Fock representation in terms of the states $\psi(n)$, satisfying the relations

$$a^\dagger a\psi(n) = n\psi(n), \qquad (\psi(m), \psi(n)) = \delta_{m,n}. \tag{2}$$

The matrix elements of $a$ and $a^\dagger$ in this representation are given by the equations

$$\begin{aligned}(\psi(m), a\psi(n)) &= n^{\frac{1}{2}}\delta_{m,n-1},\\ (\psi(m), a^\dagger\psi(n)) &= (n+1)^{\frac{1}{2}}\delta_{m,n+1}.\end{aligned} \tag{3}$$

Let us introduce the eigenstates of the destruction and creation operators:

$$a\,|z\rangle = z\,|z\rangle, \qquad \langle z|\,a^\dagger = z^*\,\langle z|.$$

These eigenstates can be expressed as linear combinations of eigenvectors, $\psi(n)$ or $\psi^\dagger(n)$, of the

* This research was supported by the Air Force Office of Scientific Research.
† Present address: Department of General Education, Nagoya University, Nagoya, Japan.
[1] R. J. Glauber, Phys. Rev. Letters **10**, 84 (1963).
[2] R. J. Glauber, Phys. Rev. **130**, 2529 (1963).
[3] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[4] E. C. G. Sudarshan, Phys. Rev. Letters **10**, 277 (1963).
[5] E. C. G. Sudarshan, in *Proceedings of the Symposium on Optical Masers* (Polytechnic Institute Press, Brooklyn, and John Wiley & Sons, Inc., New York, 1963), p. 45.
[6] L. Mandel, Phys. Rev. **134**, A10 (1964).
[7] Y. Kano, Ann. Phys. (N. Y.) **30**, 127 (1964).
[8] L. Mandel, Phys. Letters **10**, 166 (1964).

number operator defined in (2). In fact,[4,5,9,10]

$$|re^{i\theta}\rangle \equiv |z\rangle = \exp(-\tfrac{1}{2}|z|^2)\sum_{n=0}^{\infty}\frac{z^n}{(n!)^{\frac{1}{2}}}\psi(n),$$
$$\langle re^{i\theta}| \equiv \langle z| = \exp(-\tfrac{1}{2}|z|^2)\sum_{n=0}^{\infty}\frac{(z^*)^n}{(n!)^{\frac{1}{2}}}\psi^\dagger(n). \quad (4)$$

These states given by (4) form a normalized overcomplete set but they are not orthogonal,[4,9,10] since

$$\langle z \mid z'\rangle = \exp\{-\tfrac{1}{2}(|z|^2 + |z'|^2 - 2z^*z')\}. \quad (5)$$

Sudarshan[4,5] has shown that the overcompleteness of the states can be used to represent every density matrix,

$$\rho = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\rho(n, m)\psi(n)\psi^\dagger(m) \quad (6)$$

in the "diagonal" form

$$\rho = \int d^2z\,\phi(z)\,|z\rangle\langle z|, \quad (7)$$

where the "phase-space distribution function" $\phi(z)$ is given by

$$\phi(z) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{\rho(n, m)(n!m!)^{\frac{1}{2}}}{(n+m)!(2\pi r)}\exp[r^2 + i(m-n)\theta]\left\{\left(-\frac{\partial}{\partial r}\right)^{n+m}\delta(r)\right\}. \quad (8)$$

This function $\phi(z)$ is not necessarily nonnegative. Equation (8) is the sufficient condition for the density matrix $\rho$ to be expressed in the "diagonal" form (7), but is not a necessary condition. If we regard Eq. (7) as the definition of the phase-space distribution function $\phi(z)$, then we may obtain a $\phi(z)$ which is not given by Eq. (8). For example, in the case of blackbody radiation two different representations of the phase-space distribution function, $\phi(z)$ are known; one is expressed in terms of a symbolic function like Eq. (8) and the other is a Gaussian distribution.[3,11-13]

It has recently been shown[13] that a distribution function $q(z)$ may be defined, which unlike the Sudarshan distribution function $\phi(z)$ defined by (8), is nonnegative and behaves as a true probability distribution function. It may be defined as $1/\pi$ times the expectation value of the density matrix (6) with respect to the state $|z\rangle$,

$$q(z) \equiv \frac{1}{\pi}\langle z|\,\rho\,|z\rangle = \frac{1}{\pi}\sum_{n,m}\rho(n, m)\exp(-|z|^2)\cdot\frac{(z^*)^n z^m}{(n!m!)^{\frac{1}{2}}}. \quad (9)$$

This function, $q(z)$, is particularly interesting, since it plays formally the same role as a classical probability distribution function, when the statistical expectation value of an operator given by an *anti-normal order product* of destruction and creation operators, $a^\alpha(a^\dagger)^\beta$ is evaluated:

$$\begin{aligned}\mathrm{tr}(\rho a^\alpha a^{\dagger\beta} &= \mathrm{Tr}(a^{\dagger\beta}\rho a^\alpha)\\ &= \frac{1}{\pi}\int d^2z\,\langle z|\,\rho\,|z\rangle\, z^\alpha(z^*)^\beta\\ &= \int d^2z q(z) z^\alpha(z^*)^\beta. \end{aligned} \quad (10)$$

Unlike the distribution function $\phi(z)$ given in (8), this function $q(z)$ is an ordinary function, i.e., $q(z)$ is not expressed in terms of the delta function or its derivatives. Moreover, the function $q(z)$ is nonnegative and is normalized to unity in the $z$-space,

$$\int q(z)\,d^2z = 1. \quad (11)$$

We can also show that $q(z)$ is finite for all finite $|z\rangle$. The proof follows from (9):

$$\begin{aligned} q(z) \equiv \frac{1}{\pi}\langle z|\,\rho\,|z\rangle &\le \frac{1}{\pi}\sum_{n,m}|\rho(n, m)|\exp(-|z|^2)\cdot\frac{|z|^n\cdot|z|^m}{(n!m!)^{\frac{1}{2}}}\\ &< \frac{1}{\pi}\exp(-|z|^2)\sum_{n,m}\left(\frac{|z|^{2n}\cdot|z|^{2m}}{n!m!}\right)^{\frac{1}{2}}\\ &\le \frac{1}{\pi}\exp(-|z|^2)\sum_{n}\left(\frac{|z|^{2n}}{n!}\right)^{\frac{1}{2}}\sum_{m}\left(\frac{|z|^{2m}}{m!}\right)^{\frac{1}{2}}. \end{aligned}$$

The series on the right-hand side is convergent for all finite $|z|$.

If we know $q(z)$, then we can uniquely determine the density matrix. Multiplying both sides of Eq. (9) by $z^\lambda$ (where $\lambda$ is a nonnegative integer), and integrating over $\theta$, we obtain the equation

$$\int_0^{2\pi} q(z)z^\lambda\,d\theta = 2\sum_m \rho(m+\lambda, m)\exp(-|z|^2)\times\frac{|z|^{2(m+\lambda)}}{[(m+\lambda)!m!]^{\frac{1}{2}}}. \quad (12)$$

On multiplying both sides of Eq. (12) by $\exp(+|z|^2)$ and setting $|z|^2 \equiv x$, it then follows that

$$F(x) = 2\sum_m \rho(m+\lambda, m)\frac{x^{m+\lambda}}{[(m+\lambda)!m!]^{\frac{1}{2}}}, \quad (13)$$

where the function, $F(x)$, is defined by

$$F(x) \equiv e^x\int_0^{2\pi} q(z)z^\lambda\,d\theta.$$

[9] J. R. Klauder, Ann. Phys. (N. Y.) 11, 123 (1960).
[10] S. Schweber, J. Math. Phys. (N. Y.) 3, 881 (1962).
[11] L. Mandel, Phys. Letters 7, 117 (1964).
[12] C. L. Mehta and E. Wolf, Phys. Rev. 134, A1153 (1964).
[13] Y. Kano, Proc. Phys. Soc. (Japan) 19, 1555 (1964).

If we differentiate Eq. (13) $m + \lambda$ times with respect to $x$ and set $x = 0$ afterwards, then we obtain the equation

$$\left[\frac{d^{m+\lambda}}{dx^{m+\lambda}} F(x)\right]_{x=0} = 2\rho(m + \lambda, m)\left(\frac{(m + \lambda)!}{m!}\right)^{\frac{1}{2}},$$

i.e.,

$$\rho(m + \lambda, m) = \frac{1}{2}\left(\frac{m!}{(m + \lambda)!}\right)^{\frac{1}{2}}\left[\frac{d^{m+\lambda}}{dx^{m+\lambda}} F(x)\right]_{x=0}. \quad (14)$$

Therefore, we see that the matrix elements, $\rho(m, n)$, can be determined from the knowledge of $q(z)$. We can also determine the phase-space distribution function $\phi(z)$, given by (8), by substituting into (8) the value of $\rho(m, n)$ given by (14). This implies that the function $\phi(z)$ [defined by (8)] is uniquely specified by $q(z)$. However, if we take Eq. (7) as the definition of the function $\phi(z)$, then $\phi(z)$ is not always uniquely determined.

Furthermore, the matrix elements, $\rho(m, n)$, calculated from $q(z)$ can be used to determine the function, $R(z, z')$, introduced by Glauber,[3] in his recent investigation on quantum theory of coherence. This function is defined by the formula

$$R(z, z') \equiv \frac{1}{\pi}\langle z|\ \rho\ |z'\rangle$$

$$= \frac{1}{\pi}\exp\{-\tfrac{1}{2}(|z|^2 + |z'|^2)\}\sum_{n,m}\rho(n, m)\cdot\frac{(z^*)^n z'^m}{(n!m!)^{\frac{1}{2}}}. \quad (15)$$

We see that $q(z)$ is a boundary value of $R(z, z')$: $q(z) = R(z, z)$. This function, $R(z, z')$ is also useful for calculating the statistical expectation value of a *normal ordered* operator $\mathcal{O}$:

$$\mathrm{tr}\,(\rho\mathcal{O}) = \frac{1}{\pi}\int d^2 z' \int d^2 z R(z, z')\langle z'|\ \mathcal{O}\ |z\rangle. \quad (16)$$

Let us next consider the relation between the function $q(z)$ given by (9) and the function $\phi(z)$ defined by Eq. (7). From (9) and (7) it follows that

$$q(z) = \frac{1}{\pi}\int d^2\zeta\phi(\zeta) \times \exp\{-|z|^2 - |\zeta|^2 + 2(z^*\zeta + z\zeta^*)\}.$$

Setting $z = x + iy$ and $\zeta = \alpha + i\beta$ ($x$, $y$, $\alpha$ and $\beta$ all real) in the above equation we obtain the formula

$$q(z) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} d\alpha\, d\beta\, \phi(\alpha + i\beta) \times \exp\{-(x - \alpha)^2 - (y - \beta)^2\}.$$

If we change the variables of integration from $\alpha$, $\beta$ to $u$, $v$, through the equation, $x - \alpha = u$ and $y - \beta = v$, we finally obtain the following relation between $\phi(z)$ and $q(z)$:

$$q(z) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} du\, dv \times \phi((x - u) + i(y - v))e^{-(u^2+v^2)}. \quad (17)$$

The above consideration can be generalized to an arbitrary (countable) number of states. The states are now represented by a sequence of complex numbers $\{z\}$, and the Fock representation basis is labeled by a sequence of nonnegative integers $\{n\}$. The matrix elements of the density matrix can be represented by functions of two such sequences $\rho(\{n\}, \{m\})$. The statistical expectation value of any antinormal ordered operator $\mathcal{O}(\{a\}, \{a^\dagger\})$ may then be calculated by means of the formula

$$\mathrm{tr}\,[\rho\mathcal{O}(\{a\}, \{a^\dagger\})] = \int d^2\{z\}\tilde{\mathcal{O}}(\{z\}, \{z^*\})q(\{z\}), \quad (18)$$

where $\tilde{\mathcal{O}}(\{z\}, \{z^*\})$ is a function which can be obtained by substituting the eigenvalues $\{z\}$ and $\{z^*\}$ for $\{a\}$ and $\{a^\dagger\}$, respectively, in the operator $\mathcal{O}(\{a\}, \{a^\dagger\})$. The distribution function $q(\{z\})$ in (18) is given by

$$q(\{z\}) = \sum_{\{n\},\{m\}}\rho(\{n\}, \{m\})\prod_\kappa \times \exp(-|z_\kappa|^2)\cdot\frac{(z_\kappa^*)^{n_\kappa} z_\kappa^{m_\kappa}}{\pi(n_\kappa! m_\kappa!)^{\frac{1}{2}}}, \quad (19)$$

where $\kappa$ denotes a typical momentum-spin state $(\mathbf{k}, s)$.


## ACKNOWLEDGMENT

The author wishes to express his sincere thanks to Professor Emil Wolf and Professor Susumu Okubo for continued encouragement and for useful comments. He is also grateful to Professor E. C. G. Sudarshan and Dr. C. L. Mehta for helpful discussions.

# An Exactly Soluble Model Showing Ferromagnetism

SHIGETOSHI KATSURA AND SAKARI INAWASHIRO
*Department of Applied Physics, Tohoku University, Sendai, Japan*
(Received 1 June 1965)

A Hamiltonian of a one-dimensional Heisenberg model with ferromagnetic interaction is expressed in terms of fermion operators, and renormalized linked cluster expansion has been carried out. It is shown that the approximation leads to ferromagnetic behaviors similar to the molecular field approximation. A model for which the second- and higher-order terms vanish is presented and regarded as an example of quantum-mechanical systems that shows ferromagnetism rigorously. It is also noted that the model is very much like the Husimi–Temperley model or the van der Waals gas with respect to a long-range character of interaction and the vanishing of higher-order terms.

## I. INTRODUCTION

THE existence of a Curie temperature below which a spontaneous magnetization appears and antisymmetric property of the magnetization with respect to external magnetic fields are characteristic features of ferromagnetism. (An ideal case where hysteresis does not appear is considered.) It is usually supposed that these features will be derived for some type of crystal lattices by evaluating the partition function of the Heisenberg model with ferromagnetic interaction, whose Hamiltonian is given by

$$H = -2\sum_{(ij)}[J_{F\perp}(S_i^{(x)}S_j^{(x)} + S_i^{(y)}S_j^{(y)}) + J_{F\parallel}S_i^{(z)}S_j^{(z)}] - \tfrac{1}{2}g\mu_B\mathcal{H}\sum S_i^{(z)}. \quad (1)$$

Here $S_i$, $J_{F\perp}$, $J_{F\parallel}$, $g$, $\mu_B$, and $\mathcal{H}$ denote, spin operator at the $i$th site, perpendicular and parallel components of exchange integral ($J_{F\parallel}$, $J_{F\perp} > 0$), Landé $g$ factor, Bohr magneton, and external magnetic field, respectively. $\sum_{(ij)}$ means summation with respect to nearest neighbors.

Existing theories on the Heisenberg models are classified in three groups, i.e., (1) high-temperature expansion (Opechowski[1]; Kubo, Obata, and Ohno[2]; Rushbrooke and Wood[3]; Brown and Luttinger[4]); (2) low-temperature expansion [(2a) F. Bloch,[5] Dyson,[6] Oguchi,[7] Morita,[8] Wortis,[9] M. Bloch,[10] Oguchi and Honma;[11] 2(b) Frank,[12] Mannari[13]]; and (3) approximations "at large" [(3a) P. Weiss,[14] Oguchi;[15] 3(b) P. R. Weiss,[16] Kasteleijn and Kranendonk,[17] Bogoliubov and Tjablikov,[18] Englert,[19] Kawasaki and Mori,[20] Tahir-Kheli and ter Haar,[21] Stinchcombe *et al.*,[22] Callen,[23] Fujishiro, Takano, and Oguchi[24]].

In the theories in the first group, a real root of a truncated polynomial expressing inverse susceptibility is regarded as an approximate value of the Curie point under the assumption that it exists. The root, however, does or does not exist according to the degree of the truncation, and the method of the truncated inverse susceptibility fails to be applicable in a case where the inverse susceptibility has nonreal singularities inside a circle which intersects the real axis at the true Curie point in the complex plane of inverse temperature. This method, together with the Padé approximant method[25] which is powerful for the analytic continuation outside the radius of convergence, does not give a proof of the existence of the Curie point. The symmetry requirement is satisfied in the theories in the first

[1] W. Opechowski, Physica **4**, 181 (1937); **6**, 1121 (1939).
[2] R. Kubo, Y. Obata, and A. Ohno, Busseiron Kenkyu (in Japanese) No. 43, 22 (1951); No. 47, 35 (1952); No. 57, 45 (1952).
[3] G. S. Rushbrooke and P. J. Wood, Proc. Phys. Soc. (London) **A68**, 1161 (1955).
[4] H. A. Brown and J. M. Luttinger, Phys. Rev. **100**, 685 (1955).
[5] F. Bloch, Z. Physik **61**, 206 (1930).
[6] F. J. Dyson, Phys. Rev. **102**, 1217, 1230 (1956).
[7] T. Oguchi, Phys. Rev. **117**, 117 (1960).
[8] T. Morita, Progr. Theoret. Phys. (Kyoto) **20**, 728 (1958).
[9] M. Wortis, Ph.D. thesis, Harvard University (1963); Phys. Rev. **138**, A1126 (1965).
[10] M. Bloch, Phys. Rev. Letters **9**, 286 (1962).
[11] T. Oguchi and A. Honma, J. Appl. Phys. **34**, 1153 (1963).
[12] D. Frank, Z. Physik **146**, 615 (1956).
[13] I. Mannari, Progr. Theoret. Phys. (Kyoto) **19**, 201 (1958).
[14] P. Weiss, J. Phys. (4) **6**, 661 (1907).
[15] T. Oguchi, Progr. Theoret. Phys. (Kyoto) **13**, 148 (1955).
[16] P. R. Weiss, Phys. Rev. **74**, 1493 (1948).
[17] P. W. Kasteleijn and J. van Kranendonk, Physica **22**, 367 (1956).
[18] N. N. Bogoliubov and S. V. Tjablikov, Dokl. Akad. Nauk SSSR **126**, 53 (1959) [English transl.: Soviet Phys.—Doklady **4**, 589 (1959).
[19] F. Englert, Phys. Rev. Letters **5**, 192 (1960).
[20] K. Kawasaki and H. Mori, Progr. Theoret. Phys. (Kyoto) **38**, 690 (1962).
[21] R. A. Tahir-Kheli and D. ter Haar, Phys. Rev. **127**, 88 (1962).
[22] R. B. Stinchcombe, G. Horwitz, F. Englert, and R. Brout, Phys. Rev. **130**, 155 (1963).
[23] H. B. Callen, Phys. Rev. **130**, 890 (1963).
[24] T. Fujishiro, F. Takano, and T. Oguchi, J. Phys. Soc. Japan **19**, 1666 (1964).
[25] G. Baker, Jr., G. S. Rushbrooke, and H. E. Gilbert, Phys. Rev. **135**, A1272 (1964).

group, since they use low field as well as high-temperature expansion. The expansion coefficients can be obtained exactly as far as we want. All theories in the first group are essentially equivalent in the sense that results of one theory can furnish results of any other theory simply by manipulating the series, even if parameters of expansions are different from each other.

The theories in the second group are low-temperature and high-field expansions. The magnetization in vanishing magnetic field is calculated by putting $\exp(-\frac{1}{2}g\mu_B\mathfrak{H}/kT) = 1$ in these results. The theories in this group are further divided into two subgroups; (2a) the one using Bose statistics, and (2b) the other Fermi Statistics. (Excitations in Heisenberg spin systems are treated neither as bosons nor fermions in a strict meaning.) The former 2a gives a finite spontaneous magnetization at low temperatures for three dimensional lattices (sc, bcc, fcc). For one- and two-dimensional lattices, however, magnetization and susceptibility diverge (magnetization tends to minus infinity). On the other hand, the latter 2b gives a finite spontaneous magnetization at low temperatures for all one-, two-, and three-dimensional lattices. Magnetization is obtained in the form that is not antisymmetric with respect to external magnetic fields, and there are no proofs of the existence of the Curie point in the theories in the second group, both in the use of Bose and Fermi statistics.

The theories in the third group describe well the qualitative features of ferromagnetism above mentioned, a subgroup 3a for all one-, two-, and three-dimensional lattices and the other subgroup 3b for only three-dimensional lattices. It is to be noted that exact information about the model (1) is not obtained from the theories in the third group, though they give a fairly good qualitative description for both low- and high-temperature regions. It is quite doubtful whether such qualitative features are realized in one- and two-dimensional lattices as derived from the theories in subgroup 3a. Now this point is discussed.

The theorem that the one-dimensional classical system has no phase transitions was proved by Takahashi[26] for an arbitrary nearest-neighbor interaction and by Van Hove[27] for an arbitrary finite-ranged interaction. These proofs, however, do not hold for a quantum mechanical system.

The one-dimensional Heisenberg model with the ferromagnetic interaction is usually supposed not to have a spontaneous magnetization since an integral representing it diverges.[5,28] A true proof should be, however, such that the integral representing magnetization vanishes. The divergence is caused by an inadequate use of Bose statistics in treating the spin systems. Such divergence appears also in Dyson's theory[6] if it is applied to a one-dimensional problem. This fact makes it doubtful whether the kinematical interaction is really small or not even in three-dimensional lattices near and above the Curie point. Indeed Tahir-Kheli and Callen[29] and Haas and Jarrett[30] showed that Oguchi and Honma's theory[11] leads to infinitely high Curie temperature. A peculiar behavior of the magnetization near the Curie point in Bloch's theory[10] also seems to be caused by neglecting the kinematical interaction. On the other hand, similar divergences of integrals in Green function theories[18-21,23] do not lead to the divergence of the spontaneous magnetization but to the vanishing of it for one- and two-dimensional systems.

## II. RENORMALIZED LINKED CLUSTER EXPANSION

Now let us analyze an approximate theory which gives a spontaneous magnetization and a Curie temperature and which satisfies the antisymmetry requirement of the magnetization, and seek a model for which these approximate partition function holds as an exact one.

In the former paper by the authors,[31] renormalized linked cluster expansion up to second order was applied to the one-dimensional Heisenberg model with antiferromagnetic interaction and gave satisfactory results. Its first order agrees with the results in the variational treatment by Bulaevskii.[32] When the method described in the former paper is applied to the case of the one-dimensional ferromagnetic interaction, it gives a spontaneous magnetization for the one-dimensional lattice and is regarded as a theory which belongs to the subgroup 3a. This will be discussed in detail in this section. Notations and calculations are the same as in Ref. 31.

Hamiltonian (1) can be expressed in terms of

[26] H. Takahashi, Proc. Phys. Math. Soc. Japan 24, 60 (1942).
[27] L. van Hove, Physica 16, 137 (1950).
[28] R. E. Peierls, *Quantum Theory of Solids* (Oxford University Press, New York, 1955), Chap. VIII.
[29] R. A. Tahir-Kheli and H. B. Callen, J. Appl. Phys. 35, 948 (1964).
[30] C. W. Haas and H. S. Jarrett, Phys. Rev. 135, A1089 (1964).
[31] S. Inawashiro and S. Katsura, Phys. Rev. 140, A892 (1965).
[32] L. N. Bulaevskii, Zh. Eksperim. i Teor. Fiz. 43, 968 (1962) [English Transl.: Soviet Phys.—JETP. 16, 685 (1963).]

Fermion operators $a_l^+$ and $a_l$ for $l$th site of the lattice in the case of one dimension.[33-35]

$$H = -J_{F\perp}\sum_{l=1}^{N}(a_l^+a_{l+1} + a_{l+1}^+a_l) - J_{F\parallel}\sum_{l=1}^{N}(\tfrac{1}{2} - 2a_l^+a_l + 2a_l^+a_la_{l+1}^+a_{l+1}) - m\mathfrak{H}\sum_{l=1}^{N}(1 - 2a_l^+a_l), \quad (2)$$

where $m = \frac{1}{2}g\mu_B$. Hereafter the case $J_{F\perp} = J_{F\parallel} = J_F$ is treated. Transforming it into $k$-space, we have

$$H = H_0 + H_A + H_B, \quad (3)$$

$$H_0 = -\tfrac{1}{2}J_FN + \sum[\epsilon_0(k)a_k^+a_k - m\mathfrak{H}],$$

$$H_A = \sum \epsilon_1(k)a_k^+a_k, \quad (4)$$

$$H_B = -\frac{J_F}{N}\sum\sum\sum\sum V(k_1, k_2, k_3, k_4) \times a_{k1}^+a_{k2}^+a_{k3}a_{k4}, \quad (5)$$

where

$$V(k_1, k_2, k_3, k_4) = \delta_K(k_1 + k_2 - k_3 - k_4) \times [\cos(k_1 - k_4) - \cos(k_1 - k_3)], \quad (6)$$

$$\epsilon_0(k) + \epsilon_1(k) = 2J_F(1 - \cos k + h_F), \quad (7)$$

where $h_F = (m\mathfrak{H}/J_F)$. Linked cluster expansion will be carried out by regarding $H_0$ as an unperturbed part and $H_A + H_B$ as a perturbation. The functions $\epsilon_0(k)$ and $\epsilon_1(k)$ are determined in such a way that the contribution of graphs of the second and higher order which have self-closed lines vanishes. Then

$$\epsilon_0(k) = -2J_F(-h_F - 2s + p\cos k), \quad (8)$$

$$\epsilon_1(k) = -2J_F(-1 + 2s - (p - 1)\cos k). \quad (9)$$

Here $s$ and $p$ are solutions of a set of simultaneous transcendental equations

$$s = \tfrac{1}{2}B_{01}(K_F, a_F, p), \quad (10a)$$

$$p - 1 = B_{11}(K_F, a_F, p), \quad (10b)$$

where

$$a_F = h_F + 2s, \quad K_F = J_F/2kT, \quad h_F = m\mathfrak{H}/J_F,$$

$$B_{mn}(K_F, a_F, p) = \frac{1}{\pi}\int_0^\pi \cos^m k \tanh^n[2K_F(p\cos k - a_F)]\,dk. \quad (11)$$

These and further results are obtained by equating

$$J = -J_F, \qquad K = -K_F,$$

in equations in the former paper.[31] Especially the partition function up to first order

$$\frac{1}{N}\log Z = \frac{1}{\pi}\int_0^\pi \log\{2\cosh[2K_F(p\cos k - 2s - h_F)]\}\,dk - 4K_Fs^2 + K_F(1 - p)^2 \quad (12)$$

gives the magnetization, the energy and the susceptibility as

$$M/Nm = 2s, \quad (13)$$

$$E/NJ_F = \tfrac{1}{2}(p^2 - 1) - 2s^2 - 2sh_F, \quad (14)$$

$$\chi J_F/Nm^2 = 1 - 1/\{1 - 2K_F[1 - B_{02}(K_F, a_F, p)]\}. \quad (15)$$

Now the case $T \to 0$ ($K_F \to \infty$) is considered. In this limit $\tanh x \to \operatorname{sgn} x$, and

$$p - 1 = 0, \quad 1 < a_F, \quad (16)$$

$$p - 1 = -(2/\pi)(1 - a_F^2/p^2)^{\frac{1}{2}}, \quad a_F < 1.$$

$$s = \tfrac{1}{2}, \quad 1 < a_F,$$

$$s = (1/\pi)\arcsin(a_F/p), \quad a_F < 1. \quad (17)$$

Eliminating $a_F$ from (16) and (17), we have

$$s = \tfrac{1}{2}, \quad h_F > 0,$$

$$= -\tfrac{1}{2}, \quad h_F < 0.$$

When the temperature increases, the functional forms of $p$ and $s$ are modified from (16) and (17) so as to have no singularities at $a_F = 1$ (Fig. 1). The nonvanishing value of $2s$ at $h_F = 0$ ($a_F = 2s$), however, continues to exist until the gradient $(\partial 2s/\partial a_F)_{a_F=0}$ reaches 1. Accordingly the value of

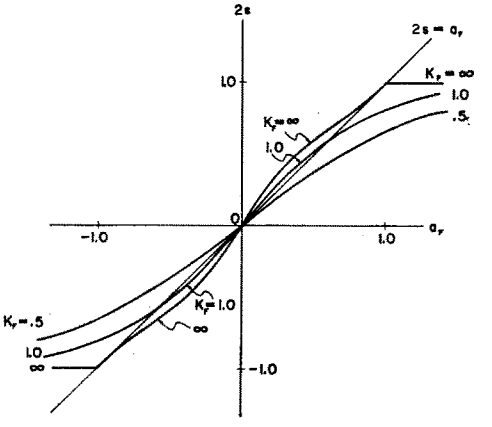


FIG. 1. Solutions of the simultaneous equations (10a) and (10b).

[33] Y. Nambu, Progr. Theoret. Phys. (Kyoto) 5, 1 (1950).
[34] S. Katsura, Phys. Rev. 127, 1508 (1962); 129, 2835 (1963).
[35] S. Katsura and S. Inawashiro, J. Math. Phys. 5, 1091 (1964).

$p(K_F, h_F = 0)$ in (13)–(17) is to be substituted by $p(K_F, a_F = 2s)$ for $T < T_c$ and by $p(K_F, a_F = 0)$ for $T > T_c$ (Fig. 2). When the $s - h_F$ curve of the

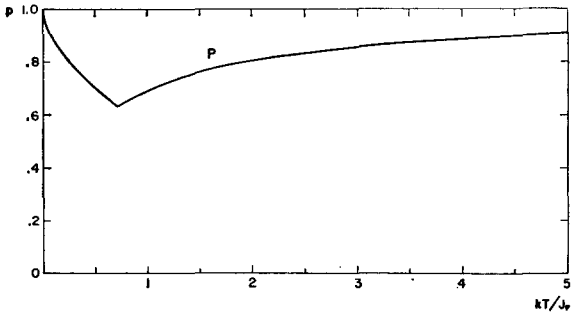


FIG. 2. The value of $p$ in the case of no magnetic field as a function of temperature.

present result is transcribed as a density–fugacity curve of a lattice gas problem, it turns out that the curve is an example in which the condensation point is not given by the smallest real positive singularity of the analytic function defined by the Mayer's fugacity series.[36,37]

The value $K_c$, corresponding to the Curie temperature, is obtained from the condition

$$(\partial 2s/\partial a_F)_{a_F=0} = 1,$$

i.e., it is given by $K_F$ that satisfies

$$\frac{1}{\pi}\int_0^{\pi} \frac{dk}{\cosh^2 [2K_F p(K_F, h_F = 0) \cos k]} = \frac{1}{2K_F}. \quad (18)$$

The value of $kT_c/J_F$ has been obtained to be $0.715 \pm 0.005$.

The spontaneous magnetization is shown in Fig. 3.

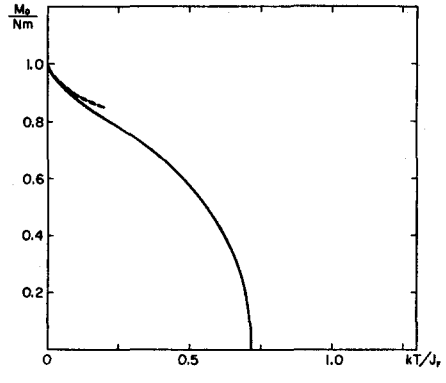


FIG. 3. Spontaneous magnetization $M_0$. — indicates the value using the leading term in (19).

[36] S. Katsura and H. Fujita, J. Chem. Phys. 19, 795 (1951); Progr. Theoret. Phys. 6, 498 (1951).
[37] S. Katsura, Advan. Phys. 12, 391 (1963).

Near $T \sim 0$,

$$\frac{M}{Nm} = 1 + \frac{1}{(2\pi)^{\frac{1}{2}}}(2^{\frac{1}{2}} - 1)\zeta(\tfrac{1}{2})\frac{1}{K_F^{\frac{1}{2}}} + O(K_F^{-1})$$
$$= 1 - \frac{0.6047}{(2\pi)^{\frac{1}{2}}}\frac{1}{K_F^{\frac{1}{2}}} + O(K_F^{-1}). \quad (19)$$

The magnetization $M(H)$ and the entropy $S(H)$ as functions of magnetic field at several temperatures are shown in Figs. 4 and 5. The former is antisymmetric and the latter symmetric with respect to external magnetic fields. Unstable parts of the van der Waals type are also shown in Figs. 4 and 5. The energy and the specific heat are shown in Figs. 6 and 7. The specific heat is finite and has a finite jump at $T = T_c$. The inverse susceptibility is shown in Fig. 8.

Singular natures of magnetization, specific heat, and susceptibility near the Curie point are given by

$$M/Nm \sim c_1[1 - T/T_C]^{\beta}, \quad (20)$$

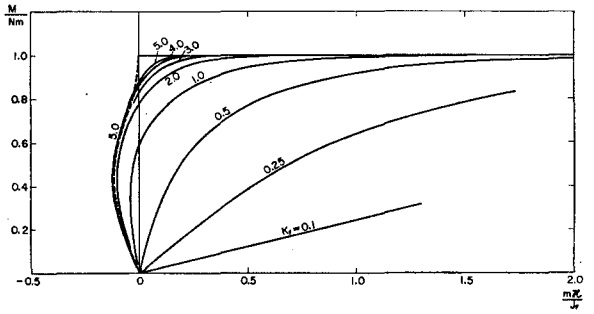


FIG. 4. Magnetization as a function of magnetic field at several temperatures. $M(H)$ is continued antisymmetrically to negative magnetic field part. Unstable van der Waals part is also shown in the figure. — indicates limit of the van der Waals part at $T = 0$.

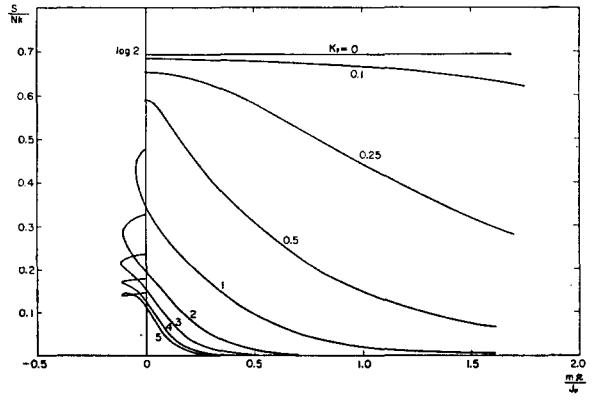


FIG. 5. Entropy as a function of magnetic field at several temperatures. $S(H)$ is continued symmetrically to negative magnetic field part.

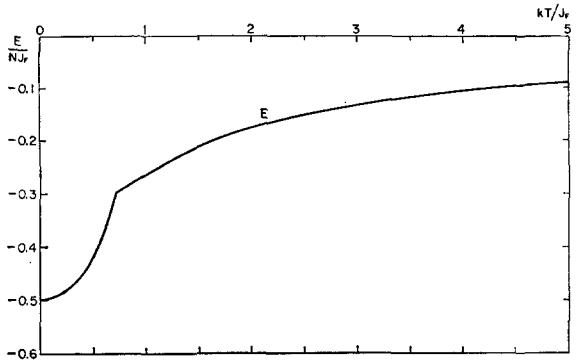


FIG. 6. Energy $E$ in the case of no magnetic field.

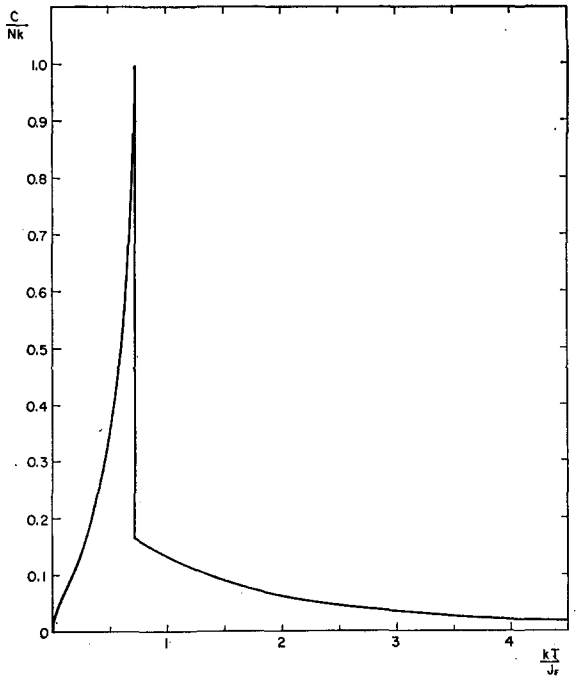


FIG. 7. Specific heat in the case of no magnetic field.

$$C/Nk \sim c_2[1 - T_C/T]^{\alpha} \quad (T \gtrsim T_C), \qquad (21)$$
$$\sim c_3[1 - T/T_C]^{\alpha'} \quad (T \lesssim T_C),$$

$$J_F\chi/Nm^2 \sim c_4[1 - T_C/T]^{-\gamma} \quad (T \gtrsim T_C), \qquad (22)$$
$$\sim c_5[1 - T/T_C]^{-\gamma'} \quad (T \lesssim T_C),$$

where

$$\alpha = \alpha' = 0, \qquad \beta = \tfrac{1}{2}, \qquad \gamma = \gamma' = 1.$$

These values of indices are the same as those of molecular field theory and satisfy the thermodynamic inequality[38,39]

[38] J. W. Essam and M. E. Fisher, J. Chem. Phys. **38**, 802 (1963).
[39] G. S. Rushbrooke, J. Chem. Phys. **39**, 842 (1963).

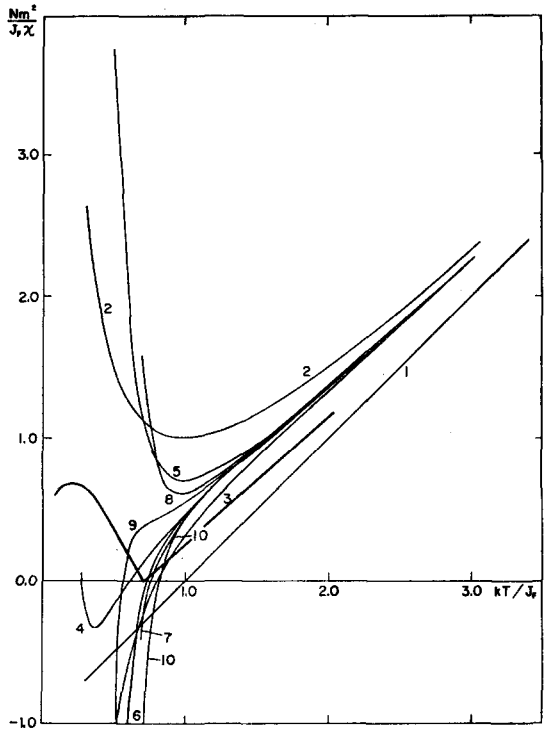


FIG. 8. Inverse susceptibility. Ordinate and abscissa denote $Nm^4/J_F\chi$ and $kT/J_F = 1/2K_F$, respectively. Heavy line denotes the value from (15), where $a_F = 0$ for $T \geq T_c$ and $a_F = 2s$ given by (10) for $T \leq T_c$, respectively. Numbers 1–10 denote values from the partial sum of the truncated polynomials of the $n$th order given in Appendix.

$$\alpha' + 2\beta + \gamma' \geq 2.$$

This transition is not of "the second kind" of Tisza but of the second order of Ehrenfest (the second kind of Landau).[40,41] In the molecular field theory the magnetization at low temperature is

$$M/Nm = 1 - 2\exp(-4K_F) + \cdots$$

in contrast to (19). It is further to be noted that the present treatment does not lead to a peculiar behavior of the magnetization near the Curie point like Bloch's[10] theory.

## III. AN EXACTLY SOLUBLE MODEL AND DISCUSSIONS

What kind of a model will give the above mentioned partition function as an exact partition function? Now we consider a replacement of

[40] L. Tisza, *Proceedings of the Conference on Phase Transformations on Solids*, edited by Smolchowski, Mayer, and Weyl (John Wiley & Sons, Inc., New York, 1951), p. 1.
[41] The derivative of the specific heat at $T = T_c$ is finite, in contrast to the singular nature expected by Landau. [L. Landau and E. Lipshitz, *Statistical Physics* (Clarendon Press, Oxford, England, 1938), Chap. XI.]

$$V(k_1, k_2, k_3, k_4)$$

$$= \delta_K(k_2 - k_3)\delta_K(k_1 - k_4)[1 - \cos(k_1 - k_3)] \quad (23)$$

instead of (6) in the perturbation $H_B$. For the first order perturbation, the replacement gives no effects. For the second- and higher-order perturbation contributions from graphs with self-closed lines vanish by the choice of $s$ and $p$ satisfying (10a) and (10b). In the other contribution from graphs without self-closed lines of the second and higher order the multiplicity of summation is lowered at least 1 and the order of magnitude reduces by a factor $1/N$ or more. In the limit $N \to \infty$ the result of the first-order perturbation becomes an exact solution of the Hamiltonian that is given by (3)–(5) and (23) with (8) and (9).

Transforming (3) with (23) into the configuration space, we have

$$H = -\frac{NJ_F}{2} + J_F \sum_{l=1}^{N} [(a_l^+ a_{l+1} + a_{l+1}^+ a_l - 2a_l^+ a_l) - m\mathcal{H}(1 - 2a_l^+ a_l)] - \frac{J_F}{N}\left[\sum_{l=1}^{N} a_l^+ a_l \sum_{l'=1}^{N} a_{l'}^+ a_{l'} + \sum_{l=1}^{N} a_l^+ a_{l+1} \sum_{l'=1}^{N} a_{l'} a_{l'+1}^+\right]. \quad (24)$$

The last term of (24) expresses a kind of infinitely long-ranged attractive interaction, which gives an interaction energy proportional to $N$, since sums over $l$ and $l'$ are taken independently.

These situations are just like the Husimi, Temperley model[42,43] or the van der Waals gas (the Kac–Uhlenheck–Hemmer model)[44,45] where only the second virial coefficient is modified from an ideal lattice gas or an ideal gas, by an infinitely weak and infinitely long-ranged attractive interaction, whose sum is finite. Though the treatment in this note has been carried out for the one-dimensional model, the qualitative conclusion holds for two and three dimensions, as it does in the Husimi–Temperley model or the van der Waals gas. [The generalization of the Hamiltonian (3)–(5) and (23) into three dimensions yields an exact solution of similar nature.] This fact suggests that a long ranged attractive interaction is favored by a phase transition in a one dimensional problem also in a quantum mechanical system. (Of course it is not a necessary condition for the three-dimensional problem.)

Whether the one-dimensional Heisenberg model has a spontaneous magnetization or not, is not yet quite rigorously solved. Though the nonexistence of spontaneous magnetization is derived from the theories in the subgroup 3b, these approximate theories do not give an exact proof. That the one-dimensional Heisenberg model has no spontaneous magnetization, may rather be inferred from results of exact calculations for finite systems by Griffiths[46] and by Bonner and Fisher[47] or from an extensive numerical analysis using the Padé approximant method by Baker, Rushbrooke, and Gilbert.[25] The reason of the nonexistence might be attributed to essential differences between one- and three-dimensional lattices. As one of the differences it is to be mentioned that bound states exist for all values of $k$ with lower energies than the corresponding spin waves in the one-dimensional model while they do not in the three dimensional model.[48,49] We hope that a theory which takes such difference into consideration explicitly will appear.

## ACKNOWLEDGMENTS

The authors wish to thank Professor T. Haseda for helpful discussions.

This work was supported in part by a research grant from the Ministry of Education of Japan and in part by a research grant from Asahi Press. The numerical calculation was carried out using NEAC 2230 in the Computing Center of Tohoku University and IBM 7090 in UNICON (University Contribution of Japan IBM).

## APPENDIX

A high temperature expansion of the susceptibility of the one-dimensional Heisenberg model was given by Baker, Rushbrooke, and Gilbert[24] up to the 10th power of $K_F$. The inversion of their series gives a high-temperature expansion of the inverse susceptibility:

[42] S. Katsura, Progr. Theoret. Phys. (Kyoto) **13**, 571 (1955); see also Ref. 37, p. 406.
[43] H. N. V. Temperley, Proc. Phys. Soc. (London) **A67**, 233 (1954).
[44] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. **4**, 216 (1963).
[45] See Ref. 37, p. 410.
[46] R. Griffiths, unpublished.
[47] J. C. Bonner and M. E. Fisher, Phys. Rev. **135**, A640 (1964).
[48] M. Wortis, Phys. Rev. **132**, 85 (1963).
[49] S. Katsura, Ann. Phys. (N. Y.) **31**, 325 (1965).

$$\frac{Nm^2}{\chi kT} = 1 - (2K_F) + (2K_F)^2 - \tfrac{2}{3}(2K_F)^3 + \tfrac{1}{8}(2K_F)^4 + \frac{29}{120}(2K_F)^5 - \frac{317}{1440}(2K_F)^6 - \frac{11}{560}(2K_F)^7 + \frac{137}{896}(2K_F)^8 - \frac{1691}{25920}(2K_F)^9 - \frac{185027}{2419200}(2K_F)^{10} + O(K_F{}^{11}). \tag{A.1}$$

Curves of truncated polynomials for $Nm^2/J_F\chi$ up to degree $n$ are also shown in Fig. 8.

# Expansion of a Function of Noncommuting Operators

KAILASH KUMAR
*Research School of Physical Sciences, The Australian National University, Canberra, Australia*
(Received 1 March 1965)

The Taylor series expansion of a function $f(A + B)$ of noncommuting operators $A$ and $B$ is written in several ways. Recursion relations, generating functions, and some explicit formulas for operator coefficients of this expansion are given.

## 1. INTRODUCTION

IN certain problems of physics[1-3] it is required to evaluate the function $f(A + B)$ when the form of $f$ is given, the spectrum and eigenfunctions of $A$ are known, and

$$[A, B] \equiv AB - BA \neq 0. \tag{1}$$

The formulas given in the works of Golden[2] and Kirzhnits[4] have not been expressed in sufficient generality and are sometimes difficult to interpret and apply. The question has received some attention from a purely mathematical point of view also. Finkelstein[5] wrote down a symbolic Taylor series for such a function in terms of a "polarization operator"[6] and gave a formula for the first differential of a function of an operator with respect to the operator. This formula was given a more concrete interpretation by Schwartz[7] who also gave the Taylor series for the case when $A$ and $B$ commute, viz.,

$$f(A + B) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(A)B^n, \quad \text{if} \quad [A, B] = 0. \tag{2}$$

The present problem is, of course, a problem in perturbation theory. The special case in which $f(A) = \exp A$ has been widely studied and is closely related to the general case. In fact the solution of the exponential problem furnishes the generating functions for the coefficient operators $C^n(A, B)$ in the formula

$$f(A + B) = \sum_{n=0}^{\infty} \frac{1}{n!} C^n(A, B) f^{(n)}(A). \tag{3}$$

If $\lambda$ is a small parameter one can further expand

$$C^n(A, \lambda B) = \sum_{r=0}^{n} \lambda^r C_r^n(A, B). \tag{4}$$

In this note we shall derive the formula (3) and give recursion relations, integral representations, and explicit formulas for the operators $C^n$ and $C_r^n$. It then becomes possible to define derivatives of an arbitrary order with respect to operators and to write down explicit formulas for them. This completes the construction of the operator Taylor series.

## 2. DERIVATION OF EQ. (3) AND RECURSION RELATIONS

No attempt will be made at mathematical precision in definitions. It is believed that the manipulations carried out below are valid for operators usually encountered in theoretical physics. The formulas hold at least in the case of finite matrices and I think that they would also be valid under the more general conditions contemplated in Ref. 7.

A function $f(A)$ of an operator $A$ is defined by means of the formula

$$f(A) = (2\pi i)^{-1} \int_c dz\, f(z) R(z, A), \tag{5}$$

where the contour $c$ goes around the singularities of the resolvent $R(z, A) = (z - A)^{-1}$ in the right sense. In the further work we shall make the abbreviations

$$(2\pi i)^{-1} \int_{c_1} \cdots (2\pi i)^{-1} \int_{c_n} \cdots \equiv \int',$$

$$R(z, A + B) \equiv R(z) = (z - A - B)^{-1}, \tag{6}$$

and

$$R(z, A) \equiv R_0(z) = (z - A)^{-1}. \tag{7}$$

We shall assume further that the Taylor expansion of $f(z)$ exists.

From (6) and (7) we have the identities

$$R(z) = R_0(z) + R_0(z)BR(z) \tag{8a}$$

$$= R_0(z) + R(z)BR_0(z) \tag{8b}$$

$$= \sum_{n=0}^{\infty} R_0(z)(BR_0(z))^n \tag{8c}$$

[1] K. J. Le Couteur, Proc. Phys. Soc. (London) **84**, 837 (1964).
[2] S. Golden, Rev. Mod. Phys. **32**, 322 (1960); Phys. Rev. **105**, 604 (1957) and **107**, 1283 (1957).
[3] L. C. R. Alfred, Phys. Rev. **121**, 1275 (1961).
[4] D. I. Kirzhnits, Zh. Eksperim. i teor. Fiz. **32**, 115 (1957) [English transl.: Soviet Phys.—JETP **5**, 64 (1957)].
[5] D. Finkelstein, Commun. Pure Appl. Math. **8**, 245 (1955).
[6] E.g., W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).
[7] J. Schwartz, Commun. Pure and Appl. Math. **8**, 371 (1955). N. Dunford and J. T. Schwartz, *Linear Operators Part I* (Interscience Publishers, Inc., New York, 1955), p. 584.

and

$$1 = \int' dz\, R(z) = \int' dz\, R_0(z), \tag{9}$$

$$R(z) = \int' dz_1\, (z - z_1)^{-1} R(z_1). \tag{10}$$

The last equation is, of course, true for any appropriate function of $z$.

Using (5) and (8a) we have

$$A + B) = \int' dz\, f(z)(R_0(z) + R(z)BR_0(z))$$

$$= f(A) + \int' dz\, dz_1\, dz_2\, f(z) \times (z - z_1)^{-1}(z - z_2)^{-1} R(z_1)BR_0(z_2).$$

On using the identity

$$(z - z_1)^{-1}(z - z_2)^{-1} = ((z - z_1)^{-1} - (z - z_2)^{-1})(z_1 - z_2)^{-1}$$

and integrating over $z$,

$$f(A + B) = f(A) + \int' dz_1\, dz_2\, \frac{f(z_1) - f(z_2)}{(z_1 - z_2)} \times R(z_1)BR_0(z_2). \tag{11}$$

Finally, on Taylor-expanding $f(z_1)$ about $z_2$ we get

$$f(A + B) = f(A) + \int' dz_1\, dz_2 \sum_{n=1}^{\infty} \frac{(z_1 - z_2)^{n-1}}{n!} \times f^{(n)}(z_2) R(z_1)BR_0(z_2). \tag{12}$$

Because of the position of $R(z_1)$ and $R_0(z_2)$ with respect to $B$, on integration with respect to $z_1$ we get a factor $(A + B)$ to the *left* of $B$ for each power of $z_1$; and on integrating with respect to $z_2$ we get a factor $A$ to the *right* of $B$ for each power of $z_2$, in addition the $f^{(n)}(z_2)$ gives a factor $f^{(n)}(A)$ to the right of $B$. That is, in the formula (3),

$$f(A + B) = \sum_{n=0}^{\infty} \frac{1}{n!} C^n f^{(n)}(A), \tag{3}$$

the coefficients $C^n$ obey the recursion relation

$$C^n = (A + B)C^{n-1} - C^{n-1}A \quad \text{or} \quad C^n = [A, C^{n-1}] + BC^{n-1}. \tag{13}$$

Using (4) in (13) we get the corresponding relations for $C_r^n$,

$$C_r^n = [A, C_r^{n-1}] + BC_{r-1}^{n-1}. \tag{14}$$

We have the special values

$$C^0 = 1, \quad C^1 = B, \quad C^2 = [A, B] + B^2, \quad C^3 = [A, [A, B]] + [A, B^2] + B[A, B] + B^3, \tag{15}$$

and

$$C_0^0 = 1, \quad C_0^1 = 0; \quad C_1^1 = B, \quad C_0^2 = 0; \quad C_1^2 = [A, B]; \quad C_2^2 = B^2. \tag{16}$$

In general,

$$C_0^n = \delta_{n0}; \quad C_n^n = B^n, \tag{17}$$

which is consistent with the requirement that when $[A, B] = 0$, Eq. (3) must reduce to the Schwartz formula (2).

### 3. EXPLICIT GENERAL FORMULA FOR $C_r^n$

It is convenient to introduce a more concise notation for representing repeated commutators with respect to $A$:

$$\{B\}^0 = B, \quad \{B\}^1 = [A, B], \quad \{B\}^2 = [A, [A, B]], \tag{18}$$

so that

$$\{B\}^{n+1} = [A, \{B\}^n], \tag{19}$$

$$\{\{B\}^n\}^r = \{B\}^{n+r}, \tag{20}$$

$$\{B + D\}^n = \{B\}^n + \{D\}^n, \tag{21}$$

$$\{BD\}^n = \sum_{r=0}^{n} \binom{n}{r} \{B\}^r \{D\}^{n-r}. \tag{22}$$

It follows that

$$C_r^n = \{C_r^{n-1}\} + BC_{r-1}^{n-1} \tag{23}$$

and

$$C_0^n = \delta_{n0}, \quad C_1^n = \{B\}^{n-1}. \tag{24}$$

From Eq. (23),

$$C_r^n = \sum_{s=0}^{n-1} B^s \{C_{r-s}^{n-1-s}\}. \tag{25}$$

Using Eq. (25) repeatedly, one can reduce the superscript until it becomes equal to the subscript. Then using the relation $C_m^m = B^m$, one gets

$$C_r^n = \sum_{s_1=0}^{r-1} \sum_{s_2=0}^{r-1-s_1} \sum_{s_3=0}^{r-1-s_1-s_2} \cdots \sum_{s_{n-r}=0}^{r-1-\sum_1^{n-r-1} s_i} \times B^{s_1}\{B^{s_2}\{B^{s_3} \cdots \{B^{r-\sum_1^{n-r} s_i}\}\}_{n-r \text{ brackets}} \cdots\}. \tag{26}$$

Alternatively,

$$C_r^n = \sum_{s=0}^{n-r} \{BC_{r-1}^{n-s-1}\}^s, \tag{27}$$

which has to be repeated $r-1$ times when the last term contains $C_1^{n-r-\sum s_i}$ for which Eq. (24) is used,

$$C_r^n = \sum_{s_1=0}^{n-r} \sum_{s_2=0}^{n-r-s_1} \sum_{s_3=0}^{n-r-s_1-s_2} \cdots \sum_{s_{r-1}=0}^{n-r-\sum_1^{r-1} s_i} \times \{B\{B\{B \cdots \{B\}^{n-r-\sum_1^r s_i}\}^{s_{r-1}}\}^{s_{r-2}} \cdots\}^{s_2}\}^{s_1}. \tag{28}$$

Formula (27) is convenient when $(n-r)$ is small and (28) is convenient when $r$ is small. Examples are

$$C_n^n = B^n, \tag{29a}$$

$$C_{n-1}^n = \sum_{s_1=0}^{n-2} B^{s_1}\{B^{n-s_1-1}\}, \tag{29b}$$

$$C_{n-2}^n = \sum_{s_1=0}^{n-3} \sum_{s_2=0}^{n-3-s_1} B^{s_1}\{B^{s_2}\{B^{n-s_1-s_2-2}\}\}, \tag{29c}$$

and

$$C_1^n = \{B\}^{n-1}, \tag{30a}$$

$$C_2^n = \sum_{s_1=0}^{n-2} \{B\{B\}^{n-2-s_1}\}^{s_1}, \tag{30b}$$

$$C_3^n = \sum_{s_1=0}^{n-3} \sum_{s_2=0}^{n-3-s_1} \{B\{B\{B\}^{n-3-s_1-s_2}\}^{s_2}\}^{s_1}. \tag{30c}$$

All these formulas could be further transformed by using Eqs. (21) and (22).

## 4. GENERATING FUNCTION

Consider the function

$$G(\lambda, t) = e^{(A+\lambda B)t}e^{-At}. \tag{31}$$

Its $n$th derivative, $d^nG(\lambda, t)/dt^n \equiv G^n(\lambda, t)$, satisfies the relation

$$G^n(1, t) = (A+B)G^{n-1}(1, t) - G^{n-1}(1, t)A. \tag{32}$$

This is identical with the recursion relation (13) among the $C^n$. Furthermore, $G^0(1, 0) = 1$ and $G^1(1, 0) = B$, so that

$$G^n(1, 0) = C^n \tag{33}$$

and

$$G(\lambda, t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} C^n(A, \lambda B). \tag{34}$$

The series expansion of $G(\lambda, t)$ in powers of $\lambda$ is the well-known perturbation formula

$$G(\lambda, t) = \sum_{r=0}^{\infty} \lambda^r G_r(t),$$

where

$$G_r(t) = \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \times \int_0^{t_{r-1}} dt_r\, B(t_r)B(t_{r-1}) \cdots B(t_1) \tag{35}$$

and

$$B(t) = e^{At}Be^{-At}.$$

It follows from (4), (34), and (35) that

$$G_r(t) = \sum_{n=r} \frac{t^n}{n!} C_r^n(A, B). \tag{36}$$

The exponential function is the generator of the coefficients in the Taylor expansion when $A$ and $B$ are scalar quantities and it is perhaps interesting to see that a generalization exists.

## 5. INTEGRAL REPRESENTATION

The coefficient of $\lambda^r$ in the expansion of $f(A+\lambda B)$ according to (8c) is given by

$$\int' dz\, f(z)R_0(z)(BR_0(z))^r = \int' dz\, dz_0\, dz_1 \cdots dz_r\, f(z) \times \frac{R_0(z_r)BR_0(z_{r-1}) \cdots BR_0(z_1)BR_0(z_0)}{(z-z_r)(z-z_{r-1}) \cdots (z-z_0)}.$$

In order that $f^{(n)}(A)$ should appear on the extreme right, expand $f(z)$ in the integral about $z_0$. Then the coefficient of $f^n(z_0)$ in the above, apart from the operator factors, is $(1/n!)T_r^n$ with

$$T_r^n(z, z_r \cdots z_0) = \frac{(z-z_0)^n}{(z-z_r)(z-z_{r-1}) \cdots (z-z_0)}, \tag{37}$$

which satisfies the recursion relation

$$T_r^n(z, z_r \cdots z_0) = (z_r - z_0)T_r^{n-1}(z, z_r \cdots z_0) + T_r^{n-1}(z, z_{r-1} \cdots z_0); \tag{38}$$

so that

$$C_r^n = \int' dz\, dz_0 \cdots dz_r\, T_r^n(z, z_r \cdots z_0) \times R_0(z_r)BR_0(z_{r-1}) \cdots BR_0(z_0) \tag{39}$$

where the relative order of $z$'s in $T_r^n$ and among the product of operators is important.

## 6. THE RIGHT COEFFICIENTS

It is equally possible to write the operator coefficients to the right of $f^{(n)}(A)$, that is, in place of formula (3) we have

$$f(A+B) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(A)\hat{C}^n. \tag{40}$$

The recursion relations are now different:

$$\hat{C}^n = -\{\hat{C}^{n-1}\} + \hat{C}^{n-1}B \tag{41}$$

and

$$\hat{C}_r^n = -\{\hat{C}_r^{n-1}\} + \hat{C}_{r-1}^{n-1}B. \tag{42}$$

The generating function in this case is

$$\hat{G}(\lambda, t) = e^{-At}e^{(A+\lambda B)t}. \tag{43}$$

Since

$$\hat{G}(\lambda, t) = e^{-At}G(\lambda, t)e^{At}, \tag{44}$$

it follows that

$$\hat{C}_r^n = \sum_{s=r}^{n} (-)^{n-s}\binom{n}{s}\{C_r^s\}^{n-s}. \tag{45}$$

To further clarify the relationship between the two formulas, note that, for any operator $X$,

$$[f(A), X] = \int' dz_1\, dz_2\, (f(z_1) - f(z_2))R_0(z_1)XR_0(z_2)$$

$$= \int' dz_1\, dz_2 \sum_{n=1}^{\infty} \frac{(z_2 - z_1)^{n+1}}{n!} \times f^{(n)}(z_1)R_0(z_1)XR_0(z_2);$$

hence,

$$Xf^{(r)} = f^{(r)}X - [f^{(r)}, X]$$

$$= \sum_{n=0}^{\infty} \frac{(-)^n}{n!} f^{(r+n)}\{X\}^n. \tag{46}$$

Then, with the help of (46) and (45), Eq. (3) can be converted directly to Eq. (40).

## 7. DIFFERENTIATION WITH RESPECT TO A MATRIX

In the foregoing we did not have to use the concept of differentiation with respect to an operator, which was one of the main interests of Refs. 5 and 7. It was shown in those works that it is possible to interpret a quantity $\partial f(A)/\partial A$ as a transformation applicable to the matrices $A$ and $B$. In other words, differentiation of an operator function with respect to its operator argument gives rise to an entity which is properly interpreted as an operator acting in the linear vector space spanned by the first type of operators.

However, it is not necessary to invoke the language of operator algebras and the new formulas can be manipulated entirely in terms of products of operators.

It is best to think of finite matrices of rank $N$ and write the indices explicitly. The matrix function $f_{ij}(A + B)$ may be considered as a function of $N^2$ quantities $(A_{kl} + B_{kl})$; and is to be differentiated with respect to $A_{kl}$. Applying the Taylor theorem for many variables we have

$$f_{ij}(A + B) = f_{ij}(A) + \sum_{a_1b_1} B_{a_1b_1}\frac{\partial f_{ij}}{\partial A_{a_1b_1}}$$

$$+ \frac{1}{2!}\sum_{a_1b_1, a_2b_2} B_{a_1b_1}B_{a_2b_2}\frac{\partial^2 f_{ij}}{\partial A_{a_1b_1}\partial A_{a_2b_2}} + \cdots. \tag{47}$$

It follows that

$$\frac{1}{r!}\sum_{\substack{a_1\cdots a_r \\ b_1\cdots b_r}} B_{a_1b_1}B_{a_2b_2}\cdots B_{a_rb_r}\frac{\partial^r f_{ij}}{\partial A_{a_1b_1}\,\partial A_{a_2b_2}\cdots\partial A_{a_rb_r}}$$

$$= \sum_{n=r}^{\infty}\frac{1}{n!}(C_r^n f^{(n)}(A))_{ij}. \tag{48}$$

The quantity multiplying the $B$'s is therefore a direct product of $(r + 1)$ matrices of type $A_{ij}$ and is independent of the matrix $B$. To find an explicit form for this operator differential coefficient we must represent $C_r^n$ as a product.

For this purpose we use the four-index quantities introduced by Zwanzig.[8] These quantities, represented by script symbols, act on ordinary two-index matrices as follows:

$$(\mathcal{A}B)_{ij} \equiv \sum_{kl}\mathcal{A}_{ij;kl}B_{kl}, \tag{49}$$

and among themselves according to the formulas

$$(\mathcal{A}_1 + \mathcal{A}_2)_{ij;kl} = (\mathcal{A}_1)_{ij;kl} + (\mathcal{A}_2)_{ij;kl}, \tag{50a}$$

$$(\mathcal{A}_1\mathcal{A}_2)_{ij;kl} = \sum_{mn}(\mathcal{A}_1)_{ij;mn}(\mathcal{A}_2)_{mn;kl}. \tag{50b}$$

The left multiplication by a matrix then corresponds to a tetradic

$$(AB)_{ij} \equiv \sum_{kl}{}^{(l)}\mathcal{A}_{ij;kl}B_{kl},$$

$$^{(l)}\mathcal{A}_{ij;kl} = A_{ik}\delta_{lj}, \tag{51a}$$

and so does the right-multiplication

$$^{(r)}\mathcal{A}_{ij;kl} = \delta_{ik}A_{lj}. \tag{51b}$$

The commutator with respect to $A$ is then represented by

$$\mathcal{A}_{ij;kl} = A_{ik}\delta_{lj} - \delta_{ik}A_{lj}. \tag{52a}$$

It follows that

$$\sum_{kl}(\mathcal{A}^n)_{ij;kl}B_{kl} = \{B\}^n. \tag{52b}$$

With these notations and abbreviating the summation restrictions in (28) we have

$$C_r^n = \sum_{\{s\}}\mathcal{A}^{s_1}B\mathcal{A}^{s_2}B\mathcal{A}^{s_3}\cdots\mathcal{A}^{n-r-\sum_1^{r-1}s_i}B \tag{53a}$$

[8] R. Zwanzig, Physica 30, 1109 (1964).

or

$$(C_r^n)_{ik} = \sum_{\{a\}\{b\}\{c\}} \sum_{\{s\}} (\mathfrak{A}^{s_1})_{ik;a_1b_1}(\mathfrak{A}^{s_2})_{b_1c_1;a_2b_2}$$

$$\times (\mathfrak{A}^{s_3})_{b_2c_2,b_3c_3} \cdots (\mathfrak{A}^{n-r-\sum_1^{r-1} s_i})_{b_rc_{r-1};a_rb_r}$$

$$\times B_{a_1b_1}B_{a_2b_2} \cdots B_{a_rb_r}. \tag{53b}$$

Comparing with (48) we have

$$\frac{\partial^r f_{ij}(A)}{\partial A_{a_1b_1}\,\partial A_{a_2b_2} \cdots \partial A_{a_rb_r}} = \sum_{n=r}^{\infty} \frac{r!}{n!} \sum_{\{a\}\{b\}\{c\}} \sum_{\{s\}}$$

$$\times {}^{(r)}f_{ij;a_0b_0}(\mathfrak{A}^{s_1})_{a_0b_0;a_1b_1}(\mathfrak{A}^{s_2})_{b_1c_1;a_2b_2}$$

$$\times (\mathfrak{A}^{s_3})_{b_2c_2;a_3b_3} \cdots (\mathfrak{A}^{n-r-\sum_1^{r-1} s_i})_{b_{r-1}c_{r-1};a_rb_r}, \tag{54}$$

where $i$, $j$, $a$, $b$, $c$ are the matrix indices and the restrictions on $s_i$ were explained in Eq. (28).

In most applications $A$ is taken as diagonal, so that

$$(\mathfrak{A}^n)_{ij;kl} = (A_i - A_j)^n \delta_{ik}\delta_{lj} \tag{55}$$

and (54) is correspondingly simplified.

To write the Taylor expansion even more concisely, note that (53b) can be looked upon as an $r$th-order scalar product

$$(C_r^n)_{ik} = (\mathbf{B}^r \overset{r}{\odot} \mathbf{C}_r^n)_{ik}$$

$$= \sum (\mathbf{B}^r)_{a_1 \cdots a_r;b_1 \cdots b_r} (\mathbf{C}_r^n)_{ik}^{a_1 \cdots a_r, b_1 \cdots b_r}, \tag{56}$$

where the definitions of $\mathbf{B}^r$ and $\mathbf{C}_r^n$ are easily extracted from (53b). The quantity $\mathbf{C}_r^n$ is independent of both the function $f$ and the matrix $B$. We can now define a row matrix $\partial_A^r$, which depends entirely on the matrix $A$,

$$\partial_A^r = r!\,(0, 0, \underbrace{0, \cdots, 0}_{r-zeros}, \mathbf{C}_r^r, \mathbf{C}_r^{r+1}, \cdots \mathbf{C}_r^n, \cdots), \tag{57}$$

and a column matrix $f(A)$ with components given by

$$\mathbf{f}(A) \to f(A), f^{(1)}(A), \frac{1}{2!}f^{(2)}(A), \cdots \frac{1}{n!}f^{(n)}(A) \cdots. \tag{58}$$

With the help of these we can write

$$f(A + \lambda B) = \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} \mathbf{B}^r \overset{r}{\odot} (\partial_A^r \mathbf{f}(A)), \tag{59}$$

which appears very similar to the ordinary Taylor expansion.

This formula also provides a new way of looking at a matrix function. We note that it need be of no consequence that the matrices $f^{(n)}(A)$ are in some way associated with a function $f(z)$. In fact, if we are given any set of matrices $g^{(n)}$; $n = 0, 1, 2 \cdots$ we can arrange them in a column like (58), putting in some zeros if necessary; then given any two matrices $A$ and $B$ the sum on the right-hand side of (59) can be formed, and this may be taken as the definition of a function $g(A + B)$ of the matrices $A$ and $B$. The corresponding quantities $g^{(n)}(A + B)$ may be obtained by starting the column with $g^{(n)}$ rather than $g^{(0)}$.

# On Expanding the Exponential

KAILASH KUMAR
*Research School of Physical Sciences, The Australian National University, Canberra, Australia*
(Received 1 March 1965)

Systematic methods for generating various expansions of the function $\exp(a + \lambda b)t$, of noncommuting operators $a$ and $b$, are presented. The usual perturbation expansion in powers of $\lambda$, and several other formulas found in the earlier literature occur as special cases. Brief remarks about relative merits and physical applications are made.

## 1. INTRODUCTION

THE exponential function of a matrix or a differential operator $A$, defined as

$$\exp A = \sum_{n=0} A^n/n!, \tag{1}$$

occurs in the formal solution for $t = 1$, of the equation

$$\partial U/\partial t = AU; \qquad U(t) = \exp(At)U(0), \tag{2}$$

where $A$ is independent of $t$.

With a suitable choice of the quantity $A$, the above equation can represent the time-dependent Schrödinger equation or the classical Liouville equation. The equation for the development of quantum mechanical density matrix and that for the time development of a Heisenberg operator can also be put in form (2) if $A$ is a super operator[1]. Zwanzig[2] has shown that the super operator can also be put in matrix form. The canonical and grand canonical partition functions have the exponential forms and formally satisfy Eq. (2). Apart from these, Eq. (2) also occurs as an expression for a system of linear differential equations.[3] The usual problem of theoretical physics is to find a suitable computational method for expressing $U(t)$ when

$$A = a + \lambda b, \tag{3}$$

where the eigenvectors of $a$ are known, $\lambda$ is small and $a$ does not commute with $b$,

$$[a, b] = ab - ba \neq 0. \tag{4}$$

Most of the work in field theory, the many-body problem and statistical mechanics, especially since 1949, (see recent textbooks or collections of reprints), has been based on the perturbation formula

$$e^{At} = e^{(a+\lambda b)t} = \sum_{n=0} \lambda^n u_n;$$

$$u_n = e^{at} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \times b(-t_1)b(-t_2) \cdots b(-t_n), \tag{5}$$

$$b(t) = e^{at} b e^{-at}.$$

The chief source of variety in the work based on this formula is the rearrangement of terms and partial summations in the power series which make use of the special properties of interactions. While this formula has received the most attention, other formulas are nonetheless known and have been used from time to time[4-6].

To mathematicians the problem of compounding exponentials of noncommuting quantities has been of interest from its origin in group theory, where it led to the following two "Baker–Hausdorff" theorems[7]:

(i) if $e^x e^y = e^z$, $z$ is a sum of repeated commutators of $x$ and $y$;

(ii) if $e^x w e^{-x} = v$, $v$ is a sum of repeated commutators of $w$ and $x$.

Related to these is a formula by Zassenhaus[7]:

$$\text{(iii)} \qquad e^{(x+y)} = e^x e^y e^{c_2} e^{c_3} e^{c_4} \cdots, \tag{6}$$

where $c_n$ is a sum of commutator products of $n$th degree formed from $x$ and $y$.

Much simplification results if the parameter $t$ is included in these formulas. For example, (ii) is well known in the form

$$e^{xt} w e^{-xt} = \sum_{n=0} \frac{t^n}{n!} [\underbrace{\cdots}_{n\text{-brackets}} [[w, x], x], \cdots x]. \tag{7}$$

The explicit form for $z$ is obtained from (i) by writing

$$x = Xt, \quad y = Yt, \quad \text{and} \quad z = \sum_{n=0} t^n Z_n$$

[1] E.g. Appendix in H. Primas, Helv. Phys. Acta **34**, 331 (1961).
[2] R. Zwanzig, Physica **30**, 1109 (1964).
[3] For example, see L. A. Pipes, *Applied Mathematics for Engineers and Physicists* (McGraw-Hill Book Company, Inc., New York, 1958).
[4] R. Peierls, Z. Physik **80**, 763 (1933).
[5] A. W. Sáenz and R. C. O'Rourke, Rev. Mod. Phys. **27**, 381 (1955).
[6] H. S. Green, Proc. Roy. Soc. (London) **A197**, 73 (1949).
[7] W. Magnus, Comm. Pure Appl. Math. **7**, 649 (1954).

and then equating the coefficients of powers of $t$ on both sides of the equation. This process is complicated enough to require machine computation in higher orders.[8]

We shall call (6) the right-running form and write it as

$$e^{(X+Y)t} = e^{Xt}e^{Yt}e^{\hat{C}_2t^2}e^{\hat{C}_3t^3}e^{\hat{C}_4t^4}\cdots. \tag{8}$$

The quantities $\hat{C}_n$ can be obtained by equating powers of $t$ in expansion of both sides.

By putting $X = a$ and $Y = \lambda b$ it follows by suitable applications of (i) and (iii) that there must be a formula of the form[9]

$$e^{At} = e^{(a+\lambda b)t} = e^{at}e^{\lambda\delta_1(t)}e^{\lambda^2\delta_2(t)}\cdots e^{\lambda^n\delta_n(t)}\cdots. \tag{9}$$

The first two terms of formula (8) coincide with the first two terms of a formula used by Sáenz and O'Rourke[5] which goes back to Peierls,[4] whereas formula (9) has some affinity with that of Green.[6]

In this note a large number of such formulas is generated in a systematic way. In particular, two sequences of formulas will be exhibited; the first member of both is formula (5) and the last members are formulas (8) and (9). Some remarks will be made on their relative merits although it is recognized that one can not go far in this direction without a full specification of the physical system. At the same time the treatment is not oriented to mathematics either, although the mathematical ramifications of these topics are also very wide. The point of view is rather that of theoretical physics of many-body problems where one seeks a scheme for generating approximations.

## 2. LEFT-RUNNING FORMULAS

The formulas (5), (6), (8), and (9) will be called right-running formulas. The property being referred to becomes clear on comparison with their left-running forms:

$$e^{At} = \sum_{n=0} \lambda^n u_n(t)$$

$$u_n(t) = \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \times b(t_n)b(t_{n-1})\cdots b(t_1)e^{at}, \tag{10}$$

$$e^{At} = \cdots e^{t^nC_n}\cdots e^{t^2C_2}e^{tb}e^{ta}, \tag{11}$$

$$e^{At} = \cdots e^{\lambda^n\sigma_n(t)}\cdots e^{\lambda^2\sigma_2(t)}e^{\lambda\sigma_1(t)}e^{ta}. \tag{12}$$

Since

$$e^{At}e^{-At} = 1,$$

it follows from (8) and (11) that

$$\hat{C}_n = (-)^{n+1}C_n, \tag{13}$$

and from (9) and (12) that

$$\delta_n(t) + \sigma_n(-t) = 0. \tag{14}$$

Formulas (13) and (14) are useful in checking calculations. It is slightly more convenient to deal with the left-running forms because usually the operators act to the right on state functions.

## 3. EXPLICIT FORMS FOR $C_n$ AND $\delta_n$

It has not been possible to obtain closed expressions for these quantities. The first few terms $C_n$ can be evaluated as indicated earlier by comparing of $t$ or by other processes indicated in references quoted by Magnus,[7]

$$C_2 = \tfrac{1}{2}[a, b], \tag{15a}$$

$$C_3 = \tfrac{1}{6}[a, [a, b]] + \tfrac{1}{3}[b, [a, b]]. \tag{15b}$$

The quantities $\sigma_n$ are obtained by equating the powers of $\lambda$ on both sides of (12) using (10),

$$\sigma_1 = u_1e^{-at} = \int_0^t dt_1\, b(t_1), \tag{16a}$$

$$\sigma_2 = u_2e^{-at} - \frac{\sigma_1^2}{2!} = \frac{1}{2}\int_0^t dt_1 \int_0^{t_1} dt_2\, [b(t_2), b(t_1)], \tag{16b}$$

$$\sigma_3(t) = u_3e^{-at} - \sigma_2\sigma_1 - \frac{\sigma_1^3}{3!} = \frac{1}{3}\int_0^t dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_3\, \{[b(t_3), [b(t_2), b(t_1)]] - [b(t_2), [b(t_1), b(t_3)]]\}. \tag{16c}$$

In the appendix a relationship between the time derivatives of $\sigma$'s is derived and it is shown that the term $\sigma_n(t)$ is an integral of a sum of $n$-fold commutators of operators $b(t)$ taken at $n$-points $t_1 \cdots t_n$.

It follows from theorems (i) and (iii) that $\sigma_n$ are sums of commutators of $a$ and $b$ with $b$ occurring $n$-times and that

$$\sigma_n(t) \sim t^{n+1} \quad \text{as} \quad t \to 0; \quad n > 1, \tag{17a}$$

$$\sigma_1(t) = bt \quad \text{as} \quad t \to 0, \tag{17b}$$

so that

$$\sigma_n(t = 0) = 0 \tag{18a}$$

and

$$\sigma_n = 0 \quad \text{if} \quad [a, b] = 0. \tag{18b}$$

[8] K. Goldberg, Duke J. Math. 23, 13 (1956).
[9] F. Fer, Bull. Classe Sci. Acad. Roy. Belg. 44, 818 (1958).

That $\sigma_n$ should turn out to be integrals of the above type is not obvious. The physical interpretation of this form has some appealing features which will be mentioned later.

Calculation of higher-order terms both $C_n$ and $\sigma_n$, rapidly becomes more complicated—possibly more so than the computation of $Z_n$ in Baker–Hausdorff formulas.

## 4. A SEQUENCE OF FORMULAS LEADING TO EQUATION (8)

Set

$$U(t) = e^{(a+\lambda b)t} = U_1(t)e^{at}; \tag{19}$$

then

$$\frac{\partial}{\partial t} U_1 = \lambda U_1 b(t); \qquad U_1(0) = 1, \tag{20a}$$

$$U_1(t) = 1 + \lambda \int_0^t dt_1\, U_1(t_1)b(t_1). \tag{20b}$$

If $U_1$ is obtained by solving (20b) by iteration and substituted in (19) then we have the formula (10) for $U(t)$. As $t \to 0$, the kernel of (20b) tends to $\lambda b$ and we have

$$U_1(t) \approx 1 + \lambda bt, \qquad t \to 0,$$

which suggests the next step, viz.,

$$U_1(t) = U_2(t)e^{\lambda bt}. \tag{21}$$

Using (20a) this gives

$$(\partial/\partial t)U_2 = U_2T_2(t); \qquad U_2(0) = 1 \tag{22a}$$

or

$$U_2(t) = 1 + \int_0^t dt_1\, U_2(t_1)T_2(t_1), \tag{22b}$$

$$T_2(t) = \lambda e^{\lambda bt}(b(t) - b)e^{-\lambda bt}. \tag{22c}$$

The next formula of the sequence is, therefore,

$$U(t) = U_2(t)e^{\lambda bt}e^{at}, \tag{23}$$

with $U_2$ obtained by solving (22b) by iteration. This is the formula discussed by Sáenz and O'Rourke.[5] Putting $\lambda = 1$, for brevity we have $T_2(t) \sim t[a, b]$ as $t \to 0$ so that

$$U_2(t) \approx 1 + \tfrac{1}{2}t^2[a, b] \equiv 1 + t^2C_2; \qquad t \to 0, \tag{24}$$

which suggests the next step, viz.,

$$U_2(t) = U_3(t)e^{t^2C_2}. \tag{25}$$

From (22a)

$$\partial U_3/\partial t = U_3T_3(t); \qquad U_3(0) = 1, \tag{26a}$$

or

$$U_3(t) = 1 + \int_0^t dt_1\, U_3(t_1)T_3(t_1) \tag{26b}$$

with

$$T_3(t) = e^{t^2C_2}T_2e^{-t^2C_2} - 2tC_2. \tag{26c}$$

We have now

$$U(t) = U_3(t)e^{t^2C_2}e^{tb}e^{ta}. \tag{27}$$

Again from (26b, c)

$$U_3(t) = 1 + \frac{t^3}{6}\{[a, [a, b]] + 2[b, [a, b]]\} + \cdots$$

$$\equiv 1 + t^3C_3 + \cdots,$$

which suggests the next formula,

$$U(t) = U_4(t)e^{t^3C_3}e^{t^2C_2}e^{tb}e^{ta}. \tag{28}$$

One can continue in this way, transferring the lowest power of $t$ to the exponent at each subsequent step. At $n$th step the formulas are

$$U(t) = U_n(t)e^{t^{n-1}C_{n-1}} \cdots e^{t^2C_2}e^{tb}e^{ta}, \tag{29}$$

$$(\partial/\partial t)U_n = U_nT_n(t); \qquad U_n(0) = 1 \tag{30a}$$

or

$$U_n(t) = 1 + \int_0^t dt\, U_n(t)T_n(t), \tag{30b}$$

$$T_n(t) = e^{t^{n-1}C_{n-1}}T_{n-1}e^{-t^{n-1}C_{n-1}} - (n-1)t^{n-2}C_{n-1}, \qquad n > 2, \tag{30c}$$

$$T_1(t) = b(t); \qquad C_1 = b, \tag{30d}$$

$$C_n = \lim_{t\to 0}\left(t^{-n}\int_0^t dt\, T_n(t)\right), \tag{30e}$$

$$T_n \sim t^{n-1} \quad \text{as} \quad t \to 0. \tag{31}$$

By carrying this process indefinitely, formula (11) clearly results and that is equivalent to formula (8).

All these formulas can be strictly valid only for small values of $t$, since at each stage the operator coefficient of the power of $t$ transferred to the exponential operator is obtained by examining the behavior of kernels for small $t$. Roughly speaking, $U_n(t)$ is obtained by repeated multiplication of $\int_0^t T_n(t)\, dt$. The lowest term in this expression is proportional to $t^n$ and contains all powers of $\lambda$ from one to $n$. Thus the convergence is limited to times $(t^n\lambda) \ll 1$ if $\lambda \ll 1$ and to times $(t\lambda) \ll 1$ if $\lambda \gg 1$, so that in either case $t$ cannot be very large.

## 5. GENERALIZATION OF A THEOREM OF SÁENZ AND O'ROURKE

Any $U_n$ can be approximated by a finite power series

$$U_n \doteq [U_n(t)]^m = \sum_{r=0}^{m} M_r(t), \tag{32a}$$

$$M_r(t) = \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{r-1}} dt_r \times T_n(t_r) \cdots T_n(t_1). \tag{32b}$$

Then the approximation

$$[U_n(t)]^m e^{t^{n-1}C_{n-1}} \cdots e^{bt} e^{at} \tag{33}$$

always contains all the terms of the lower approximation

$$[U_{n-r}(t)]^m e^{t^{n-r-1}C_{n-r-1}} \cdots e^{bt} e^{at}. \tag{34}$$

The proof is obtained by taking $m = 0, 1, 2, \cdots$, etc. and comparing the terms.

This theorem was enunciated by Sáenz and O'Rourke[5] for $n = 2$, and was used by them to draw the conclusion that the approximation to the formula (23) with $n = 2$ is always better than the corresponding approximation to the formula (11) with $n = 1$. This conclusion is not very sound. The experience with perturbation theory has shown that there are certain classes of terms which, if taken together, cancel each other or otherwise produce reasonable behavior leading to a good approximation, but if all of the group is not kept together then the approximation fails. These groupings of terms depend crucially on the detailed physical properties of the system and hence the mere fact that one approximation contains more terms than another tells us nothing about their relative merits.

## 6. A SEQUENCE OF FORMULAS LEADING TO EQ. (9)

An improvement over the formulas of the previous sequence can be obtained if at each step instead of transferring a single power of $t$ to the exponential we transfer whole power series (operator functions of $t$) to the exponential. One needs some criterion for selecting these power series. It is likely that considerations of the physical situation would be very useful in this connection. The very minimum of the physical consideration furnishes us with the parameter $\lambda$, which already leads to interesting results. It is not necessary here to assume that $\lambda$ is small.

As before the first step is to write

$$U(t) = \mathfrak{U}_1(t)e^{at}; \qquad \mathfrak{U}_1(t) = U_1(t)$$

which leads to the perturbation formula (11). In the next step we transfer all the terms linear in $\lambda$ to the exponential and write

$$\mathfrak{U}_1 = \mathfrak{U}_2 \exp\left[\lambda \int_0^t dt_1\, b(t_1)\right] \equiv \mathfrak{U}_2 e^{\lambda\sigma_1} \tag{35}$$

which gives

$$\frac{\partial}{\partial t}\mathfrak{U}_2 = \mathfrak{U}_2\mathfrak{J}_2; \qquad \mathfrak{U}_2(0) = 1, \tag{36a}$$

$$\mathfrak{J}_2(t) = \left[\frac{\lambda^2}{2!}[\sigma_1, b] + \frac{\lambda^3}{3!}\{[\sigma_1^2, b] + [\sigma_1, \sigma_1 b]\} \cdots\right] e^{-\lambda\sigma_1} \tag{36b}$$

which gives for $\mathfrak{U}_2$

$$\mathfrak{U}_2 = 1 + \frac{\lambda^2}{2!}\int_0^t dt_1 \int_0^{t_1} dt_2\, [b(t_2), b(t_1)] + \cdots = 1 + \lambda^2\sigma_2(t) + \cdots, \tag{37}$$

from where we take the next step,

$$\mathfrak{U}_2 = \mathfrak{U}_3 e^{\lambda^2\sigma_2(t)}. \tag{38}$$

Proceeding in this way at $n$th step, the equations are

$$U(t) = \mathfrak{U}_n(t) e^{\lambda^{n-1}\sigma_{n-1}(t)} \cdots e^{\lambda^2\sigma_2(t)} e^{\lambda\sigma_1(t)} e^{at}, \tag{39}$$

$$\frac{\partial}{\partial t}\mathfrak{U}_n(t) = \mathfrak{U}_n(t)\mathfrak{J}_n(t); \qquad \mathfrak{U}_n(0) = 1 \tag{40a}$$

or

$$\mathfrak{U}_n(t) = 1 + \int_0^t dt_1\, \mathfrak{U}_n(t_1)\mathfrak{J}_n(t_1) \tag{40b}$$

with

$$\mathfrak{J}_n(t) = e^{\lambda^{n-1}\sigma_{n-1}(t)}\mathfrak{J}_{n-1}e^{-\lambda^{n-1}\sigma_{n-1}(t)} + e^{\lambda^{n-1}\sigma_{n-1}(t)}\frac{\partial}{\partial t}e^{-\lambda^{n-1}\sigma_{n-1}(t)}; n > 2, \tag{40c}$$

$$\mathfrak{J}_1(t) = \lambda b(t), \tag{40d}$$

and

$$\sigma_n(t) = \lim_{\lambda\to 0}\left(\lambda^{-n}\int_0^t \mathfrak{J}_n(t)\, dt\right). \tag{40e}$$

On carrying on the process indefinitely, Eq. (12) evidently results, which is equivalent to Eq. (9).

The second stage of this sequence is similar to the second step of a procedure described by Green[6] for solving the density matrix equation.

The first term in the power series of $\mathfrak{J}_n(t)$ is proportional to $\lambda^n t^{n+1}$ so that the series for $\mathfrak{U}_n$ can be valid for time $t \ll \lambda^{-1}$ for large $n$. At least for small values $\lambda$ the present sequence may be considered an improvement over the previous sequence.

## 7. COMPARISON OF THE TWO SEQUENCES: $\lim t \to \infty$

The main reason for considering the present sequence an improvement over the previous one is the greater facility it provides for physical intuition. If $a$ and $b$ were appropriate quantum mechanical operators, then in the language of quantum mechanics $b(t) = e^{at}be^{-at}$ would be the time-dependent Heisenberg operator for physical quantity $b$ corresponding to the "Hamiltonian" $a/i$. Comparing the two sequences one sees that, whereas in the first case one was dealing with the instantaneous values of the operators at time $t = 0$, in the present case one deals instead with the integrals over the whole interval zero to $t$.

The fact that one deals with integrals rather than instantaneous values leads one to expect that average effects will be more adequately represented by the second formula. For instance, it is frequently required to find the behavior of $U(t)$ or some related quantity as $t \to \infty$. For this case difficulties of ordinary perturbation formula are well known and the formulas of Sec. 4 offer no advantage. In particular, in the latter case it is not easy to see how various powers of $t$ in the formula are to lead to a finite result in the limit. In contrast, the formulas of Sec. 6 make it very plausible that such a limit could exist and be finite, at the very least they show that it would happen provided integrals of $b(t)$ and various commutators exist, since the exponential operators always exist if their exponents do.

In formulas of both sequences, the approximations obtained by putting $U_n = 1$ or $\mathfrak{U}_n = 1$ give unitary operators if $a$ and $b$ are anti-Hermitian. This property is desirable in some physical problems.

## 8. DERIVATIVES OF THE EXPONENTIAL WITH RESPECT TO $\lambda$ AND THEIR RELATION TO $\sigma$'s

Since

$$e^{A\tau}e^{At}e^{-A\tau} = e^{At}, \tag{41}$$

Eq. (11) can be converted to

$$e^{At} = \cdots e^{t^n\mathbf{C}_n(\tau)} \cdots e^{t^2\mathbf{C}_2(\tau)}e^{t\mathbf{b}(\tau)}e^{t\mathbf{a}(\tau)} \tag{42}$$

with

$$\mathbf{X}(\tau) = e^{A\tau}Xe^{-A\tau}, \tag{43}$$

which means that the operator $\exp At$ can be determined completely, provided the operators $\mathbf{a}(\tau)$ and $\mathbf{b}(\tau)$ are given for *any* single value of $\tau$, the point $\tau = 0$ having no special significance. This result again could be anticipated from physical considerations. A similar modification of other formulas could be made.

From (12) we obtain

$$e^{(A+\epsilon b)t} = \cdots e^{\epsilon^n\sigma_n'(t)} \cdots e^{\epsilon\sigma_1'(t)}e^{tA}, \tag{44}$$

where $\sigma_n'$ has the same structure as $\sigma_n$ except that $b(t)$ is everywhere replaced by $\mathbf{b}(t)$; for instance,

$$\sigma_1'(t) = \int_0^t dt_1\, \mathbf{b}(t_1). \tag{45}$$

Then using the basic definition of the derivative one gets with $A = a + \lambda b$,

$$\frac{d}{d\lambda}e^{At} = \lim_{\epsilon\to 0}\frac{e^{(A+\epsilon b)t} - e^{At}}{\epsilon} = \int_0^t dt_1\, \mathbf{b}(t_1)e^{At}. \tag{46}$$

This is a much more compact expression than would be obtained by differentiating the power series (5) or (10). The same power series for the derivative can also be obtained from (46). It follows in general that

$$\frac{d^n}{d\lambda^n}e^{At} = \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_n} dt_n\, \mathbf{b}(t_n) \times \mathbf{b}(t_{n-1}) \cdots \mathbf{b}(t_1)e^{At}, \tag{47}$$

a result which makes sense immediately since Eq. (10) is actually the Taylor expansion of $\exp At$.

Introducing the quantities

$$d_n(t) = [(d^n/d\lambda^n)e^{At}]e^{-At}, \tag{48}$$

we obtain from Eqs. (16)

$$\sigma_1' = d_1, \tag{49a}$$

$$\sigma_2' = \frac{1}{2!}(d_2 - d_1^2), \tag{49b}$$

$$\sigma_3' = \frac{1}{3!}(d_3 - 3d_2d_1 + 2d_1^3), \tag{49c}$$

$$\sigma_4' = \frac{1}{4!}(d_4 - 4d_3d_1 - 3d_2^2 + 9d_2d_1^2 + 3d_1^2d_2 - 6d_1^4). \tag{49d}$$

These terms have a certain resemblance to the derivatives $d^n(\ln x)/d\lambda^n$, when $x$ is put equal to $e^{At}$ *after* differentiation. Evidently when $a$ and $b$ commute $d_n(t) = b^nt^n$, and all $\sigma_n'$ vanish for $n > 1$, as required.

If for some particular problem one can assume that a particular derivative is small then by associating a parameter with it, it would be possible to generate new sequences of approximations in a manner very similar to the ones already illustrated in Secs. 4 and 6.

## 9. OTHER SEQUENCES

The essential process is to write $T_n = T_n' + T_n''$ in

$$\partial U_n/\partial t = U_nT_n \tag{50}$$

and set

$$U_n = U_{n+1}e^{T_n'}. \tag{51}$$

Then $U_{n+1}$ is given by

$$\begin{aligned}(\partial/\partial t)U_{n+1} &= U_{n+1}T_{n+1}, \qquad U_{n+1}(0) = 1;\\ T_{n+1} &= e^{T_n'}T_n e^{-T_n'} + e^{T_n'}(\partial/\partial t)e^{-T_n'}.\end{aligned} \tag{52}$$

The choice of separation of $T_n'$ is quite arbitrary at each step so that a great variety of formulas is possible. Green[6] has considered a sequence in which at every alternate step $T_n'$ is the diagonal part of the kernel. Although he deals with the solution of density matrix, a comparison is however possible. It appears that he has not taken into account the $t$-dependence of the diagonal parts in his scheme; otherwise such a scheme may indeed be useful for some purposes.

## 10. THE EVOLUTION OPERATOR OF QUANTUM MECHANICS

If $H_0$ is the noninteracting part of the Hamiltonian and $V$ the interaction then the evolution operator[10] in the interaction picture is defined as

$$\begin{aligned}U(t, t') &= e^{(it/\hbar)H_0}e^{(-it/\hbar)H}e^{(+it'/\hbar)H}e^{(-it'/\hbar)H_0}\\ &\equiv U_1^{-1}(t)U_1(t'),\end{aligned} \tag{54}$$

provided $U_1$ is given by (19) and

$$\begin{aligned}H_0 &= \frac{i}{\hbar}a; \qquad V = \frac{i}{\hbar}b;\\ H &\equiv H_0 + \lambda V = \frac{\hbar}{i}A.\end{aligned} \tag{55}$$

The integral equation satisfied by $U(t, t')$ is

$$U(t, t') = 1 - \frac{i}{\hbar}\int_{t'}^{t} d\tau\, V(\tau)U(\tau, t'), \tag{56}$$

which is entirely analogous to the equations previously considered since $t'$ does not play any essential role. In formulas of type (8), instead of powers of $t$ alone we shall have combinations of powers of $t$ and $t'$ which arise from integrals of the type $\int_{t'}^{t} t^n dt$. But the formulas similar to those in Sec. 6 will retain their compact form; in fact, we shall have

$$\begin{aligned}U(t, t') = \cdots \exp\left[\frac{1}{2}\left(-\frac{i}{\hbar}\right)^2\int_{t'}^{t} dt_1\int_{t'}^{t_1} dt_2\right.\\ \left.\times\,[V(t_1), V(t_2)]\right]\exp\left[-\frac{i}{\hbar}\int_{t'}^{t} V(t_1)\,dt_1\right].\end{aligned} \tag{57}$$

[10] For example, see S. S. Schweber, H. A. Bethe and F. de Hoffman; *Mesons and Fields* (Row, Peterson and Company, Evanston, Illinois, 1955), Vol. I.

For problems in which the perturbation changes neither too suddenly nor too slowly this formula could perhaps be a better basis of discussion than the ordinary perturbation theory. The feature which promises a more rapid convergence is the appearance of the commutators rather than merely the products of $V(t)$—apart from the fact that they occur as arguments in exponential functions. This is not to say that actual evaluation of these operators will be easier.

The same formula with different limits of integration describes the $S$-matrix, as

$$S = U(+\infty, -\infty). \tag{58}$$

## 11. A PROBLEM IN STATISTICAL MECHANICS

The type of advantage offered by these formulas is illustrated by means of a simple example. The only remarkable thing is that an approximation suitable to a given situation is obtained. We consider only the first term, and say nothing about what may happen on including higher terms. The practical interest of the example is quite limited.

Consider a one-dimensional classical motion of a particle in an external field. Then the Hamiltonian is

$$H = E = \tfrac{1}{2}p^2 + v(q); \qquad (m = 1) \tag{59}$$

and the Liouville operator

$$L = \frac{\partial H}{\partial p}\frac{\partial}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial}{\partial p}. \tag{60}$$

The development of a classical ensemble is described by the operator

$$\exp(tL) = \exp\left(t(p\partial_q - \partial_q v)\partial_p\right) \tag{61}$$

so that $a \equiv p\partial_q$; $\lambda b = -(\partial_q v)\partial_p$; hence

$$\begin{aligned}b(t) &= -e^{tp\partial_q}(\partial_q v)\partial_p e^{-tp\partial_q}\\ &= -\sum_{n=0}^{\infty}\frac{t^n p^n}{n!}(\partial_q^{n+1}v)[-t\partial_q + \partial_p],\end{aligned} \tag{62}$$

since

$$[\partial_q, (\partial_q v)] = (\partial_q^2 v),$$

from which

$$\begin{aligned}\int_0^t b(t_1)\,dt_1 = \frac{1}{p}\left[tv(q + pt) - \frac{1}{p}\int_q^{q+pt} dq\, v(q)\right]\partial_q\\ + \frac{1}{p}[v(q) - v(q + pt)]\partial_p.\end{aligned} \tag{63}$$

The coefficient of $\partial_q$, although somewhat complicated, would presumably be small in suitable cir-

cumstance; so that, in the case $v(\pm\infty)$ vanishes, we have

$$\exp\left[\delta\int_0^\infty b(t_1)\,dt_1\right] \cong \exp\left[(v/p)\partial_p\right]. \tag{64}$$

This operator has the interesting properties

$$e^{(v/p)\partial_p}p \equiv \left[1 + \frac{v}{p}\partial_p + \frac{1}{2!}\frac{v}{p}\partial_p\frac{v}{p}\partial_p + \cdots\right]p = (p^2 + 2v)^{\frac{1}{2}} \tag{65}$$

and

$$e^{(v/p)\partial_p}e^{-\frac{1}{2}p^2} = e^{-\frac{1}{2}p^2 - v(q)}. \tag{66}$$

In case it can be shown that the other terms from $\exp(tL)$ give no contributions when applied to these functions, simple interpretations can be given. Thus (65) is an expression of conservation of energy $\frac{1}{2}p_\infty^2 = \frac{1}{2}p_0^2 + v(q_0)$ and (66) shows that a Maxwellian distribution of momenta at a time $t$ is changed, under the influence of interactions, to a Maxwellian distribution of energies at $t = \infty$.

## APPENDIX

Making use of the identity

$$\left(\frac{d}{dt}e^Q\right)e^{-Q} = \dot{Q} + \frac{1}{2!}[Q, \dot{Q}] + \frac{1}{3!}[Q, [Q, \dot{Q}]] + \cdots = \int_0^1 d\alpha\, e^{\alpha Q}\dot{Q}e^{-\alpha Q}, \tag{A1}$$

in Eq. (40c) we have

$$\mathfrak{I}_n(t) = \left\{\mathfrak{I}_{n-1}(t) + \lambda^{n-1}[\sigma_{n-1}, \mathfrak{I}_{n-1}] + \frac{\lambda^{2n-2}}{2!}[\sigma_{n-1}, [\sigma_{n-1}, \mathfrak{I}_{n-1}]] + \cdots\right\} - \left\{\lambda^{n-1}\dot{\sigma}_{n-1} + \frac{\lambda^{2n-2}}{2!}[\sigma_{n-1}, \dot{\sigma}_{n-1}] + \cdots\right\}. \tag{A2}$$

Since by Eq. (40e)

$$\mathfrak{I}_{n-1} = \lambda^{n-1}\dot{\sigma}_{n-1} + \cdots,$$

the leading term in $\mathfrak{I}_n$ is proportional to $\lambda^n$, as expected. In addition, it is seen that the coefficients of the powers $\lambda^n$ to $\lambda^{2n-3}$ in $\mathfrak{I}_n$ and $\mathfrak{I}_{n-1}$ are identical. That is, coefficients of $\lambda^{n+1}$ to $\lambda^{2n-1}$ in $\mathfrak{I}_{n+1}$ and $\mathfrak{I}_n$ are identical, so that the coefficients of $\lambda^{2n-2}$ and $\lambda^{2n-1}$ generated in computing $\mathfrak{I}_n$ are carried along unchanged until they become the first terms in $\mathfrak{I}_{2n-2}$ and $\mathfrak{I}_{2n-1}$. Thus we have the important result that a computation of $\mathfrak{I}_n$ for $n \geq 2$, gives us the quantities $\sigma_{2n-2}$ and $\sigma_{2n-1}$! Of course, the calculation of $\mathfrak{I}_1$ only yields $\sigma_1$.

A recursive relation between derivatives of $\sigma_n$ can be established directly as follows: Consider a solution of the form

$$U = \cdots e^{Q_n}\cdots e^{Q_2}e^{Q_1}e^{Q_0}. \tag{A3}$$

Then

$$\frac{dU}{dt} = UA = U(q_0 + e^{-Q_0}q_1e^{Q_0} + e^{-Q_0}e^{-Q_1}q_2e^{Q_1}e^{Q_0} + \cdots), \tag{A4}$$

where

$$q_n = e^{-Q_n}\frac{d}{dt}e^{Q_n} = \int_{-1}^0 d\alpha\, e^{\alpha Q_n}\dot{Q}_n e^{-\alpha Q_n}. \tag{A5}$$

The solution (A3) can be constructed by solving the equation

$$A = q_0 + e^{-Q_0}q_1e^{Q_0} + e^{-Q_0}e^{-Q_1}q_2e^{Q_1}e^{Q_0} + \cdots \tag{A6}$$

as follows: For arbitrary operator $B$,

$$\begin{aligned}
&\dot{Q}_0 = B, \quad Q_0 = \int_0^t B\,dt, \quad q_0 = \int_{-1}^0 d\alpha\, e^{\alpha Q_0}B\,e^{-\alpha Q_0},\\
&\dot{Q}_1 = e^{Q_0}(A - q_0)e^{-Q_0}; \qquad Q_1 = \int_0^t \dot{Q}_1\,dt,\\
&\dot{Q}_2 = e^{Q_1}(\dot{Q}_1 - q_1)e^{-Q_1},\\
&\qquad\qquad\cdots\cdots\cdots\cdots\\
&\dot{Q}_n = e^{Q_{n-1}}\dot{Q}_{n-1}e^{-Q_{n-1}} - \int_0^1 d\alpha\, e^{\alpha Q_{n-1}}\dot{Q}_{n-1}e^{-\alpha Q_{n-1}}.
\end{aligned} \tag{A7}$$

This relation is independent of $A$ and $B$ and is determined entirely by the form (A3) and the above process of solution. The method also holds[9] when $A$ is a function of $t$—a case we have not considered in the rest of this paper.

If we start with $B = a$, $Q_n$ are identical with $\sigma_n$ and it follows from (A7) and Eq. (7) of text that all $\sigma_n$, for $n > 1$, are formed from integrals involving only the commutators of $b(t)$.

# Note on Orthogonal Polynomials which are "Invariant in Form" to Rotations of Axes

C. D. MALDONADO
*Northrop Space Laboratories, Hawthorne, California*
(Received 19 April 1965)

The precedure of Bhatia and Wolf for constructing orthogonal sets whose elements are "invariant in form" with respect to rotations of axes, is extended to include sets which are defined over the entire two-dimensional plane. Use was made of the Gram–Schmidt process to derive general expressions for generating elements of many unique and complete sets corresponding to different circularly symmetric weight functions for two cases, where in the first case the elements are only functions of the real variables $x$ and $y$ while in the other case they are also functions of the real variable $r = (x^2 + y^2)^{\frac{1}{2}}$. These general expressions were used to obtain two new, unique and complete sets corresponding to Gaussian and exponential weight functions, respectively. The radial polynomials for these two sets were found to be closely related to the Laguerre polynomials. The generating functions for these radial polynomials are also given.

## I. INTRODUCTION

A PROCEDURE for constructing the Zernike polynomials[1] which form a complete set for the interior of the unit circle and which are "invariant in form" with respect to rotations of axes about the origin of coordinates has been presented by Bhatia and Wolf.[2] The fundamental role which these polynomials have played in the diffraction theory of optical aberrations is well known[3–5] and recently, Herlitz[6,7] has used them in obtaining a solution to the integral equation which relates the emission coefficient to the emitted spectral intensity for the nontrivial case of asymmetrical light sources. An extension of Herlitz's method to include asymmetrical light sources of infinite extent has been carried out[8] by making use of orthogonal polynomials which are defined over the entire plane and whose properties are analogous to those of Zernike. The procedure for constructing these and other related sets of orthogonal polynomials will represent the main topic of discussion in this paper.

The procedure is basically one which makes use of the Gram–Schmidt[9] process to construct the elements of such sets by orthogonalizing linearly independent functions whose properties are consistent with those as predicted by the procedure of Bhatia and Wolf. In this manner we arrive at general expressions for generating the elements of many unique and complete sets corresponding to different circularly symmetric weight functions for two separate cases where in one case the elements or polynomials of the sets are functions of the real variables $x$ and $y$ while in the other case they are also functions of the real variable $r = (x^2 + y^2)^{\frac{1}{2}}$.

Two new, unique and complete sets of orthogonal polynomials corresponding to Gaussian and exponential weight functions, respectively, were obtained from these general expressions and the explicit results for the radial polynomials of these sets were found to be closely related to the Laguerre polynomials. The generating functions for these radial polynomials are presented and use is made of them to obtain an expression which relates the radial polynomials of both sets.

## II. ORTHOGONAL SETS

We present here a procedure for constructing polynomials (not necessarily real) which form complete sets on the entire two dimensional plane and whose elements are "invariant in form" with respect to rotation of axes. Before this is done, it is important that the meaning of the term "invariant in form" be clarified. For this we refer to the paper by Bhatia and Wolf in which they have coined the phrase "invariant in form" to mean that when any rotation

$$\begin{aligned} x' &= x \cos\varphi + y \sin\varphi, \\ y' &= -x \sin\varphi + y \cos\varphi, \end{aligned} \tag{2.1}$$

[1] F. Zernike, Physica 1, 689 (1934).
[2] A. B. Bhatia and E. Wolf, Proc. Cambridge Phil. Soc. 50, 40 (1954).
[3] B. R. A. Nijboer, Physica 10, 679 (1947).
[4] B. R. A. Nijboer, Physica 13, 605 (1947).
[5] M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, Ltd., London, 1959).
[6] S. I. Herlitz, Ark. Fysik 23, 571 (1963).
[7] S. I. Herlitz, Addendum to Ref. 6 (March 1963).
[8] C. D. Maldonado (to be published).
[9] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), Vol. I.

is applied, each element of the set, which we shall denote by $V(x, y)$, is transformed into an element of the same form, i.e.,

$$V(x, y) = G(\varphi)V(x', y'), \tag{2.2}$$

$G(\varphi)$ being a continuous function with period $2\pi$ in the angle of rotation $\varphi$ and satisfies the boundary condition $G(0) = 1$. Also in their paper, Bhatia and Wolf prove (see Theorem 1) that if $V(x, y)$ is a polynomial of degree $n$ then it can be expressed in polar coordinates $(r, \varphi)$ as

$$V_n^l(r\cos\varphi, r\sin\varphi) = R_n^l(r)\exp(il\varphi), \tag{2.3}$$

where $l$ is an integer positive, negative or zero and $R_n^l(r)$ is a polynomial in $r$ of degree $n$, containing no power of $r$ lower than $|l|$; moreover, $R_n^l(r)$ is an even or odd polynomial according as $l$ is even or odd.

Actually one can show that there is an infinity of such sets whose elements are in accord with this theorem, depending on the manner with which the linearly independent functions

$$r^{|l|}\exp(il\varphi), r^{|l|+2}\exp(il\varphi), r^{|l|+4}\exp(il\varphi), \cdots \tag{2.4}$$

are ordered when the polynomials, $V_n^l(r\cos\varphi, r\sin\varphi)$, are constructed from them by the Gram–Schmidt process. Of this infinity of admissible sets we will concern ourselves with that set which is obtained by orthogonalizing the sequence of functions in the order as shown in Eq. (2.4), since according to Theorem 2, which Bhatia and Wolf prove in their paper, it represents the one and only set which contains a polynomial for each permissible pair of values of $n$ (degree) and $l$ (angular dependence), i.e., for integral values of $n$ and $l$ such that $n \geq 0$, $l \gtrless 0$, $n \geq |l|$ and $n - |l|$ is even. It is important to note that this uniqueness theorem is true only relative to the weight function which is chosen for carrying out the Gram–Schmidt process.

The Gram–Schmidt process for constructing the elements of this unique set as an ascending sequence of polynomials in the degree $n$ by forming linear combinations of the functions in the order as shown in Eq. (2.4) and requiring that the elements of the set be normalized and orthogonal with respect to a circularly symmetric weight function $K(r)$, i.e.,

$$\int_0^\infty \int_0^{2\pi} d\varphi\, dr\, r V_n^l(r\cos\varphi, r\sin\varphi) \times \tilde{V}_\beta^\alpha(r\cos\varphi, r\sin\varphi)K(r) = \delta_n^\beta \delta_l^\alpha, \tag{2.5}$$

where $\sim$ denotes complex conjugate and $\delta_n^\beta$ and $\delta_l^\alpha$ are Kronecker symbols, is a straightforward iterative scheme which we can compactly write as

$$V_{m+2k}^{\pm m}(r\cos\varphi, r\sin\varphi) = \exp(\pm im\varphi)R_{m+2k}^{\pm m}(r), \tag{2.6}$$

where

$$R_{m+2k}^{\pm m}(r) = (1/2\pi)^{\frac{1}{2}} r^m C_k^m S_k^m(r), \tag{2.7}$$

$$C_k^m = \left\{\int_0^\infty dr\, r^{2m+1}[S_k^m(r)]^2 K(r)\right\}^{-\frac{1}{2}}, \tag{2.8}$$

$$S_k^m(r) = r^{2k} + \sum_{p=1}^{k} A_{k-p}^{m,p} S_{k-p}^m(r), \tag{2.9}$$

$$A_{k-p}^{m,k} = -\left(\int_0^\infty dr\, r^{2(m+k)+1} S_{k-p}^m(r)K(r)\right) \times \left(\int_0^\infty dr\, r^{2m+1}[S_{k-p}^m(r)]^2 K(r)\right)^{-1}, \tag{2.10}$$

$m \equiv |l|$ and $n = m + 2k$ for $m, k = 0, 1, 2, 3, \cdots$. For a given value of $n$ the unique set generated by Eq. (2.6) contains $(n + 1)(n + 2)/2$ linearly independent polynomials of degree $\leq n$. From this it follows that every monomial $x^i y^j$ ($i \geq 0$, $j \geq 0$ integers) and consequently every polynomial may be expressed as a linear combination of a finite number of elements of the set. Then by Weierstrass' theorem[9] on approximations by polynomials we conclude that the set is complete.

### A. Unique Set for $K(r) = \exp(-r^2)$

There are many unique and complete sets, one for each specified weight function, which are derivable from Eq. (2.6); however, we will restrict our attention to that set which corresponds to the circularly symmetric function $K(r) = \exp(-r^2)$, since this set was instrumental in extending the method of Herlitz to include asymmetrical light sources of infinite extent.[8] The procedure for constructing this set is an iterative scheme which makes use of Eqs. (2.6) to (2.10), starting from $k = 0, 1, 2, 3, \cdots$, etc. for a given value of $m$. The computational details are lengthy and the end result for the radial polynomials is an explicit expression of the form

$$R_{m+2k}^{\pm m}(r) = (-1)^k(1/\pi)^{\frac{1}{2}}(k!/(m + k)!)^{\frac{1}{2}} r^m L_k^m(r^2), \tag{2.11}$$

where[10]

$$L_k^m(r^2) = \sum_{s=0}^{k} (-1)^s \frac{(m + s)!}{(k - s)!\,(m + s)!\,s!}(r^2)^s \tag{2.12}$$

denotes the associated Laguerre polynomial of argument $r^2$. A few of these radial polynomials normalized

[10] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. II, Chap. 10.

TABLE I. Radial polynomials $\pi^{\frac{1}{2}}[(m+k)!/k!]^{\frac{1}{2}}R^{\pm m}_{m+2k}(r)$ for $m, k \leq 3$.

| | $k$ | | | |
|---|---|---|---|---|
| $m$ | 0 | 1 | 2 | 3 |
| 0 | 1 | $r^2 - 1$ | $\frac{1}{2}r^4 - 2r^2 + 1$ | $\frac{1}{6}r^6 - \frac{3}{2}r^4 + 3r^2 - 1$ |
| 1 | $r$ | $r^3 - 2r$ | $\frac{1}{2}r^5 - 3r^3 + 3r$ | $\frac{1}{6}r^7 - 2r^5 + 6r^3 - 4r$ |
| 2 | $r^2$ | $r^4 - 3r^2$ | $\frac{1}{2}r^6 - 4r^4 + 6r^2$ | $\frac{1}{6}r^8 - \frac{5}{2}r^6 + 10r^4 - 10r^2$ |
| 3 | $r^3$ | $r^5 - 4r^3$ | $\frac{1}{2}r^7 - 5r^5 + 10r^3$ | $\frac{1}{6}r^9 - 3r^7 + 15r^5 - 20r^3$ |

as $\pi^{\frac{1}{2}}[(m+k)!/k!]^{\frac{1}{2}}R^{\pm m}_{m+2k}(r)$ are compiled in Table I for $m, k \leq 3$.

We now make use of the well-known generating function for the Laguerre polynomials given by[10]

$$\sum_{k=0}^{\infty} z^k L_k^m(x) = (1-z)^{-(m+1)} \exp\left(\frac{xz}{(z-1)}\right), \quad (2.13)$$

which is valid provided $|z| < 1$, to arrive at a generating function for the radial polynomials of Eq. (2.11). To do this we set $x = r^2$ in Eq. (2.13) and solve for $L_k^m(r^2)$ in terms of $R^{\pm m}_{m+2k}(r)$ from Eq. (2.11) then upon substitution into Eq. (2.13) we obtain

$$\sum_{k=0}^{\infty} \left(\frac{(m+k)!}{k!}\right)^{\frac{1}{2}} z^k R^{\pm m}_{m+2k}(r) = \left(\frac{1}{\pi}\right)^{\frac{1}{2}} r^m (1+z)^{-(m+1)} \exp\left(\frac{zr^2}{(1+z)}\right) \quad (2.14)$$

as the desired expression for the generating function for the radial polynomials, $R^{\pm m}_{m+2k}(r)$.

## III. RELATED ORTHOGONAL SETS

Up to this point in the discussion we have restricted our attention to orthogonal sets whose elements are "invariant in form" with respect to rotations of axes and are functions of the real variables $x$ and $y$; now we wish to consider related sets whose elements are also functions of the real variable $r = (x^2 + y^2)^{\frac{1}{2}}$.

The procedure for constructing the elements of such sets is similar to that as given in the previous section. For example, the "invariant in form" property allows us to use Theorem 1 (Ref. 1) to write a typical element of any set, which we take to be a polynomial of degree $n$ in $x$, $y$ and $r$ and denote it as $W(x, y, r)$, in polar coordinates as

$$W_n^l(r\cos\varphi, r\sin\varphi, r) = Q_n^l(r)\exp(il\varphi), \quad (3.1)$$

where again $l$ is an integer positive, negative or zero and $Q_n^l(r)$ is a polynomial in $r$ of degree $n$ such that it contains no powers of $r$ lower than $|l|$. The polynomial $Q_n^l(r)$ differs from the previous radial polynomial $R_n^l(r)$, in that it no longer is an even or an odd polynomial according as $l$ is even or odd.

It can be shown that there is an infinity of such sets corresponding to the infinite possible ways with which the sequence of linearly independent functions

$$r^{|l|}\exp(il\varphi), r^{|l|+1}\exp(il\varphi), r^{|l|+2}\exp(il\varphi), \cdots \quad (3.2)$$

can be ordered when the elements of a set are constructed from them by the Gram–Schmidt process. From this infinity of allowable sets we will again restrict our attention to that set whose elements are obtained by orthogonalizing the sequence of functions in the order as shown in Eq. (3.2), because according to Theorem 2 (Ref. 1) it represents the one and only set which contains a polynomial for each permissible pair of values $n$ and $l$, i.e., for integral values of $n$ and l such that $n \geq 0$, $l \gtrless 0$, and $n \geq |l|$.

The Gram–Schmidt process for generating the elements of this unique set such that they are orthonormal with respect to a circularly symmetric weight function $K(r)$, i.e.,

$$\int_0^{\infty}\int_0^{\infty} d\varphi\, dr\, r W_n^l(r\cos\varphi, r\sin\varphi, r) \times \bar{W}_\beta^\alpha(r\cos\varphi, r\sin\varphi, r)K(r) = \delta_n^\beta \delta_l^\alpha, \quad (3.3)$$

is an iterative scheme which is compactly given by the following expression:

$$W^{\pm m}_{m+k}(r\cos\varphi, r\sin\varphi, r) = \exp(\pm im\varphi)Q^{\pm m}_{m+k}(r), \quad (3.4)$$

where

$$Q^{\pm m}_{m+k}(r) = (1/2\pi)^{\frac{1}{2}} r^m B_k^m P_k^m(r), \quad (3.5)$$

$$B_k^m = \left(\int_0^{\infty} dr\, r^{2m+1}[P_k^m(r)]^2 K(r)\right)^{-\frac{1}{2}}, \quad (3.6)$$

$$P_k^m(r) = r^k + \sum_{p=1}^{k} D_{k-p}^{m,p} P_{k-p}^m(r), \quad (3.7)$$

$$D_{k-p}^{m,p} = -\left(\int_0^{\infty} dr\, r^{2m+k+1} P_{k-p}^m(r)K(r)\right) \times \left(\int_0^{\infty} dr\, r^{2m+1}[P_{k-p}^m(r)]^2 K(r)\right)^{-1}, \quad (3.8)$$

$m \equiv |l|$ and $n = m + k$ for $m, k = 0, 1, 2, 3, \cdots$. Now one can easily verify that for a given value of $n$ the set generated by Eq. (3.4) contains $(n+1)^2$ linearly independent polynomials of degree $\leq n$; and hence by a similar argument based on Weierstrass' theorem on approximations by polynomials it can be shown that the set is complete.

### A. Unique Set for $K(r) = \exp(-r)$

Of the many complete and unique sets corresponding to different weight functions, we use as an illustrative example the set which corresponds to the weight function $K(r) = \exp(-r)$. For this chosen weight function the Gram–Schmidt process of Eq. (3.4) yields the following explicit expression:

$$Q^{\pm m}_{m+k}(r) = (-1)^k(1/2\pi)^{\frac{1}{2}} \times [k!/(2m+k+1)!]^{\frac{1}{2}} r^m L_k^{2m+1}(r), \qquad (3.9)$$

for generating the radial polynomials of this unique set, where $L_k^{2m+1}(r)$ is the associated Laguerre polynomial of Eq. (2.12) with $m$ and $r^2$ replaced by $2m+1$ and $r$, respectively. In Table II a few of these radial polynomials normalized as

$$(2\pi)^{\frac{1}{2}}[(2m+k+1)!/k!]^{\frac{1}{2}} Q^{\pm m}_{m+k}(r)$$

have been compiled for values of $m$ and $k \leq 3$.

TABLE II. Normalized radial polynomials $(2\pi)^{\frac{1}{2}}[(2m+k+1)!/k!]^{\frac{1}{2}} Q^{\pm m}_{m+k}(r)$ for $m, k \leq 3$.

| | $k$ | | | |
|---|---|---|---|---|
| $m$ | 0 | 1 | 2 | 3 |
| 0 | 1 | $r - 2$ | $\frac{1}{2}r^2 - 3r + 3$ | $\frac{1}{6}r^3 - 2r^2 + 6r - 4$ |
| 1 | $r$ | $r^2 - 4r$ | $\frac{1}{2}r^3 - 5r^2 + 10r$ | $\frac{1}{6}r^4 - 3r^3 + 15r^2 - 20r$ |
| 2 | $r^2$ | $r^3 - 6r^2$ | $\frac{1}{2}r^4 - 7r^3 + 21r^2$ | $\frac{1}{6}r^5 - 4r^4 + 28r^3 - 56r^2$ |
| 3 | $r^3$ | $r^4 - 8r^3$ | $\frac{1}{2}r^5 - 9r^4 + 36r^3$ | $\frac{1}{6}r^6 - 5r^5 + 45r^4 - 120r^3$ |

The generating function for these radial polynomials is obtained by replacing $x$ and $m$ in Eq. (2.13) by $r$ and $2m+1$, respectively, then substituting $L_k^{2m+1}(r)$ in terms of $Q^{\pm m}_{m+k}(r)$ from Eq. (3.9); this gives the result:

$$\sum_{k=0}^{\infty} \left(\frac{(2m+k+1)!}{k!}\right)^{\frac{1}{2}} z^k Q^{\pm m}_{m+k}(r) = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} r^m (1+z)^{-(2m+2)} \exp\left(\frac{zr}{(1+z)}\right). \qquad (3.10)$$

Also, if we replace $r$ by $r^2$ in this latter expression and $m$ by $2m+1$ in Eq. (2.14), then compare the resultant expressions we obtain

$$2^{\frac{1}{2}} r Q^{\pm m}_{m+k}(r^2) = R^{\pm(2m+1)}_{2m+2k+1}(r) \qquad (3.11)$$

as a relationship which exists between the radial polynomials of the two unique sets which we have chosen to consider in this paper.

### ACKNOWLEDGMENT

The author is indebted to Professor Alfredo Baños, Jr. for several helpful discussions.

# Half-Space Neutron Transport with Linearly Anisotropic Scattering

N. J. McCormick* and I. Kuščer*
*Department of Nuclear Engineering, The University of Michigan, Ann Arbor, Michigan*
(Received 12 April 1965)

The method developed by Case is used to solve four time-independent, one-speed problems for neutron transport in a homogeneous medium where the scattering function is linear in the cosine of the scattering angle. The solutions to the albedo, Milne, Green's function, and constant isotropic source problems for a half-space are facilitated by the use of half-range bi-orthogonality relations between the eigenfunctions of the homogeneous transport equation. Expressions are also derived for the emerging angular densities and the densities and net currents on the surface of the half-space.

## I. INTRODUCTION

THE Case approach to solving neutron transport problems utilizes an expansion of the neutron angular density in terms of the eigenfunctions of the homogeneous transport equation. The set of eigenfunctions was first shown to be complete for the case of isotropic scattering of one-speed neutrons.[1] Using these results, answers to many problems were obtained.[1-5] The completeness theorem for one-speed neutrons was extended to the case of anisotropic scattering by Mika[6] and explicit results for the Milne problem with linearly anisotropic scattering were found by Shure and Natelson.[7]

Recently, orthogonality relations between the eigenfunctions were observed[8] which simplified the solution of one-speed problems with isotropic scattering. An extension to the case of linearly anisotropic scattering was also indicated at that time. It is this approach which is followed here.

An attempt towards further generalization has recently been made.[9] It appears that the eigenfunctions generally obey a set of bi-orthogonality relations of the same form as mentioned in Ref. 8. If the scattering function is of order $N$ in the cosine of the scattering angle, the eigenfunction $\varphi_\nu(\mu)$ and its "adjoint" $\tilde{\varphi}_\nu(\mu)$ differ by a term $\frac{1}{2}c\nu B(\nu, \mu)$, where $B$ is a polynomial of order $(N-1)$ in both variables. A rigorous proof of these two statements is still lacking. In any event, the computation of $B(\nu, \mu)$ would in general be very tedious. Only in two cases is the situation comparatively simple: in that of linearly anisotropic scattering ($N = 1$) and in that of a nonabsorbing medium with $N = 2$.[9] We will restrict ourselves to the first case.

Sections II–IV deal with an absorbing medium. After presenting the bi-orthogonality relations (Sec. II), we apply them to four standard half-space problems (Sec. III): the albedo, Milne, Green's function, and constant isotropic source problems. A special calculation (Sec. IV) leads to simplified formulas for the emerging angular densities and related quantities. In Sec. V, the first three problems are solved for a nonabsorbing medium.

## II. DEFINITIONS AND BI-ORTHOGONALITY RELATIONS

The transport equation to be solved, written in the usual notation,[6,7] is

$$\left[\mu \frac{\partial}{\partial x} + 1\right]\psi(x, \mu) = \frac{c}{2}\int_{-1}^{1} [1 + b\mu\mu']\psi(x, \mu')\, d\mu' + q, \qquad (1)$$

where $\psi(x, \mu)$ is the azimuthal integral of the angular density. Here $-1 \leq b \leq 1$, and we choose $c < 1$, deferring the case $c = 1$ until Sec. V. For the homogeneous equation, separation of variables is achieved through the ansatz

$$\psi(x, \mu) = e^{-x/\nu}\varphi_\nu(\mu), \qquad (2)$$

where the eigenfunctions $\varphi_\nu(\mu)$ are normalized such that

$$\int_{-1}^{1} \varphi_\nu(\mu)\, d\mu = 1. \qquad (3)$$

* Present address: Institute of Physics, University of Ljubljana, Ljubljana, Yugoslavia.

[1] K. M. Case, Ann. Phys. (N. Y.) 9, 1 (1960).
[2] K. M. Case, *Recent Developments in Neutron Transport Theory*, Michigan Memorial Phoenix Project Report, The University of Michigan (1961).
[3] M. R. Mendelson and G. C. Summerfield, J. Math. Phys. 5, 668 (1964).
[4] G. J. Mitsis, Nucl. Sci. Eng. 17, 55 (1963).
[5] N. J. McCormick and M. R. Mendelson, Nucl. Sci. Eng. 20, 462 (1964).
[6] J. R. Mika, Nucl. Sci. Eng. 11, 415 (1961).
[7] F. Shure and M. Natelson, Ann. Phys. (N. Y.) 26, 274 (1964).
[8] I. Kuščer, N. J. McCormick, and G. C. Summerfield, Ann. Phys. 30, 411 (1964).
[9] N. J. McCormick, "One-Speed Neutron Transport Problems in Plane Geometry," Ph.D. thesis, The University of Michigan (1964).

There are two kinds of eigenfunctions. The continuum modes, belonging to the interval $-1 < \nu < 1$, are given by[7]

$$\varphi_\nu(\mu) = \tfrac{1}{2}c\nu\, d(\nu\mu)\, \mathrm{P}\,[(\nu-\mu)^{-1}] + \lambda(\nu)\delta(\nu-\mu), \tag{4}$$

$$\lambda(\nu) = 1 - \frac{c\nu}{2}\,\mathrm{P}\int_{-1}^{1}\frac{d(\nu\mu)}{\nu-\mu}\,d\mu$$

$$= d(\nu^2)[1 - c\nu\tanh^{-1}\nu] - b(1-c)^2\nu^2, \tag{5}$$

$$d(\nu\mu) = 1 + b(1-c)\nu\mu. \tag{6}$$

[The symbol P in Eq. (4) is a reminder that we must take the Cauchy principal value of any integral over $\nu$ or $\mu$.] Furthermore, there are two discrete modes,

$$\varphi_\pm(\mu) = \tfrac{1}{2}c\nu_0\, d(\pm\nu_0\mu)/(\nu_0 \mp \mu), \tag{7}$$

belonging to the two real roots,[6] $\pm\nu_0$, of the equation

$$\Lambda(\pm\nu_0) = 0, \tag{8}$$

where

$$\Lambda(z) = 1 - \frac{cz}{2}\int_{-1}^{1}\frac{d(z\mu)}{z-\mu}\,d\mu$$

$$= d(z^2)[1 - cz\tanh^{-1}(1/z)] - b(1-c)^2z^2. \tag{9}$$

It will be useful to know the derivative of $\Lambda(z)$ at $z = \nu_0$:

$$\Lambda'(\nu_0) = \frac{cd(\nu_0^2)}{\nu_0(\nu_0^2-1)} - \frac{(1-c)d(3\nu_0^2)}{\nu_0 d(\nu_0^2)}. \tag{10}$$

We also infer from Eq. (9) that

$$\Lambda(\infty) = (1-c)(1-\tfrac{1}{3}cb). \tag{11}$$

The value of $\lambda(\nu)$ is related to the boundary values of $\Lambda(z)$ on the cut $(-1, 1)$ by the equation

$$\Lambda^\pm(\nu) \equiv \lim_{\epsilon\to 0^+}\Lambda(\nu \pm i\epsilon)$$

$$= \lambda(\nu) \pm i\tfrac{1}{2}\pi c\nu d(\nu^2), \qquad -1 < \nu < 1. \tag{12}$$

As was pointed out in Ref. 8, instead of orthogonality relations we now have bi-orthogonality relations for the eigenfunctions $\varphi_\nu(\mu)$ and $\varphi_\pm(\mu)$, with "adjoints" of the form

$$\tilde{\varphi}_\nu(\mu) = \varphi_\nu(\mu) + \tfrac{1}{2}c\nu B,$$
$$\tilde{\varphi}_\pm(\mu) = \varphi_\pm(\mu) \pm \tfrac{1}{2}c\nu_0 B. \tag{13}$$

These relations are proved in much the same way as those for isotropic scattering, and at the same time the value of the constant $B$ is derived.

The method involves the use of the functions[7]

$$X(z) = \frac{1}{1-z}\exp\left[\frac{1}{2\pi i}\int_0^1 \ln\frac{\Lambda^+(\mu)}{\Lambda^-(\mu)}\frac{d\mu}{\mu-z}\right], \tag{14}$$

$$\gamma(\mu) = \frac{c\mu}{2}\frac{X^+(\mu)}{\Lambda^+(\mu)} = \frac{c\mu}{2}\frac{X^-(\mu)}{\Lambda^-(\mu)}$$

$$= \tfrac{1}{2}c\mu[\Lambda(\infty)(\nu_0^2-\mu^2)X(-\mu)]^{-1}, \quad 0 \le \mu \le 1, \tag{15}$$

and of the moments of the latter,

$$\gamma_n = \int_0^1 \gamma(\mu)\mu^n\,d\mu. \tag{16}$$

In addition, we need the identities[7]

$$X(z)X(-z) = \Lambda(z)/(\nu_0^2 - z^2)\Lambda(\infty), \tag{17}$$

$$X(\nu_0)X(-\nu_0) = -\Lambda'(\nu_0)/2\nu_0\Lambda(\infty), \tag{18}$$

$$X^2(0) = 1/\nu_0^2\Lambda(\infty), \tag{19}$$

$$\int_0^1 \frac{\gamma(\mu)\,d(\mu^2)}{\mu-z}\,d\mu = X(z), \tag{20}$$

$$\mathrm{P}\int_0^1 \frac{\gamma(\mu)\,d(\mu^2)}{\nu-\mu}\,d\mu$$

$$= -(2/c\nu)\lambda(\nu)\gamma(\nu), \qquad 0 < \nu < 1, \tag{21}$$

$$\int_0^1 \gamma(\mu)\,d(\mu^2)\,d\mu = \lim_{z\to\infty}[-zX(z)] = 1, \tag{22}$$

$$\gamma_{-1} + b(1-c)\gamma_1 = X(0), \tag{23}$$

$$d(\nu_0^2)\,d(\bar{\nu}^2) = (1-c)/\gamma_0^2\Lambda(\infty), \tag{24}$$

where

$$\bar{\nu} = \gamma_1/\gamma_0. \tag{25}$$

These identities help us to show that bi-orthogonality in $0 < \mu < 1$ among the set $\varphi_+(\mu)$, $\varphi_\nu(\mu)$, $0 < \nu < 1$, and the adjoints is produced by the weight function $(\nu_0 - \mu)\gamma(\mu)$. We also find that $B$ must be chosen as

$$B = b(1-c)(\nu_0 - \bar{\nu})/d(\nu_0\bar{\nu}). \tag{26}$$

Let us note that the quoted weight function is closely related to Chandrasekhar's $H$-function[7]:

$$(\nu_0 - \mu)\gamma(\mu) = \tfrac{1}{2}c\mu[\Lambda(\infty)]^{-1/2}H(\mu). \tag{27}$$

The bi-orthogonality relations, and a set of related formulas useful in applications, are listed below (where $0 < \nu < 1$ and $0 < \nu' < 1$). In order to save space, we use in some of the formulas the symbol $\xi$ for either $\nu'$ or $\nu_0$. Correspondingly, $\varphi_\xi(\mu)$ denotes either a continuum eigenfunction or $\varphi_+(\mu)$. Products of two singular eigenfunctions will be understood in the same sense as in Ref. 8.

$$\int_0^1 \varphi_\nu(\mu)\tilde{\varphi}_{\nu'}(\mu)(\nu_0-\mu)\gamma(\mu)\,d\mu$$

$$= (\nu_0-\nu)\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu)\delta(\nu-\nu'), \tag{28}$$

$$\int_0^1 \varphi_\nu(\mu)\tilde{\varphi}_+(\mu)(\nu_0-\mu)\gamma(\mu)\,d\mu = 0, \tag{29}$$

$$\int_0^1 \varphi_+(\mu)\tilde{\varphi}_\nu(\mu)(\nu_0 - \mu)\gamma(\mu)\,d\mu = 0, \tag{30}$$

$$\int_0^1 \varphi_+(\mu)\tilde{\varphi}_+(\mu)(\nu_0 - \mu)\gamma(\mu)\,d\mu = -(\tfrac{1}{2}c\nu_0)^2 X(\nu_0)\,d(\nu_0^2), \tag{31}$$

$$\int_0^1 \varphi_{-\nu}(\mu)\tilde{\varphi}_\xi(\mu)(\nu_0 - \mu)\gamma(\mu)\,d\mu = (\tfrac{1}{2}c\nu)^2 \frac{\tilde{\varphi}_\xi(-\nu)}{\Lambda(\infty)(\nu_0 - \nu)\gamma(\nu)}, \tag{32}$$

$$\int_0^1 \varphi_-(\mu)\tilde{\varphi}_\xi(\mu)(\nu_0 - \mu)\gamma(\mu)\,d\mu = c\nu_0\xi X(-\nu_0)\tilde{\varphi}_+(-\xi), \tag{33}$$

$$\int_0^1 \varphi_\xi(\mu)\tilde{\varphi}_{-\nu}(\mu)(\nu_0 - \mu)\gamma(\mu)\,d\mu = \left(\frac{c\nu}{2}\right)^2 \frac{\varphi_\xi(-\nu)}{\Lambda(\infty)(\nu_0 - \nu)\gamma(\nu)}, \tag{34}$$

$$\int_0^1 \varphi_{-\nu'}(\mu)\tilde{\varphi}_{-\nu}(\mu)(\nu_0 - \mu)\gamma(\mu)\,d\mu = \frac{c\nu}{2}\frac{1}{\Lambda(\infty)}\left[\frac{c\nu}{2}\frac{\varphi_{\nu'}(\nu)}{(\nu_0 - \nu)\gamma(\nu)} - \frac{c\nu'}{2}\frac{\tilde{\varphi}_{\nu'}(\nu)}{(\nu_0 - \nu')\gamma(\nu')}\right], \tag{35}$$

$$\int_0^1 \varphi_-(\mu)\tilde{\varphi}_{-\nu}(\mu)(\nu_0 - \mu)\gamma(\mu)\,d\mu = \frac{c\nu}{2}\left[\frac{c\nu}{2}\frac{\varphi_+(\nu)}{\Lambda(\infty)(\nu_0 - \nu)\gamma(\nu)} - 2\nu_0 X(-\nu_0)\tilde{\varphi}_+(\nu)\right], \tag{36}$$

$$\int_0^1 \tilde{\varphi}_\xi(\mu)(\nu_0 - \mu)\gamma(\mu)\,d\mu = \tfrac{1}{2}c\xi(1 - c)/\Lambda(\infty)\gamma_0\,d(\nu_0\bar{\nu}), \tag{37}$$

$$\int_0^1 \tilde{\varphi}_{-\nu}(\mu)(\nu_0 - \mu)\gamma(\mu)\,d\mu = \frac{c\nu}{2}\frac{1}{\Lambda(\infty)}\left[\frac{c\nu}{2}\frac{1}{(\nu_0 - \nu)\gamma(\nu)} - \frac{1 - c}{\gamma_0\,d(\nu_0\bar{\nu})}\right], \tag{38}$$

$$\int_0^1 \tilde{\varphi}_\xi(\mu)(\nu_0 - \mu)\gamma(\mu)\frac{d\mu}{\mu} = \tfrac{1}{2}c[\Lambda(\infty)]^{-1/2}(1 + B\xi), \tag{39}$$

$$\int_0^1 \tilde{\varphi}_{-\nu}(\mu)(\nu_0 - \mu)\gamma(\mu)\frac{d\mu}{\mu} = \frac{c}{2}\left\{\frac{1 - B\nu}{[\Lambda(\infty)]^{1/2}} - \frac{c\nu}{2}\frac{1}{\Lambda(\infty)(\nu_0 - \nu)\gamma(\nu)}\right\}. \tag{40}$$

Through the use of these relations, we can express the results for typical half-space problems in terms of the functions $X$ or $\gamma$ and the moments of the latter. The numerical evaluation of these functions, for any given $c$ and $b$, may be performed by iteration of the nonlinear integral equation of Shure and Natelson.[7]

## III. SOLUTIONS TO STANDARD HALF-SPACE PROBLEMS

We wish to determine the angular density for the (a) albedo, (b) Milne, (c) Green's function, and (d) constant isotropic source problems, all for the half-space $x \geq 0$ and $c < 1$. These problems are defined by the following source and boundary conditions:

$$q = \begin{cases} 0, & \text{(a),(b)} \\ \delta(x - x_0)\delta(\mu - \mu_0), & \text{(c)} \\ 1, & \text{(d)} \end{cases} \tag{41}$$

$$[\psi(x, \mu)]_{x\to\infty} \to \begin{cases} \text{bounded}, & \text{(a),(c),(d)} \\ \varphi_-(\mu)e^{x/\nu_0}, & \text{(b)} \end{cases} \tag{42}$$

$$\psi(0, \mu) = \begin{cases} \delta(\mu - \mu_0), & \mu > 0, \quad \text{(a)} \\ 0, & \mu > 0. \quad \text{(b)–(d)} \end{cases} \tag{43}$$

The desired solution is expanded in terms of the eigensolutions (4) and (7). If $q \not\equiv 0$, one further term has to be added, namely the solution of the corresponding infinite-medium problem. Thus we see that the conditions (41) and (42) are met by the expansion

$$\psi(x, \mu) = f(x, \mu) + a_+\varphi_+(\mu)e^{-x/\nu_0} + \int_0^1 A(\nu)\varphi_\nu(\mu)e^{-x/\nu}\,d\nu, \tag{44}$$

where

$$f(x, \mu) = \begin{cases} 0, & \text{(a)} \\ \varphi_-(\mu)e^{x/\nu_0}, & \text{(b)} \\ G_\infty(x_0, \mu_0 \to x, \mu), & \text{(c)} \\ 1/(1 - c). & \text{(d)} \end{cases} \tag{45}$$

The function $G_\infty(x_0, \mu_0 \to x, \mu)$ is the solution to the infinite-medium Green's function problem for linearly anisotropic scattering[6]:

$$G_\infty(x_0, \mu_0 \to x, \mu) = 2\varphi_\pm(\mu_0)\varphi_\pm(\mu)e^{-|x - x_0|/\nu_0}/c\nu_0^2\Lambda'(\nu_0)\,d(\nu_0^2) + \int_0^1 \frac{\varphi_{\pm\nu}(\mu_0)\varphi_{\pm\nu}(\mu)e^{-|x - x_0|/\nu}}{\nu\Lambda^+(\nu)\Lambda^-(\nu)}\,d\nu, \qquad x \gtrless x_0. \tag{46}$$

The boundary conditions (43), applied to (44), lead to an equation of the form (for $\mu > 0$)

$$a_+\varphi_+(\mu) + \int_0^1 A(\nu)\varphi_\nu(\mu)\,d\nu = \psi(\mu), \tag{47}$$

from which the expansion coefficients $a_+$ and $A(\nu)$ must be determined. Here $\psi(\mu)$ is a known function, namely

$$\psi(\mu) \equiv \psi(0, \mu) - f(0, \mu)\,\psi(\mu) = \begin{cases} \delta(\mu - \mu_0), & \text{(a)} \\ -\varphi_-(\mu), & \text{(b)} \\ -G_\infty(x_0, \mu_0 \to 0, \mu), & \text{(c)} \\ -1/(1-c). & \text{(d)} \end{cases} \tag{48}$$

Use of the half-range bi-orthogonality relations of Sec. II immediately gives the following results:

$$a_+ = R/[-(\tfrac{1}{2}c\nu_0)^2 X(\nu_0)\, d(\nu_0^2)], \tag{49}$$

$$R = \begin{cases} (\nu_0 - \mu_0)\gamma(\mu_0)\tilde{\varphi}_+(\mu_0), & \text{(a)} \\ -c\nu_0^2 X(-\nu_0)\tilde{\varphi}_+(-\nu_0), & \text{(b)} \\ -\left[X(-\nu_0)\tilde{\varphi}_+(-\nu_0)\,\dfrac{2\varphi_-(\mu_0)e^{-x_0/\nu_0}}{\Lambda'(\nu_0)\, d(\nu_0^2)}\right. & \\ \left. +\displaystyle\int_0^1 \left(\frac{c\nu'}{2}\right)^2 \frac{\tilde{\varphi}_+(-\nu')}{\Lambda(\infty)(\nu_0 - \nu')\gamma(\nu')}\,\frac{\varphi_{-\nu'}(\mu_0)e^{-x_0/\nu'}}{\nu'\Lambda^+(\nu')\Lambda^-(\nu')}\,d\nu'\right], & \text{(c)} \\ -\dfrac{c\nu_0}{2}\,\dfrac{1}{\Lambda(\infty)\gamma_0\, d(\nu_0\bar{\nu})}, & \text{(d)} \end{cases} \tag{50}$$

$$A(\nu) = S/[(\nu_0 - \nu)\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu)], \tag{51}$$

$$S = \begin{cases} (\nu_0 - \mu_0)\gamma(\mu_0)\tilde{\varphi}_\nu(\mu_0), & \text{(a)} \\ -c\nu_0\nu X(-\nu_0)\tilde{\varphi}_+(-\nu), & \text{(b)} \\ -\left[\nu X(-\nu_0)\tilde{\varphi}_+(-\nu)\,\dfrac{2\varphi_-(\mu_0)e^{-x_0/\nu_0}}{\nu_0\Lambda'(\nu_0)\, d(\nu_0^2)}\right. & \\ \left. +\displaystyle\int_0^1 \left(\frac{c\nu'}{2}\right)^2 \frac{\tilde{\varphi}_\nu(-\nu')}{\Lambda(\infty)(\nu_0 - \nu')\gamma(\nu')}\,\frac{\varphi_{-\nu'}(\mu_0)e^{-x_0/\nu'}}{\nu'\Lambda^+(\nu')\Lambda^-(\nu')}\,d\nu'\right], & \text{(c)} \\ -\dfrac{c\nu}{2}\,\dfrac{1}{\Lambda(\infty)\gamma_0\, d(\nu_0\bar{\nu})}. & \text{(d)} \end{cases} \tag{52}$$

The above expansion coefficients for the Milne problem were obtained in a different way by Shure and Natelson.[7]

The solutions of the four problems are now complete since the angular density is known from (44), (45), and (49) through (52). The neutron densities and net currents, defined by

$$\rho(x) = \int_{-1}^{1} \psi(x, \mu)\, d\mu, \qquad j(x) = \int_{-1}^{1} \mu\psi(x, \mu)\, d\mu, \tag{53}$$

are easily obtained by integration of (44) and use of (3):

$$\rho(x) = \int_{-1}^{1} f(x, \mu)\, d\mu + a_+ e^{-x/\nu_0} + \int_0^1 A(\nu)e^{-x/\nu}\, d\nu, \tag{54}$$

$$j(x) = \int_{-1}^{1} \mu f(x, \mu)\, d\mu + (1 - c)\left[a_+\nu_0 e^{-x/\nu_0} + \int_0^1 A(\nu)\nu e^{-x/\nu}\, d\nu\right].$$

With the above results, one is able to obtain an expression for the Milne problem extrapolation distance, $z_0$, that is the distance from the surface of the half-space at which the asymptotic density vanishes. We see that

$$\rho_{a*}(x) \equiv e^{x/\nu_0} + a_+ e^{-x/\nu_0} \tag{55}$$

vanishes at $x = -z_0$, with

$$z_0 = \tfrac{1}{2}\nu_0 \ln(-a_+^{-1}). \tag{56}$$

Equations (49) and (50)(b) and the identity

$$\tilde{\varphi}_+(-\nu_0) = \tfrac{1}{4}c\, d(\nu_0^2)\, d(-\nu_0\bar{\nu})/d(\nu_0\bar{\nu}), \tag{57}$$

lead to

$$z_0 = \tfrac{1}{2}\nu_0 \ln[-X(\nu_0)\, d(\nu_0\bar{\nu})/X(-\nu_0)\, d(-\nu_0\bar{\nu})]. \tag{58}$$

## IV. SURFACE QUANTITIES

For the emerging angular distribution from the half-space, we need to evaluate the expansion (44) for $\mu < 0$. Its value for $\mu > 0$ is given by Eq. (43). The switching from positive to negative $\mu$ is performed by a trick; we multiply both sides of Eq. (47) by $\tilde{\varphi}_{-\mu'}(\mu)(\nu_0 - \mu)\gamma(\mu)d\mu$ and integrate. Equation (34) immediately helps us to the general result (for $\mu > 0$):

$$\psi(0, -\mu) = f(0, -\mu) + \left(\frac{2}{c\mu}\right)^2 \Lambda(\infty)(\nu_0 - \mu)\gamma(\mu) \int_0^1 \psi(\mu')\tilde{\varphi}_{-\mu}(\mu')(\nu_0 - \mu')\gamma(\mu')\, d\mu'. \tag{59}$$

For our four problems, the integral here, as well as the integrals yielding the surface densities and net currents, are all contained in the formulas of Sec. II. We thus arrive at the following results[9]:

$$\psi(0, -\mu) = \begin{cases} \dfrac{2}{c\mu_0}\dfrac{2}{c\mu}\Lambda(\infty)(\nu_0 - \mu_0)\gamma(\mu_0)(\nu_0 - \mu)\gamma(\mu)\tilde{\varphi}_{-\mu_0}(\mu), & \text{(a)} \\ \dfrac{4\nu_0}{c\mu}\Lambda(\infty)X(-\nu_0)(\nu_0 - \mu)\gamma(\mu)\tilde{\varphi}_+(\mu), & \text{(b)} \\ (\nu_0 - \mu)\gamma(\mu)\left\{\dfrac{2}{c\nu_0}\dfrac{2}{c\mu}\dfrac{2\Lambda(\infty)X(-\nu_0)}{\Lambda'(\nu_0)\, d(\nu_0^2)}\varphi_+(-\mu_0)\tilde{\varphi}_+(\mu)e^{-x_0/\nu_0}\right. & \\ \quad \left. + \displaystyle\int_0^1 \frac{\varphi_\nu(-\mu_0)\tilde{\varphi}_\nu(\mu)e^{-x_0/\nu}}{\mu(\nu_0 - \nu)\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu)}\, d\nu\right\}, & \text{(c)} \\ \dfrac{2}{c\mu}\dfrac{(\nu_0 - \mu)\gamma(\mu)}{\gamma_0\, d(\nu_0\bar{\nu})}. & \text{(d)} \end{cases} \tag{60}$$

$$\rho(0) = \begin{cases} \dfrac{2}{c\mu_0}[\Lambda(\infty)]^{\frac{1}{2}}(\nu_0 - \mu_0)\gamma(\mu_0)(1 - B\mu_0), & \text{(a)} \\ 2\nu_0[\Lambda(\infty)]^{\frac{1}{2}}X(-\nu_0)(1 + B\nu_0), & \text{(b)} \\ 2[\Lambda(\infty)]^{\frac{1}{2}}X(-\nu_0)(1 + B\nu_0)\dfrac{2\varphi_+(-\mu_0)e^{-x_0/\nu_0}}{c\nu_0\Lambda'(\nu_0)\, d(\nu_0^2)} & \\ \quad + \frac{1}{2}c[\Lambda(\infty)]^{-\frac{1}{2}}\displaystyle\int_0^1 \frac{\varphi_\nu(-\mu_0)(1 + B\nu)e^{-x_0/\nu}}{(\nu_0 - \nu)\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu)}\, d\nu, & \text{(c)} \\ \dfrac{2}{c}\left[\dfrac{1}{\gamma_0[\Lambda(\infty)]^{\frac{1}{2}}\, d(\nu_0\bar{\nu})} - 1\right]. & \text{(d)} \end{cases} \tag{61}$$

$$j(0) = \begin{cases} 2(1 - c)(\nu_0 - \mu_0)\gamma(\mu_0)/c\gamma_0\, d(\nu_0\bar{\nu}), & \text{(a)} \\ -2\nu_0^2(1 - c)X(-\nu_0)/\gamma_0\, d(\nu_0\bar{\nu}), & \text{(b)} \\ -\dfrac{1 - c}{\gamma_0\, d(\nu_0\bar{\nu})}\left\{2X(-\nu_0)\dfrac{2\varphi_+(-\mu_0)e^{-x_0/\nu_0}}{c\Lambda'(\nu_0)\, d(\nu_0^2)}\right. & \\ \quad \left. + \displaystyle\int_0^1 \frac{c\nu}{2}\frac{\varphi_\nu(-\mu_0)e^{-x_0/\nu}}{\Lambda(\infty)(\nu_0 - \nu)\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu)}\, d\nu\right\}, & \text{(c)} \\ -2(\nu_0 - \bar{\nu})/cd(\nu_0\bar{\nu}). & \text{(d)} \end{cases} \tag{62}$$

Equation (60)(b) was obtained earlier by Shure and Natelson.[7] Equation (60)(a) agrees with the result of Chandrasekhar[10] after the notation is converted.[7]

The result (60)(a) for the albedo problem shows also that $\psi(0, -\mu)/\mu_0$ is a symmetric function of $\mu$ and $\mu_0$, in agreement with the reciprocity theorem.[10] Moreover, a more general form of this theorem[11] leads to the conclusion that $\psi(0, -\mu)$ for the Green's function problem differs from the value of $\psi(x_0, -\mu)/\mu_0$ for the albedo problem only in the interchangement of the variables $\mu$ and $\mu_0$. This is verified by Eq. (60)(c) and the albedo problem results.

[10] S. Chandrasekhar, *Radiative Transfer* (Dover Publications, Inc., New York, 1960).

[11] K. M. Case, Rev. Mod. Phys. **29**, 651 (1957).

## V. THE NONABSORBING MEDIUM

This case is treated separately because for $c = 1$ the two zeros $z = \pm\nu_0$ of $\Lambda(z)$ merge at infinity. When $c \to 1$, we observe that

$$\nu_0 \to [(1-c)(3-b)]^{-1/2}, \quad \gamma_0 \to 1, \quad \gamma_{-1} \to \sqrt{3}. \tag{63}$$

The orthogonality relations of Sec. II must be divided by $\nu_0$ before taking the limit $c \to 1$. With the addition of some auxiliary formulas, the set of needed relations becomes

$$\int_0^1 \varphi_\nu(\mu)\varphi_{\nu'}(\mu)\gamma(\mu)\,d\mu = \gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu)\delta(\nu-\nu'), \tag{64}$$

$$\int_0^1 \varphi_{-\nu}(\mu)\varphi_{\nu'}(\mu)\gamma(\mu)\,d\mu = \tfrac{3}{4}\nu^2\varphi_{\nu'}(-\nu)/\gamma(\nu), \tag{65}$$

$$\int_0^1 \varphi_{-\nu}(\mu)\varphi_{-\nu'}(\mu)\gamma(\mu)\,d\mu = \tfrac{3}{4}\nu'\varphi_\nu(\nu')\left[\frac{\nu'}{\gamma(\nu')} - \frac{\nu}{\gamma(\nu)}\right], \tag{66}$$

$$\int_0^1 \varphi_\nu(\mu)\gamma(\mu)\mu\,d\mu = -\tfrac{1}{2}\nu, \tag{67}$$

$$\int_0^1 \varphi_{-\nu}(\mu)\gamma(\mu)\mu\,d\mu = \tfrac{1}{2}\nu - \tfrac{3}{4}\nu^3/\gamma(\nu), \tag{68}$$

$$\int_0^1 \varphi_\nu(\mu)\gamma(\mu)\,d\mu = 0, \tag{69}$$

$$\int_0^1 \varphi_{-\nu}(\mu)\gamma(\mu)\,d\mu = \tfrac{3}{4}\nu^2/\gamma(\nu), \tag{70}$$

$$\int_0^1 \varphi_\nu(\mu)\gamma(\mu)\frac{d\mu}{\mu} = \tfrac{1}{2}\sqrt{3}, \tag{71}$$

$$\int_0^1 \varphi_{-\nu}(\mu)\gamma(\mu)\frac{d\mu}{\mu} = \tfrac{1}{2}\sqrt{3} - \tfrac{3}{4}\nu/\gamma(\nu). \tag{72}$$

Nothing here depends upon the anisotropy parameter $b$. All the functions involved and all the formulas are the same as for isotropic scattering.

Since the two discrete modes used before have now become identical, we choose a new basis which includes the following two solutions of the homogeneous transport equation:

$$\psi_1(x,\mu) = \tfrac{1}{2}, \quad \psi_2(x,\mu) = \tfrac{3}{2}[(1-\tfrac{1}{3}b)x - \mu]. \tag{73}$$

The expansion (44) is replaced by

$$\psi(x,\mu) = f(x,\mu) + \tfrac{1}{2}a_1 + \int_0^1 A(\nu)\varphi_\nu(\mu)e^{-x/\nu}\,d\nu. \tag{74}$$

For the (a) albedo, (b) Milne, and (c) Green's function problems, $f(x,\mu)$ will be taken as follows:

$$f(x,\mu) = \begin{cases} 0, & \text{(a)} \\ \psi_2(x,\mu), & \text{(b)} \\ G'_\infty(x_0, \mu_0 \to x, \mu). & \text{(c)} \end{cases} \tag{75}$$

By the choice of (75)(b), the solution of the Milne problem has been renormalized to unit net current. The Green's function for the infinite medium is chosen such that its value is finite at $x \to +\infty$, namely

$$G'_\infty(x_0, \mu_0 \to x, \mu) = \begin{cases} \psi_2(x_0, -\mu_0) + \cdots, & x > x_0, \\ \psi_2(x,\mu) + \cdots, & x < x_0, \end{cases} \tag{76}$$

where the dots indicate the same integral as in Eq. (46).

The final results for the three problems then follow through application of the above formulas and are given in the following list:

$$a_1 = \begin{cases} 2\gamma(\mu_0), & \text{(a)} \\ 3\bar{\nu}, & \text{(b)} \\ 3\bar{\nu} - \dfrac{3}{2}\displaystyle\int_0^1 \frac{\nu'\varphi_{-\nu'}(\mu_0)e^{-x_0/\nu'}}{\gamma(\nu')\Lambda^+(\nu')\Lambda^-(\nu')}\,d\nu', & \text{(c)} \end{cases} \tag{77}$$

$$A(\nu) = \begin{cases} \gamma(\mu_0)\varphi_\nu(\mu_0)/\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu), & \text{(a)} \\ -3\nu/4\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu) & \text{(b)} \\ -\dfrac{3}{4\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu)}\left\{\nu\right. & \\ \left. + \displaystyle\int_0^1 \frac{\nu'\varphi_\nu(-\nu')\varphi_{-\nu'}(\mu_0)e^{-x_0/\nu'}}{\gamma(\nu')\Lambda^+(\nu')\Lambda^-(\nu')}\,d\nu'\right\}, & \text{(c)} \end{cases} \tag{78}$$

$$\psi(0,-\mu) = \begin{cases} 4\gamma(\mu_0)\gamma(\mu)\varphi_{-\mu_0}(\mu)/3\mu\mu_0, & \text{(a)} \\ \gamma(\mu)/\mu, & \text{(b)} \\ \dfrac{\gamma(\mu)}{\mu}\left\{1 + \displaystyle\int_0^1 \frac{\varphi_\nu(\mu)\varphi_{-\nu}(\mu_0)e^{-x_0/\nu}}{\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu)}\,d\nu\right\}, & \text{(c)} \end{cases} \tag{79}$$

$$\rho(0) = \begin{cases} (2/\sqrt{3})\gamma(\mu_0)/\mu_0, & \text{(a)} \\ \sqrt{3}, & \text{(b)} \\ \sqrt{3}\left\{1 + \dfrac{1}{2}\displaystyle\int_0^1 \frac{\varphi_{-\nu}(\mu_0)e^{-x_0/\nu}}{\gamma(\nu)\Lambda^+(\nu)\Lambda^-(\nu)}\,d\nu\right\}, & \text{(c)} \end{cases} \tag{80}$$

$$j(0) = \begin{cases} 0, & \text{(a)} \\ -1. & \text{(b),(c)} \end{cases} \tag{81}$$

Again the results (most of them known) are the same as for isotropic scattering. The only term dependent upon $b$ is the $f(x,\mu)$ in the cases (b) and (c).

This agrees with the general observation[7] that $\psi(x, \mu)$ and the corresponding solution for $b = 0$ differ only by a term $\frac{1}{2}bjx$ if the medium is nonabsorbing.

The asymptotic part of the density for the Milne problem is

$$\rho_{as}(x) \equiv (3 - b)x + 3\bar{\nu}, \tag{82}$$

so

$$z_0 = (1 - \tfrac{1}{3}b)^{-1}\bar{\nu}. \tag{83}$$

This result is well-known: $(1 - \frac{1}{3}b)^{-1}$ is the transport mean free path and $\bar{\nu} = 0.710446$.

## ACKNOWLEDGMENTS

We are indebted to Professor G. C. Summerfield and Professor F. C. Shure for helpful discussions.

# An Expansion Method for Treating Singular Perturbation Problems

ALI HASAN NAYFEH
*Heliodyne Corporation, Los Angeles, California*
(Received 20 May 1965)

Cochran's method for treating singular perturbation problems is shown to give, in some cases, expansions which are not uniformly valid. This method is modified and extended to give uniformly valid expansions. The new method is applied to the problem of heat transfer in a duct to give a solution in agreement with that obtained by the WKBJ method.

## 1. INTRODUCTION

MANY physical problems can be formulated mathematically as boundary-value problems consisting of differential equations $L(y; x; \epsilon) = 0$ and boundary conditions $B(x; \epsilon) = 0$ depending on a parameter $\epsilon$. Many of these problems cannot be solved exactly for an arbitrary $\epsilon$, and hence their solutions might be sought by perturbation methods. If there exists an $\epsilon = \epsilon_0$ ($\epsilon$ can be normalized such that $\epsilon_0 = 0$) for which the reduced boundary-value problem $L(y; x; 0) = 0$ and $B(x; 0) = 0$ can be solved exactly or more easily, one attempts to construct the solution of the full problem by simply perturbing the solution of the reduced problem for small $\epsilon$. Unfortunately, many of these straightforward expansions are not uniformly valid because they break down in certain regions called regions of nonuniformity. A problem in which the straightforward perturbation expansion is not uniformly valid is called a singular perturbation problem.

Many methods have been devised for treating such problems. In this paper, Cochran's method[1] is investigated. It is shown that this method gives, in some cases, expansions which are not uniformly valid. This method is modified and extended to give uniformly valid expansions. The new method is applied to the problem of heat transfer in a duct and the result is compared with that obtained by using the WKBJ method.[2]

## 2. COCHRAN'S METHOD

In order to effect a uniformly valid expansion of a boundary value problem $L(y; x; \epsilon) = 0$ and $B(x; \epsilon) = 0$ which has a nonuniformity in the region

$$x - \mu = O[\phi(\epsilon)], \quad \lim_{\epsilon \to 0} \phi(\epsilon) = 0,$$

Cochran transforms the ordinary differential equation into a partial differential equation in terms of two new variables

$$\xi = x, \tag{2.1a}$$

$$\eta = g(x)/\phi(\epsilon), \tag{2.1b}$$

where $g(\mu) = 0$ to reflect the nonuniformity. He hypothesizes that there exists an asymptotic representation of the above problem of the form

$$y(x; \epsilon) = \tilde{y}(\xi, \eta; \epsilon) = \sum_{n=0}^{\infty} \Delta_n(\epsilon) y_n(\xi, \eta) \tag{2.2}$$

where each $y_n(\xi, \eta)$ is bounded independently of $\epsilon$ for all $\xi$ and $\eta$ and

$$\lim_{\epsilon \to 0} \frac{\Delta_n(\epsilon)}{\Delta_{n-1}(\epsilon)} = 0. \tag{2.3}$$

To be able to modify and extend this method, we will apply it to two simple examples.

### Ordinary Differential Equation—Initial-Value Problem

Consider the problem

$$\epsilon^2(F'' + 2F') + F = 0, \tag{2.4a}$$

$$F(0) = 0, \qquad F'(0) = -1, \tag{2.4b}$$

for which the exact solution is

$$F = e^{-x} \cos(\epsilon^{-2} - 1)^{\frac{1}{2}}x. \tag{2.5}$$

The straightforward perturbation expansion of Eq. (2.4a) for $\epsilon \ll 1$ is not valid in the region $x = O(\epsilon)$. Following Cochran, we assume that

$$\tilde{F} = \sum_{n=0}^{\infty} \epsilon^n F_n(\xi, \eta), \tag{2.6}$$

where $\xi = x$, $\eta = [g(x)]/\epsilon$, and each $F_n$ is bounded for all $\xi$, and $\eta$. Substituting the expansion (2.6) in Eq. (2.4a) and equating the coefficients of each power of $\epsilon$ to zero, we obtain

$$g'^2 F_{0\eta\eta} + F_0 = 0, \tag{2.7}$$

$$g'^2 F_{1\eta\eta} + F_1 + 2g'F_{0\xi\eta} + (g'' + 2g')F_{0\eta} = 0, \tag{2.8}$$

[1] J. Cochran, Ph.D. thesis, Stanford University (1962).
[2] A. Erdelyi, *Asymptotic Expansions* (Dover Publications, Inc., 1955), pp. 78–107.

$$g_2' F_{2\eta\eta} + F_2 + 2g'F_{1\xi\eta} + (g'' + 2g')F_{1\eta} + 2F_{0\xi} + F_{0\xi\xi} = 0. \quad (2.9)$$

The general solutions of Eqs. (2.7) and (2.8) are

$$F_0 = A_0(\xi) \cos(\eta/g') + B_0(\xi) \sin(\eta/g') \quad (2.10)$$

$$F_1 = \left\{A_1(\xi) - \left[\left(\frac{A_0}{g'}\right)' + (g'' + g')\frac{A_0}{g'^2}\right]\eta + \frac{B_0 g''}{2g'^3}\eta^2\right\}\cos\left(\frac{\eta}{g'}\right) + \left\{B_1(\xi) - \left[\left(\frac{B_0}{g'}\right)' + (g'' + g')\frac{B_0}{g'^2}\right]\eta - \frac{A_0 g''}{2g'^3}\eta^2\right\}\sin\left(\frac{\eta}{g'}\right). \quad (2.11)$$

The condition that $F_1$ is bounded for all $\eta$ necessitates that the coefficients of $\eta^2$ and $\eta$ should vanish. Since $A_0$ and $B_0 \not\equiv 0$,

$$g'' = 0. \quad (2.12a)$$

Using the fact that $g(0) = 0$, we get

$$g = cx, \quad (2.12b)$$

where $c$ is a constant which can be taken to be unity without loss of generality. Then, the vanishing of the coefficient of $\eta$ gives

$$A_0' + A_0 = 0, \quad (2.13a)$$

$$B_0' + B_0 = 0. \quad (2.13b)$$

Hence,

$$A_0 = a_0 e^{-\xi} \quad (2.14a)$$

$$B_0 = b_0 e^{-\xi} \quad (2.14b)$$

where $a_0$ and $b_0$ are arbitrary constants. Thus,

$$F_0 = e^{-\xi}[a_0 \cos\eta + b_0 \sin\eta], \quad (2.15)$$

$$F_1 = A_1(\xi)\cos\eta + B(\xi)\sin\eta. \quad (2.16)$$

From Eq. (2.9), we find that

$$F_2 = [A_2(\xi) - (A_1' + A_1 + \tfrac{1}{2}b_0 e^{-\xi})\eta]\cos\eta + [B_2(\xi) - (B_1' + B_1 - \tfrac{1}{2}a_0 e^{-\xi})\eta]\sin\eta. \quad (2.17)$$

Since $F_2$ is bounded for all $\eta$,

$$A_1' + A_1 + \tfrac{1}{2}b_0 e^{-\xi} = 0, \quad (2.18a)$$

$$B_1' + B_1 - \tfrac{1}{2}a_0 e^{-\xi} = 0. \quad (2.18b)$$

Therefore,

$$A_1 = (a_1 - \tfrac{1}{2}b_0\xi)e^{-\xi}, \quad (2.19a)$$

$$B_1 = (b_1 + \tfrac{1}{2}a_0\xi)e^{-\xi}, \quad (2.19b)$$

where $a$ and $b$, are arbitrary constants. Thus,

$$F = e^{-x}\left\{\left(a_0 \cos\frac{x}{\epsilon} + b_0 \sin\frac{x}{\epsilon}\right) + \epsilon\left[(a_1 - \tfrac{1}{2}b_0 x)\cos\frac{x}{\epsilon} + (b_1 + \tfrac{1}{2}a_0 x)\sin\frac{x}{\epsilon}\right]\right\}. \quad (2.20)$$

The expansion (2.20) is not uniformly valid because $\epsilon F_1$ and $F_0$ are of the same order in the region $x = O(\epsilon^{-1})$.

**Partial Differential Equation—Boundary-Value Problem**

Consider the problem

$$\epsilon u_{xx} + u_{yy} + u_x = 0, \quad (2.21)$$

$$u(0, y) = f_1(y), \quad (2.22a)$$

$$u(1, y) = f_2(y), \quad (2.22b)$$

$$u(x, 0) = g_1(x), \quad (2.22c)$$

$$u(x, 1) = g_2(x). \quad (2.22d)$$

If $\epsilon \downarrow 0$, Eq. (2.21) changes from an elliptic to a parabolic equation. The reduced equation cannot satisfy all of the boundary conditions and hence it is of the singular perturbation type. The nonuniformity is in the region $x = O(\epsilon)$.

In order to apply Cochran's method, we let

$$u = \tilde{u}(\xi, \zeta, \eta; \epsilon) = \sum_{n=0}^{\infty} \epsilon^n u_n(\xi, \zeta, \eta), \quad (2.23)$$

where

$$\xi = x, \quad (2.24a)$$

$$\zeta = y, \quad (2.24b)$$

$$\eta = g(x, y)/\epsilon, \qquad g(0, y) = 0. \quad (2.24c)$$

Substituting the series (2.23) into Eq. (2.21) and equating each of the coefficients of $\epsilon^{-2}$, $\epsilon^{-1}$, and $\epsilon^0$ to zero, we obtain

$$g_\zeta^2 u_{0\eta\eta} = 0, \quad (2.25)$$

$$g_\zeta^2 u_{0\eta\eta} + g_\xi u_{0\eta} + 2g_\zeta u_{0\zeta\eta} + g_{\zeta\zeta}u_{0\eta} + g_\zeta^2 u_{1\eta\eta} = 0, \quad (2.26)$$

$$g_\zeta^2 u_{1\eta\eta} + g_\xi u_{1\eta} + u_{0\xi} + 2g_\xi u_{0\xi\eta} + g_{\xi\xi}u_{0\eta} + u_{0\zeta\zeta} + 2g_\zeta u_{1\zeta\eta} + g_{\zeta\zeta}u_{1\eta} + g_\zeta^2 u_{2\eta\eta} = 0. \quad (2.27)$$

The solutions of these equations are

$$g = g(\xi), \quad (2.28)$$

$$u_0 = A_0(\xi, \zeta) + B_0(\xi, \zeta)e^{-(\eta/g')}, \quad (2.29)$$

$$u_1 = -(A_{0\xi} + A_{0\xi\xi})\frac{\eta}{g'} + A_1(\xi, \zeta) + \left[-\frac{B_0 g''}{2g'^3}\eta^2 + (B_{0\xi\xi} - B_{0\xi})\frac{\eta}{g'} + B_1(\xi, \zeta)\right]e^{-(\eta/g')}, \tag{2.30}$$

where $A_0$, $B_0$, $A_1$, and $B_1$ are smooth functions of $\xi$ and $\zeta$. The conditions that $u_0$ and $u_1$ are bounded for all $\eta$ necessitate that

$$g' > 0, \tag{2.31a}$$

$$A_{0\xi} + A_{0\xi\xi} = 0. \tag{2.31b}$$

The expansion obtained above still contains two arbitrary functions: $B_0(\xi, \zeta)$ and $g(\xi)$.

### 3. PROPOSED METHOD

As shown in the preceding section, Cochran's method does not give uniformly valid expansions beyond the first term of some singular perturbation problems. Also, the condition that each $y_n(\xi, \eta)$ is bounded for all $\xi$ and $\eta$ is not always sufficient to determine the arbitrary functions which appear in the expansions.

In order to develop a method for treating singular perturbation problems, we note that their uniformly valid asymptotic expansions cannot be expansions in the Poincaré sense[3] because they are expansions in sequences of $\epsilon$, whose coefficients depend upon $\epsilon$ as well as upon $x$. However, as illustrated by the two examples which we have considered, the dependence of the coefficients on $\epsilon$ is in the combination which gives rise to a new variable. Thus, if we introduce a suitable variable $\eta = \eta(x; \epsilon)$ in addition to the variable $\xi = x$, a uniformly valid asymptotic expansion in the Poincaré sense may be found for a problem which has one nonuniformity, but now the coefficients of the sequences are functions of the two independent variables $\xi$ and $\eta$. Hence, the introduction of $\eta$ transforms the problem of obtaining a uniformly valid asymptotic expansion of a function $y(x; \epsilon)$ into finding the straightforward asymptotic expansion of $\tilde{y}(\xi, \eta; \epsilon) = y(x; \epsilon)$.

The method consists of the following four steps:

First, we introduce a new independent variable $\eta$ in addition to $\xi = x$. If the straightforward expansion breaks down in the region $(x - \mu) = O[\phi(\epsilon)]$, where $\mu < \infty$, and $\lim_{\epsilon\to 0} \phi(\epsilon) = 0$,

$$\eta = \eta(x; \epsilon) = \sum_{n=0}^{\infty} \delta_n(\epsilon)\eta_n(x). \tag{3.1}$$

Here $\delta_0(\epsilon) = [1/\phi(\epsilon)]$ and $\eta(\mu) = 0$ in order to reflect the nonuniformity, and $\delta_n(\epsilon)$ for $n \geq 1$ will be determined in the course of analysis. If $\mu$ is infinite, and the nonuniformity is in the region where

$$x\phi(\epsilon) = O(1), \qquad \lim_{\epsilon\to 0} \phi(\epsilon) = 0,$$

we introduce first a new variable $\bar{x}$ such that

$$\bar{x} = x\phi(\epsilon).$$

This transformation moves the nonuniformity to a finite region and can be treated as above.

Second, the differential equation is transformed into a partial differential equation of the two independent variables $\xi$ and $\eta$. The functions of the independent variable $x$ which appear in the original equation are expressed in terms of $\xi$ except those which reflect the nonuniformity.

Third, we assume that there exists an asymptotic representation of $y(x; \epsilon)$ in the form

$$y(x; \epsilon) = \tilde{y}(\xi, \eta; \epsilon) = \sum_{n=0}^{\infty} \Delta_n(\epsilon)y_n(\xi, \eta), \tag{3.2}$$

where

$$\lim_{\epsilon\to 0} [\Delta_n(\epsilon)/\Delta_{n-1}(\epsilon)] = 0, \tag{3.3}$$

$$y_0(\xi, \eta),\ y_n(\xi, \eta)/y_{n-1}(\xi, \eta) < \infty, \tag{3.4}$$

for all $\xi = x$ and $\eta = \eta(x; \epsilon)$ where $x$ is in the domain of interest.[4] The condition (3.4) is the mathematical expression of the fact that the expansion (3.2) is regular in the whole domain of interest.

Fourth, we substitute the expansion (3.2) in the transformed partial differential equation and equate coefficients of each function of $\epsilon$ to zero. We solve the resulting equations. The solution will contain arbitrary functions of the variable $\xi$. They will be determined by imposing the condition (3.4). Now, we apply this new method to the two examples previously discussed.

#### Example 1

To effect a uniformly valid expansion of (2.4) we let

$$\xi = x, \tag{3.5a}$$

$$\eta = \frac{g_0(x)}{\epsilon} + \delta_1(\epsilon)g_1(x) + \cdots, \tag{3.5b}$$

[3] J. G. Van der Corput, *Asymptotic Expansion I. Fundamental Theorems of Asymptotics* (Department of Mathematics, University of California, Berkeley, 1954).

[4] This condition is similar to the elimination of secular terms used by Kryloff and Bogoliuboff in their variation of parameters method for treating nonlinear oscillation problems. See N. Kryloff and N. Bogoliuboff, *Introduction to Nonlinear Mechanics*, (Princeton University Press, Princeton, New Jersey, 1947).

where $\eta(0) = 0$ because the nonuniformity is at the origin and is of order $\epsilon$. Thus Eq. (2.4a) transforms into

$$\epsilon^2(F_{\xi\xi} + 2\eta' F_{\xi\eta} + \eta'' F_\eta + \eta'^2 F_{\eta\eta} + 2F_\xi + 2\eta' F_\eta) + F = 0. \quad (3.6)$$

We assume that

$$F = \sum_{n=0}^{\infty} \epsilon^n F_n(\xi, \eta). \quad (3.7)$$

Since the nonuniformity in Eq. (2.20) appeared in the determination of $A_1(\xi)$ and $B_1(\xi)$ from Eq. (2.17), the function $g_1(x)$ must appear in the latter equation and hence $\delta_1(\epsilon) = \epsilon$. Substituting the series (3.7) in Eq. (3.6) and collecting coefficients of equal powers of $\epsilon$, we get equations for $F_0$, $F_1$, and $F_2$. Solving these equations and using the condition that $[F_n(\xi, \eta)/F_{n-1}(\xi, \eta)] < \infty$ for all $\eta$, we get a result which deviates from Cochran's at Eq. (2.17) because our method allows for the appearance of $g_1'(\xi)$. Thus instead of Eqs. (2.18), we obtain

$$2A_1' + 2A_1 + b_0(1 + 2g_1')e^{-\xi} = 0, \quad (3.8a)$$

$$2B_1' + 2B_1 - a_0(1 + 2g_1')e^{-\xi} = 0. \quad (3.8b)$$

The solutions of Eqs. (3.8) are

$$A_1 = a_1 e^{-\xi} - \tfrac{1}{2}b_0(\xi + 2g_1)e^{-\xi}, \quad (3.9a)$$

$$B_1 = b_1 e^{-\xi} + \tfrac{1}{2}a_0(\xi + 2g_1)e^{-\xi}, \quad (3.9b)$$

where $a_1$ and $b_1$ are arbitrary constants. The condition that $F_1(\xi, \eta)/F_0(\xi, \eta)$ is bounded for all $\xi$ leads to

$$g_1 = -\tfrac{1}{2}\xi, \quad (3.10)$$

because

$$\eta(0) = g_1(0) = 0.$$

Therefore

$$F = e^{-\xi}(a \cos \eta + b \sin \eta) + \cdots, \quad (3.11)$$

where $a$ and $b$ are arbitrary constants and

$$\eta = (x/\epsilon) - \tfrac{1}{2}\epsilon x + \cdots, \quad (3.12a)$$

$$\xi = x. \quad (3.12b)$$

Using the boundary conditions $F(0) = 1$ and $F'(0) = -1$, we get finally

$$F = e^{-x} \cos\left(\frac{x}{\epsilon} - \frac{1}{2}\epsilon x\right) + O(\epsilon^3), \quad (3.13)$$

because

$$\frac{1}{\epsilon^3}\left|e^{-x}\left[\cos\left(\frac{1}{\epsilon^2} - 1\right)^{\frac{1}{2}}x - \cos\left(\frac{x}{\epsilon} - \frac{1}{2}\epsilon x\right)\right]\right| \leq M, \quad (3.14)$$

where $M$ is independent of $\epsilon$.

**Example 2**

The application of the new method to this example determines the functions $B_0(\xi, \zeta)$ and $g(\xi)$. From Eqs. (2.29 and 2.30), $u_1(\xi, \zeta, \eta)/u_0(\xi, \zeta, \eta)$ is bounded for all $\eta = [g(x)]/\epsilon$, $0 \leq x \leq 1$ if

$$g'' = 0, \quad (3.15)$$

$$B_{0\xi\zeta} - B_{0\xi} = 0. \quad (3.16)$$

Since $g(0) = 0$ and $g' > 0$, the solution of Eq. (3.15) is

$$g = c\xi, \quad (3.17)$$

where $c$ is an arbitrary positive constant which can be taken to be unity without loss of generality. Therefore, to first order

$$u = A_0(x, y) + B_0(x, y)e^{-x/\epsilon}, \quad (3.18)$$

where

$$\begin{aligned} &A_{0x} + A_{0yy} = 0, && B_{0x} - B_{0yy} = 0, \\ &A_0(1, y) = f_2(y), && B_0(x, 0) = 0, \\ &A_0(x, 0) = g_1(x), && B_0(x, 1) = 0, \\ &A_0(x, 1) = g_2(x), && B_0(0, y) = f_1(y) - A_0(0, y). \end{aligned} \quad (3.19)$$

## 4. HEAT TRANSFER IN A DUCT

Consider the heat transfer in a fluid with constant density which is in steady laminar or in steady mean turbulent motion in a flat duct. The velocity profile is fully developed and the temperature is uniform when it suddenly enters a flat duct whose wall is kept at a different constant temperature. The mathematical description of the problem in non-dimensional quantities is

$$v(x)\frac{\partial T}{\partial z} = \frac{\partial}{\partial x}\left[p(x)\frac{\partial T}{\partial x}\right], \quad (4.1a)$$

$$p(x) = (1 + \epsilon_H \Pr/\nu) > 0, \quad (4.1b)$$

$$T(z, x) = 1 \quad \text{for} \quad z < 0, \quad (4.2a)$$

$$T(z, \pm 1) = 0 \quad \text{for} \quad z > 0. \quad (4.2b)$$

Here, $T$ is the temperature, $v$ is the velocity, $\nu$ is the kinematic viscosity, Pr is the Prandtl number, $\epsilon_H$ is the eddy diffusivity for heat, $z$ is the longitudinal coordinate, and $x$ is the normal coordinate. Letting $T(z,x) = y(x)Z(z)$ and separating variables, we get the following eigenvalue problem

$$[p(x)y']' + \lambda^2 v(x)y = 0, \quad (4.3a)$$

$$y(1) = y(-1) = 0. \quad (4.3b)$$

Since $v(x)$ vanishes at the wall, it has the form

$$v(x) = (1 - x^2)f(x), \qquad f(x) > 0. \quad (4.3c)$$

The eigenvalue problem obtained above is called the Graetz problem.[5,6] Although we derived the Graetz problem for turbulent flow, the laminar problem can be obtained from (4.3) by letting $f(x) = p(x) = 1$. The problem considered here is the determination of the first-order asymptotic expansions of the eigenfunctions and eigenvalues for $\lambda \gg 1$ in (4.3) by using the method proposed in this paper. The laminar case has been investigated by Sellars, Tribus, and Klein[6] by using the WKBJ method. Since the problem is symmetric, it suffices to determine the solution in [0, 1] and the boundary condition $y(-1) = 0$ can be replaced by $y'(0) = 0$.

The straightforward perturbation expansion which can be obtained by a Liouville's transformation[2] is not valid in the neighborhoods of the zeros of $v(x)$. The problem considered has a turning point at $x = 1$. In order to determine the extent of the region of nonuniformity, we let $(1 - x)\lambda^{\alpha} = z$ where $z = O(1)$ in Eq. (4.1a), and choose $\alpha$ such that the first term is of the same order as the second. This is so if $\alpha = (\frac{2}{3})$.

In order to effect a first-order uniformly valid expansion in [0, 1] using the proposed method, we let

$$y = y_0(\xi, \eta) + \lambda^{-2/3}y_1(\xi, \eta) + \cdots, \quad (4.4a)$$

$$\xi = x, \quad (4.4b)$$

$$\eta = \lambda^{2/3}h(x), \quad (4.4c)$$

where

$$h(x) = (1 - x)g(x), \quad g(x) > 0. \quad (4.4d)$$

Equation (4.1a) is thus transformed into a partial differential equation in terms of $\xi$ and $\eta$. The functions of the independent variable $x$ which appear in Eq. (4.1a) are expressed in terms of $\xi$ except $(1 - x)$ because it reflects the nonuniformity and is replaced by $[\eta\lambda^{-2/3}]/[g(x)]$. Substituting the series (4.4a) in Eq. (4.1a) and equating each of the coefficients of $\lambda^{4/3}$ and $\lambda^{2/3}$ to zero, we obtain

$$L(y_0) = h'^2py_{0\eta\eta} + \frac{(1 + \xi)f}{g}\eta y_0 = 0, \quad (4.5)$$

$$L(y_1) = -(ph'' + p'h')y_{0\eta} - 2h'py_{0\xi\eta}. \quad (4.6)$$

The solution of Eq. (4.5) is

$$y_0 = A_0(\xi)\eta^{\frac{1}{2}}J_{\frac{1}{3}}[\gamma(\xi)\eta^{\frac{3}{2}}] + B_0(\xi)\eta^{\frac{1}{2}}J_{-\frac{1}{3}}[\gamma(\xi)\eta^{\frac{3}{2}}], \quad (4.7a)$$

where

$$\gamma = \pm\frac{2}{3}\left[\frac{(1 + \xi)f}{gph'^2}\right]^{\frac{1}{2}}. \quad (4.7b)$$

Substituting for $y_0$ in Eq. (4.6), we get

$$L(y_1) = -\frac{\partial}{\partial\eta}\{\eta^{\frac{1}{2}}[2A_0'h'p + (ph'' + p'h')A_0]J_{\frac{1}{3}} + \eta^{\frac{1}{2}}[2B_0'h'p + (ph'' + p'h')B_0]J_{-\frac{1}{3}}\} + \gamma'\frac{\partial}{\partial\eta}[2h'p\eta^2(A_0J_{\frac{1}{3}}' + B_0J_{-\frac{1}{3}}')]. \quad (4.8)$$

For $y_1/y_0$ to be bounded for all $\eta(x; \lambda)$, $0 \leq x \leq 1$, the inhomogeneous part in Eq. (4.8) must vanish. Thus,

$$\gamma' = 0, \quad (4.9a)$$

$$2h'pA_0' + (ph'' + p'h')A_0 = 0, \quad (4.9b)$$

$$2h'pB_0' + (ph'' + p'h')B_0 = 0. \quad (4.9c)$$

The solutions of these equations are

$$\gamma = c, \quad (4.10a)$$

$$A_0 = a_0/(h'p)^{\frac{1}{2}}, \quad (4.10b)$$

$$B_0 = b_0/(h'p)^{\frac{1}{2}}, \quad (4.10c)$$

where $a_0$, $b_0$, and $c$ are arbitrary constants. The constant $c$ can be taken to be unity without loss of generality. Furthermore, the negative sign must be taken in Eq. (4.7b) so that $g(x) > 0$. Hence, the solution of Eq. (4.7b) is

$$h^{\frac{3}{2}} = -\int_1^x \left[\frac{(1 - t^2)f(t)}{p(t)}\right]^{\frac{1}{2}} dt = u(x). \quad (4.11)$$

Thus,

$$y = \frac{u^{\frac{1}{2}}}{[(1 - x^2)fp]^{\frac{1}{4}}}[aJ_{\frac{1}{3}}(\lambda u) + bJ_{-\frac{1}{3}}(\lambda u)], \quad (4.12)$$

where $a$ and $b$ are arbitrary constants.

Imposing the boundary condition $y(1) = 0$, we find that $b = 0$, and hence,

$$y = \frac{au^{\frac{1}{2}}}{[(1 - x^2)fp]^{\frac{1}{4}}}J_{\frac{1}{3}}(\lambda u). \quad (4.13)$$

Near $x = 0$, $\lambda u$ is large and hence $y$ can be represented asymptotically as

$$y = \left(\frac{2}{\lambda}\right)^{\frac{1}{2}}a[(1 - x^2)fp]^{-\frac{1}{4}}\cos(\lambda u - \tfrac{5}{12}\pi). \quad (4.14)$$

Since $y'(0) = 0$, $f(x)$ and $p(x)$ are even,

$$\sin[\lambda u(0) - \tfrac{5}{12}\pi] = 0, \quad (4.15a)$$

[5] M. Jakob, *Heat Transfer* (John Wiley & Sons, Inc., New York, 1949), Vol. 1, pp. 451–480.

[6] J. R. Sellars, M. Tribus, and J. S. Klein, Trans. ASME **78**, 441 (1956).

or

$$\lambda_n = (n + \tfrac{5}{12})\pi/u(0), \quad n = 0, 1, 2, \cdots . \tag{4.15b}$$

For the laminar case,

$$u(0) = -\int_1^0 (1 - t^2)^{\frac{1}{2}}\, dt = \tfrac{1}{4}\pi. \tag{4.16a}$$

Hence,

$$\lambda_n = 4n + \tfrac{5}{3}, \quad n = 0, 1, 2, \cdots . \tag{4.16b}$$

Strictly speaking, the eigenvalues in (4.16b) are only valid for large $n$ because the entire expansion is obtained for $\lambda \gg 1$. The eigenvalues (4.16b) agree with those obtained by using the WKBJ method[6] ($A - 25$). On the other hand, our eigenfunctions are given by a single uniformly valid expansion in [0, 1], while those obtained by the WKBJ method are given by two expansions—one is valid in the middle of the duct and the other is valid near the wall. However, our eigenfunctions reduce to those obtained by the WKBJ method near the wall and in the middle of the duct, respectively.

## 5. CONCLUSION

Cochran's method for treating singular perturbation problems yields expansions which may contain arbitrary functions. The application of this method can give expansions which are not uniformly valid. This method is modified and extended to give uniformly valid expansions. The new method is applied to the Graetz problem which arises in the problem of heat transfer in a duct. The eigenvalues thus obtained agree with those obtained using the WKBJ method.

Since the major criterion we used in determining the arbitrary functions which appeared in our expansions is $[y_n(\xi, \eta)/y_{n-1}(\xi, \eta)] < \infty$ for all $\xi$ and $\eta(x; \epsilon)$ such that $x$ is in the domain of interest, the proposed expansion method might break down if the equation admits solutions which are unbounded in the domain of interest.[7] For example, the proposed method is inapplicable to the problem of determining the asymptotic representation of the solution bounded at the origin for $\lambda \gg 1$ in

$$xy'' + y' + \lambda^2 x(1 - x^2)y = 0.$$

Thus, this method has to be extended or modified further in order to apply to such problems.

## ACKNOWLEDGMENT

The author acknowledges the stimulating discussions of Dr. Milton Van Dyke, Dr. Gordon Latta, and Dr. I-Dee Chang. This study originated at the Department of Aeronautics and Astronautics, Stanford University, and was partially supported by the Air Force Office of Scientific Research under contract No. AF 49(638)-1274.

[7] A. H. Nayfeh, Ph.D thesis, Stanford University (1964)

# Generalized Solutions for Massless Free Fields and Consequent Generalized Conservation Laws

R. F. O'Connell and D. R. Tompkins
*Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana*
(Received 4 March 1965)

Generalized solutions to the equations for a massless free field with arbitrary spin are written down. It is shown that they lead immediately to generalizations of all the usual conservation laws.

## INTRODUCTION

NEW conservation laws for the electromagnetic field have been discovered recently by Lipkin.[1] Morgan[2] has shown Lipkin's conserved third-rank tensor is a special case of a generalized energy–momentum tensor density. The procedure adopted by Morgan was to construct a generalized tensor by analogy with the usual expression for the conventional $T_{\mu\nu}$. He then showed that the generalized tensor was divergence-free. We will show instead that a possibly more basic concept is the consideration of generalized solutions to the usual field equations. This immediately leads to generalized conservation laws. We will then demonstrate how the conserved quantities of Lipkin and Morgan may be simply obtained as particular cases of our general formalism. Furthermore, in addition to the conserved energy–momentum tensor and current displayed by Morgan, we will show that other generalized conserved quantities may be obtained because our procedure admits a generalization of all the usual conserved quantities.

## GENERALIZED SOLUTIONS AND GENERALIZED CONSERVATION LAWS

The equation of a massless free field may be written down in the formalism of Pauli–Fierz[3] or Hammer–Good.[4] For our purpose it is more convenient to use the method of the latter authors who show that the wave equation may be written in the form[5]

$$(S_k/S)\frac{\partial\psi(x)}{\partial x_k}+\frac{\partial\psi(x)}{\partial t}=0, \tag{1}$$

where $S$ is the angular momentum matrix for arbitrary spin $S = \frac{1}{2}, 1, \frac{3}{2} \cdots$ and $\psi$ is a $(2S+1)$ component spinor. As an auxiliary condition, the only solutions retained are those which satisfy the Bargmann–Wigner[6] criteria of having their spins parallel or antiparallel to the momentum. Good,[7] for example, explicitly shows that Maxwell's equations may be written in this form.

Consider any general operator $V$ so selected that $V\psi(x)$ is a generalized solution of Eq. (1). Thus, we can write

$$(S_k/S)\frac{\partial\psi'(x)}{\partial x_k}+\frac{\partial\psi'(x)}{\partial t}=0 \tag{2}$$

and

$$(S_k/S)\frac{\partial\psi''(x)}{\partial x_k}+\frac{\partial\psi''(x)}{\partial t}=0, \tag{3}$$

where

$$\psi'=V'\psi \quad \text{and} \quad \psi''=V''\psi, \tag{4}$$

and where $V'$ and $V''$ are so selected that $\psi'$ and $\psi''$ satisfy Eq. (1).

Good[7] has shown how all the usual conservation laws can be derived quite simply from Eq. (1) without using Lagrangian's or Noether's Theorem. He considers infinitesimal transformations of coordinates[8]

$$x_\mu \rightarrow x_\mu''' = x_\mu + \delta x_\mu,$$

and the corresponding transformations of the wavefunction

$$\psi(x) \rightarrow \psi'''(x''').$$

He then introduces a general operator $O$ defined by

$$\psi'''(x)=O\psi(x).$$

This enables him to deduce the conservation laws

[1] D. M. Lipkin, J. Math. Phys. **5**, 696 (1964).
[2] T. A. Morgan, J. Math. Phys. **5**, 1659 (1964).
[3] M. Fierz, Helv. Phys. Acta **12**, 3 (1939); M. Fierz and W. Pauli, Proc. Roy. Soc. (London) **A173**, 211 (1939).
[4] C. L. Hammer and R. H. Good, Jr., Phys. Rev. **108**, 882 (1957).
[5] In our units $\hbar = c = 1$. Furthermore, Greek indices run from 1–4 and Latin indices from 1–3.
[6] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. US **34**, 211 (1948).
[7] R. H. Good, Jr., Phys. Rev. **105**, 1914 (1957).
[8] In order to avoid confusion with our already defined single-prime quantities, we use triple primes here instead of the single primes used by Good.

$$\frac{\partial}{\partial x_k}(\psi^+(S_k/S)O\psi) + \frac{\partial}{\partial t}(\psi^+O\psi) = 0. \quad (5)$$

Corresponding to the various coordinate transformations of displacement, rotation, etc., Good obtains appropriate expressions for $O$. The various operators $O$ when substituted into Eq. (5) give rise to the usual conservation laws of energy–momentum, angular momentum, and so on.

Instead of considering Eq. (1) as Good did, we will consider Eqs. (2) and (3). This enables us to deduce a generalization of Good's conservation laws. We obtain

$$\frac{\partial}{\partial x_k}(\psi'^+(S_k/S)O\psi'') + \frac{\partial}{\partial t}(\psi'^+O\psi'') = 0, \quad (6)$$

which constitute our generalized conservation laws. We notice that the usual conservation laws contain only $\psi$; whereas our generalized conservation laws contain both $\psi'$ and $\psi''$.

As an example, we will discuss in detail the generalization of the energy–momentum tensor $T_{\mu\nu}$ for the electro-magnetic field. The generalization consists of replacing $E$ and $H$ by $E'('')$ and $H'('')$. Thus, Maxwell's equations for the free electromagnetic field in vacuum may be written

$$\epsilon_{ijk}\frac{\partial}{\partial x_j}E'_k = -\frac{\partial}{\partial t}H'_i, \quad (7a)$$

$$\epsilon_{ijk}\frac{\partial}{\partial x_j}H'_k = \frac{\partial}{\partial t}E'_i, \quad (7b)$$

$$(\partial/\partial x_i)E'_i = 0, \quad (7c)$$

$$(\partial/\partial x_i)H'_i = 0, \quad (7d)$$

and we have a similar set with the primes replaced by double primes. Associated with $\psi'$ is a generalized antisymmetric field tensor $F'_{\mu\nu}$ and a generalized potential $A'_\mu$ defined by

$$\psi'_k = i(-\delta_{4\mu}\delta_{k\nu} + \tfrac{1}{2}\epsilon_{k\mu\nu})F'_{\mu\nu}, \quad (8a)$$

and

$$F'_{\mu\nu} = \frac{\partial A'_\nu}{\partial x_\mu} - \frac{\partial A'_\mu}{\partial x_\nu}, \quad (8b)$$

and we have similar quantities with the primes replaced by double primes. This immediately enables us to construct[9] a conserved energy–momentum tensor $\tilde{T}_{\mu\nu}$ given by

$$\tilde{T}_{\mu\nu} = -\tfrac{1}{2}(F'_{\mu\gamma}F''_{\mu\nu} + F'^*_{\mu\gamma}F''^*_{\gamma\nu}), \quad (9)$$

where $F'^*_{\mu\nu}$ is the dual of $F'_{\mu\nu}$. In the particular case of

$$V' = \partial_{\alpha_1}\partial_{\alpha_2}\cdots\partial_{\alpha_n}; \quad V'' = \partial_{\beta_1}\partial_{\beta_2}\cdots\partial_{\beta_m}, \quad (10)$$

we see that $\tilde{T}_{\mu\nu}$ is identical with Morgan's conserved tensor $T_{\mu\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m}$. We wish to emphasize two points. First of all, Morgan's tensor is a particular case of our general $\tilde{T}_{\mu\nu}$, as given by Eq. (9). Secondly, Morgan proceeded by initially postulating a $T_{\mu\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m}$ and then showing that it is conserved while our approach emphasizes the basic concept of generalized solutions to the field equations. From this concept a conserved $\tilde{T}_{\mu\nu}$ follows immediately in a similar manner to the derivation of the usual $T_{\mu\nu}$ from the field equations with the usual solutions.

Morgan has shown that Lipkin's third-order tensor is a particular case of $T_{\mu\nu\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m}$, but it will be instructive to derive Lipkin's results explicitly by our method. To do this it will be convenient to rewrite Eqs. (7) in vector notation,

$$\nabla\times\mathbf{E}' = -\partial\mathbf{H}'/\partial t, \quad (11a)$$

$$\nabla\times\mathbf{H}' = \partial\mathbf{E}'/\partial t, \quad (11b)$$

$$\nabla\cdot\mathbf{E}' = 0, \quad (11c)$$

$$\nabla\cdot\mathbf{H}' = 0. \quad (11d)$$

In the usual case we derive a conserved tensor $T_{\mu\nu}$ which, when written in component form, expresses the laws of conservation of energy density and momentum density. These laws are more readily identifiable when the components of $T_{\mu\nu}$ are expressed in terms of field quantities. The generalizations of these equations are easily shown to be

$$(\partial/\partial t)\{\mathbf{E}'\cdot\mathbf{E}'' + \mathbf{H}'\cdot\mathbf{H}''\} + \nabla\cdot\{(\mathbf{E}'\times\mathbf{H}'') + (\mathbf{E}''\times\mathbf{H}')\} = 0, \quad (12)$$

and

$$(\partial/\partial t)\{(\mathbf{E}'\times\mathbf{H}'') + (\mathbf{E}''\times\mathbf{H}')\}_i + \partial/\partial x_j\{(\mathbf{E}'\cdot\mathbf{E}'' + \mathbf{H}'\cdot\mathbf{H}'')\,\delta_{ij} - (E'_iE''_j + H'_iH''_j)\} = 0. \quad (13)$$

In the process of deriving Eq. (12), if we display components of the fields, we can obtain the result

$$(\partial/\partial t)\{E'_iE''_i + H'_iH''_i\} = \{[E'_i(\nabla\times\mathbf{H}'')_i - H'_i(\nabla\times\mathbf{E}'')_i] + [E''_i(\nabla\times\mathbf{H}')_i - H''_i(\nabla\times\mathbf{E}')_i]\}. \quad (14)$$

We obtain two further equations from Eq. (14) by first replacing $E'$ and $H'$ by $(\nabla\times\mathbf{E}')$ and $(\nabla\times\mathbf{H}')$,

[9] For details of this construction, see for example, J. L. Synge, *Relativity: The Special Theory* (North-Holland Publishing Company, Amsterdam, 1956), p. 323.

respectively, and second by replacing $E''$ and $H''$ by $(\nabla \times \mathbf{E}'')$ and $(\nabla \times \mathbf{H}'')$, respectively. Adding these two equations together, and using the wave equation, leads to the conservation law:

$$(\partial/\partial t)\{([(\nabla \times \mathbf{E}')_i E_i'' + (\nabla \times \mathbf{H}')_i H_i''] + [E_i'(\nabla \times \mathbf{E}'')_i + H_i'(\nabla \times \mathbf{H}'')_i]) + (i \leftrightarrow j)\} + (\partial/\partial x_k)\{[(E_i' H_{i,k}'' - H_i' E_{i,k}'') + (E_i'' H_{i,k}' - H_i'' E_{i,k}')] + (i \leftrightarrow j)\} = 0. \quad (15)$$

We now set

$$E' = (\nabla \times \nabla \cdots \nabla \times)_{n \text{ times}} \mathbf{E}, \quad (16a)$$

and

$$H' = (\nabla \times \nabla \cdots \nabla \times)_{n \text{ times}} \mathbf{H}, \quad (16b)$$

and, in a similar manner, define $E''$ and $H''$ which are the same as the single primed quantities except that an index $m$ replaces $n$.

We will now show that Lipkin's results are particular cases of the above. For ease in comparison we use Lipkin's definition of $Z^{\mu\nu\rho}$ [as given by his Eqs. (17)–(22)].

Setting $n = 0$ and $m = 1$ in Eq. (12), we obtain

$$(\partial/\partial t)Z^{000} + (\partial/\partial x_k)Z^{00k} = 0; \quad (17)$$

setting $n = 0$ and $m = 1$ in Eq. (13), we obtain

$$(\partial/\partial t)Z^{0j0} + (\partial/\partial x_k)Z^{0jk} = 0; \quad (18)$$

setting $n = m = 0$ in Eq. (15), and making use of Eq. (17), we obtain

$$(\partial/\partial t)Z^{ij0} + (\partial/\partial x_k)Z^{ijk} = 0; \quad (19)$$

Eqs. (17)–(19) constitute all of Lipkins conserved quantities.

In summary, we have shown that generalized solutions of the usual wave equations are possibly a more fundamental concept than generalized conservation laws and that the latter readily follow once we have established the existence of the former.

*Note Added in Proof.* In subsequent publications[10,11] we demonstrated the existence of generalized conservation laws for free fields with mass, and examined the physical interpretation of these conservation laws.

[10] R. F. O'Connell and D. R. Tompkins, Nuovo Cimento **38**, 1088 (1963).
[11] R. F. O'Connell and D. R. Tompkins, Nuovo Cimento **39**, 391 (1965).

# Reduction of a One-Loop Feynman Diagram with $n$ Vertices in $m$-Dimensional Lorentz Space

BENGT PETERSSON
*Department of Theoretical Physics, University of Lund, Lund, Sweden*
(Received 26 April 1965)

An explicit formula is given for the evaluation of a diagram with one loop and $n$ vertices in $m$-dimensional Lorentz space ($n > m$). The result is given as a sum of terms each corresponding to a one-loop diagram with $m$ vertices and with a coefficient which can be obtained from rules analogous to the rules of residue calculus.

## INTRODUCTION

THE standard method of evaluating the contributions from Feynman diagrams is to introduce the so-called "Feynman auxilary variables." For a diagram with a large number of lines this method gives rise to rather extensive integrations. In fact, if $n > m$, we increase the number of integrations by using this method.

In this paper we study those diagrams consisting of a single loop with an arbitrary number of vertices. Such a diagram gives rise to an integral of the form

$$F(z_{ik}, a_k) = \int \left[ dq \Big/ \prod_{k=1}^{n} [(q - p_k)^2 + a_k - i\epsilon] \right], \quad (1)$$

$$z_{ik} = -(p_i - p_k)^2 = (p_{i0} - p_{k0})^2 - (\bar{p}_i - \bar{p}_k)^2. \quad (2)$$

Here, the differences of the vectors $p_i (j = 1, \cdots, n)$ are the external energy momentum vectors of the diagram and the $a_i$ denote the squares of the masses of the particles corresponding to the internal lines of the polygon.

In two-dimensional Lorentz space, Källén and Toll[1] have, by explicit integration of (1), obtained the following reduction formula:

$$F(z_{ik}, a_k) = \int \left[ d^2q \Big/ \prod_{k=1}^{n} [(q - p_k)^2 + a_k - i\epsilon] \right]$$
$$= \frac{1}{2} \sum_{i<k} \left\{ 1 \Big/ \prod_{\substack{l=1 \\ l \neq i,k}}^{n} [(q_{ik}^{+} - p_l)^2 + a_l] \right.$$
$$\left. + 1 \Big/ \prod_{\substack{l=1 \\ l \neq i,k}}^{n} [(q_{ik}^{-} - p_l)^2 + a_l] \right\}$$
$$\times \int \frac{d^2q}{[(q - p_i)^2 + a_i - i\epsilon][(q - p_k)^2 + a_k - i\epsilon]}, \quad (3)$$

where $q_{ik}^{\pm}$ are the two vectors which satisfy

$$(q_{ik} - p_i)^2 + a_i = 0,$$
$$(q_{ik} - p_k)^2 + a_k = 0. \quad (4)$$

The integrals in the sum are characteristic of a "bubble" diagram with two external lines. In $m$ dimensions with $m > 2$ [$(m - 1)$ space- and 1 time-dimension], an explicit evaluation of (1) becomes rather complicated. However, Halpern[2] has given a reduction formula for a diagram with $m + 1$ lines, and Brown[3] has shown how to reduce a diagram with $n$ lines to a sum of diagrams with $n - 1$ lines if $n \geq m + 2$. (Here and below "diagram" means "one-loop diagram.") Using their results, we show that the following simple generalization of formula (2) is valid in $m$ dimensions. Let $P$ be a set of $m$ vectors $p_l$ such that the Gram determinants of their differences are different from zero. It will be shown that there are exactly two vectors $q_P^{\pm}$ which satisfy

$$(q_P - p_l)^2 + a_l = 0 \quad (m \text{ equations}). \quad (5)$$

For convenience we define

$$[i(q)] = [(q - p_i)^2 + a_i]. \quad (6)$$

Then we have the following reduction formula:

$$\int \frac{d^mq}{\prod_{k=1}^{n} [k(q) - i\epsilon]} = \frac{1}{2} \sum_{P} \left\{ \frac{1}{\prod_{l \notin P} [l(q_P^{+})]} \right.$$
$$\left. + \frac{1}{\prod_{l \notin P} [l(q_P^{-})]} \right\} \times \int \frac{d^mq}{\prod_{k \in P} [k(q) - i\epsilon]}. \quad (7)$$

The integrals in the sum correspond to $m$-dimensional diagrams with $m$ lines. We remark that the diagram with four lines in four dimensions has been explicitly evaluated by Wu[4].

[1] G. Källén and J. Toll, J. Math. Phys. **6**, 299 (1965).
[2] F. R. Halpern, Phys. Rev. Letters **10**, 310 (1963).
[3] L. M. Brown, Nuovo Cimento **22**, 178 (1961).
[4] A. C. Wu, Kgl. Danske Videnskab. Selskab, Mat.—Fys. Medd. **33**, No. 3 (1961).

## I. PROOF FOR THE CASE $n = m + 1$

### A. Reduction Procedure

As was mentioned above, Halpern has given a formula for the reduction of the $m$-dimensional diagram with $m + 1$ lines to a sum of diagrams with $m$ lines. For the convenience of the reader, a summary of his proof is given here. He uses the representation of the diagram as an integral over the Feynman variables. We give the well-known transformation formula here, to fix our conventions.

$$\int \left[ d^m q \Big/ \prod_{k=1}^{n} [(q - p_k)^2 + a_k - i\epsilon] \right] = (-1)^{-s} \pi^{m/2} \times \Gamma(s) i \int_0^1 \cdots \int_0^1 d\alpha_1 \cdots d\alpha_n \frac{\delta(1 - \sum \alpha_i)}{D^s}, \quad (8)$$

where $s = n - \frac{1}{2}m$, and $D = \sum_{i<k} z_{ik}\alpha_i\alpha_k - \sum a_k\alpha_k$. The condition $\sum \alpha_i = 1$ allows us to write $D$ as a homogeneous polynomial of second degree in $\alpha$

$$D = \tfrac{1}{2} \sum_{j,k} \psi_{jk}\alpha_j\alpha_k, \quad (9)$$

$$\psi_{jk} = z_{jk} - a_j - a_k, \; j \neq k; \quad \psi_{jj} = -2a_j.$$

For the reduction, Halpern[5] uses a generalized version of Stokes's theorem which, in our case, reads

$$\int_0^1 \cdots \int_0^1 d\alpha_1 \cdots d\alpha_n \, \delta(1 - \sum_j \alpha_j) \sum_{i=1}^{n} \frac{\partial A_i}{\partial \alpha_i} = -\sum_{i=1}^{n} \int_0^1 \cdots \int_0^1 d\alpha_1 \cdots d\alpha_n \times \delta(1 - \sum_j \alpha_j) A_i \delta(\alpha_i), \quad (10)$$

where $A_i$ are $n$ quantities which are functions of $\alpha$ and fulfill the condition

$$\sum_{i=1}^{n} A_i = 0. \quad (11)$$

To be able to apply Stokes's theorem to our problem, we must find quantities $A_i$ such that

$$\text{(a)} \quad \sum_{i=1}^{n} \frac{\partial A_i}{\partial \alpha_i} = \frac{1}{D^s}; \qquad \text{(b)} \quad \sum_{i=1}^{n} A_i = 0. \quad (12)$$

The construction of these $A_i$ proceeds in two steps. First, quantities $\Lambda_i$ are defined which fulfil condition (a), and afterwards we derive from these quantities $A_i$ which also satisfy condition (b).

[5] F. R. Halpern and W. Wilson, "Poles in Feynman Diagrams with Several Loops" (preprint). In Appendix I of this paper there is a proof of Stokes's theorem as it is stated here in formula (10).

We have the linear relations

$$\sum_{i=1}^{n} \alpha_i = 1, \quad (13)$$

$$\sum_{i=1}^{n} \psi_{ij}\alpha_i = \frac{\partial D}{\partial \alpha_j}, \qquad j = 1, 2 \cdots n. \quad (14)$$

As it is obvious that nontrivial solutions $\alpha_i$ of Eqs. (13) and (14) exist, the determinant

$$\Delta = \begin{vmatrix} 1 & 1 & \cdots & 1 & 1 \\ \psi_{11} & \psi_{12} & \cdots & \psi_{1n} & \partial D/\partial \alpha_1 \\ \vdots & \vdots & & \vdots & \vdots \\ \psi_{1n} & \psi_{2n} & \cdots & \psi_{nn} & \partial D/\partial \alpha_n \end{vmatrix} \quad (15)$$

is equal to zero. The determinant $\Delta$ can be expanded in the form

$$\Delta = \sum_{i=1}^{n} P_i \frac{\partial D}{\partial \alpha_i} - \psi = 0, \quad (16)$$

where $P_i$ and $\psi$ are minors corresponding to the last column in $\Delta$. If one defines

$$\Lambda_i = P_i/(1 - s)\psi D^{s-1}, \quad (17)$$

one gets

$$\sum_{i=1}^{n} \frac{\partial \Lambda_i}{\partial \alpha_i} = \frac{1}{D^s} \quad \text{(note that } P_i \text{ and } \psi \text{ are independent of } \alpha\text{)}. \quad (18)$$

However, the sum $\sum_{i=1}^{n} \Lambda_i$ is not equal to zero in general and, therefore, the $\Lambda_i$ cannot replace the $A_i$ in Eq. (10). A satisfactory choice of $A_i$ is

$$A_i = \Lambda_i \sum_{j=1}^{n} \alpha_j - \alpha_i \sum_{j=1}^{n} \Lambda_j. \quad (19)$$

Then, we have, trivially,

$$\sum_{i=1}^{n} A_i = 0. \quad (20)$$

If we are careful not to put $\sum \alpha_i = 1$ before differentiation and use the fact that $D$ is homogeneous in $\alpha_i$ to apply Euler's theorem, we get

$$\sum_{i=1}^{n} \frac{\partial A_i}{\partial \alpha_i} = \frac{1}{D^s} + [2s - n - 1] \frac{\sum P_j}{(1 - s)\psi D^{s-1}}. \quad (21)$$

If $n = m + 1$, we have $2s - n - 1 = 0$, so that, in this case, one gets the result

$$\int_0^1 \cdots \int_0^1 d\alpha_1 \cdots d\alpha_{m+1} \frac{\delta(1 - \sum_j \alpha_j)}{D^{1+m/2}} = \sum_{i=1}^{m+1} \frac{2P_i}{m\psi} \int_0^1 \cdots \int_0^1 d\alpha_1 \cdots d\alpha_{m+1} \times \frac{\delta(\alpha_i)\delta(1 - \sum_j \alpha_j)}{D^{m/2}}. \quad (22)$$

In terms of the original integrals, we get

$$\int \frac{d^m q}{\prod_{k=1}^{m+1}[(q-p_k)^2+a_k-i\epsilon]} = -\sum_{i=1}^{m+1}\frac{P_i}{\psi}\int\left[d^m q[(q-p_i)^2 + a_i - i\epsilon]\Big/\prod_{k=1}^{m+1}[(q-p_k)^2+a_k-i\epsilon]\right]. \quad (23)$$

We close this section with a few remarks about the algebraic structure of $P_i$ and $\psi$. If we subtract the first column from all other columns and the first row from all other rows, $\psi$ can be written in the form

$$\psi = \tilde{\psi} = \begin{vmatrix} -2a_1 & z_{12}+a_1-a_2 & \cdots & z_{1n}+a_1-a_n \\ z_{12}+a_1-a_2 & -2z_{12} & \cdots & z_{2n}-z_{12}-z_{1n} \\ z_{13}+a_1-a_3 & z_{23}-z_{12}-z_{13} & \cdots & z_{3n}-z_{13}-z_{1n} \\ \vdots & \vdots & & \vdots \\ z_{1n}+a_1-a_n & z_{2n}-z_{12}-z_{1n} & \cdots & -2z_{1n} \end{vmatrix}, \quad (24)$$

$$\tilde{\psi}_{ij} = \tilde{\psi}_{ji}. \quad (24a)$$

Using the same procedure, we can transform $P_i$ to the form

$$P_i = (-1)^{i+1}\tilde{\psi}^{1i} \qquad i \neq 1, \quad (25)$$

where $\tilde{\psi}^{1i}$ is the determinant obtained from $\tilde{\psi}$ by suppressing row one and column $i$.

**B. Comparison between the Coefficients in (7) and (23).**

Let $Q$ be an arbitrary vector in $m$-dimensional Lorentz space and $(p_{12}\cdots p_{1m})$ a set of vectors such that their Gram determinant is different from zero, and let one of them, e.g., $p_{12}$ be timelike, i.e.,

$$G(p_{12}\cdots p_{1m}) = \det(p_{1i}\cdot p_{1j}) \neq 0, \quad (26a)$$

$$p_{12}^2 < 0. \quad (26b)$$

Because of (26a) the vectors $(p_{12}\cdots p_{1m})$ are linearly independent. The vector $Q$ can be expanded as

$$Q = \frac{1}{G(p_{12}\cdots p_{1m})}\left\{\begin{vmatrix} 0 & -p_{12} & \cdots & -p_{1m} \\ Q\cdot p_{12} & p_{12}^2 & \cdots & p_{12}\cdot p_{1m} \\ \vdots & \vdots & & \vdots \\ Q\cdot p_{1m} & p_{12}\cdot p_{1m} & \cdots & p_{1m}^2 \end{vmatrix} + \begin{vmatrix} Q & p_{12} & \cdots & p_{1m} \\ Q\cdot p_{12} & p_{12}^2 & \cdots & p_{12}\cdot p_{1m} \\ \vdots & \vdots & & \vdots \\ Q\cdot p_{1m} & p_{12}\cdot p_{1m} & \cdots & p_{1m}^2 \end{vmatrix}\right\}. \quad (27)$$

We introduce a vector $\beta$ perpendicular to $(p_{12}\cdots p_{1m})$. Because $p_{12}$ is timelike, $\beta$ is necessarily spacelike. Further, we require

$$\beta^2 = 1. \quad (27a)$$

The set $(p_{12}\cdots p_{1m}, \beta)$ spans the whole $m$-dimensional space. The vector given by the second determinant in (27) is perpendicular to $(p_{12}\cdots p_{1m})$. Thus it is zero or parallel to $\beta$. We write $Q$ as follows:

$$G(p_{12}\cdots p_{1m})Q = \begin{vmatrix} 0 & -p_{12}\cdots -p_{1m} \\ Q\cdot p_{12} & \\ \vdots & G(p_{12}\cdots p_{1m}) \\ Q\cdot p_{1m} & \end{vmatrix} \pm [G(Q, p_{12}\cdots p_{1m})G(p_{12}\cdots p_{1m})]^{\frac{1}{2}}\beta. \quad (28)$$

Let $q_P$ be a vector which satisfies

$$\begin{aligned} &(q_P - p_1)^2 + a_1 = 0, \\ &\vdots \qquad\qquad (m \text{ equations}) \\ &(q_P - p_m)^2 + a_m = 0. \end{aligned} \quad (29)$$

There are exactly two such vectors $q_P^{\pm}$. We define $Q_P^{\pm} = q_P^{\pm} - p_1$. Equations (29) become

$$\begin{aligned} &Q_P^2 + a_1 = 0 \\ &(Q_P + p_{1l})^2 + a_l = 0 \quad p_{1l} \in (p_{12}\cdots p_{1m}). \end{aligned} \quad (30)$$

We expand $Q_P$ by means of formula (28) and, by virtue of this, we can write our original coefficients in (7) in terms of determinants. Let $p_{1i}$ be a vector such that $p_{1i} \notin (p_{12}\cdots p_{1m})$. Consider the expression

$$G(p_{12}\cdots p_{1m})[(Q_P + p_{1i})^2 + a_i] = -2\left\{\begin{vmatrix} -\frac{1}{2}(Q_P^2 + p_{1i}^2 + a_i) & p_{1i}\cdot p_{12}\cdots p_{1i}\cdot p_{1m} \\ Q_P\cdot p_{12} & \\ \vdots & G(p_{12}\cdots p_{1m}) \\ Q_P\cdot p_{1m} & \end{vmatrix} \pm [G(Q, p_{12}\cdots p_{1m})G(p_{12}\cdots p_{1m})]^{\frac{1}{2}}\beta\cdot p_{1i}\right\}. \quad (31)$$

We can expand $p_{1i}$ in the same way as $Q_P$ and, since the sign of $\beta$ is not fixed, we choose it so that

$$\beta\cdot p_{1i} = [G(p_{1i}, p_{12}\cdots p_{1m})/G(p_{12}\cdots p_{1m})]^{\frac{1}{2}}. \quad (32)$$

Then we get

$$G(p_{12}\cdots p_{1m})\{[(Q_P^+ + p_{1i})^2 + a_i] + [(Q_P^- + p_{1i})^2 + a_i]\} = -4\begin{vmatrix} -\frac{1}{2}(Q_P^2+p_{1i}^2+a_i) & p_{1i}\cdot p_{12} \cdots p_{1i}\cdot p_{1m} \\ Q_P\cdot p_{12} & \\ \vdots & G(p_{12}\cdots p_{1m}) \\ Q_P\cdot p_{1m} & \end{vmatrix} \equiv 2P_i'. \tag{33}$$

We can identify $P_i'$ with the determinant $P_i$ in (25). We use Eqs. (30) and the definition of $z_{ij}$ (2)

(a) $Q_P^2 = -a_1$,

(b) $2Q_P\cdot p_{1l} = z_{1l} + a_1 - a_l$, $p_{1l} \in (p_{12}\cdots p_{1m})$,

(c) $2p_{1i}\cdot p_{1j} = z_{ij} - z_{1i} - z_{1j}$. (34)

If these equalities are substituted into (33), we get

$$P_i' = -\begin{vmatrix} z_{1i}+a_1-a_i & z_{2i}-z_{12}-z_{1i} & \cdots & z_{im}-z_{1i}-z_{1m} \\ z_{12}+a_1-a_2 & -2z_{12} & \cdots & z_{2m}-z_{12}-z_{1m} \\ \vdots & & & \vdots \\ z_{1m}+a_1-a_m & z_{2m}-z_{12}-z_{1m} & \cdots & -2z_{1m} \end{vmatrix} = (-1)^{i+1}\bar{\psi}^{1i} = P_i, \qquad i \neq 1. \tag{35}$$

With trivial modifications of the proof above we can deal with the case $i = 1$. To compare the determinant $\psi$ in (24) with our coefficients, we substitute Eqs. (34) in (24) and interchange the $i$th row and column with the second row and column. We then obtain $\psi$ as a function of $Q_P$ and $p_{ij}$

$$\psi = \bar{\psi} = 2\begin{vmatrix} Q_P^2 & -\frac{1}{2}(Q_P^2 + p_{1i}^2 + a_i) & Q_P\cdot p_{12} \cdots Q_P\cdot p_{1m} \\ -\frac{1}{2}(Q_P^2 + p_{1i}^2 + a_i) & p_{1i}^2 & p_{1i}\cdot p_{12} \cdots p_{1i}\cdot p_{1m} \\ Q_P\cdot p_{12} & p_{1i}\cdot p_{12} & \\ \vdots & \vdots & G(p_{12}\cdots p_{1m}) \\ Q_P\cdot p_{1m} & p_{1i}\cdot p_{1m} & \end{vmatrix}. \tag{36}$$

Next, we use a theorem from the elementary theory of determinants which provides us with a correspondence between minors in $\psi^{-1}$ and minors in $\psi$:

$$\psi^{-1} = \left|\begin{array}{c|c} \Delta & \\ \hline & \end{array}\right| \qquad \psi = \left|\begin{array}{c|c} & \\ \hline & \Delta' \end{array}\right|.$$

With the notation above we have

$$\Delta\psi = \Delta'. \tag{37}$$

In our case we take $\Delta$ as

$$\Delta = \begin{vmatrix} (\psi^{-1})_{11} & (\psi^{-1})_{12} \\ (\psi^{-1})_{12} & (\psi^{-1})_{22} \end{vmatrix} = \frac{1}{\psi^2}\begin{vmatrix} 2G(p_{1i}, p_{12}\cdots p_{1m}) & P_i \\ P_i & 2G(Q_P, p_{12}\cdots p_{1m}) \end{vmatrix}. \tag{38}$$

It is easy to see that the quantities in $\Delta$ are the corresponding minors of (36). Hence we get

$$4G(p_{1i}, p_{12}\cdots p_{1m})G(Q_P, p_{12}\cdots p_{1m}) - P_i^2 = \psi G(p_{12}\cdots p_{1m}) \tag{39}$$

because $\Delta' = G(p_{12}\cdots p_{1m})$.

As our final result we get

$$[(Q_P^{\pm} + p_{1i})^2 + a_i] = \frac{1}{G(p_{12}\cdots p_{1m})} \times \{P_i \pm 2[G(Q_P, p_{12}\cdots p_{1m})G(p_{1i}, p_{12}\cdots p_{1m})]^{\frac{1}{2}}\} \tag{40}$$

and, using (39),

$$\frac{1}{2}\left\{\frac{1}{(Q_P^+ + p_{1i})^2 + a_i} + \frac{1}{(Q_P^- + p_{1i})^2 + a_i}\right\} = \frac{1}{2}\frac{(Q_P^+ + p_{1i})^2 + a_i + (Q_P^- + p_{1i})^2 + a_i}{[(Q_P^+ + p_{1i})^2 + a_i][(Q_P^- + p_{1i})^2 + a_i]} = \frac{1}{2}\frac{G(p_{12}\cdots p_{1m})2P_i}{P_i^2 - 4G(Q_P, p_{12}\cdots p_{1m})G(p_{1i}, p_{12}\cdots p_{1m})} = (-1)\frac{P_i}{\psi} \quad \text{Q.E.D.}$$

The identification is made for real $p_i$ subject to the conditions (26a) and (26b). As the coefficients are analytic functions of the external invariants, the result can be analytically continued from there.

## II. PROOF FOR THE CASE $n \geq m + 2$

We use a result given by Brown,[3] which provides us with a reduction formula from $n$ to $n - 1$. Consider the function for the diagram with $n$ lines:

$$F(z_{ik}, a_k) = \int \left[ d^m q \Big/ \prod_{k=1}^{n} [k(q)] \right].$$

If $n \geq m + 2$, we can always find a set of numbers $b_i$ such that

$$\sum_{i=1}^{n} b_i = 0, \tag{41}$$

$$\sum_{i=1}^{n} b_i p_i = 0. \tag{42}$$

Then,

$$\sum_{i=1}^{n} b_i [i(q)] = \sum_{i=1}^{n} b_i (p_i^2 + a_i) \tag{43}$$

and this sum is independent of $q$. We use a solution of (41) and (42) in which

$$b_{m+3} = b_{m+4} = \cdots = b_n = 0.$$

$$F = \int \frac{d^m q}{\prod_{k=1}^{n} [k(q)]} = \frac{\sum_{i=1}^{m+2} b_i}{\sum_{i=1}^{m+2} b_i [j]} \int \frac{[i(q)]\, d^m q}{\prod_{k=1}^{n} [k(q)]}. \tag{44}$$

We use this reduction formula to prove formula (7) by induction. Suppose (7) is valid for all diagrams with $n - 1$ lines. We have already proved that (7) is valid for $n - 1 = m + 1$.

$$\therefore \int \frac{d^m q}{\prod_{\substack{k=1 \\ k \neq i}}^{n} [k]} = \frac{1}{2} \sum_{\substack{P \\ i \in P}} \left\{ \frac{1}{\prod_{\substack{l \in P \\ l \neq i}} [l(q_P^+)]} + \frac{1}{\prod_{\substack{l \in P \\ l \neq i}} [l(q_P^-)]} \right\} \int \frac{d^m q}{\prod_{k \in P} [k(q)]}$$

$$= \frac{1}{2} \sum_{\substack{P \\ i \in P}} \left\{ \frac{[i(q_P^+)]}{\prod_{l \in P} [l(q_P^+)]} + \frac{[i(q_P^-)]}{\prod_{l \in P} [l(q_P^-)]} \right\} \int \frac{d^m q}{\prod_{k \in P} [k(q)]}. \tag{45}$$

If we substitute this in (44), we get

$$F = \frac{1}{2} \frac{\sum_{i=1}^{m+2} b_i}{\sum_{i=1}^{m+2} b_i [j]} \sum_{\substack{P \\ i \in P}} \left\{ \frac{[i(q_P^+)]}{\prod_{l \in P} [l(q_P^+)]} + \frac{[i(q_P^-)]}{\prod_{l \in P} [l(q_P^-)]} \right\} \int \frac{d^m q}{\prod_{k \in P} [k(q)]}. \tag{46}$$

If we interchange the sums over $P$ and $i$ and use the fact that

$$\sum b_i [i(q)] = \sum b_i [i(q_P^+)] = \sum b_i [i(q_P^-)], \tag{47}$$

we get

$$F = \frac{1}{2} \sum_{P} \left\{ \frac{1}{\prod_{l \in P} [l(q_P^+)]} + \frac{1}{\prod_{l \in P} [l(q_P^-)]} \right\} \int \frac{d^m q}{\prod_{k \in P} [k(q)]}$$

which is the desired result.

*Note added in proof.* After this article was submitted, the author obtained a preprint by D. B. Melrose, "Reduction of Feynman Diagrams," in which similar results are proved with the aid of different techniques.

## ACKNOWLEDGMENTS

I wish to express my deep gratitude to Professor Gunnar Källén, who suggested this investigation, for his most generous help and advice. I also wish to thank Professor Torsten Gustafson for his kind interest and encouragement.

# The Froissart–Gribov Continuation and Reggeon Unitarity Conditions. I*

J. C. POLKINGHORNE†
*International Atomic Energy Agency, International Centre for Theoretical Physics, Trieste, Italy*
(Received 6 May 1965)

An identity is derived which is useful in the discussion of unitary integrals. It is used to discuss single Regge pole insertions in the Froissart–Gribov continuation and with the help of perturbation-theory models a derivation is given of reggeon unitarity conditions. The general form suggested by Gribov, Pomeranchuk, and Ter-Martirosyan is confirmed. Finally cancellation mechanisms are discussed and their relation to the mechanism generating Regge poles is emphasized.

## 1. INTRODUCTION

THE existence of cuts in the Regge plane[1-4] has considerably complicated the question of extracting the high-energy behavior of scattering amplitudes. Gribov, Pomeranchuk, and Ter-Martirosyan[5] have endeavoured to determine the form of the diffraction peak resulting from a superposition of the leading singularities generated by a Pomeranchuk pole. An important element in their analysis is the expression for the discontinuity around a Regge cut for which they propose an appealing unitarity-like form which they derive from a heuristic discussion of unitarity in the $s$-channel. Some of the details of their proposal are found not to be confirmed by a perturbation theory model discussed by Swift.[6]

The purpose of this paper is to explore the discontinuity formula for Regge cuts from a different point of view. The method is essentially an extension of that given in I and is based upon the Froissart–Gribov definition of the Regge amplitude

$$b^{\pm}(l, s) = \frac{k^{-2l}}{\pi} \int_{t_0}^{\infty} \frac{dt'}{2k^2} Q_l\left(1 + \frac{t'}{2k^2}\right) \times [A_t(s, t') \pm A_u(s, t')], \quad (1)$$

where $k$ is the center-of-mass momentum associated with $s$, and $A_t$, $A_u$ are the discontinuities across the t and u normal thresholds respectively. These discontinuities are themselves, by generalized unitarity, bilinear expressions in terms of scattering amplitudes. Thus the possibility of an iterative scheme exists similar to that which generates Landau singularities[7,8]: an assumption about the high-energy behavior of amplitudes will lead to a high-energy behavior of $A_t$ and $A_u$, which will then give singularities in $l$ in (1). However, since it is the high $t'$ behavior of $A_t$ and $A_u$ which is in question, the relevant expressions will contain contributions from intermediate states with all possible numbers of particles. Thus a knowledge of the high-energy properties of production–process amplitudes is necessary to start the iteration. Nothing useful is known rigorously about this problem but perturbation-theory models have illustrated the type of behavior which may be expected.[9] In this paper, therefore, we shall use a blend of rigorous methods together with ansatz suggested by perturbation theory. Only single Regge poles will be considered. Multiple poles, which are a feature of production amplitudes, will be dealt with in a second paper.

In Sec. 2, we derive a useful identity. It is used in Sec. 3 to evaluate the effect of certain Regge pole insertions. The resulting cuts and their discontinuities are evaluated in Sec. 4. In Sec. 5 the vitally important case of insertions associated with crossed lines is considered. Part of the integration is evaluated using the identity. The remainder is evaluated by a device which uses perturbation theory as a means for evaluating unitary integrals. In Sec. 6 it is emphasized that the cancellations occurring when many insertion terms are summed depend for their possibility on a mechanism which generates Regge poles in such a manner as to associate them with negative integers in the weak coupling limit. Finally in Sec. 7 it is shown that the sum over all single pole insertions produces a cut whose discontinuity is of the Gribov, Pomeranchuk, and Ter-Martirosyan type.

* The research reported in this document has been sponsored in part by U. S. Air Force Office of Scientific Research under grant AF EOAR 65-36, through the European Office of Aerospace Research (OAR) U. S. A. F.

† On leave from Department of Applied Mathematics and Theoretical Physics, University of Cambridge.

[1] D. Amati, S. Fubini, and A. Stanghellini, Phys. Letters 1, 29 (1962).

[2] J. C. Polkinghorne, Phys. Rev. **128**, 2459 (1962); referred to as I.

[3] S. Mandelstam, Nuovo Cimento **30**, 1127, 1148 (1963).

[4] J. C. Polkinghorne, J. Math. Phys. (N. Y.) 4, 1393 (1963).

[5] V. N. Gribov, I. Ya. Pomeranchuk, and K. A. Ter-Martirosyan Phys. Rev. **139**, B184 (1965); and to be published.

[6] A. R. Swift, J. Math. Phys. (N. Y.) (to be published).

[7] J. C. Polkinghorne, Nuovo Cimento **23**, 360; **25**, 901 (1962).

[8] H. P. Stapp, Phys. Rev. **125**, 2139 (1962).

[9] I. G. Halliday and J. C. Polkinghorne, Phys. Rev. **132**, 852 (1963); J. C. Polkinghorne, Nuovo Cimento **36**, 857 (1965).

## 2. A USEFUL IDENTITY[10]

We consider

$$I = \frac{1}{\pi}\int_{\Delta=0}^{\infty} \frac{Q_l(1 + t/2k^2)}{[\Delta(s, t; m^2, t_1, t_2)]^{\frac{1}{2}}} \frac{dt}{2k^2}. \quad (2)$$

The function $\Delta^{-\frac{1}{2}}$ is the Jacobian for the transformation from an integral over the loop momentum to an integral over invariants[11] for the loop of Fig. 1. It is given by

$$\Delta = \det(p_i p_j), \quad (3)$$

where $p_i$ $(i = 1, \cdots 4)$ are the momenta in the four lines of the loop.

We take $l$ as integral and later use Carlson's theorem to extend the result to nonintegral $l$. Then (2) may be rewritten, using the inverse of the Froissart–Gribov procedure, as

$$I = \frac{1}{2}\int_{-1}^{+1} P_l(z)\, dz \left[\frac{2i}{-4\pi^2}\int \frac{dt_1'\, dt_2'}{[\Delta(s, t; m^2, t_1', t_2')]^{\frac{1}{2}}} \times \frac{1}{t_1' - t_1}\frac{1}{t_2' - t_2}\right], \quad (4)$$

where $z = 1 + t/2k^2$ and $F$, the function in the square brackets, is constructed to have a spectral function associated with its $t$-cut which is just $\Delta^{-\frac{1}{2}}$. Now the integral in $F$ is (to within a factor of $i$, since $\Delta < 0$ in the physical region) just an $s$-channel phase-space integral so that (4) may be rewritten as

$$\frac{k}{2\pi^2(k^2 + m^2)^{\frac{1}{2}}}\left(\frac{1}{2}\int_{-1}^{+1}\frac{dt_1'}{2k^2}\frac{P_l(z_1')}{t_1' - t_1}\right) \times \left(\frac{1}{2}\int_{-1}^{+1}\frac{dt_2'}{2k^2}\frac{P_l(z_2')}{t_2' - t_2}\right), \quad (5)$$

giving as our final identity

$$I = \frac{1}{2k^2}\frac{1}{\pi^2}\frac{1}{4k(k^2 + m^2)^{\frac{1}{2}}} \times Q_l(1 + t_1/2k^2)Q_l(1 + t_2/2k^2). \quad (6)$$

We now deduce from Carlson's theorem that the identity (6) which we have proved for integral $l$ is also true for nonintegral $l$.

It is possible to derive similar identities by the same method which apply when not all the masses are taken equal to $m$ as are indicated in Fig. 1.

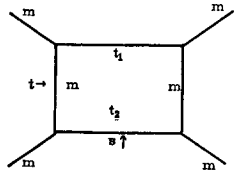


FIG. 1. The loop associated with $\Delta$.

## 3. REGGE POLES IN UNITARY INTEGRALS

The identity (6) provides a simple means of evaluating the effect of Regge pole contributions in the unitary integrals appearing in $A_t$ and $A_u$. We can consider the term diagrammatically represented in Fig. 2. It corresponds to a term in the unitary sum for $A_t$ with the internal solid lines representing particles on the mass shell. The two wavy lines represent Regge poles with trajectories $\alpha(s_1)$, $\alpha(s_2)$ inserted in the two scattering amplitudes of the unitary integral. The asymptotic variable of the Regge pole is $t_{12}'$, the invariant associated with the sum of the momenta in the two sets of lines in the figure which carry $t_1'$, and $t_2'$, respectively.

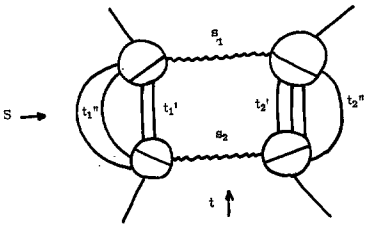


FIG. 2. A possible Regge pole insertion in $A_t$.

It is convenient when considering the unitary integral to represent it in terms of integrals over invariants[11]. The dual diagram, or momentum diagram, of Fig. 2 is given in Fig. 3. Sets of lines carrying momenta in which we are interested are

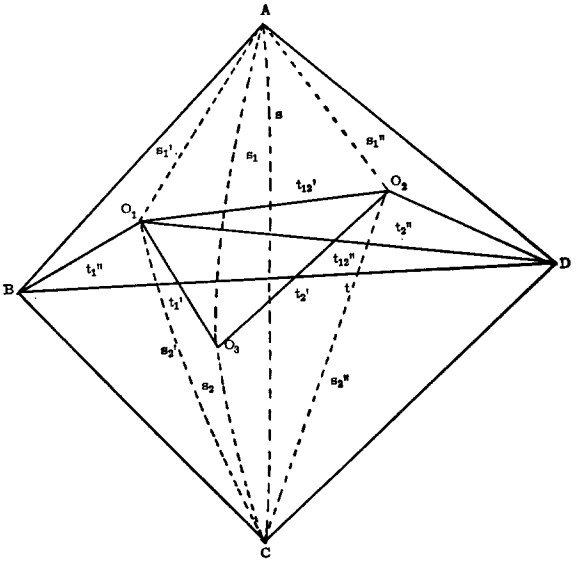


FIG. 3. The dual diagram of Fig. 2.

[10] Formulas similar to this have also been given by V. N. Gribov, Zh. Eksperim. i Teor. Fiz. 41, 1962 (1961) [English transl.: Soviet Phys.—JETP 14, 1395 (1962)]; R. Omnès, Nuovo Cimento 25, 806 (1962).

[11] I. T. Drummond, Nuovo Cimento 29, 720 (1963).

represented by a single line, corresponding to the sum of the momenta of the set. It is clear that Fig. 3 can be constructed by constructing the successive simplexes A B C D $O_1$, A $O_1$ C D $O_2$, A $O_1$ C $O_2$ $O_3$. This means that the unitary integral contains the integrations

$$\int \frac{dt_1' \, dt_2' \, dt_1'' \, dt_2'' \, dt_{12}' \, dt_{12}'' \, ds_1 \, ds_2 \, ds_1' \, ds_2' \, ds_1'' \, ds_2''}{[\Delta(s, t; t_1'', t_{12}'', s_1', s_2')]^{\frac{1}{2}}[\Delta(s, t_{12}''; t_{12}', t_2'', s_1'', s_2'')]^{\frac{1}{2}}[\Delta(s, t_{12}'; t_1', t_2', s_1, s_2)]^{\frac{1}{2}}}, \tag{7}$$

where $\Delta$ is defined by (3) for each of the three loops corresponding to the three simplexes. There are also additional integrations over further subinvariants within the sets of lines making up the invariants $t_1''$, $t_2''$, $t_1'$, $t_2'$, respectively. We do not need to consider these explicitly. Neither, since we are interested only in the structure, do we need to keep a precise tally of numerical factors in the manipulations that follow and these factors will in fact be omitted.

We regard the identity (6) as evaluating the Regge transform of $F$. Then using the Mandelstam form[12] of the Sommerfeld–Watson transform we may write

$$F \sim \int_C \frac{dl \, (2l+1) Q_l(t_1) Q_l(t_2) Q_{-l-1}(t)}{\lambda^{\frac{1}{2}}(s, s_1, s_2) \cos \pi l} + \cdots, \tag{8}$$

where the succinct notation $Q_l(t_1)$ has been adopted, the factor $\lambda^{\frac{1}{2}}$ with

$$\lambda = s^2 + s_1^2 + s_2^2 - 2ss_1 - 2ss_2 - 2s_1s_2, \tag{9}$$

is the correct generalization[11] of the kinematic factor $4k(k^2+m^2)^{\frac{1}{2}}$ in (6) to the unequal mass case, and we have omitted writing down explicitly the terms which cancel the poles of sec $\pi l$ since they are not relevant to our discussion. The contour $C$ runs parallel to the imaginary axis to the right of $l = -1$.

We are not, however, interested in $F$ but in $\Delta^{-\frac{1}{2}}$, its spectral function across the $t$-cut. From (9) this must be representable in the form

$$\Delta^{-\frac{1}{2}} \sim \int_C \frac{dl \, (2l+1) Q_l(t_1) Q_l(t_2) Q_{-l-1}(t) \sin \pi l}{\lambda^{\frac{1}{2}}(s, s_1, s_2) \cos \pi l}. \tag{10}$$

As a check on the consistency of the discussion we may note that the poles of the $Q$ functions at $l = -1$ produce for the leading asymptotic behavior of (10)

$$\sim t^{-1}\lambda^{-\frac{1}{2}}(s; s_1, s_2), \tag{11}$$

which is easily seen to be correct by comparison with the explicit form of $\Delta^{-\frac{1}{2}}$.

We now introduce into the unitary integral (6) representations of the form of (10) for the three $\Delta^{-\frac{1}{2}}$ functions involved, labeling the integration parameters as $l_1$, $l_2$, $l_3$, respectively. The integrand will also contain the Regge pole terms corresponding to Fig. 2 whose leading behaviors will be

$$\begin{aligned} &t_{12}'^{\alpha(s_1)} \beta(s_1, t_1'', t_1') \beta(s_1, t_2'', t_2'), \\ &t_{12}'^{\alpha(s_2)} \beta(s_2, t_1'', t_1') \beta(s_2, t_2'', t_2'). \end{aligned} \tag{12}$$

The whole expression is now substituted into the Froissart–Gribov definition for $a(l, s)$ and the $t_1$, $t_{12}''$, $t_{12}'$ integrations performed since the supposed dependence of the integrand on these variables is now explicitly exhibited. The result for the leading behavior is

$$\begin{aligned} &\int \frac{ds_1' \, ds_2' \, ds_1'' \, ds_2'' \, ds_1 \, ds_2 \, dt_1' \, dt_2' \, dt_1'' \, dt_2'' \, d\xi}{\lambda^{\frac{1}{2}}(s; s_1', s_2')\lambda^{\frac{1}{2}}(s; s_1'', s_2'')\lambda^{\frac{1}{2}}(s; s_1, s_2)}, \\ &\int dl_1 \frac{f_1(l, l_1)}{l - l_1} \sin \pi l_1 Q_{l_1}(t_1''), \\ &\int dl_2 \frac{f_2(l_1, l_2)}{l_1 - l_2} \sin \pi l_2 Q_{l_2}(t_2''), \\ &\int dl_3 \frac{f_3(l_2, l_3)}{l_2 - l_3 - \alpha(s_1) - \alpha(s_2)} \sin \pi l_3 Q_{l_3}(t_1') Q_{l_3}(t_2'), \\ &\beta(s_1, t_1'', t_1') \beta(s_1, t_2'', t_2') \beta(s_2, t_1'', t_1') \beta(s_2, t_2'', t_2'), \end{aligned} \tag{13}$$

where $f_1$, $f_2$, $f_3$ are known functions arising from the $t$, $t_{12}''$ and $t_{12}'$ integrations, respectively. Their properties which are relevant for our discussion are that $f_1$ has a pole at $l = -1$, $f_2$ a pole at $l_1 = -1$, and $f_3$ a pole at $l_2 = -1$. The collective symbol $\xi$ stands for the integrals over all the subinvariants not explicitly shown.

The leading singularity in $l$ of (13) is to be obtained by moving all the contours as far as possible to the left. The limits to these distortions will arise from the explicit denominators in (13), from the poles of $f_1$, $f_2$, $f_3$ mentioned above, and from the singularities in $l_1$, $l_2$, $l_3$ arising from the $t_1''$, $t_2''$, $t_1'$, $t_2'$ integrations. In order to evaluate this last effect it is necessary to know something of the properties of the $\beta$'s. For this we turn to perturbation theory models.

## 4. THE CUT CONTRIBUTIONS

The terms in which we are interested will be a sum of expressions of type (13) taken over the different numbers of lines which can occur in the sets of momenta whose sums give $t_1''$, $t_2''$, $t_1'$, $t_2'$. Such a term will contain within it an expression of the form

[12] S. Mandelstam, Ann. Phys. (N. Y.) 19, 254 (1962).

$$\sum \int dt_1' \, d\xi_1' Q_{l_*}(t_1') \beta(s_1, t_1'', t_1') \beta(s_2, t_2'', t_1'), \qquad (14)$$

where $\int d\xi_1'$ is taken on the subinvariants of the set of momenta whose sum is $t_1'$ and the summation is taken over the number of momenta in the set. A form like (14) looks very much like a generalized Froissart–Gribov definition for the continuation of an amplitude whose dependence on $t_1'$ has been replaced by a partial wave projection, and one might therefore expect that it would give singularities in $l_3$ corresponding to Regge poles, etc.

That this is not the case can be seen by the study of some simple perturbation theory models. Figure 4 illustrates a simple diagram corresponding to the type of contribution we are considering. The lines carrying $t_1''$ and $t_2''$ are joined to the remainder of the diagram by single lines since otherwise their momenta would be coupled to those appearing in $t_1'$ and $t_2'$ and so would be associated with the Regge pole[9]. This structure ensures that no Regge pole appears in the $t$ or $t_{12}''$ integration.

The combination of two poles of this type in a unitary integral gives a $t_1'$ factor to insert in (14) of the type shown in Fig. 5. Increasing the number of lines carrying the $t_1'$ variable will not produce in Fig. 5 the iteration which is necessary to produce Regge poles, etc. In the terminology of the high-energy behavior of perturbation theory the number of minimal $d$-lines in Fig. 5 is not increased by increasing the number of lines carrying $t_1'$. This feature illustrated by Figs. 4 and 5 is perfectly general. The sets of lines carrying $t_1'$ and $t_2'$ must always be joined to subdiagrams corresponding to $\beta$-functions and so in the unitary integral appear sandwiched between these subdiagrams in a non-iterative way. Therefore, as far as the variable $l_3$ is concerned, the leading singularity is given simply by the poles of the $Q_{l_*}$ functions in (13). When account is taken of the factor $\sin \pi l_3$ this just gives a simple pole at $l_3 = -1$.

The leading singularity may now be extracted

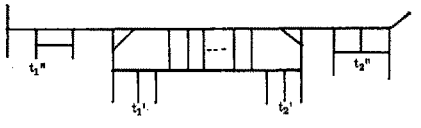


FIG. 4. A typical diagram contributing a Regge pole of the type considered.

from (13). The poles give $l = l_1 = l_2$, $l_3 = -1$ so that the singularity is

$$\int \frac{ds_1 \, ds_2}{\lambda^{\frac{1}{2}}(s, s_1, s_2)} \frac{F_l^{(1)} F_l^{(2)}}{l - \alpha(s_1) - \alpha(s_2) + 1}, \qquad (15)$$

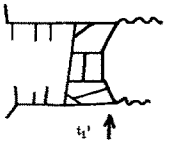


FIG. 5. The $\beta$ product in $t_1'$ corresponding to two diagrams like Fig. 4.

where

$$F_l^{(1)} = f_1(l, l) f_3^{\frac{1}{2}}(l_1, -1) \sin \pi l \times \int \frac{ds_1' \, ds_2' \, dt_1' \, d\xi_1' \, d\xi_1'' \, dt_1''}{\lambda^{\frac{1}{2}}(s, s_1', s_2')} \times Q_l(t_1') \beta(s_1, t_1'', t_1') \beta(s_2, t_1'', t_1'), \qquad (16)$$

and $F_l^{(2)}$ is similarly defined with $t_1''$ replaced by $t_2''$.

Singularities in $l$ of the functions $F^{(1)}$ and $F^{(2)}$ will arise from the $t_1''$ and $t_2''$ integrations, respectively. The restrictions on the possible singularities which applied to the similar integrations over $t_1'$ and $t_2'$ are no longer operative as Fig. 4 illustrates. Thus the functions $F_l^{(1)}$, $F_l^{(2)}$ may be expected to possess the complete range of Regge poles, Regge cuts and essential singularities.

The form (15) is similar to the expression proposed by Gibov, Pomeranchuk, and Ter-Martirosyan[5]. It differs from theirs only in the fact that they state that $F^{(1)}$ and $F^{(2)}$ should be evaluated on opposite sides of the $l$-cut. When account is taken of the fact that $F^{(1)}$ and $F^{(2)}$ themselves contain the cut singularity so that their discontinuities must also be added in, it is easy to see that the resulting sum of discontinuities then reduces to their prescription for the complete discontinuity around the cut.[13]

So far however, we have only discussed the combination of Regge poles which both correspond to the same asymptotic variable $t_{12}'$. In fact however the contributions from all such diagrams of this type must cancel among themselves for they correspond to diagrams which in perturbation theory will not have the necessary pinch contributions present for positive Feynman parameters.[4] Thus if (15) is to have more than purely formal interest it must be extended to a more general case. This is done in the next section.

## 5. CROSSED LINE INSERTIONS

We consider now insertions of the type illustrated by Fig. 6. Again the particular connection of the lines carrying $t_1''$ and $t_2''$ to the remainder of the diagram is dictated by the requirement that the Regge poles manifest themselves only in the com-

[13] Cf. normal threshold discontinuities evaluated in perturbation theory. Symbolically the structure is $\Delta = FF + FFF + \ldots = F(F + FF + \ldots) = F(F + \Delta)$.

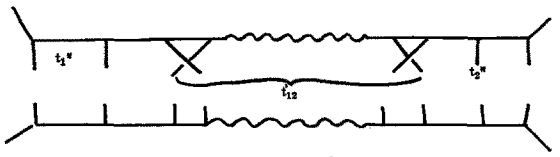

FIG. 6. A typical crossed line insertion.

plex of integrations corresponding to the variables carrying $t'_{12}$. A particularly simple crossed line structure has been drawn in Fig. 6 for illustrative purposes, but it will become clear that our methods can be applied to the general case.

The first steps in the analysis of Fig. 6 are as before. The unitary integrals are represented by integrals over momenta and the identity (6) is used. The integrals over $t$ and $t'_{12}$ can be performed, and we are left with an expression of the structure

$$\begin{aligned} &\int \frac{ds'_1\, ds'_2\, ds''_1\, ds''_2\, d\xi}{\lambda^{\frac{1}{2}}(s; s'_1, s'_2)\lambda^{\frac{1}{2}}(s; s''_1, s''_2)}\,, \\ &\int dl_1 \frac{f_1(l, l_1)}{l - l_1} \sin \pi l_1 \int dt''_1 Q_{l_1}(t''_1), \\ &\int dl_2 \frac{f_2(l_1, l_2)}{l_1 - l_2} \sin \pi l_2 \int dt''_2 Q_{l_2}(t''_2), \\ &\int dt'_{12} Q(t'_{12}) \cdot A_1 A_2, \end{aligned} \tag{17}$$

where $A_1A_2$ stands for the product of the two amplitudes combined in the unitary integral. The method becomes unusable however when the $t'_{12}$ integration has to be considered. The invariants of interest can no longer be represented in dual diagrams which can be constructed out of successive simplexes.

Instead of attempting directly to unravel the resulting kinematic complexities, we use Feynman integrals as a device for the evaluation of unitary integrals. The subintegrations of the $t'_{12}$ complex in Fig. 6 may be related to the discontinuity of an integral associated with Fig. 7 across the $t'_{12}$ cut in which the four crossed lines carrying $q_1 \cdots q_4$ are put on the mass shell. This integral is not yet in simple Feynman form for instead of the usual propagator we must associate with line 1 a term which we represent in the form

$$(q_1 + q_3)^{2\alpha(s_1''')} \equiv \int ds_1 (q_1 + q_3)^{2\alpha(s_1)} \delta(s_1 - s_1'''). \tag{18}$$

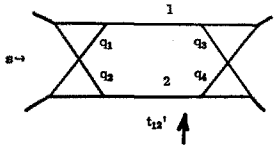

FIG. 7. The perturbation diagram associated with the $t_{12}'$ integration in Fig. 6.

The invariant $s_1'''$ is the squared momentum associated with line 1, and the reason for choosing the oblique expression (18) will become apparent. A similar factor is associated with the line 2. We now interpret the $\delta$-functions in these two factors as arising from taking the discontinuity of Fig. 7 across the normal threshold in $s$ at $(s_1 + s_2)^2$, where $s_1$ and $s_2$ are treated as being the internal masses associated with the lines 1 and 2. The integrations over $s_1$ and $s_2$ will be performed after all the other integrations associated with the diagram. These latter integrations we shall call the diagram integral.

We shall evaluate the asymptotic form of the discontinuity across $s = (s_1 + s_2)^2$ by taking the discontinuity of the asymptotic form of the diagram integral. The diagram integral is now in Feynman form with a propagator

$$\frac{(q_1 + q_3)^{2\alpha(s_1)}}{s_1''' - s_1} \tag{19}$$

for the line 1, and a similar expression for 2. The object of our manipulation has been to get an integral in which $s_1$, and hence the exponent $\alpha(s_1)$, is effectively constant since standard symmetric integration methods can then be applied. In evaluating the asymptotic form, the effect of the numerator in (19) must be taken into account using methods developed for the discussion of Feynman integrals with spin.[14] The numerators will provide factors proportional to $t^{\alpha(s_1)}$ and $t^{\alpha(s_2)}$ and the pinch[4] between the two crosses will give $t^{-1}$. This is for the diagram integral itself. It is now necessary to take two discontinuities. The first is across the $t'_{12}$ cut. The integrals corresponding to the function on either side of this cut differ by having different contours for the Feynman parameters. The difference between the two contours will also contain the pinch so that the $t^{-1}$ behavior is retained. Thus the asymptotic form of this discontinuity is found to be

$$\int \frac{ds_1'''\, ds_2'''}{\lambda^{\frac{1}{2}}(s, s_1''', s_2''')} \frac{f^2(s, s_1''', s_2''')}{(s_1''' - s_1)(s_2''' - s_2)} \times (t'_{12})^{\alpha(s_1)+\alpha(s_2)-1}. \tag{20}$$

The functions $f$ are calculable from a knowledge of the difference of the two contours referred to, but we do not need them explicitly here. The final discontinuity to be taken is across the normal threshold $s = (s_1 + s_2)^2$. This gives

$$\frac{f^2(s, s_1, s_2)}{\lambda^{\frac{1}{2}}(s, s_1, s_2)} (t'_{12})^{\alpha(s_1)+\alpha(s_2)-1}. \tag{21}$$

[14] J. C. Polkinghorne, J. Math. Phys. (N. Y.) 5, 1491 (1964).

Finally, when the $s_1$ and $s_2$ integrations are performed, (21) gives a contribution which when inserted into the $t'_{12}$ integration gives an $l_2$ singularity

$$\int \frac{f^2\, ds_1\, ds_2}{\lambda^{\frac{1}{2}}(s, s_1, s_2)} \frac{1}{l_2 - \alpha(s_1) - \alpha(s_2) + 1}. \qquad (22)$$

The combination of (17) and (22) will then give an $l$ cut which can be written in the form (15), with $f$'s incorporated into the definitions of $F_l^{(1)}$ and $F_l^{(2)}$.

Although we have discussed the particular case illustrated by Fig. 6, it is possible to replace the crosses by any diagrams with third spectral functions since all such pairs of diagrams give the typical pinch contribution resulting in a $t^{-1}$ asymptotic behavior in the spinless case.

The method may also be applied to the evaluation of contributions associated with Regge poles in different asymptotic variables but without crossed lines. There is no pinch on the physical side of the $t'_{12}$-cut, but on the lower side the $\alpha$-contours corresponding to the lines put on the mass shell are distorted back to negative values and a pinch contribution becomes possible (cf. Sec. 3 of Ref. 4). Therefore a pinch generated $t^{-1}$ factor will be obtained for the discontinuity. Of course the sum of all contributions of this particular type must give no cut because they correspond to planar diagrams.[3,4]

## 6. COMPARISON OF DIFFERENT INSERTIONS

We know that the diagram of Fig. 8 does not have a cut[3,15]. It corresponds to many different insertions of the type considered above and the sum of their cut contributions must just give zero. In this section we wish to emphasise an important property which makes this possible. By doing so we are essentially amplifying a remark made at the end of Ref. 15.

The contributions from the two-particle and three-particle intermediate states illustrated in Fig. 8 are two of the possible insertions. These must partly cancel to help give the cut-free result quoted above. At first sight this does not seem possible since the ladder generating the pole in the three-particle intermediate state has one less rung than that occurring in the two particle state and so the two contributions appear to be of different orders in the coupling constant and thus incapable of any general cancellation. However, the ladder for the three-particle state is associated with a subintegration over a variable, $t'$ say. If the ladder has $n$-rungs its leading behaviour will be $t'^{-1}(\ln t')^{n-1}$, and its contribution to the unitary integral will appear in the form

$$\int dt' Q_l(t') \cdot t'^{-1}(\ln t')^{n-1} \sim \frac{1}{(l+1)^{n+1}}. \qquad (23)$$

The appearance of $(n+1)$ rather than $n$ as the exponent on the right-hand side of (23) is due to pole of $Q_l$ at $l = -1$. This has the effect of making the ladder of $n$-rungs associated with $t'$ equivalent to a ladder of $(n+1)$ rungs associated with $t$. Thus the two-particle and three-particle states give contributions of the same order, and this is easily generalized to the $n$-particle states, so that cancellation is possible. From this discussion we see that the possibility of cancellation is intimately connected with the existence of Regge poles generated in a way which associates them with the negative integers at which the $Q_l$ functions have poles. The cancellation would not be expected to occur for any other sort of Regge pole, if such could exist.

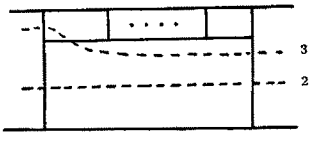


FIG. 8. A diagram not giving a cut.

## 7. CONCLUSION

We have shown that each Regge pole insertion leads to a cut whose discontinuity is given by (15). The functions $F_l^{(1)}$ and $F_l^{(2)}$ are defined by integrations over the variables $t'_1$ and $t'_2$. In summing over all the possible insertions, we must sum over all the possible structures which can be inserted into their subintegrations. In order to uncouple the momenta in $t'_1$ and $t'_2$ from the Regge poles we noted that these structures are joined to the $t'_{12}$ part of the diagram by pairs of lines so that the form is that illustrated by Fig. 9. A sum over all these possible insertions will give a discontinuity of the form (15) also. Thus the form suggested by Gribov, Pomeranchuk and Ter-Martirosyan[5] is confirmed by this analysis.

However the detailed properties of the functions $F_l^{(1)}$ and $F_l^{(2)}$ cannot be determined without taking

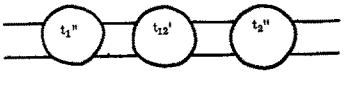


FIG. 9. The insertion structure corresponding to the Regge poles considered.

[15] J. C. Polkinghorne, Phys. Letters 4, 24 (1963).

into account the effect of cancellations when these sums are taken. In fact the example discussed by Swift[6] shows that these properties differ from those which one might simply expect.

The question of cuts generated by multiple pole insertions will be taken up in a subsequent paper.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor Abdus Salam and the International Atomic Energy Association for hospitality at the International Centre for Theoretical Physics, Trieste, during which this paper was completed.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 12 DECEMBER 1965

# Critical Störmer Conditions in Quadrupole and Double Ring-Current Fields

EUGENE W. URBAN
*National Aeronautics and Space Administration, George C. Marshall Space Flight Center*
*Huntsville, Alabama*
(Received 17 December 1964)

A theoretical study has been made of the behavior of the critical Störmer pass points for general axially symmetric magnetic configurations. A topological method has been derived to predict the occurrence of the critical Störmer conditions for charged-particle exclusion. This analytic technique, when applied to geomagnetically interesting fields, should be a useful aid to the understanding of experimental data. The method is applied here to three magnetic geometries: double ring currents with parallel dipole moments and with antiparallel dipole moments, and the axial magnetic quadrupole. The topology of the regions representing allowed motion is treated systematically and the behavior of the critical pass points is illustrated in typical Störmer plots. For the quadrupole and the antiparallel ring system critical pass points are found to occur only out of the equatorial plane. For the parallel ring system, critical points can occur in or out of the equatorial plane. For certain special conditions as many as three simultaneous critical pass points are found, and two simultaneous points occur for a wide range of parameters.

## INTRODUCTION

SEVERAL extensions of the Störmer problem[1] of allowed motion of unbound, charged particles have been made in connection with cosmic ray effects and the shielding of current systems. Treiman[2] and Ray,[3] following suggestions of Chapman and Ferraro,[4] treated the combination of a magnetic dipole and a ring current. Others (Akasofu and Cain,[5] Akasofu and Chapman,[6] Akasofu and Lin,[7] Kellogg and Winckler,[8]) have further considered this problem. Levy[9] and a group at the Lockheed Research Laboratories[10] considered the applications of Störmer's criteria to a single-turn current loop and, also in connection with space shielding, Tooper[11] has studied the Störmer problem of a right, circular solenoid.

The present work, a treatment of the Störmer problem of three simple axially symmetric magnetic systems, reveals some remarkable properties of the Störmer forbidden regions, which could have application to cosmic-ray and trapped-radiation studies. The systems are not considered in combination with a dipole field, and the results, therefore, cannot be directly compared with geophysical data. However, the behavior of the Störmer regions is sufficiently complex that this intermediate step is valuable for a reasonably simple interpretation of later studies of the possible geomagnetic effects. Two of the three field geometries considered here are those due to two parallel, coaxial current loops with currents in the same sense (parallel dipole

[1] C. Störmer, *The Polar Aurora* (Clarendon Press, Oxford, England, 1955).
[2] S. B. Treiman, J. Geophys. Res. **89**, 130 (1953).
[3] E. C. Ray, Phys. Rev. **101**, 1142 (1956).
[4] S. Chapman and V. Ferraro, Terr. Mag. Atmos. Elec., **38**, 79 (1933); **45**, 245 (1940).
[5] S. I. Akasofu and J. C. Cain, J. Geophys. Res. **67**, 4078 (1962).
[6] S. I. Akasofu and S. J. Chapman, J. Geophys. Res. **66**, 1321 (1961).
[7] S. I. Akasofu and W. C. Lin, J. Geophys. Res. **68**, 973 (1963).
[8] P. J. Kellogg and J. R. Winckler, J. Geophys. Res. **66**, 3991 (1961).
[9] R. H. Levy, Avco-Everett Research Laboratory Research Report 106 (1961).
[10] C. F. Kooi, Lockheed Technical Report 8-94-64-2 (1964).
[11] R. F. Tooper, ASD-TDR-63-194 (1963).

into account the effect of cancellations when these sums are taken. In fact the example discussed by Swift[6] shows that these properties differ from those which one might simply expect.

The question of cuts generated by multiple pole insertions will be taken up in a subsequent paper.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor Abdus Salam and the International Atomic Energy Association for hospitality at the International Centre for Theoretical Physics, Trieste, during which this paper was completed.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 12 DECEMBER 1965

# Critical Störmer Conditions in Quadrupole and Double Ring-Current Fields

EUGENE W. URBAN
*National Aeronautics and Space Administration, George C. Marshall Space Flight Center*
*Huntsville, Alabama*
(Received 17 December 1964)

A theoretical study has been made of the behavior of the critical Störmer pass points for general axially symmetric magnetic configurations. A topological method has been derived to predict the occurrence of the critical Störmer conditions for charged-particle exclusion. This analytic technique, when applied to geomagnetically interesting fields, should be a useful aid to the understanding of experimental data. The method is applied here to three magnetic geometries: double ring currents with parallel dipole moments and with antiparallel dipole moments, and the axial magnetic quadrupole. The topology of the regions representing allowed motion is treated systematically and the behavior of the critical pass points is illustrated in typical Störmer plots. For the quadrupole and the antiparallel ring system critical pass points are found to occur only out of the equatorial plane. For the parallel ring system, critical points can occur in or out of the equatorial plane. For certain special conditions as many as three simultaneous critical pass points are found, and two simultaneous points occur for a wide range of parameters.

## INTRODUCTION

SEVERAL extensions of the Störmer problem[1] of allowed motion of unbound, charged particles have been made in connection with cosmic ray effects and the shielding of current systems. Treiman[2] and Ray,[3] following suggestions of Chapman and Ferraro,[4] treated the combination of a magnetic dipole and a ring current. Others (Akasofu and Cain,[5] Akasofu and Chapman,[6] Akasofu and Lin,[7] Kellogg and Winckler,[8]) have further considered this problem. Levy[9] and a group at the Lockheed Research Laboratories[10] considered the applications of Störmer's criteria to a single-turn current loop and, also in connection with space shielding, Tooper[11] has studied the Störmer problem of a right, circular solenoid.

The present work, a treatment of the Störmer problem of three simple axially symmetric magnetic systems, reveals some remarkable properties of the Störmer forbidden regions, which could have application to cosmic-ray and trapped-radiation studies. The systems are not considered in combination with a dipole field, and the results, therefore, cannot be directly compared with geophysical data. However, the behavior of the Störmer regions is sufficiently complex that this intermediate step is valuable for a reasonably simple interpretation of later studies of the possible geomagnetic effects. Two of the three field geometries considered here are those due to two parallel, coaxial current loops with currents in the same sense (parallel dipole

[1] C. Störmer, *The Polar Aurora* (Clarendon Press, Oxford, England, 1955).
[2] S. B. Treiman, J. Geophys. Res. **89**, 130 (1953).
[3] E. C. Ray, Phys. Rev. **101**, 1142 (1956).
[4] S. Chapman and V. Ferraro, Terr. Mag. Atmos. Elec., **38**, 79 (1933); **45**, 245 (1940).
[5] S. I. Akasofu and J. C. Cain, J. Geophys. Res. **67**, 4078 (1962).
[6] S. I. Akasofu and S. J. Chapman, J. Geophys. Res. **66**, 1321 (1961).
[7] S. I. Akasofu and W. C. Lin, J. Geophys. Res. **68**, 973 (1963).
[8] P. J. Kellogg and J. R. Winckler, J. Geophys. Res. **66**, 3991 (1961).
[9] R. H. Levy, Avco-Everett Research Laboratory Research Report 106 (1961).
[10] C. F. Kooi, Lockheed Technical Report 8-94-64-2 (1964).
[11] R. F. Tooper, ASD-TDR-63-194 (1963).

moments) and in the opposite sense (antiparallel dipole moments). The third field configuration is the axially symmetric magnetic quadrupole which represents the limiting case for the antiparallel current loop system when, for constant loop dipole moment, system dimensions become infinitesimally small.

The parallel loop case, could represent two ring currents in the northern and southern hemispheres, respectively, or, equivalently and possibly more significantly, a system of counterflowing and somewhat separated electron and proton ring currents near the equatorial plane. The antiparallel case is included for comparison.

## STÖRMER PROBLEM

A brief review of the general Störmer problem and its application to the dipole case is in order. We proceed as usual from either the Lorentz force equation or the Hamiltonian representation for a charged particle of relativistic mass $m$, and charge $e$, moving in an axially symmetric magnetic field. The field is defined by a vector potential $\mathbf{A} = A_\phi\hat{\phi}$, where $\phi$ is the azimuth angle about the axis of symmetry. The coordinate system is shown in Fig. 1.

The nonrelativistic Hamiltonian is given by

$$\begin{aligned} H &= \frac{1}{2m}(\mathbf{p} - e\mathbf{A})^2 \\ &= \frac{1}{2m}\left[p_r^2 + \frac{p_\theta^2}{r^2} + \left(\frac{p_\phi}{r\sin\theta} - eA_\phi\right)^2\right]. \end{aligned} \tag{1}$$

The relativistic form leads to the same equations of motion if no particle energy losses occur. Applying Hamilton's canonical equations of motion, we obtain

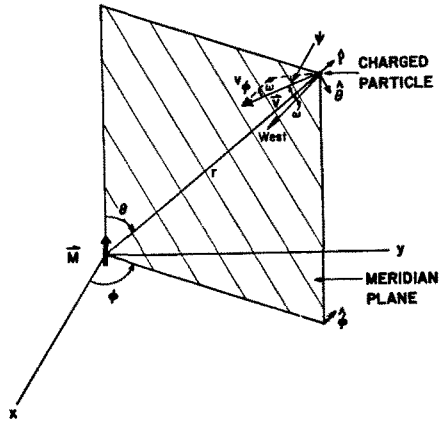


FIG. 1. Spherical coordinate system.

$$\begin{aligned} p_r &= m\dot{r}, \\ p_\theta &= mr^2\dot{\theta}, \\ p_\phi &= mr^2\sin^2\theta\dot{\phi} + er\sin\theta A_\phi, \\ \dot{p}_r &= \frac{p_\theta^2}{mr^3} + \frac{1}{m}\left(\frac{p_\phi}{r\sin\theta} - eA_\phi\right)\left(\frac{p_\phi}{r^2\sin\theta} + e\frac{\partial A_\phi}{\partial r}\right), \\ \dot{p}_\theta &= \frac{1}{m}\left(\frac{p_\phi}{r\sin\theta} - eA_\phi\right)\left(\frac{p_\phi\cos\theta}{r\sin^2\theta} + e\frac{\partial A_\phi}{\partial\theta}\right), \\ \dot{p}_\phi &= 0. \end{aligned} \tag{2}$$

The last member of this equation states that the component of angular momentum about the symmetry ($z$) axis is conserved. Rewriting the third member of (2) in terms of this constant of the motion, $P_\phi$, which we define to be $-2\gamma p$, we obtain

$$mr\sin\theta\dot{\phi} = -2\gamma p/r\sin\theta - eA_\phi. \tag{3}$$

A second constant of the motion is the energy or speed of the particle as evidenced by the fact that

$$\dot{H} = 0, \tag{4}$$

which implies that $H = E =$ constant. The $\phi$ component of the particle's velocity is $r\sin\theta\dot{\phi}$; the angle $\omega$ between west and the velocity vector is, therefore, given by

$$\cos\omega = -v_\phi/v = -r\sin\theta\dot{\phi}/v \equiv Q. \tag{5}$$

The velocity vector lies on a cone of directions, the so-called Störmer cone, with vertex at the observer and axis along the east–west direction. It is apparent that, for real motion,

$$|\cos\omega| = |Q| \leq 1. \tag{6}$$

We now introduce a normalizing unit of length, $C_{st}$, the Störmer distance, which is the radius of the trapped, circular orbit of a particle of charge e and momentum $p$ in the equatorial plane of a dipole of moment $M$. It is given by

$$C_{st}^2 = \mu_0 eM/4\pi p. \tag{7}$$

With this parameter we define dimensionless quantities

$$\begin{aligned} \rho &= r/C_{st}, \\ \lambda &= a/C_{st}, \\ \bar{\gamma} &= \gamma/C_{st}, \\ \bar{A}_\phi &= (ea^2/pC_{st}^2)A_\phi. \end{aligned} \tag{8}$$

Substitution of (5), (7), and (8) into (3) leads finally to

$$Q = (2\bar{\gamma}/\rho\sin\theta) + (\bar{A}_\phi/\lambda^2). \tag{9}$$

## MAGNETIC DIPOLE

Application of (9) to the dipole shielding problem has been discussed in considerable detail elsewhere.[1,3,12,13] If the vector potential for the dipole is normalized by means of the previously defined parameters,

$$\bar{A}_\phi = \lambda^2(\sin\theta/\rho^2). \tag{10}$$

Upon substitution into (9), we have

$$Q = 2\bar{\gamma}/\rho \sin\theta + \sin\theta/\rho^2. \tag{11}$$

The free parameter remaining in the equation is $\bar{\gamma}$, the normalized impact parameter. Plots of this equation are then made in $\rho$ and $\theta$ for $Q = \pm 1$ and for various values of $\bar{\gamma}$. These plots represent meridian-plane cross sections of the space around the dipole and are of course independent of azimuth angle, $\phi$. By the symmetry of the dipole field and the form of (11), the plots are symmetric under reflection in the equatorial plane. A typical set of these well known plots is shown in Fig. 2. The so-called forbidden regions in which $|Q| > 1$ and into which, therefore, particles cannot move are cross-hatched. Regions in which particles are allowed are unshaded. For $0 > \bar{\gamma} > -1.00$, a single allowed region connects the origin with infinity. When $\bar{\gamma} = -1.00$ the boundary $Q = -1$ just reaches the equatorial plane, and for all $\bar{\gamma} \leq -1.00$ there exists an inner allowed region which is separated from the outer allowed region by an arm of the outer forbidden region. In theory, each point of the allowed region is accessible to particles arriving from infinity. However, particles cannot cross the forbidden region and, consequently, unbound particles cannot penetrate into the inner allowed region. Likewise, particles which somehow find themselves in the inner allowed region are, in the ideal case, permanently trapped there. The actual trajectory of a particular particle can be determined only by numerical integration of the complete set of equations of motion, (2), but we are assured that whatever the particle's initial conditions, $\bar{\gamma}$, e, $m$, $E$, the trajectory, however complicated, will remain in the outer allowed region defined by that $\bar{\gamma}$.

As $\bar{\gamma}$ increases monotonically from large negative

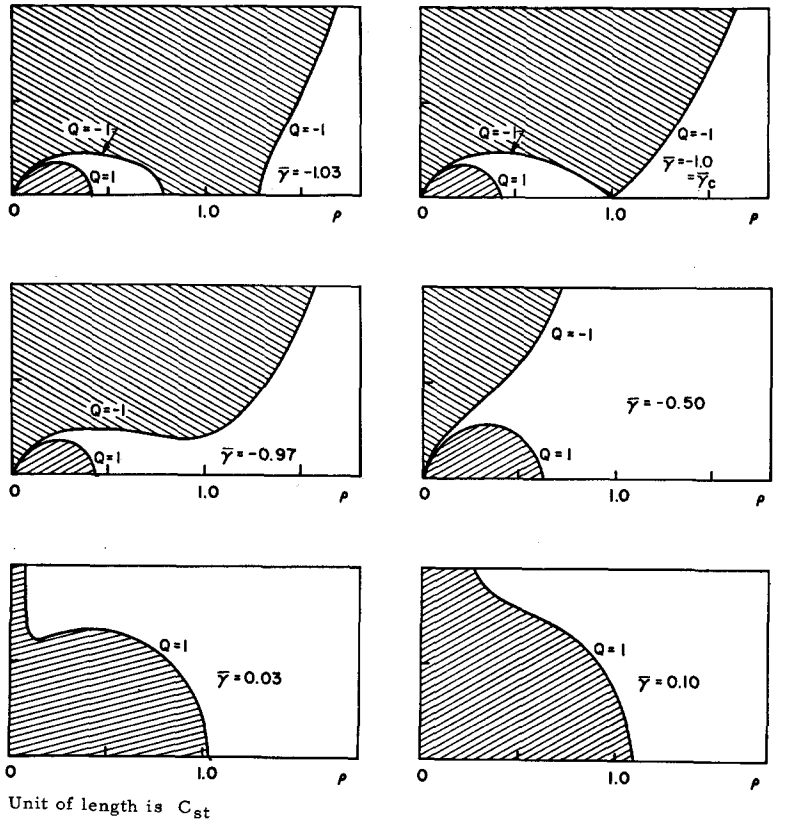


FIG. 2. Forbidden regions for the magnetic dipole.

[12] M. S. Vallarta, *Handbuch der Physik*, edited by S. Flugge (Springer–Verlag, Berlin, 1961), Vol. XLVI, pp. 88–129.
[13] T. H. Johnson, Rev. Mod. Phys. **10**, 193 (1938).

values through $-1.00$, then zero, to large positive values, the boundary $Q = 1$ of the inner forbidden region expands monotonically. We can, therefore, conclude that the critical condition, $\bar{\gamma} = -1.00$, specifies the largest inner allowed region totally forbidden to particles from infinity. In other words, the particular inner forbidden region, outlined by the curve $Q = 1$ which occurs at $\bar{\gamma} = -1.00$, lies entirely within all inner forbidden regions for $\bar{\gamma} > -1.0$ (outer forbidden region open). Upon substitution of $Q = 1$, and $\bar{\gamma} = -1.00$, Eq. (11) gives the boundary in $\rho$ and $\theta$ of the totally shielded region. The minimum energy of a particle reaching a $(r, \theta)$ in space [or $(r_e, \theta)$ on the surface of the earth] can then be found from (7). Alternatively, similar substitution of $Q = 0$ would define minimum energy for vertical incidence.

## GEOGRAPHICAL ANALOG OF THE $Q$ SURFACE

It is of interest to examine in general the nature of the critical point which defines the condition of maximum exclusion of unbound particles. Störmer[1] pointed out that the $Q$ = constant curves can be looked upon as equipotential or contour lines in a geographical analog. If we define a new parameter,

$$\tilde{Q} = Q^2 = [2\bar{\gamma}/\rho \sin\theta + \bar{A}_\phi/\lambda^2]^2, \tag{12}$$

we can easily show from (2) that

$$\begin{aligned} \frac{1}{2}\frac{\partial \tilde{Q}}{\partial \rho} &= -\frac{C_{st}^2}{v^2}(\ddot{\rho} - \rho\dot{\theta}^2) \\ \frac{1}{2\rho}\frac{\partial \tilde{Q}}{\partial \theta} &= -\frac{C_{st}^2}{v^2}(\rho\ddot{\theta} + 2\dot{\rho}\dot{\theta}). \end{aligned} \tag{13}$$

If we now assume that $\frac{1}{2}\tilde{Q}$ is some scalar potential function, the force experienced by an entity upon which the potential field acts is given by

$$\mathbf{F} = -\nabla(\tfrac{1}{2}\tilde{Q}), \tag{14}$$

and we can write

$$\begin{aligned} F_\rho &= -\partial/\partial\rho(\tfrac{1}{2}\tilde{Q}) = n(\ddot{\rho} - \rho\dot{\theta}^2), \\ F_\theta &= -\frac{1}{\rho}\frac{\partial}{\partial\theta}(\tfrac{1}{2}\tilde{Q}) = n(\rho\ddot{\theta} + 2\dot{\rho}\dot{\theta}). \end{aligned} \tag{15}$$

A comparison of (13) and (15) leads to the conclusion that we can treat the motion of a charged particle in the meridian plane as the motion of a pseudo particle with "mass" $n = C_{st}^2/v^2$ in a "gravitational" potential field

$$V = \tfrac{1}{2}\tilde{Q}, \tag{16}$$

whose equipotential or contour lines are given by curves of constant $\tilde{Q}$ or, therefore, by curves of constant $Q$. The force on the pseudo particle in this "landscape" is perpendicular to the equipotentials and is directed toward the curve $Q = 0$; the pseudo particle cannot move into a region in which $\tilde{Q} > 1$ or $|Q| > 1$. The projection on the $Q = 0$ plane of the path of a pseudo particle in this geographical analog will be the projection in the meridian plane of the trajectory of a real particle with equivalent initial conditions.

## SADDLE-POINT CONDITION

Störmer employed this analog for intuitive interpretation of the trajectories he calculated, but in the case of magnetic geometries more complex than a pure dipole, another important use exists. From Fig. 3, a plot of $Q$ = constant for $\bar{\gamma} = -1.00$, it appears that the so-called "pass point," by way of which an unbound particle approaches the vicinity of the dipole when $\bar{\gamma} > -1.00$, is of the form of a saddle point in the $Q$ surface. If this is so, then the critical pass or saddle point at $\bar{\gamma}_c = -1.00$, defining first exclusion of unbound particles from the inner allowed region, should be derivable from the mathematical conditions for a saddle point and the fact that $Q = -1$ there.

A saddle point is defined for some function $F(\rho, \theta)$ by the following conditions:

$$\begin{aligned} \partial F/\partial\rho &= 0, \\ (1/\rho)(\partial F/\partial\theta) &= 0, \end{aligned} \tag{17}$$

$$\left[\frac{1}{\rho}\frac{\partial^2 F}{\partial\rho\,\partial\theta}\right]^2 - \left[\frac{\partial^2 F}{\partial\rho^2}\right]\left[\frac{1}{\rho^2}\frac{\partial^2 F}{\partial\theta^2}\right] > 0. \tag{18}$$

Applying (17) to (11) we obtain

$$-\frac{2\bar{\gamma}_c}{\rho_c^2 \sin\theta_c} + \frac{1}{\lambda^2}\frac{\partial \bar{A}_\phi}{\partial\rho}\bigg|_{\substack{\rho=\rho_c \\ \theta=\theta_c}} = 0, \tag{19}$$

$$-\frac{2\bar{\gamma}_c \cos\theta_c}{\rho_c \sin^2\theta_c} + \frac{1}{\lambda^2}\frac{\partial \bar{A}_\phi}{\partial\theta}\bigg|_{\substack{\rho=\rho_c \\ \theta=\theta_c}} = 0. \tag{20}$$

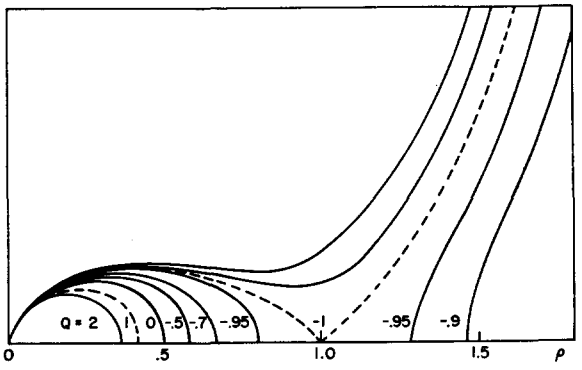


FIG. 3. Contour map of curves; $Q$ = constant for $\bar{\gamma} = -1.00$, dipole.

The condition that $Q = -1$ gives a third equation,

$$\frac{2\bar{\gamma}_0}{\rho_0 \sin \theta_0} + \frac{1}{\lambda^2} \bar{A}_\phi \Big|_{\substack{\rho=\rho_0 \\ \theta=\theta_0}} = -1. \tag{21}$$

We now have three simultaneous equations which may, in principle, be solved in closed form to give $\rho_0$, $\theta_0$, and $\bar{\gamma}_0$. The final test is, then, that the inequality (18) is also satisfied.

Applying (19)–(21) to the dipole, the well known critical saddle point values are obtained:

$$\bar{\gamma}_0 = -1.00, \qquad \rho_0 = 1.00, \qquad \theta_0 = \tfrac{1}{2}\pi; \tag{22}$$

the inequality is of course satisfied by these values. The dipole is an almost trivial application of the saddle point condition. The simple symmetry of the vector potential ensures that $\partial Q/\partial\theta = 0$ everywhere on the equatorial plane. The same is true for the somewhat more complex cases of the single current ring[9] and the dipole plus ring current.[3] The obvious approach is to assume that $\theta_0 = \frac{1}{2}\pi$ and to solve (19) and (21) for $\rho_0$ and $\bar{\gamma}_0$. Previous treatments have employed just this approach, but without noting the existence of a saddle point. However, there are more complex magnetic geometries which possess critical points for $Q = -1$ out of the equatorial plane, and which may also have interesting saddle points on the $Q = 1$ boundary. For these cases all three equations (19)–(21) and the inequality must be used.

## VECTOR POTENTIAL OF SINGLE AND DOUBLE RING CURRENTS

Consider a circular ring of current of infinitesimal cross section, oriented parallel to the equatorial plane, with radius $a$, and center at a distance $b$ from the origin along the axis of symmetry, and carrying a current $I$. The vector potential of this ring at a point $(r, \theta)$ is given by

$$A_\phi = \frac{\mu_0 I a^{\frac{1}{2}}}{2\pi} \frac{k^3 C(k)}{(r \sin \theta)^{\frac{1}{2}}}, \tag{23}$$

where

$$k^2 = \frac{4ar \sin \theta}{(r \sin \theta + a)^2 + (r \cos \theta - b)^2},$$

$$C(k) = 1/k^4[(2 - k^2)K(k) - 2E(k)],$$

and $K$, $E$, and $C$ are complete elliptic integrals.[14] The vector potential of a combination of two such loops, located at $+b$ and $-b$ on the axis of symmetry, is given by

$$A_\phi = \frac{\mu_0}{2\pi(\rho \sin \theta)^{\frac{1}{2}}} [I_1 a_1^{\frac{1}{2}} k_1^3 C_1(k_1) \pm I_2 a_2^{\frac{1}{2}} k_2^3 C_2(k_2)], \tag{24}$$

the upper sign corresponding to currents of the same sense (parallel) the lower sign to currents of the opposite sense (antiparallel). Also,

$$k_{1,2}^2 = \frac{4a_{1,2} r \sin \theta}{(r \sin \theta + a_{1,2})^2 + (r \cos \theta \mp b)^2}, \tag{25}$$

the subscript 1 and the upper sign referring in both parallel and antiparallel cases to the upper loop, and the subscript 2 and the lower sign to the lower loop. If $a_1 = a_2 = a$ and $I_1 = I_2 = \frac{1}{2}I$, and if we define a new dimensionless parameter $\eta = b/C_{sl}$, we obtain

$$\bar{A}_\phi = \frac{1}{\pi} \frac{\lambda^{\frac{1}{2}}}{(\rho \sin \theta)^{\frac{1}{2}}} [k_1^3 C_1 \pm k_2^3 C_2],$$

$$k_{1,2}^2 = \frac{4\lambda\rho \sin \theta}{(\rho \sin \theta + \lambda)^2 + (\rho \cos \theta \mp \eta)^2}. \tag{26}$$

This will reduce, respectively, to a single loop carrying current $I$, or to no current at all, when $\eta = 0$.

## AXIAL QUADRUPOLE

If we take the limit as $\rho \gg \lambda$ of the vector-potential expression for a single loop, (23), by means of the series approximation for $C$ as given by Jahnke and Emde,[14] we obtain (10) as should be expected. Similarly if we take the limit as $\rho \gg \lambda$ and $\rho \gg \eta$ in (26), the upper sign leads again to (10), while the lower sign leads to

$$\bar{A}_\phi = \frac{3\lambda^2 \eta \sin 2\theta}{2\rho^3}, \tag{27}$$

the dimensionless vector potential of an axial magnetic quadrupole oriented along the $z$ axis. Consideration of the Störmer problem for this configuration leads to some interesting conclusions. To remove the infinitesimal factor $\eta$ in (27), we redefine $C_{sl}$ in terms of the quadrupole moment $2bM$ rather than the dipole moment $M$. $C_{sl}$ is no longer the radius of gyration of a particle trapped in a circular orbit in the equatorial plane, but its continued use is convenient. This redefinition leads to the replacement of $\eta$ by the factor $\frac{1}{2}$ and

$$\bar{A}_\phi = \frac{3\lambda^2 \sin 2\theta}{4\rho^3}. \tag{28}$$

If we insert (28) into (9) and apply the saddle-point condition (19)–(21), we find that

$$\theta_0 = 63.433°, \qquad \rho_0 = 1.063, \qquad \bar{\gamma}_0 = -0.7125.$$

[14] E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, Inc., New York, 1945).

A series of typical Störmer forbidden-zone plots are shown in Fig. 4. They are not symmetric in the equatorial plane, and here for the first time is a critical point out of the equatorial plane in the first (and second) quadrant. Note that for $\bar{\gamma} > 0$ the curves $Q = \pm 1$ play the same roles (boundaries of the inner and outer forbidden regions, respectively) as they do for the dipole and single current loop cases. At $\bar{\gamma} = 0$ the $Q = -1$ curve has not disappeared as it did for the dipole and the regions are symmetric. For $\bar{\gamma} > 0$ the $Q = \pm 1$ curves have exchanged roles, the regions repeat their behavior for $\bar{\gamma} < 0$, and another critical point exists at

$$\theta_c^* = 116.567°,$$

$$\rho_c^* = 1.063 = \rho_c,$$

$$\bar{\gamma}_c^* = 0.7125 = -\bar{\gamma}_c.$$

It can readily be seen that when $\bar{\gamma} = \bar{\gamma}_c$ the inner allowed region defined by $Q = 1$ is a maximum shielded region. Moreover this same region obviously lies entirely within the outer forbidden region which exists for $\bar{\gamma} \geq \bar{\gamma}_c^*$. The same argument applies to the inner allowed region which occurs for $\bar{\gamma} = \bar{\gamma}_c^*$. We may, therefore, conclude that the symmetric region, composed of both inner forbidden regions, is totally shielded for all $\bar{\gamma}$ such that $|\bar{\gamma}| \geq |\bar{\gamma}_c|$.

## DOUBLE RING CURRENTS

If in (26) we make the further parameter definition that $\mu = \eta/\lambda = b/a$ and $\rho' = \rho/\lambda = r/a$ and substitute, we obtain

$$\bar{A}_\phi = \frac{1}{\pi} \frac{1}{(\rho' \sin \theta)^{\frac{1}{2}}} [k_1^3 C_1 \pm k_2^3 C_2], \tag{29}$$

where

$$k_{1,2}^2 = \frac{4\rho' \sin \theta}{(\rho' \sin \theta + 1)^2 + (\rho' \cos \theta \mp \mu)^2}, \tag{30}$$

and (9) becomes

$$Q = \frac{2\bar{\gamma}}{\rho' \lambda \sin \theta} + \frac{k_1^3 C_1 \pm k_2^3 C_2}{\pi \lambda^2 (\rho' \sin \theta)^{\frac{1}{2}}}. \tag{31}$$

The two new parameters have the significance that the loops lie in the intersection of a cylinder of radius $\rho' = 1$ and a cone of half-angle $\alpha = \cot^{-1}\mu$, the $z$ axis being the axis of both cylinder and cone. It is important to keep in mind the difference in meaning between the double signs in (29) and those in (30). Application of the saddle-point condition to (34) leads to three simultaneous equations in six variables:

$$Q\lambda^2 = \frac{1}{2\pi(\rho_c' \sin \theta_c)^{\frac{1}{2}}} \left\{ G_\pm + \frac{\mu^2 - \rho_c'^2 + 1}{4\rho_c' \sin \theta_c} H_\pm \right\}, \tag{32}$$

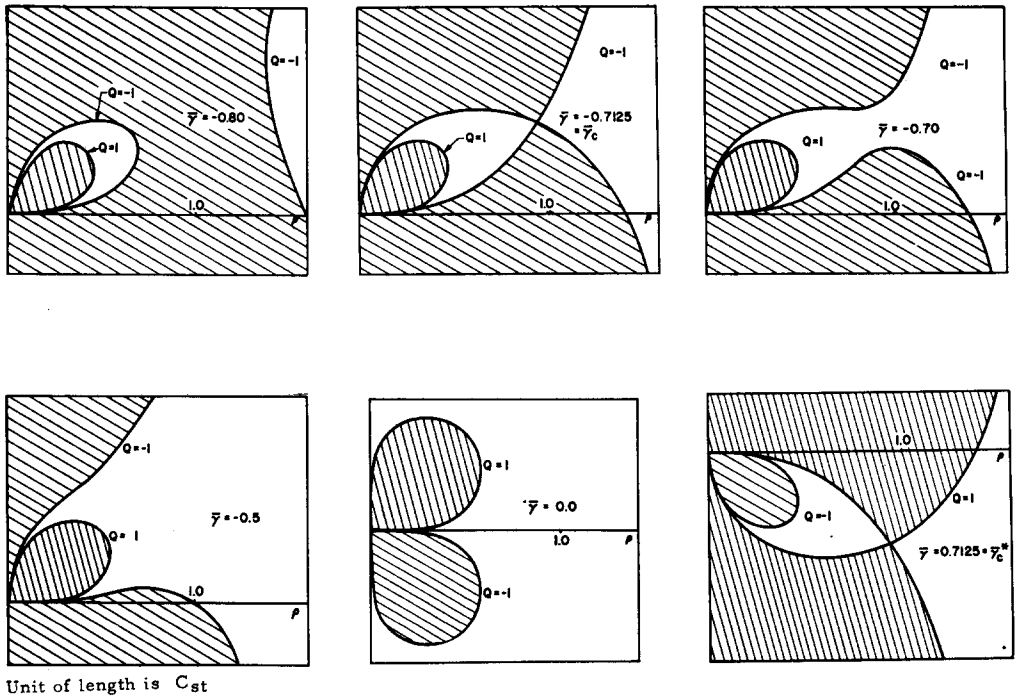


FIG. 4. Forbidden regions for the magnetic quadrupole.

$$-\bar{\gamma}_0\lambda = \frac{(\rho_0' \sin\theta_0)^{\frac{1}{2}}}{4\pi}\left\{G_\pm - \frac{\mu^2 - \rho_0'^2 + 1}{4\rho_0' \sin\theta_0} H_\pm\right\}, \quad (33)$$

$$-\bar{\gamma}_0\lambda = \frac{(\rho_0' \sin\theta_0)^{\frac{1}{2}}}{4\pi}\left\{G_\pm - \frac{\mu^2 + \rho_0'^2 + 1}{4\rho_0' \sin\theta_0} H_\pm + \frac{\mu}{2 \sin\theta_0 \cos\theta_0} H_\mp\right\}, \quad (34)$$

where $G_\pm = k_1^3 C_1 \pm k_2^3 C_2$, $H_\pm = [k_1^5/(1 - k_1^2)](D_1 - C_1) \pm [k_2^5/(1 - k_2^2)](D_2 - C_2)$, $D_i = D(k_i)$ is another complete elliptic integral,[14] and $H_\mp$ is like $H_\pm$ but with an inversion of the double signs. Equation (33) defines the condition $\partial Q/\partial\rho' = 0$ while (34) comes from $\partial Q/\partial\theta = 0$. The inequality condition has been derived from (18) but is not shown.

## PARALLEL CASE

Our problem now consists of choosing the upper signs in (32)–(34) and finding consistent solutions to these three equations and the inequality for the outer forbidden region boundary $Q = -1$,—or for the inner forbidden region boundary $Q = 1$,—for a range of values of $\theta_0$, $\rho_0'$, $\bar{\gamma}_0$, $\mu$, and $\lambda$. The most straightforward method is to specify $\mu$, assume a $\theta_0$, and solve for $\rho_0'$, $\lambda$, and $\bar{\gamma}_0$. The results are shown in Fig. 5 which is a set of curves of constant $\mu$, plotted as a function of $\lambda$ and $\bar{\gamma}_0$. For clarity the smooth variation of $\rho_0'$ along the curves is not indicated. The limiting curve for $\mu = 0$, given by Levy,[9] is included.

Each curve of constant $\mu \neq 0$ is actually composed of three branches. The branches for $\mu = 1.00$ are typical and their behavior is shown in Figs. 5 and 6. One branch (I) is defined by $\theta_0 \neq \frac{1}{2}\pi$ and the variation of $\theta_0$ is indicated. The critical $\theta$ varies between a minimum given by the half-angle, $\alpha = \cot^{-1}\mu$, of the cone on which the loop lies, and a maximum which approaches the limit $\frac{1}{2}\pi$. Each point on branch I is a simultaneous solution of (32)–(34) and the inequality.

The remaining two branches give critical points in the equatorial plane, $\theta_0 = \frac{1}{2}\pi$. Equations (33) and (34) do not intersect for $\theta = \frac{1}{2}\pi$; these two branches define simultaneous solutions of only (32) and (33) and the inequality. The symmetry of the vector potential guarantees that for $\theta = \frac{1}{2}\pi$, $\partial Q/\partial\rho' = 0$ for all $\rho_0'$, and the lack of information from (34) is unimportant. One of the two equatorial-plane branches (II) specifies inner saddle points on the $Q = -1$ equipotential, as in Fig. 7(a). While this inner saddle point is not important to our unbound radiation problem, it is interesting. The third branch (III) defines outer critical points which correspond to and resemble the single critical point found in the dipole and single loop cases.[9]

The properties of the forbidden regions change rather rapidly near the common intersection of the three branches; this area is seen in greater detail in Fig. 6. $\rho_0'$ increases monotonically along branches II and III from points A to D to H, etc. A few representative Störmer plots taken from the three branches are shown in Fig. 7.

In the region from B to D on II, the inequality is not satisfied, indicating that a small relative

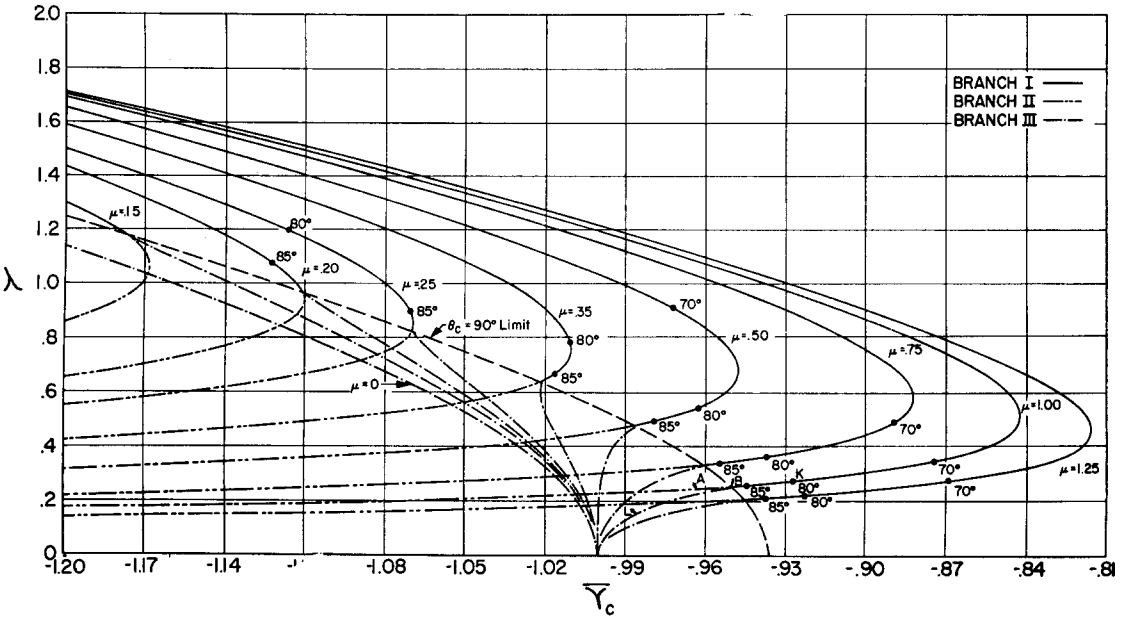


FIG. 5. Plot of $\lambda$ vs $\bar{\gamma}_0$ as a function of $\mu$ for $Q = -1$; double parallel ring currents.

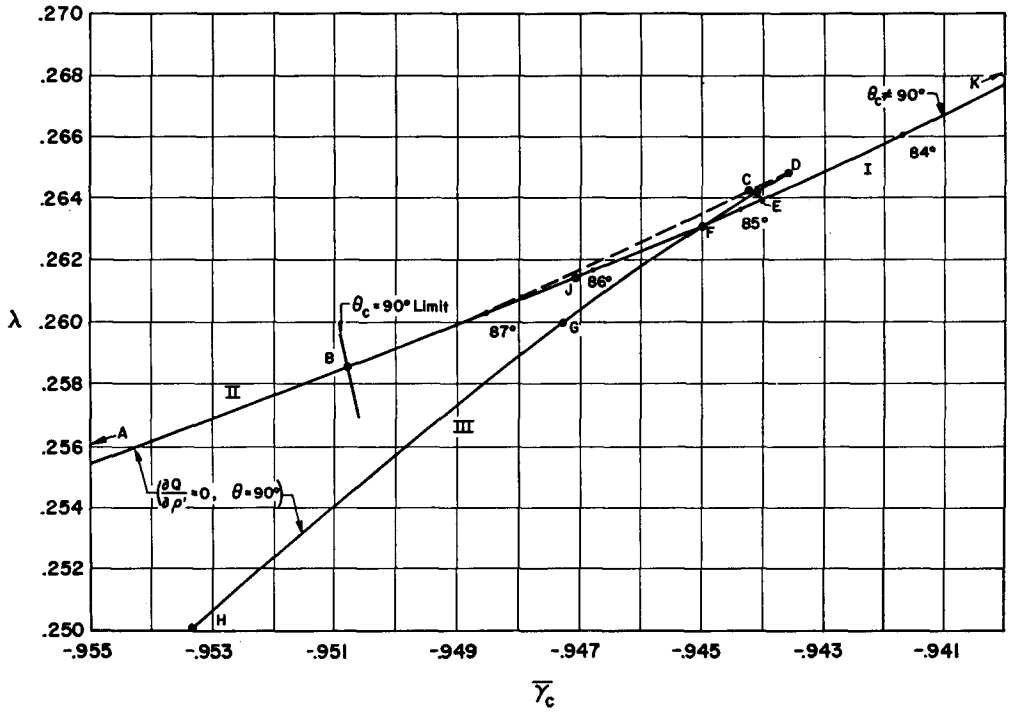


FIG. 6. Detail of $\mu = 1.00$ curve for $Q = -1$, double parallel ring currents.

maximum or minimum (but no saddle point) exists on the $Q = -1$ curve. This is demonstrated in Fig. 7(b) as a disconnected, infinitesimal $Q = -1$ region in the equatorial plane. At point F two simultaneous saddle points exist, an inner one at $\theta_c = 85.128°$ and an outer one at $\theta_c = 90°$. Figure 7(c) illustrates this situation which is analogous to dipole plus ring behavior noted by Ray[3] and Kellogg and Winckler.[8] Four separate allowed regions and three separate forbidden regions can be seen to exist for this con-

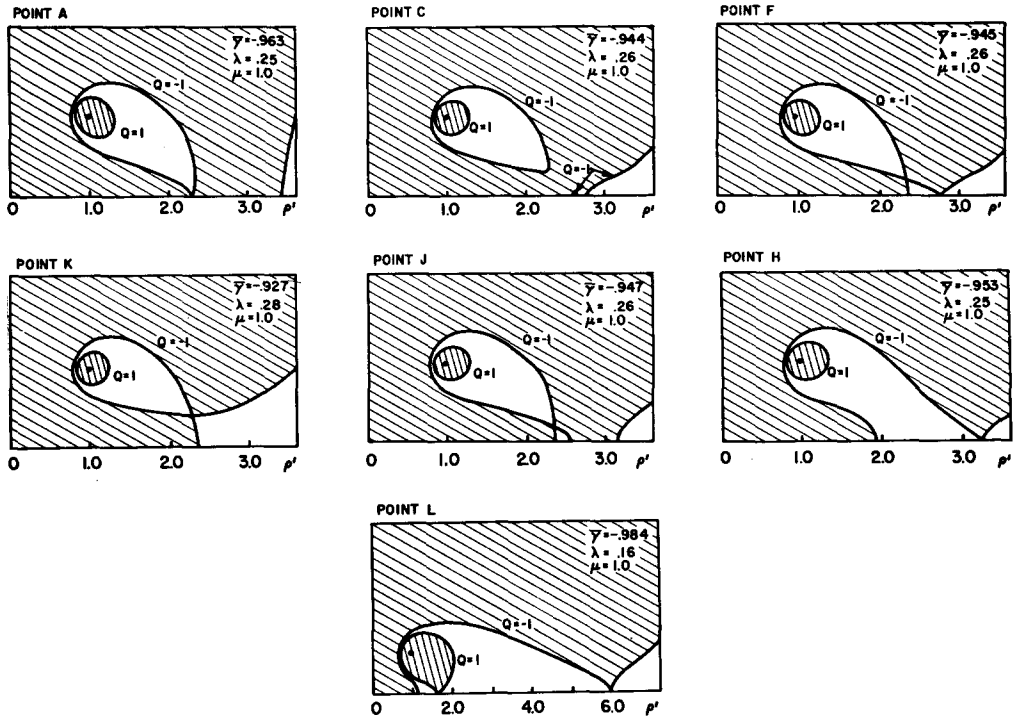


FIG. 7. Critical forbidden regions in the $(\rho, \theta)$ plane for double parallel ring currents, $\mu = 1.00$.

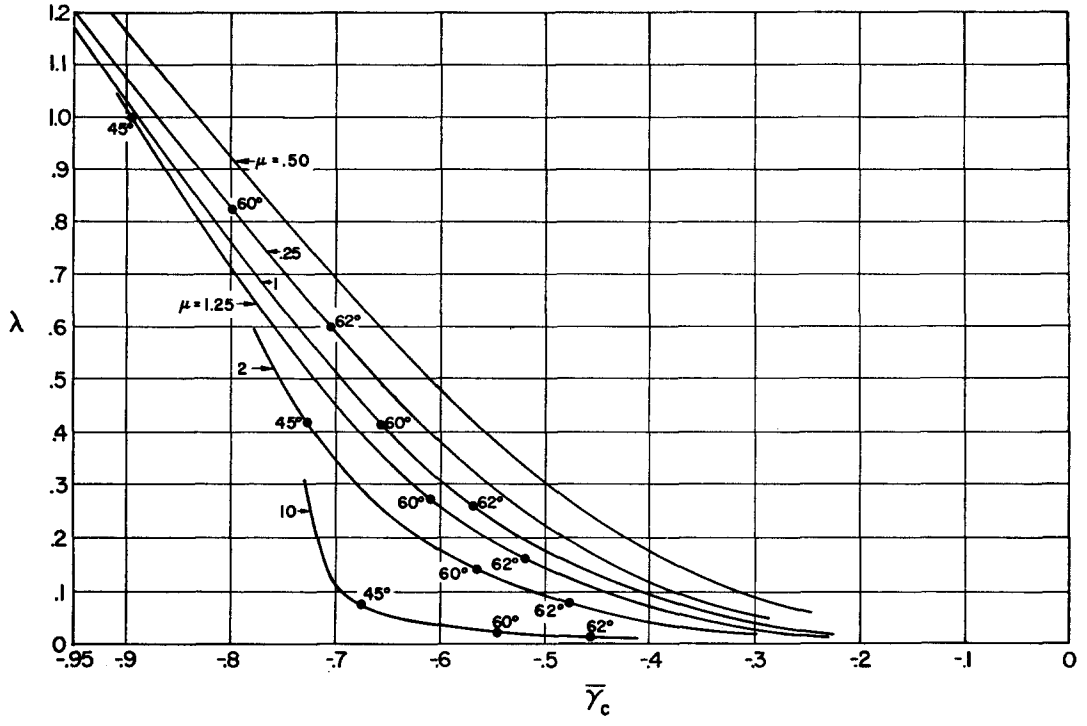


FIG. 8. Plot of λ vs $\bar{\gamma}_c$ as a function of $\mu$ for $Q = -1$; double antiparallel ring currents.

dition. If we move away from F along one of the four possible critical paths, one or the other of the critical points will disappear. Moving along branch I toward J or along branch III toward E produces isolated critical points which do not, even when open, allow entrance of unbound particles. Moving along branch I toward K or along branch III toward G produces outer critical points which, when slightly open, allow

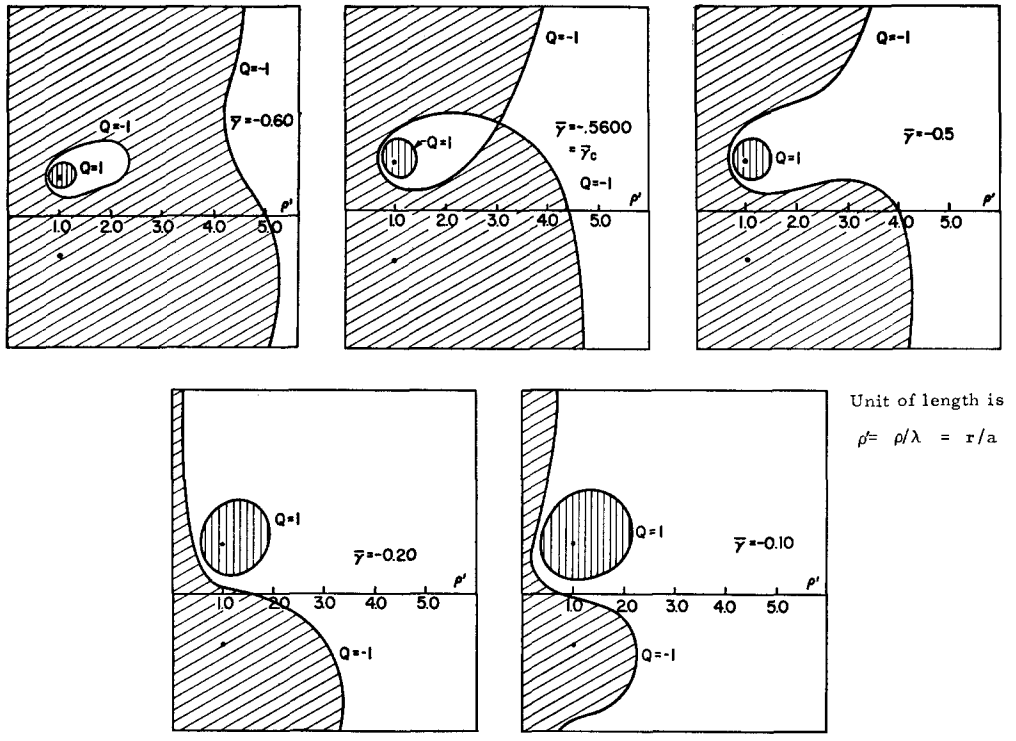


FIG. 9. Forbidden regions for double antiparallel ring currents, $\mu = 1.00$, $\lambda = 0.25$.

unbound particles to enter the innermost allowed region. The important critical parameters are, therefore, given by the composite broken curve H to F to K, etc. The ($\partial Q/\partial \rho' = 0$, $Q = 1$) curve is not shown, but it intersects branch III in another interesting point (L) which defines simultaneous inner and outer forbidden region saddle points in the equatorial plane [Fig. 7(g)].

## ANTIPARALLEL CASE

For the antiparallel current case the lower sign in (29) is chosen, and we proceed as before. The asysmmetry of the configuration indicates that, as with the quadrupole, no critical points can exist in the equatorial plane. The resulting curves, shown in Fig. 8, are much simpler than those for the parallel-current case. Typical forbidden-region plots for $\mu = 1.00$ are given in Fig. 9. The same conclusions concerning maximum shielded regions developed for the quadrupole apply to this configuration. If $\bar{\gamma}_0$ defines the critical point for some combination of $\mu$ and $\lambda$, then the total maximum shielded region is twice the inner forbidden region for that $\bar{\gamma}_0$; it is totally shielded for all $\bar{\gamma}$ such that $|\bar{\gamma}| \geq |\bar{\gamma}_0|$.

## CONCLUSION

The preceding technique for studying the shielding due to complex, axi-symmetric magnetic configurations will certainly be no more exact in predicting the behavior of the allowed cone for these magnetic configurations than is the simple Störmer method for the dipole. However, knowledge of the existence of multiple critical points, of multiple allowed and forbidden regions, and of the lack of symmetry in the equatorial plane of the Störmer plots for certain fields should be valuable in interpretation of more exact treatments which employ Lemaitre–Vallarta[12] theory and which account for shadow effects.

Immediate applications of this saddle-point method to configurations of possible geomagnetic interest are planned. These include dipole plus quadrupole and dipole plus double ring currents. It can be predicted from the above calculations that the dipole plus quadrupole will not have critical pass points in the equatorial plane for any nonzero ratio of quadrupole moment to dipole moment. The dipole plus double ring current combination will have multiple critical points whose behavior will be even more complicated than in the double loop case.

# Isotopic Space, Complex Conjugation, and $U_3$ Symmetry for Particles

S. TEITLER
*U. S. Naval Research Laboratory, Washington, D. C.*
(Received 3 May 1965)

The inherent $U_4$ symmetry over spinor spaces of the regular decomposition indicated previously is discussed from the viewpoint of a single spinor space. It is noted that this symmetry does not yet include isotopic spin. Isotopic space is introduced by considering that the space–time vectors and its vector Clifford algebra are imbedded in the algebra of a six-dimensional Euclidean space. Three of the latter's dimensions are identified with ordinary space, and three are identified with isotopic space. The "time vector" corresponds to the pseudoscalar element in isospace, and the imaginary unit is expressed as a linear combination of the bivectors in isospace. The relation among complex conjugation, time inversion, and space inversion is thereby clarified. Some comments on the application of these ideas to particle symmetry discussions are made. In particular one obtains as the first extension of the $SU_2$ symmetry of isotopic spin, a $U_3$ symmetry for particles.

## INTRODUCTION

IN previous work[1], we have used the vector Clifford algebra, $C_{16}$, generated with the basis vectors of Lorentz space–time $L_4$ for the description of one-particle fields in which 4-spinors are identified as elements of the minimal ideals of $C_{16}$. In II and III, these minimal ideals were enumerated and classified. Further, in III, two classes of unitary symmetries inherent in this formulation were indicated. These are (1) the $SU_4$ symmetry among the components of a given spinor (which is in general not compatible with Lorentz symmetry) and (2) the $U_4$ symmetry among the spinor spaces belonging to the regular decomposition of $C_{16}$ (which is compatible with Lorentz symmetry). In this paper, we shall not concern ourselves with the unitary symmetry of class (1), i.e., the spin unitary symmetry.

As may be recalled from III, the regular decomposition consisted of expressing any member of $C_{16}$ in terms of the four minimal ideal bases generated from the mutually annihilating idempotents which may be formed from the basis elements of a four-component commuting subalgebra of $C_{16}$. The $U_4$ symmetry over these minimal ideals (spinor spaces) leads to the invariance of the regular current density

$$J^\mu = \sum_\alpha (\psi^{\alpha\dagger} e^\mu \psi^\alpha)_s \qquad (\alpha = a, b, c, d), \tag{1.1}$$

where the sum is over the four spinor spaces so that each $\psi^\alpha$ is a four-component element of a minimal left ideal (or spinor). The subscript s means the scalar part of the Clifford product.

This $U_4$ symmetry over the spinor spaces of the regular decomposition means that we consider unitary combinations over spinors from each space each of which satisfies the Dirac-like equation in its respective space

$$[\kappa - i \sum_\mu e^\mu(\partial_\mu - iqA_\mu)]\psi^\alpha = 0. \tag{1.2}$$

Here $\kappa = mc/\hbar$ and $q$ stands for the (algebraic charge)/$\hbar c$. However, each of these spinors has an image in every other space of the regular decomposition which we obtain by transformation of its minimal ideal basis (and possibly complex conjugation of its coefficients). The image however does not in general satisfy the same Dirac-like equation.

Thus, consider for specificity, a standard type minimal left ideal basis

$$\begin{aligned} l_0^a &= \tfrac{1}{4}(1 - e^0 + ie^{12} - ie^{012}), \\ l_1^a &= \tfrac{1}{4}(-e^{13} + e^{013} + ie^{23} - ie^{023}), \\ l_2^a &= \tfrac{1}{4}(-ie^3 - ie^{03} + e^{123} + e^{0123}), \\ l_3^a &= \tfrac{1}{4}(-ie^1 - ie^{01} - e^2 - e^{02}). \end{aligned} \tag{1.3}$$

Then, for example, a suitable regular decomposition may be generated from (1.3) by the operations of complex conjugation $K$, $(l_\mu^b)$; the combination of time inversion $T$, space inversion $P$ and $K$, $(l_\mu^c)$; and finally the combination of $T$ and $P$, $(l_\mu^d)$. Parenthetically it should be noted that we consider these operations, except for $K$, to leave the spinor coefficients unchanged. These coefficients may be considered as functions of space–time coordinate labels which are the quantity of measurement along a given set of coordinate axes. The operations of space or time inversion, or any other change of coordinate axes under which the form of the spinor coefficient is unchanged means that its coordinate label arguments now refer to a measurement along

[1] S. Teitler, Nuovo Cimento Suppl. 3, 1 (1965). (designated I); those papers designated II and III are as yet unpublished.

the new set of axes. Thus for a spinor coefficient $\psi(x_0, \cdots)$, where $x_0$ is a position along the $e^0$-axes, time inversion of $L_4$ basis vectors leaves the same $\psi(x_0, \cdots)$; but now the argument provides a position along the $(-e^0)$ axes.

Returning now to our discussion of the relation among spinors in different spinor spaces of the regular decomposition we note for example that a spinor with basis $l_\mu^a$ which satisfies the Dirac-like equation has an image in the spinor space with basis $l_\mu^a$ which has complex conjugate coefficients and satisfies the charge conjugate Dirac-like equation. It may be recalled that in II, we defined $TPK$ as the charge conjugation operator since the application of $TPK$ to the Dirac-like equation yielded the charge conjugate equation albeit generally in another spinor space. We are following a somewhat different procedure here. By starting with a spinor with a "charge conjugate" basis which satisfies the Dirac-like equation, we obtain its image in the original spinor space which has complex conjugate coefficients and corresponds to the usual charge conjugate spinor. Similarly the image in the spinor space with basis $l_\mu^a$ obtained from the application of $TP$ to spinors satisfying Eq. (1.2) in the spinor space with basis $l_\mu^d$ satisfies the mass conjugate Dirac-like equation. Finally the $K$ image of spinors satisfying Eq. (1.2) in the $l_\mu^b$ spinor space satisfies the mass–charge conjugate Dirac-like equation in the $l_\mu^a$ spinor space.

For other types of minimal ideals, e.g., the Weyl and Majorana, the operations $K$, $TPK$, and $TP$ do not generate the complete set of minimal ideal bases of the regular decomposition and the operations $P$, $PK$, $T$, and $TK$ may be used. Note that for a standard-type minimal ideal these latter operations give a complete set but with a phase shift of $e^{i\pi}$ between the $(l_0, l_1)$ and $(l_2, l_3)$ components relative to the operations $(1, K, TPK, TP)$. In the original spinor space, their images satisfy the correspondingly altered Dirac-like equation (see Table I). One might consider a unitary symmetry over the eight (possibly redundant) spinor classes in a single spinor space which satisfy, respectively, the eight Dirac-like equations in order to allow for the equivalent of complete regular decomposition independently of the choice of minimal ideal type. However, as we shall argue below in Sec. III, the physically expected unitary symmetry should be restricted to at most $U_4$.

We note that while the operations 1, $TP$, $TPK$, $K$, etc. provide for spinors which satisfy respectively the ordinary, the mass conjugate, the charge conjugate, and the combined charge–mass conjugate Dirac-like equations and their space inverted counterparts, there seems to be no consistent way to obtain isotopic charge change in this manner. It seems then that despite the plethora of inherent unitary symmetries revealed in the $C_{16}$ description of spinor fields, we must look elsewhere for isotopic spin. Fortunately, there is a natural way to extend the formalism which is motivated by the desire to treat, in a unified manner, operations involving both space–time and complex conjugation and to remove the indefinite metric as a starting point. Thus we seek to widen our viewpoint (and our algebra) to allow consideration of $K$ on a similar basis to $T$ and $P$. We accomplish this by following the ideas of a previous discussion[2] concerning the basis of space–time in the next section.

TABLE I. Dirac-like operators for the respective images in the original or generating spinor space.

| Operation | Image Dirac-like operator |
|---|---|
| 1 | $O_D \equiv \left[\kappa - i\sum_{\mu=0}^{3} e^\mu(\partial_\mu - iqA_\mu)\right]$; $\kappa = \frac{mc}{\hbar}$, $\frac{q}{\hbar c} \to q$ |
| $TPK$ | $O_D^{TPK} \equiv \left[\kappa - i\sum_{\mu=0}^{3} e^\mu(\partial_\mu + iqA_\mu)\right]$ |
| $TP$ | $O_D^{TP} \equiv \left[\kappa + i\sum_{\mu=0}^{3} e^\mu(\partial_\mu - iqA_\mu)\right]$ |
| $K$ | $O_D^{K} \equiv \left[\kappa + i\sum_{\mu=0}^{3} e^\mu(\partial_\mu + iqA_\mu)\right]$ |
| $P$ | $O_D^{P} \equiv \left[\kappa - ie^0(\partial_0 - iqA_0) + i\sum_{k=1}^{3} e^k(\partial_k - iqA_k)\right]$ |
| $PK$ | $O_D^{PK} \equiv \left[\kappa + ie^0(\partial_0 + iqA_0) - i\sum_{k=1}^{3} e^k(\partial_k + iqA_k)\right]$ |
| $TK$ | $O_D^{TK} \equiv \left[\kappa - ie^0(\partial_0 + iqA_0) + i\sum_{k=1}^{3} e^k(\partial_k + iqA_k)\right]$ |
| $T$ | $O_D^{T} \equiv \left[\kappa + ie^0(\partial_0 - iqA_0) - i\sum_{k=1}^{3} e^k(\partial_k - iqA_k)\right]$ |

## II. CONCERNING THE BASIS FOR SPACE-TIME

While it has been very fruitful to view the time basis element as a vector akin to space vectors, there has always been the difficulty of mixing apples and oranges arising because of the indefinite metric of $L_4$.[3] From the formal viewpoint of vector Clifford

[2] S. Teitler (unpublished).
[3] See, e.g., C. Lanczos, Phys. Rev. 134, B476 (1964).

algebras however, one may consider that time is not necessarily a vector in a physical vector space but rather some other algebraic basis element. Then space–time would be an artificial construct suitable in the description of certain physical phenomena but not completely suitable for all. Of course the introduction of such an interpretation could only be justified if it encompassed all previous results and provided insight into some other puzzles; most particularly here, the relation between the imaginary unit $i$ in $C_{16}$ and space–time, and the nature of complex conjugation in $C_{16}$.

We wish then to imbed the space–time vectors in a Clifford *algebra* generated with the basis of some vector space. Since one of our conditions is the removal of an indefinite metric as a starting point, we consider the generating space to be Euclidean. We may then identify the three space vectors with three of these Euclidean basis vectors. The smallest such vector space in whose algebra there is a satisfactory "time vector" is the six-dimensional space, $E_6$, having the basis vectors $v^\mu$ ($\mu = a, b, c, 1, 2, 3$) with metric $[\epsilon\delta^{\mu\nu}]$ possessing only nonvanishing diagonal elements all equal to $\epsilon$ ($\epsilon$ equals 1 or $-1$). The $v^\mu$'s satisfy the anticommutation relations

$$v^\mu v^\nu + v^\nu v^\mu = 2\epsilon\delta^{\mu\nu} \quad (\mu, \nu = a, b, c, 1, 2, 3). \tag{2.1}$$

The algebra generated using the $E_6$ basis vectors possesses sixty-four basis elements $1; v^\mu; v^{\mu\nu}, (\mu < \nu); v^{\lambda\mu\nu}, (\lambda < \mu < \nu); v^{\kappa\lambda\mu\nu}, (\kappa < \lambda < \mu < \nu); v^{\iota\kappa\lambda\mu\nu}, (\iota < \kappa < \lambda < \mu < \nu); v^{abc123}$, where all Greek indices run from $a$ to 3, and we have chosen increasing order of indices in the basis elements of the algebra for specificity. This algebra $C_{64}$ possesses one scalar element, six vector elements, fifteen bivector elements, six vector elements, fifteen bivector elements, twenty trivector elements, fifteen quadrivector elements, six quintivector elements, and one pseudoscalar element.

As indicated above we identify a triplet of vectors of $E_6$ as the three basis vectors of ordinary space. We assume triplets do not overlap so there are only two such choices e.g. $(v^a, v^b, v^c)$ or $(v^1, v^2, v^3)$. Thus we are assuming that the six-dimensional Euclidean space should be considered as the direct sum of two three-dimensional Euclidean spaces each of which might correspond to ordinary space. For definiteness let us choose $(v^1, v^2, v^3)$ as the unit vectors of ordinary space. Then $v^{abc}$ may be cast in the role of "time vector" since it anticommutes with the $v^j$ ($j = 1, 2, 3$) and has self-product equal to $-\infty \rightarrow \epsilon$.

Then using $(v^{abc}, v^1, v^2, v^3)$ one can generate a subalgebra of $C_{64}$ which is isomorphic to $C_{16}$ when we identify $e^0 = v^{abc}$, $e^j = v^j$ ($j = 1, 2, 3$), and make the same choice for $\epsilon$, i.e. $\epsilon = -1$. Further, in our present six-dimensional approach we note that the bivectors $(v^{ab}, v^{ac}, v^{bc})$ all have self-product equal to $-1$ and individually commute with all the elements of the subalgebra generated with $(v^{abc}, v^1, v^2, v^3)$. Then we may expect that any of these bivectors or a suitable combination of them may be used to play the role of the imaginary unit. Indeed for $\epsilon = \mp 1$, $(1, v^{ab}, \pm v^{ac}, v^{bc})$ form the basis of a quaternion algebra $Q_A$, so that $C_{64}$ may be viewed as a quaternion-valued sixteen-element algebra which includes a sixteen-element complex algebra as a subalgebra corresponding to a thirty-two-element real subalgebra.

In restricting our attention to the complex algebra in which the bivectors $(v^{ab}, v^{ac}, v^{bc})$ are used in the description of the complex unit, there seems to be no reason to single out one of them to be identified with $i$ in $C_{16}$. Accordingly we consider $i$ to be a linear combination

$$i = \alpha v^{ab} + \beta v^{ac} + \gamma v^{bc}, \tag{2.2}$$

where

$$\alpha^2 + \beta^2 + \gamma^2 = 1.$$

Note that $i$ and $v^{abc}$ are invariant under a rotation of axes in $(v^a, v^b, v^c)$-space so that the formulation using $E_6$ is only specified to within a rotation in the $(v^a, v^b, v^c)$ subspace.

Consider now a reversion in the algebraic indices generated using the $v^a$, $v^b$, $v^c$ basis vectors. This means the opposite ordering of algebra of the $a$, $b$, $c$ indices, e.g., $abc \rightarrow cba$, etc. This may also be viewed as an inversion of the "vector" components of the quaternion space $Q_A$. Then

$$i \rightarrow \alpha v^{ba} + \beta v^{ca} + \gamma v^{cb} = -i, \tag{2.3a}$$

but also

$$e^0 = v^{abc} \rightarrow v^{cba} = -e^0. \tag{2.3b}$$

Thus a reversion in the $(a, b, c)$ indices corresponds to $TK$. However, since $e^0$ is assigned as the pseudoscalar in the $(v^a, v^b, v^c)$ subalgebra, we identify $T$ with inversion in the $(v^a, v^b, v^c)$-subspace. Then we obtain for complex conjugation alone, a combination of (1), inversion of the $(v^a, v^b, v^c)$-subspace and (2), reversion of its vector Clifford algebra (or inversion of the quaternion space $Q_A$).

Since the operation of complex conjugation is so involved in the change of sign of charge, it is reasonable to suppose that other operations involving the

charge are also intimately connected with the ($v^a$, $v^b$, $v^c$)-subspace. Indeed the invariance of the formalism to rotations in the ($v^a$, $v^b$, $v^c$)-subspace, indicates that the latter may be identified with isotopic space. We make this identification so that $E_6$ may then be considered as made up of two three-dimensional spaces, one which is ordinary space and one isotopic space. The relative juxtaposition of these two spaces and their algebras are involved in the operations of $K$, $TPK$, $TP$, etc. Further another significant aspect of the present formalism is that it provides a link between the two spaces by identifying the pseudoscalar in the ($v^a$, $v^b$, $v^c$) subspace as the "time vector" used in the description of dynamics in the ($v^1$, $v^2$, $v^3$) subspace. In this way we see how physical fields need both spaces for their description and how the intimate relation between space–time and complex quantities may be viewed in a unified manner.

## III. PARTICLE UNITARY SYMMETRY

Throughout our discussions of the application of $C_{16}$ to fields, we have consistently considered homogeneous Lorentz transformations of the basis vectors of $L_4$ and the corresponding transformations of the basis elements of $C_{16}$. We have required Lorentz covariance with respect to these transformations. Moreover, we have coupled in, through the dynamical requirement that the fields satisfy a given Dirac-like equation, a covariance with respect to inhomogeneous Lorentz transformations. However when we consider the entire set of Dirac-like operators as in Table I, we may expect this dynamical extension to the inhomogeneous Lorentz symmetry for the fields is no longer valid in that they include a relative sign change among space and time derivatives, respectively. Since we do not wish to upset this dynamical extension we restrict possible unitary symmetry among spinors of a given spinor space to the set satisfying either ($O_D$, $O_D^K$, $O_D^{TPK}$, $O_D^{TP}$) or its parity counterpart but not both. For specificity we consider the set containing $O_D$. This has possible unitary summetry $U_4$.

We may further reduce this set by restricting our considerations to those Dirac-like equations with the same relative sign of its mass factor, i.e. a separation of particles and antiparticles. This reduces our consideration to a $U_2$ unitary symmetry between spinor solutions to, say, $O_D$ and $O_D^{TPK}$.

We wish to couple this $U_2$ symmetry to isotopic spin symmetry. We note, however, that there is a redundancy between the $O_D$ and $O_D^{TPK}$ solutions when the charge vanishes. Hence, the isotopic spin symmetry and the $U_2$ spinor symmetry are not independent. We expect then a combined unitary symmetry over only three types of spinor solutions, one each when $q = 0, \pm|q|$ in $O_D$. This means we have arrived at a basis for the occurrence of at least the usual *ad hoc* $SU_3$ symmetry for particles.[4] However using an argument similar to the one used in the discussion of the $U_4$ symmetry for the complete regular current density, we would arrive at $U_3$ symmetry here. Indeed there are indications that the $U_3$ symmetry for particles is the appropriate first extension of isotopic symmetry even from *ad hoc* considerations.[5]

It should be emphasized, especially to the reader who has reservations concerning the embedding of space–time in the algebra generated using $E_6$, that the $U_3$ particle symmetry obtained here relies only on the occurrence of isotopic spin plus the inherent unitary symmetry of the spinors arising from the invariance of the regular current density. Thus it is independent of the model of isotopic space and, in particular, independent of the details of Sec. II.

This $U_3$ symmetry may be extended to include antiparticles or in other words, the full $U_4$ symmetry possible with the ($O_D$, $O_D^{TPK}$, $O_D^{TP}$, $O_D^K$) sets of solutions rather than the $U_2$ symmetry used above. Also there is the further extension possible by encompassing the spin unitary symmetry among the components of the individual spinors (see III), as in the work initiated by Gürsey, Radicati, and Pais.[6]

---

[4] See, e.g., M. Gell-Mann and Y. Ne 'eman, *The Eightfold Way* (W. A. Benjamin, Inc., New York, 1964).

[5] See, e.g., S. Okubo, C. Ryan, and R. E. Marshak, Nuovo Cimento **34**, 759 (1964).

[6] F. Gürsey and L. A. Radicati, Phys. Rev. Letters **13**, 173 (1964); A. Pais, *ibid.*, 175 (1964).

# Spin-Matrix Polynomials and the Rotation Operator for Arbitrary Spin*

T. A. WEBER AND S. A. WILLIAMS
*Institute for Atomic Research and Department of Physics, Iowa State University, Ames, Iowa*
(Received 18 January 1965)

A technique for the expansion of an arbitrary analytic function of a spin matrix is developed in terms of a complete set of polynomials based on the characteristic equations of the spin matrices. The expansion coefficients are determined by use of the ascending difference operator from the calculus of finite differences. The expansions are valid not only for functions of a spin matrix but also for functions of a complex variable. In the latter case the eigenvalues of the spin matrices are the zeros of the polynomials. In either case the expansion coefficients are the same although of course in the case of a spin matrix the series terminates. As an example the rotation operator for arbitrary spin is developed via its functional analog in terms of these polynomials, and the region of convergence of the series to the function is investigated.

## I. INTRODUCTION

It is well known that an analytic function of spin matrices can be written as a polynomial in the spin matrices because there are $(2s+1)^2$ linearly independent matrices which may be formed from the spin matrices and their products for fixed spin $s$.

In particular, a function of one component of the spin can be expanded in a polynomial of degree $2s$ in that spin component. This is a direct consequence of the Cayley–Hamilton theorem[1] which states that a square matrix satisfies its own characteristic equation. Since the matrices associated with the components of the spin operator for fixed spin have the same eigenvalues, and since these eigenvalues are repeated in the characteristic equations for higher spin, it is convenient to develop a set of polynomials of a complex variable based upon the characteristic equations. An analytic function can then be expanded in terms of these polynomials and the calculus of finite differences applied to determine the coefficients of expansion. In general, for the expansion of an arbitrary analytic function this results in an infinite series, but the series terminates if the variable is replaced by a spin matrix. As an application, the function $e^{iz\theta}$ is developed in terms of these polynomials. It is shown that the series, for $|\theta| < \pi$ with any $z$, represents the function $e^{iz\theta}$. In the special case that $z$ is a component of the spin operator, the series terminates, and the resulting polynomial represents the rotation operator.

## II. SPIN MATRIX POLYNOMIALS

Consider the set of polynomials

$$\begin{aligned} W_1 &= z, \\ W_3 &= (z-1)z(z+1), \\ W_5 &= (z-2)(z-1)z(z+1)(z+2), \\ &\vdots \end{aligned} \tag{1}$$

If $z$ is a spin matrix associated with integral spin $s$, then the polynomial $W_{2s+1}$ is equal to zero and all subsequent polynomials are zero by the Cayley–Hamilton Theorem. This occurs since $W_{2s+1} = 0$ is the characteristic equation for spin $s$ and all higher-degree polynomials contain this polynomial as a factor. We shall call this the termination property of the polynomials. Thus this set of polynomials is suitable for the expansion of an analytic function of integral spin matrices. However, to obtain a complete set, a set of polynomials of even degree must be added to the set given in Eq. (1). The most general way of completing the set of polynomials which preserves the termination property is to multiply the set given in Eq. (1) by $z - c$. The constant $c$ can be taken equal to zero since it will just reproduce the original set of polynomials; so we choose the set

$$1,\ W_{2n+1} \equiv (z+n)!/(z-n-1)!, \qquad zW_{2n+1}; \qquad n = 0, 1 \cdots . \tag{2}$$

For half-integral spin, consider the set of polynomials

$$\begin{aligned} W_0 &= 1, \\ W_2 &= (z-\tfrac{1}{2})(z+\tfrac{1}{2}), \\ W_4 &= (z-\tfrac{3}{2})(z-\tfrac{1}{2})(z+\tfrac{1}{2})(z+\tfrac{3}{2}), \\ &\vdots \end{aligned} \tag{3}$$

* Contribution No. 1649. Work was performed in the Ames Laboratory of the U. S. Atomic Energy Commission.
[1] See, for example, Bernard Friedman, *Principles and Techniques of Applied Mathematics* (John Wiley & Sons, Inc., New York, 1956), p. 92.

If $z$ is a spin matrix associated with half-integral spin $s$, then again $W_{2s+1}$ and all subsequent polynomials are zero. To complete the set of polynomials applicable to half-integer spin, we append the above set with the polynomials $zW_{2n}$. Our complete set of polynomials for half-integer spin is then

$$W_{2n} \equiv (z + n - \tfrac{1}{2})!/(z - \tfrac{1}{2} - n)!, \qquad zW_{2n}; \qquad n = 0, 1, 2 \cdots . \tag{4}$$

Although the even polynomials $W_{2n}$ of Eq. (3) and odd polynomials $W_{2n+1}$ of Eq. (1) form a complete set in a very natural way, we choose instead to augment each in the way indicated in order to preserve the termination property. For this reason we shall treat the expansion of an arbitrary analytic function for integral spin and half-integral spin separately. Thus we expand an arbitrary analytic function

$$f(z) = c + \sum_{n=0}^{\infty} a_n z W_{2n+1}(z) + \sum_{n=0}^{\infty} b_n W_{2n+1}(z) \tag{5}$$

for integral spin, and

$$f(z) = \sum_{n=0}^{\infty} \alpha_n W_{2n}(z) + \sum_{n=0}^{\infty} \beta_n z W_{2n}(z) \tag{6}$$

for half-integral spin.

## III. EXPANSION COEFFICIENTS

It is apparent that one can determine the expansion coefficients of Eqs. (5) and (6) by evaluating the function at the successive zeros of the polynomials which may be conveniently written as

$$W_{2n+1}(n - k) = 0, \qquad k = 0, 1, 2 \cdots 2n \tag{7a}$$

and

$$W_{2n}(n - \tfrac{1}{2} - k) = 0, \quad k = 0, 1, \cdots 2n - 1. \tag{7b}$$

For example, setting $z = 0$ in Eq. (5) yields

$$c = f(0)$$

since all the polynomials vanish. Then with $z = \pm 1$, $W_3$, $W_5$, etc. vanish, and we have

$$a_0 = \tfrac{1}{2}[f(1) + f(-1)] - f(0),$$
$$b_0 = \tfrac{1}{2}[f(1) - f(-1)].$$

In principle all the coefficients can be obtained by continuing in this fashion. It is to be noted that the coefficients are equally applicable if $z$ is a complex variable or if $z$ is a spin matrix component; in the latter case one need compute only a finite number of coefficients.

In order to determine the coefficients in a systematic manner, we introduce the unit step ascending difference operator defined by[2]

$$\nabla g(z) = g(z) - g(z - 1), \tag{8}$$

where $g$ is any function of $z$. The $n$th ascending difference is

$$\nabla^n g(z) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} g(z - k), \tag{9}$$

where $\binom{n}{k}$ is the binomial coefficient. Clearly the ascending difference of an $n$th degree polynomial is a polynomial of degree $n - 1$. Therefore, if $g(z)$ is a polynomial of degree less than $n$, then $\nabla^n g(z) \equiv 0$. Consider

$$\nabla^{2k+1} W_{2n+1}(z) = \sum_{j=0}^{2k+1} (-1)^j \binom{2k+1}{j} W_{2n+1}(z - j), \tag{10}$$

which is nonzero only for $n \geq k$. Setting $z = k$, this becomes

$$\nabla^{2k+1} W_{2n+1}(k) = \sum_{j=0}^{2k+1} (-1)^j \binom{2k+1}{j} W_{2n+1}[n - (n - k + j)]. \tag{11}$$

From Eq. (7a) it follows that the terms in the sum with $j = k - n, k - n + 1, \cdots, k + n$ vanish; hence, the lower limit in the sum of Eq. (11) is $j = k + n + 1$. But since the lower limit must be equal to or less than the upper limit, we conclude that $k \geq n$. For $k > n$ the left-hand side of Eq. (11) vanishes identically, and therefore, we write

$$\nabla^{2k+1} W_{2n+1}(k) = (-1)^{2n+1} W_{2n+1}(-n - 1)\, \delta_{nk} = (2n + 1)!\, \delta_{nk}. \tag{12}$$

Similarly,

$$\nabla^{2k+2}(k + 1) W_{2n+1}(k + 1) = (2n + 2)!\, \delta_{nk}. \tag{13}$$

Decomposing an arbitrary analytic function $f(z)$ into even and odd functions, it follows from Eq. (5) that

$$f_e(z) = f_e(0) + \sum_{n=0}^{\infty} a_n z W_{2n+1}(z) \tag{14}$$

and

$$f_0(z) = \sum_{n=0}^{\infty} b_n W_{2n+1}(z). \tag{15}$$

[2] See, for example, D. R. Hartree, *Numerical Analysis* (Clarendon Press, Oxford, England, 1952), p. 53.

Using Eqs. (12) and (13), the expansion coefficients are

$$a_n = \nabla^{2n+2} f_e(n+1)/(2n+2)! \tag{16}$$

and

$$b_n = \nabla^{2n+1} f_0(n)/(2n+1)!. \tag{17}$$

The expansions given in Eqs. (14) and (15), together with the coefficients given in Eqs. (16) and (17), are suitable for cases of interest involving integral spin.

In a similar manner one shows

$$\nabla^{2k} W_{2n}(k - \tfrac{1}{2}) = (2n)!\, \delta_{kn} \tag{18}$$

and

$$\nabla^{2k+1}(k + \tfrac{1}{2}) W_{2n}(k + \tfrac{1}{2}) = (2n+1)!\, \delta_{kn}. \tag{19}$$

Using Eqs. (18) and (19), the expansion coefficients of Eq. (6) are

$$\alpha_n = \nabla^{2n} f_e(n - \tfrac{1}{2})/(2n)! \tag{20}$$

and

$$\beta_n = \nabla^{2n+1} f_0(n + \tfrac{1}{2})/(2n+1)!, \tag{21}$$

where we have again decomposed $f(z)$ into

$$f_e(z) = \sum_{n=0}^{\infty} \alpha_n W_{2n}(z), \tag{22}$$

$$f_0(z) = \sum_{n=0}^{\infty} \beta_n z W_{2n}(z). \tag{23}$$

The expansions given in Eqs. (22) and (23) have the termination property when applied to half-integral spin.

As with all such expansions, one must investigate not only whether the series converges, but whether the series converges to the function on the left. This will be considered in Sec. V for the expansions of $\sin z\theta$ and $\cos z\theta$.

## IV. ROTATION OPERATOR FOR ARBITRARY SPIN

The rotation operator for arbitrary spin is

$$R = e^{i\theta s} = \cos \theta z + i \sin \theta z, \tag{24}$$

where $z = \hat{\beta}\cdot\mathbf{s}$, $\hat{\beta}$ is the direction of rotation, $\theta$ is the angle of rotation, and $\mathbf{s}$ is the spin operator $(s_x, s_y, s_z)$. In the following we shall restrict the discussion to integral spin and only state the results for half-integral spin. In the derivations we shall regard $z$ as a complex variable since the coefficients are independent of the nature of $z$. Using Eqs. (14) and (16),

$$\cos \theta z = 1 + \sum_{n=0}^{\infty} \frac{\nabla^{2n+2} \cos[\theta(n+1)]}{(2n+2)!} z W_{2n+1}(z). \tag{25}$$

But from Eq. (9),

$$\begin{aligned}
&\nabla^{2n+2} \cos[\theta(n+1)] \\
&= \sum_{k=0}^{2n+2} (-1)^k \binom{2n+2}{k} \cos[\theta(n+1-k)] \\
&= \mathrm{Re} \sum_{k=0}^{2n+2} (-1)^k \binom{2n+2}{k} e^{i\theta(n+1-k)} \\
&= \mathrm{Re}\,(e^{i\frac{1}{2}\theta} - e^{-i\frac{1}{2}\theta})^{2n+2} \\
&= (-1)^{n+1} 2^{n+1} (1 - \cos\theta)^{n+1}.
\end{aligned} \tag{26}$$

Therefore, Eq. (25) becomes

$$\cos \theta z = 1 + \sum_{n=0}^{\infty} \frac{(-1)^{n+1} 2^{n+1}}{(2n+2)!} \times (1 - \cos\theta)^{n+1} z W_{2n+1}(z). \tag{27}$$

Similarly, using Eqs. (15) and (17) we obtain

$$\sin \theta z = \sum_{n=0}^{\infty} \frac{(-1)^n 2^n}{(2n+1)!} \sin\theta (1 - \cos\theta)^n W_{2n+1}(z). \tag{28}$$

One can also easily obtain the expansion of $\sin \theta z$ by differentiating the expansion of $\cos \theta z$, Eq. (27), with respect to $\theta$. In order to verify that differentiation of the expansion of $\sin \theta z$ yields the expansion of $\cos \theta z$, it is convenient to make use of the identity

$$z^2 W_{2n+1}(z) = (n+1)^2 W_{2n+1}(z) + W_{2n+3}(z). \tag{29}$$

Upon differentiation of Eq. (28) with respect to $\theta$, we have

$$z \cos \theta z = \sum_{n=0}^{\infty} \frac{(-1)^n 2^n}{(2n+1)!} [n + (n+1)\cos\theta] \times (1 - \cos\theta)^n W_{2n+1}(z). \tag{30}$$

Multiplying Eq. (27) by $z$ and using Eq. (29), we obtain

$$\begin{aligned}
z \cos \theta z &= W_1(z) + \sum_{n=0}^{\infty} \frac{(-1)^{n+1} 2^{n+1}}{(2n+2)!} (1 - \cos\theta)^{n+1} \\
&\quad \times \{(n+1)^2 W_{2n+1}(z) + W_{2n+3}(z)\} \\
&= \sum_{n=0}^{\infty} c_n(\theta) W_{2n+1}(z).
\end{aligned} \tag{31}$$

Simple algebraic manipulation shows that the $c_n(\theta)$ are equal to the expansion coefficients of Eq. (30). The above differential properties will be used in the discussion of the convergence of the expansions.

For half-integral spin the expansions are

$$\cos \theta z = \sum_{n=0}^{\infty} \frac{(-1)^n 2^{2n}}{(2n)!} \cos \tfrac{1}{2}\theta (\sin \tfrac{1}{2}\theta)^{2n} W_{2n}(z) \qquad (32)$$

and

$$\sin \theta z = \sum_{n=0}^{\infty} \frac{(-1)^n 2^{2n+1}}{(2n+1)!} (\sin \tfrac{1}{2}\theta)^{2n+1} z W_{2n}(z). \qquad (33)$$

In the case that $z$ is $\hat{\beta}\cdot\mathbf{s}$, the upper limit of Eqs. (27) and (28) is $n = s - 1$, while in Eqs. (32) and (33) the upper limit is $n = s - \frac{1}{2}$ because of the termination property. For example, with $s = \frac{1}{2}$ we have, from Eqs. (32) and (33), the familiar result

$$e^{i\theta\hat{\beta}\cdot\mathbf{s}} = \cos \tfrac{1}{2}\theta + 2i\hat{\beta}\cdot\mathbf{s} \sin \tfrac{1}{2}\theta \qquad (34)$$

or, in terms of the Pauli spin matrices $\boldsymbol{\sigma} = 2\mathbf{s}$,

$$e^{i\frac{1}{2}\theta\hat{\beta}\cdot\boldsymbol{\sigma}} = \cos \tfrac{1}{2}\theta + i\hat{\beta}\cdot\boldsymbol{\sigma} \sin \tfrac{1}{2}\theta. \qquad (35)$$

## V. CONVERGENCE OF THE sin θz AND cos θz EXPANSIONS

Again we will discuss the case of integral spin; the results for half-integral spin are the same. Consider the expansions of the cos $\theta z$ and sin $\theta z$ as given by Eqs. (27) and (28). The ratio test shows that the series are absolutely convergent if $|\sin \frac{1}{2}\theta| < 1$ for arbitrary complex $z$. Hence, either series, considered as a power series in sin $\frac{1}{2}\theta$ has a radius of convergence of one in the complex sin $\frac{1}{2}\theta$ plane. This region of convergence can now be mapped into the complex $\theta$ plane, but the mapping is one-to-many. Consider the region in the complex $\theta$ plane which contains the point $\theta = 0$ (i.e., region containing $-\pi \leq \theta \leq \pi$, $\theta$ real). For the point $\theta = 0$, it is obvious that the series given in Eqs. (27) and (28) converge to the cosine and sine functions evaluated at that point. In view of the fact that differentiation of the sine expansion gives the cosine expansion and vice versa, as discussed in Sec. IV, it is clear that the analytic function represented by either expansion and all of its derivatives equals the function on the left and all of its derivatives at $\theta = 0$. Hence, the expansion is equal to the function it perports to represent in the open region of convergence in the complex $\theta$ plane containing the point $\theta = 0$. None of the other regions of convergence in the complex $\theta$ plane contain such a point for arbitrary $z$. In particular for $\theta$ real, the expansions given by Eqs. (27), (28), (32), and (33) converge to the appropriate function only in the open interval $-\pi < \theta < \pi$ for arbitrary complex $z$.

If $z$ is an integer, then Eqs. (27) and (28) terminate and therefore converge to cos $\theta z$ and sin $\theta z$, respectively, for all $\theta$. Similar remarks apply to Eqs. (32) and (33) if $z$ is a half-integer. It should be noted that Eq. (28), with $z$ an integer, and Eq. (33), for $z$ a half-integer, correspond to two of Dwight's formulas[3] for sin $n\theta$. The rotation operator as a polynomial in the spin matrices may also be deduced[4] by using these formulas of Dwight.

## VI. CONCLUSION

It is obvious that one can apply the techniques presented here to the problem of a function of an arbitrary square matrix $A$. Of course, if the eigenvalues do not differ by fixed amounts, the ascending difference operator will be of little use. However, one can construct a set of polynomials in the following way: 1, $(A - \lambda_1)(A - \lambda_2)$, $\cdots$, where $\lambda_1, \lambda_2, \cdots$ are the eigenvalues. The highest nonvanishing polynomial will be of degree one less than the order of the matrix, $n$. These obviously form a complete set so far as functions of the matrix $A$ are concerned. In order to determine the coefficients, one successively evaluates the function at $\lambda_1, \lambda_2, \cdots$ determining one coefficient at each substitution, rather than solving $n$ simultaneous equations as is usually done.[5] The extra conditions necessary to determine the coefficients for the case of repeated roots are obtained from the derivatives of the function.

One application of the above techniques is the development in terms of the polynomials $W_n$ of the Lorentz transformation operator for particles of arbitrary spin. With this operator one can, for example, obtain the Hamiltonian for the particle.[6] The present explicit calculations are limited by mathematical complexity to small values of the spin. The authors are presently developing the explicit form for arbitrary spin.

*Note added in proof.* R. van Wageninger, Nucl. Phys. **60**, 250 (1964), gives the results obtained by a computer calculation of the spin-dependent coefficients in an expansion of the rotation operator in powers of $\hat{\beta}\cdot\mathbf{s}$ for $s = \frac{1}{2}$ to $s = 6$.

## ACKNOWLEDGMENT

We are pleased to acknowledge many helpful discussions with our colleagues, in particular Professor R. H. Good, Jr. and Professor J. M. Keller.

[3] H. B. Dwight, *Table of Integrals and Other Mathematical Data* (The Macmillan Company, New York, 1957), 3rd ed., formulas 403.11 and 403.13.
[4] S. Weinberg, Phys. Rev. **133**, B 1320 (1964).
[5] Reference 1, pp. 119–121.
[6] D. L. Weaver, C. L. Hammer, and R. H. Good, Jr., Phys. Rev. **135**, B 241 (1964).

# A Model of Interacting Radiation and Matter. II*

CHARLES R. WILLIS
*Boston University, Boston, Massachusetts*
(Received 3 February 1965)

We extend the investigation of the long time behavior of a model consisting of $N$ two-level atoms interacting with radiation to include dissipation, pumping, and center-of-mass motion. We show that when the effective expansion parameter $\tilde{\gamma}N\tilde{\gamma}$ is small, the self-consistent field approximation remains the solution to lowest order in $\tilde{\gamma}N\tilde{\gamma}$. The inclusion of the center-of-mass motion introduces another dimensionless parameter $\beta$ into the theory. We show that when this parameter is small, the electromagnetic field amplitude varies slowly on the time scale of the center-of-mass motion. We solve the equations of motion in this slowly varying limit by an extension of the Bogoliuboff–Kryloff theory of quasilinearity to the problem of time-dependent integral kernals. We find the unique stable stationary state and show that in the slowly varying limit the stationary state is approached independent of initial conditions. We calculate the frequency shift to second order in $\beta$. The first-order frequency shift is the same as that calculated by Lamb. We compare our steady-state solution with recent experiments with lasers. We include the effect of collisions in the steady state.

## I. INTRODUCTION

IN this paper we extend our model[1] of interacting radiation and matter to include dissipation, pumping, and center-of-mass motion. In I we showed that the lowest-order solution to our model of $N$ stationary two-level systems in the presence of radiation is the self-consistent field approximation which we refer to as SCFA. In the absence of center-of-mass motion, we found there was only one dimensionless dynamical constant in the theory, namely

$$\tilde{\gamma}N\tilde{\gamma} \equiv \mathfrak{N}e^2 \, |\langle a| \, \boldsymbol{\varepsilon}\cdot\mathbf{r} \, |b\rangle|^2 \, (2\pi/\hbar\Omega) = (4\pi)^{-1}\mathfrak{N}r_0\lambda^2 \equiv \alpha^2,$$

where $\mathfrak{N}$ is the number of two-level systems per unit volume, $\langle a| \, \boldsymbol{\varepsilon}\cdot\mathbf{r} \, |b\rangle$ is the dipole matrix element between the two levels of the atom, $\Omega$ is the cavity frequency, $r_0$ is the classical electron radius, and $\lambda$ is the wavelength of the radiation. The second equality follows from a sum rule or equivalently from evaluating the matrix elements with harmonic oscillator wavefunctions.

If we restore dimensions we find

$$\tilde{\gamma}N\tilde{\gamma}\omega_0^2 = \alpha^2\omega_0^2 \equiv \omega_L^2 = (\mathfrak{N}\pi e^2/m) = (\omega_p/2)^2,$$

where $\omega_0$ is the atomic frequency and $\omega_p$ is the plasma frequency. The close relationship between $\omega_L$ and $\omega_p$ is not coincidental. They both are measures of the collective response of a system of sources in interaction with the electromagnetic field through Maxwell's equations. The plasma frequency in this context arises from the self-consistent interaction of charges with transverse electromagnetic waves. The laser frequency $\omega_L$ arises from the self-consistent interaction of induced dipoles with the electromagnetic field.

When we consider the center-of-mass motion of our two-level systems, we must distinguish between solid and gaseous behavior because these introduce quite different characteristic times into the problem. In this paper we consider gaseous lasers, so the largest center-of-mass characteristic frequency is $\omega_D$, the Doppler width. Thus, we have a second dimensionless constant $\beta \equiv \alpha(\omega_0/\omega_D) = (\omega_L/\omega_D)$ which measures the magnitude of the effect of the center-of-mass motion. For most gas lasers, $\beta$ is much less than one, which is fortunate because $\beta$ small means that the field amplitudes vary slowly compared with the center-of-mass motion. Since $\alpha$ is very small for gas lasers, the SCFA is valid and, in addition, with $\beta$ small we show that our model of $N$ two-level systems satisfies a generalization of the quasi-linear theory of Kryloff–Bogoliuboff.[2] If $\beta$ were much larger than one the center-of-mass motion would be slow compared with the times over which energy is exchanged between radiation and matter and we would have relaxation oscillations.

We introduce dissipation and a pump, phenomenologically, as linear terms in the operator equations of motion. With dissipation we have three more characteristic times which are nondynamical. The relaxation time for the diagonal matrix elements of the matter density matrix is $T_1$, the corresponding time for the off-diagonal matrix elements is $T_2$, and the relaxation time for the radiation is $T_r$. We show

* The research reported in this paper was sponsored in part by the Air Force Cambridge Research Laboratories, Office of Aerospace Research under contract No. AF 19(628)–2460.

[1] C. R. Willis, J. Math. Phys. 5, 1241 (1964), hereafter referred to as I.

[2] N. Kryloff and N. Bogoliuboff, *Introduction to Nonlinear Mechanics* (Princeton University Press, Princeton, New Jersey, 1947).

that the steady state of the system is a very sensitive function of $T_2$.

In Sec. II we introduce the Hamiltonian, derive the operator equations of motion, and formally eliminate the matter dipole operators. Although our treatment of the Hamiltonian in Sec. II is self-contained, we refer the reader to I for a more thorough treatment.

In Sec. III we find the steady-state solutions for the energy of the electromagnetic field and for the steady-state population inversion. For the steady state we require only that $\alpha \ll 1$.

We solve the time-dependent equations to second order in $\beta$ in Secs. IV and V. We show that no matter how the system is started when the pump power is at or above threshold, the system approaches the unique stable steady state derived in Sec. III. We show further that a full solution to all orders in $\beta$ does not affect the steady state but only the rate at which the steady state is approached. However, the frequency shift is a power series in $\beta$, and we calculate it to second order. For small $\beta$ the series converges very rapidly.

In Sec. VI we discuss the steady-state electromagnetic energy density as a function of cavity frequency and compare our results with recent experiments.[3,4]

In the Appendix we derive the effect of the center-of-mass motion including collisions on the behavior of the laser. In particular, we show that the introduction of collisions effectively decreases $T_2$.

## II. HAMILTONIAN OF THE MODEL AND THE EQUATIONS OF MOTION

We consider $N$ two-level systems 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

$$h(N) = \frac{\hbar\omega_0}{2}\sum_{\alpha}^{N}\hat{\sigma}_\alpha + H_{\rm cm}(\mathbf{X}, \mathbf{P}), \qquad (2.1)$$

where

$$\hat{\sigma}_\alpha \equiv 2\sigma_\alpha^\dagger\sigma_\alpha - 1, \qquad [\sigma_\alpha, \sigma_\alpha^\dagger]_+ = 1,$$
$$[\sigma_\alpha, \sigma_\alpha]_+ = [\sigma_\alpha^\dagger, \sigma_\alpha^\dagger]_+ = 0,$$
$$[\sigma_\alpha^\dagger, \sigma_\beta^\dagger] = [\sigma_\alpha^\dagger, \sigma_\beta] = [\sigma_\alpha, \sigma_\beta] = 0 \quad \text{for} \quad \alpha \neq \beta,$$

and where $\mathbf{X} = (X_1 \cdots X_N)$, $\mathbf{P} = (P_1 \cdots P_N)$. The plus subscript indicates an anticommutator. The vanishing of the commutators for different atoms represents the fact that we are treating the atoms as distinguishable; i.e., we are assuming the density of atoms to be sufficiently low so that the overlap of their wavefunctions is negligible and thus the effects of symmetry may be neglected.

In this section we do not need the properties of $H_{\rm cm}$; therefore, we defer a thorough analysis of it to the Appendix.

A convenient representation of the matter system is a Kronecker product of single-particle spaces. In the single-atom space we may represent the operators in the following form:

$$\sigma_\alpha^\dagger = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}, \qquad \sigma_\alpha = \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}, \qquad \hat{\sigma}_\alpha = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

The Hamiltonian for the electromagnetic field is

$$H_f = \sum_k \hbar\Omega_k(a_k^\dagger a_k + \tfrac{1}{2}), \qquad (2.2)$$

where

$$[a_k, a_{k'}^\dagger] = \delta_{kk'}, \qquad a_k = (2\hbar\Omega_k)^{-\frac{1}{2}}(p_k - i\Omega_k q_k),$$
$$a_k^\dagger = (2\hbar\Omega_k)^{-\frac{1}{2}}(p_k + i\Omega_k q_k).$$

The vector potential is

$$\mathbf{A}(x, t) = c(4\pi)^{\frac{1}{2}}\sum_k \boldsymbol{\varepsilon}_k q_k(t)E_k(x),$$

where $E_k(x)$ is the $k$th cavity eigenfunction and $\boldsymbol{\varepsilon}_k$ is a unit vector in the plane of polarization.

The interaction Hamiltonian, $H_i$, is

$$H_i = -\frac{1}{c}\int j\cdot A\, d^3x = -\sum_k\sum_\alpha^N \{\sigma_\alpha\langle b|\, j_k(X_\alpha)\,|a\rangle + \sigma_\alpha^\dagger\langle a|\, j_k(X_\alpha)\,|b\rangle\}q_k E_k(X_\alpha), \qquad (2.3)$$

where $X_\alpha$ is the center of mass of the $\alpha$th atom and

$$\langle a|\, j_k(X_\alpha)\,|b\rangle = (e/2m)[E_k(X_\alpha)]^{-1}\int \psi_a^*(x) \times \{\boldsymbol{\varepsilon}_k\cdot\mathbf{p}E_k(X_\alpha + x) + E_k(X_\alpha + x)\boldsymbol{\varepsilon}_k\cdot\mathbf{p}\}\psi_b(x)\, d^3x,$$

where $\mathbf{p} \equiv (\hbar/i)\nabla_x$. In the dipole approximation we have

$$\langle a|\, j_k(X_\alpha)\,|b\rangle = (e/m)\langle a|\, \boldsymbol{\varepsilon}_k\cdot\mathbf{p}\,|b\rangle,$$

and our interaction Hamiltonian is

$$H_i = -(e/m)\sum_k\sum_\alpha E_k(X_\alpha)q_k \times \{\sigma_\alpha\langle a|\, \boldsymbol{\varepsilon}_k\cdot\mathbf{p}\,|b\rangle + \langle b|\, \boldsymbol{\varepsilon}_k\cdot\mathbf{p}\,|a\rangle\sigma_\alpha^\dagger\}. \qquad (2.4a)$$

If $\omega_0 \approx \Omega_k$, then $H_i$ becomes

$$H_i = \sum_k\sum_\alpha \gamma_k\Gamma_k(X_\alpha)p_k\mu_\alpha, \qquad (2.4b)$$

where we have used

$$i\omega_0 q_k = p_k, \qquad \mu_\alpha \equiv \sigma_\alpha + \sigma_\alpha^\dagger,$$
$$\langle a|\, \boldsymbol{\varepsilon}_k\cdot\mathbf{p}\,|b\rangle = im\omega_0\langle a|\, \boldsymbol{\varepsilon}_k\cdot\mathbf{x}\,|b\rangle.$$

[3] R. A. McFarlane, W. R. Bennett, and W. E. Lamb, Appl. Phys. Letters 2, 189 (1963).
[4] A. Szöke and A. Javan, Phys. Rev. Letters 10, 521 (1963).

The definitions of $\gamma_k$ and $\Gamma_k(X_\alpha)$ are $\gamma_k = e\langle a|\ \boldsymbol{\varepsilon}_k\cdot\mathbf{x}\ |b\rangle(4\pi/V)^{\frac{1}{2}}$ and $\Gamma_k(X_\alpha) \equiv E_k(X_\alpha)V^{\frac{1}{2}}$, where $V$ is the volume.

Although Eq. (2.4b) is no easier to solve than Eq. (2.4a), we use Eq. (2.4b) to facilitate comparison of our results with results of semiphenomenological theories which in effect use Eq. (2.4b).

When we express $H_i$ in terms of dimensionless operators and variables, we obtain

$$\frac{H_i}{\hbar\omega_0} = \frac{H_d}{\hbar\omega_0} + \frac{H_{nd}}{\hbar\omega_0} = \sum_\alpha \sum_k \tilde{\gamma}_k\Gamma_k(X_\alpha)\{a_k^\dagger\sigma_\alpha + a_k\sigma_\alpha^\dagger\} + \sum_\alpha \sum_k \tilde{\gamma}_k\Gamma_k(X_\alpha)\{a_k^\dagger\sigma_\alpha^\dagger + a_k\sigma_\alpha\}, \quad (2.5)$$

where $\tilde{\gamma}_k = (\hbar\omega_0)^{-1}(\hbar\Omega_k/2)^{\frac{1}{2}}\gamma_k$ is dimensionless and essentially independent of $k$ because $\Omega_k \approx \omega_0$ and $\langle a|\ \boldsymbol{\varepsilon}_k\cdot\mathbf{x}\ |b\rangle$ does not depend on the magnitude of $k$, only its direction. In I we retained $H_{nd}$, and we found it made no contribution in lowest order. We drop $H_{nd}$ here for no other reason than to make later calculations clearer.

Since we are considering a single mode, we drop the $k$ subscript from $\tilde{\gamma}_k$, $\Gamma_k$, $a_k$, and $a_k^\dagger$, combine Eqs. (2.1), (2.2), and (2.5), and obtain

$$H_N = h(N) + H_{em} + H_f + \hbar\omega_0\tilde{\gamma}\sum_\alpha^N \Gamma(X_\alpha) \times [a^\dagger\sigma_\alpha + a\sigma_\alpha^\dagger]. \quad (2.6)$$

We obtain the equations of motion for the radiation and matter operators when we use Eq. (2.6), the commutation relations, and the definition of the time derivative of an operator,

$$i\hbar(\partial O/\partial t) = [O, H_N].$$

The equations of motion for $a$, $a^\dagger$, $\sigma$, and $\sigma^\dagger$ are

$$i\dot{a} - \Omega a = \tilde{\gamma}\omega_0\sum_\alpha^N \sigma_\alpha\Gamma(X_\alpha), \quad (2.7a)$$

$$i\dot{\sigma}_\alpha - \omega_0\sigma_\alpha = -\tilde{\gamma}\omega_0 a\hat{\sigma}_\alpha\Gamma(X_\alpha), \quad (2.7b)$$

$$i\dot{a}^\dagger + \Omega a^\dagger = -\tilde{\gamma}\omega_0\sum_\alpha^N \sigma_\alpha^\dagger\Gamma(X_\alpha), \quad (2.7c)$$

$$i\dot{\sigma}_\alpha^\dagger + \omega_0\sigma_\alpha^\dagger = \tilde{\gamma}\omega_0 a^\dagger\hat{\sigma}_\alpha\Gamma(X_\alpha). \quad (2.7d)$$

From the rigorous equations of motion, Eqs. (2.7a)–(2.7d), we observe that, given a formal solution to Eqs. (2.7a) and (2.7b), we can write down the formal solution to Eqs. (2.7c) and (2.7d) by taking the complex conjugate of the solution for Eqs. (2.7a) and (2.7b). Also we see that more natural variables than $\sigma_\alpha$ and $\sigma_\alpha^\dagger$ are

$$\xi_\alpha^k = \sigma_\alpha\Gamma_k(X_\alpha), \qquad (\xi_\alpha^k)^\dagger = \sigma_\alpha^\dagger\Gamma_k(X_\alpha). \quad (2.8)$$

If $\Gamma_k(X_\alpha) = \sin kX_\alpha$, then $\xi_\alpha^k$ is the operator for the $k$th Fourier sine transform of the operator $\sigma_\alpha$.

When the definitions, Eq. (2.8), are substituted in Eqs. (2.7a) and (2.7b) and when Eq. (2.7b) is multiplied by $\Gamma(X_\alpha)$, we obtain

$$i\dot{a} + i(a/T_r) - \Omega a = \tilde{\gamma}\omega_0\sum_\alpha \xi_\alpha, \quad (2.9a)$$

$$i\dot{\xi}_\alpha + i\frac{\xi_\alpha}{T_2} - i\xi_\alpha\frac{\dot{\Gamma}(X_\alpha)}{\Gamma(X_\alpha)} - \omega_0\xi_\alpha = -\tilde{\gamma}\omega_0 a\hat{\sigma}_\alpha\Gamma^2(X_\alpha), \quad (2.9b)$$

where we drop the $k$ subscript because we are dealing with a single mode. The term $\dot{\Gamma}$ represents the time derivative of $\Gamma$ which is time-dependent because $X_\alpha(t)$ is an "operator" in the $6N$-dimensional classical phase space of the center-of-mass variables. "Operator" here means that $X_\alpha(t)$ is not an average value but is the instantaneous value of the position of the $\alpha$th atom at time $t$. As discussed in the Appendix, we treat the center of mass classically. We could equally well treat the center of mass quantum mechanically by keeping track of the order of the operators $\Gamma$ and $\dot{\Gamma}$.

We also inserted two of our three phenomenological relaxation times. $T_r$ is the radiation relaxation time of the matter free cavity usually written as $(\nu/2Q)^{-1}$ and $T_2$ represents the relaxation time of the offdiagonal matrix elements of the density matrix which is analogous to the $T_2$ of paramagnetic resonance theory.

We solve Eq. (2.9b) for $\xi_\alpha$;

$$\begin{aligned}\xi_\alpha &= i\tilde{\gamma}\omega_0\int^t \{e^{-\nu_2(t-t')}e^{i\Delta t'}b(t')\hat{\sigma}_\alpha(t')\Gamma^2[X_\alpha(t')]\} \times \exp\left[\int_{t'}^t \frac{\partial\log\Gamma(t'')}{\partial t''}\,dt''\right]dt' \\ &= i\tilde{\gamma}\omega_0\int^t \Gamma[X_\alpha(t)]\Gamma[X_\alpha(t')] \times e^{-\nu_2(t-t')}e^{i\Delta t'}\hat{\sigma}_\alpha(t')b(t')\,dt', \end{aligned} \quad (2.10)$$

where

$$\Delta \equiv \omega_0 - \Omega, \quad \nu_2 \equiv T_2^{-1}, \quad b \equiv ae^{i\Omega t}, \quad \tilde{\xi} \equiv e^{i\omega_0 t}\xi.$$

When we substitute Eq. (2.10) in Eq. (2.9a), we obtain

$$\dot{b} + \frac{b}{T_r} = \tilde{\gamma}^2\omega_0^2\sum_\alpha^N\int^t \Gamma(\alpha, t)\Gamma(\alpha, t')e^{-\nu_2(t-t')} \times e^{-i\Delta(t-t')}\hat{\sigma}_\alpha(t')b(t')\,dt'. \quad (2.11a)$$

The complex conjugate of Eq. (2.11a) is

$$\dot{b}^\dagger + \frac{b^\dagger}{T_r} = \tilde{\gamma}^2\omega_0^2\sum_\alpha^N\int^t \Gamma(\alpha, t)\Gamma(\alpha, t')e^{-\nu_2(t-t')} \times e^{i\Delta(t-t')}\hat{\sigma}_\alpha(t')b^\dagger(t')\,dt', \quad (2.11b)$$

where $\Gamma(\alpha, t) \equiv \Gamma[X_\alpha(t)]$.

The equation of motion for $\hat{\sigma}_\alpha$ is

$$i\frac{\partial\hat{\sigma}_\alpha}{\partial t} = 2\tilde{\gamma}\omega_0\{\xi_\alpha^\dagger b e^{i\Delta t} - \xi_\alpha b^\dagger e^{-i\Delta t}\} - \frac{i(\hat{\sigma}_\alpha - \hat{\sigma}_0)}{T_1}, \tag{2.12}$$

where we have added the combined dissipation and pumping term. The term, $T_1$, is the relaxation time for the diagonal matrix elements of the density matrix and $N\hat{\sigma}_0$ is the population difference produced by the pump in the absence of radiation. We leave the equation for the time derivative of $\hat{\sigma}_\alpha$ in the form of Eq. (2.12). However, if we substituted Eq. (2.10) in Eq. (2.12) we would have reduced the $(3N + 2)$ equations for $\xi_\alpha$, $\xi_\alpha^\dagger$, $a$, $a^\dagger$, and $\hat{\sigma}_\alpha$ to the $(N + 2)$ equations for $\hat{\sigma}_\alpha$, $b$, and $b^\dagger$. These equations are still functions of $6N$ center-of-mass variables through the $\Gamma$'s. This is as far as one can go with a rigorous solution. In the next section we introduce an approximation procedure that permits us to solve the equations.

## III. STEADY-STATE SOLUTION

In I we showed that the SCFA is the solution to the quantum mechanical Liouville equation to lowest order in $\tilde{\gamma}N\tilde{\gamma}$ in the absence of initial particle–field correlations and neglecting the center-of-mass motion. We showed that when one went beyond the first order in $\tilde{\gamma}N\tilde{\gamma}$ particle–field correlations appeared and average values of operators were no longer sufficient for a solution.

Since we are now considering the motion of the center of mass we have three sets of variables: (1) electromagnetic field variables; (2) internal degree of freedom variables; and (3) center-or-mass variables. Hence, we have three different kinds of correlations possible. However, since initial correlations are dissipated by the irreversibility and since the pump is too crude to maintain subtle correlations, the only way that correlations can be maintained is by the interaction Hamiltonian. Thus, any correlations that appear must be of the order of $(\tilde{\gamma}N\tilde{\gamma})^2$ or higher. This, of course, is the usual case; plasmas and lasers are exceptional in that, due to the long-range nature of the force, there is a SCFA which depends only on average values with no correlations and which appears to only first order in $\tilde{\gamma}N\tilde{\gamma}$.

The effects of the center-of-mass motion, although important, are small [in the ratio of $(\omega_D/\omega_0) \sim 10^{-6}$], so we expect the SCFA will still be a solution when the atoms are moving. The statement of the SCFA, i.e., "each atom sees all the $(N - 1)$ other atoms to the lowest order through the electromagnetic field," remains true even though the atoms are moving. Their motion will affect the value of the field but not the fact that each atom sees the average field. It is true that new correlations appear between the state of motion of the atoms and the internal state of the atoms, but these are all of order $(\tilde{\gamma}N\tilde{\gamma})^2$ or higher.

Since the derivation of the extended SCFA is a straightforward generalization of Sec. III of I, we do not give a formal derivation here but obtain the equations directly by taking a trace of Eqs. (2.11a), (2.11b), (2.12), and (2.9a) with a product density matrix. The density matrix consists of a product of $N$ one-particle density matrices with $N$ one-particle center of mass distribution functions and with the density matrix of the electromagnetic field. This is completely analogous to writing down the Vlasov equation directly instead of starting with the $N$-particle Liouville equation and deriving the Vlasov equation for the case where there are no initial correlations.

When we take the trace of Eqs. (2.11a) and (2.11b) in the extended SCFA, we obtain

$$\langle\dot{b}\rangle + (\langle b\rangle/T_r) = \omega_L^2\int^t K(t,t')\bar{e}^{\nu_a(t-t')}\,\bar{e}^{i\Delta(t-t')}\,\langle\hat{\sigma}(t')\rangle\langle b(t')\rangle\,dt' \tag{3.1a}$$

$$\langle\dot{b}^\dagger\rangle + (\langle b^\dagger\rangle/T_r) = \omega_L^2\int^t K(t,t')e^{-\nu_a(t-t')}e^{i\Delta(t-t')}\langle\hat{\sigma}(t')\rangle\langle b^\dagger(t')\rangle\,dt', \tag{3.1b}$$

where $\omega_L^2 \equiv \tilde{\gamma}N\tilde{\gamma}\omega_0^2$ and $K(t, t') \equiv \langle\Gamma(t)\Gamma(t')\rangle$. The symbol $\langle O\rangle$ represents the trace of $O$ over the product of $N$ one-particle density matrices with $N$ one-particle distribution functions for the center of mass and with a density matrix for the electromagnetic field.

We obtain the equation of motion for $\langle\hat{\sigma}\rangle$ of Eq. (2.12) with the product density matrix

$$i\langle\dot{\hat{\sigma}}\rangle = 2\tilde{\gamma}\omega_0\{\langle\xi^\dagger\rangle\langle b\rangle e^{i\Delta t} - \langle\xi\rangle\langle b^\dagger\rangle e^{-i\Delta t}\} + (i/T_1)(\hat{\sigma}_0 - \langle\hat{\sigma}\rangle). \tag{3.2}$$

In order to eliminate $\langle\xi^\dagger\rangle$ and $\langle\xi\rangle$, we take the trace of Eq. (2.9a) with the density matrix in the SCFA, and we obtain

$$\begin{aligned} i\langle\dot{b}\rangle + i(\langle b\rangle/T_r) &= \omega_0\tilde{\gamma}N\langle\xi\rangle e^{-i\Delta t}, \\ i\langle\dot{b}^\dagger\rangle + i(\langle b^\dagger\rangle/T_r) &= -\omega_0\tilde{\gamma}N\langle\xi^\dagger\rangle e^{i\Delta t}. \end{aligned} \tag{3.3}$$

When we combine Eqs. (3.2) and (3.3), we obtain

$$\frac{\partial}{\partial t}\left\{\langle\hat{\sigma}\rangle + \frac{2}{N}\langle b^\dagger\rangle\langle b\rangle\right\} = \frac{\hat{\sigma}_0 - \langle\hat{\sigma}\rangle}{T_1} - \frac{4}{N}\frac{1}{T_r}\langle b^\dagger\rangle\langle b\rangle. \tag{3.4}$$

If there were no dissipation Eq. (3.4) would be a constant of the motion which for $\omega_0 = \Omega$ is the energy. The fact that Eq. (3.4) is independent of the center-of-mass variables is a consequence of the SCFA which makes the solution of the equations much easier.

Since $\langle b^\dagger \rangle$ and $\langle b \rangle$ are complex numbers, we find it convenient to define the following quantities

$$\langle b(t) \rangle \equiv f(t)e^{-i\phi(t)} \quad \text{and} \quad \langle b^\dagger(t) \rangle = f(t)e^{i\phi(t)}.$$

When we substitute the above definitions in Eqs. (3.1a) and (3.1b) and add the resulting equations, we obtain

$$\dot{f} + (f/T_r) = \omega_L^2 \int^t K(t, t')e^{-\nu_2(t-t')}\langle \hat{\sigma}(t') \rangle f(t')$$

$$\times \cos [\Delta(t - t') + \phi(t') - \phi(t)]\, dt'. \qquad (3.5)$$

In the steady state the time derivatives of $f$ and $\langle \hat{\sigma} \rangle$ vanish and the time derivative of $\phi$ is a constant. The reason that we have a frequency shift, $\omega_s = \dot{\phi}_s$, in the steady state is that we introduced the $b$ variables by splitting off the unperturbed cavity frequency $\Omega$, not the actual operating frequency $\Omega' = \Omega + \omega_s$. When we set $\dot{f} = 0$ in Eq. (3.5) and let $t \to \infty$, we obtain the steady-state population difference per atom

$$\langle \hat{\sigma} \rangle_s = \left\{ \omega_L^2 T_r \int_0^\infty K_s(\tau)e^{-\nu_2\tau} \cos \Delta'\tau\, d\tau \right\}^{-1}, \qquad (3.6)$$

where $\Delta' \equiv \omega_0 - \Omega - \omega_s = \Delta - \omega_s$ is the difference between the atomic frequency and the actual cavity frequency $\Omega'$. We determine the steady-state frequency shift in Sec. IV. The expression $K_s(\tau)$ is equal to $K_s(t - t')$, where the subscript $s$ indicates steady state, and we use the fact that $K(t, t')$ depends only on the time difference in the steady state as we show in the Appendix.

When we set the time derivative in Eq. (3.4) equal to zero we obtain

$$f_s^2 = (\hat{\sigma}_0 - \langle \hat{\sigma} \rangle_s)(NT_r/4T_1). \qquad (3.7)$$

We conclude this section with a discussion of Eq. (3.4). The general solution of Eq. (3.4) is

$$\langle \hat{\sigma} \rangle + (2/N)f^2 = \hat{\sigma}_0[1 - e^{-t/T_1}]$$

$$- \frac{[2T_1 - T_r]}{T_1T_r} \int^t e^{-\tau/T_1}(2/N)f^2(t - \tau)\, d\tau. \qquad (3.8)$$

If the approach to the steady state is slow compared with $T_1$, then for $t > T_1$, we obtain

$$\langle \hat{\sigma}(t) \rangle + \frac{4}{N}\frac{T_1}{T_r} f^2(t) = \hat{\sigma}_0. \qquad (3.9)$$

We show in Sec. IV that almost always the approach to the steady state is slow compared with $T_1$. Since the steady-state value of Eq. (3.9) is the same as the steady-state equation (3.7), we are guaranteed Eq. (3.9) approaches the correct steady state. Consequently, Eq. (3.9) is usable even if the approach to the steady state is not rapid compared with $T_1$ except for those situations that depend on the detailed wiggles of the approach to the steady state.

## IV. SOLUTION OF THE TIME DEPENDENT EQUATIONS

The solution of the three first-order nonlinear integral equations, Eqs. (3.1a), (3.1b), and (3.4), constitutes the extended SCFA solution to the quantum mechanical Liouville equation where initial correlations are absent or dissipate rapidly. Under these conditions the solution of the three equations is a rigorous solution to order $\tilde{\gamma}N\tilde{\gamma}$. We cannot solve these equations exactly, not so much because of the nonlinearity, but because $K(t, t')$ has a transcendental time dependence even in the simplest case of free particle motion. In a future publication we will show the nonlinear equation can be solved rigorously in closed form for stationary atoms.

As we mentioned in the Introduction, there is an additional dimensionless constant, $\beta = \alpha(\omega_0/\omega_D)$ when we consider the motion of the atoms. Thus, if $\beta$ is very small we can expand in powers of $\beta$ or, conversely, in powers of $\beta^{-1}$. Fortunately, $\beta$ is of the order of 0.1 or less for most gas lasers. Small $\beta$ means physically that the field amplitudes vary slowly compared with the center-of-mass motion and, consequently, the fields do not see the instantaneous velocity of the center of mass but respond to certain integrals of the center-of-mass velocity distribution. However, the exact steady state is a sensitive functional of the center-of-mass velocity distribution function as we show in Sec. VI.

When we assume that $\langle b \rangle$, $\langle b^\dagger \rangle$, and $\langle \hat{\sigma} \rangle$ are slowly varying with respect to the characteristic time scale of the center-of-mass motion, Eqs. (3.1a) and (3.1b) reduce to the following form:

$$\langle \dot{b} \rangle + \frac{\langle b \rangle}{T_r} = \frac{\omega_L^2}{\omega_D}\langle \hat{\sigma} \rangle\langle b \rangle \int_0^{\omega_D t} K(t, t')e^{-\bar{\nu}_2 t'}e^{-i\bar{\Delta}t'}\, dt', \qquad (4.1a)$$

$$\langle \dot{b}^\dagger \rangle + \frac{\langle b^\dagger \rangle}{T_r} = \frac{\omega_L^2}{\omega_D}\langle \hat{\sigma} \rangle\langle b^\dagger \rangle \int_0^{\omega_D t} K(t, t')e^{-\bar{\nu}_2 t'}e^{i\bar{\Delta}t'}\, dt', \qquad (4.1b)$$

where $\bar{\nu}_2 \equiv \nu_2/\omega_D$ and $\bar{\Delta} = \Delta/\omega_D$ are dimensionless variables. We assumed that the largest frequency

associated with the center-of-mass motion is $\omega_D$. If there is any center-of-mass frequency greater than $\omega_D$, it should be used instead of $\omega_D$. The choice of $\omega_D$ does not imply free-particle motion. The term, $K(t, t')$, includes collisions. If in Eqs. (4.1a) and (4.1b) we temporarily introduce a dimensionless time $\tau \equiv \omega_L t$, the dimensionless variable measuring the rate of change of $\langle b \rangle$ and $\langle b^\dagger \rangle$ is $\beta = (\omega_L/\omega_D)$. This is the basis for our assertion that $\beta \ll 1$ implies slow variation. In the next section we show that the systematic solution to Eqs. (3.1a) and (3.1b) is a power series in $\beta$. The solution to lowest order in $\beta$ is shown in Eqs. (4.1a) and (4.1b).

When we substitute Eq. (3.8) in Eqs. (4.1a) and (4.1b), we obtain

$$\dot{f} + \frac{f}{T_r} = \frac{\omega_L^2}{\omega_D} \langle \hat{\sigma} \rangle f \int_0^{\omega_D t} K(t, t') \cos \bar{\Delta} t' e^{-\bar{\nu}_s t'} \, dt', \tag{4.2a}$$

$$\dot{\phi} = \frac{\omega_L^2}{\omega_D} \langle \hat{\sigma} \rangle \int_0^{\omega_D t} K(t, t') \sin \bar{\Delta} t' e^{-\bar{\nu}_s t'} \, dt'. \tag{4.2b}$$

We observe that the requirement of slow variation on $\langle b \rangle$ and $\langle b^\dagger \rangle$ implies the slow variation of $f$ and $\phi$. We simplify Eq. (4.2a) by observing that for $t > \omega_D^{-1}$ the upper limit on the integral approaches infinity and $K(t, t')$ approaches its steady-state value, $K_s(t - t')$. Consequently, for $t > \omega_D^{-1}$, Eq. (4.2a) with the help of Eq. (3.6) reduces to the following form:

$$\dot{f} = f(A - Bf^2), \tag{4.3}$$

where

$$A \equiv \frac{F_c^0 \omega_L^2 \hat{\sigma}_0}{\omega_D} - \frac{1}{T_r}, \qquad B \equiv 2F_c^0 \frac{T_1}{T_r} \frac{2}{N} \frac{\omega_L^2}{\omega_D}$$

and where

$$F_{(s)}^{n(c)} = \int_0^\infty \tau^n K_s(\tau) \begin{pmatrix} \cos \bar{\Delta}\tau \\ \sin \bar{\Delta}\tau \end{pmatrix} e^{-\bar{\nu}_s \tau} \, d\tau. \tag{4.4}$$

The solution of Eq. (4.3) is

$$f(t) = \frac{(A/B)^{\frac{1}{2}} f(0) e^{At}}{\{(A/B) + f^2(0)[e^{2At} - 1]\}^{\frac{1}{2}}}. \tag{4.5}$$

Equation (4.5) states that, no matter what the initial field intensity is if $A > 0$, then for $t \gg A^{-1}$, $f(t)$ approaches $(A/B)^{\frac{1}{2}}$. However, from Eq. (3.7) we see that $(A/B)^{\frac{1}{2}}$ is just $f_s$. If $A < 0$, then $f$ approaches the only other steady state which is $f_s = 0$; i.e., there is zero field. Consequently, the condition $A = 0$ is the threshold condition.

The threshold condition is

$$\mathfrak{N} \langle \hat{\sigma} \rangle_s = \frac{\mathfrak{N} \omega_D}{\omega_L^2 T_r F_c^0(\bar{\Delta}', \bar{\nu}_2)} = \mathfrak{N} \hat{\sigma}_0. \tag{4.6}$$

Equation (4.6) gives the population inversion per unit volume, $\mathfrak{N}\hat{\sigma}_0$, required by the pump to start an oscillation. Since $\omega_L^2$ is proportional to $\mathfrak{N}$, the middle term of Eq. (4.6) is independent of $\mathfrak{N}$. Any increase in pump power beyond the threshold condition, Eq. (4.6), goes directly into the electromagnetic field as we see by rewriting Eq. (3.7) in the form

$$4 \frac{T_1}{T_r} \frac{f_s^2}{V} = \mathfrak{N}(\hat{\sigma}_0 - \langle \hat{\sigma} \rangle_s), \tag{4.7}$$

where $V$ is the volume. Since $\mathfrak{N} \langle \hat{\sigma} \rangle_s$ is independent of $\mathfrak{N}$, the pump must first produce a population inversion density $\mathfrak{N} \langle \hat{\sigma} \rangle_s$ given by Eq. (4.6). This corresponds to zero electromagnetic field. As the pump power is increased the electromagnetic energy density increases linearly with pump power while $\mathfrak{N} \langle \hat{\sigma} \rangle_s$ remains essentially constant. There may be a slight dependence on the pump power of $K_s(\tau)$ through its dependence on the collision frequency which is dependent on pump power.

The above discussion refers to the case where the number of two-level systems per unit volume $\mathfrak{N}$ is fixed and given. Then the threshold condition, Eq. (4.6), is met by increasing $\hat{\sigma}_0$, the fractional inversion per atom. For gas lasers we do not have permanent two-level systems, but they are created by the pump. So for gas lasers we take $\hat{\sigma}_0$ equal to one, and the threshold condition becomes the required number of two-level systems per unit volume $\mathfrak{N}$ the pump must produce to start oscillating. An increase in pump power above threshold, in effect, leads to an increase in the number of two-level systems. If laser action took place between two levels, one of which was the ground state, the number of two-level systems would be the same as the number of atoms. However, as is usually the case, when both initial and final states are excited states, the number of two-level systems is the number of atoms excited to either of these two levels.

When $t > A^{-1}$, we can obtain the first-order frequency shift directly from Eq. (4.2b), which becomes

$$\dot{\phi} = \frac{\omega_L^2}{\omega_D} \langle \hat{\sigma} \rangle_s \int_0^\infty K_s(\tau) e^{-\bar{\nu}_s \tau} \sin \bar{\Delta}\tau \, d\tau = \frac{1}{T_r} \frac{F_s^0}{F_c^0}, \tag{4.8}$$

where we use Eq. (4.4). The first-order frequency shift, Eq. (4.8), is equivalent to Lamb's[5] first-order frequency shift in the case where we assume $K_s(\tau)$ depends only on free-particle motion.

The condition that we can use the simpler Eq. (3.10) for the energy instead of Eq. (3.9) is

[5] W. E. Lamb, Phys. Rev. 134, A1429 (1964).

$$\frac{T_1}{T_r}\left[\frac{\hat{\sigma}_0}{\langle\hat{\sigma}\rangle_s} - 1\right] < 1.$$

For a typical gas laser this condition is valid for pump power up to a hundred times threshold. If it is violated we use Eq. (3.9) which would not change the first-order frequency shift and would not affect the steady state. It would make a small correction to the second-order frequency shift and to the rate of approach to the steady state.

In the next section we show that as you go to higher and higher powers of $\beta$ the steady state is unaffected, only the rate of approach changes. However, the frequency shift $\phi$ is a power series in $\beta$ and depends on the rate of approach to the steady state in the second and higher orders of $\beta$.

We have shown that, although the nonlinearity is small, it serves two important functions. First, out of all possible periodic solutions to the linearized equations it selects only one, namely the steady-state equation (4.7). The second function is that it fixes the form of the steady state independent of the dynamics of the approach. The first function requires the nonlinearity as measured by $\beta$ to be small in the sense of the Bogoliuboff–Kryloff[2] quasi-linear theory. The second function is completely independent of the smallness of $\beta$.

## V. CALCULATION OF FREQUENCY SHIFT TO SECOND ORDER

In this section we demonstrate the generalization of our lowest-order calculation to higher powers of $\beta$. In the first-order calculation we replace $\langle\hat{\sigma}(t-\tau)\rangle\langle b(t-\tau)\rangle$ in Eqs. (3.1a) and (3.1b) by $\langle\hat{\sigma}(t)\rangle\langle b(t)\rangle$. In order to go to an arbitrary order we expand $\langle\hat{\sigma}(t-\tau)\rangle\langle b(t-\tau)\rangle$ in a Taylor series

$$\langle\hat{\sigma}(t-\tau)\rangle\langle b(t-\tau)\rangle = \langle\hat{\sigma}(t)\rangle\langle b(t)\rangle - \tau\frac{\partial}{\partial t}[\langle\hat{\sigma}(t)\rangle\langle b(t)\rangle] + \frac{\tau^2}{2}\frac{\partial^2}{\partial t^2}[\langle\hat{\sigma}(t)\rangle\langle b(t)\rangle] + \cdots . \tag{5.1}$$

When the first two terms of the Taylor series, Eq. (5.1), are substituted in Eq. (3.1a), we obtain

$$\langle\dot{b}\rangle + \frac{\langle b\rangle}{T_r} = \frac{\omega_L^2}{\omega_D}\langle\hat{\sigma}(t)\rangle\langle b(t)\rangle\int^{\omega_D t} K(t,t')e^{-i\bar{\Delta}t'}e^{-i\bar{\gamma}_a t'}\,dt' - \left(\frac{\omega_L}{\omega_D}\right)^2\langle b(t)\rangle\frac{\partial\langle\hat{\sigma}(t)\rangle}{\partial t}\int^{\omega_D t} t'K(t,t')e^{-i\bar{\Delta}t'}e^{-\bar{\gamma}_a t'}\,dt' - \left(\frac{\omega_L}{\omega_D}\right)^2\langle\hat{\sigma}(t)\rangle\frac{\partial\langle b(t)\rangle}{\partial t}\int^{\omega_D t} t'K(t,t')e^{-i\bar{\Delta}t'}e^{-\bar{\gamma}_a t'}\,dt'. \tag{5.2}$$

For $t > \omega_D^{-1}$ we obtain

$$\left[1 + \left(\frac{\omega_L}{\omega_D}\right)^2 F^1\langle\hat{\sigma}\rangle\right]\langle\dot{b}\rangle + \frac{\langle b\rangle}{T_r} = \frac{\omega_L^2}{\omega_D}F^0\langle\hat{\sigma}\rangle\langle b\rangle - \left(\frac{\omega_L}{\omega_D}\right)^2 F^1\langle b\rangle\frac{\partial\langle\hat{\sigma}\rangle}{\partial t}, \tag{5.3}$$

where

$$F^{(n)} = \int_0^\infty \tau^n K_s(\tau)e^{-i\bar{\Delta}\tau}e^{-\bar{\gamma}_a\tau}\,d\tau.$$

The corresponding equation for $\langle b^\dagger\rangle$ is

$$\left[1 + \left(\frac{\omega_L}{\omega_D}\right)^2 (F^1)^*\langle\hat{\sigma}\rangle\right]\langle\dot{b}^\dagger\rangle + \frac{\langle b^\dagger\rangle}{T_r} = \frac{\omega_L^2}{\omega_D}(F^0)^*\langle\hat{\sigma}\rangle\langle b^\dagger\rangle - \left(\frac{\omega_L}{\omega_D}\right)^2(F^1)^*\langle b^\dagger\rangle\frac{\partial\langle\hat{\sigma}\rangle}{\partial t}. \tag{5.4}$$

We obtain the equation of motion for the amplitude $f(t)$ when we add Eqs. (5.3) and (5.4),

$$\left[1 + \left(\frac{\omega_L}{\omega_D}\right)^2 F_c^1\left(\langle\hat{\sigma}\rangle - \frac{8}{N}\frac{T_1}{T_r}f^2\right)\right]\dot{f} + \frac{f}{T_r} = \frac{\omega_L^2}{\omega_D}f\langle\hat{\sigma}\rangle F_c^0, \tag{5.5}$$

where we have used Eq. (4.4).

When we set $\dot{f} = 0$ in Eq. (5.5) we obtain the steady-state condition, Eq. (3.6), for $\langle\hat{\sigma}\rangle_s$. The steady-state equation (4.7) is the same to all orders in $\beta$. No matter how many terms in Eq. (5.1) we retain, they contain only time derivatives of $f$ multiplied by functions of $f$ and $\langle\hat{\sigma}\rangle$. Consequently, in the steady state where all time derivatives vanish we always obtain Eq. (3.6). When we retain higher and higher powers of $\beta$, we do not affect the steady state but we obtain a more accurate description of the approach to the steady state. This is analogous to the nonlinear Boltzmann equation where the solution of the steady-state condition is local equilibrium independent of the force law but where the nonlinear approach to the steady state is very complicated and extremely sensitive to the force law.

When we subtract Eq. (5.4) from Eq. (5.2) and consider times $t > \omega_D^{-1}$, we obtain

$$\dot{\phi} = \frac{\omega_L^2}{\omega_D}\langle\hat{\sigma}(t)\rangle F_s^0\left[1 - \left(\frac{\omega_L}{\omega_D}\right)^2\langle\hat{\sigma}(t)\rangle F_c^1\right] - \left(\frac{\omega_L}{\omega_D}\right)^2 F_s^1\frac{\partial\langle\hat{\sigma}\rangle}{\partial t}.$$

If we now let $t \to \infty$ and use Eq. (3.6), we obtain

$$\dot{\phi} = \frac{F_s^0}{F_c^0}\frac{1}{T_r}\left[1 - \frac{F_c^1}{F_c^0}\frac{1}{\omega_D T_r}\right]. \tag{5.6}$$

Equation (5.6) is the phase shift correct to second order in $\beta$. If we retain $n$ terms in Eq. (5.1) the $n$th-order phase shift is a sum of terms in which the $n$th term is proportional to $F_c^n$. Since the steady-state

condition only requires $\phi$ to be a constant, we have to retain an infinite number of terms in Eq. (5.1) to get the rigorous frequency shift in the steady state. If $\beta$ is small the series converges rapidly. Since $(F_c^1/F_c^0)$ is of order one, the second term on the right-hand side of Eq. (5.6) is less than .001 in typical gas lasers at room temperature.

## VI. STEADY STATE OF A GAS LASER

The steady-state electromagnetic energy density is a transcendental function of $\bar{\nu}_2$ and $\bar{\Delta}$. Although we are unable to investigate the function analytically, we can obtain many of its properties qualitatively. In particular, we show that the structure of the energy density versus cavity frequency curve is a sensitive function of $\bar{\nu}_2$.

When we combine Eqs. (4.7) and (4.6), we obtain

$$\frac{4T_1}{T_r}\frac{f_s^2}{V} = \mathfrak{N} - \frac{\mathfrak{N}(\omega_D/\omega_L)}{\omega_L T_r F_c^0(\bar{\nu}_2, \bar{\Delta})}, \tag{6.1}$$

where we have set $\hat{\sigma}_0 = 1$. We consider a gas laser with free-particle motion, thus $K_s(\tau)$ in $F_c^0$ is $\exp(-\frac{1}{2}\tau^2)$. When $\bar{\nu}_2 = 0$, Eq. (6.1) is the usual Doppler profile. Since $F_c^0$ is an even function of $\bar{\Delta}$, the curve is a symmetrical function of $\Omega$ about $\omega_0$. We obtain the cutoff frequency $\Omega^*$ by setting $f_s = 0$ in Eq. (6.1). The energy density has relative maximum at $\bar{\Delta} = 0$ for all $\bar{\nu}_2$. However, as $\bar{\nu}_2$ increases this relative maximum grows considerably less than the Doppler curve at $\bar{\Delta} = 0$. Consequently, as we show below there can be—and is—a dip in the center of the Doppler curve if $\bar{\nu}_2$ is sufficiently large.

The necessary condition that Eq. (6.1) has a minimum is that $\partial F_c^0/\partial\bar{\Delta}$ vanishes and the sufficient condition is that $\partial^2 F_c^0/\partial\bar{\Delta}^2$ is greater than zero. We find there is a minimum for $\bar{\Delta}^2 \ll 1$ which we denote by $\bar{\Delta}_m$,

$$\bar{\Delta}_m^2 = 6\left[\frac{d_2(\bar{\nu}_2)}{d_4(\bar{\nu}_2)}\right] = 6\left[\frac{1-\bar{\nu}_2^2}{3+5\bar{\nu}_2^2+\bar{\nu}_2^4}\right], \tag{6.2}$$

where

$$d_{2n}(\bar{\nu}_2) = \int_0^\infty \tau^{2n} e^{-\bar{\nu}_2\tau}\, d\tau.$$

For Eq. (6.2) to be true $\bar{\Delta}_m$ must be much smaller than one, which means that $\bar{\nu}_2$ must be much greater than two. However, such a large $\bar{\nu}_2$ requires a very large pump power and is unobservable.

If we drop the condition that $\bar{\Delta}_m$ be very small and ask only that $|\bar{\Delta}_m| \lesssim 1$, the condition on $\bar{\nu}_2$ becomes considerably less severe. The exact location of the minimum requires the solution of the transcendental equation

$$\int_0^\infty \tau \sin \bar{\Delta}_m\tau e^{-\bar{\nu}_2\tau} e^{-\tau^2/2}\, d\tau = 0. \tag{6.3}$$

When $\bar{\nu}_2$ is $O(.1)$, $|\bar{\Delta}_m|$ is less than or equal to one. In this region the minimum condition is very sensitive to changes in $\bar{\nu}_2$. We show the qualitative structure of the energy density curve in this region in Fig. 1.

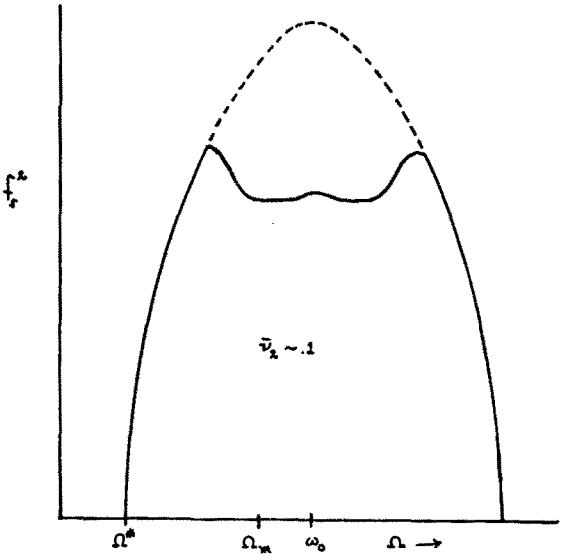


FIG. 1. Dimensionless energy density as a function of cavity frequency $\Omega$ for $\bar{\nu}_2 \sim 0.1$ and for $|\Omega^*| > |\Omega_m|$. The curve without the dip is the Doppler profile. If $|\Omega_m| \gg |\Omega^*|$ the only observable effect is a flattening of the Doppler peak.

If $\bar{\nu}_2$ is very much smaller than one, there is no solution for $|\bar{\Delta}_m| \lesssim 1$. In the He–Ne gas laser under normal conditions $\bar{\nu}_2 \sim .02$.

When we expand Eq. (6.3) to first order in $\bar{\nu}_2$, we obtain

$$\bar{\Delta}_m \exp(-\bar{\Delta}_m^2/2) = \bar{\nu}_2 \frac{\partial}{\partial\bar{\Delta}} \int_0^\infty \tau \cos\bar{\Delta}\tau \exp(-\tau^2/2)\, d\tau \Big|_{\bar{\Delta}=\bar{\Delta}_m} \equiv \bar{\nu}_2 \frac{\partial}{\partial\bar{\Delta}} W(\bar{\Delta}) \Big|_{\bar{\Delta}=\bar{\Delta}_m}. \tag{6.4}$$

There is no analytic expression for $W(\bar{\Delta})$; however, for $\bar{\Delta} > 1$, $W(\bar{\Delta})$ approaches $-(\bar{\Delta})^{-2}$. Thus, for $\bar{\Delta} > 1$, we obtain

$$\bar{\Delta}_m^4 \exp(-\tfrac{1}{2}\bar{\Delta}_m^2) = 2\bar{\nu}_2.$$

The solution corresponding to a minimum occurs for $\bar{\Delta} > 2$. Consequently, for $\bar{\nu}_2 \ll 1$, there are symmetrical dips in the energy density curve as shown in Fig. 2. However, $\bar{\Delta} > 2$ means the dips occur for $\Omega$ greater than twice the Doppler width. As a result

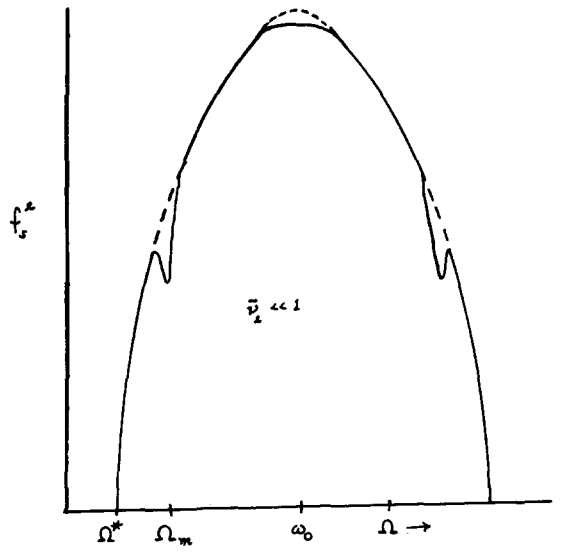


FIG. 2. Dimensionless energy density as a function of cavity frequency $\Omega$ for $\bar{\nu}_2 \ll 1$ and for $|\Omega^*| > |\Omega_m|$. The curve without the side dips is the Doppler profile. If $|\Omega_m| > |\Omega^*|$ then only the peak above the side dips is observable.

the side dips will usually be unobservable because the power requirement is too high and because multimode operation would mask the dips.

The relaxation frequency $\bar{\nu}_2$ is a purely phenomenological constant representing relaxation of the off-diagonal matrix elements of the density matrix for the internal degrees of freedom. We have shown that the structure of the energy density as a function of $\Omega$ is very sensitive to changes in $\bar{\nu}_2$. Collisions have the qualitative effect of increasing $\bar{\nu}_2$ because, as we show in the Appendix, the presence of collisions changes $\bar{\nu}_2$ to $\bar{\nu}_2 + \bar{\nu}_c$, where $\bar{\nu}_c$ is the collision frequency. Consequently, any change in conditions that increase $\bar{\nu}_c$ will effectively increase $\bar{\nu}_2$. Since we expect $\bar{\nu}_c$ to depend at least weakly on pump power, the effective $\bar{\nu}_2$ will depend on pump power.

As a result of the above discussion we see that a center dip may not appear in the energy density curve, but if $\bar{\nu}_2$ is increased sufficiently a dip will appear. In recent experiments[3,4] no dip was observed until $\bar{\nu}_2$ was increased (by changing isotope mass) and then the dip appeared. Also, in addition to the usual power dependence, the dip structure has a power dependence which might be due to $\bar{\nu}_c$'s power dependence.

## VII. DISCUSSION

We have shown that the SCFA for a system of $N$ two-level systems remains valid when we include the center-of-mass motion, dissipation, and pumping. The nonlinearity, although small, plays two fundamental roles. The nonlinearity determines the steady state and, dynamically, the nonlinearity is responsible for picking out the steady state as the only allowed solution of the manifold of solutions of the linearized theory. Bogoliuboff–Kryloff[2] call this behavior quasi-linear. The dominant condition necessary for quasi-linearity is $\beta \ll 1$. In our quasi-linear theory the steady-state conditions determine the steady-state amplitude $f_s$ and the steady-state population inversion $\langle \sigma \rangle_s$. A dynamical solution to all orders in $\beta$ affects only the rate at which these steady-state values are approached. Consequently, as long as $\beta \ll 1$, a solution to all orders in $\beta$ reduces essentially to a calculation of the frequency shift to all orders in $\beta$.

We observe that if simple rate equations have the correct steady-state values for $f_s^2$ and $\langle \sigma \rangle_s$ built into them and if the dynamics of the approach to the steady state are not determined, then the only difference in the single mode steady-state predictions will be that the SCFA equations determine the frequency shift and the rate equations do not. However, we emphasize that we proved that as a result of the nonlinearity the steady state is approached rapidly; whereas, in the simple rate equations the approach to steady state is built into the form of the equations by assumption.

If $\beta \gg 1$, the radiation and matter would exchange large amounts of energy before the center-of-mass motion has an appreciable effect. Consequently, we would have relaxation oscillations. Since relaxation oscillations are not observed in gas lasers, we may infer that $\beta$ is of order one or smaller. This dependence of the solution on $\beta$ is very similar to the Van der Pol equation. When the nonlinearity is small, the Van der Pol equation behaves quasi-linearly. When the nonlinearity is large, the Van der Pol equation has relaxation oscillations. To emphasize the similarity of our equations with the Van der Pol equation, we point out that Eq. (4.3) is the quasi-linear theory amplitude equation for the Van der Pol equation. However, there are differences. In the Van der Pol equation the first-order frequency shift is zero; whereas, we obtain the frequency shift given in Eq. (4.8).

We are now carrying out the multimode case. Since the number of frequency shifts the theory must predict increases, the difference between the SCFA and the simple rate equations increases. In the present paper the pump and the dissipation are treated as fixed numbers. Since they are really stochastic variables, we are also generalizing the present formalism to include noise. Above the threshold the

main effect of noise is to give our steady-state oscillation line a width. The shape of the line depends on the amplitude of the noise and on the noise correlation time. Below threshold the problem is much more difficult because there is competition between noise and particle–field correlations.

## APPENDIX

The Hamiltonian for the center of mass is

$$H_{\rm cm} = \frac{1}{2m}\sum_{\alpha}^{N} P_\alpha^2 + \frac{1}{2}\sum_{\alpha}^{N}\sum_{\beta}^{N} V(X_\alpha - X_\beta) + \sum_{\alpha}^{N}\sum_{i}^{M} U(X_\alpha - \eta_i),$$

where we are considering a model of $N$ active atoms and $M$ pump atoms. The potential energy between active atoms is $V(X_\alpha - X_\beta)$ and $U(X_\alpha - \eta_i)$ is the potential energy of interaction between the $\alpha$th active atom and the $i$th pump atom located at the point $\eta_i$.

In order to obtain the equation of motion for the distribution function of the center-of-mass coordinates, we take the trace over all variables (except the center-of-mass variables) in the equation of motion for the density matrix of the entire system. The result is

$$\frac{\partial F_N(X_1 \cdots P_N)}{\partial t} + \sum_\alpha \frac{P_\alpha}{m}\frac{\partial F_N}{\partial X} - \sum_\alpha \frac{\partial}{\partial X_\alpha}\Big[\sum_\beta V(X_\alpha - X_\beta)\Big]\frac{\partial F_N}{\partial P_\alpha} - \sum_\alpha \frac{\partial}{\partial X_\alpha}\Big[\sum_i^M U(X_\alpha - \eta_i)\Big]\frac{\partial F_N}{\partial P_\alpha} - \gamma\hbar\omega_0 \sum_\alpha [\langle a^\dagger \sigma_\alpha\rangle + \langle a\sigma_\alpha^\dagger\rangle]\frac{\partial \Gamma}{\partial X_\alpha}\frac{\partial F_N}{\partial P_\alpha} = 0. \quad \text{(A.1)}$$

Except for the last term in Eq. (A.1) we have the classical Liouville equation for $N$ atoms interacting through $V(X_\alpha - X_\beta)$ and interacting with $M$ pump atoms through $U(X_\alpha - \eta_i)$. The last term in Eq. (A.1) represents the effect of the radiation processes on the center-of-mass motion, and it is small compared to the other terms in the equation. We neglect this term because the lowest order to which it contributes is $(\tilde{\gamma}N\tilde{\gamma})^2$, and it makes no contribution to the SCFA.

If we neglected the interaction of the atoms among themselves, we could reduce Eq. (A.1) to a linear equation for $F_1$, the one-particle distribution function. This is often a good approximation since collisions of active atoms with pump atoms are often $10^3$ times more frequent than collisions with active atoms. However, in any case, the density of atoms is sufficiently low so that a Boltzmann description is valid and Eq. (A.1) reduces to the following nonlinear equation for $F_1$,

$$\frac{\partial F_1(X, V, t)}{\partial t} + V\frac{\partial F_1}{\partial X} = J_{11}(F_1) + J_{12}(F_1), \quad \text{(A.2)}$$

where

$$J_{11}(F_1) = \int_{-\infty}^{\infty} d^3V_2 \int_0^{2\pi} d\epsilon \int_0^{\infty} db\, b\, |V - V_2| \times \{F_1(V')F_1(V_2') - F_1(V)F_1(V_2)\},$$

$$J_{12}(F_1) = \int_{-\infty}^{\infty} d^3V_2 \int_0^{2\pi} d\epsilon \int_0^{\infty} db\, b\, |V - V_2| \times \{F_1(V')\mathfrak{F}_1(V_2') - F_1(V)\mathfrak{F}_1(V_2)\},$$

and where $\mathfrak{F}_1(X, V, t)$ is the distribution function for the pump atoms which we assume we know. $V'$ and $V_2'$ is the solution of the two-body problem for initial velocities $V$ and $V_2$. The impact parameter is $b$. We can linearize the nonlinear term in Eq. (A.2), and we obtain

$$\frac{\partial F_1(X, V, t)}{\partial t} + V\frac{\partial F_1}{\partial X} = \int L_{11}(V, V')F_1(X, V', t)\, d^3V' + \int L_{12}(V, V')F_1(X, V', t)\, d^3V', \quad \text{(A.3)}$$

where the term with $L_{11}$ is the linearized Boltzmann equation. The term with $L_{12}$ is the linear Boltzmann operator for collisions with pump atoms and is a function of $\mathfrak{F}(X, V, t)$, the density of pump atoms.

The functional of the center-of-mass motion needed in Sec. III is

$$K_s(t, t') \equiv \langle \Gamma(t)\Gamma(t')\rangle = \int \Gamma(X')\mathcal{G}[X'V' \mid XV, (t - t')] \times \Gamma(X)F_1^0(XV)\, d^3X\, d^3V\, d^3X'\, d^3V', \quad \text{(A.4)}$$

where $F_1^0$ is the steady-state solution of Eq. (A.3) and $\mathcal{G}[X'V' \mid XV, (t - t')]$ is the Green's function of Eq. (A.3). The Green's function depends only on the time difference. For free particles we can write Eq. (A.4) in the alternate form,

$$K_s(\tau) = \int \Gamma(X - V\tau)\Gamma(X)F_1^0(X, V)\, d^3X\, d^3V. \quad \text{(A.5)}$$

When we neglect collisions, take $F_1^0$ to be the Maxwell–Boltzmann distribution, and consider rec-

tangular geometry, we obtain

$$K_s(\tau) = 2\left(L^{-3}\int \sin^2 kX\, d^3X\right)\exp(-\omega_D^2\tau^2/2) \approx \exp(-\omega_D^2\tau^2/2),$$

where $\omega_D^2 \equiv k^2(k_BT)m^{-1}$. We have replaced $L^{-1}\int \sin^2 X\, dX$ by $\frac{1}{2}$ since the remainder is proportional to $(\lambda/2L)$ where $\lambda$ is the wavelength of the cavity mode and $L$ is the length of the cavity. In most gas lasers the ratio is of the order of $10^{-6}$. However, in multimode operation we have terms that are proportional to the difference of the wavenumbers, $k - k'$, where $k'$ is a nearby mode. Such terms are sensitive to large scale density variations including density gradients at the walls. Thus, in multimode situations we must be more careful about ignoring the spatial dependence of $F_1^0$.

If we replace the collision operators in the right-hand side of Eq. (A.3) by a relaxation time approximation, the Green's function is

$$\mathcal{G}[kV \mid kV', \tau] = \delta^3(V - V')e^{-\tau/\tau_c}[\cos(kV\tau)].$$

Then $K_s(\tau)$ is

$$K_s(\tau) = \exp(-\omega_D^2\tau^2/2)\exp(-\nu_c\tau),$$

where $\nu_c$ is the collision frequency. The inclusion of a collision frequency is thus formally the same as replacing $\nu_2$ by $(\nu_2 + \nu_c)$ in the SCFA.

# Null Fields in Einstein–Maxwell Field Theory

M. WYMAN AND R. TROLLOPE
*University of Alberta, Edmonton, Alberta, Canada*
(Received 12 April 1965)

The role of the electromagnetic null field, within the framework of classical relativistic electrodynamics, has never fully established. The object of this paper is to present a large class of exact solutions to the Einstein–Maxwell null field equations with the hope that the physical implications of these solutions may lead to a more thorough understanding of these fields.

In the present analysis, it appears that the nature of the solutions depends on the form of the propagation vector. In the cases for which the divergence of this vector vanishes, an extensive class of solutions is obtained and particular examples are considered. The analysis of the case when the divergence is nonzero is far from complete and only a particular solution is given.

## 1. INTRODUCTION

In its original form, Einstein's theory of gravitation is based on the representation of the gravitational field by a symmetric tensor whose components, $g_{ij}$, are determined from the field equations

$$R_{ij} - \tfrac{1}{2}g_{ij}R = -kT_{ij}, \tag{1.1}$$

where $R_{ij}$ is the Ricci tensor based on $g_{ij}$, $R = g^{ij}R_{ij}$, $k$ is a constant and $T_{ij}$ is the energy–momentum tensor. It is not difficult to incorporate into this theory Maxwell's electromagnetic field theory. This can be done by writing the energy–momentum tensor in the form

$$T_{ij} = T_{ij}^{(1)} + T_{ij}^{(2)}, \tag{1.2}$$

where $T_{ij}^{(1)}$ represents the gravitational and $T_{ij}^{(2)}$ the electromagnetic contributions to the energy-momentum tensor. In the empty space surrounding material bodies, $T_{ij}^{(1)} \equiv 0$, and, in such an instance, we need only obtain the form of $T_{ij}^{(2)}$. It has been shown[1] that under certain circumstances, the form of the electromagnetic energy-momentum tensor is

$$T_{ij}^{(2)} = -\tfrac{1}{4}g_{ij}F_{\alpha\beta}F^{\alpha\beta} + F_{i\alpha}F^{\alpha}_{j}, \tag{1.3}$$

where $F_{ij}$ is a skew-symmetric tensor satisfying the equations

$$F_{ij;k} + F_{jk;i} + F_{ki;j} = 0, \tag{1.4}$$

$$F^{ij}{}_{;j} = 0, \tag{1.5}$$

where the semi-colon denotes covariant differentiation based on the Christoffel symbols $\left\{ \begin{matrix} i \\ jk \end{matrix} \right\}$.

Since a choice of units can be made so that $k = -1$, it is usual to take

$$R_{ij} - \tfrac{1}{2}g_{ij}R = F_{i\alpha}F^{\alpha}_{j} - \tfrac{1}{4}g_{ij}F_{\alpha\beta}F^{\alpha\beta}, \tag{1.6}$$

$$F_{ij;k} + F_{jk;i} + F_{ki;j} = 0, \tag{1.7}$$

$$F^{ij}{}_{;j} = 0, \tag{1.8}$$

to be the field equations required to describe both gravitational and electromagnetic field phenomena in the empty space surrounding material bodies.

In 1924, Rainich[2] showed that, if $R_{\alpha\beta}R^{\alpha\beta} \neq 0$ then the Einstein–Maxwell equations are equivalent to

$$R = 0, \qquad R_{i\alpha}R^{\alpha j} = \tfrac{1}{4}\delta^{j}_{i}R_{\alpha\beta}R^{\alpha\beta}, \tag{1.9}$$

$$\alpha_{i;j} - \alpha_{j;i} = 0, \tag{1.10}$$

where

$$\alpha^{i} \equiv \frac{e^{ijkl}R_{j\alpha;k}R^{\alpha}_{l}}{(-g)^{\frac{1}{2}}R_{\alpha\beta}R^{\alpha\beta}}, \tag{1.11}$$

$e^{ijkl}$ being the well-known permutation symbol.

Fields for which $R_{\alpha\beta}R^{\alpha\beta} = 0$ have now become known as null fields. Rainich felt that the invariant $R_{\alpha\beta}R^{\alpha\beta}$ should behave like a regular function of a complex variable. If such was the case, then $R_{\alpha\beta}R^{\alpha\beta} \neq 0$ in any domain of space-time would imply it was non-zero except for a finite number of points in any finite region. This argument is not very convincing because it is not true, for example, for the invariant $R = g^{ij}R_{ij}$, which is zero outside matter and non-zero at points inside matter. As a second example, one has in Newtonian gravitational theory the fact that the potential inside a spherical shell is constant and outside it is a variable.

It is clear, however, that if null fields cannot be excluded, then the Rainich formulation is not equivalent to the Einstein–Maxwell formulation of a combined gravitational and electromagnetic field theory.

The Rainich result was rediscovered by Misner and Wheeler[3] who were attempting to reformulate

[1] A. S. Eddington, *The Mathematical Theory of Relativity*, (Cambridge University Press, London, 1937), 2nd ed., p. 132.

[2] G. Y. Rainich, Trans. Am. Math. Soc. **27**, 106, (1925).

[3] C. W. Misner and J. A. Wheeler, Ann. Phys. **2**, 525, (1957).

many of the fundamental concepts of classical physics. They noted, of course, the difficulty that arises in the consideration of null fields and expressed the hope that such fields might be incorporated into the Rainich result by a limiting process or by a more general formulation of a similar type. Peres[4] proved that the first hope could not be obtained and Hlavatý[5] proved that a general formulation along the lines of the Rainich result does exist which does not exclude null fields.

Null fields have been the subject of many investigations and, indeed, have been the subject of several controversies. The subject of some of these controversies has been whether or not one could exclude such fields on physical or other grounds.

Our own interest in null fields does not stem from the above-mentioned controversies, nor for that matter, in an interest in incorporating them into a framework of the Rainich type. As will appear later, there seems very little reason why one should abandon the original formulation given by Einstein. With this attitude our interest in null fields is almost entirely due to the fact that it is possible to obtain a large class of solutions to the field equations which determine null fields. Indeed, it may be that we are very close to being able to give a complete set of explicit solutions. The major objective of the present paper is to establish these solutions.

Misner and Wheeler adopted the name "an already unified field theory" to describe the Rainich formulation of the field theory. Since a physical theory does not stand or fall on the name given to the theory, the discussion we give concerning the right of this theory to be called a unified field theory must be considered as belonging only to the field of semantics. For this reason, the second section is in no way pertinent to the major objective of the present paper. Since several authors seem to have accepted the Rainich formulation as a unified field theory, it seems appropriate to at least voice some objections to this classification of the Rainich–Einstein–Maxwell field theory.

## 2. AN ALREADY UNIFIED FIELD THEORY

The name unified field theory would seem to imply the existence of one or more field theories in which two or more distinct fields play fundamental roles. It may be possible to formulate a theory in which these two fields become different manifestations of one and the same field. Einstein did this in the special theory by unifying the electrostatic and electromagnetic fields into one field.

In building a unified field theory it is important to determine the number of components the fundamental field quantity will have in any given coordinate system. This is not a trivial problem because it is not necessarily true that the number of components of the unified field should equal the total number of components the two fields had in the older theories. We need only site the example of the generalization of Newtonian gravitational theory, depending on one potential function; and Maxwell's electromagnetic theory depending on four potential functions, to the Einstein–Maxwell theory in which there are fourteen potential functions. Having determined the form of mathematical representation of the field, one must choose field equations by means of which the components of the field become determined. Normally these equations will be partial differential equations of some specified order. The question of whether this new theory should now be considered as a theory of only one field, or a theory of two fields in which interaction may take place, or, for that matter, still two distinct theories of two fields, with no interaction, is not an easy one to answer. Indeed, it is doubtful if all physicists would answer the question in the same way.

Misner and Wheeler adopt the point of view that the fundamental field has ten components, and is represented by the metric tensor $g_{ij}$. The field equations (1.9)–(1.11) are a mixture of second- and fourth-order partial differential equations. Misner and Wheeler go on to show that physical phenomena that were formerly classified as either gravitational or electromagnetic could now be explained in terms of one field, namely the metric field $g_{ij}$. In this sense, one might say that a unified field theory has been produced. This is, however, a point of view we find difficult to accept. In accepting a given theory as a unified field theory, it would seem to us that it is more important to prove that certain equivalent forms of the theory do not exist than it is to prove that the theory can be placed into a certain specified form. For example, if we have two distinct theories based on real field equations $L = 0$ and $M = 0$, then we would hardly say a unification takes place if ultimately in some new theory the field equations take the form $L^2 + M^2 = 0$.

If we indeed leave aside the question of null fields, then the physical content of any theory based on Eqs. (1.9)–(1.11) cannot be different from the physical content based on the field equations (1.6)–

[4] A. Peres, Phys. Rev. 118, 1105, (1960).
[5] V. Hlavatý, J. Math. Phys. (Cambridge) 40, 1, (1961).

(1.8). From our point of view, the question of whether the theory is indeed a unified field theory must be answered in this formulation as well as the Rainich formulation.

Let us for a moment drop Eq. (1.8) and consider only

$$R_{ij} - \tfrac{1}{2}g_{ij}R = F_{i\alpha}F_j^{\alpha} - \tfrac{1}{4}g_{ij}F_{\alpha\beta}F^{\alpha\beta}, \quad (2.1)$$

$$F_{ij;k} + F_{jk;i} + F_{ki;j} = 0. \quad (2.2)$$

If we use a well-known calculation involving the conservation identities for the electromagnetic energy–momentum tensor, we find

$$R^{\alpha}{}_{i;\alpha} - \tfrac{1}{2}R_{,i} = F^{\alpha\beta}{}_{;\beta}F_{\alpha i}. \quad (2.3)$$

However the Bianchi identity implies that the left-hand side of (2.3) is identically zero, therefore

$$F^{\alpha\beta}{}_{;\beta}F_{\alpha i} = 0. \quad (2.4)$$

If $|F_{ij}| \neq 0$ then (2.4) implies

$$F^{\alpha\beta}{}_{;\beta} = 0. \quad (2.5)$$

Although $|F_{ij}| = 0$ is a necessary condition for null fields, it is not a sufficient condition. Hence the exclusion of fields for which $|F_{ij}| = 0$ excludes more than null fields. In any event, it would indeed seem that the condition $F^{\alpha\beta}{}_{;\beta} = 0$ might, under certain circumstances, legitimately be considered as part of the equations of integrability of (2.1) and (2.2). In fact, under rather mild restrictions on the class of solutions of (2.1) and (2.2) which will be of interest, the same argument can be extended to fields for which $|F_{ij}| = 0$. The argument depends on a limiting process.

Although (2.2) is seemingly expressed in terms of both fields $F_{ij}$ and $g_{ij}$, this is not the case. Since $F_{ij}$ is skew-symmetric, (2.2) can be written

$$F_{ij,k} + F_{jk,i} + F_{ki,j} = 0,$$

the comma denoting partial differentiation. But this simply implies the existence of a vector $\varphi_i$ such that

$$F_{ij} = \varphi_{i,j} - \varphi_{j,i} \quad (2.6)$$

and this, of course, is quite independent of the metric $g_{ij}$. It therefore would seem to us that in this form the theory should be considered to be a theory of two distinct fields in which interaction takes place, and should not be considered a unified field theory.

Although there is nothing sacred about second-order partial differential equations, the possibility of a theory demanding field equations of the fourth order is disturbing. The number of possible tensors one can form from $g_{ij}$ and its derivatives multiplies rapidly as the order of the derivatives increases. One can give rather simple satisfying reasons why the Einstein–Maxwell field equations should be chosen if we restrict ourselves to second-order equations. On the other hand, it is extremely doubtful if one could give similar reasons why the Rainich formulation should be chosen over all possible theories which could be advanced when the freedom of fourth-order partial differential equations is allowed.

Finally, the replacement of a given set of field equations by a set of equations that are essentially equations of integrability would we believe make it possible to call any theory, involving the interaction of two fields, a unified field theory. For example, consider the following simple generalization of Einstein's gravitational theory for empty space. Suppose as before $g_{ij}$ is the metric tensor, and $a_{ij}$ is a second symmetric tensor representing a field of some kind. Consider the field equations

$$a_{ij;k} = 0, \quad (2.7)$$

$$R_{ij} = a_{ij}, \quad (2.8)$$

where covariant differentiation is based on the Christoffel symbols. Obviously, $a_{ij} = 0$ gives the usual form of Einstein's gravitational theory, and in fact $a_{ij} = \lambda g_{ij}$, $\lambda$ constant, is a solution which yields the cosmological form of this theory. Other solutions for $a_{ij}$ also exist. In any event, (2.7) and (2.8) are completely equivalent to the equations of integrability

$$R_{ij;k} = 0, \quad (2.9)$$

equations which are independent of $a_{ij}$. This is of course a trivial and artificial example. It does, however, illustrate the point we wish to make. There must be more to a unified field theory than the use of mere algebra and calculus to show that certain forms of equations exist.

We are only too well aware of the fact that we have not indeed given any definition of a unified field theory. We have already indicated that this is not easily accomplished, and, fortunately, was not necessary for the given discussion. At best, the present section is a mild protest against the use of the term unified field theory for this particular theory. Regardless of its classification or form, the Einstein–Maxwell theory is still of interest and, due to the nonlinear character of the field equations, explicit solutions are difficult to obtain. Any procedure which generates infinite classes of such solutions, corresponding either to null or non-null fields is of course of interest in that physical implications

of the theory can then be investigated in some detail.

### 3. THE FUNDAMENTAL TETRAD

In the following analysis, it is assumed that we are dealing with a Riemannian space of signature $(-, +, +, +)$. For reference purposes we list the following:

$$R_{ij} = [\log(-g)^{\frac{1}{2}}]_{,ij} - \left\{\begin{matrix}\alpha\\ij\end{matrix}\right\}_{,\alpha} + \left\{\begin{matrix}\alpha\\i\beta\end{matrix}\right\}\left\{\begin{matrix}\beta\\\alpha j\end{matrix}\right\} - \left\{\begin{matrix}\alpha\\ij\end{matrix}\right\}[\log(-g)^{\frac{1}{2}}]_{,\alpha}, \tag{3.1}$$

$$\left\{\begin{matrix}i\\jk\end{matrix}\right\} = \tfrac{1}{2}g^{i\alpha}(g_{k\alpha,j} + g_{\alpha j,k} - g_{jk,\alpha}), \tag{3.2}$$

$$g = \det(g_{ij}). \tag{3.3}$$

It is also convenient to introduce the dual of $F_{ij}$ which is denoted by ${}^*F_{ij}$ and defined by

$${}^*F_{ij} = \tfrac{1}{2}(-g)^{\frac{1}{2}}e_{ijkl}F^{kl}. \tag{3.4}$$

In terms of $F_{ij}$ and ${}^*F_{ij}$ Eqs. (1.6)–(1.8) may be written

$$R_{ij} - \tfrac{1}{2}g_{ij}R = F_{i\alpha}F^{\alpha}_{j} + {}^*F_{i\alpha}{}^*F^{\alpha}_{j}, \tag{3.5}$$

$$F^{ij}{}_{;j} = 0, \qquad {}^*F^{ij}{}_{;j} = 0. \tag{3.6}$$

An electromagnetic field is null if the Lorentz invariants $\mathbf{E}\cdot\mathbf{H}$ and $E^2 - H^2$ are identically zero. The tensor equivalents of these conditions are

$$F_{ij}F^{ij} = 0, \qquad {}^*F_{ij}F^{ij} = 0. \tag{3.7}$$

Hence, Eqs. (3.5)–(3.7) together with the identity $g_{ij;k} = 0$ form the complete set of field equations which determine null fields.

If $P$ is an arbitrary but fixed point in space-time, it is always possible to choose a reference frame such that the expressions

$$g_{ij} = \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}, \tag{3.8}$$

$$F_{ij} = \begin{bmatrix} 0 & 0 & -k & 0\\ 0 & 0 & k & 0\\ k & -k & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}, \tag{3.9}$$

are valid at $P$. The last expression follows from the null conditions (3.7); $k$ is an unknown constant. Furthermore, the coordinate transformation

$$x^1 = k\bar{x}^2 - (\bar{x}^1/2k), \qquad x^2 = k\bar{x}^2 + (\bar{x}^1/2k), \tag{3.10}$$

can be used to produce a reference frame in which

$$g_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}, \qquad F_{ij} = \begin{bmatrix} 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0\\ -1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}. \tag{3.11}$$

Equations (3.11) can be used to define a tetrad of basis vectors. It is clear that $F_{ij}$ has only one eigenvalue, namely zero. Also, it defines at $P$ the eigenvectors

$$w^i = (0, w^2, 0, 0), \tag{3.12}$$

$$h^i = (0, \eta, 0, 1), \tag{3.13}$$

where $w^2$ and $\eta$ are arbitrary constants. The vector $w^i$ is null and is uniquely determined up to a multiplicative constant. The unit vector $h^i$ is determined only up to an additive multiple of $w^i$. If $e_i$ is a unit vector which is orthogonal to $w^i$ and $h^i$, then it has the form

$$e_i = (\xi, 0, 1, 0), \tag{3.14}$$

where $\xi$ is arbitrary. The vector $e_i$ can be used to obtain a particular vector in the direction of $w^i$. This vector is denoted by $w_i$ and defined by

$$w_i \equiv F_{ij}e^j = (1, 0, 0, 0). \tag{3.15}$$

The basis can be completed by the addition of a second null vector. Denoting this vector by $a_i$ and defining it by the relations

$$a^iw_i = 1, \qquad a^ia_i = 0, \qquad a^ie_i = 0, \qquad a^ih_i = 0, \tag{3.16}$$

it follows that its components are given by

$$a_i = \left(\frac{-\xi^2 - \eta^2}{2}, 1, -\xi, -\eta\right). \tag{3.17}$$

It can be shown that $\xi$ and $\eta$ may be transformed away without affecting the form of (3.11), i.e., there is a reference frame such that the covariant components of the basis vectors are given by

$$w_i = (1, 0, 0, 0), \qquad a_i = (0, 1, 0, 0), \qquad e_i = (0, 0, 1, 0), \qquad h_i = (0, 0, 0, 1). \tag{3.18}$$

Hence, from (3.11), (3.4) and (3.5) one can conclude

$$g_{ij} = w_ia_j + w_ja_i + e_ie_j + h_ih_j, \tag{3.19}$$

$$F_{ij} = w_ie_j - w_je_i, \tag{3.20}$$

$${}^*F_{ij} = w_ih_j - w_jh_i, \tag{3.21}$$

$$R_{ij} = 2w_iw_j. \tag{3.22}$$

The field quantities have now been expressed in terms of the basis vectors, which in turn are determined by (3.22), Maxwell's equations:

$$w^i{}_{;i}e_j + w^i e_{j;i} - e^i{}_{;i}w_j - e^i w_{j;i} = 0, \quad (3.23)$$

$$w^i{}_{;i}h_j + w^i h_{j;i} - h^i{}_{;i}w_j - h^i w_{j;i} = 0, \quad (3.24)$$

and the identities

$$g_{ij;k} = 0, \quad (3.25)$$

$$R^i{}_{j;i} = 2(w^i w_j)_{;i} = 0. \quad (3.26)$$

Equations (3.25) and (3.26), respectively, yield

$$(w^i w_i)_{;j} = 2w^i w_{i;j} = 0, \quad (3.27)$$

$$w^i{}_{;i}w_j + w^i w_{j;i} = 0. \quad (3.28)$$

If these expressions are evaluated at the point $P$, then from (3.18) and (3.11) it follows that

$$w_{2;i} = (0, 0, 0, 0), \quad (3.29)$$

$$w_{i;2} = (-w^i{}_{;i}, 0, 0, 0). \quad (3.30)$$

Also, if (3.23) and (3.24) are evaluated at $P$ then, for $j = 3$ and 4, the results are

$$w^i{}_{;i} = w_{3;3} = 0, \quad (3.31)$$

$$e_{4;2} - w_{4;3} = 0, \quad (3.32)$$

$$h_{3;2} - w_{3;4} = 0, \quad (3.33)$$

$$w^i{}_{;i} - w_{4;4} = 0. \quad (3.34)$$

But since $e_i$ is orthogonal to $h_i$ we must have

$$e_{4;2} + h_{3;2} = 0. \quad (3.35)$$

Equations (3.29)–(3.35) may be expressed by the following matrix:

$$w_{i;j} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ 0 & 0 & 0 & 0 \\ \alpha_3^* & 0 & -\alpha_2 & \alpha_5 \\ \alpha_4^* & 0 & -\alpha_5 & -\alpha_2 \end{bmatrix}, \quad (3.36)$$

where the $\alpha$'s are the invariants,

$$\alpha_1 \equiv w_{i;j}a^i a^j, \quad (3.37)$$

$$\alpha_2 \equiv w_{i,j}a^i w^j = -w^i{}_{,i}, \quad (3.38)$$

$$\alpha_3 \equiv w_{i,j}a^i e^j, \quad (3.39)$$

$$\alpha_4 = w_{i,j}a^i h^i, \quad (3.40)$$

$$\alpha_3^* = w_{i,j}e^i a^i, \quad (3.41)$$

$$\alpha_4^* \equiv w_{i;j}h^i a^i, \quad (3.42)$$

$$\alpha_5 \equiv w_{i;j}e^i h^i, \quad (3.43)$$

evaluated at $P$. From (3.36) the following covariant expression can be obtained:

$$w_{i,j} - w_{j,i} = \alpha_2(w_i a_j - w_j a_i) + (\alpha_3 - \alpha_3^*)F_{ij} + (\alpha_4 - \alpha_4^*)^*F_{ij} + \alpha_5(e_i h_j - e_j h_i). \quad (3.44)$$

[Since the Christoffel symbol is a symmetric connection, the semi-colons on the left-hand side of (3.43) have been replaced by commas.] If (3.44) is multiplied by $w_k$ and similar equations are obtained by permuting the $i$, $j$, and $k$, the equations may be added to obtain

$$(w_{i,j} - w_{j,i})w_k + (w_{j,k} - w_{k,j})w_i + (w_{k,i} - w_{i,k})w_j = \alpha_5(e_i h_j w_k + e_j h_k w_i + e_k h_i w_j - h_i e_j w_k - h_j e_k w_i - h_k e_i w_j). \quad (3.45)$$

The above expression shows that $\alpha_5$ plays a dominant role in determining the form of $w_i$. From the theory of ordinary differential equations it is known that the vanishing of $\alpha_5$ is a necessary and sufficient condition for $w_i$ to be a product of a scalar and a gradient. In such cases it has been found, that a reference frame can be introduced in which the metric tensor can be expressed in terms of four unknown functions. Furthermore, the form of the metric is such that the determining field equations are simplified to such an extent, that a general solution is obtainable under rather weak restrictions.

## 4. THE EQUATIONS OF INTEGRABILITY

The basis vectors are subject to the following integrability conditions:

$$w_{i;jk} - w_{i;kj} = w^\alpha R_{\alpha ijk}, \quad (4.1)$$

$$a_{i;jk} - a_{i;kj} = a^\alpha R_{\alpha ijk}, \quad (4.2)$$

$$e_{i;jk} - e_{i;kj} = e^\alpha R_{\alpha ijk}, \quad (4.3)$$

$$h_{i;jk} - h_{i;kj} = h^\alpha R_{\alpha ijk}. \quad (4.4)$$

The above equations can be evaluated at $P$, thereby expressing all of the components of $R_{ijkl}$ in terms of the $\alpha$'s and other fundamental invariants. For example, a major restriction on $\alpha_2$ and $\alpha_3$ can be obtained from the equation

$$R_{22} = R^\alpha_{22\alpha} = 2w_2 w_2 = 0. \quad (4.5)$$

It is easily seen that the standard identities for $R_{ijkl}$ imply that (4.5) is equivalent to

$$R_{2332} + R_{2442} = 0, \quad (4.6)$$

which by (4.1) becomes

$$w_{3;32} - w_{3;23} + w_{4;42} - w_{4;24} = 0. \quad (4.7)$$

A straightforward computation will show that (4.7) reduces to

$$\alpha_{2,2} = 2\alpha_2^2 - \alpha_5^2. \tag{4.8}$$

This result is significant in that it yields a sufficient condition for the vanishing of $\alpha_5$, viz., that the divergence of $w_i$ be identically zero. Hence, it would seem that the problem could best be discussed by a consideration of the two special cases $w^i{}_{;i} = 0$ and $w^i{}_{;i} \neq 0$.

## 5. THE METRIC TENSOR

If it is assumed that $\alpha_5$ vanishes which is certainly the case when $w^i{}_{;i} = 0$, then by (3.45) it is possible to put

$$w_i = e^{\tau} w_{,i}, \tag{5.1}$$

where $\tau$ and $w$ are unknown invariants. It is possible to choose a reference frame such that the contravariant vector $w^i$ has the form

$$w^i = (0, e^{\tau}, 0, 0), \tag{5.2}$$

at all points in space. In this reference frame $w$ must be independent of $x^2$. Also, since $w_i$ is not identically zero, it must be a function of at least one of $x^1$, $x^3$, or $x^4$. There is no loss of generality in assuming that $w$ is a function of $x^1$, and then the coordinate transformations

$$\begin{aligned} \bar{x}^1 &= w(x^1, x^3, x^4), & x^{-2} &= x^2, \\ \bar{x}^3 &= x^3, & \bar{x}^4 &= x^4, \end{aligned} \tag{5.3}$$

leave $w^i$ unchanged, and simplifies $w_i$ to

$$w_i = e^{\tau}(1, 0, 0, 0). \tag{5.4}$$

The vector $e_i$ is related to $w_i$ by the Maxwellian equation

$$F_{ij,k} + F_{jk,i} + F_{ki,j} = 0, \tag{5.5}$$

which may be written explicitly as

$$\begin{aligned} w_i(e_{j,k} - e_{k,j}) + w_j(e_{k,i} - e_{i,k}) \\ + w_k(e_{i,j} - e_{j,i}) = e_i(w_{j,k} - w_{k,j}) \\ + e_j(w_{k,i} - w_{i,k}) + e_k(w_{i,j} - w_{j,i}). \end{aligned} \tag{5.6}$$

If this equation is considered in a reference frame in which (5.4) holds, then putting $i = 1$ we have

$$e^{\tau}(e_{j,k} - e_{k,j}) = e^{\tau}(e_k\tau_{,j} - e_j\tau_{,k}), \tag{5.7}$$

where it is implicit that $j$ or $k$ cannot equal one. Equation (5.7) may be rewritten to read

$$(e^{\tau}e_j)_{,k} - (e^{\tau}e_k)_{,j} = 0, \tag{5.8}$$

from which we can conclude that

$$e_i = e^{-\tau}\mu_{,i} + \xi w_i \tag{5.9}$$

where $\xi$ is an unknown and $\mu$ is some function independent of $x^2$. Since the additive multiple of $w_i$ does not effect either $F_{ij}$ or $g_{ij}$ there is no loss of generality in putting $\xi = 0$.

A similar argument can be applied to $^*F_{ij}$ and a corresponding result for $h_i$ can be obtained, i.e., we may assume that

$$e_i = e^{-\tau}\mu_{,i}, \qquad h_i = e^{-\tau}\nu_{,i}, \tag{5.10}$$

where $\mu$ and $\nu$ are independent of $x^2$.

It is clear that the coordinate transformation

$$\bar{x}^1 = x^1, \qquad \bar{x}^2 = x^2, \qquad \bar{x}^3 = \mu, \qquad \bar{x}^4 = \nu, \tag{5.11}$$

is admissible and that it does not effect the form of $w^i$ and $w_i$. In the new reference frame, $e_i$ and $h_i$ simplify to

$$e_i = (0, 0, e^{-\tau}, 0), \tag{5.12}$$

$$h_i = (0, 0, 0, e^{-\tau}). \tag{5.13}$$

It seems that little information is available concerning the form of $a_i$, but it is known that $w^i a_i = 1$. Hence, the most general expression for $a_i$ must be

$$a_i = e^{-\tau}(\alpha, 1, \beta, \gamma), \tag{5.14}$$

where $\alpha$, $\beta$, and $\gamma$ are unknown functions. Finally, (5.4), (5.12), (5.13), and (5.14) can be combined with (3.19) to yield

$$g_{ij} = \begin{bmatrix} 2\alpha & 1 & \beta & \gamma \\ 1 & 0 & 0 & 0 \\ \beta & 0 & e^{-2\tau} & 0 \\ \gamma & 0 & 0 & e^{-2\tau} \end{bmatrix}. \tag{5.15}$$

For reference purposes we present the following equations which are a consequence of (5.15):

$$g^{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & \sigma & -\beta e^{2\tau} & -\gamma e^{2\tau} \\ 0 & -\beta e^{2\tau} & e^{2\tau} & 0 \\ 0 & -\gamma e^{2\tau} & 0 & e^{2\tau} \end{bmatrix}, \tag{5.16}$$

$$\log(-g)^{\frac{1}{2}} = -2\tau, \tag{5.17}$$

$$\sigma \equiv -2\alpha + e^{2\tau}(\beta^2 + \gamma^2). \tag{5.18}$$

## 6. THE FIELD EQUATIONS

The assumption

$$w^i{}_{;i} = 0, \tag{6.1}$$

implies the existence of a reference frame such that the metric tensor is given by (5.15). Furthermore,

condition (6.1) gives the extra restriction that $\tau$ is independent of $x^2$. This being the case, the nonzero Christoffel symbols are

$$\begin{Bmatrix}1\\11\end{Bmatrix} = -\alpha_{,2}, \quad \begin{Bmatrix}1\\13\end{Bmatrix} = -\tfrac{1}{2}\beta_{,2}, \quad \begin{Bmatrix}1\\14\end{Bmatrix} = -\tfrac{1}{2}\gamma_{,2},$$

$$\begin{Bmatrix}2\\11\end{Bmatrix} = \alpha_{,1} - \sigma\alpha_{,2} - \beta e^{2\tau}(\beta_{,1} - \alpha_{,3}) - \gamma e^{2\tau}(\gamma_{,1} - \alpha_{,4}),$$

$$\begin{Bmatrix}2\\12\end{Bmatrix} = \alpha_{,2} - \tfrac{1}{2}e^{2\tau}(\beta\beta_{,2} + \gamma\gamma_{,2}),$$

$$\begin{Bmatrix}2\\13\end{Bmatrix} = \alpha_{,3} - \tfrac{1}{2}\sigma\beta_{,2} + \beta\tau_{,1} - \tfrac{1}{2}\gamma e^{2\tau}(\gamma_{,3} - \beta_{,4}),$$

$$\begin{Bmatrix}2\\14\end{Bmatrix} = \alpha_{,4} - \tfrac{1}{2}\sigma\gamma_{,2} + \gamma\tau_{,1} - \tfrac{1}{2}\beta e^{2\tau}(\beta_{,4} - \gamma_{,3}),$$

$$\begin{Bmatrix}2\\23\end{Bmatrix} = \tfrac{1}{2}\beta_{,2}, \quad \begin{Bmatrix}2\\24\end{Bmatrix} = \tfrac{1}{2}\gamma_{,2},$$

$$\begin{Bmatrix}2\\33\end{Bmatrix} = \beta_{,3} + e^{-2\tau}\tau_{,1} + \beta\tau_{,3} - \gamma\tau_{,4}, \tag{6.2}$$

$$\begin{Bmatrix}2\\34\end{Bmatrix} = \tfrac{1}{2}(\gamma_{,3} + \beta_{,4}) + \beta\tau_{,4} + \gamma\tau_{,3},$$

$$\begin{Bmatrix}2\\44\end{Bmatrix} = \gamma_{,4} + e^{-2\tau}\tau_{,1} - \beta\tau_{,3} + \gamma\tau_{,4},$$

$$\begin{Bmatrix}3\\11\end{Bmatrix} = e^{2\tau}(\beta\alpha_{,2} + \beta_{,1} - \alpha_{,3}),$$

$$\begin{Bmatrix}3\\12\end{Bmatrix} = \tfrac{1}{2}e^{2\tau}\beta_{,2}, \quad \begin{Bmatrix}3\\13\end{Bmatrix} = \tfrac{1}{2}e^{2\tau}\beta\beta_{,2} - \tau_{,1},$$

$$\begin{Bmatrix}3\\14\end{Bmatrix} = \tfrac{1}{2}e^{2\tau}(\beta\gamma_{,2} + \beta_{,4} - \gamma_{,3}),$$

$$\begin{Bmatrix}3\\33\end{Bmatrix} = -\tau_{,3}, \quad \begin{Bmatrix}3\\34\end{Bmatrix} = -\tau_{,4}, \quad \begin{Bmatrix}3\\44\end{Bmatrix} = \tau_{,3};$$

$$\begin{Bmatrix}4\\11\end{Bmatrix} = e^{2\tau}(\gamma\alpha_{,2} + \gamma_{,1} - \alpha_{,4}),$$

$$\begin{Bmatrix}4\\12\end{Bmatrix} = \tfrac{1}{2}e^{2\tau}\gamma_{,2}, \quad \begin{Bmatrix}4\\14\end{Bmatrix} = \tfrac{1}{2}e^{2\tau}\gamma\gamma_{,2} - \tau_{,1},$$

$$\begin{Bmatrix}4\\13\end{Bmatrix} = \tfrac{1}{2}e^{2\tau}(\tau\beta_{,2} + \gamma_{,3} - \beta_{,4}),$$

$$\begin{Bmatrix}4\\33\end{Bmatrix} = \tau_{,4}, \quad \begin{Bmatrix}4\\34\end{Bmatrix} = -\tau_{,3}, \quad \begin{Bmatrix}4\\44\end{Bmatrix} = -\tau_{,4},$$

From (3.1), (3.22), and (5.4), the field equations are

$$(-2\tau)_{,ij} - \begin{Bmatrix}\alpha\\ij\end{Bmatrix}_{,\alpha} + \begin{Bmatrix}\alpha\\i\beta\end{Bmatrix}\begin{Bmatrix}\beta\\\alpha j\end{Bmatrix} + 2\begin{Bmatrix}\alpha\\ij\end{Bmatrix}\tau_{,\alpha} = 2e^{2\tau}\delta^1_i\delta^1_j, \tag{6.3}$$

and in turn,

$$R_{12} = -\alpha_{,22} + \tfrac{1}{2}e^{2\tau}[(\beta_{,2})^2 + (\gamma_{,2})^2 - \beta_{,23} - \gamma_{,24}] = 0, \tag{6.4}$$

$$R_{22} \equiv 0, \tag{6.5}$$

$$R_{23} = -\tfrac{1}{2}\beta_{,22} = 0, \tag{6.6}$$

$$R_{24} = -\tfrac{1}{2}\gamma_{,22} = 0, \tag{6.7}$$

$$R_{33} = -\nabla^2\tau - (\beta_{,3} + \beta\tau_{,3} - \gamma\tau_{,4})_{,2} + \tfrac{1}{2}(\beta_{,2})^2 = 0, \tag{6.8}$$

$$R_{34} = -[\tfrac{1}{2}(\gamma_{,3} + \beta_{,4}) + \beta\tau_{,4} + \gamma\tau_{,3}]_{,2} + \tfrac{1}{2}\beta_{,2}\gamma_{,2} = 0, \tag{6.9}$$

$$R_{44} = -\nabla^2\tau - (\gamma_{,4} - \beta\tau_{,3} + \gamma\tau_{,4})_{,2} + \tfrac{1}{2}(\gamma_{,2})^2 = 0, \tag{6.10}$$

where

$$\nabla^2\tau \equiv \tau_{,33} + \tau_{,44}. \tag{6.11}$$

It is apparent from (6.6) and (6.7) that

$$\beta = \xi x^2 + \xi_0, \tag{6.12}$$

$$\gamma = \eta x^2 + \eta_0, \tag{6.13}$$

where the $\xi$'s and $\eta$'s are functions of $x^1$, $x^3$, and $x^4$. When $\xi^2 + \eta^2 \neq 0$, a straightforward calculation shows that $\xi$, $\eta$ are given by

$$\xi = \frac{-2\varphi_{,3}\nabla^2\varphi}{\varphi_{,3}^2 + \varphi_{,4}^2}, \qquad \eta = \frac{-2\varphi_{,4}\nabla^2\varphi}{\varphi_{,3}^2 + \varphi_{,4}^2}, \tag{6.14}$$

where $\varphi(x^1, x^3, x^4)$ is a solution of

$$\nabla^2\log\,[\nabla^2\varphi(\varphi_{,3}^2 + \varphi_{,4}^2)] + \frac{2(\nabla^2\varphi)^2}{\varphi_{,3}^2 + \varphi_{,4}^2} = 0. \tag{6.15}$$

It can be verified that this equation is the only restriction that must be placed on $\varphi$, i.e., the other field equations which involve $\xi$ and $\eta$ are satisfied identically if $\varphi$ is a solution of (6.15). The most general solution for $\varphi$ involves four arbitrary functions. It is easy to show that if $\varphi(x^3, x^4)$ is one solution of (4.15) then $\varphi(\mu, \nu)$, where $\mu = \mu(x^3 + ix^4)$ and $\nu = \nu(x^3 - ix^4)$ are arbitrary functions, is also a solution. However, it seems unlikely that the general solution can be presented in any reasonably simple form. For example, if it is assumed that $\varphi = \varphi(x^3)$ then (6.15) becomes

$$\{\log\,[\varphi_{,33}(\varphi_{,3})^2]\}_{,33} + \frac{2(\varphi_{,33})^2}{\varphi_{,3}^2} = 0. \tag{6.16}$$

It is possible to integrate (6.16) and express $x^3$ as an explicit function of $\log\,\varphi_{,3}$. Unfortunately, the expression is too complicated to be inverted in order

to obtain an explicit expression for $\varphi_{,3}$. It is easily seen that two particular solutions of (6.16) are

$$\varphi_{,3} = x^3, \qquad \varphi_{,3} = (x^3)^{\frac{1}{2}}. \tag{6.17}$$

It is shown later that these solutions actually correspond to rather general situations.

Up to this point no mention has been made of the field equations

$$R_{11} = 2e^{2\tau}, \qquad R_{13} = R_{14} = 0. \tag{6.18}$$

These equations turn out to be extremely complicated and it seems unlikely that a general consideration of these equations will lead to anything useful. Hence, in the following sections we, by placing rather weak restrictions on $\alpha$, $\beta$, and $\gamma$, obtain various rigorous classes of solutions of the field equations.

## 7. CASE I. $\xi^2 + \eta^2 = 0$

For this case the field equations are

$$R_{44} = R_{33} = -\nabla^2\tau = 0, \tag{7.1}$$

$$R_{22} = R_{23} = R_{24} = R_{34} \equiv 0, \tag{7.2}$$

$$R_{12} = -\alpha_{,22} = 0, \tag{7.3}$$

$$R_{13} = -\alpha_{,32} - \tfrac{1}{2}(\gamma_{,3} - \beta_{,4})_{,4} = 0, \tag{7.4}$$

$$R_{14} = -\alpha_{,42} - \tfrac{1}{2}(\beta_{,4} - \gamma_{,3})_{,3} = 0, \tag{7.5}$$

$$\begin{aligned} R_{11} = {} & \nabla^2\alpha - (2\alpha - \beta^2 - \gamma^2)\alpha_{,22} \\ & -\alpha_{,2}(\beta_{,3} + \gamma_{,4}) - \tfrac{1}{2}(\beta_{,4} - \gamma_{,3})^2 \\ & -2\beta\alpha_{,23} - 2\gamma\alpha_{,24} - (\beta_{,3} + \gamma_{,4})_{,1} \\ & -2e^{2\tau}, \end{aligned} \tag{7.6}$$

where

$$\nabla^2 \equiv \frac{\partial^2}{(\partial x^3)^2} + \frac{\partial^2}{(\partial x^4)^2}. \tag{7.7}$$

The general solution to (7.3)–(7.5) is

$$\alpha_{,2} = -\bar{\gamma}_{,4} = -\bar{\beta}_{,3}, \tag{7.8}$$

$$\tfrac{1}{2}(\gamma_{,3} - \beta_{,4}) = \bar{\gamma}_{,3} = -\bar{\beta}_{,4}, \tag{7.9}$$

where $\bar{\beta}$ is an arbitrary harmonic function and $\bar{\gamma}$ is its harmonic conjugate. Since

$$(\gamma - \bar{\gamma})_{,3} = (\beta - \bar{\beta})_{,4}, \tag{7.10}$$

there exists a function $\theta(x^1, x^3, x^4)$ such that

$$\beta - \bar{\beta} = \theta_{,3}, \qquad \gamma - \bar{\gamma} = \theta_{,4}. \tag{7.11}$$

The coordinate transformation

$$\bar{x}^1 = x^1, \qquad \bar{x}^2 = x^2 + \theta, \qquad \bar{x}^3 = x^3, \qquad \bar{x}^4 = x^4, \tag{7.12}$$

puts the metric tensor into the form

$$\bar{g}_{ij} = \begin{bmatrix} 2\bar{\alpha} & 1 & \bar{\beta} & \bar{\gamma} \\ 1 & 0 & 0 & 0 \\ \bar{\beta} & 0 & e^{-2\tau} & 0 \\ \bar{\gamma} & 0 & 0 & e^{-2\tau} \end{bmatrix}. \tag{7.13}$$

If the bars are now omitted, then field equations (7.4) and (7.5) can be replaced by

$$\begin{aligned} \beta_{,3} &= \gamma_{,4} = -\alpha_{,2}, \\ \beta_{,4} &= -\gamma_{,3}. \end{aligned} \tag{7.14}$$

By virtue of (5.14), Eq. (5.6) may be written

$$\nabla^2\alpha = 2(\gamma\gamma_{,3} - \beta\beta_{,3} + \beta_{,1})_{,3} + 2e^{\tau}, \tag{7.15}$$

or, alternatively,

$$\nabla^2\alpha = 2(\beta\beta_{,4} - \gamma\gamma_{,4} + \gamma_{,1})_{,4} + 2e^{\tau}. \tag{7.16}$$

Placing

$$U = \gamma\gamma_{,3} - \beta\beta_{,3} + \beta_{,1}, \tag{7.17}$$

and

$$V = \beta\beta_{,4} - \gamma\gamma_{,4} + \gamma_{,1}, \tag{7.18}$$

we may write

$$\nabla^2\alpha = U_{,3} + V_{,4} + 2e^{2\tau}. \tag{7.19}$$

It is easily verified that $U$ and $V$ are harmonic conjugates and therefore satisfy the relation

$$\tfrac{1}{2}\nabla^2(x^3U + x^4V) = U_{,3} + V_{,4}. \tag{7.20}$$

Hence the general solution to (7.19) is

$$\alpha = \tfrac{1}{2}(x^3U + x^4V) - x^2\beta_{,3} + \sigma \tag{7.21}$$

where $\sigma = \sigma(x^1, x^3, x^4)$ is the general solution of

$$\nabla^2\sigma = 2e^{2\tau}. \tag{7.22}$$

The above results may be summarized as follows:

(i) There is a reference frame such that the metric tensor takes the form

$$g_{ij} = \begin{bmatrix} 2\alpha & 1 & \beta & \gamma \\ 1 & 0 & 0 & 0 \\ \beta & 0 & e^{-\tau} & 0 \\ \gamma & 0 & 0 & e^{-\tau} \end{bmatrix}. \tag{7.23}$$

(ii) $\beta(x^1, x^3, x^4)$ and $\tau(x^1, x^3, x^4)$ are arbitrary harmonic functions in the variables $x^3$ and $x^4$, $\gamma(x^1, x^3, x^4)$ is the harmonic conjugate of $\beta$.

(iii) Once $\beta$, $\gamma$ and $\tau$ are prescribed, $\alpha$ is given by

$$\alpha = \tfrac{1}{2}[x^3(\gamma\gamma_{,3} - \beta\beta_{,3} + \beta_{,1}) + x^4(\beta\beta_{,4} - \gamma\gamma_{,4} + \gamma_{,1})]$$
$$- x^2\beta_{,3} + \tfrac{1}{2}\iint e^{2\tau}\, dz\, d\bar{z} + \alpha^*, \qquad (7.24)$$

where $z = x^3 + ix^4$ and $\alpha^*(x^1, x^3, x^4)$ is another arbitrary harmonic function.

(iv) From (3.20), (5.4) and (5.12), the electromagnetic tensor is given by

$$F_{ij} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \qquad (7.25)$$

(v) Since the metric tensor involves three arbitrary functions while the electromagnetic tensor remains fixed, it would appear that a complete knowledge of $F_{ij}$ does not uniquely determine the metric of the space. Furthermore, it may be noted that the field equations place no restrictions on the variable $x^1$.

### 8. CASE II. $\xi^2 + \eta^2 \neq 0$, $\alpha_{,22} = 0$

From (6.4), (6.6), and (6.7) we infer that

$$\alpha = \tfrac{1}{2}\alpha_{,22}(x^2)^2 + \alpha_2 x^2 + \alpha_0, \qquad (8.1)$$

where the $\alpha$'s on the right hand side are independent of $x^2$. Equation (8.1) together with (6.12) and (6.13) can be used to obtain the following decompositions:

$$R_{11} = (x^2)^2 R_{11}^{(2)} + x^2 R_{11}^{(1)} + R_{11}^{(0)}, \qquad (8.2)$$
$$R_{13} = x^2 R_{13}^{(1)} + R_{13}^{(0)}, \qquad (8.3)$$
$$R_{14} = x^2 R_{14}^{(1)} + R_{14}^{(0)}, \qquad (8.4)$$

the $R$'s on the right being independent of $x^2$. If $\alpha_{,22} = 0$, then the equation $R_{11}^{(2)} = 0$ reduces to

$$\xi_{,4} - \eta_{,3} = 0. \qquad (8.5)$$

Hence, there exists a $u(x^1, x^3, x^4)$ such that

$$\xi = u_{,3}, \qquad \eta = u_{,4}. \qquad (8.6)$$

The function $u$ is determined by (6.4) which reads

$$\nabla^2 u = (u_{,3})^2 + (u_{,4})^2, \qquad (8.7)$$

implying that $e^{-u}$ is a harmonic function. The transformation

$$\bar{x}^3 = e^{-u}, \quad \bar{x}^4 = \text{harmonic conjugate of } e^{-u}, \qquad (8.8)$$

preserves the form of $g_{ij}$, but in the new reference frame we will have

$$\xi = -1/x^3, \qquad \eta = 0. \qquad (8.9)$$

It is now possible to determine $\tau$ [Eqs. (6.8)–(6.10)] and the result is

$$\tau = \tfrac{1}{4}\log x^3 + \tau_0(x^1). \qquad (8.10)$$

It can be shown that there is no loss of generality in putting $\tau_0$ and $\eta_0$ equal to zero. Thus far we have shown

$$\alpha = \alpha_{,2}x^2 + \alpha_0, \qquad \beta = (-x^2/x^3) + \xi_0, \qquad (8.11)$$
$$\tau = \tfrac{1}{4}\log x^3, \qquad \gamma = 0.$$

The equations $R_{13}^{(0)} = R_{14}^{(0)} = 0$ yield

$$\alpha_{,23} = \tfrac{1}{2}(x^3)^{\frac{1}{2}}\xi_{0,44}, \qquad (8.12)$$

$$\alpha_{,24} = -\tfrac{1}{2}(x^3)^{\frac{1}{2}}\left(\frac{3}{2x^3}\xi_{0,4} + \xi_{0,43}\right), \qquad (8.13)$$

and by direct integration

$$\alpha_{,2} = -\tfrac{1}{2}(x^3)^{\frac{1}{2}}\left(\frac{3}{2x^3}\xi_0 + \xi_{0,3}\right), \qquad (8.14)$$

an arbitrary function having been removed by a transformation of the form

$$x^2 = \bar{x}^2 + f(x^1, x^3). \qquad (8.15)$$

If (8.14) is substituted into (8.12), the result is

$$\nabla^2 v - v_{,3}/x^3 = 0, \qquad (8.16)$$

where

$$v \equiv \xi_0(x^3)^{\frac{1}{2}}. \qquad (8.17)$$

The general solution of (8.16) is

$$v = x^3 \int_0^\pi [F(x^3\cos\varphi + ix^4)$$
$$+ G(x^3\cos\varphi - ix^4)]\cos\varphi\, d\varphi \qquad (8.18)$$

where $F$ and $G$ are arbitrary functions of their arguments. Hence,

$$\xi_0 = \frac{1}{(x^3)^{\frac{1}{2}}}\int_0^\pi (F + G)\cos\varphi\, d\varphi, \qquad (8.19)$$

$$\alpha_{,2} = -\frac{1}{2}\int_0^\pi (F' + G')\, d\varphi, \qquad (8.20)$$

where, in the usual way, the primes denote total differentiation.

The only remaining field equation which does not reduce to an identity is $R_{11}^{(0)} = 2e^{2\tau}$. A tedious computation shows that this is equivalent to

$$\nabla^2\alpha_0 - \frac{\alpha_{0,3}}{x^3} + \frac{\alpha_0}{(x^3)^2} = 2 + \xi_{0,3}\alpha_{,2}$$
$$+ 2\xi_0\alpha_{,23} + \xi_{0,13} + \frac{(x^3)^{\frac{1}{2}}}{2}(\xi_{0,4})^2. \qquad (8.21)$$

Since $\alpha_{,2}$ and $\xi_0$ are known (see (8.19) and (8.20)) the general solution of this equation is

$$\alpha_0 = x^3 \int_0^\tau [F^*(x^3 \cos\varphi + ix^4) + G^*(x^3 \cos\varphi - ix^4)]\, d\varphi + \alpha_0^* \qquad (8.22)$$

where $F^*$ and $G^*$ are arbitrary. $\alpha_0^*$ is any particular integral of (8.21).

The metric is now determined in terms of arbitrary functions. However, many of the functions which were transformed away, such as $\eta_0$, $\tau_0$, etc., will reappear in the expression for the electromagnetic tensor.

### 9. CASE III. $\xi^2 + n^2 \neq 0$, $\nabla^2\tau = 0$

In the discussion of Case II, it was assumed that $\alpha_{,22}$ was zero. The fact that this is not necessarily the case can be shown by making the ad hoc assumption that $\tau$ is harmonic. For if this is the case, then $\tau$ can be removed from the metric by a coordinate transformation. Hence, we may assume that there is a reference frame such that

$$g_{ij} = \begin{bmatrix} 2\alpha & 1 & \beta & \gamma \\ 1 & 0 & 0 & 0 \\ \beta & 0 & 1 & 0 \\ \gamma & 0 & 0 & 1 \end{bmatrix}, \qquad (9.1)$$

$$w_i = (e^\tau, 0, 0, 0). \qquad (9.2)$$

As before, the dependence of $\alpha$, $\beta$ and $\gamma$ on $x^2$ is as follows:

$$\alpha = \tfrac{1}{2}\alpha_{,22}(x^2)^2 + \alpha_2 x^2 + \alpha_0, \qquad \beta = x^2\xi + \xi_0, \qquad \gamma = x^2\eta + \eta_0. \qquad (9.3)$$

Field equations (6.8)–(6.10) reduce to

$$\xi_{,3} = \tfrac{1}{2}\xi^2, \qquad \eta_{,4} = \tfrac{1}{2}\eta^2, \qquad \xi_{,4} + \eta_{,3} = \xi\eta. \qquad (9.4)$$

Assuming $\xi\eta \neq 0$ [if such is not the case then the final result, i.e., Eq. (9.13) follows immediately], Eqs. (9.4) imply

$$\xi = \frac{2}{\varphi(x^4) - x^3}, \qquad \eta = \frac{2}{\psi(x^3) - x^4}, \qquad (9.5)$$

where $\varphi$ and $\psi$ are solutions of

$$\psi_{,3}(\varphi - x^3)^2 + \varphi_{,4}(\psi - x^4)^2 + 2(\psi - x^4)(\varphi - x^3) = 0. \qquad (9.6)$$

For an arbitrary $x^3$, let $x^4 = \psi(x^3)$. Equation (9.6) implies

$$\psi_{,3}(\varphi - x^3)^2 = 0. \qquad (9.7)$$

Since $\xi\eta \neq 0$ then we must have $\psi_{,3}\varphi_{,4} \neq 0$. Hence

$$\psi(\varphi(x^4)) = x^4, \qquad (9.8)$$

a result which can be used to show that the general solution to (9.6) is

$$\psi = ax^3 + b, \qquad \varphi = \frac{1}{a}(x^4 - b), \qquad (9.9)$$

where $a$ and $b$ can be arbitrary functions of $x^1$. There is no loss of generality if we put $b = 0$, in which case we have

$$\xi = \frac{2a}{x^4 - ax^3}, \qquad \eta = \frac{2}{ax^3 - x^4}. \qquad (9.10)$$

Since

$$\xi\, dx^3 + \eta\, dx^4 = -2\left(\frac{a dx^3 - dx^4}{ax^3 - x^4}\right), \qquad (9.11)$$

a rotation of the form

$$\bar{x}^3 = \frac{ax^3 - x^4}{(a^2+1)^{\frac{1}{2}}}, \qquad \bar{x}^4 = \frac{x^3 + ax^4}{(a^2+1)^{\frac{1}{2}}}, \qquad (9.12)$$

can be used to obtain a reference frame in which

$$\xi = -2/x^3, \qquad \eta = 0. \qquad (9.13)$$

As in the previous case, it is possible to justify taking $\eta_0 = 0$.

The net results of the above transformations are

$$\alpha_{,22} = 1/(x^3)^2, \qquad \beta = (-2x^2/x^3) + \xi_0, \qquad \gamma = 0. \qquad (9.14)$$

[The expression for $\alpha_{,22}$ follows from (6.4)].

The functions $\xi_0$ and $\alpha_2$ can be obtained from the equations $R_{13} = R_{14} = 0$. The results are

$$\alpha_2 = -\tfrac{1}{2}(2\xi_0/x^3 + \xi_{0,3}), \qquad (9.15)$$

$$\nabla^2\xi_0 + 2\xi_{0,3}/x^3 = 0. \qquad (9.16)$$

The solution to (9.16) is

$$\xi_0 = H(x^1, x^3, x^4)/x^3, \qquad (9.17)$$

where $H$ is an arbitrary harmonic function. Once $H$ is specified $\alpha_2$ is given by

$$\alpha_2 = \frac{-1}{2(x^3)^2}(x^3 H)_{,3}. \qquad (9.18)$$

The final equation to be satisfied is $R_{11} = 2e^{2\tau}$. This equation determines $\alpha_0$, and the resulting differential equation is

$$\nabla^2\alpha_0 - \frac{2\alpha_{0,3}}{x^3} + \frac{2\alpha_0}{(x^3)^2} = 2e^{2\tau} + \alpha_2\xi_{0,3} + \tfrac{1}{2}(\xi_{0,4})^2 + \xi_{0,13}, \qquad (9.19)$$

$\tau$ being an arbitrary harmonic function. If we put

$$u = \frac{\alpha_0}{x^3}, \tag{9.20}$$

equation (9.19) may be put in the form

$$\nabla^2 u = \frac{1}{x^3} F(x^1, x^3, x^4), \tag{9.21}$$

where $F$ represents the right-hand side of (9.19). The solution to (9.21) is

$$u = \iint F\left(x^1, \frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2i}\right) dz\, d\bar{z} + \psi(z) + \varphi(\bar{z}), \tag{9.22}$$

where

$$z = x^3 + ix^4, \qquad \bar{z} = x^3 - ix^4, \tag{9.23}$$

and the functions $\varphi$, $\psi$ are arbitrary.

This completes the analysis of Case II as $\alpha$, $\beta$ are now explicitly expressed in terms of arbitrary functions while $\gamma$ is identically zero. In order to complete the discussion relative to $w^i{}_{;i} = 0$, it would be necessary to consider the final case

$$\nabla^2 \tau \neq 0, \qquad \alpha_{,22} \neq 0. \tag{9.24}$$

For this case, we have no suggestion for a possible method of attack which might lead to a general solution.

## 10. THE CASE $w^i{}_{;i} \neq 0$

The consideration of this case is far from complete, for if $w^i{}_{;i} \neq 0$ then the equations of integrability become extremely unwieldy. However, we shall, by a particular example, show that mathematical solutions of the field equations do exist for which $w^i{}_{;i} \neq 0$.

If we assume that $w_{i;j}$, given in general by (3.36), has $\alpha_3 = \alpha_4 = \alpha_3^* = \alpha_4^* = 0$, then

$$w_{i,j} = \begin{bmatrix} \alpha_1 & \alpha_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\alpha_2 & \alpha_5 \\ 0 & 0 & -\alpha_5 & -\alpha_2 \end{bmatrix}. \tag{10.1}$$

It is possible to show that $\alpha_5$ is zero. The calculations leading to this result are rather tedious and will not be given. With $\alpha_5 = 0$ the analysis of Sec. 5 is applicable and the metric tensor can be put in the form given in (5.15). It can be shown that the field equations are satisfied identically if

$$g_{ij} = \begin{bmatrix} 2\alpha & 1 & \beta & \gamma \\ 1 & 0 & 0 & 0 \\ \beta & 0 & (x^2)^2 & 0 \\ \gamma & 0 & 0 & (x^2)^2 \end{bmatrix}, \tag{10.2}$$

$$F_{ij} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{10.3}$$

where

$$\beta = (x^2)^2(\bar{\alpha}x^3 + \bar{\gamma}x^4), \tag{10.4}$$

$$\gamma = (x^2)^2(\bar{\alpha}x^4 - \bar{\gamma}x^3), \tag{10.5}$$

$$\alpha = -\bar{\alpha}x^2 - \frac{\bar{\beta}}{x^2} + \frac{\beta^2 + \gamma^2}{(x^2)^2}, \tag{10.6}$$

$$\bar{\beta} = e^{3A} \int e^{-3A}\, dx^1, \tag{10.7}$$

$$A = \int \bar{\alpha}\, dx^1, \tag{10.8}$$

$\bar{\alpha}$ and $\bar{\gamma}$ being arbitrary functions of $x^1$. We do not imply, however, that Eqs. (10.2)–(10.8) are the general solution when $\alpha_5 = 0$.

## 11. SUMMARY OF RESULTS

It has been shown that the form of the propagation null vector $w_i$ depends on the invariant $\alpha_5$. The vanishing of $\alpha_5$, which is equivalent to

$$R_{ij;k} F^{i\alpha} {}^*F^{k\beta} = 0, \tag{11.1}$$

implies that $w_i$ takes the form $e^{\tau} w_{,i}$. It then follows that a reference frame exists such that $g_{ij}$ is given by

$$g_{ij} = \begin{bmatrix} 2\alpha & 1 & \beta & \gamma \\ 1 & 0 & 0 & 0 \\ \beta & 0 & e^{-2\tau} & 0 \\ \gamma & 0 & 0 & e^{-2\tau} \end{bmatrix}. \tag{11.2}$$

The general solution for $\alpha$, $\beta$, $\gamma$ and $\tau$ has not been obtained. However, classes of solutions are available for the following special cases:

**(A)** $\quad w^i{}_{;i} = 0, \qquad (\beta_{,2})^2 + (\gamma_{,3})^2 = 0$

Solutions of this type have been discussed by various authors. Peres,[4] and later Takeno,[6] considered cases where the metric was equivalent to (11.2) provided that $\beta = \gamma = \tau = 0$. Also, Pandya and Vaidya[7] presented a particular solution which is similar to the solution given here.

The general solution for this case is:

---

[6] H. Takeno, Tensor 11, 99 (1961).

[7] I. M. Pandya and P. C. Vaidya, Proc. Natl. Inst. Sci. India **A27**, 620, (1961).

(i) $\beta(x^1, x^3, x^4)$ and $\tau(x^1, x^3, x^4)$ are arbitrary functions, harmonic in the variables $x^3$ and $x^4$.
(ii) $\gamma(x^1, x^3, x^4)$ is the harmonic conjugate of $\beta$.
(iii) Once $\beta$, $\gamma$, and $\tau$ are specified,

$$\alpha \equiv \tfrac{1}{2}[x^3(\gamma\gamma_{,3} - \beta\beta_{,3} + \beta_{,1}) + x^4(\beta\beta_{,4} - \gamma\gamma_{,4} + \gamma_{,1})] - x^2\beta_{,3} + \frac{1}{2}\iint e^{2\tau}\, dz\, d\bar{z} + \alpha^*, \tag{11.3}$$

where $z = x^3 + ix^4$, and $\alpha^*(x^1, x^3, x^4)$ is an arbitrary harmonic function.

(iv) The electromagnetic field is given by

$$F_{ij} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{11.4}$$

One of the simpler solutions of this type is obtained by putting

$$\tau = \beta = \gamma = \alpha^* = 0, \qquad \alpha = \tfrac{1}{2}[(x^3)^2 + (x^4)^2]. \tag{11.5}$$

The coordinate transformation

$$t = (x^2 - x^1)/\sqrt{2}, \quad x = (x^2 + x^1)/\sqrt{2}, \quad y = x^3, \quad z = x^4, \tag{11.6}$$

puts the line-element $ds^2$ into the form

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 + \tfrac{1}{2}(y^2 + z^2)(dx - dt)^2, \tag{11.7}$$

while $F_{ij}$ becomes

$$F_{ij} = \begin{bmatrix} 0 & 0 & -1/\sqrt{2} & 0 \\ 0 & 0 & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{11.8}$$

Hence this special case is of the form discussed by Peres.

**(B)** $w^i{}_{;i} = 0, \quad (\beta_{,2})^2 + (\gamma_{,2})^2 \neq 0, \quad \alpha_{,22} = 0$

For this case it is shown that a coordinate system exists for which

$$\alpha = \alpha_{,2}x^2 + \alpha_0, \quad \beta = (-x^2/x^3) + \xi_0, \quad \tau = \tfrac{1}{4}\log x^3, \quad \gamma = 0. \tag{11.9}$$

The functions $\xi_0$ and $\alpha_{,2}$ can be expressed in terms of arbitrary functions $F(x)$ and $G(x)$ in the following manner:

$$\xi_0 = \frac{1}{(x^3)^{\frac{1}{2}}}\int_0^{\pi} \{F(x^3\cos\varphi + ix^4) + G(x^3\cos\varphi - ix^4)\}\cos\varphi\, d\varphi, \tag{11.10}$$

$$\alpha_{,2} = -\frac{1}{2}\int_0^{\pi} \{F'(x^3\cos\varphi + ix^4) + G'(x^3\cos\varphi - ix^4)\}\, d\varphi. \tag{11.11}$$

Once $F$ and $G$ are specified, a particular integral $\alpha_0^*$ must be computed for equation (8.21). The general solution for this equation is then

$$\alpha_0 = x^3\int_0^{\pi} \{F^*(x^3\cos\varphi + ix^4) + G^*(x^3\cos\varphi - ix^4)\}\, d\varphi + \alpha_0^*, \tag{11.12}$$

where $F^*$ and $G^*$ are arbitrary functions of a single argument.

It seems unlikely that one can obtain an explicit expression for $F_{ij}$ as $e_i$ is obtained by inverting the arbitrary harmonic function $e^{-u}$ [see (8.8)]. However it is not difficult to generate particular solutions. For example if we put

$$u = -\log x^3, \tag{11.13}$$

$$F(\sigma) = G(\sigma) = a\sigma + b, \tag{11.14}$$

$$F^*(\sigma) = G^*(\sigma) = a^*\sigma + b^*, \tag{11.15}$$

where $a$, $a^*$, $b$ and $b^*$ are arbitrary functions of $x^1$, the results are

$$g_{ij} = \begin{bmatrix} 2\alpha_0 - 2\pi a x^2 & 1 & \dfrac{-x^2}{x^3} + \pi a(x^3)^{\frac{1}{2}} & 0 \\ 1 & 0 & 0 & 0 \\ \dfrac{-x^2}{x^3} + \pi a(x^3)^{\frac{1}{2}} & 0 & \dfrac{1}{(x^3)^{\frac{1}{2}}} & 0 \\ 0 & 0 & 0 & \dfrac{1}{(x^3)^{\frac{1}{2}}} \end{bmatrix}, \tag{11.16}$$

where

$$\alpha_0 = 2(x^3)^2 + \pi b^* x^3 + 2\pi(a_{,1} - \pi a^2)(x^3)^{\frac{3}{2}}. \tag{11.17}$$

$F_{ij}$ is given by (11.4).

**(C)** $w^i{}_{;i} = 0, \quad (\beta_{,2})^2 + (\gamma_{,2})^2 = 0, \quad \nabla^2\tau = 0$

Although $\tau$ can be removed from the metric by a coordinate transformation, it remains as an arbitrary function in $w_i = e^{\tau}(x^1)_{,i}$. Hence, it will probably appear in the electromagnetic tensor. The

most general metric satisfying the above conditions is

$$g_{ij} = \begin{bmatrix} 2\alpha & 1 & \beta & 0 \\ 1 & 0 & 0 & 0 \\ \beta & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{11.18}$$

$$\alpha = \frac{(x^2)^2}{2(x^3)^2} - \frac{x^2(x^3H)_{,3}}{2(x^3)^2} + x^3(u^* + u), \tag{11.19}$$

$$\beta = \frac{-2x^2}{x^3} + \frac{H}{x^3}. \tag{11.20}$$

$H(x^1, x^3, x^4)$ and $u(x^1, x^3, x^4)$ are arbitrary functions, harmonic in the variables $x^3$ and $x^4$. $u^*$ is a particular integral of (9.21).

A particular solution for this case can be obtained if we put $\tau = 0$ and

$$H = ax^3x^4, \tag{11.21}$$

where $a(x^1)$ is arbitrary. A straightforward computation yields

$$\alpha = \frac{(x^2)^2}{2(x^3)^2} - a\frac{x^2x^4}{x^3} + a^2(x^4)^2 + (2 - 3/2a^2)(x^3)^2(\log x^3 - 1) + x^3u, \tag{11.22}$$

$$\beta = \frac{-2x^2}{x^3} + ax^4. \tag{11.23}$$

$F_{ij}$ is given by (11.4).

**(D) $w^i{}_{;i} \neq 0$, $w_i$ restricted as in (10.1)**

The form of the solution is given in Eqs. (10.2)–(10.8) and demonstrates that mathematical solutions for this case do exist. Also the fact that $w_i = e^\tau w_{,i}$ seems to contradict a result by Peres,[8] viz.

$$w_i = \varphi w_{,i} \quad \text{implies} \quad w^i{}_{;i} = 0.$$

The simplest solution of this type is obtained when $\bar{\alpha} = \bar{\gamma} = 0$, $\bar{\beta} = x^1$. The metric is then given by

$$g_{ij} = \begin{bmatrix} -\frac{2x^1}{x^2} & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & (x^2)^2 & 0 \\ 0 & 0 & 0 & (x^2)^2 \end{bmatrix}. \tag{11.24}$$

$F_{ij}$ is again given by (11.14).

## 12. CONCLUSION

In the present paper we have tried to outline procedures by means of which several infinite classes of solutions of the Einstein–Maxwell field equations may be obtained. By examining in detail some of the particular solutions, it is our hope that one might gain some insight into the physical role of the electromagnetic null field. It seems possible that some of the solutions might be used, by a variety of procedures, to produce classes of solutions to the relativistic field equations that determine purely gravitational fields.

[8] A. Peres, Ann. Phys. (N. Y.) 14, 419, (1961).

# Perturbation Method in the Diffraction of Electromagnetic Waves by Arbitrarily Shaped Penetrable Obstacles*

C. YEH
*Electrical Engineering Department, University of Southern California, Los Angeles, California*
(Received 16 November 1964)

The perturbation technique which is based on a Taylor series expansion of the boundary conditions at the perturbed boundary is extended to consider the problem of the diffraction of waves by a dielectric object with perturbed boundary. Since this approach attacks the complete boundary-value problem, the result is valid for the near zone as well as for the far zone and is valid for all frequencies. In a way of illustration, the problem of the diffraction of electromagnetic waves by a dielectric cylinder with perturbed boundary is treated. A specific example on the scattering of plane waves by a dielectric elliptic cylinder with small eccentricity is given. Numerical results are also computed for this specific example and are compared with those obtained from the exact solution.

## I. INTRODUCTION

EXACT solutions of boundary-value problems in the theory of electromagnetic wave diffraction are available only for certain specific bodies of relatively simple shape.[1,2] For example, the available exact solutions for cylindrical bodies without sharp edges are limited to those with circular, elliptic or parabolic cross-sections. The diffraction of waves by a conducting or dielectric sphere, by dielectric coated spheres and by a perfectly conducting disk are the few three dimensional problems that have been solved rigorously. The need for approximate methods to treat the more general cases of diffraction from arbitrarily shaped obstacles is quite apparent. The variational principles[3,4] provide a very powerful tool in obtaining approximate expression for the scattering cross section; but it is not possible to derive from the variational principles a description of the electromagnetic fields. Furthermore, the success of the variational approach depends to a great extent on the trial function. At low frequencies, the Rayleigh method[5,6] is very successful. However, the solutions of Laplace's equation are still required. At very high frequencies, the treatment of diffraction problems by geometric and physical optics techniques developed by Fock[7] and Keller[8] is very successful. An approximate or perturbation method in the medium frequency range still remains to be found.

In the present work the boundary perturbation technique[9,10] which is based on a Taylor expansion of the boundary conditions at the perturbed boundary will be extended to consider the problem of the diffraction of waves by a dielectric object with perturbed boundary. Since this approach attacks the complete boundary-value problem, the perturbation solution for the field components is valid for the near zone (i.e., near the obstacle) as well as for the far zone and is valid for all frequencies. In a way of illustration, the problem of the diffraction of electromagnetic waves by a dielectric cylinder with perturbed boundary will be treated. A specific example on the scattering of plane waves by a dielectric elliptic cylinder with small eccentricity will be given. The more involved case of the diffraction by a dielectric sphere with perturbed boundary can be solved in a similar manner.[11]

It is hoped that this perturbation approach will not only find applications in microwave and plasma physics but also in collision theory, acoustics, meteorology and astrophysics.[12]

## II. THE PERTURBATION SOLUTION

It is assumed that an arbitrarily shaped dielectric cylinder which has a permittivity $\epsilon_1$ and a per-

* Supported in part by the U. S. Naval Ordnance Test Station, Pasadena.

[1] R. King and T. T. Wu, *The Scattering and Diffraction of Waves* (Harvard University Press, Cambridge, Massachusetts, 1959).
[2] C. J. Bouwkamp, Repts. Progr. Phys. 17, 35 (1954).
[3] P. M. Morse and H. Feshback, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953).
[4] H. Levine and J. Schwinger, *Theory of Electromagnetic Waves* (Interscience Publishers, Inc., New York, 1951).
[5] Lord Rayleigh, Phil. Mag. 44, 28 (1897).
[6] A. F. Stevenson, J. Appl. Phys. 24, 1134 (1953).
[7] V. A. Fock, Phys. (USSR) 10, 130, 399 (1946).
[8] J. B. Keller, J. Opt. Soc. Am. 52, 102 (1962).
[9] Reference 3, p. 1052.
[10] P. C. Clemmow and V. H. Weston, Proc. Roy. Soc. (London) A264, 246 (1961).
[11] C. Yeh, Phys. Rev. 135, A1193 (1964).
[12] See H. C. van de Hulst's article in *Electromagnetic Scattering* (Pergamon Press, Ltd., Oxford, 1963).

meability $\mu_1$, is embedded in a homogeneous dielectric medium ($\epsilon_0$, $\mu_0$). The boundary of the cross section of the dielectric cylinder (Fig. 1) takes the

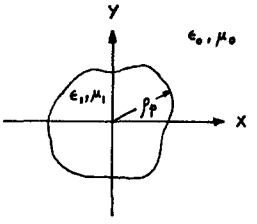


FIG. 1. The arbitrarily shaped penetrable cylinder.

shape of a perturbed circle which may be expressed by the following equation:

$$\rho_p = \rho_0[1 + \delta f_1(\phi) + \delta^2 f_2(\phi) + \cdots], \qquad (1)$$

where $\rho_0$ is the radius of the unperturbed circle, $\delta$ is a smallness parameter, and $f_n(\phi)$ are arbitrary, single-valued, continuous functions satisfying the condition

$$\sum_{n=1}^{\infty} |\delta^n f_n(\phi)| < 1, \qquad 0 \leq \phi \leq 2\pi.$$

The cylindrical coordinates $(\rho, \phi, z)$ are used.

Let the given exciting field (which need not necessarily be a plane wave) be denoted by $\mathbf{E}^{(i)}, \mathbf{H}^{(i)}$, the scattered field by $\mathbf{E}^{(s)}, \mathbf{H}^{(s)}$, and the field inside the dielectric body by $\mathbf{E}^{(t)}, \mathbf{H}^{(t)}$. The zeroth-order solution will be designated by a subscript 0, the first-order solution by subscript 1, etc. Hence, the resultant scattered fields and the resultant transmitted fields inside the body are, respectively,

$$\mathbf{E}^{(s)} = \mathbf{E}_0^{(s)} + \delta\mathbf{E}_1^{(s)} + \delta^2\mathbf{E}_2^{(s)} + \cdots,$$
$$\mathbf{H}^{(s)} = \mathbf{H}_0^{(s)} + \delta\mathbf{H}_1^{(s)} + \delta^2\mathbf{H}_2^{(s)} + \cdots, \qquad (2)$$

and

$$\mathbf{E}^{(t)} = \mathbf{E}_0^{(t)} + \delta\mathbf{E}_1^{(t)} + \delta^2\mathbf{E}_2^{(t)} + \cdots,$$
$$\mathbf{H}^{(t)} = \mathbf{H}_0^{(t)} + \delta\mathbf{H}_1^{(t)} + \delta^2\mathbf{H}_2^{(t)} + \cdots. \qquad (3)$$

The higher-order solutions are generated from known zeroth-order solution; i.e., $\mathbf{E}^{(i)}, \mathbf{H}^{(i)}, \mathbf{E}_0^{(s)}, \mathbf{H}_0^{(s)}, \mathbf{E}_0^{(t)}$, and $\mathbf{H}_0^{(t)}$ are assumed known qualities. For the sake of clarity and simplicity, only the first-order solution will be carried out in detail. The higher-order solution can be obtained in a similar fashion.

The boundary conditions require the continuity of tangential electric and magnetic fields at the boundary surface $\rho = \rho_p$:

$$\mathbf{n} \times [\mathbf{E}^{(i)}(\rho_p, \phi, z) + \mathbf{E}^{(s)}(\rho_p, \phi, z)] = \mathbf{n} \times \mathbf{E}^{(t)}(\rho_p, \phi, z), \qquad (4)$$

$$\mathbf{n} \times [\mathbf{H}^{(i)}(\rho_p, \phi, z) + \mathbf{H}^{(s)}(\rho_p, \phi, z)] = \mathbf{n} \times \mathbf{H}^{(t)}(\rho_p, \phi, z), \qquad (5)$$

where $\mathbf{n}$ is a unit vector outward normal to the boundary surface and can be written as

$$\mathbf{n} \simeq \mathbf{e}_\rho - \delta \frac{\partial f_1}{\partial \phi} \mathbf{e}_\phi \qquad (6)$$

to the first order in $\delta$ in cylindrical coordinates. $\mathbf{e}_\rho$ and $\mathbf{e}_\phi$ are, respectively, the unit vectors in $\rho$ and $\phi$ directions. $f_1(\phi)$ has been defined in Eq. (1). Carrying out the vector operations and expressing Eqs. (4) and (5) to the first order in $\delta$ in component form with the help of Eqs. (2) and (3), one obtains

$$\mathbf{e}_\rho: \quad \delta \frac{\partial f_1}{\partial \phi} [E_z^{(i)}(\rho_p, \phi, z) + E_{0z}^{(s)}(\rho_p, \phi, z)] = \delta \frac{\partial f_1}{\partial \phi} E_{0z}^{(t)}(\rho_p, \phi, z), \qquad (7)$$

$$\mathbf{e}_\phi: \quad E_z^{(i)}(\rho_p, \phi, z) + E_{0z}^{(s)}(\rho_p, \phi, z) + \delta E_{1z}^{(s)}(\rho_p, \phi, z) = E_{0z}^{(t)}(\rho_p, \phi, z) + \delta E_{1z}^{(t)}(\rho_p, \phi, z), \qquad (8)$$

$$\mathbf{e}_z: \quad E_\phi^{(i)}(\rho_p, \phi, z) + E_{0\phi}^{(s)}(\rho_p, \phi, z) + \delta\left\{E_{1\phi}^{(s)}(\rho_p, \phi, z) + \frac{\partial f_1}{\partial \phi}[E_\rho^{(i)}(\rho_p, \phi, z) + E_{0\rho}^{(s)}(\rho_p, \phi, z)]\right\} = E_{0\phi}^{(t)}(\rho_p, \phi, z) + \delta\left[E_{1\phi}^{(t)}(\rho_p, \phi, z) + \frac{\partial f_1}{\partial \phi} E_{0\rho}^{(t)}(\rho_p, \phi, z)\right], \qquad (9)$$

$$\mathbf{e}_\rho: \quad \delta \frac{\partial f_1}{\partial \phi} [H_z^{(i)}(\rho_p, \phi, z) + H_{0z}^{(s)}(\rho_p, \phi, z)] = \delta \frac{\partial f_1}{\partial \phi} H_{0z}^{(t)}(\rho_p, \phi, z), \qquad (10)$$

$$\mathbf{e}_\phi: \quad H_z^{(i)}(\rho_p, \phi, z) + H_{0z}^{(s)}(\rho_p, \phi, z) + \delta H_{1z}^{(s)}(\rho_p, \phi, z) = H_{0z}^{(t)}(\rho_p, \phi, z) + \delta H_{1z}^{(t)}(\rho_p, \phi, z), \qquad (11)$$

$$\mathbf{e}_z: \quad H_\phi^{(i)}(\rho_p, \phi, z) + H_{0\phi}^{(s)}(\rho_p, \phi, z) + \delta\left\{H_{1\phi}^{(s)}(\rho_p, \phi, z) + \frac{\partial f_1}{\partial \phi}[H_\rho^{(i)}(\rho_p, \phi, z) + H_{0\rho}^{(s)}(\rho_p, \phi, z)]\right\} = H_{0\phi}^{(t)}(\rho_p, \phi, z) + \delta\left[H_{1\phi}^{(t)}(\rho_p, \phi, z) + \frac{\partial f_1}{\partial \phi} H_{0\rho}^{(t)}(\rho_p, \phi, z)\right]. \qquad (12)$$

Equations (7) and (10) are satisfied by the zeroth-order solution. We now expand the above functions in Eqs. (8), (9), (11), and (12) to order $\delta$ in Taylor series about the unperturbed boundary $\rho = \rho_0$, obtaining

$$E_z^{(i)}(\rho_0, \phi, z) + E_{0z}^{(s)}(\rho_0, \phi, z) - E_{0z}^{(t)}(\rho_0, \phi, z) = \delta\{E_{1z}^{(t)}(\rho_0, \phi, z) - E_{1z}^{(s)}(\rho_0, \phi, z)$$

$$- \rho_0 f_1[E_z^{(i)'}(\rho_0, \phi, z) + E_{0z}^{(s)'}(\rho_0, \phi, z) - E_{0z}^{(t)'}(\rho_0, \phi, z)]\}, \quad (13)$$

$$E_\phi^{(i)}(\rho_0, \phi, z) + E_{0\phi}^{(s)}(\rho_0, \phi, z) - E_{0\phi}^{(t)}(\rho_0, \phi, z) = \delta\Big\{E_{1\phi}^{(t)}(\rho_0, \phi, z) - E_{1\phi}^{(s)}(\rho_0, \phi, z) - \rho_0 f_1[E_\phi^{(i)'}(\rho_0, \phi, z) + E_{0\phi}^{(s)'}(\rho_0, \phi, z) - E_{0\phi}^{(t)'}(\rho_0, \phi, z)] - \frac{\partial f_1}{\partial \phi}[E_\rho^{(i)}(\rho_0, \phi, z) + E_{0\rho}^{(s)}(\rho_0, \phi, z) - E_{0\rho}^{(t)}(\rho_0, \phi, z)]\Big\}, \quad (14)$$

$$H_z^{(i)}(\rho_0, \phi, z) + H_{0z}^{(s)}(\rho_0, \phi, z) - H_{0z}^{(t)}(\rho_0, \phi, z) = \delta\{H_{1z}^{(t)}(\rho_0, \phi, z) - H_{1z}^{(s)}(\rho_0, \phi, z) - \rho_0 f_1[H_z^{(i)'}(\rho_0, \phi, z) + H_{0z}^{(s)'}(\rho_0, \phi, z) - H_{0z}^{(t)'}(\rho_0, \phi, z)]\}, \quad (15)$$

$$H_\phi^{(i)}(\rho_0, \phi, z) + H_{0\phi}^{(s)}(\rho_0, \phi, z) - H_{0\phi}^{(t)}(\rho_0, \phi, z) = \delta\Big\{H_{1\phi}^{(t)}(\rho_0, \phi, z) - H_{1\phi}^{(s)}(\rho_0, \phi, z) - \rho_0 f_1[H_\phi^{(i)'}(\rho_0, \phi, z) + H_{0\phi}^{(s)'}(\rho_0, \phi, z) - H_{0\phi}^{(t)'}(\rho_0, \phi, z)] - \frac{\partial f_1}{\partial \phi}[H_\rho^{(i)}(\rho_0, \phi, z) + H_{0\rho}^{(s)}(\rho_0, \phi, z) - H_{0\rho}^{(t)}(\rho_0, \phi, z)]\Big\}, \quad (16)$$

where the prime signifies the derivative of the function with respect to $\rho_0$. The left-hand sides of the above equations are equal to zero by virtue of the zeroth-order solution. Hence, the right-hand sides of the above equations must vanish identically. Rearranging and combining Eqs. (13) and (14) gives

$$[E_{1z}^{(s)}(\rho_0, \phi, z) - E_{1z}^{(t)}(\rho_0, \phi, z)]\mathbf{e}_\phi + [E_{1\phi}^{(s)}(\rho_0, \phi, z) - E_{1\phi}^{(t)}(\rho_0, \phi, z)]\mathbf{e}_z = u_1(\rho_0, \phi, z)\mathbf{e}_\phi + u_2(\rho_0, \phi, z)\mathbf{e}_z, \quad (17)$$

and combining Eqs. (15) and (16) gives

$$[H_{1z}^{(s)}(\rho_0, \phi, z) - H_{1z}^{(t)}(\rho_0, \phi, z)]\mathbf{e}_\phi + [H_{1\phi}^{(s)}(\rho_0, \phi, z) - H_{1\phi}^{(t)}(\rho_0, \phi, z)]\mathbf{e}_z = v_1(\rho_0, \phi, z)\mathbf{e}_\phi + v_2(\rho_0, \phi, z)\mathbf{e}_z, \quad (18)$$

where

$$u_1(\rho_0, \phi, z) = \rho_0 f_1[E_{0z}^{(t)'}(\rho_0, \phi, z) - E_z^{(i)'}(\rho_0\ \phi, z) - E_{0z}^{(s)'}(\rho_0, \phi, z)],$$

$$u_2(\rho_0, \phi, z) = \rho_0 f_1[E_{0\phi}^{(t)'}(\rho_0, \phi, z) - E_\phi^{(i)'}(\rho_0, \phi, z) - E_{0\phi}^{(s)'}(\rho_0, \phi, z)] + \frac{\partial f_1}{\partial \phi}[E_{0\rho}^{(t)}(\rho_0, \phi, z) - E_\rho^{(i)}(\rho_0, \phi, z) - E_{0\rho}^{(s)'}(\rho_0, \phi, z)],$$

$$v_1(\rho_0, \phi, z) = \rho_0 f_1[H_{0z}^{(t)'}(\rho_0, \phi, z) - H_z^{(i)'}(\rho_0, \phi, z) - H_{0z}^{(s)'}(\rho_0, \phi, z)],$$

$$v_2(\rho_0, \phi, z) = \rho_0 f_1[H_{0\phi}^{(t)'}(\rho_0, \phi, z) - H_\phi^{(i)'}(\rho_0, \phi, z) - H_{0\phi}^{(s)'}(\rho_0, \phi, z)] + \frac{\partial f_1}{\partial \phi}[H_{0\rho}^{(t)}(\rho_0, \phi, z) - H_\rho^{(i)}(\rho_0, \phi, z) - H_{0\rho}^{(s)'}(\rho_0, \phi, z)]. \quad (19)$$

It is noted that the resultant fields given by Eqs. (2) and (3) must satisfy the wave equation. It is therefore clear that each term in Eqs. (2) and (3) must separately satisfy the wave equation. Consequently, the general expressions for the longitudinal components of $\mathbf{E}_1^{(s)}$, $\mathbf{H}_1^{(s)}$, $\mathbf{E}_1^{(t)}$, and $\mathbf{H}_1^{(t)}$, that are appropriate to the present problem, are[13]

$$E_{1z}^{(s)} = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} A_n H_n^{(1)}[(k_0^2 - h^2)^{\frac{1}{2}}\rho]e^{in\phi}e^{-ihz}\, dh, \quad (20)$$

$$H_{1z}^{(s)} = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} B_n H_n^{(1)}[(k_0^2 - h^2)^{\frac{1}{2}}\rho]e^{in\phi}e^{-ihz}\, dh, \quad (21)$$

$$E_{1z}^{(t)} = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_n J_n[(k_1^2 - h^2)^{\frac{1}{2}}\rho]e^{in\phi}e^{-ihz}\, dh, \quad (22)$$

$$H_{1z}^{(t)} = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} D_n J_n[(k_1^2 - h^2)^{\frac{1}{2}}\rho]e^{in\phi}e^{-ihz}\, dh, \quad (23)$$

where $H_n^{(1)}[(k_0^2 - h^2)^{\frac{1}{2}}\rho]$ and $J_n[(k_1^2 - h^2)^{\frac{1}{2}}\rho]$ are, respectively, Hankel and Bessel functions, and $k_0^2 = \omega^2\mu_0\epsilon_0$ and $k_1^2 = \omega^3\mu_1\epsilon_1$. $A_n$, $B_n$, $C_n$, and $D_n$ are yet unknown arbitrary constants that can be determined from Eqs. (17) and (18) using the orthogonality properties of the trigonomatric functions. The transverse components of $\mathbf{E}_1^{(s)}$, $\mathbf{H}_1^{(s)}$, $\mathbf{E}_1^{(t)}$, and $\mathbf{H}_1^{(t)}$ can be obtained from Maxwell's equations with the help of Eqs. (20) through (23).

Substituting the expressions for $\mathbf{E}_1^{(s)}$, $\mathbf{H}_1^{(s)}$, $\mathbf{E}_1^{(t)}$, and $\mathbf{H}_1^{(t)}$ into Eqs. (17) and (18), and making use of the orthogonality properties of the trigonometirc functions, one obtains

$$A_n H_n^{(1)}[(k_0^2 - h^2)^{\frac{1}{2}}\rho_0] - C_n J_n[(k_1^2 - h^2)^{\frac{1}{2}}\rho_0] = \frac{1}{2\pi}\int_0^{2\pi} S_1(\rho_0, \phi)e^{-in\phi}\, d\phi, \quad (24)$$

$$A_n \frac{\partial}{\partial \rho_0} H_n^{(1)}[(k_0^2 - h^2)^{\frac{1}{2}}\rho_0] - C_n \frac{\mu_0}{\mu_1}\frac{\partial}{\partial \rho_0} J_n[(k_1^2 - h^2)^{\frac{1}{2}}\rho_0]$$

[13] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941).

$$= \frac{-i\omega\mu_0}{2\pi} \int_0^{2\pi} T_2(\rho_0, \phi)e^{-in\phi}\, d\phi, \qquad (25)$$

$$B_n H_n^{(1)}[(k_0^2 - h^2)^{\frac{1}{2}}\rho_0] - D_n J_n[(k_1^2 - h^2)^{\frac{1}{2}}\rho_0] = \frac{1}{2\pi} \int_0^{2\pi} T_1(\rho_0, \phi)e^{-in\phi}\, d\phi, \qquad (26)$$

$$B_n \frac{\partial}{\partial\rho_0} H_n^{(1)}[(k_0^2 - h^2)^{\frac{1}{2}}\rho_0] - D_n \frac{\epsilon_0}{\epsilon_1} \frac{\partial}{\partial\rho_0} J_n[(k_1^2 - h^2)^{\frac{1}{2}}\rho_0] = \frac{i\omega\epsilon_0}{2\pi} \int_0^{2\pi} S_2(\rho_0, \phi)e^{-in\phi}\, d\phi, \qquad (27)$$

where $S_1$, $S_2$, $T_1$, and $T_2$ are defined as follows:

$$u_1(\rho_0, \phi, z) = \int_{-\infty}^{\infty} S_1(\rho_0, \phi)e^{-ihz}\, dh,$$

$$u_2(\rho_0, \phi, z) = \int_{-\infty}^{\infty} S_2(\rho_0, \phi)e^{-ihz}\, dh,$$

$$v_1(\rho_0, \phi, z) = \int_{-\infty}^{\infty} T_1(\rho_0, \phi)e^{-ihz}\, dh,$$

$$v_2(\rho_0, \phi, z) = \int_{-\infty}^{\infty} T_2(\rho_0, \phi)e^{-ihz}\, dh. \qquad (28)$$

$u_1$, $u_2$, $v_1$ and $v_2$ are given by Eq. (19). The coefficients $A_n$, $B_n$, $C_n$, and $D_n$ can be found readily from the above equations. Substituting these coefficients back to Eqs. (20)–(23) gives the first-order correction to the electromagnetic fields due to the departure of the boundary surface from a perfect circular cylinder with radius $\rho_0$. Higher-order corrections can be found successively in the same manner. It is interesting to note that in general the perturbed wave will have all components of electromagnetic fields even if the incident wave is a pure TE wave ($E_z^{(i)} = 0$) or a pure TM wave ($H_z^{(i)} = 0$).

## III. THE SCATTERING OF PLANE WAVES BY A DIELECTRIC ELLIPTIC CYLINDER

As an example of the application of the theory derived in Sec. II, the problem of the scattering of plane waves by a dielectric elliptic cylinder with small eccentricity will be considered. It is assumed that the incident plane wave with its electric vector polarized in the $z$ direction is propagating in the direction of the positive $x$ axis. The equation of an ellipse is given by

$$\rho_p = \frac{\rho_0}{[1 - 2\delta \sin^2(\phi - \phi_0)]^{\frac{1}{2}}}, \qquad (29)$$

where

$$\delta = \frac{1}{2}\left[1 - \left(\frac{\rho_0}{\rho_0 + \Delta\rho_0}\right)^2\right], \qquad (30)$$

$2\rho_0$ and $2(\rho_0 + \Delta\rho_0)$ are the lengths of the two axes of the ellipse. $\phi_0$ is the angle between the $x$ axis and the major axis of the ellipse if $\delta < 0$, and it is the angle between the $x$ axis and the minor axis of the ellipse if $\delta > 0$. For small eccentricity, one has

$$\rho_p \simeq \rho_0[1 + \delta \sin^2(\phi - \phi_0)]. \qquad (31)$$

Comparing Eqs. (31) and (1) gives

$$f_1(\phi) = \sin^2(\phi - \phi_0). \qquad (32)$$

The unperturbed solution to the problem of the scattering of normally incident plane ($E$) wave by a dielectric circular cylinder is well known:

$$\mathbf{E}^{(i)} = \sum_{n=-\infty}^{\infty} (i)^n J_n(k_0\rho)e^{in\phi}\mathbf{e}_z, \qquad (33a)$$

$$\mathbf{H}^{(i)} = \frac{1}{i\omega\mu_0} \sum_{n=-\infty}^{\infty} (i)^n \times \left[\frac{in}{\rho} J_n(k_0\rho)\mathbf{e}_\rho - \frac{d}{d\rho} J_n(k_0\rho)\mathbf{e}_\phi\right]e^{in\phi}, \qquad (33b)$$

$$\mathbf{E}_0^{(s)} = \sum_{n=-\infty}^{\infty} a_n(i)^n H_n^{(1)}(k_0\rho)e^{in\phi}\mathbf{e}_z, \qquad (34a)$$

$$\mathbf{H}_0^{(s)} = \frac{1}{i\omega\mu_0} \sum_{n=-\infty}^{\infty} a_n(i)^n \times \left[\frac{in}{\rho} H_n^{(1)}(k_0\rho)\mathbf{e}_\rho - \frac{d}{d\rho} H_n^{(1)}(k_0\rho)\mathbf{e}_\phi\right]e^{in\phi}, \qquad (34b)$$

$$\mathbf{E}_0^{(t)} = \sum_{n=-\infty}^{\infty} b_n(i)^n J_n(k_1\rho)e^{in\phi}\mathbf{e}_z, \qquad (35a)$$

$$\mathbf{H}_0^{(t)} = \frac{1}{i\omega\mu_1} \sum_{n=-\infty}^{\infty} b_n(i)^n \times \left[\frac{in}{\rho} J_n(k_1\rho)\mathbf{e}_\rho - \frac{d}{d\rho} J_n(k_1\rho)\mathbf{e}_\phi\right]e^{in\phi}, \qquad (35b)$$

with

$$a_n = \frac{\nu_n(k_1\rho_0)J_n'(k_0\rho_0) - (\epsilon_1\mu_0/\epsilon_0\mu_1)^{\frac{1}{2}}J_n(k_0\rho_0)J_n'(k_1\rho_0)}{(\epsilon_1\mu_0/\epsilon_0\mu_1)^{\frac{1}{2}}H_n^{(1)}(k_0\rho_0)J_n'(k_1\rho_0) - H_n^{(1)'}(k_0\rho_0)J_n(k_1\rho_0)}, \qquad (36a)$$

$$b_n = \frac{J_n'(k_0\rho_0)H_n^{(1)}(k_0\rho_0) - J_n(k_0\rho_0)H_n^{(1)'}(k_0\rho_0)}{(\epsilon_1\mu_0/\epsilon_0\mu_1)^{\frac{1}{2}}H_n^{(1)}(k_0\rho_0)J_n'(k_1\rho_0) - H_n^{(1)'}(k_0\rho_0)J_n(k_1\rho_0)}. \qquad (36b)$$

The prime in the above expressions denotes differentiation with respect to $k_0\rho_0$ or $k_1\rho_0$ as appropriate.

To find the first-order perturbation solution, we first substitute Eqs. (33) through (35) into Eq. (19)

obtaining

$$u_1(\rho_0, \phi) = \rho_0 f_1 \sum_{p=-\infty}^{\infty} \alpha_p e^{ip\phi},$$

$$u_2(\rho_0, \phi) = 0,$$

$$v_1(\rho_0, \phi) = 0,$$

$$v_2(\rho_0, \phi) = \rho_0 f_1 \sum_{p=-\infty}^{\infty} \beta_p e^{ip\phi} + \frac{\partial f_1}{\partial \phi} \sum_{p=-\infty}^{\infty} \gamma_p e^{ip\phi}, \quad (37)$$

where

$$\alpha_p = (i)^p[b_p k_1 J_p'(k_1\rho_0) - k_0 J_p'(k_0\rho_0) - a_p k_0 H_p^{(1)'}(k_0\rho_0)],$$

$$\beta_p = (i)^p\left[\frac{-b_p k_1^2}{i\omega\mu_1} J_p''(k_1\rho_0) + \frac{k_0^2}{i\omega\mu_0} J_p''(k_0\rho_0) + \frac{a_p k_0^2}{i\omega\mu_0} H_p^{(1)''}(k_0\rho_0)\right],$$

$$\gamma_p = (i)^p\left[\frac{pb_p}{\omega\mu_1\rho_0} J_p(k_1\rho_0) - \frac{p}{\omega\mu_0\rho_0} J_p(k_0\rho_0) - \frac{pa_p}{\omega\mu_0\rho_0} H_p^{(1)}(k_0\rho_0)\right]. \quad (38)$$

The expansion coefficients for the first-order perturbation fields are then found by putting expressions (37) into Eqs. (24)–(27) and carrying out the integration involving the angular functions. One has

$$A_n H_n^{(1)}(k_0\rho_0) - C_n J_n(k_1\rho_0) = \chi_n,$$

$$A_n(\rho_0 k_0) H_n^{(1)'}(k_0\rho_0) - C_n(k_1\rho_0)\left(\frac{\mu_0}{\mu_1}\right) J_n'(k_1\rho_0) = \eta_n,$$

$$B_n = 0, \qquad D_n = 0, \quad (39)$$

where

$$\chi_n = \tfrac{1}{2}[\rho_0\alpha_n - \tfrac{1}{2}\rho_0\alpha_{n+2}e^{-2i\phi_0}],$$

$$\eta_n = -\tfrac{1}{2}i\omega\mu_0[\rho_0^2\beta_n - \tfrac{1}{2}\rho_0^2\beta_{n+2}e^{-2i\phi_0} + i\rho_0\gamma_{n+2}e^{2i\phi_0}]. \quad (40)$$

The following expressions have been used:

$$\int_0^{2\pi} \sin^2(\phi - \phi_0) e^{i(p-n)\phi}\, d\phi = \pi[\delta_{p,n} - \tfrac{1}{2}e^{-2i\phi_0}\delta_{2,(p-n)}], \quad (41)$$

$$\int_0^{2\pi} \sin^2(\phi - \phi_0) e^{i(p-n)\phi}\, d\phi = i\pi e^{2i\phi_0}\delta_{2,(p-n)}, \quad (42)$$

with

$$\delta_{r,m} = 0, \quad r \neq m, \qquad = 1, \quad r = m.$$

Solving Eqs. (39) gives

$$A_n = \frac{\eta_n J_n(k_1\rho_0) - (k_1\rho_0)(\mu_0/\mu_1)\chi_n J_n'(k_1\rho_0)}{k_0\rho_0 H_n^{(1)'}(k_0\rho_0) J_n(k_1\rho_0) - (k_1\rho_0)(\mu_0/\mu_1) J_n'(k_1\rho_0) H_n^{(1)}(k_0\rho_0)}, \quad (43)$$

$$C_n = \frac{\chi_n(k_0\rho_0) H_n^{(1)'}(k_0\rho_0) - \eta_n H_n^{(1)}(k_0\rho_0)}{k_0\rho_0 H_n^{(1)}(k_0\rho_0) J_n(k_1\rho_0) - (k_1\rho_0)(\mu_0/\mu_1) J_n'(k_1\rho_0) H_n^{(1)}(k_0\rho_0)}. \quad (44)$$

Hence, the scattered fields correct to the first order in $\delta$ are

$$\mathbf{E}^{(s)} = \sum_{n=-\infty}^{\infty} [(i)^n a_n + \delta A_n] H_n^{(1)}(k_0\rho) e^{in\phi}\mathbf{e}_z, \quad (45)$$

$$\mathbf{H}^{(s)} = \frac{1}{i\omega\mu_0} \sum_{n=-\infty}^{\infty} [(i)^n a_n + \delta A_n] \times \left[\frac{in}{\rho} H_n^{(1)}(k_0\rho)\mathbf{e}_\rho - \frac{d}{d\rho} H_n^{(1)}(k_0\rho)\mathbf{e}_\phi\right] e^{in\phi}, \quad (46)$$

Of particular interest is the behavior of the backscattering cross section which is defined as the ratio of the total power scattered by a fictitious isotropic scatterer which scatters energy in all directions with intensity equal to that scattered directly back toward the source by the actual scattering object, to the incident power per unit area on the scatterer; i.e.,

$$\sigma_B^E = \lim_{\rho\to\infty} 2\pi\rho \frac{|E_z^{(s)}|^2}{|E_z^{(i)}|^2} \quad \text{at} \quad \phi = \pi, \quad (47)$$

or, using Eq. (45),

$$\sigma_B^E = \frac{4}{k_0}\left|\sum_{n=-\infty}^{\infty} [(i)^n a_n + \delta A_n] \exp i\left(\frac{n\pi}{2} - \frac{\pi}{4}\right)\right|^2. \quad (48)$$

Simplifying gives

$$\sigma_B^E = \frac{4}{k_0}\left\{\left|\sum_{n=-\infty}^{\infty} (i)^n a_n \exp i\left(\frac{n\pi}{2} - \frac{\pi}{4}\right)\right|^2 + \delta\left[\left(\sum_{n=-\infty}^{\infty} (i)^n a_n \exp i\left(\frac{n\pi}{2} - \frac{\pi}{4}\right)\right) \times \left(\sum_{n=-\infty}^{\infty} A_n \exp i\left(\frac{n\pi}{2} - \frac{\pi}{4}\right)\right)^* + \left(\sum_{n=-\infty}^{\infty} (i)^n a_n \exp i\left(\frac{n\pi}{2} - \frac{\pi}{4}\right)\right)^* \times \left(\sum_{n=-\infty}^{\infty} A_n \exp i\left(\frac{n\pi}{2} - \frac{\pi}{4}\right)\right)\right]\right\}, \quad (49)$$

where the star above the series indicates the complex conjugate of the function. The first term on the

right-hand side of the above equation represents the back-scattering cross section of an unperturbed circular cylinder, while the other term corresponds to the first-order correction due to small eccentricity.

To qualitatively illustrate how the solutions behave, the backscattering cross sections as a function of frequency are computed. Numerical computations are carried out using the high-speed IBM 7090 computer. It is assumed that $\epsilon_1/\epsilon_0 = 2.0$ and $\mu_1/\mu_0 = 1.0$. Two cases of perturbed cylindrical shape are considered: $\delta = 0.1, 0.05$. Different angles of incidence are used. Results are shown in Figs. 2 through 4. Numerical investigation shows that the first order perturbation solution should be good approximation to the exact solution for $|\delta| \leq 0.05$.

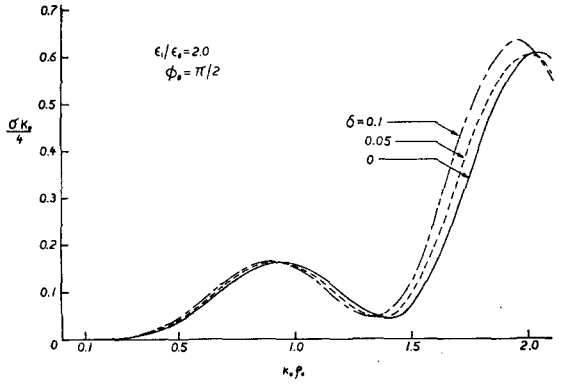


FIG. 2. The normalized back-scattering cross section for an elliptical dielectric cylinder. The direction of the incident wave is parallel to the major axis of the ellipse. $k_0\rho_0$ is the normalized semi-minor axis.

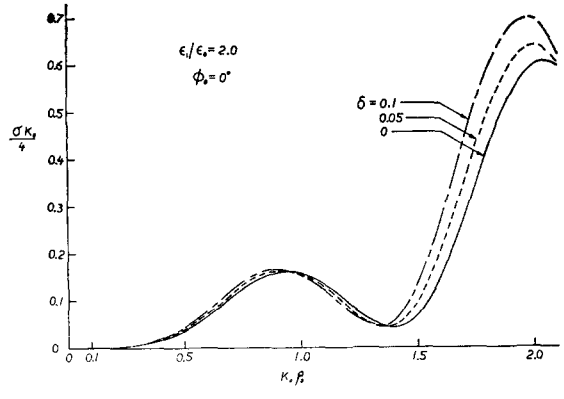


FIG. 3. The normalized back-scattering cross section for an elliptical dielectric cylinder. The direction of the incident wave is parallel to the minor axis of the ellipse. $k_0\rho_0$ is the normalized semi-minor axis.

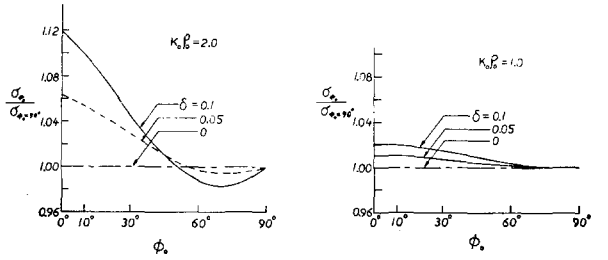


FIG. 4. Relative back-scattering cross section as a function of the direction of the incident wave.

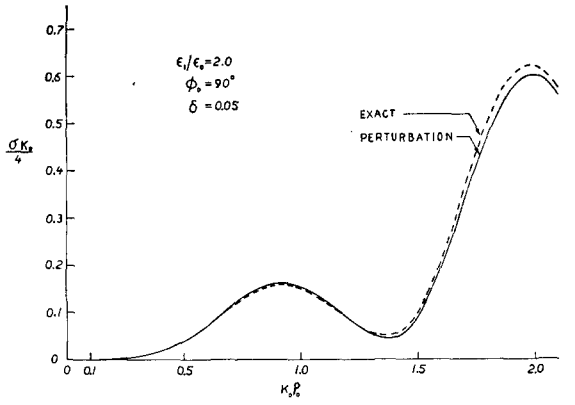


FIG. 5. Comparison between the normalized back-scattering cross section for an elliptical dielectric cylinder obtained from exact solution and that obtained according to the perturbation method.

It is always desirable to compare the results obtained from the approximate approach to available exact results. Numerical computations are therefore carried out from the exact formal solution of the diffraction of plane waves by dielectric elliptic cylinder. The exact solution is given in terms of infinite series of Mathieu functions.[14] The comparison between the exact results and the perturbation results is shown in Fig. 5. It can be seen that the agreement is very good. However it is expected that for flatter elliptical cross section, higher-order perturbation solutions must be included.

## ACKNOWLEDGMENT

The use of the computing facilities at the Western Data Processing Center at University of California, Los Angeles is gratefully acknowledged.

[14] C. Yeh, J. Math. Phys. 4, 65 (1963).

# Erratum: The Clebsch–Gordan Series for $SU(3)$

SIDNEY COLEMAN

[J. Math. Phys. 5, 1343 (1964)]
(Received 22 June 1965)

IN Eq. (2), the symbol inside the summation sign should be $(n - i, n' - j; m - j, m' - i)$. Also, in the last paragraph on p. 1343, $(n, m; n', m')$ should be $(n, n'; m, m')$.

I am indebted to S. R. Deans for calling these errors to my attention.

# Erratum: Transformation from a Linear Momentum to an Angular Momentum Basis for Particles of Zero Mass and Finite Spin

H. E. MOSES

[J. Math. Phys. **6**, 928 (1965)]
(Received 13 July 1965)

Equation (AI.1) on page 939 should read

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^{n+\beta}}{2^n(n+\alpha+\beta)!}\frac{d^{n+\alpha+\beta}}{dx^{n+\alpha+\beta}}[(1+x)^{n+\alpha}(1-x)^{n+\beta}].$$

The differential equation for $Q(x)$ on page 939 which follows Eq. (AI.5) should read

$$(1-x^2)[d^2Q(x)/dx^2] + [\beta - \alpha - (\alpha+\beta+2)x]\,dQ(x)/dx + n(n+\alpha+\beta+1)Q(x) = 0.$$