JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Correlation Inequalities for Coupled Oscillators*

HARVEY S. LEFF
Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106

(Received 7 August 1970)

Correlations in a system of N classical, coupled oscillators are studied, with a view toward obtaining a more complete understanding of Griffiths-type inequalities. The potential energy is assumed to be $U = \frac{1}{2}\sum_i\sum_j J_{ij}x_ix_j - \sum H_{kk}x_k$, with $-\infty \leq x_i \leq \infty$ and $J_{ij} = J_{ji}$. The $N \times N$ matrix with elements $\{J_{ij}\}$ is assumed positive definite. Sufficiency conditions for the correlation functions to satisfy Griffiths-type inequalities are found to be: (i) $K_{ij} \geq 0$ for all i, j, where $\mathbf{K} = \mathbf{J}^{-1}$ and (ii) $\sum_j K_{lj}H_j \geq 0$ for all l. The class of systems obeying (i) and (ii) contains those for which $J_{ij} \leq 0$, $i \neq j$, and $H_k \geq 0$ for all k; these are direct analogs of Ising ferromagnets. It is proved that a necessary and sufficient condition for Griffiths-type inequalities to hold for arbitrary $\{H_k \geq 0\}$ is simply (i) above. The sufficiency conditions (i) and (ii) are broader than those available to date for Ising models (only the sufficiency condition of ferromagnetic coupling is known). The necessary and sufficient condition (i) has no known Ising counterpart at present.

I. INTRODUCTION

The mathematical description of Ising ferromagnets has been enhanced in recent years by the consideration of inequalities for spin correlation functions. Such inequalities were proved first by Griffiths[1] for spin-$\frac{1}{2}$ systems with pairwise forces. They have been generalized to include the spin-$\frac{1}{2}$ case[2,3] and the arbitrary spin case[4,5] with many-body forces. They have been useful tools in the rigorous proofs of various statistical mechanical and thermodynamic properties for Ising ferromagnets,[6–11] and have been investigated to some extent for quantum and classical Heisenberg models,[12–15] and other models.[5]

In the present paper, Griffiths-type inequalities are studied for classical, coupled-oscillator systems. Such inequalities are found to hold under well defined circumstances. The model is tractable enough so that: (i) rather broad sufficiency conditions for their validity can be established, and (ii) a necessary and sufficient condition for their existence can be obtained. This provides the only nontrivial example (known to the author) for which such extensive conditions regarding the existence of Griffiths-type inequalities are known. For Ising models, ferromagnetic coupling is known to be a sufficient condition for Griffiths' inequalities, but broader sufficiency conditions are lacking and necessary conditions have yet to be ascertained.

In view of the still incomplete understanding[16] of the nature and scope of such inequalities, the establishment of sharp criteria for their existence, even for oversimplified models such as the present one, may shed light on the general picture. It is felt that this hope justifies publication of the present findings. The mathematical techniques employed here differ substantially from those employed in previous correlation inequality studies. Some novel aspects of matrix analysis and the theory of cumulant expansions are central to the discussion. Perhaps such techniques will find further utility in the consideration of other, related problems.

In Sec. II, the coupled-oscillator model is introduced and its canonical partition function is derived. In Sec. III, the correlation functions are introduced and their connection with the cumulants of the canonical probability distribution is discussed. The Griffiths-type correlation inequalities are investigated in Sec. IV; the major results are stated succinctly in Theorems 1–3. In Sec. V, certain monotonicity properties, which are consequences of the correlation inequalities, are elucidated for the thermodynamic potentials. Section VI contains a summary and interpretation of the results. This section closes with a brief discussion of the corresponding quantum mechanical problem, which leads to some interesting and apparently unsolved matrix-theoretic problems.

II. THE MODEL

The basic model under consideration is a collection of N particles in one dimension,[17] coupled together such that each one executes oscillations about its own equilibrium position. The coordinates $(x_1, \cdots, x_N)$ are measured relative to the N equilibrium positions. The potential energy is assumed to have the form

$$U(\mathbf{x}) = \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij}x_ix_j \geq 0, \tag{1}$$

where $-\infty < x_i < \infty$ for $i = 1, \cdots, N$. $\mathbf{x}$ represents the N-dimensional vector with Cartesian components $(x_1, \cdots, x_N)$, and equality in (1) occurs only for $\mathbf{x} = (0, \cdots, 0) \equiv \mathbf{0}$. The real coefficients $\{J_{ij}\}$ are assumed to satisfy the symmetry condition

$$J_{ij} = J_{ji} \quad \text{for all pairs } (i, j). \tag{2}$$

The right hand side of (1) may be thought of as a truncated Taylor series expansion about $\mathbf{x} = \mathbf{0}$, using the convention $U(\mathbf{0}) = 0$. Equation (2) then results from equality of the mixed second derivatives of the full potential energy (assuming the latter is continuous and has continuous first and second derivatives). Properties (1) and (2) imply that the matrix $\mathbf{J}$ of the coupling coefficients $\{J_{ij}\}$ is a real-symmetric, *positive-definite* matrix.[18]

Although a formal resemblance between (1) and the Ising model Hamiltonian for pairwise forces is evident, it is helpful to delineate explicitly some of the similarities and differences for the two models: (i) Both are bilinear forms in their variables. (ii) Here, each x_i varies continuously from $-\infty$ to $+\infty$, while for the Ising model, each x_i either takes on a discrete set of values (quantum case) or a continuum of values in some bounded domain (classical case). (iii) For the Ising model, the Hamiltonian is preceded by a minus sign, while (1) contains no such prefactor. [If a minus sign *were* included in (1), the inequality $U \geq 0$ would then make $\mathbf{J}$ a *negative*-definite matrix.] Thus, the case for which $J_{ij} \leq 0$, $i \neq j$, here corresponds directly to the *ferromagnetic* Ising model. This point is clarified further in Sec. IV. (iv) Since $\mathbf{J}$ is positive definite here, $J_{ii} > 0$ for $i = 1, \cdots, N$. For the Ising model, J_{ii} is usually chosen to be zero for all i. However, at least for the spin-$\frac{1}{2}$ case, *if* all the J_{ii} are chosen to be nonzero for the Ising model,the Griffiths inequalities are unaffected. (v) The *positive-definite* character of $\mathbf{J}$ here does *not* imply that all off-diagonal elements of $\mathbf{J}$ have the same algebraic sign. In the usual treatment of Ising ferromagnets, on the other hand, one chooses $J_{ij} \geq 0$ for $i \neq j$ and does not specify that the corresponding matrix is positive- (or negative-) definite.[19]

The coupled-oscillator system can be modified so as to simulate external magnetic field interactions by using the total potential energy

$$U_H(\mathbf{x}) = U(\mathbf{x}) - \sum_{k=1}^{N} H_k x_k, \tag{3}$$

where H_k is the "magnetic field" acting on particle k. The condition $H_k = H_l$ for all $k \neq l$ (uniform magnetic field) can be used in the final results, if so desired, but it is convenient in their derivation to work with the set $\{H_k\}$.

The canonical partition function Z for the system described by (1)–(3) is

$$\begin{aligned} Z(\mathbf{H}) &= (\Lambda^N N!)^{-1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} d\mathbf{x}^N \exp\left[-\beta U_H(\mathbf{x})\right] \\ &= (\Lambda^N N!)^{-1} Z_c(\mathbf{H}). \end{aligned} \tag{4}$$

The last line defines the configurational partition function, $Z_c(\mathbf{H})$. Λ is the thermal wavelength, $\Lambda = (2\pi m/\beta h^2)^{\frac{1}{2}}$, which depends nontrivially on the inverse temperature β and the mass m of the oscillators; h is Planck's constant.

For $\mathbf{H} = (H_1, \cdots, H_N) = \mathbf{0}$, $Z_c(\mathbf{0})$ can be evaluated explicitly by transforming to a basis in which $\mathbf{J}$ is diagonal. The result,[20] which is well known, is

$$Z_c(\mathbf{0}) = (2\pi/\beta)^{N/2} (\det \mathbf{J})^{-\frac{1}{2}}. \tag{5}$$

Notice that $\det \mathbf{J} > 0$ since $\mathbf{J}$ is positive-definite. For $\mathbf{H} \neq \mathbf{0}$, a similar procedure is possible. Let $\mathbf{R}$ denote the real-orthogonal matrix which diagonalizes $\mathbf{J}$, i.e.,

$$\mathbf{RJR}^{-1} = \mathbf{D}, \tag{6}$$

where $\mathbf{D}$ is a diagonal matrix whose elements $D_{ii} = \lambda_i^2 > 0$ are the eigenvalues of $\mathbf{J}$.[21] Define $\mathbf{y} = \mathbf{Rx}$, so that, using an obvious inner product notation,

$$U_H(\mathbf{x}) = \tfrac{1}{2}(\mathbf{y}, \mathbf{Dy}) - (\mathbf{y}, \mathbf{RH}) \equiv \tilde{U}_H(\mathbf{y}). \tag{7}$$

Since $\mathbf{R}$ is real-orthogonal, the Jacobian of the $\mathbf{x} \to \mathbf{y}$ transformation is unity, and

$$Z_c(\mathbf{H}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} d\mathbf{y}^N \exp\left[-\beta \tilde{U}_H(\mathbf{y})\right]. \tag{8}$$

Rewriting (7) as

$$\tilde{U}_H(\mathbf{y}) = \tfrac{1}{2} \sum_{i=1}^{N} \lambda_i^2 [y_i - (\mathbf{RH})_i \lambda_i^{-2}]^2 - \tfrac{1}{2} \sum_{k=1}^{N} (\mathbf{RH})_k^2 \lambda_k^{-2}, \tag{9}$$

and defining new variables $z_i = y_i - (\mathbf{RH})_i \lambda_i^{-2}$, it is clear that

$$\begin{aligned} Z_c(\mathbf{H}) &= Z_c(\mathbf{0}) \exp\left[\tfrac{1}{2}\beta \sum_{i=1}^{N} (\mathbf{RH})_i^2 \lambda_i^{-2}\right] \\ &= Z_c(\mathbf{0}) \exp\left[\tfrac{1}{2}\beta \sum_{j=1}^{N} \sum_{k=1}^{N} K_{jk} H_j H_k\right], \end{aligned} \tag{10}$$

where

$$\mathbf{K} \equiv \mathbf{J}^{-1} \tag{11}$$

is the inverse of $\mathbf{J}$, i.e., the matrix with eigenvalues $\{\lambda_i^{-2}\}$.

Notice that the existence of $\mathbf{K}$ is insured by the positive definiteness of $\mathbf{J}$. Also, $\mathbf{K}$ is positive definite since its eigenvalues are all positive, and it is symmetric because $\mathbf{J}$ is. That is, the transpose of $\mathbf{JK} = \mathbf{KJ} = \mathbf{I}$ is $\mathbf{K}^T\mathbf{J} = \mathbf{JK}^T = \mathbf{I}$; thus $\mathbf{K}^T = \mathbf{K}$.

III. CORRELATION FUNCTIONS AND CUMULANTS

The correlation function $\langle x_l \rangle$ is defined by

$$\begin{aligned}\langle x_l \rangle &= [Z_c(\mathbf{H})]^{-1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} d\mathbf{x}^N (x_l) \exp [-\beta U_H(\mathbf{x})] \\ &= \beta^{-1} \frac{\partial}{\partial H_l} [\ln Z_c(\mathbf{H})] \\ &= \sum_{j=1}^{N} K_{lj} H_j \\ &= (\mathbf{KH})_l . \end{aligned} \tag{12}$$

The symmetry of $\mathbf{K}$ was employed in conjunction with Eq. (10) in obtaining the third line in (12).

The binary correlation function $\langle x_l x_m \rangle$ can be written

$$\begin{aligned}\langle x_l x_m \rangle &= [Z_c(\mathbf{H})]^{-1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} d\mathbf{x}^N (x_l x_m) \\ &\quad \times \exp [-\beta U_H(\mathbf{x})] \\ &= -(1 + \delta_{lm}) \beta^{-1} \left(\frac{\partial}{\partial J_{lm}} \ln Z_c(\mathbf{H}) \right). \end{aligned} \tag{13}$$

The factor $(1 + \delta_{lm})$, where δ_{lm} is the Kronecker delta, arises because of the symmetry of $\mathbf{J}$. Using (5) and (10), the right-hand side of (13) can be evaluated explicitly. This process, though instructive, is somewhat tedious and is carried out in the appendix. An equivalent and more direct approach to obtain $\langle x_l x_m \rangle$ is to recognize that, due to the structure of $U_H(\mathbf{x})$,

$$\frac{\partial \langle x_l \rangle}{\partial H_m} = \beta [\langle x_l x_m \rangle - \langle x_l \rangle \langle x_m \rangle]. \tag{14}$$

Using (12), it is seen that

$$\frac{\partial \langle x_l \rangle}{\partial H_m} = K_{lm}, \tag{15}$$

whereupon (14) reduces to

$$\langle x_l x_m \rangle = \beta^{-1} K_{lm} + (\mathbf{KH})_l (\mathbf{KH})_m . \tag{16}$$

Of course, the analysis beginning with (13) (see Appendix) also leads to (16).

In analogy with (14), higher order correlation functions can be obtained generally by differentiating lower order functions. The differentiations can be carried out with respect to components of $\mathbf{H}$ or elements of $\mathbf{J}$. For example,

$$\frac{\partial \langle x_l x_m \rangle}{\partial H_n} = \beta (\langle x_l x_m x_n \rangle - \langle x_l x_m \rangle \langle x_n \rangle). \tag{17}$$

Using (16), the left hand side of (17) can be evaluated explicitly, in which case (17) can be rearranged to read

$$\begin{aligned}\langle x_l x_m x_n \rangle &= \langle x_l x_m \rangle \langle x_n \rangle + \langle x_l x_n \rangle \langle x_m \rangle \\ &\quad + \langle x_m x_n \rangle \langle x_l \rangle - 2 \langle x_l \rangle \langle x_m \rangle \langle x_n \rangle. \end{aligned} \tag{18}$$

An equivalent procedure for obtaining $\langle x_l x_m x_n \rangle$ is to write

$$\frac{\partial \langle x_n \rangle}{\partial J_{lm}} = -\beta [\langle x_l x_m x_n \rangle - \langle x_l x_m \rangle \langle x_n \rangle] (1 + \delta_{lm})^{-1}. \tag{19}$$

Using (12) and (A6), one again is led to (18).

The above discussion can be generalized and extended in scope in the following way. Let P denote the set of all particle indices $(1, 2, \cdots, N)$ and let A denote a subset of P; i.e., $A \subset P$. Further, define X_A by

$$X_A = \prod_{i \in A} x_i . \tag{20}$$

The various correlation functions $\langle X_A \rangle$ can be generated conveniently in terms of the cumulants[22] of the canonical probability distribution. The cumulants may be defined as follows. For a set of N real numbers $\{\xi_k\}$, $1 \leq k \leq N$, consider the "moment generating function"

$$\begin{aligned}&\left\langle \exp \left[\sum_{k=1}^{N} \xi_k x_k \right] \right\rangle \\ &\qquad = \sum_{\nu_1=0}^{\infty} \cdots \sum_{\nu_N=0}^{\infty} \left[\prod_{j=1}^{N} \frac{\xi_j^{\nu_j}}{\nu_j !} \right] \langle x_1^{\nu_1} \cdots x_N^{\nu_N} \rangle \\ &\qquad \equiv \exp \left\{ \sum_{\{\nu_j\}}{}' \left[\prod_{j=1}^{N} \frac{\xi_j^{\nu_j}}{\nu_j !} \right] \langle x_1^{\nu_1} \cdots x_N^{\nu_N} \rangle_c \right\}. \end{aligned} \tag{21}$$

In the last line, each ν_i runs from zero to infinity, but the primed summation excludes the term for which $\nu_i = 0$ for *every* i, $1 \leq i \leq N$. For a given set of integers $(\nu_1, \cdots, \nu_N)$ (not all zero), (21) defines the cumulants $\langle x_1^{\nu_1} \cdots x_N^{\nu_N} \rangle_c$. If $\nu_i = 1$ for $i \in A$ and $\nu_i = 0$ for $i \notin A$, then the latter cumulant is written simply as $\langle X_A \rangle_c$. Clearly, since

$$\langle X_A \rangle = \left[\prod_{i \in A} \frac{\partial}{\partial \xi_i} \right] \left\langle \exp \left[\sum_{k=1}^{N} \xi_k x_k \right] \right\rangle \Bigg|_{\xi=0}, \tag{22}$$

$\langle X_A \rangle$ can be related to the cumulants, using (21). In fact, it turns out[22] that each correlation function $\langle X_A \rangle$ is expressible in terms of the corresponding cumulant $\langle X_A \rangle_c$ and products of lower order cumulants; similarly $\langle X_A \rangle_c$ can be expressed in terms of $\langle X_A \rangle$ and products of lower order correlation functions. For example,

$$\langle x_l \rangle = \langle x_l \rangle_c \tag{23}$$

and

$$\langle x_l x_m \rangle = \langle x_l x_m \rangle_c + \langle x_l \rangle_c \langle x_m \rangle_c . \tag{24}$$

It can be shown[23] that, in general,

$$\langle X_A \rangle = \sum'' \prod_i \langle X_{B_i} \rangle_c, \tag{25}$$

where the doubly-primed summation is over all partitions of A into distinct sets B_i, whose union $\bigcup B_i$ is A itself. For each term in the series, i runs over the number of elements in the corresponding partition.

Equation (25) is instrumental in proving the theorems in the next section. A further property which is needed is: For the coupled-oscillator model,

$$\left\langle \prod_{i=1}^{N} x_i^{\nu_i} \right\rangle_c = 0 \quad \text{for} \quad \sum_{i=1}^{N} \nu_i \geq 3. \tag{26}$$

This property is well known (being most commonly stated for $H_k = 0$, $1 \leq k \leq N$) to statisticians familiar with the multivariate normal distribution.[24] It can be proved as follows. With the aid of (10), the moment generating function of (21) can be evaluated explicitly. The result is

$$\left\langle \exp\left[\sum_{k=1}^{N} \xi_k x_k\right] \right\rangle = \frac{Z_c(\mathbf{H} + \beta^{-1}\boldsymbol{\xi})}{Z_c(\mathbf{H})} = \frac{\exp\left[\frac{1}{2}\beta \sum_{j=1}^{N}\sum_{k=1}^{N} K_{jk}(H_j + \beta^{-1}\xi_j)(H_k + \beta^{-1}\xi_k)\right]}{\exp\left[\frac{1}{2}\beta \sum_{j=1}^{N}\sum_{k=1}^{N} K_{jk} H_j H_k\right]} \tag{27}$$

It follows from (21) that

$$\sum_{(\nu_i)}' \left[\prod_{j=1}^{N} \frac{\xi_j^{\nu_j}}{\nu_j!}\right] \langle x_1^{\nu_1} \cdots x_N^{\nu_N} \rangle_c = \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} K_{jk}(\beta^{-1}\xi_j\xi_k + 2\xi_j H_k). \tag{28}$$

Equating coefficients of $(\Pi_j \xi_j^{\nu_j})$ on each side of (28), it is immediately clear that (26) is true.

IV. CORRELATION INEQUALITIES

Having developed the basic mathematical formalism in the last section, it is now possible to obtain the desired Griffiths-type correlation inequalities, which are:

GTI: $\langle X_A \rangle \geq 0$ for all $A \subset P$.

GTIIa: $\partial \langle X_A \rangle / \partial H_n \geq 0$ for all $A \subset P$ and $1 \leq n \leq N$.

GTIIb: $\partial \langle X_A \rangle / \partial J_{mn} \leq 0$ for all $A \subset P$ and $1 \leq n, m \leq N$.

GTIIa and GTIIb are referred to collectively simply as GTII. The major results which follow are stated in theorem form, for precision.

Theorem 1: For a classical system of coupled oscillators defined by (1)–(3), sufficient conditions for GTI and GTII are that the inverse matrix $\mathbf{K} = \mathbf{J}^{-1}$ and the external field $\mathbf{H}$ be such that: (i) $K_{lm} \geq 0$ for $1 \leq l \leq N$, $1 \leq m \leq N$ and (ii) $\sum_{j=1}^{N} K_{lj} H_j \geq 0$ for $1 \leq l \leq M$.

Proof: From (26) it follows that the only nonzero cumulants are $\langle x_l \rangle_c$ and $\langle x_l x_m \rangle_c$ for $1 \leq m, l \leq N$. Furthermore, since $\langle x_l \rangle_c = \langle x_l \rangle$, it follows from (12) that

$$\langle x_l \rangle_c = \sum_{j=1}^{N} K_{lj} H_j. \tag{29}$$

From (12), (16), and (24) one obtains

$$\langle x_l x_m \rangle_c = \beta^{-1} K_{lm}. \tag{30}$$

Clearly, if (29) and (30) are nonnegative for $1 \leq l$, $m \leq N$, then, due to (25), all moments $\langle X_A \rangle$ are guaranteed to be nonnegative. This proves the statement of Theorem 1 regarding GTI.

From (29), (30), and (A6) one has

$$\frac{\partial \langle x \rangle_{lc}}{\partial H_n} = K_{ln}, \tag{31}$$

$$\frac{\partial \langle x_l x_m \rangle_c}{\partial H_n} = 0, \tag{32}$$

$$\frac{\partial \langle x_l \rangle_c}{\partial J_{mn}} = -(1 - \delta_{mn})[K_{lm}\langle x_n \rangle_c + K_{ln}\langle x_m \rangle_c] - \delta_{mn} K_{lm} \langle x_m \rangle_c, \tag{33}$$

$$\frac{\partial \langle x_l x_m \rangle_c}{\partial J_{nr}} = -\beta^{-1}[(K_{ln}K_{mr} + K_{lr}K_{mn})(1 - \delta_{nr}) + K_{ln}K_{mn}\delta_{nr}]. \tag{34}$$

Due to (25), derivatives of $\langle X_A \rangle$ consist of derivatives of sums of products of 1- and 2-particle cumulants. But (31)–(34) together with the nonnegativity of (29) and (30) [conditions (i) and (ii)] insure that each resulting term is nonnegative for H_n-derivatives and nonpositive for J_{mn}-derivatives. This proves the statement regarding GTII.

Remarks: The condition $K_{lm} \geq 0$ for all pairs (l, m) does *not* imply the same property for $\mathbf{J}$, nor does it imply that the off-diagonal elements of $\mathbf{J}$ are all negative. For example, for $N = 3$, consider

$$\mathbf{J} = \begin{pmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{pmatrix}, \quad \mathbf{K} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}.$$

Here, $\det \mathbf{J} = 2160$ and all subdeterminants $J^{(k)} = \det (J_{ij})$, $1 \leq i, j \leq k$ are positive; thus, $\mathbf{J}$ is positive definite.[25] The diagonal elements of $\mathbf{J}$ are all positive, as they must be, while $J_{12} = J_{21} < 0$, $J_{23} = J_{32} < 0$, but $J_{13} = J_{31} > 0$. It is concluded that the condition $K_{ij} \geq 0$ for all pairs (i, j) does *not* demand that the off-diagonal elements of $\mathbf{J}$ all have the same algebraic sign. Furthermore, by reversing the choices of $\mathbf{J}$ and $\mathbf{K}$ above, one sees that if $\mathbf{J}$ has all nonnegative elements, $\mathbf{K}$ need not have all nonnegative elements, and therefore Theorems 1 and 2 do not apply in general when $J_{ij} \geq 0$ for all pairs (i, j).

On the other hand, suppose $J_{ij} \leq 0$ for all pairs $i \neq j$, and that $J_{ii} > 0$ for all i. (Recall that $J_{ij} \leq 0$, $i \neq j$, corresponds to *ferromagnetic* coupling in the Ising model.) Then one may appeal to the following lemma.

Lemma 1 (Stieltjes–Ostrowski): If $\mathbf{J}$ is an $N \times N$ positive-definite matrix with the property $J_{ij} \leq 0$ for $i \neq j$, then $\mathbf{K} = \mathbf{J}^{-1}$ has the property $K_{lm} \geq 0$ for $1 \leq l \leq N$, $1 \leq m \leq N$.

Remarks: This lemma was proved first by Stieltjes[26] for positive-definite, real matrices, and only for strict inequalities ($J_{ij} < 0$, $i \neq j$, $K_{lm} > 0$). The generalization to $J_{ij} \leq 0$ for $i \neq j$ is due to Ostrowski.[27] Ostrowski's proof is less restrictive in that it holds for a class of real matrices ("M-matrices") which contains the class of positive definite matrices. [A real matrix is an M-matrix if: (i) its diagonal elements are positive, (ii) its off-diagonal elements are nonpositive, (iii) its determinant and all its principal subdeterminants are positive.[28]]

Theorem 2: For the system of Theorem 1 a sufficient set of conditions for GTI and GTII is

(i) $J_{ij} \leq 0$, $i \neq j$

and

(ii) $\sum_{l=1}^{N} K_{jl} H_l \geq 0$ for $1 \leq j \leq N$.

Proof: By Lemma 1, (i) implies that $K_{lm} \geq 0$ for all pairs (l, m). Thus, (i) and (ii) reduce to the sufficiency conditions of Theorem 1.

Remarks: A special case of (ii) is $H_l \geq 0$ for $1 \leq l \leq N$. This contains the uniform field case: $H_l = H \geq 0$, $1 \leq l \leq N$.

Theorem 3: For the system of Theorem 1, a necessary and sufficient condition for GTI and GTII to hold for all $H_k \geq 0$, $1 \leq k \leq N$, is $K_{lm} \geq 0$ for $1 \leq l \leq N$, $1 \leq m \leq N$.

Proof: The sufficiency part is clear from Theorem 1. The necessity can be seen as follows. If $\langle x_l \rangle = \sum K_{lj} H_j \geq 0$ for all $H_j \geq 0$ and for all l, then $K_{lj} \geq 0$ for all pairs (l, j). This is necessary since, for example, if $K_{lj} < 0$ for $j = p$, then one could choose $H_j = 0$ for $j \neq p$ and $H_p > 0$, whereupon the contradiction $\langle x_l \rangle < 0$ would occur. Thus $K_{lj} \geq 0$ for all pairs (l, j) is necessary for $\langle x_l \rangle \geq 0$ for all l. Furthermore, the nonnegativity of $\mathbf{K}$ is necessary for

$$\langle x_l x_m \rangle = \beta^{-1} K_{lm} + (\sum K_{li} H_i)(\sum K_{mj} H_j) \geq 0.$$

To see this, choose $H_j = 0$ for all j, whereupon $\langle x_l x_m \rangle = \beta^{-1} K_{lm}$. This completes the proof.

Remarks: Theorem 3 holds, in particular, for the uniform field case $H_k = H$ for $1 \leq k \leq N$, and all $H \geq 0$. A further special case of interest is that for which $H \equiv 0$.

The following theorem is included here because it provides a necessary and sufficient condition on the matrix $\mathbf{J}$ such that $\mathbf{K} = \mathbf{J}^{-1}$ has all positive elements. The theorem is relatively complicated to state and its proof is nontrivial. The interested reader is directed to the paper by Fiedler and Pták[29] for the details. In Sec. VI the physical implications of this necessary and sufficient condition are examined.

Theorem 4 (Fiedler–Pták): Let $\mathcal{S}_0$ denote the class of real matrices $\mathbf{B}$ with the property that there exists a vector $\mathbf{x}^0 \neq 0$ with components $x_i^0 \geq 0$, $1 \leq i \leq N$, for which the components $(\mathbf{B}\mathbf{x}^0)_l \geq 0$, $1 \leq l \leq N$. Let $\mathcal{M}$ denote the class of all real matrices $\mathbf{J}$ such that: (i) $\mathbf{J} \in \mathcal{S}_0$ and (ii) no submatrix obtained from $\mathbf{J}$ by omitting at least one column thereof belongs to $\mathcal{S}_0$. Then, $\mathbf{J}^{-1}$ has all positive elements if and only if $\det(\mathbf{J}) \neq 0$ and $\mathbf{J} \in \mathcal{M}$.

V. THERMODYNAMIC RAMIFICATIONS

In this section several monotonicity properties of the thermodynamic potentials are proved for coupled-oscillator systems for which $J_{ij} \leq 0$ and $H_j \geq 0$ for $1 \leq i, j \leq N$. These properties are implied by Lemma 1 and Theorem 2 of Sec. IV, and are summarized in Theorems 5 and 6 below. A similar investigation without the above restrictions on $\mathbf{J}$ and $\mathbf{H}$ is possible (but it is less transparent) and is left as an exercise for the interested reader.

Theorem 5: For the system of Theorem 1, with $J_{ij} \leq 0$, $i \neq j$ and $H_k \geq 0$, $1 \leq k \leq N$, the average total energy and Helmholtz free energy are monotone decreasing functions of each H_k, $1 \leq k \leq N$. The entropy is independent of the set $\{H_k\}$.

Proof: From (1), (3), (12), and (16) one obtains

$$\langle U_H \rangle = \tfrac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N} J_{kl}\langle x_k x_l\rangle - \sum_{k=1}^{N} H_k\langle x_k\rangle$$
$$= N/2\beta - \tfrac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N} K_{kl}H_kH_l. \quad (35)$$

Therefore, using (12), (35), and Theorem 2,

$$\frac{\partial\langle U_H\rangle}{\partial H_n} = -\langle x_n\rangle \leq 0 \quad \text{for} \quad 1 \leq n \leq N. \quad (36)$$

The average total energy, $\langle E\rangle$ is simply $\langle U_H\rangle$ plus the average kinetic energy, which is easily seen to be $N/2\beta$. Therefore,

$$\frac{\partial\langle U_H\rangle}{\partial H_n} = \frac{\partial\langle E\rangle}{\partial H_n} \leq 0 \quad \text{for} \quad 1 \leq n \leq N. \quad (37)$$

The Helmholtz free energy $F = -\beta^{-1}\ln Z$ satisfies

$$\frac{\partial F}{\partial H_n} = -\langle x_n\rangle \leq 0 \quad \text{for} \quad 1 \leq n \leq N. \quad (38)$$

The entropy is $S = k\beta(\langle E\rangle - F)$ and, using (36)–(38), it becomes clear that

$$\frac{\partial S}{\partial H_n} = 0 \quad \text{for} \quad 1 \leq n \leq N. \quad (39)$$

The results (37)–(39) establish the desired theorem.

Remarks: The monotonicity properties for the energy and Helmholtz free energy are shared by a general class of ferromagnetic Ising models.[8] For the latter, the entropy is also a monotone decreasing function of the external field components. Here, this is also the case *technically*. (S is both monotone decreasing and increasing!) However, the present entropy behavior is clearly far removed from that of the Ising model. The entropy constancy is discussed further in Sec. VI. Notice that, for $\mathbf{H} = 0$, one obtains $\langle E\rangle = N/\beta$, which is consistent with the well-known equipartition theorem.

Theorem 6: For the system of Theorem 1, with $J_{ij} \leq 0$, $i \neq j$, and $H_k \geq 0$, $1 \leq k \leq N$, the average total energy and Helmholtz free energy are monotone increasing functions, and the entropy is a monotone decreasing function of the coupling coefficients $\{J_{ij}\}$.

Proof: From (35), using (A6), one obtains

$$\frac{\partial\langle U_H\rangle}{\partial J_{mn}} = \tfrac{1}{2}(2 - \delta_{mn})\langle x_n\rangle\langle x_m\rangle \geq 0. \quad (40)$$

The inequality follows from Theorem 2. It follows that

$$\frac{\partial\langle U_H\rangle}{\partial J_{mn}} = \frac{\partial\langle E\rangle}{\partial J_{mn}} \geq 0 \quad \text{for} \quad 1 \leq m, n \leq N. \quad (41)$$

Furthermore, it is a straightforward exercise to show that

$$\frac{\partial F}{\partial J_{mn}} = \tfrac{1}{2}(2 - \delta_{mn})\langle x_m x_n\rangle \geq 0 \quad \text{for} \quad 1 \leq m, n \leq N. \quad (42)$$

Again, the inequality follows from Theorem 2. Finally, using (40)–(42), one sees that

$$\frac{\partial S}{\partial J_{mn}} = -\tfrac{1}{2}k(2 - \delta_{mn})K_{mn} \leq 0 \quad \text{for} \quad 1 \leq m, n \leq N. \quad (43)$$

The results (41)–(43) establish Theorem 6.

Remarks: For ferromagnetic Ising models, the energy, Helmholtz free energy, and entropy are all monotone decreasing functions of the (positive) coupling coefficients.[8] The first two properties are actually shared in essence by the present model, the opposite sense of the inequalities being due solely to the absence of a minus one prefactor in $U(\mathbf{x})$, Eq. (1). The entropy property here is identical in form with that for Ising ferromagnets.

VI. DISCUSSION

A. Major Results

The present analysis yields new results regarding the nature and scope of Griffiths-type inequalities. The content of Theorem 2 provides a direct analog of Griffiths' first (GTI) and second (GTII) inequalities for Ising ferromagnets. Theorem 1 provides broad sufficiency conditions for Griffiths-type inequalities and Theorem 3 provides a necessary and sufficient condition for these inequalities. The sense of the inequality GTIIb is opposite to that for Ising models, but this is due simply to the absence of a prefactor minus one in the zero field potential energy, Eq. (1). Of course, the coupled oscillator model is more simple and less interesting than the Ising model.[30] On the other hand, it is apparently the only nontrivial model for which such detailed information regarding correlation inequalities is known. In view of the strong resemblance of its potential energy structure and resulting correlation inequalities with their Ising model counterparts, it seems reasonable to use it in an attempt to deepen the understanding of correlation inequalities. Presumably, corresponding sufficiency conditions for the Ising model, which are broader than simply the criterion of ferromagnetic coupling, and also necessary and sufficient conditions, will be

found ultimately. Perhaps the present results will aid or at least provide further stimulus for such efforts.

B. Physical Interpretation

For a 1-dimensional coupled oscillator system, the condition $J_{mn} \leq 0$ for $m \neq n$ is physically reasonable since the force on particle m due to its pair interaction with particle n is $F_{m,n} = -J_{mn}x_n$. This is expected to be positive (negative) when $x_n > 0$ ($x_n < 0$), which implies $J_{mn} < 0$. Thus, if one particle is shifted to the right all other particles with which it interacts tend to shift to the right as well. This statement makes the inequality $\langle X_A \rangle \geq 0$, which means in essence that the oscillator positions are correlated *positively*, intuitively reasonable. The effect of a positive field H_k is to shift *each* average value $\langle x_l \rangle$ to the right, which tends to enhance the *positive* correlations between positions. These statements give physical meaning to GTI and GTIIa.

Since an algebraic increase of an off-diagonal element, $J_{mn} < 0$, decreases the magnitude of the force $F_{m,n} = -J_{mn}x_n$ for a given value of x_n, it is expected that $\langle x_m x_n \rangle$ decreases correspondingly. Similarly, one expects the general correlation functions $\langle X_A \rangle$ to decrease under a decrease of some J_{mn}. This explains GTIIb in physical terms.

C. Entropy and Order

The thermodynamic monotonicity properties can also be argued on physical grounds. Consider here the entropy property. Since an increase in J_{mn} for $m \neq n$ ($J_{mn} < 0$) decreases the magnitude $|J_{mn}|$, this corresponds to weaker coupling between particles m and n. This, in turn, tends to increase the localization of the particles in the neighborhoods of their respective equilibrium sites; this increased localization may be thought of as increased order. Similarly, the increase of a (positive) diagonal coefficient J_{ii} increases the localization of particle i relative to its equilibrium position. Again, the order increases. These statements are consistent with Theorem 5, using the idea that an increase in order is reflected by a decrease in entropy.[31,32] An increase in a component $H_k \geq 0$ shifts each particle's average position to the right (not necessarily all by the same amount) and has no net effect on the systems order; thus S is unchanged under such a variation (Theorem 5).

D. Fiedler–Pták Theorem

It seems likely that, on the basis of Lemma 1, the condition $K_{ij} \geq 0$ for all pairs (i,j) is implied for a large class of matrices $\mathbf{J}$, which have "predominantly nonpositive, off-diagonal elements." For example, the 3×3 matrix illustrated in Sec. IV has this property by virtually any reasonable definition thereof. (One such criterion is, say, $\sum_{i<j}\sum J_{ij} < 0$; for the 3×3 example this sum equals -186.) Some such condition would presumably be *sufficient* for the nonnegativity of $\mathbf{K}$. On the other hand, Theorem 4 gives *necessary* and *sufficient* conditions on $\mathbf{J}$ for the positivity of $\mathbf{K}$ (i.e., $K_{ij} > 0$, $1 \leq i, j \leq N$). It is natural to ask what physical information, if any, can be extracted from the highly abstract statement of this theorem. It is shown below that the above intuitive reasoning is correct and that Theorem 4 has a pleasing interpretation.

Using the notation of Theorem 4, note first that every positive-definite matrix $\mathbf{J}$ is contained in the set $\mathcal{S}_0$. This point is proved in the paper by Fiedler and Pták.[29] This means that for the coupled oscillator problem there exists a nonzero $\mathbf{x}^0$, with nonnegative components, such that $\sum_j J_{ij}x_j^0 \geq 0$ for $1 \leq i \leq N$. Notice, however, that for the zero external field case, the net force on particle i is

$$F_i = -\frac{\partial U}{\partial x_i} = -\sum_{j=1}^{N} J_{ij}x_j. \tag{44}$$

Therefore, the statement $\mathbf{J} \in \mathcal{S}_0$ means that there exists a nonequilibrium configuration $\mathbf{x}^0$ for which no oscillator lies to the left of its equilibrium site, and for which the net force on *each* particle is either zero or is directed toward the left, i.e.,

$$F_i^0 \leq 0 \quad \text{for} \quad 1 \leq i \leq N. \tag{45}$$

It is instructive to examine condition (45) for the special case for which $J_{ij} \leq 0$, $i \neq j$. Then,

$$F_i^0 = -|J_{ii}x_i^0| + \sum_{j \neq i} |J_{ij}x_j^0|, \tag{46}$$

and (45) means: For some nonequilibrium configuration of the above type, the force on each particle due to the binding to its equilibrium site is not exceeded by the net force due to the remaining $N - 1$ oscillators. The former is directed to the left and the latter to the right.

Now, returning to the statement of Theorem 4, let $\mathcal{R}$ denote a *proper* subset of the index set P, and suppose column r is removed from $\mathbf{J}$ for each $r \in \mathcal{R}$. The condition $\mathbf{J} \notin \mathcal{S}_0$ means that for each set $\mathcal{R} \subset P$

$$F_i^0(\mathcal{R}) \equiv -\sum_{j \notin \mathcal{R}} J_{ij}x_j^0 > 0 \tag{47}$$

for at least one value of i, for *every* nonequilibrium set

$\{x_i^0 \geq 0\}$. This may be interpreted as a condition for "predominantly nonpositive off-diagonal elements," confirming the above intuitive expectations regarding such a condition.

For the special case, $J_{ij} \leq 0$, $i \neq j$, one has

$$F_i^0(\mathcal{R}) = -|J_{ii}x_i^0| + \sum_{\substack{j \notin \mathcal{R} \\ j \neq i}} |J_{ij}x_j^0|, \quad \text{for} \quad i \notin \mathcal{R}, \quad (48)$$

and

$$F_i^0(\mathcal{R}) = \sum_{j \notin \mathcal{R}} |J_{ij}x_j^0| \qquad \text{for} \quad i \in \mathcal{R}. \quad (49)$$

If (46) is nonpositive for all $i \notin \mathcal{R}$, for some nonequilibrium configuration $\{x_i^0 \geq 0\}$, so is (48), since nonnegative terms have been deleted from (46). However, (49) is intrinsically nonnegative for all $i \in \mathcal{R}$, and for *any* nontrivial choice of the set $\{x_i^0 \geq 0\}$. For this case, it is clear that for every $\mathcal{R} \subset P$, $\mathbf{J}^{\mathcal{R}}$ (i.e., $\mathbf{J}$, with columns r, $r \in \mathcal{R}$, removed) does not lie in $\mathcal{S}_0$. By definition then, $\mathbf{J} \in \mathcal{M}$ and Theorem 4 applies. Physically, (49) means: The force on the ith oscillator, excluding the coupling to its equilibrium site, and excluding any number of other couplings, is nonnegative for $\{x_j^0 \geq 0\}$ and $\mathbf{x}^0 \neq 0$.

Conditions similar to (45) and (47) may also be necessary and sufficient for Griffiths' Ising model inequalities, but this is left as an open question here.

E. A Converse Problem

A question which is related to the converse of Theorem 4 has been posed by Kelly and Sherman[2] in the context of the Ising model. In the present context, their question reads: *What conditions on the correlation functions* $\langle X_A \rangle$ imply that $J_{ij} \leq 0$ for $i \neq j$? According to (25), (26), (29), and (30), all correlation functions are expressible in terms of the elements of $\mathbf{K}$ and $\mathbf{H}$. Therefore, the Kelly–Sherman question for the present model is related to the following.

Question: What conditions on a real-symmetric, positive-definite matrix $\mathbf{K}$ imply that its inverse $\mathbf{J}$ satisfies $J_{ij} \leq 0$ for $i \neq j$?

F. Quantum Mechanical Extension

It is natural to ask whether or not the present results apply to quantum mechanical systems with potential energy given by (1)–(3). Although the answer to this question is unknown at present, the problem leads to some interesting mathematical questions which are sketched briefly below. Note first that, since each particle's position and momentum no longer commute, the quantum coupled oscillator model is *not* expected to be a suitable analog for the Ising model.

As in the classical case, the magnetic field contributions occur as a multiplicative factor in the canonical partition function. Denoting the partition function by Z_{qm} (qm stands for quantum-mechanical),

$$Z_{\text{qm}}(\mathbf{H}) = Z_{\text{qm}}(\mathbf{0}) \exp\left[\tfrac{1}{2}\beta \sum_{j=1}^{N} \sum_{k=1}^{N} K_{jk}H_jH_k\right]. \quad (50)$$

The zero external field case, involving $Z_{\text{qm}}(\mathbf{0})$, has been discussed in part by Ford, Kac, and Mazur[34] in a paper dealing primarily with time-dependent phenomena for coupled oscillators. They found that for particles of unit mass

$$\langle x_l x_m \rangle_{\text{qm}} = \beta^{-1} L_{lm}, \quad (51)$$

where $\mathbf{L}$ is a real-symmetric, positive-definite matrix which commutes with $\mathbf{J}$ and $\mathbf{K}$, and which has eigenvalues

$$\mu_i^2 = b\lambda_i^{-1} \coth(b\lambda_i), \quad (52)$$

where

$$b = \hbar\beta/2. \quad (53)$$

They showed also that the binary momentum-momentum and momentum-position correlation functions satisfy

$$\langle p_l p_m \rangle_{\text{qm}} = \beta^{-1}(KL)_{lm} \quad (54)$$

and

$$\langle p_l x_m \rangle_{\text{qm}} = -\langle x_l p_m \rangle_{\text{qm}} = \tfrac{1}{2}i\hbar\delta_{lm}. \quad (55)$$

For small values of the parameter b (classical limit), it is clear that

$$\mu_i^2 \approx \lambda_i^{-2} + \tfrac{1}{3}b^2. \quad (56)$$

This implies that, to order b^2,

$$\mathbf{L} = \mathbf{K} + \tfrac{1}{3}b^2 I. \quad (57)$$

Therefore, $\mathbf{L}$ has all nonnegative elements if K does (or even if K has very small, negative off-diagonal elements). It follows that for sufficiently small values of b, a sufficient condition for (51) and (54) to be nonnegative is that $\mathbf{K}$ be a nonnegative matrix. In turn, a sufficiency condition for $\mathbf{K}$ to be nonnegative is that $\mathbf{J}$ have nonpositive off-diagonal elements (Lemma 1).

For very large values of the parameter b (extreme quantum limit),

$$\mu_i^2 \approx b\lambda_i^{-1} \quad (58)$$

and, apparently,

$$\mathbf{L} \approx b\mathbf{K}^{\frac{1}{2}}, \quad (59)$$

where $\mathbf{K}^{\frac{1}{2}}$ is the unique (positive-definite) square root matrix of $\mathbf{K}$. If $\mathbf{K}^{\frac{1}{2}}$ has all nonnegative elements, then the same is true for $\mathbf{K}$. Therefore, if for very large values of b, $\langle x_l x_m \rangle_{\text{qm}} \geq 0$ for all (l, m), then the

corresponding classical result is implied. A meaningful discussion of the converse problem requires an answer to the following.

Question: Suppose the positive-definite matrix **K** has the property $K_{ij} \geq 0$ for all pairs (i, j). What then is implied about the algebraic signs of $(K^{\frac{1}{2}})_{ij}$ for $i \neq j$?

For intermediate values of b one is led to questions of the following nature. Suppose that **K** and **L** are the two commuting, real-symmetric matrices defined above, with eigenvalues $\{\lambda_i^{-2}\}$ and $\{b\lambda_i^{-1} \coth b\lambda_i\}$, respectively.

Question (i)*:* If **K** has all nonnegative elements, what then is implied about the algebraic signs of L_{ij} for $i \neq j$?

Question (ii)*:* If $\mathbf{J} = \mathbf{K}^{-1}$ has nonpositive, off-diagonal elements, what then is implied about the algebraic signs of $(L^{-1})_{ij}$ for $i \neq j$?

Question (iii)*:* The matrix $\mathbf{L} - \mathbf{K}$ has eigenvalues $\{b\lambda_i^{-1}\mathfrak{L}(b\lambda_i)\}$, where $\mathfrak{L}(x)$ is the Langevin function: $\mathfrak{L}(x) \equiv \coth(x - x^{-1})$. If **K** has all nonnegative elements, what then is implied about the algebraic signs of $(\mathbf{L} - \mathbf{K})_{ij}$ for $i \neq j$?

It appears that an understanding of correlation inequalities for quantum mechanical, coupled oscillators requires answers to well defined, but apparently nontrivial, mathematical questions such as those above.

Note added in proof: (a) S. Sherman [J. Math. Phys. **11,** 2480 (1970)] has recently obtained necessary and sufficient conditions on the correlation functions to guarantee that an *Ising magnet* with two-body interactions, in a nonnegative external magnetic field, is a ferromagnet, i.e., $J_{ij} \geq 0$. This relates to Sec. VIE of the present paper.

(b) For the coupled-oscillator problem, suppose that $K_{lm} \leq 0$ for $l \neq m$. Then, by Lemma 1, $J_{ij} \geq 0$ for all i, j. Therefore, due to Eq. (30), a sufficient condition for "antiferromagnetic coupling" $(J_{ij} \geq 0)$ for the coupled-oscillator problem is $\langle x_l x_m \rangle_c = \langle x_l x_m \rangle - \langle x_l \rangle \langle x_m \rangle \leq 0$ for $l \neq m$. Furthermore, Theorem 4 provides necessary and sufficient conditions on the matrix with elements $\langle x_l x_m \rangle_c$ for "antiferromagnetic coupling."

ACKNOWLEDGMENTS

The author would like to thank Prof. P. R. Zilsel, Prof. P. L. Taylor, Dr. J. Czika, and Dr. K. Y. Millard for informative and stimulating discussions on various aspects of this work.

APPENDIX

Here the objective is to derive Eq. (16), beginning with Eq. (13). From (5), (10), and (13)

$$\langle x_l x_m \rangle = (2\beta)^{-1}(1 + \delta_{lm}) \times \left[\frac{\partial \ln(\det \mathbf{J})}{\partial J_{lm}} - \beta \sum_{j=1}^{N} \sum_{k=1}^{N} H_j H_k \left(\frac{\partial K_{jk}}{\partial J_{lm}}\right)\right]. \tag{A1}$$

The first derivative on the right hand side can be written

$$\frac{\partial \ln(\det \mathbf{J})}{\partial J_{lm}} = (\det \mathbf{J})^{-1} \sum_{i=1}^{N} \det \mathbf{J}'(i), \tag{A2}$$

where $\mathbf{J}'(i)$ is obtained from **J** by replacing the ith column, $\{J_{ki}; i = 1, \cdots, N\}$ by $\{\partial J_{ki}/\partial J_{lm}; k = 1, \cdots, N\}$. Therefore $\det \mathbf{J}'(i) = 0$ unless $i = l$ or $i = m$. But $\det \mathbf{J}'(l)$ and $\det \mathbf{J}'(m)$ are simply the ml and lm cofactors, respectively, of **J** [if $l = m$, of course, only *one* term in (A2) is nonzero], and it is well known that $\mathbf{J}^{-1} = \mathbf{K}$ is such that $K_{ml} = lm$ cofactor of $\mathbf{J} \times (\det \mathbf{J})^{-1}$. Furthermore, since **J** is symmetric, so is **K**. (See the last paragraph of Sec. II.) Therefore,

$$\frac{\partial \ln(\det \mathbf{J})}{\partial J_{lm}} = (2 - \delta_{lm})K_{lm}. \tag{A3}$$

In order to evaluate the second set of terms on the right-hand side of (A1), one needs $\partial K_{jk}/\partial J_{lm}$. This can be obtained by first differentiating

$$\sum_{k=1}^{N} J_{rk} K_{kj} = \delta_{rj} \tag{A4}$$

with respect to J_{lm}, which gives

$$-\sum_{k=1}^{N} J_{rk}(\partial K_{kj}/\partial J_{lm}) = \sum_{k=1}^{N} K_{kj}[(\delta_{rl}\delta_{km} + \delta_{rm}\delta_{kl})(1 - \delta_{lm}) + \delta_{rl}\delta_{kl}\delta_{lm}]. \tag{A5}$$

Multiplication of (A5) by K_{ri} and subsequent summation over r yields

$$\frac{\partial K_{ij}}{\partial J_{lm}} = -(K_{il}K_{jm} + K_{im}K_{jl})(1 - \delta_{lm}) - K_{il}K_{jl}\delta_{lm}. \tag{A6}$$

Using (A6), and (A3) in (A1), one obtains (16).

* Supported in part by the U.S. Atomic Energy Commission.

[1] R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967).

[2] D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968).

[3] J. Ginibre, Phys. Rev. Letters **23**, 828 (1969).

[4] R. B. Griffiths, J. Math. Phys. **10**, 1559 (1969).

[5] J. Ginibre, Commun. Math. Phys. **16**, 310 (1970).

[6] See Refs. 1 and 3, and especially the articles cited in the latter.

[7] M. E. Fisher, Phys. Rev. **180**, 594 (1969).

[8] H. S. Leff, Phys. Rev. A **7**, 2368 (1970).

[9] D. C. Mattis and M. Plischke, J. Math. Phys. **10**, 1107 (1969).

[10] R. H. T. Yeh, Phys. Letters **30A**, 297 (1969).

[11] R. B. Griffiths, C. A. Hurst, and S. Sherman, J. Math. Phys. **11**, 790 (1970).

[12] C. A. Hurst and S. Sherman, Phys. Rev. Letters **23**, 1357 (1969).

[13] R. H. T. Yeh, J. Math. Phys. **11**, 1317 (1970); Phys. Rev. Letters **23**, 1241 (1969).

[14] R. B. Griffiths, Phys. Rev. **1**, 3883 (1970).

[15] T. Asano, Phys. Rev. Letters **24**, 1409 (1970).

[16] Quantum Heisenberg models and antiferromagnetic Ising models provide further examples for which a detailed understanding is lacking.

[17] The restriction to one dimension is not necessary in what follows. It is, however, convenient with regard to both notation and physical interpretation. The results generalize to a ν-dimensional space if the one-dimensional position coordinates $(x_1, \cdots, x_N)$ are replaced by the νN Cartesian components $(x_1, \cdots, x_{\nu N})$.

[18] R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill, New York, 1968), p. 40.

[19] The author is aware of several treatments of the Ising model which do explicitly utilize the assumption of positive-definiteness for the matrix of the coupling coefficients. See A. J. F. Siegert, *Statistical Physics—1962 Brandeis Lectures*, Vol. 3 (Benjamin, New York, 1963), p. 159. The mathematical analysis is related to the present one in some ways and suggests a possible link between the two problems which deserves further investigation. See also A. J. F. Siegert and D. J. Vezetti, J. Math. Phys. **9**, 2173 (1968); C. J. Thompson, A. J. F. Siegert, and D. J. Vezetti, J. Math. Phys. **11**, 1018 (1970).

[20] See, for example, Ref. 18, p. 96.

[21] A necessary and sufficient condition for **J** to be positive definite is that $\lambda_i^2 > 0$ for $i = 1, \cdots, N$. See Ref. 18, p. 54.

[22] For a discussion of cumulants for distributions of a single variable see H. Cramér, *Mathematical Methods of Statistics* (Princeton U.P., Princeton, N.J., 1961), p. 186. The single variable case and a generalization to the bivariate and multivariate cases is contained in M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Hafner, N.Y., 1963), Chap. 3. The multivariate case is discussed further in: E. Meeron, J. Chem. Phys. **27**, 1238 (1957); R. Kubo, J. Phys. Soc. Japan **17**, 1100 (1962); I. Z. Fisher, *Statistical Theory of Liquids* (University of Chicago Press, Chicago, 1964), see the supplement by S. A. Rice and P. Gray; J. Yvon, *Correlations and Entropy in Classical Statistical Mechanics* (Pergamon, New York, 1969), Sec. 2.6.

[23] See the works of Meeron, Kubo, Fisher, and Yvon, Ref. 22.

[24] See, for example, the book of Kendall and Stuart, Ref. 22, p. 349.

[25] The conditions $J^{(k)} > 0$, $1 \leq k \leq N$, are necessary and sufficient conditions for **J** to be positive definite. See Ref. 18, p. 74.

[26] T. J. Stieltjes, Acta. Math. **9**, 385 (1886/87). A proof may be found in Ref. 18, p. 153.

[27] A. Ostrowski, Comment. Math. Helveticii **30**, 175 (1956); **10**, 69 (1937) Ostrowski does not refer to the work of Stieltjes and it is not clear that he was aware of it.

[28] An extensive study of M-matrices (in English) may be found in M. Fiedler and V. Pták, Czech. Math. J. **12**, 382 (1962).

[29] M. Fiedler and V. Pták, Num. Math. **9**, 163 (1966).

[30] For example, if the cumulant $\langle x_k x_l x_m \rangle_c$ for the Ising model were identically zero the magnetic susceptibility would be linear in the (uniform) external field, or trivially constant. This, of course, is not so in general. See Ref. 11 for a discussion of the convexity of the magnetization.

[31] For a further discussion of this point, see H. S. Leff, Am. J. Phys. **37**, 548 (1969); R. Baierlein, Am. J. Phys. **36**, 625 (1968). See also Ref. 8.

[32] H. Falk and M. Suzuki, Phys. Rev.B **1**, 3051 (1970); H. Falk, to be published. See also the book of Yvon, Ref. 22.

[33] For completeness, it is pointed out that an essentially different theorem, supplying necessary and sufficient conditions for a matrix to have a nonnegative inverse, has been given recently by P. G. Ciarlet, Lin. Alg. Appl. **1**, 139 (1968).

[34] G. W. Ford, M. Kac, and P. Mazur, J. Math. Phys. **6**, 504 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

A Scalar-Tensor Theory of Gravitation in a Modified Riemannian Manifold

D. K. SEN AND K. A. DUNN*
Department of Mathematics, The University of Toronto, Toronto, Ontario, Canada

(Received 15 October 1970)

A new scalar–tensor theory of gravitation is formulated in a modified Riemannian manifold in which both the scalar and tensor fields have intrinsic geometrical significance. This is in contrast to the well-known Brans–Dicke theory where the tensor field alone is geometrized and the scalar field is alien to the geometry. The static spherically symmetric solution of the exterior field equations is worked out in detail.

1. INTRODUCTION

Recently, there seems to be a renewed and serious interest in a scalar–tensor theory of gravitation, and attempts are under way to check experimentally (using the Mariner satellites) the theory suggested by Brans and Dicke,[1] as opposed to the purely tensor theory of Einstein. There is, however, a fundamental difference between the theory of Einstein and that of Brans and Dicke. With the general theory of relativity Einstein has introduced in physics a new principle: *the principle of geometrization of physics.* The general theory, in fact, succeeds in geometrizing the phenomenon of gravitation by abandoning the flat space–time of special theory and identifying the metric tensor of a Riemannian space–time with the gravitational potential. It is well-known,[2] however, that it is possible to construct meaningful theories of gravitation in a flat space–time, where gravitation, described

* Supported in part by the U.S. Atomic Energy Commission.

[1] R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967).

[2] D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968).

[3] J. Ginibre, Phys. Rev. Letters **23**, 828 (1969).

[4] R. B. Griffiths, J. Math. Phys. **10**, 1559 (1969).

[5] J. Ginibre, Commun. Math. Phys. **16**, 310 (1970).

[6] See Refs. 1 and 3, and especially the articles cited in the latter.

[7] M. E. Fisher, Phys. Rev. **180**, 594 (1969).

[8] H. S. Leff, Phys. Rev. A **7**, 2368 (1970).

[9] D. C. Mattis and M. Plischke, J. Math. Phys. **10**, 1107 (1969).

[10] R. H. T. Yeh, Phys. Letters **30A**, 297 (1969).

[11] R. B. Griffiths, C. A. Hurst, and S. Sherman, J. Math. Phys. **11**, 790 (1970).

[12] C. A. Hurst and S. Sherman, Phys. Rev. Letters **23**, 1357 (1969).

[13] R. H. T. Yeh, J. Math. Phys. **11**, 1317 (1970); Phys. Rev. Letters **23**, 1241 (1969).

[14] R. B. Griffiths, Phys. Rev. **1**, 3883 (1970).

[15] T. Asano, Phys. Rev. Letters **24**, 1409 (1970).

[16] Quantum Heisenberg models and antiferromagnetic Ising models provide further examples for which a detailed understanding is lacking.

[17] The restriction to one dimension is not necessary in what follows. It is, however, convenient with regard to both notation and physical interpretation. The results generalize to a ν-dimensional space if the one-dimensional position coordinates $(x_1, \cdots, x_N)$ are replaced by the νN Cartesian components $(x_1, \cdots, x_{\nu N})$.

[18] R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill, New York, 1968), p. 40.

[19] The author is aware of several treatments of the Ising model which do explicitly utilize the assumption of positive-definiteness for the matrix of the coupling coefficients. See A. J. F. Siegert, *Statistical Physics—1962 Brandeis Lectures*, Vol. 3 (Benjamin, New York, 1963), p. 159. The mathematical analysis is related to the present one in some ways and suggests a possible link between the two problems which deserves further investigation. See also A. J. F. Siegert and D. J. Vezetti, J. Math. Phys. **9**, 2173 (1968); C. J. Thompson, A. J. F. Siegert, and D. J. Vezetti, J. Math. Phys. **11**, 1018 (1970).

[20] See, for example, Ref. 18, p. 96.

[21] A necessary and sufficient condition for **J** to be positive definite is that $\lambda_i^2 > 0$ for $i = 1, \cdots, N$. See Ref. 18, p. 54.

[22] For a discussion of cumulants for distributions of a single variable see H. Cramér, *Mathematical Methods of Statistics* (Princeton U.P., Princeton, N.J., 1961), p. 186. The single variable case and a generalization to the bivariate and multivariate cases is contained in M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Hafner, N.Y., 1963), Chap. 3. The multivariate case is discussed further in: E. Meeron, J. Chem. Phys. **27**, 1238 (1957); R. Kubo, J. Phys. Soc. Japan **17**, 1100 (1962); I. Z. Fisher, *Statistical Theory of Liquids* (University of Chicago Press, Chicago, 1964), see the supplement by S. A. Rice and P. Gray; J. Yvon, *Correlations and Entropy in Classical Statistical Mechanics* (Pergamon, New York, 1969), Sec. 2.6.

[23] See the works of Meeron, Kubo, Fisher, and Yvon, Ref. 22.

[24] See, for example, the book of Kendall and Stuart, Ref. 22, p. 349.

[25] The conditions $J^{(k)} > 0$, $1 \leq k \leq N$, are necessary and sufficient conditions for **J** to be positive definite. See Ref. 18, p. 74.

[26] T. J. Stieltjes, Acta. Math. **9**, 385 (1886/87). A proof may be found in Ref. 18, p. 153.

[27] A. Ostrowski, Comment. Math. Helveticii **30**, 175 (1956); **10**, 69 (1937) Ostrowski does not refer to the work of Stieltjes and it is not clear that he was aware of it.

[28] An extensive study of M-matrices (in English) may be found in M. Fiedler and V. Pták, Czech. Math. J. **12**, 382 (1962).

[29] M. Fiedler and V. Pták, Num. Math. **9**, 163 (1966).

[30] For example, if the cumulant $\langle x_k x_l x_m \rangle_c$ for the Ising model were identically zero the magnetic susceptibility would be linear in the (uniform) external field, or trivially constant. This, of course, is not so in general. See Ref. 11 for a discussion of the convexity of the magnetization.

[31] For a further discussion of this point, see H. S. Leff, Am. J. Phys. **37**, 548 (1969); R. Baierlein, Am. J. Phys. **36**, 625 (1968). See also Ref. 8.

[32] H. Falk and M. Suzuki, Phys. Rev.B **1**, 3051 (1970); H. Falk, to be published. See also the book of Yvon, Ref. 22.

[33] For completeness, it is pointed out that an essentially different theorem, supplying necessary and sufficient conditions for a matrix to have a nonnegative inverse, has been given recently by P. G. Ciarlet, Lin. Alg. Appl. **1**, 139 (1968).

[34] G. W. Ford, M. Kac, and P. Mazur, J. Math. Phys. **6**, 504 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

A Scalar-Tensor Theory of Gravitation in a Modified Riemannian Manifold

D. K. SEN AND K. A. DUNN*
Department of Mathematics, The University of Toronto, Toronto, Ontario, Canada

(Received 15 October 1970)

A new scalar–tensor theory of gravitation is formulated in a modified Riemannian manifold in which both the scalar and tensor fields have intrinsic geometrical significance. This is in contrast to the well-known Brans–Dicke theory where the tensor field alone is geometrized and the scalar field is alien to the geometry. The static spherically symmetric solution of the exterior field equations is worked out in detail.

1. INTRODUCTION

Recently, there seems to be a renewed and serious interest in a scalar–tensor theory of gravitation, and attempts are under way to check experimentally (using the Mariner satellites) the theory suggested by Brans and Dicke,[1] as opposed to the purely tensor theory of Einstein. There is, however, a fundamental difference between the theory of Einstein and that of Brans and Dicke. With the general theory of relativity Einstein has introduced in physics a new principle: *the principle of geometrization of physics.* The general theory, in fact, succeeds in geometrizing the phenomenon of gravitation by abandoning the flat space–time of special theory and identifying the metric tensor of a Riemannian space–time with the gravitational potential. It is well-known,[2] however, that it is possible to construct meaningful theories of gravitation in a flat space–time, where gravitation, described

either by a tensor field or even a scalar field, does not have any geometrical significance whatsoever. In the theory of Brans and Dicke we have a mixture of these two approaches because, whereas the tensor field is identified with the metric tensor of a Riemannian geometry, the scalar field remains alien to the geometry. It would therefore seem more in the spirit of Einstein's principle of geometrization to have a scalar–tensor theory of gravitation, where both the scalar and tensor fields have more or less intrinsic geometrical significance.

In this paper we suggest the framework of such a theory. While attempting to unify gravitation and electromagnetism in a single space–time geometry, Weyl[3] showed how one can introduce a *vector* field with an intrinsic geometrical significance. Recall that the geometrical structure of a differentiable manifold is determined by (i) an affine connection characterized (in a local coordinate system x^λ) by its components $\Gamma^\mu_{\alpha\beta}$, which are defined by the change due to infinitesimal parallel transfer of a vector ξ^μ from a point $P(x^\mu)$ to $P'(x^\mu + dx^\mu)$,

$$\delta\xi^\mu = -\Gamma^\mu_{\alpha\beta}\xi^\alpha\, dx^\beta,^4 \tag{1.1}$$

and (ii) a metrical connection characterized by the metric fundamental tensor $g_{\mu\lambda}$ which is defined by the *measure* of length ξ of ξ^μ,

$$\xi = g_{\mu\lambda}\xi^\mu\xi^\lambda. \tag{1.2}$$

Riemannian geometry is characterized by the symmetry of $\Gamma^\mu_{\alpha\beta}$, that is, $\Gamma^\mu_{\alpha\beta} = \Gamma^\mu_{\beta\alpha}$ and by the condition that the length of a vector should not change under parallel transfer, i.e.,

$$\delta\xi = \delta(g_{\mu\lambda}\xi^\mu\xi^\lambda) = 0. \tag{1.3}$$

By this condition the Riemannian connection is uniquely determined by the metric tensor

$$\Gamma^\mu_{\alpha\beta} = \left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\}, \tag{1.4}$$

where on the right-hand side are Christoffel symbols of the second kind. Note the asymmetry between (1.1) and (1.3). Weyl generalized the notion of a Riemannian metrical connection in the following manner.

With each vector ξ^μ there is an associated measure of length given by the quadratic form $\xi = g_{\mu\lambda}\xi^\mu\xi^\lambda$. Two vectors ξ^μ and η^μ have therefore the same length if and only if $\xi = \eta$. Now the quadratic form is completely determined only when one specifies a nonzero scale (or gauge) factor of proportionality. At every point of the manifold one has, therefore, the possibility of a change of scale (i.e., a recalibration) or a *gauge transformation*. In a Weyl manifold it is not sufficient for the metrical connection to have a measure determination at every point; every point must also be metrically related to the surrounding neighborhood. The concept of metrical relationship is analogous to that of affine relationship. The metrical relationship is determined by specifying the change in the measure of length of a vector due to an infinitesimal parallel transfer. There exists a geodesic gauge in which there is no change in the measure of length of ξ^μ transferred parallel from $P(x^\mu)$ to $P'(x^\mu + dx^\mu)$. In an arbitrary gauge, however, ξ is assumed to change [in contrast to (1.3)] according as

$$\delta\xi = -\xi\phi_\mu\, dx^\mu, \tag{1.5}$$

where ϕ_μ is a vector function characterizing the manifold. The metrical connection of a Weyl manifold is therefore characterized by *two* independent quantities $g_{\mu\lambda}$ and ϕ_μ, relative to a reference system (= coordinate system + gauge). If one makes a gauge transformation, that is, recalibrates all lengths, $\xi \to \bar{\xi} = \lambda\xi$, where $\lambda = \lambda(x)$ in general, ϕ_μ and $g_{\mu\lambda}$ transform in the following manner:

$$g_{\mu\lambda} \to \bar{g}_{\mu\lambda} = \lambda g_{\mu\lambda}; \quad \phi_\mu \to \phi_\mu - \lambda^{-1}\lambda_{,\mu}, \tag{1.6}$$

where $\lambda_{,\mu} = \partial\lambda/\partial x^\mu$. The components of affine connection $\Gamma^\mu_{\alpha\beta}$ of a Weyl manifold are now determined through (1.1) and (1.5) by $g_{\mu\lambda}$ and ϕ_μ:

$$\Gamma^\mu_{\alpha\beta} = \left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\} + \tfrac{1}{2}(\delta^\mu_\alpha\phi_\beta + \delta^\mu_\beta\phi_\alpha - g_{\alpha\beta}\phi^\mu), \tag{1.7}$$

where $\phi^\mu = g^{\mu\lambda}\phi_\lambda$. In Weyl's unified field theory the vector field ϕ_μ is identified with the vector potential of an electromagnetic field. Weyl's hypothesis of nonintegrability of length transfer had been criticized by Einstein because it implies that the frequency of spectral lines emitted by atoms would not remain constant but would depend on their past history.

Lyra[5] has suggested a modification of Riemannian geometry which may also be considered as a modification of Weyl's geometry. In Lyra's geometry Weyl's concept of gauge, which is essentially a metrical concept, is modified by the introduction of a gauge function in the structureless manifold.

The displacement vector **PP'** between two neighboring points $P(x^\mu)$ and $P'(x^\mu + dx^\mu)$, in Lyra's geometry, has the components $\xi^\mu = x^0\, dx^\mu$, where $x^0(x)$ is a nonzero gauge function. The coordinate system x^μ together with x^0 form a *reference system* $(x^0; x^\mu)$. A general transformation of reference systems is given by

$$x^\mu \to x^{\mu'} = x^{\mu'}(x^\lambda), \quad (x^0; x^\mu) \to (x^{0'}; x^{\mu'})$$

with

$$A^{\mu'}_\mu \equiv \frac{\partial x^{\mu'}}{\partial x^\mu}, \quad \det A^{\mu'}_\mu \neq 0. \tag{1.8}$$

Under (1.8) a multicomponent tensor $\xi^{\rho_1\cdots\rho_s}_{\sigma_1\cdots\sigma_r}$ transforms as follows:

$$\xi^{\rho_1'\cdots\rho_s'}_{\sigma_1'\cdots\sigma_r'} = \lambda^{s-r} A^{\rho_1'}_{\rho_1}\cdots A^{\rho_s'}_{\rho_s} A^{\sigma_1}_{\sigma_1'}\cdots A^{\sigma_r}_{\sigma_r'}\xi^{\rho_1\cdots\rho_s}_{\sigma_1\cdots\sigma_r}, \quad (1.9)$$

where $\lambda = x^{0'}/x^0$. The factor λ^{s-r} arises as a consequence of the introduction of the gauge function.

In an affinely connected manifold the components of the affine connection $\Gamma^\mu_{\alpha\beta}$ can be considered to arise as a consequence of general coordinate transformations in the following manner.[6] Let us suppose that, in a local coordinate system x^μ, a vector ξ^μ is constant—that is, $\xi^\mu_{,\lambda} = 0$. Then, in another arbitrary local coordinate system $x^{\mu'}$, we have in the usual case ($\xi^\mu = A^\mu_{\mu'}\xi^{\mu'}$)

$$\xi^{\mu'}_{,\lambda'} + \Gamma^{\mu'}_{\nu'\lambda'}\xi^{\nu'} = 0, \quad (1.10)$$

where

$$\Gamma^{\mu'}_{\nu'\lambda'} = -A^\mu_{\nu'}A^{\mu'}_{\mu,\lambda'}, \quad A^{\mu'}_{\mu,\lambda'} = \frac{\partial A^{\mu'}_\mu}{\partial x^{\lambda'}}.$$

Another way of expressing the fact that ξ^μ is constant would be to say that (1.10) is valid in all coordinate systems, but that $\Gamma^\mu_{\nu\lambda} = 0$ in the original coordinate system x^μ. The transition from an integrable affinely connected manifold to a nonintegrable one (thereby from ordinary derivative to covariant derivative) is made by assuring that $\Gamma^\mu_{\nu\lambda} \neq 0$ in general.

Now a vector ξ^μ in Lyra's geometry transforms as

$$\xi^{\mu'} = \lambda A^{\mu'}_\mu \xi^\mu. \quad (1.11)$$

If $\xi^\mu_{,\lambda} = 0$ in the reference system $(x^0; x^\mu)$, then, in the reference system $(x^{0'}; x^{\mu'})$ we have

$$(1/x^{0'})\xi^{\mu'}_{,\lambda'} + \Gamma^{\mu'}_{\nu'\lambda'}\xi^{\nu'} - \tfrac{1}{2}\phi_{\lambda'}\xi^{\mu'} = 0, \quad (1.12)$$

where

$$\Gamma^{\mu'}_{\nu'\lambda'} = \frac{-1}{x^{0'}} A^\mu_{\nu'}A^{\mu'}_{\mu,\lambda'}; \quad \phi_{\lambda'} = \frac{1}{x^0}\frac{\partial \ln \lambda^2}{\partial x^{\lambda'}}.$$

The parallel transfer of a vector ξ^μ in Lyra's geometry is therefore given by

$$\delta\xi^\mu = -\tilde{\Gamma}^\mu_{\alpha\beta}\xi^\alpha x^0\, dx^\beta, \quad (1.13)$$

where

$$\tilde{\Gamma}^\mu_{\alpha\beta} = \Gamma^\mu_{\alpha\beta} - \tfrac{1}{2}\delta^\mu_\alpha\phi_\beta.$$

Note that the $\tilde{\Gamma}^\mu_{\alpha\beta}$ are not symmetric although $\Gamma^\mu_{\alpha\beta} = \Gamma^\mu_{\beta\alpha}$. The components of the generalized affine connection are thus characterized not only by $\Gamma^\mu_{\alpha\beta}$ but also by ϕ_β which appears as a natural consequence of the formal introduction of the gauge function in the structureless space.

The metric or the measure of length of the displacement vector $\xi^\mu = x^0\, dx^\mu$ between two points $P(x^\mu)$ and $P'(x^\mu + dx^\mu)$ is given by the absolute invariant (that is, invariant under both gauge and coordinate transformations)

$$ds^2 = g_{\mu\lambda}x^0\, dx^\mu x^0\, dx^\lambda, \quad (1.14)$$

where $g_{\mu\lambda}$ is a symmetric tensor of second rank. The parallel transfer of length in Lyra's geometry is integrable (as in Riemannian geometry) in contrast to Weyl's geometry; that is,

$$\delta(g_{\mu\lambda}\xi^\mu\xi^\lambda) = 0. \quad (1.15)$$

From (1.13) and (1.15) it follows that

$$\Gamma^\mu_{\alpha\beta} = (1/x^0)\left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\} + \tfrac{1}{2}(\delta^\mu_\alpha\phi_\beta + \delta^\mu_\beta\phi_\alpha - g_{\alpha\beta}\phi^\mu). \quad (1.16)$$

Thus apart from the factor $1/x^0$, (1.16) is identical with the components of affine connection in Weyl's geometry, not, however, $\tilde{\Gamma}^\mu_{\alpha\beta}$. In Lyra's geometry we thus have a vector field with an intrinsic geometrical significance without the inconvenience of nonintegrability of length transfer of Weyl's geometry.

In the next section we give an *invariant* formulation of Lyra's geometry. In Sec. 3 we take over the geometry as the framework for a theory of gravitation. The field equations are derived from a variational principle constructed from the scalar curvature. Variation with respect to $g_{\mu\lambda}$ furnishes the basic field equations, whereas variation with respect to ϕ_α constrains ϕ_α to be basically the gradient of the gauge function x^0. The basic field equations then contain only the metric *tensor* $g_{\mu\lambda}$ and the *scalar* x^0 and turn out to be a special case of the Brans–Dicke field equations, if x^0 is identified with the scalar ϕ in the Brans–Dicke theory. We next consider the static spherically symmetric solution in some detail.

2. A MODIFIED RIEMANNIAN MANIFOLD

Let M be a connected, second countable Hausdorff space. By a local *reference system* on M we shall mean a triple (U_i, ψ_i, f_i) where

(i) U_i is an open subset of M,

(ii) ψ_i is a homeomorphism of U_i onto an open subset of $\mathbb{R}^n$ for some integer n, and

(iii) $f_i: U_i \to \mathbb{R} - \{0\}$ is a (nonzero) *gauge* function on U_i.

We shall say M is an n-dimensional C^∞ Lyra manifold if there exists a collection of local reference systems $\{(U_i, \psi_i, f_i);\ i \in I$, an index set$\}$ on M such that

(i) $\bigcup_{i\in I} U_i = M$,

(ii) whenever $U_i \cap U_j \neq \phi$, the maps $\psi_i \circ \psi_j^{-1}$ and $\psi_j \circ \psi_i^{-1}$ are C^∞ on their domains of definition,

(iii) for each $i \in I$ the map $f_i \circ \psi_i^{-1}$ is a C^∞ function on $\psi_i(U_i)$, and

(iv) the collection is maximal with respect to (i), (ii), and (iii).

The notions of C^∞ functions and tangent space $T_m(M)$ at a point $m \in M$ can be introduced in the usual

manner[7] on a Lyra manifold M. Consider now a reference system (U_i, ψ_i, f_i) on M. If $m \in U_i$, let $\psi_i(m) = \{x^1(m), \cdots, x^n(m)\} = x$ be the local coordinates of m. In local coordinates let the gauge function $f_i \circ \psi_i^{-1}: \psi_i(U_i) \to \mathbb{R} - \{0\}$ be denoted by $x^0: x \mapsto x^0(x)$, which, together with the coordinate map, form (in local coordinates) the reference system $(x^0; x^\mu)$, $\mu = 1, \cdots, n$. A local reference system induces a natural basis

$$\left\{\tilde{e}_\mu(m) = [x^0(x)]^{-1}\left(\frac{\partial}{\partial x^\mu}\right)_x\right\} \quad \text{in } T_m(M),$$

so that a tangent vector $\zeta \in T_m(M)$ can be written as $\zeta = \alpha^\mu \tilde{e}_\mu(m)$, $\alpha^\mu \in \mathbb{R}$.

Let $B(M) = \{(m, \zeta) \mid m \in M, \zeta \in T_m(M)\}$ and

$$\pi: B(M) \to M$$

be the projection map defined by $\pi(m, \zeta) = m$. The mapping $\tilde{\psi}_i: (m, \zeta) \mapsto (x^1, \cdots, x^n; \alpha^1, \cdots, \alpha^n)$ is then a homeomorphism of $\pi^{-1}(U_i) \subset B(M)$ onto an open subset of $\mathbb{R}^{2n}$. Let (U_j, ψ_j, f_j) be another local reference system such that $U_i \cap U_j \neq \phi$ and $(m, \zeta) \in \pi^{-1}(U_i \cap U_j)$. If $\psi_j(m) = \{x^{1'}(m), \cdots, x^{n'}(m)\} = x'$, $x^{0'} = f_j \circ \psi_j^{-1}$, and

$$\tilde{\psi}_j(m, \zeta) = (x^{1'}, \cdots, x^{n'}; \alpha^{1'}, \cdots, \alpha^{n'}),$$

where $\zeta = \alpha^{\mu'} \tilde{e}_{\mu'}(m)$, it follows that

$$\alpha^{\mu'} = \left(\frac{x^0(x)}{x^{0'}(x')}\right)^{-1} A^{\mu'}_\mu(m)\alpha^\mu, \quad A^{\mu'}_\mu(m) = \frac{\partial x^{\mu'}}{\partial x^\mu}\bigg|_m, \tag{2.1}$$

and the transformation

$$(x^1, \cdots, x^n; \alpha^1, \cdots, \alpha^n) \mapsto (x^{1'}, \cdots, x^{n'}; \alpha^{1'}, \cdots, \alpha^{n'})$$

is C^∞. $B(M)$ can thus be made into a $2n$-dimensional C^∞ manifold, with coordinate neighborhoods

$$\{[\pi^{-1}(U_i), \tilde{\psi}_i]\}.$$

We shall call $B(M)$ the Lyra tangent bundle of the Lyra manifold M. A C^∞ vector field on an open subset U of M is a C^∞ map $\chi: U \to B(M)$ such that $\pi \circ \chi$ is the identity map on U. In a local reference system χ can be expressed as $\chi = \chi^\mu(x)\tilde{e}_\mu$, where $\tilde{e}_\mu = (x^0)^{-1}x^{0-1}\partial/\partial x^\mu$ and the components $\chi^\mu(x)$ transform as

$$\chi^{\mu'} = \lambda A^{\mu'}_\mu \chi^\mu, \quad \lambda = x^{0'}/x^0. \tag{2.2}$$

Tensor fields can be defined in an analogous manner, and it can thus be seen that, under transformation of local reference systems (1.8), the components of a multicomponent tensor field transform as in (1.9).

We shall denote by $T(M)$ the linear space of all C^∞ vector fields and by $C^\infty(M)$ the set of all C^∞ functions on M. An affine connection on a Lyra manifold can be defined in the normal way, that is, as a bilinear mapping $\nabla: T(M) \times T(M) \to T(M)$, such that, for every $f \in C^\infty(M)$ and arbitrary X, Y, $Z \in T(M)$, the following two conditions are satisfied [writing $\nabla_X Y$ for $\nabla(X, Y)$]:

$$\begin{aligned} &\text{(i)} \quad \nabla_{fX} Y = f\nabla_X Y, \\ &\text{(ii)} \quad \nabla_X fY = f\nabla_X Y + X(f)Y. \end{aligned} \tag{2.3}$$

In a local reference system (U_i, ψ_i, f_i) the connection is specified by C^∞ functions $\tilde{\Gamma}^\mu_{\alpha\beta}$ where

$$\nabla_{\tilde{e}_\beta} \tilde{e}_\alpha = \tilde{\Gamma}^\mu_{\alpha\beta} \tilde{e}_\mu. \tag{2.4}$$

In another local reference system we have

$$\nabla_{\tilde{e}_{\beta'}} \tilde{e}_{\alpha'} = \nabla_{\lambda^{-1}A_{\beta'}{}^\beta \tilde{e}_\beta}(\lambda^{-1}A^\alpha_{\alpha'}\tilde{e}_\alpha) = \tilde{\Gamma}^{\mu'}_{\alpha'\beta'}\tilde{e}_{\mu'}.$$

The properties (2.3) then imply the following transformation law for the $\tilde{\Gamma}^\mu_{\alpha\beta}$;

$$\tilde{\Gamma}^{\mu'}_{\alpha'\beta'} = \lambda^{-1}A^{\mu'}_\mu A^\alpha_{\alpha'}A^\beta_{\beta'}\tilde{\Gamma}^\mu_{\alpha\beta} + (x^0)^{-1}A^{\mu'}_\lambda A^\lambda_{\alpha',\beta'} - \tfrac{1}{2}(x^{0'})^{-1}\delta^{\mu'}_{\alpha'}A^\lambda_{\beta'}(\ln \lambda^2)_{,\lambda}.$$

If we set

$$\tilde{\Gamma}^\mu_{\alpha\beta} = \Gamma^\mu_{\alpha\beta} - \tfrac{1}{2}\delta^\mu_\alpha \phi_\beta, \tag{2.5}$$

we obtain the following transformation laws:

$$\begin{aligned} \Gamma^{\mu'}_{\alpha'\beta'} &= \lambda^{-1}(A^{\mu'}_\mu A^\alpha_{\alpha'}A^\beta_{\beta'}\Gamma^\mu_{\alpha\beta} + (x^0)^{-1}A^{\mu'}_\lambda A^\lambda_{\alpha',\beta'}), \\ \phi_{\beta'} &= \lambda^{-1}A^\beta_{\beta'}\{\phi_\beta + (x^0)^{-1}(\ln \lambda^2)_{,\beta}\}, \end{aligned} \tag{2.6}$$

In a local reference system the connection is thus characterized by the two sets of C^∞ functions $\Gamma^\mu_{\alpha\beta}$ and ϕ_β, which transform as in (2.6).

Given a connection ∇ on a Lyra manifold, we can define a pseudoconnection $\bar{\nabla}$ as follows. Let $\bar{T}(M)$ denote the set of all vector-valued functions on M whose components transform as a vector field under coordinate transformations only (but not necessarily under gauge transformations). Then $\bar{\nabla}$ is a bilinear mapping $\bar{\nabla}: T(M) \times T(M) \to \bar{T}(M)$ defined in the following way. If in a local reference system ∇ is specified by $\Gamma^\mu_{\alpha\beta} - \frac{1}{2}\delta^\mu_\alpha\phi_\beta$, then $\bar{\nabla}_{\tilde{e}_\beta}\tilde{e}_\alpha = \Gamma^\mu_{\alpha\beta}\tilde{e}_\mu$. The torsion of the pseudoconnection $\bar{\nabla}$ is a mapping $\mathrm{Tor}_{\bar{\nabla}}: T(M) \times T(M) \to T(M)$ given by

$$\mathrm{Tor}_{\bar{\nabla}}(X, Y) = \bar{\nabla}_X Y - \bar{\nabla}_Y X - [X, Y], \tag{2.7}$$

where $[X, Y]$ is the Lie Bracket of X and Y. By a Lyra connection on M we shall mean a connection ∇ with $\mathrm{Tor}_{\bar{\nabla}} \equiv 0$ which implies that $\Gamma^\mu_{\alpha\beta} = \Gamma^\mu_{\beta\alpha}$. Note that this does *not* mean that the original connection ∇ is symmetric.

An autoparallel (or geodesic) of a connection ∇ is a curve $s \to x(s)$ whose tangent vector is transferred parallel to itself. In a local reference system it is a curve given by $x^\mu(s)$, with tangent vector components

$\xi^\mu = x^0(dx^\mu/ds)$, satisfying the differential equation

$$x^0 \frac{d^2x^\mu}{ds^2} + (x^0)^2\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} - \tfrac{1}{2}(x^0)^2(\phi_\alpha - \overset{\circ}{\phi}_\alpha)\frac{dx^\alpha}{ds}\frac{dx^\mu}{ds} = 0, \quad (2.8)$$

where

$$\overset{\circ}{\phi}_\alpha = (x^0)^{-1}[\ln (x^0)^2]_{,\alpha}. \quad (2.9)$$

The covariant derivative of a mixed tensor field component $\xi^{\rho_1\cdots\rho_r}_{\sigma_1\cdots\sigma_s}$ in a local reference system is given by the corresponding formula in Riemannian geometry except that $\Gamma^\mu_{\alpha\beta}$ is replaced by $\tilde{\Gamma}^\mu_{\alpha\beta}$ and $\partial/\partial x^\alpha$ by $(x^0)^{-1}\partial/\partial x^\alpha$, that is

$$\nabla_\alpha \xi^{\rho_1\cdots\rho_r}_{\sigma_1\cdots\sigma_s} = (x^0)^{-1}\xi^{\rho_1\cdots\rho_r}_{\sigma_1\cdots\sigma_s,\alpha} = \sum_{\lambda=1}^{r}\Gamma^{\rho_\lambda}_{\nu\alpha}\xi^{\cdots\nu\cdots}_{\cdots\cdots} - \sum_{\lambda=1}^{s}\Gamma^{\nu}_{\sigma_\lambda\alpha}\xi^{\cdots\cdots}_{\cdots\nu\cdots} - \tfrac{1}{2}(r-s)\phi_\alpha\xi^{\cdots}_{\cdots}. \quad (2.10)$$

The curvature tensor of a Lyra connection ∇ can be regarded as a mapping

$$K: T(M) \times T(M) \times T(M) \to T(M)$$

given by

$$(X, Y, Z) \to K(X, Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z. \quad (2.11)$$

In a local reference system if we set

$$K(\tilde{e}_\alpha, \tilde{e}_\beta)\tilde{e}_\lambda = K^{\beta\mu}_{\lambda\alpha\beta}\tilde{e}_\mu, \quad (2.12)$$

the components of the curvature tensor are given by

$$K^\mu_{\lambda\alpha\beta} = (x^0)^{-2}\left(\frac{\partial(x^0\tilde{\Gamma}^\mu_{\lambda\beta})}{\partial x^\alpha} - \frac{\partial(x^0\tilde{\Gamma}^\mu_{\lambda\alpha})}{\partial x^\beta}\right) + \tilde{\Gamma}^\mu_{\rho\alpha}\tilde{\Gamma}^\rho_{\lambda\beta} - \tilde{\Gamma}^\mu_{\rho\beta}\tilde{\Gamma}^\rho_{\lambda\alpha}. \quad (2.13)$$

The curvature tensor K can be written as the sum of two tensors $K = R^* + \Phi$, where

$$R^*(X, Y)Z = \bar{\nabla}_X\bar{\nabla}_Y Z - \bar{\nabla}_Y\bar{\nabla}_X Z - \bar{\nabla}_{[X,Y]}Z \quad (2.14)$$

and

$$\Phi(X, Y)Z = K(X, Y)Z - R^*(X, Y)Z. \quad (2.15)$$

The components of R^* and Φ in a local reference system are

$$R^{*\mu}_{\lambda\alpha\beta} = (x^0)^{-1}[\Gamma^\mu_{\lambda\beta,\alpha} - \Gamma^\mu_{\lambda\alpha,\beta}] + \Gamma^\mu_{\rho\alpha}\Gamma^\rho_{\lambda\beta} - \Gamma^\mu_{\rho\beta}\Gamma^\rho_{\lambda\alpha} - \tfrac{1}{2}(\overset{\circ}{\phi}_\alpha\Gamma^\mu_{\lambda\beta} - \overset{\circ}{\phi}_\beta\Gamma^\mu_{\lambda\alpha}), \quad (2.16)$$

$$\Phi^\mu_{\lambda\alpha\beta} = \tfrac{1}{2}\delta^\mu_\lambda\{(x^0)^{-1}[\phi_{\alpha,\beta} - \phi_{\beta,\alpha}] - \tfrac{1}{2}(\overset{\circ}{\phi}_\alpha\phi_\beta - \overset{\circ}{\phi}_\beta\phi_\alpha)\}, \quad (2.17)$$

so that

$$K^\mu_{\lambda\alpha\beta} = R^{*\mu}_{\lambda\alpha\beta} + \Phi^\mu_{\lambda\alpha\beta}. \quad (2.18)$$

(2.16) and (2.17) are, for $n > 2$, uniquely determined by the curvature tensor (2.13). $R^{*\mu}_{\lambda\alpha\beta}$ and $\Phi^\mu_{\lambda\alpha\beta}$ have the following symmetry properties:

$$R^{*\mu}_{\lambda\alpha\beta} + R^{*\mu}_{\lambda\beta\alpha} = 0, \quad \Phi_{\alpha\beta} + \Phi_{\beta\alpha} = 0, \quad (2.19)$$

where we have set $\Phi^\mu_{\lambda\alpha\beta} = \frac{1}{2}\delta^\mu_\lambda\Phi_{\alpha\beta}$. $K^\mu_{\lambda\alpha\beta}$ also satisfies the usual Bianchi identities

$$\nabla_\alpha K^\mu_{\lambda\beta\gamma} + \nabla_\beta K^\mu_{\lambda\gamma\alpha} + \nabla_\gamma K^\mu_{\lambda\alpha\beta} = 0. \quad (2.20)$$

A metric structure on M is specified by a symmetric tensor field of type (0, 2), whose components in a local reference system are given by

$$g_{\mu\lambda} = g(\tilde{e}_\mu, \tilde{e}_\lambda). \quad (2.21)$$

The usual *metric* is therefore

$$ds^2 = (x^0)^2 g_{\mu\lambda}\, dx^\mu\, dx^\lambda. \quad (2.22)$$

As in Riemannian geometry a "*Riemannian*" Lyra connection ∇ is uniquely determined by the metric field g if one assumes

$$Z[g(X, Y)] = g(\nabla_Z X, Y) + g(X, \nabla_Z Y) \quad (2.23)$$

for every X, Y, $Z \in T(M)$. This, together with the condition $\mathrm{Tor}_\nabla \equiv 0$, determines a unique Lyra connection in terms of the metric tensor. In a local reference system (2.23) and $\mathrm{Tor}_\nabla \equiv 0$ imply (1.16), i.e.,

$$\Gamma^\mu_{\alpha\beta} = (x^0)^{-1}\begin{Bmatrix}\mu\\ \alpha\beta\end{Bmatrix} + \tfrac{1}{2}(\delta^\mu_\alpha\phi_\beta + \delta^\mu_\beta\phi_\alpha - g_{\alpha\beta}\phi^\mu). \quad (2.24)$$

Lyra's geometry is therefore characterized by the two fundamental entities $g_{\mu\lambda}$ and ϕ_α, the latter arising naturally as a result of the formal introduction of gauge in the geometrically structureless C^∞ manifold.

A geodesic of the metrical connection is an extremal curve $x^\mu = x^\mu(t)$ given by

$$\delta\int ds = \delta\int\left((x^0)^2 g_{\mu\lambda}\frac{dx^\mu}{dt}\frac{dx^\lambda}{dt}\right)^{\frac{1}{2}} dt = 0. \quad (2.25)$$

The Euler–Lagrange equations for the geodesics turn out to be[8] (taking $t = s$)

$$\frac{d^2x^\mu}{ds^2} + \begin{Bmatrix}\mu\\ \alpha\beta\end{Bmatrix}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} + \frac{x^0}{2}(\delta^\mu_\alpha\overset{\circ}{\phi}_\beta + \delta^\mu_\beta\overset{\circ}{\phi}_\alpha - g_{\alpha\beta}\overset{\circ}{\phi}{}^\mu)\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = 0. \quad (2.26)$$

On the other hand, in view of (2.24), Eq. (2.8) for an autoparallel curve becomes

$$x^0\frac{d^2x^\mu}{ds^2} + \left[(x^0)^{-1}\begin{Bmatrix}\mu\\ \alpha\beta\end{Bmatrix} + \tfrac{1}{2}(\delta^\mu_\alpha\phi_\beta + \delta^\mu_\beta\phi_\alpha - g_{\alpha\beta}\phi^\mu)\right](x^0)^2\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = \tfrac{1}{2}(\phi_\alpha - \overset{\circ}{\phi}_\alpha)(x^0)^2\frac{dx^\alpha}{ds}\frac{dx^\mu}{ds}. \quad (2.27)$$

Thus in Lyra's geometry, in contrast to the situation in Riemannian geometry, the autoparallels associated with the affine connection do not coincide with the geodesics which arise from the metric. This is, by the

way, also the case in Weyl's geometry. A comparison of (2.26) with (2.27) shows that a sufficient condition that the two types of curves be the same is

$$\phi_\alpha = \overset{\circ}{\phi}_\alpha. \tag{2.28}$$

The above condition is invariant under gauge transformations because $\overset{\circ}{\phi}_\alpha$ transforms exactly as ϕ_α when $x^0 \to x^{0'}$.

A Lyra manifold where (2.28) is valid, is therefore characterized by $g_{\mu\lambda}$ and the scalar x^0.

If one substitutes (2.24) in the expression (2.13) for the components of the curvature tensor and sets $\mu = \beta$, one obtains for the contracted curvature tensor ($n = 4$)

$$\begin{aligned} K_{\lambda\alpha} &= K^{\beta}_{\lambda\alpha\beta} = K_{(\lambda\alpha)} + K_{[\lambda\alpha]}\,, \\ K_{(\lambda\alpha)} &= K_{(\alpha\lambda)} \\ &= (x^0)^{-2}R_{\lambda\alpha} + \tfrac{1}{2}(x^0)^{-1}(\phi_{\lambda;\alpha} + \phi_{\alpha;\lambda}) \\ &\quad + \tfrac{1}{2}(x^0)^{-1}g_{\lambda\alpha}\phi^{\beta}_{;\beta} - \tfrac{1}{2}\phi_\lambda\phi_\alpha + \tfrac{1}{4}(\overset{\circ}{\phi}_\lambda\phi_\alpha + \overset{\circ}{\phi}_\alpha\phi_\lambda) \\ &\quad + \tfrac{1}{2}g_{\lambda\alpha}\phi_\beta\phi^\beta + \tfrac{1}{4}g_{\lambda\alpha}\overset{\circ}{\phi}_\beta\phi^\beta\,, \qquad (2.29) \\ K_{[\lambda\alpha]} &= -K_{[\alpha\lambda]} \\ &= \tfrac{1}{2}(x^0)^{-1}(\phi_{\lambda,\alpha} - \phi_{\alpha,\lambda}) - \tfrac{1}{4}(\overset{\circ}{\phi}_\lambda\phi_\alpha - \overset{\circ}{\phi}_\alpha\phi_\lambda), \end{aligned}$$

where $R_{\lambda\alpha}$ is the Ricci tensor of Riemannian geometry and $\phi_{\lambda;\alpha}$ the covariant derivative of ϕ_λ with respect to the Christoffel symbols $\{^{\mu}_{\alpha\beta}\}$. For $n = 4$, the curvature scalar is given by

$$\begin{aligned} K = g^{\lambda\alpha}K_{\lambda\alpha} &= (x^0)^{-2}R + 3(x^0)^{-1}\phi^{\alpha}_{;\alpha} \\ &\quad + \tfrac{3}{2}\phi_\alpha\phi^\alpha + \tfrac{3}{2}\overset{\circ}{\phi}_\alpha\phi^\alpha, \qquad (2.30) \end{aligned}$$

where R is the Riemannian curvature scalar. If we now choose the normal gauge $x^0 = 1$, then (2.30) becomes

$$K = R + 3\phi^{\alpha}_{;\alpha} + \tfrac{3}{2}\phi_\alpha\phi^\alpha, \tag{2.31}$$

which is identical with the corresponding curvature scalar of Weyl's geometry.

3. A SCALAR-TENSOR THEORY OF GRAVITATION

We now consider a four-dimensional Lyra manifold, endowed with a hyperbolic (i.e., with signature $+, -, -, -$) metric tensor, as the framework for a scalar–tensor theory of gravitation. We refer to Dicke's paper[1] for various arguments for a scalar–tensor theory of gravitation.

The exterior (vacuum) field equations are to be derived from a variational principle

$$\delta\int \mathfrak{W}\, dx^1 \cdots dx^4 = 0, \tag{3.1}$$

where the integrand is an *absolute* invariant. The simplest variational principle involving the curvature tensor is

$$\delta\int K(-g)^{\frac{1}{2}}x^0\, dx^1 \cdots x^0\, dx^4 = 0, \tag{3.2}$$

where the gauge factors make the integrand an absolute invariant under both coordinate and gauge transformations. Other possible invariants are

$$\mathfrak{W} = W(-g)^{\frac{1}{2}}(x^0)^4.$$

$$W = \begin{cases} g^{\lambda\alpha}K_{(\lambda\alpha)}, \\ K_{[\lambda\alpha]}K^{[\lambda\alpha]}, \\ aK + bg^{\lambda\alpha}K_{(\lambda\alpha)} + cK_{[\lambda\alpha]}K^{[\lambda\alpha]}, \\ \qquad\qquad a, b, c \text{ constants}. \end{cases} \tag{3.3}$$

In this paper we shall consider only (3.2) and heuristically examine its consequences. If we substitute (2.30) in (3.2) and consider independent variations of $g_{\alpha\beta}$ and ϕ_α, we get

$$\begin{aligned} &\int\{(x^0)^2\delta[R(-g)^{\frac{1}{2}}] + 3(x^0)^3\delta[\phi^{\alpha}_{;\alpha}(-g)^{\frac{1}{2}}] \\ &\qquad + \tfrac{3}{2}(x^0)^4\delta[\phi_\alpha\phi^\alpha(-g)^{\frac{1}{2}}] \\ &\qquad + \tfrac{3}{2}(x^0)^4\delta[\overset{\circ}{\phi}_\alpha\phi^\alpha(-g)^{\frac{1}{2}}]\}\, dx^1 \cdots dx^4 \\ &= \int\{-[R^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}R + \tfrac{3}{2}(x^0)^2\phi^\alpha\phi^\beta \\ &\qquad - \tfrac{3}{4}(x^0)^2g^{\alpha\beta}\phi_\nu\phi^\nu - \tfrac{3}{4}(x^0)^2g^{\alpha\beta}\overset{\circ}{\phi}_\nu\phi^\nu \\ &\qquad + \tfrac{3}{2}(x^0)^2\overset{\circ}{\phi}{}^\alpha\phi^\beta](x^0)^2(-g)^{\frac{1}{2}}\delta g_{\alpha\beta} \\ &\qquad + [3\phi^\alpha + \tfrac{3}{2}\overset{\circ}{\phi}{}^\alpha](x^0)^4(-g)^{\frac{1}{2}}\delta\phi_\alpha\}\, dx^1 \cdots dx^4 \\ &= 0. \end{aligned} \tag{3.4}$$

The exterior field equations are therefore

$$\begin{aligned} &R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R + \tfrac{3}{2}(x^0)^2\phi_\alpha\phi_\beta - \tfrac{3}{4}(x^0)^2g_{\alpha\beta}\phi_\nu\phi^\nu \\ &\qquad - \tfrac{3}{4}(x^0)^2g_{\alpha\beta}\overset{\circ}{\phi}_\nu\phi^\nu + \tfrac{3}{2}(x^0)^2\overset{\circ}{\phi}_\alpha\phi_\beta = 0, \qquad (3.5) \\ &\qquad 3\phi_\alpha + \tfrac{3}{2}\overset{\circ}{\phi}_\alpha = 0. \qquad (3.6) \end{aligned}$$

The two sets of equations can be combined into the following single set of equations:

$$\begin{aligned} &R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R - \omega(x^0)^{-2}x^0_{,\alpha}x^0_{,\beta} \\ &\qquad + \tfrac{1}{2}\omega(x^0)^{-2}g_{\alpha\beta}x^0_{,\nu}x^{0,\nu} = 0, \end{aligned} \tag{3.7}$$

in view of (2.9), where we have set $\omega = \frac{3}{2}$.

The interior (matter) field equations are then simply

$$\begin{aligned} &R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R - \omega(x^0)^{-2}x^0_{,\alpha}x^0_{,\beta} + \tfrac{1}{2}\omega(x^0)^{-2}g_{\alpha\beta}x^0_{,\nu}x^{0,\nu} \\ &\qquad = -[8\pi G/(x^0)^2]T_{\alpha\beta}. \end{aligned} \tag{3.8}$$

If we compare (3.8) with the interior field equations of Brans and Dicke[1,9]

$$\begin{aligned} &R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R - \omega\phi^{-2}\phi_{,\alpha}\phi_{,\beta} + \tfrac{1}{2}\omega\phi^{-2}g_{\alpha\beta}\phi_{,\nu}\phi^{,\nu} \\ &\qquad = -(8\pi/c^2\phi)T_{\alpha\beta} + \phi^{-1}(\phi_{,\alpha;\beta} - g_{\alpha\beta}\Box\phi), \end{aligned} \tag{3.9}$$

we see that our theory can be regarded as a special case of (3.9), where the Brans–Dicke scalar function ϕ

satisfies

$$\phi_{,\alpha;\beta} - g_{\alpha\beta}\Box\phi = 0, \tag{3.10}$$

and the Brans–Dicke constant $\omega = \frac{3}{2}$. However, as we shall see later, the geodesic equations of motion are different in our case, so that the difference between our theory and the Brans–Dicke theory is not so straightforward.

We next consider the static spherically symmetric solution of the exterior field equations (3.7). So let

$$x^0 = x^0(r)$$

and

$$g_{\alpha\beta} = \begin{pmatrix} e^{\nu} & 0 & 0 & 0 \\ 0 & -e^{\lambda} & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2\sin^2\theta \end{pmatrix}, \tag{3.11}$$

where

$$\lambda = \lambda(r), \quad \nu = \nu(r)$$

in a "polar" coordinate system (t, r, θ, ϕ). Equations (3.7) then reduce to the following set:

$$-\nu'/r - (1 - e^{\lambda})/r^2 - (\omega/2)f(r) = 0, \tag{3.12}$$

$$-r^2e^{-\lambda}[\nu''/2 - \lambda'\nu'/4 + \nu'^2/4 + (\nu' - \lambda')/2r] + (\omega/2)r^2e^{-\lambda}f(r) = 0, \tag{3.13}$$

$$e^{\nu-\lambda}[-\lambda'/r + (1 - e^{\lambda})/r^2] - (\omega/2)e^{\nu-\lambda}f(r) = 0, \tag{3.14}$$

where we have put

$$f(r) = [x^{0\prime}(r)/x^0(r)]^2. \tag{3.15}$$

From (3.12) and (3.14) we have

$$\nu' + \lambda' + \omega r f(r) = 0, \tag{3.16}$$

or

$$\lambda' = -\nu' - \omega r f(r). \tag{3.17}$$

Substituting (3.17) in (3.13), we get

$$r\nu'' + 2\nu' + r\nu'^2 + (\omega/2)\nu' r^2 f(r) = 0, \tag{3.18}$$

which can be solved for ν in terms of r and $f(r)$. Setting $\mu = e^{\nu}$, we get from (3.18)

$$\mu'' + \mu' g(r) = 0, \tag{3.19}$$

where

$$g(r) = -[2/r + (\omega/2)rf(r)]. \tag{3.20}$$

Then

$$\mu = e^{\nu} = D + C\phi(r), \tag{3.21}$$

where

$$\phi(r) = \int \exp\left(\int g(r)\,dr\right) dr \tag{3.22}$$

and D and C are integration constants.

On the other hand, integration of (3.17) gives a relationship between ν and λ,

$$e^{\lambda} = A \exp\left(-\int \omega r f(r)\,dr\right) e^{-\nu}, \tag{3.23}$$

where A is another integration constant. Equations (3.20) and (3.22) enable us to express everything in terms of $\phi(r)$ and its derivatives, as follows:

$$\exp\left(-\int \omega r f(r)\,dr\right) = r^4\phi'(r), \tag{3.24}$$

$$(\omega/2)f(r) = -2/r^2 - \phi''(r)/r\phi'(r). \tag{3.25}$$

Thus

$$e^{\lambda} = Ar^4[\phi'(r)]^2/[D + C\phi(r)]. \tag{3.26}$$

Finally, substituting (3.21), (3.25), and (3.26) in (3.12), we obtain the following nonlinear differential equation for $\phi(r)$:

$$(D + C\phi)(1 + r\phi''/\phi') + Ar^4\phi'^2 - Cr\phi' = 0$$

or

$$(D + C\phi)(r\phi')' + Ar^4\phi'^3 - Cr\phi'^2 = 0. \tag{3.27}$$

Apart from the special Schwarzschild solution, we have been unable to obtain a solution of (3.27) in a closed form. However, a formal series solution of the form $\phi = \sum_{n=0}^{\infty} a_n r^{-n}$ exists, although the question of convergence of such a solution remains open. Details of the solution and recurrence formula for a_n are relegated to the Appendix.

The general solution of the form $\phi = \sum_{n=0}^{\infty} a_n r^{-n}$ is as follows:

$$\begin{gathered} a_0 \text{ arbitrary}, \\ a_1 \neq 0, \quad Aa_1^2 = D + Ca_0, \\ a_2 = 0, \\ a_3 \text{ arbitrary}, \end{gathered} \tag{3.28}$$

a_n, $n > 3$, are determined in terms of a_0 and a_3 by a well-defined recurrence relation.

If we now impose the usual boundary condition at infinity, i.e., $e^{\nu} \to 1$ as $r \to \infty$, we get

$$D + Ca_0 = 1. \tag{3.29}$$

Then, also

$$e^{\lambda} = \frac{Ar^4(a_1^2/r^4 + \cdots + \cdots)}{D + C(a_0 + a_1/r + \cdots + \cdots)} \to 1,$$

as $r \to \infty$. A special case of (3.28) is

$$\begin{aligned} &a_0 \text{ arbitrary}, \\ &a_1 = \pm[(D + Ca_0/A)]^{\frac{1}{2}}, \\ &a_n = 0, \quad n > 1, \end{aligned} \tag{3.30}$$

which together with (3.29) gives the Schwarzschild solution

$$\begin{aligned} e^{\nu} &= 1 \pm C/\sqrt{Ar}, \\ e^{\lambda} &= 1/(1 \pm C/\sqrt{Ar}). \end{aligned} \tag{3.31}$$

From (3.15) and (3.25) we get for the gauge function

$$x^0(r) = \text{const} \times \exp\int\{-[(4/\omega r^2) + (2/\omega r)(\phi''/\phi')]\}^{\frac{1}{2}}\,dr. \quad (3.32)$$

In the Schwarzschild case, thus, $x^0 = \text{constant}$, since $\phi''/\phi' = -2/r$. This is, of course, also clear from the field equations (3.12)–(3.14).

It is interesting to note that e^ν does not contain any r^{-2} term. The effect of r^{-n}, $n > 2$, terms on gravitational red shift (which is determined solely by e^ν) experiments is, of course, too small to be detected at present.

To discuss both the problem of perihelion precession and bending of light rays, it is necessary to have an equation of motion of test bodies. As we have seen in Sec. 2, the geodesics of metrical connection are not identical with the autoparallels of affine connection. In view of (3.6), however, the two are practically the same, i.e., they are given by

$$\frac{d^2x^\mu}{ds^2} + \begin{Bmatrix}\mu\\ \alpha\beta\end{Bmatrix}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} + \beta x^0 \overset{\circ}{\phi}_\alpha \frac{dx^\mu}{ds}\frac{dx^\alpha}{ds} - \alpha\frac{x^0}{2} g_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds}\overset{\circ}{\phi}^\mu = 0, \quad (3.33)$$

where for geodesics $\alpha = \beta = 1$ and for autoparallels $\beta = \frac{1}{4}$, $\alpha = -\frac{1}{2}$. Long ago Eddington[10] considered the effect of non-Schwarzschild terms in the usual geodesic equations of general relativity. The combined effect of non-Schwarzschild terms in the metric and the extra terms in (3.33) are being investigated.

In conclusion we would like to point out that the variational principle (3.2) and the consequent field equations are probably too simple to provide an alternative to either the general theory of relativity or Brans–Dicke theory, which is significant from an experimental point of view. In this paper we have strived merely to provide a framework for an alternative theory should new experimental results make the general theory unsatisfactory.

ACKNOWLEDGMENTS

This work was supported in part by the National Research Council of Canada under an operating research grant. We are grateful to W. D. Halford for some useful ideas.

APPENDIX

Consider a series solution of (3.27) of the form

$$\phi = \sum_{n=0}^{\infty} a_n r^{-n}. \quad (A1)$$

Then

$$\phi' = -\sum_{n=2}^{\infty}(n-1)a_{n-1}r^{-n}, \quad (A2)$$

$$(r\phi')' = \sum_{n=2}^{\infty}(n-1)^2 a_{n-1}r^{-n}, \quad (A3)$$

$$(D + C\phi)(r\phi')' = \sum_{n=2}^{\infty} D(n-1)^2 a_{n-1}r^{-n} + \sum_{n=2}^{\infty}\left(\sum_{k=0}^{n-2} Ca_k(n-k-1)^2 a_{n-k-1}\right)r^{-n}, \quad (A4)$$

$$Cr\phi'^2 = \sum_{n=3}^{\infty} C\left(\sum_{k=2}^{n-1}(k-1)a_{k-1}(n-k)a_{n-k}\right)r^{-n}, \quad (A5)$$

$$\phi'^3 = -\left(\sum_{n=2}^{\infty}(n-1)a_{n-1}r^{-n}\right) \times \left[\sum_{n=4}^{\infty}\left(\sum_{k=2}^{n-2}(k-1)(n-k-1)a_{k-1}a_{n-k-1}\right)r^{-n}\right], \quad (A6)$$

$$Ar^4\phi'^3 = -\sum_{n=2}^{\infty} A\left[\sum_{l=2}^{n}[(l-1)a_{l-1}] \times \left(\sum_{k=2}^{n-l+2}(k-1)(n-l-k+3)a_{k-1}a_{n-l-k+3}\right)\right]r^{-n}. \quad (A7)$$

Substituting (A4), (A5), and (A7) in (3.27), we get, on equating coefficients of

r^{-2}: $a_1(D + Ca_0) - Aa_1^3 = 0$,
or a_0 arbitrary, $a_1 \neq 0$,
$Aa_1^2 = D + Ca_0$;

r^{-3}: $a_2[-2(D + Ca_0)] = 0$,
or $a_2 = 0$;

r^{-4}: $a_3[9(D + Ca_0) - 9Aa_1^2] = 0$,
or a_3 arbitrary;

r^{-n}, $n > 4$; the following recurrence relation for a_n $(n \geq 4)$:

$$a_{n-1}[(D + Ca_0)(n-1)(n-4)] - Aa_1\sum_{k=3}^{n-1}(k-1)(n-k+1)a_{k-1}a_{n-k+1} - A\sum_{l=3}^{n-1}[(l-1)a_{l-1}] \times \left(\sum_{k=2}^{n-l+2}(k-1)(n-l-k+3)a_{k-1}a_{n-l-k+3}\right) - C\sum_{l=2}^{n-1}a_{n-l}a_{l-1}(n-l)(2l-n-1) = 0. \quad (A8)$$

* National Research Council of Canada 1967 Science Predoctoral Fellow.

[1] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961); R. H. Dicke, "Experimental Relativity," in *Relativity, Groups and Topology*, C. M. DeWitt and B. S. DeWitt, Eds. (Gordon and Breach, New York, 1964).

[2] A. N. Whitehead, *The Principle of Relativity* (Cambridge U.P., Cambridge, 1922); G. D. Birkhoff, Proc. Natl. Acad. Sci. (U.S.) **29**, 231 (1943); A. Papapetrou, Proc. Roy. Irish Acad. **2**, 11 (1948).

[3] H. Weyl, Sitzber. Preuss. Akad. Wiss., 465 (1918).

[4] The usual summation convention is used throughout.

[5] G. Lyra, Math. Z. **54**, 52 (1951); E. Scheibe, Math. Z. **57**, 65 (1952); D. K. Sen, Z. Physik **149**, 311 (1957).

[6] E. Schrödinger, *Space-Time Structure* (Cambridge U.P., Cambridge, 1954), p. 27.

[7] See, for example, N. Hicks, *Notes on Differential Geometry* (Van Nostrand, Princeton, N.J., 1965).

[8] D. K. Sen, Can. Math. Bull. **3**, 255 (1960).

[9] In Brans–Dicke's equation ω plays a role analogous to G^{-1}. In (3.8) G can be absorbed in x^0.

[10] A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U.P., Cambridge, 1930), 2nd Ed.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Magnetic Corrections to the Boltzmann Distribution Function

R. B. THOMAS, JR.
RCA Astro-Electronics Division, Princeton, New Jersey

(Received 26 August 1970)

By means of an expansion in powers of the magnetic field, where classical statistics are assumed, an expression is found for the gauge-independent equilibrium distribution properly corresponding to the ordinary Boltzmann transport equation. The validity of the new distribution is demonstrated by determining the current density induced in a simple system of free electrons simultaneously accelerated by crossed electric and magnetic fields and comparing the current with that obtained using the gauge-dependent density matrix formalism. To terms of second order in the magnetic field, the results are shown to be in exact agreement.

1. INTRODUCTION

In preceding work the relationship between the quantum theory of electrical transport in a magnetic field and the corresponding Boltzmann equation was established on the basis of a simple model.[1–3] By making use of operator methods alone, it was shown that the exact gauge-dependent Liouville equation for the density operator could be transformed into a completely gauge-independent equation satisfied by a new density operator. The diagonal elements of this new gauge-independent density operator were found to satisfy the ordinary Boltzmann transport equation if certain higher-order corrections are neglected.[4,5]

The transformation to the gauge-independent transport formalism is apparently accomplished without approximation, and the physical content of the resulting gauge-independent transport equations should be identical in all respects with that implied by the initial gauge-dependent equations. To clarify the role of the magnetic field in its relationship to the Boltzmann transport equation and at the same time demonstrate more explicitly the equivalence of the gauge-dependent and independent transport formalisms, an investigation of electrical transport in a simple system is carried out. The system considered is that of free electrons simultaneously accelerated by constant, uniform, and mutually perpendicular electric and magnetic fields. Interactions of the electrons with one another and with a scattering potential are neglected. For simplicity the electrons are assumed to obey ordinary Maxwell–Boltzmann statistics, but the treatment can readily be extended to include Fermi–Dirac statistics.

In Sec. 2 an expression for the gauge-independent equilibrium distribution is developed in ascending powers of the magnetic field from its known relationship to the gauge-dependent equilibrium density operator. Using this distribution function, we find an approximate solution to the Boltzmann equation in Sec. 3 and use it to calculate the resulting current density to terms of second order in the magnetic field. On the basis of the gauge-dependent density matrix formalism, the same current is determined exactly in Sec. 4 and found to be identical with that obtained in Sec. 3.

2. GAUGE-INDEPENDENT EQUILIBRIUM DISTRIBUTION

Consider a collection of noninteracting free electrons moving in the presence of constant, uniform electric and magnetic fields. The complete Hamiltonian for each electron in this system is

$$H_T = H_B + H_E, \tag{2.1}$$

where H_B is the Hamiltonian of a free electron in a uniform magnetic field and H_E the interaction with

* National Research Council of Canada 1967 Science Predoctoral Fellow.

[1] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961); R. H. Dicke, "Experimental Relativity," in *Relativity, Groups and Topology*, C. M. DeWitt and B. S. DeWitt, Eds. (Gordon and Breach, New York, 1964).

[2] A. N. Whitehead, *The Principle of Relativity* (Cambridge U.P., Cambridge, 1922); G. D. Birkhoff, Proc. Natl. Acad. Sci. (U.S.) **29**, 231 (1943); A. Papapetrou, Proc. Roy. Irish Acad. **2**, 11 (1948).

[3] H. Weyl, Sitzber. Preuss. Akad. Wiss., 465 (1918).

[4] The usual summation convention is used throughout.

[5] G. Lyra, Math. Z. **54**, 52 (1951); E. Scheibe, Math. Z. **57**, 65 (1952); D. K. Sen, Z. Physik **149**, 311 (1957).

[6] E. Schrödinger, *Space-Time Structure* (Cambridge U.P., Cambridge, 1954), p. 27.

[7] See, for example, N. Hicks, *Notes on Differential Geometry* (Van Nostrand, Princeton, N.J., 1965).

[8] D. K. Sen, Can. Math. Bull. **3**, 255 (1960).

[9] In Brans–Dicke's equation ω plays a role analogous to G^{-1}. In (3.8) G can be absorbed in x^0.

[10] A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U.P., Cambridge, 1930), 2nd Ed.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Magnetic Corrections to the Boltzmann Distribution Function

R. B. THOMAS, JR.
RCA Astro-Electronics Division, Princeton, New Jersey

(Received 26 August 1970)

By means of an expansion in powers of the magnetic field, where classical statistics are assumed, an expression is found for the gauge-independent equilibrium distribution properly corresponding to the ordinary Boltzmann transport equation. The validity of the new distribution is demonstrated by determining the current density induced in a simple system of free electrons simultaneously accelerated by crossed electric and magnetic fields and comparing the current with that obtained using the gauge-dependent density matrix formalism. To terms of second order in the magnetic field, the results are shown to be in exact agreement.

1. INTRODUCTION

In preceding work the relationship between the quantum theory of electrical transport in a magnetic field and the corresponding Boltzmann equation was established on the basis of a simple model.[1–3] By making use of operator methods alone, it was shown that the exact gauge-dependent Liouville equation for the density operator could be transformed into a completely gauge-independent equation satisfied by a new density operator. The diagonal elements of this new gauge-independent density operator were found to satisfy the ordinary Boltzmann transport equation if certain higher-order corrections are neglected.[4,5]

The transformation to the gauge-independent transport formalism is apparently accomplished without approximation, and the physical content of the resulting gauge-independent transport equations should be identical in all respects with that implied by the initial gauge-dependent equations. To clarify the role of the magnetic field in its relationship to the Boltzmann transport equation and at the same time demonstrate more explicitly the equivalence of the gauge-dependent and independent transport formalisms, an investigation of electrical transport in a simple system is carried out. The system considered is that of free electrons simultaneously accelerated by constant, uniform, and mutually perpendicular electric and magnetic fields. Interactions of the electrons with one another and with a scattering potential are neglected. For simplicity the electrons are assumed to obey ordinary Maxwell–Boltzmann statistics, but the treatment can readily be extended to include Fermi–Dirac statistics.

In Sec. 2 an expression for the gauge-independent equilibrium distribution is developed in ascending powers of the magnetic field from its known relationship to the gauge-dependent equilibrium density operator. Using this distribution function, we find an approximate solution to the Boltzmann equation in Sec. 3 and use it to calculate the resulting current density to terms of second order in the magnetic field. On the basis of the gauge-dependent density matrix formalism, the same current is determined exactly in Sec. 4 and found to be identical with that obtained in Sec. 3.

2. GAUGE-INDEPENDENT EQUILIBRIUM DISTRIBUTION

Consider a collection of noninteracting free electrons moving in the presence of constant, uniform electric and magnetic fields. The complete Hamiltonian for each electron in this system is

$$H_T = H_B + H_E, \tag{2.1}$$

where H_B is the Hamiltonian of a free electron in a uniform magnetic field and H_E the interaction with

the external electric field. We have

$$H_B = \frac{1}{2m}\mathbf{P}^2 \tag{2.2}$$

in which $\mathbf{P} \equiv \mathbf{p} - (e/c)\mathbf{A}(\mathbf{r})$ and

$$H_E = -eE_\alpha x_\alpha. \tag{2.3}$$

In (2.2), e is the (algebraic) charge of the electron and $\mathbf{A}(\mathbf{r})$ is the vector potential. In (2.3), the E_α, $\alpha = 1, 2, 3$, are the components of the external electric field and the repeated index α implies a summation over α. The exact density operator for the system, ρ_T, is determined by the gauge-dependent Liouville equation

$$i\hbar\frac{\partial\rho_T}{\partial t} = [H_T, \rho_T], \tag{2.4a}$$

with H_T given by (2.1). The expectation value of any observable quantity represented by the operator $M(\mathbf{r}, \mathbf{p} - (e/c)\mathbf{A}, t)$ is then given by

$$\bar{M} = \mathrm{Tr}\,\{\rho_T M\}. \tag{2.4b}$$

It was previously shown that the gauge-dependent Eqs. (2.4) can be replaced by the gauge-independent equations

$$i\hbar\frac{\partial\tilde{\rho}_T}{\partial t} = [H_0 + H_E, \tilde{\rho}_T] - \frac{e}{2mc}\{(\mathbf{p}\times\mathbf{B})\cdot[\mathbf{r}, \tilde{\rho}_T] + [\mathbf{r}, \tilde{\rho}_T]\cdot(\mathbf{p}\times\mathbf{B})\}, \tag{2.5a}$$

and

$$\bar{M} = \mathrm{Tr}\,\{\tilde{\rho}_T\tilde{M}\}, \tag{2.5b}$$

where $\tilde{\rho}_T$ is the new density operator, $H_0 = \mathbf{p}^2/2m$ is just the free electron Hamiltonian, and $\tilde{M} = (\mathbf{r}, \mathbf{p}, t)$ is the appropriate operator corresponding to the observable.[1] In carrying out the transformation of Eqs. (2.4) into Eqs. (2.5), it is useful to write ρ_T in the form of an integral operator

$$\rho_T = \int_{-\infty}^{\infty} e^{i\mathbf{P}\cdot\boldsymbol{\xi}} R_T(\boldsymbol{\xi}, \mathbf{r}, \mathbf{B}, t)\, d\boldsymbol{\xi}, \tag{2.6}$$

which can be taken to define the gauge-independent function $R_T(\boldsymbol{\xi}, \mathbf{r}, \mathbf{B}, t)$. Then $\tilde{\rho}_T$ can be written

$$\tilde{\rho}_T = \int_{-\infty}^{\infty} e^{i\mathbf{p}\cdot\boldsymbol{\xi}} R_T(\boldsymbol{\xi}, \mathbf{r}, \mathbf{B}, t)\, d\boldsymbol{\xi}, \tag{2.7}$$

where $R_T(\boldsymbol{\xi}, \mathbf{r}, \mathbf{B}, t)$ is identical with the corresponding function appearing in (2.6).

The usual procedure in treating problems of electrical transport is to assume the existence of a well-defined equilibrium distribution in the absence of an electric field and determine by means of perturbation theory the change in this distribution resulting from the application of an electric field.[6] Using this procedure and limiting ourselves to terms of first order in the electric field, we write

$$\tilde{\rho}_T = \tilde{\rho}_E + \tilde{\rho}, \tag{2.8}$$

where $\tilde{\rho}_E$ is linear in the electric field and $\tilde{\rho}$ is the new gauge-independent equilibrium density operator which satisfies the equation

$$0 = [H_0, \tilde{\rho}] - \frac{e}{2mc}\{(\mathbf{p}\times\mathbf{B})\cdot[\mathbf{r}, \tilde{\rho}] + [\mathbf{r}, \tilde{\rho}]\cdot(\mathbf{p}\times\mathbf{B})\}, \tag{2.9}$$

since $\tilde{\rho}$ is of course independent of time. Following Kohn and Luttinger,[5] we set

$$E_\alpha = E_\alpha^0 e^{st}, \tag{2.10}$$

where s is a suitably chosen frequency parameter so that E_α is zero at $t = -\infty$ and reaches its full value E_α^0 at $t = 0$. Then $\tilde{\rho}_E$ must satisfy the initial condition

$$\tilde{\rho}_E\,(t = -\infty) = 0; \tag{2.11}$$

Eqs. (2.5a) and (2.11) can now be satisfied by taking

$$\tilde{\rho}_E = fe^{st}, \tag{2.12}$$

where f is independent of time. The quantity f is the correction to the equilibrium density operator at $t = 0$, which is what we want. Choosing a plane wave representation in which H_0 is diagonal and assuming that $s \ll \omega_0$, we find that the diagonal matrix elements of f satisfy the equation

$$0 = e\mathbf{E}^0\cdot\nabla_k\tilde{\rho}_k + \frac{e\hbar}{mc}(\mathbf{k}\times\mathbf{B})\cdot\nabla_k f_k, \tag{2.13}$$

where we have set $\tilde{\rho}_k \equiv \tilde{\rho}_{kk}$ and $f_k \equiv f_{kk}$. Except for the absence of a potential scattering term, Eq. (2.13) has the form of the ordinary Boltzmann equation, where f_k is just the correction to the distribution function linear in the electric field.

Before Eq. (2.13) can be solved, $\tilde{\rho}$ must be known. In most treatments of electrical transport based on the Boltzmann equation, $\tilde{\rho}$ is incorrectly taken to be simply the Boltzmann distribution function in the absence of a magnetic field, it being assumed that the effects of the magnetic field are properly taken into account by the magnetic field term

$$(e\hbar/mc)(\mathbf{k}\times\mathbf{B})\cdot\nabla_k f_k.$$

Actually $\tilde{\rho}$ itself depends on the magnetic field, allowing the possibility of magnetic effects not encountered in the usual treatment of electrical transport based on the Boltzmann equation. While $\tilde{\rho}$ must satisfy Eq. (2.9), this requirement is not sufficient to determine its exact form. However, $\tilde{\rho}$ can be found from the implied relationship between $\tilde{\rho}_T$ and ρ_T as expressed by Eqs. (2.6) and (2.7). Let

$$\rho_T = \rho_E + \rho, \tag{2.14}$$

where ρ_E is linear in the electric field and ρ is the exact gauge-dependent equilibrium density operator satisfying Eq. (2.4a) at $t = -\infty$. From statistical mechanics we know

$$\rho = Ke^{-\beta \mathbf{P}^2/2m}, \tag{2.15a}$$

where $\beta = 1/kT$ and

$$K = \frac{1}{V}\left(\frac{2\pi\hbar^2\beta}{m}\right)^{\frac{3}{2}} \frac{2kT}{\hbar\omega_0} \sin\frac{\hbar\omega_0}{2kT}. \tag{2.15b}$$

Placing (2.15) in (2.6) gives the following equation for determining $R(\boldsymbol{\xi}, \mathbf{B})$:

$$Ke^{-\beta \mathbf{P}^2/2m} = \int_{-\infty}^{\infty} e^{i\mathbf{P}\cdot\boldsymbol{\xi}} R(\boldsymbol{\xi}, \mathbf{B})\, d\boldsymbol{\xi}. \tag{2.16}$$

Once $R(\boldsymbol{\xi}, \mathbf{B})$ is known, $\tilde{\rho}$ can be found directly from (2.7).

In determining $R(\boldsymbol{\xi}, \mathbf{B})$ it is convenient to choose the Landau gauge $\mathbf{A}(\mathbf{r}) = (-yB, 0, 0)$, which gives a magnetic field $\mathbf{B}$ directed in the positive z direction. Setting

$$R(\boldsymbol{\xi}, \mathbf{B}) = KS_{xy}(\xi_x, \xi_y)S_z(\xi_z), \tag{2.17}$$

we can factor Eq. (2.16) at once into the pair of operator equations

$$e^{-\beta(P_x^2+P_y^2)/2m} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(P_x\xi_x+P_y\xi_y)} S_{xy}(\xi_x, \xi_y)\, d\xi_x\, d\xi_y \tag{2.18}$$

and

$$e^{-\beta P_z^2/2m} = \int_{-\infty}^{\infty} e^{iP_z\xi_z} S_z(\xi_z)\, d\xi_z, \tag{2.19}$$

since P_z commutes with both P_x and P_y. To solve for $S_z(\xi_z)$, form the matrix of (2.19) using a plane wave representation. Inverting the result gives

$$S_z(\xi_z) = \left(\frac{m}{2\pi\beta}\right)^{\frac{1}{2}} e^{-m\xi_z^2/2\beta}, \tag{2.20}$$

which is identical with the expression obtained directly from (2.19) treating P_z as if it were an ordinary continuous variable.

Finding $S_{xy}(\xi_x, \xi_y)$ is somewhat more difficult since P_x and P_y do not commute. It is useful to begin by factoring the exponential operators appearing in (2.18). First consider the left-hand side of (2.18); this can be factored by means of the relationship[7]

$$e^{p^2+q^2} = (\sec 2c)^{\frac{1}{2}} e^{(p^2/2c)\tan 2c} e^{-(pq/c)\ln\cos 2c} e^{(q^2/2c)\tan 2c}, \tag{2.21}$$

which holds whenever the operators p and q commute with their commutator $[q, p] \equiv c$. Taking $p = i(\beta/2m)^{\frac{1}{2}}P_x$ and $q = i(\beta/2m)^{\frac{1}{2}}P_y$, we find $c = \frac{1}{2}i\beta\hbar\omega_0$, which, being a constant, commutes with p and q. The exponential operator on the right-hand side of Eq. (2.18) can be factored by making use of the relationship[8]

$$e^{u+v} = e^u e^v e^{-\frac{1}{2}[u,v]}, \tag{2.22}$$

where again u and v are two operators which commute with their commutator $[u, v]$. Taking $u = i\omega_0 my\xi_x$ and $v = i\xi_y p_y$, we find $[u, v] = -i\omega_0 m\hbar\xi_x\xi_y$, which does in fact commute with u and v as defined here. Rewrite (2.18), making use of (2.21) and (2.22). Now assume that $\beta\hbar\omega_0 \ll 1$, a condition that is well satisfied for a wide range of interesting temperatures and magnetic fields, and expand the results in powers of $\beta\hbar\omega_0$, taking care to preserve the order of noncommuting operators. This gives

$$\begin{aligned} e^{-\beta P_x^2/2m}\Big(1 &- \frac{i\beta^2\hbar\omega_0}{2m} P_xP_y \\ &- \frac{\beta^2\hbar^2\omega_0^2}{4} + \frac{\beta^3\hbar^2\omega_0^2 P_x^2}{6m} - \frac{\beta^4\hbar^2\omega_0^2}{8m^2}(P_xP_y)^2 \\ &+ \frac{\beta^3\hbar^2\omega_0^2}{6m} P_y^2\Big) e^{-\beta P_y^2/2m} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iP_x\xi_x} e^{iP_y\xi_y} \\ &\times [1 + \tfrac{1}{2}i\hbar\omega_0 m\xi_x\xi_y + \tfrac{1}{2}(\tfrac{1}{2}i\hbar\omega_0 m\xi_x\xi_y)^2] \\ &\times \sum_{n=0} S_{xy}^{(n)}(\xi_x, \xi_y)(\beta\hbar\omega_0)^n\, d\xi_x\, d\xi_y. \end{aligned} \tag{2.23}$$

Here we have let

$$S_{xy}(\xi_x, \xi_y) = \sum_{n=0} S_{xy}^{(n)}(\xi_x, \xi_y)(\beta\hbar\omega_0)^n, \tag{2.24}$$

in which the coefficients $S_{xy}^{(n)}(\xi_x, \xi_y)$ are assumed to be independent of the magnetic field. In (2.23) collect terms of the same order in $\beta\hbar\omega_0$ and set their respective sums equal to zero. This gives the following set of equations for determining the functions $S_{xy}^{(n)}(\xi_x, \xi_y)$:

$$e^{-\beta P_x^2/2m} e^{-\beta P_y^2/2m} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iP_x\xi_x} e^{iP_y\xi_y} S_{xy}^{(0)}(\xi_x, \xi_y)\, d\xi_x\, d\xi_y, \tag{2.25a}$$

$$\begin{aligned} \frac{i\beta^2}{2m} e^{-\beta P_x^2/2m} P_xP_y e^{-\beta P_y^2/2m} &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iP_x\xi_x} e^{iP_y\xi_y} \\ &\times [\tfrac{1}{2}\,\mathrm{im}\,\xi_x\xi_y S_{xy}^{(0)}(\xi_x, \xi_y) + \beta S_{xy}^{(1)}(\xi_x, \xi_y)]\, d\xi_x\, d\xi_y, \end{aligned} \tag{2.25b}$$

$$\begin{aligned} \tfrac{1}{2}\beta^2 e^{-\beta P_x^2/2m}\Big[\frac{1}{2} &- \frac{\beta}{3m}(P_x^2 + P_y^2) \\ &+ \frac{1}{4}\frac{\beta^2}{m^2}(P_xP_y)^2\Big] e^{-\beta P_y^2/2m} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iP_x\xi_x} e^{iP_y\xi_y} [\tfrac{1}{8}m^2\xi_x^2\xi_y^2 S_{xy}^{(0)}(\xi_x, \xi_y) \\ &- \tfrac{1}{2}\,\mathrm{im}\,\beta\xi_x\xi_y S_{xy}^{(1)}(\xi_x, \xi_y) - \beta^2 S_{xy}^{(2)}(\xi_x, \xi_y)]\, d\xi_x\, d\xi_y. \end{aligned} \tag{2.25c}$$

To solve these equations, we begin by setting $S_{xy}^{(0)}(\xi_x, \xi_y) = S_x^{(0)}(\xi_x)S_y^{(0)}(\xi_y)$. Then (2.25a) can be separated and solved giving

$$S_x^{(0)}(\xi_x) = \left(\frac{m}{2\pi\beta}\right)^{\frac{1}{2}} e^{-m\xi_x^2/2\beta}, \tag{2.26a}$$

$$S_y^{(0)}(\xi_y) = \left(\frac{m}{2\pi\beta}\right)^{\frac{1}{2}} e^{-m\xi_y^2/2\beta}. \tag{2.26b}$$

By using this result, the first term on the right-hand side of (2.25b) can be evaluated. It turns out to be identical with the left-hand term, which implies that $S_{xy}^{(1)} = 0$. In this case the second term on the right-hand side of (2.25c) is also zero. Make use of the expression for $S_{xy}^{(0)}$ to evaluate the first term on the right-hand side of (2.25c). In the resulting equation replace $(P_xP_y)^2$ with $P_x^2P_y^2$ by neglecting a term of third order in the magnetic field. Then

$$e^{-\beta P_x^2/2m}\left(\frac{\beta}{2m}(P_x^2 + P_y^2) - 1\right)e^{-\beta P_y^2/2m} = 8\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iP_x\xi_x}e^{iP_y\xi_y}S_{xy}^{(2)}\,d\xi_x\,d\xi_y, \tag{2.27}$$

which has the solution

$$S_{xy}^{(2)} = -\frac{1}{24}\left(\frac{m}{2\pi\beta}\right)\left(1 + \frac{m}{\beta}(\xi_x^2 + \xi_y^2)\right)e^{-(m/2\beta)(\xi_x^2+\xi_y^2)}. \tag{2.28}$$

Substituting our results into (2.17) and expanding $K(\beta)$ to the appropriate order, we find

$$R(\boldsymbol{\xi}, \mathbf{B}) = \frac{\hbar^3}{V}\left(1 - \frac{1}{24}\frac{m}{\beta}(\hbar\omega_0\beta)^2(\xi_x^2 + \xi_y^2)\right)e^{-(m/2\beta)\xi^2}, \tag{2.29}$$

which satisfies the condition $[R(\xi, \mathbf{B})_{kk}]_{\xi=0} = \hbar^3/V$ derived previously.[4]

Making use of (2.29), we find from (2.7) that the gauge-independent equilibrium density operator is just

$$\tilde{\rho} = \frac{1}{V}\left(\frac{2\pi\hbar^2\beta}{m}\right)^{\frac{3}{2}} \times \left[1 + \frac{(\hbar\omega_0\beta)^2}{24}\left(\frac{\beta}{m}(p_x^2 + p_y^2) - 2\right)\right]e^{-\beta\mathbf{p}^2/2m}. \tag{2.30}$$

This satisfies Eq. (2.9) as it must. In a plane-wave representation, $\tilde{\rho}$ has only diagonal elements,

$$\tilde{\rho}_{kk'} = \frac{1}{V}\left(\frac{2\pi\hbar^2\beta}{m}\right)^{\frac{3}{2}}\left[1 + \frac{(\hbar\omega_0\beta)^2}{24} \times \left(\frac{\beta\hbar^2}{m}(k_x^2 + k_y^2) - 2\right)\right]e^{-\beta\hbar^2\mathbf{k}^2/2m}\delta_{k,k'}, \tag{2.31}$$

from which one can readily show that the trace of $\tilde{\rho}$ is unity.

3. CURRENT DENSITY

As a simple example of the present theory, which can be treated exactly by other means, we calculate the current density produced by mutually perpendicular electric and magnetic fields chosen to lie along the positive x and z axes, respectively. Then Eq. (2.13) becomes

$$0 = eE_x^0\frac{\partial\tilde{\rho}_k}{\partial k_x} + \hbar\omega_0\left(k_y\frac{\partial}{\partial k_x} - k_x\frac{\partial}{\partial k_y}\right)f_k. \tag{3.1}$$

To solve this equation, assume that f_k can be written in the form

$$f_k = -\frac{\hbar}{V}\left(\frac{2\pi\hbar^2\beta}{m}\right)^{\frac{3}{2}}\frac{eE_x^0\beta}{\omega_0 m}Q(k_x^2, k_y^2, k_z^2)k_y, \tag{3.2}$$

where Q is an unknown function depending only on the square of each component of the wave vector $\mathbf{k}$. Substituting (2.31) and (3.2) into Eq. (3.1) shows that

$$Q = \left[1 + \frac{1}{24}(\hbar\omega_0\beta)^2\left(\frac{\beta\hbar^2}{m}(k_x^2 + k_y^2) - 4\right)\right]e^{-\beta\hbar^2\mathbf{k}^2/2m}, \tag{3.3}$$

which is seen to have the form required of Q.

The gauge-independent velocity operator is just p/m.[1] Then according to (2.5b) the current density is given by

$$\bar{\mathbf{J}} = \frac{en_0}{m}\mathrm{Tr}\,\{\tilde{\rho}_T\mathbf{p}\} = \frac{en_0\hbar}{m}\sum_k f_k\mathbf{p}, \tag{3.4}$$

where n_0 is the electron density, the equilibrium distribution contributing nothing to the current. Substitute (3.2) into (3.4) and carry out the indicated sum over $\mathbf{k}$. Only the y component of the current density does not vanish, giving

$$\bar{J}_y = -\frac{en_0\hbar^2}{(2\pi)^3 m}\left(\frac{2\pi\hbar^2\beta}{m}\right)^{\frac{3}{2}}\frac{eE_x^0}{\omega_0}\frac{\beta}{m} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}k_y^2\left[1 + \frac{1}{24}(\omega_0\hbar\beta)^2 \times \left(\frac{\beta\hbar^2}{m}(k_x^2 + k_y^2) - 4\right)\right]e^{-\beta\hbar^2\mathbf{k}^2/2m}\,dk_x\,dk_y\,dk_z = -en_0c\frac{E_x^0}{B}, \tag{3.5}$$

a result which is good to second order in $\hbar\omega_0\beta$. Actually (3.5) is identical with the exact result (to terms linear in $\mathbf{E}^0$) obtained using the gauge-dependent density operator formalism; this is shown in the following section.

4. GAUGE-DEPENDENT DENSITY MATRIX FORMALISM

It is of interest to compare our result (3.5) for the current density with the exact value which can be found using the gauge-dependent density matrix formalism.[9] Again following Kohn and Luttinger,[5] we take

$$\rho_E = ge^{st}, \tag{4.1}$$

where g is linear in the electric field and s is a suitably chosen frequency. We are then interested in g at $t \equiv 0$, when the electric field (2.10) attains its full value $\mathbf{E}^0$. To obtain an equation for g, combine (4.1) with (2.14) and substitute the result into Eq. (2.4a). This gives

$$i\hbar sg = \frac{1}{2m}[\mathbf{P}^2, g] - eE_\alpha^0[x_\alpha, \rho], \tag{4.2}$$

if terms of second order in the electric field are neglected.

To solve Eq. (4.2), choose a representation $\varphi_\nu(\mathbf{r})$ in which H_B is diagonal: $H_B\varphi_\nu = E_\nu\varphi_\nu$. Taking the gauge $\mathbf{A}(\mathbf{r}) = (0, xB, 0)$, we have

$$\varphi_\nu(\mathbf{r}) = (L_yL_z)^{-\frac{1}{2}}u_n(x + x_0)e^{ik_y y}e^{ik_z z}. \tag{4.3}$$

Here ν represents the set of quantum numbers (n, k_y, k_z) and $u_n(x + x_0)$ is the harmonic oscillator wavefunction centered at $x_0 = -k_y/\alpha^2$, where $\alpha^2 = m\omega_0/\hbar$. Form the matrix of (4.2), using (4.3), and solve the resulting equation for $g_{\nu\nu'}$. We find

$$\begin{aligned} g_{\nu\nu'} &= -eE_x^0\frac{\rho_\nu - \rho_{\nu'}}{E_\nu - E_{\nu'}}[x]_{\nu\nu'} \\ &= -eE_x^0\frac{\rho_\nu - \rho_{\nu'}}{E_\nu - E_{\nu'}}\frac{1}{\alpha}\left[\left(\frac{n+1}{2}\right)^{\frac{1}{2}}\delta_{n+1,n'}\right. \\ &\quad \left. + \left(\frac{n}{2}\right)^{\frac{1}{2}}\delta_{n-1,n'} - \alpha x_0\delta_{n,n'}\right]\delta_{k_y,k_{y'}}\delta_{k_z,k_{z'}}, \end{aligned} \tag{4.4}$$

after inserting an explicit expression for the matrix $[x]_{\nu\nu'}$ and again assuming $s \ll \omega_0$.

The gauge-dependent velocity operator is given by the commutator of the position coordinate $\mathbf{r}$ with the Hamiltonian H_B:

$$\mathbf{v} = \frac{i}{\hbar}[H_B, \mathbf{r}] = \frac{1}{m}\mathbf{P}. \tag{4.5}$$

To obtain the current density, multiply (4.5) by the electronic charge density n_0e and insert it along with (4.4) into Eq. (2.4b). The only nonvanishing contribution to the current density is

$$\begin{aligned} \bar{J}_y &= \frac{en_0}{m}\operatorname{Tr}\{\rho_T P_y\} \\ &= 4n_0e\left(\frac{\hbar\omega_0}{2m}\right)^{\frac{1}{2}}\sum_n\sum_{k_y}\sum_{k_z}(n+1)^{\frac{1}{2}}g_{nk_yk_z,n+1\,k_yk_z} \\ &= -en_0c\frac{E_x^0}{B}, \end{aligned} \tag{4.6}$$

which is seen to be identical with our previous result (3.5).

5. SUMMARY AND DISCUSSION

The current density induced by the application of an electric field to electrons moving in the presence of a uniform magnetic field is found by means of a previously established gauge-independent transport formalism. In attempting to solve for the gauge-independent distribution function by the usual procedure of finding the correction to the equilibrium distribution linear in the electric field, it becomes apparent that the equation for the gauge-independent density operator is not sufficient to determine the precise form of the needed equilibrium distribution. An approximate expression for the gauge-independent equilibrium distribution is derived from its known relationship to the gauge-dependent equilibrium density operator. From this expression the correction to the distribution function linear in the electric field is found and used to calculate the current density, which turns out to be identical with the exact result obtained using the gauge-dependent transport formalism.

In the usual treatment of transport problems based on the Boltzmann equation, the equilibrium distribution is taken to be the ordinary Boltzmann distribution function. In the absence of potential scattering, this is equivalent to taking

$$\tilde{\rho} = Ke^{-\beta\mathbf{p}^2/2m}, \tag{5.1}$$

with $K^{-1} = \operatorname{Tr}\tilde{\rho}$, even when a magnetic field is present. While Eq. (2.9) is satisfied by (5.1), it is equally well satisfied by any function of $\mathbf{p}^2$. It is clear in any case that (5.1) can not be correct when a magnetic field is present since it depends in no way on the magnetic field. Furthermore, any attempt to solve Eq. (2.9) leads at once to the conclusion that the equation itself is not sufficient to determine the equilibrium distribution. This is true even if one takes (5.1) to be the principal contribution and then attempts to include the effects of the magnetic field as a perturbation. More information about $\tilde{\rho}$ is required and this is supplied by the known connection between ρ and $\tilde{\rho}$ through the function $R(\boldsymbol{\xi}, \mathbf{r}, \mathbf{B}, t)$. ρ itself is, of course, known from statistical mechanics. In determining $\tilde{\rho}$ from ρ, we find that $\tilde{\rho}$ has as its principal term just (5.1), as expected [see Eq. (2.30)], but contains additional terms depending on the magnetic field. In the particularly simple problem we have chosen to treat, it turns out that the contributions to the current density arising from the additional magnetic terms in the distribution function exactly cancel one another. Indeed, even with the

limited number of terms we have retained in the expansion of the distribution function, the supposed approximate result for the current density is found to be identical with the exact result at least up to terms linear in the electric field. The inclusion of additional terms in the series expression for the gauge-independent distribution function would probably not change this result. The identity of our approximate expression for the current density with the exact result is very possibly fortuitous and related to the fact that the same current is also given by a calculation based entirely on the principles of classical particle mechanics.

If we used a similar approach, it would be of interest to investigate further the influence of a magnetic field on electrical transport by considering a more realistic system in which potential scattering occurs, particularly scattering by a periodic potential. Such an investigation has been undertaken but is complicated by the necessity of including several higher-order terms in the transport equation.

ACKNOWLEDGMENT

The author wishes to thank Professor J. M. Luttinger for helpful comments on the work presented here.

[1] R. B. Thomas, Jr., Phys. Rev. **171**, 827 (1968).
[2] J. Mertsching and H. W. Streitwolf, Phys. Status Solidi **21**, K167 (1967).
[3] R. B. Stinchcombe, Proc. Phys. Soc. (London) **78**, 275 (1961).
[4] R. B. Thomas, Jr., Phys. Rev. **152**, 138 (1966).
[5] W. Kohn and J. M. Luttinger, Phys. Rev. **108**, 590 (1957).
[6] See, for example, F. Blatt, "Theory of Mobility of Electrons in Solids," in *Solid State Physics* (Academic, New York, 1957), Vol. 4, p. 214.
[7] H. T. Davis, *The Theory of Linear Operators* (Principia, Bloomington, Ind., 1936), Chap. 4.
[8] E. Merzbacher, *Quantum Mechanic*, (Wiley, New York, 1961). p. 162.
[9] P. N. Argyres and L. M. Roth, Phys. Chem. Solids **12**, 89 (1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Continuity of Automorphisms of Quasilocal Algebra

ETANG CHEN
Department of Mathematics, East Carolina University, Greenville, North Carolina

(Received 31 March 1970)

We show that automorphisms of quasilocal algebra $\mathfrak{A}$, a C^*-algebra with identity on Hilbert space $\mathfrak{H}$, is extendable to the weak operator closure of $\mathfrak{A}$ under a certain physical assumption.

Given a topological group G and a C^*-algebra $\mathfrak{A}$ with identity acting on a Hilbert space $\mathfrak{H}$, let α be a homomorphic mapping from G into Aut $(\mathfrak{A})$, the group of $*$-automorphisms of $\mathfrak{A}$, and α_g be the image of $g \in G$ under α, i.e., $\alpha: g \to \alpha_g$ for $g \in G$ and $\alpha_g \in$ Aut $(\mathfrak{A})$. Suppose that $g \to \langle \xi, \alpha_g(A)\eta \rangle$ is continuous on G for all $A \in \mathfrak{A}$ and all $\xi, \eta \in \mathfrak{H}$. Let $\mathfrak{A}^-$ be the weak-operator closure of $\mathfrak{A}$; then there are two problems: (i) Can each α_g be extended to a $*$-automorphism of $\mathfrak{A}^-$ and (ii) is $g \to \langle \xi, \alpha_g(A)\eta \rangle$ continuous on G for all $A \in \mathfrak{A}^-$ and $\xi, \eta \in \mathfrak{H}$? Aarens solved these problems when the mapping $(g, A) \to \alpha_g(A)$ of $G \times \mathfrak{A}$ into $\mathfrak{A}$ satisfies a mild joint continuity condition.[1] Recently, Kallman[2] assumed (i) and showed that (ii) is automatically correct if G has a complete metric topology. Kadison and Ringrose[3] showed that (i) is true if

$$\|\alpha_g - \iota\| < 2, \tag{1}$$

where ι is the identity automorphism of $\mathfrak{A}$.

However, (1) seems to be a very strong condition for physical problems. In this short note we will give another condition based on physical motivation, such that (i) is true for quasilocal algebra of observables in quantum field theory, which is the original interest to study the continuity of automorphisms of C^*-algebras.[1]

We recall some well-known local structures of quasilocal algebra of observables: For every bounded open region O of Minkowski space-time M, there is a local von Neumann algebra $R(O)$ on a (separable) Hilbert space $\mathfrak{H}$, such that $R(O_1) \subseteq R(O_2)$ if $O_1 \subseteq O_2$. Let Q be the union of all local von Neumann algebras $R(O)$, i.e., $Q = \bigcup_{i \in I} R(O_i)$, where I is some index set such that $\bigcup_{i \in I} O_i = M$. We choose to designate, under the name *quasilocal algebra of observables*, the C^*-algebra $\mathfrak{A}$ obtained as the norm-operator closure of Q, denoted by $\bar{Q}^n$, i.e., $\mathfrak{A} = \bar{Q}^n$.

The symmetries of a physical theory within this framework are represented by a $*$-automorphism α_g of $\mathfrak{A}$, such that each local algebra $R(O)$ is mapped isomorphically onto another local algebra $R(O_g)$,

$$\alpha_g: R(O) \to \alpha_g(R(O)) = R(O_g),$$

where O_g is the image of O "shifted by $g \in G$."

limited number of terms we have retained in the expansion of the distribution function, the supposed approximate result for the current density is found to be identical with the exact result at least up to terms linear in the electric field. The inclusion of additional terms in the series expression for the gauge-independent distribution function would probably not change this result. The identity of our approximate expression for the current density with the exact result is very possibly fortuitous and related to the fact that the same current is also given by a calculation based entirely on the principles of classical particle mechanics.

If we used a similar approach, it would be of interest to investigate further the influence of a magnetic field on electrical transport by considering a more realistic system in which potential scattering occurs, particularly scattering by a periodic potential. Such an investigation has been undertaken but is complicated by the necessity of including several higher-order terms in the transport equation.

ACKNOWLEDGMENT

The author wishes to thank Professor J. M. Luttinger for helpful comments on the work presented here.

[1] R. B. Thomas, Jr., Phys. Rev. **171**, 827 (1968).
[2] J. Mertsching and H. W. Streitwolf, Phys. Status Solidi **21**, K167 (1967).
[3] R. B. Stinchcombe, Proc. Phys. Soc. (London) **78**, 275 (1961).
[4] R. B. Thomas, Jr., Phys. Rev. **152**, 138 (1966).
[5] W. Kohn and J. M. Luttinger, Phys. Rev. **108**, 590 (1957).
[6] See, for example, F. Blatt, "Theory of Mobility of Electrons in Solids," in *Solid State Physics* (Academic, New York, 1957), Vol. 4, p. 214.
[7] H. T. Davis, *The Theory of Linear Operators* (Principia, Bloomington, Ind., 1936), Chap. 4.
[8] E. Merzbacher, *Quantum Mechanic*, (Wiley, New York, 1961). p. 162.
[9] P. N. Argyres and L. M. Roth, Phys. Chem. Solids **12**, 89 (1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Continuity of Automorphisms of Quasilocal Algebra

ETANG CHEN
Department of Mathematics, East Carolina University, Greenville, North Carolina

(Received 31 March 1970)

We show that automorphisms of quasilocal algebra $\mathfrak{A}$, a C^*-algebra with identity on Hilbert space $\mathfrak{H}$, is extendable to the weak operator closure of $\mathfrak{A}$ under a certain physical assumption.

Given a topological group G and a C^*-algebra $\mathfrak{A}$ with identity acting on a Hilbert space $\mathfrak{H}$, let α be a homomorphic mapping from G into Aut $(\mathfrak{A})$, the group of $*$-automorphisms of $\mathfrak{A}$, and α_g be the image of $g \in G$ under α, i.e., $\alpha: g \to \alpha_g$ for $g \in G$ and $\alpha_g \in$ Aut $(\mathfrak{A})$. Suppose that $g \to \langle \xi, \alpha_g(A)\eta \rangle$ is continuous on G for all $A \in \mathfrak{A}$ and all $\xi, \eta \in \mathfrak{H}$. Let $\mathfrak{A}^-$ be the weak-operator closure of $\mathfrak{A}$; then there are two problems: (i) Can each α_g be extended to a $*$-automorphism of $\mathfrak{A}^-$ and (ii) is $g \to \langle \xi, \alpha_g(A)\eta \rangle$ continuous on G for all $A \in \mathfrak{A}^-$ and $\xi, \eta \in \mathfrak{H}$? Aarens solved these problems when the mapping $(g, A) \to \alpha_g(A)$ of $G \times \mathfrak{A}$ into $\mathfrak{A}$ satisfies a mild joint continuity condition.[1] Recently, Kallman[2] assumed (i) and showed that (ii) is automatically correct if G has a complete metric topology. Kadison and Ringrose[3] showed that (i) is true if

$$\|\alpha_g - \iota\| < 2, \tag{1}$$

where ι is the identity automorphism of $\mathfrak{A}$.

However, (1) seems to be a very strong condition for physical problems. In this short note we will give another condition based on physical motivation, such that (i) is true for quasilocal algebra of observables in quantum field theory, which is the original interest to study the continuity of automorphisms of C^*-algebras.[1]

We recall some well-known local structures of quasilocal algebra of observables: For every bounded open region O of Minkowski space-time M, there is a local von Neumann algebra $R(O)$ on a (separable) Hilbert space $\mathfrak{H}$, such that $R(O_1) \subseteq R(O_2)$ if $O_1 \subseteq O_2$. Let Q be the union of all local von Neumann algebras $R(O)$, i.e., $Q = \bigcup_{i \in I} R(O_i)$, where I is some index set such that $\bigcup_{i \in I} O_i = M$. We choose to designate, under the name *quasilocal algebra of observables*, the C^*-algebra $\mathfrak{A}$ obtained as the norm-operator closure of Q, denoted by $\bar{Q}^n$, i.e., $\mathfrak{A} = \bar{Q}^n$.

The symmetries of a physical theory within this framework are represented by a $*$-automorphism α_g of $\mathfrak{A}$, such that each local algebra $R(O)$ is mapped isomorphically onto another local algebra $R(O_g)$,

$$\alpha_g: R(O) \to \alpha_g(R(O)) = R(O_g),$$

where O_g is the image of O "shifted by $g \in G$."

In the present article we shall study only automorphisms of a quasilocal algebra with the above local structures. Before the formulation of the main result below, some other properties of local observables will be discussed.

Given a vector state (a vector state ω is a state of $\mathfrak{A}$ such that $\omega(A) = \langle \xi, A\eta \rangle$ for $A \in \mathfrak{A}$ and $\xi, \eta \in \mathfrak{H}$.) ω of $\mathfrak{A}$ and $\omega(A^*A)$ the probability of ω through $A \in \mathfrak{A}$ with $\|A\| \leq 1$, then ω can be determined by the collection of probabilities $\omega(A^*A)$ for all observables A. However, in an actual experiment, e.g., a monitoring experiment, one can obtain only a finite number of probabilities $\omega(A_i^*A_i)$ for a finite set of observables A_i, $i = 1, 2, \cdots, n$ (indeed, there are only a finite number of measurements, apparatus, and experiments). And, moreover, due to the limitation of accuracy in experiments, one is unable to find exactly $\omega(A_i^*A_i)$ but p_i within a finite accuracy, viz.,

$$|p_i - \omega(A_i^*A_i)| < \epsilon_i, \tag{2}$$

for $i = 1, 2, \cdots, n$, where the ϵ_i are experimental errors. Expression (2) indicates that there are no actual experiments which enable us to specify a definite state ω, but only a weak $*$-neighborhood of ω in $\mathfrak{A}^*$, the dual space of $\mathfrak{A}$.[4]

If we want to determine an observable $A \in \mathfrak{A}$ with $\|A\| \leq 1$ in some vector states ω_i, instead of fixing a state through observables, we have also to find the set of probabilities $\omega_i(A^*A)$ for all vector states ω_i. Similarly to the above restrictions and limitations in actual experiments, what can be obtained from experimental measurements is not exactly $\omega_i(A^*A)$ but q_i such that

$$|q_i - \omega_i(A^*A)| < \epsilon_i \tag{3}$$

for $i = 1, 2, \cdots, n$ and some ϵ_i (the errors in experiments). Expression (3) gives a weak-operator neighborhood of A in $\mathfrak{A}$, given by the weak-operator topology of $\mathfrak{B}(\mathfrak{H})$, the set of all bounded operators in $\mathfrak{H}$. Hence we are also unable to find an observable exactly, but only its weak-operator neighborhood.

Another physical interpretation of (3) can be given in terms of *effect* and *ensemble* (Gesamheit) introduced by Ludwig[5] as follows, although the notions of effect and ensemble contain more than observables and vector states: For an ideal effect A and a finite set of ensembles ω_i, $i = 1, 2, \cdots, n$, one can always find a real effect T such that $q_i = \omega_i(T^*T)$, satisfying (3) for a given experimental error ϵ_i. A finite set of ensembles ω_i indicates, physically, that there are only a finite number of apparatus and experiments. "One can always find a real effect T" means that one can always construct an instrument in experiment to measure a real effect. Therefore, (3) is in fact the notion of *physical approximation* of ideal effects given in Ref. 5.

Suppose that $A \in R(O)$ with $\|A\| \leq 1$. Since all experiments (including the locations of instruments and measurements) are operated in the finite region O, then the experimental results (real effects) are obtained from O. Mathematically, this means that there are real effects T in $R(O)$ with $\|T\| \leq 1$ such that (3) is satisfied. Therefore, there is an open weak-operator neighborhood of A belonging to $R_1(O)$, the unit ball of $R(O)$, i.e., A is an interior point of $R_1(O)$. Let $\mathring{R}_1(O)$ be the interior of $R_1(O)$ with respect to the weak-operator topology of $\mathfrak{B}(\mathfrak{H})$; then the conclusion of the above discussion is $A \in \mathring{R}_1(O)$.

An observable $A \in Q_1$, the unit ball of Q, has I-*property* ("I" for interior) if $A \in \mathring{R}_1(O)$ for some finite region $O \subset M$. Furthermore, a subset W of Q_1 has I-property if, for each $A \in W$, there is some finite region $O \subset M$ such that $A \in \mathring{R}_1(O)$. Consequently, if W has I-property, then

$$W \subseteq \bigcup_{i \in I} \mathring{R}_1(O_i)$$

for some index set I.

The purpose of this article is to show the following.

Theorem: Let α_g be an automorphism of quasilocal algebra $\mathfrak{A}$ as defined above. Suppose each weakly closed subset of Q_1, the unit ball of Q, has I-property; then α_g is extendable to $\mathfrak{A}^-$.

Before we prove the main theorem, we need two lemmas.

Lemma 1: Let α_g be an automorphism of quasilocal algebra $\mathfrak{A}$. Then α_g is ultraweakly continuous on $\mathfrak{A}$, if α_g is ultraweakly continuous on Q.

Proof: Let ϕ be a normal state with $\|\Phi\| = 1$ on $\mathfrak{A}^-$, and $\epsilon > 0$. Since $\bar{Q}^n = \mathfrak{A}$, then for each $A \in \mathfrak{A}$ there exists $T \in Q$ such that $\|A - T\| \leq \epsilon/3$. Moreover, by the assumption, $|\phi[\alpha_g(T)] - \phi[\alpha_{g_0}(T)]| \leq \epsilon/3$ for $T \in Q$ and $g_0 \in G$. Hence

$$\begin{aligned} |\phi[\alpha_g(A)] - \phi[\alpha_{g_0}(A)]| &\leq |\phi[\alpha_g(A - T)]| + |\phi[\alpha_g(T)] - \phi[\alpha_{g_0}(T)]| \\ &\quad + |\phi[\alpha_{g_0}(T - A)]| \\ &\leq 2\|A - T\| + \epsilon/3 \leq \epsilon. \end{aligned}$$

Therefore, α_g is ultraweakly continuous on $\mathfrak{A}$.

Lemma 2: Let $Q = \bigcup_{i \in I} R(O_i)$ for $i \in I$, where $\bigcup_{i \in I} O_i = M$. If α_g is an automorphism of quasilocal

algebra $\mathfrak{A}$ such that

$$\alpha_g: R(O) \to \alpha_g(R(O)) = R(O_g)$$

is a isomorphic mapping, then α_g is ultraweakly continuous on Q if each weakly closed subset of Q_1, the unit ball of Q, has I-property.

Proof: Let W be a weakly closed subset of Q_1. By assumption, W has I-property; hence $W \subseteq \bigcup_{i\in I} \mathring{R}_1(O_i)$ for some index set I. Furthermore, W is compact in the weak-operator topology; thus $W \subseteq \bigcup_{i\in I} \mathring{R}_1(O_i)$ for some finite index subset $J \subset I$. Hence

$$W \subseteq \bigcup_{i\in J} R_1(O_i)$$

for finite subset J of I.

For each $A \in W$ by assumption there is some O_i such that $R_1(O_i)$ is a weak-operator neighborhood of A and, since W is weakly closed, the intersection of W and $R_1(O_i)$ is nonempty, i.e., $W \cap R_1(O_i) = V_i \neq \varnothing$. Also, V_i is weakly closed, because $R_1(O_i)$ is also weakly closed. Clearly, V_i is a weakly closed subset of $R_1(O_i)$, and $W \cap \bigcup_{i\in J} R_1(O_i) = \bigcup_{i\in J} V_i$ for finite index set J. In fact, we have

$$W = \bigcup_{i\in J} V_i$$

for finite index set J.

Moreover, α_g is a isomorphic mapping of $R(O_{g^{-1}})$ onto $R(O)$, where $O_{g^{-1}}$ is the inverse image of O under $g \in G$; then α_g is ultraweakly bicontinuous,[6] or weakly continuous on the unit ball. Hence, for $V_i \subseteq R_1(O_i)$, $\alpha_g^{-1}(V_i)$ is weakly closed, and for finite index set J

$$\alpha_g^{-1}(W) = \bigcup_{i\in J} \alpha_g^{-1}(V_i)$$

is therefore weakly closed, which implies α_g is weakly continuous on Q_1 and thus ultraweakly continuous on Q.

The proof of the theorem is trivial now. From Lemma 3 of Ref. 3 we know that α_g is extendable to $\mathfrak{A}^-$ if α_g is ultraweakly bicontinuous on $\mathfrak{A}$. Since $(\alpha_g)^{-1} = \alpha_{g^{-1}}$, we need only to show that α_g is ultraweakly continuous on $\mathfrak{A}$, which follows directly from Lemmas 1 and 2.

A final remark is given here to show that for $G = R^n$, if automorphism α_x of a quasilocal algebra $\mathfrak{A}$ satisfies (1), then the observable has I-property.

Since α_x satisfies (1), α_x is in a norm-continuous 1-parameter subgroup of Aut $(\mathfrak{A})$.[3] Let $t \to \alpha_{tx}$ be the norm-continuous 1-parameter group. If $A \in Q_1$, then $A \in R_1(O_0)$ for some $O_0 \subset M$. Let O be a finite region of M such that $O_0 \subseteq O$; then $R(O_0) \subseteq R(O)$ and $A \in R_1(O)$. For each $x \in R^n$ and all sufficiently small t, O_0 can be shifted by tx such that $O_0 + tx \subseteq O$; then $\alpha_{tx}[R(O_0)] \subseteq R(O)$. Hence, for each $B \in R(O)'$,

$$[B, \alpha_{tx}(A)] = 0 \tag{4}$$

for small t. Furthermore, due to the norm-continuity of $t \to \alpha_{tx}$, there is a derivation δ of $\mathfrak{A}$ such that $\alpha_{tx} = \exp(t\delta)$. Applying the same method given by Kadison,[7] we obtain

$$[B, \alpha_{tx}(A)] = [B, A] + t[B, \delta(A)] + \tfrac{1}{2}t^2[B, \delta^2(A)] + \cdots = 0$$

for small t, so that $[B, \delta^n(A)] = 0$ for $n = 0, 1, \cdots$. Thus (4) is true for all t and each $x \in R^n$, which implies that $\alpha_{tx}(A) \in R(O)$ for all t and each $x \in R^n$, or more precisely, $\alpha_{tx}(A) \in R_1(O)$ for all t and each $x \in R^n$. Moreover, since $\|\alpha_{tx} - \iota\| < 2$, we have

$$|\langle \xi_i, [\alpha_{tx}(A) - A]\eta_i\rangle| < 2$$

for ξ_i, $\eta_i \in \mathfrak{H}$ with $\|\xi_i\| < 1$, $\|\eta_i\| \leq 1$. Therefore, there is a weak-operator neighborhood $U(A; \xi_i, \eta_i)$ of A given by

$$U(A; \xi_i, \eta_i) = \{\alpha_{tx}(A); |\langle \xi_i, [\alpha_{tx}(A) - A]\eta_i\rangle| < 2 \text{ for all } t \text{ and } x \in R^n, \xi_i, \eta_i \in \mathfrak{H} \text{ with } \|\xi_i\| \leq 1, \|\eta_i\| \leq 1\}$$

which is contained in $R_1(O)$. Thus A is an interior point of $R_1(O)$, or $A \in \mathring{R}_1(O)$ for some $O \subset M$.

Conversely, it is clear to see that I-property does not imply (1).

ACKNOWLEDGMENT

The author is indebted to the referee for helpful suggestions.

[1] J. F. Aarens, Commun. Math. Phys. **7**, 332 (1968).
[2] R. R. Kallman, Commun. Math. Phys. **14**, 13 (1969).
[3] R. V. Kadison and J. R. Ringrose, Commun. Math. Phys. **4**, 32 (1967).
[4] R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).
[5] G. Ludwig, Z. Naturforsch. **22a**, 1303 (1967).
[6] J. Dixmier, *Les algèbres d'opérateurs dans l'espace hilbertien* (Gauthier-Villars, Paris, 1957), Corollaire 1, p. 57.
[7] R. Kadison, Commun. Math. Phys. **4**, 258 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Weight Lowering Operators and the Multiplicity-Free Isoscalar Factors for the Group R_5

S. J. ALIŠAUSKAS AND A. P. JUCYS
Institute of Physics and Mathematics of the Academy of Sciences of Lithuanian SSR, Vilnius, USSR

(Received 19 May 1970)

Expressions for the matrix elements of powers of infinitesimal operators of R_5 are obtained in a basis reduced according to $R_5 \supset (SU_2 \times SU_2)$. Operators producing the basis functions of any weight from those of highest weight are considered. The coupling and recoupling procedures for the semistretched and the other multiplicity-free cases are considered. By use of projection operators the expressions for the isoscalar factors of both kinds of semistretched Clebsch–Gordan coefficients of R_5, as well as a new formula for the 9*j*-coefficient of SU_2 involving only three summation parameters, are obtained. Methods of obtaining the normalized isoscalar factor coupling the basis functions of two symmetrized representations of R_5 are discussed.

I. INTRODUCTION

Several papers[1–5] devoted to the problems of the theory of representations of the group R_5, of interest to physicists, appeared in the 1960's. In Ref. 1 the group R_5 is considered together with the other Lie groups of the second rank. In Ref. 2 some expressions for isoscalar factors (i.f.'s) of the Clebsch–Gordan coefficients (CGC's), in the basis reduced with respect to $SU_2 \times SU_2$, are obtained. In Ref. 3 the polynomial basis for the irreducible representations is derived. In Refs. 3 and 4 the expressions for matrix elements of infinitesimal operators (i.o.'s) are considered. In Ref. 4 the invariants and a family of identities relating various polynomials of degree four in the i.o.'s are found. In Ref. 5 the bases reduced with respect to R_3 are treated.

In Ref. 6 the properties of the basis functions as well as of the i.f.'s under the elements of the substitution group[7] (of parameters labeling the representations) are established; and expressions, in terms of SU_2 quantities, have been found for the stretched and semistretched i.f.'s of the first kind [Eqs. (20) and (29), respectively]. The corresponding expression for the semistretched i.f.'s of the second kind has not been given there, because it involves nine summation parameters and therefore is not of practical importance. The main goal of this paper is to work out more powerful techniques of dealing with the basis functions and i.f.'s, enabling us to obtain expressions for the i.f.'s involving a greatly reduced number of summation parameters.

In the next section we obtain the expressions for the matrix elements of powers of i.o.'s using methods analogous to those used in the theory of SU_3.[8,9] In Sec. III we construct the general weight lowering operator expressed in terms of products of powers of the infinitesimal operators. This appears to be more useful than that expressed in terms of powers of "step-up" and "step-down" operators of Ref. 2, because in our case the results of Sec. II may be applied with great success.

In Sec. IV the standard semistretched coupling schemes are defined together with those nonstandard ones which can be obtained from the standard ones with the help of the elements of the substitution group mentioned above. In Sec. V, by use of the weight lowering operators of Sec. III and of the properties of the quantities of representations of SU_2,[10] the expression for the semistretched i.f.'s of the second kind is deduced, which involves only three summation parameters instead of nine as obtained by the methods used in Ref. 6. A quantity having higher symmetry than the i.f.'s itself is singled out from this expression.

Section VI is devoted to the consideration of the multiplicity-free coupling of two symmetrized vectors as well as of a symmetrized vector and symmetrized spinor, the main attention being paid to the normalization problem. In solving this last problem, the above-mentioned symmetrical quantity was applied in carrying out the summation involved in the special case of the recoupling coefficient needed in the techniques used. In Sec. VII, we examine the semistretched recoupling matrices which are expressible in terms of recoupling matrices of SU_2.

In the Appendix a new expression is deduced for the 9*j* coefficient of SU_2 containing only three summation parameters. The techniques of the theory of representations of R_5 are used for this purpose. This possibility emerges from the fact that quantities of R_5 are expressed in terms of SU_2 as is done to full extent in Ref. 6 as well as in this paper.

II. INFINITESIMAL OPERATORS AND THE MATRIX ELEMENTS OF THEIR POWERS

We take the infinitesimal operators, as is done by Hecht,[2] and express them in terms of those of SU_4

with slight differences in normalization. So we write

$$H_1 = \tfrac{1}{2}(E_{11} - E_{22}), \quad F_{+0} = E_{12}, \quad F_{-0} = E_{21}, \quad \text{(1a)}$$

$$H_2 = \tfrac{1}{2}(E_{33} - E_{44}), \quad F_{0+} = E_{34}, \quad F_{0-} = E_{43}, \quad \text{(1b)}$$

$$T_{++} = -E_{14} - E_{32}, \quad T_{--} = E_{41} + E_{23},$$
$$T_{+-} = E_{13} - E_{42}, \quad T_{-+} = E_{31} - E_{24}. \quad \text{(1c)}$$

Operators (1a) and (1b) are the infinitesimal operators of subgroups SU_2, and those of (1c) constitute the bispinor of R_4, which transforms like a spinor with respect to each SU_2.

The matrix elements of (1a) and (1b) are obtained from the theory of representations of SU_2, and those of (1c) are expressed in terms of CGC's of the subgroups SU_2 and reduced matrix elements obtained from formulas (16) of Ref. 2, omitting the factor $\frac{1}{2}$ and the expression $J_m(J_m + 2) + \Lambda_m(\Lambda_m + 1)$ in the denominator, because of the difference in the normalization and the definition of reduced matrix elements.

Under the term of the power of i.o.'s we understand the coupling of a certain number of equal operators into the tensor of highest rank and the highest (or lowest) projection of this rank. The matrix elements of powers of (1a) and (1b) are known from the theory of SU_2.

The general reduced matrix element is obtained by calculating this matrix element

$$\left\langle \begin{matrix} \langle K\Lambda\rangle \\ I', M + \alpha\, J + \alpha, J + \alpha \end{matrix} \right| T^{2\alpha}_{++} \left| \begin{matrix} \langle K\Lambda\rangle \\ IMJJ \end{matrix} \right\rangle. \quad (2)$$

We obtain the SU_2 states $|IM\rangle$ and $|JN\rangle$ from the states $|II\rangle$ and $|JJ\rangle$ with the help of the operators

$$\left(\frac{(I+M)!}{(2I)!\,(I-M)!}\right)^{\frac{1}{2}} F_{-0}^{I-M}, \quad (3a)$$

$$\left(\frac{(J+N)!}{(2J)!\,(J-N)!}\right)^{\frac{1}{2}} F_{0-}^{J-N}. \quad (3b)$$

F_{-0}^{I-M} in (3a) indicates that the operator F_{-0} is to be applied $I - M$ times, successively, to the state $|II\rangle$. The same holds for (3b).

We make use of the transposition formula

$$T^a_{++}F^b_{-0} = \sum_{u,v} \frac{a!\,b!\,(-1)^v}{u!\,(v-u)!\,(a-u-v)!\,(b-v)!} \times F^{b-v}_{-0} T^{v-u}_{-+} T^{a-u-v}_{++} F^{u}_{0+}, \quad (4)$$

obtained by using the two first equations of the formulas (2.11) of Ref. 8. Further, we use particular cases of (2) with $M = I$ (so that $I' = I + \alpha$) and $M + \alpha = -I'$ (so that $I' = I - \alpha$). Under these conditions the summation in the corresponding expressions disappears.

The result for the reduced matrix element of (2) (after omitting the Clebsch–Gordan coefficients of SU_2) is the following expression:

$$\left\langle \begin{matrix} \langle K\Lambda\rangle \\ I'J + \alpha \end{matrix} \right\| T^{2\alpha} \left\| \begin{matrix} \langle K\Lambda\rangle \\ IJ \end{matrix} \right\rangle = \left(\frac{(2\alpha)!\,(2I+1)(2J+1)!}{(2J'+1)!}\right)^{\frac{1}{2}} \frac{P(K\Lambda I'J')}{P(K\Lambda IJ)\nabla(\alpha II')}. \quad (5)$$

Here $J' = J + \alpha$, and $\nabla(abc)$ is defined by Eq. (16) of Ref. 6. The new quantity P is defined as follows:

$$P(K\Lambda IJ) = \left(\frac{(K+\Lambda+I+J+2)!\,(K+\Lambda-I+J+1)!\,(K-\Lambda+I+J+1)!}{(K+\Lambda+I-J+1)!\,(K+\Lambda-I-J)!} \times \frac{(K-\Lambda-I+J)!\,(I+J+\Lambda-K)!}{(K-\Lambda+I-J)!}\right)^{\frac{1}{2}}. \quad (6)$$

Further, we obtain a formula similar to (4) with F_{0-} instead of F_{-0} and make use of the reduced matrix element stretched with respect to the subgroup SU_2. This enables us to obtain the general expression for the reduced matrix element. It has the form

$$\left\langle \begin{matrix} \langle K\Lambda\rangle \\ I'J' \end{matrix} \right\| T^{2\alpha} \left\| \begin{matrix} \langle K\Lambda\rangle \\ IJ \end{matrix} \right\rangle = \frac{[(2I+1)(2J+1)]^{\frac{1}{2}}(J+J'-\alpha)!\,\nabla(\alpha II')\nabla(\alpha JJ')}{P(K\Lambda I'J')P(K\Lambda IJ)} \times \sum_{i,j} \frac{(-1)^{J-I'+\alpha+i-j}(2i+1)P^2(K\Lambda ij)}{(2j+1)!\,(\alpha+J+J'-2j)!\,\nabla^2(j-J',i,I')\nabla^2(j-J,i,I)}. \quad (7)$$

Unfortunately, the expression (7) is not symmetric with respect to the parameters I and J (as well as I' and J'). It simplifies considerably for the symmetric representations $\langle\Lambda\Lambda\rangle$ or $\langle KO\rangle$.

It is to be taken into account that the contragredient operators are $-T_{++}$ and T_{--}, a property which has been used in comparing the matrix elements of powers of i.o.'s. Moreover, the powers of i.o.'s generally are irreducible tensor operators of SU_2 but not of R_5 itself, as was the case in considering the group SU_3 (cf. Ref. 9).

III. THE GENERAL WEIGHT LOWERING OPERATOR

The formulas (8) of Ref. 2 allow us to change the weight by one step. However, it is very useful to have operators allowing us to deduce the state of any weight from the state of highest weight by one single operation instead of repeated application of some operator many times. The reason is that the separate parts of "step-up" and "step-down" operators do not commute, and repeated application gives very awkward expressions.

For SU_3 such an operator is constructed in the form of a polynomial in i.o.'s. For R_5 it is more difficult to do in this way, because the coefficients in the polynomial are very complicated.

It is easy to find states satisfying the condition $I + J = K + \Lambda$, by acting on the state of highest weight with the operator $T_{-+}^{2\alpha}$, the result being a state of nondegenerate weight. Consequently,

$$\left|\begin{matrix}\langle K\Lambda\rangle \\ K-\alpha, K-\alpha, \Lambda+\alpha, \Lambda+\alpha\end{matrix}\right\rangle = \left(\frac{(2K-2\Lambda-2\alpha)!}{(2\alpha)!\,(2K-2\Lambda)!}\right)^{\frac{1}{2}} T_{-+}^{2\alpha} \left|\begin{matrix}\langle K\Lambda\rangle \\ KK\Lambda\Lambda\end{matrix}\right\rangle \quad (8)$$

since, in this special case, T_{-+} acts on the state as i.o.'s of SU_2.

The next step is to construct the states with arbitrary I and J. First, we observe that the following expansion holds:

$$\left|\begin{matrix}\langle K\Lambda\rangle \\ IIJJ\end{matrix}\right\rangle = \sum_{\alpha,\beta} Q[\langle K\Lambda\rangle IJ, \alpha\beta] F_{-0}^{K-\alpha-\beta-I} F_{0-}^{\Lambda+\alpha-\beta-J} T_{--}^{2\beta} \times \left|\begin{matrix}\langle K\Lambda\rangle \\ K-\alpha, K-\alpha, \Lambda+\alpha, \Lambda+\alpha\end{matrix}\right\rangle. \quad (9)$$

In order to find the coefficients Q in (9), we operate with the step-up operators F_{+0} or F_{0+} on both sides of this last equation. The result on the left-hand side must be zero. By the use of commutators of F_{+0} and F_{0+} with the operators of the right-hand side of (9), we obtain recurrence formulas for Q, which together with the boundary conditions (the degree of the polynomial in i.o.'s cannot be negative) show that the coefficients Q must be proportional to the expression

$$[(2\beta)!]^{-1}[(K+I-\alpha-\beta+1)!\,(K-I-\alpha-\beta)!\,(\Lambda+J+\alpha-\beta+1)!\,(\Lambda-J+\alpha-\beta)!]^{-\frac{1}{2}} \times \begin{bmatrix} I & J & K-\Lambda \\ K-\alpha-\beta+1 & -\Lambda-\alpha+\beta-1 & K-\Lambda-2\alpha \end{bmatrix}. \quad (10)$$

Here, the quantity in the square bracket is a nonstandard CGC of SU_2, because the absolute values of two of the projections exceed the values of the corresponding angular momenta. For this CGC we use the expression of Van der Waerden [the first of Eqs. (13.1) of Ref. 10] and normalize (10). For this purpose, we multiply (9) by the state contragredient to the state of the left-hand side, the result being equal to unity, and carry out the calculation of matrix elements of the right-hand side. Then by the use of (3) we obtain

$$\begin{aligned}\left|\begin{matrix}\langle K\Lambda\rangle \\ IMJN\end{matrix}\right\rangle &= B[\langle K\Lambda\rangle IJ]\nabla(K-\Lambda, I, J)\frac{(I+J-K+\Lambda)!}{(I+J+K-\Lambda+1)!} \\ &\times \left(\frac{(2I+1)(2J+1)(I+M)!\,(J+N)!\,(2K-2\Lambda)!}{(I-M)!\,(J-N)!}\right)^{\frac{1}{2}} \\ &\times \sum_{\alpha,\beta,u} \frac{[(2\alpha)!\,(2K-2\Lambda-2\alpha)!]^{\frac{1}{2}}(-1)^{2\beta}}{(2\beta)!\,(K-I-\alpha-\beta)!\,(\Lambda-J+\alpha-\beta)!\,u!\,(2\alpha-u)!} \\ &\times \frac{(K-I-\alpha-\beta+u)!}{(K-\Lambda+I-J-u)!\,(K-\Lambda-I+J-2\alpha+u)!\,(\Lambda+J-\alpha-\beta+u+1)!} \\ &\times F_{-0}^{K-M-\alpha-\beta} F_{0-}^{\Lambda-N+\alpha-\beta} T_{--}^{2\beta} \left|\begin{matrix}\langle K\Lambda\rangle \\ K-\alpha, K-\alpha, \Lambda+\alpha, \Lambda+\alpha\end{matrix}\right\rangle, \end{aligned} \quad (11)$$

$B[\langle K\Lambda\rangle IJ]$ being defined by Eq. (15) of Ref. 6.

The expression (11) can be obtained in another form by using other expressions for the nonstandard CGC of SU_2 in (9). One of those forms, useful in practice, is obtained by using the Fock–Racah expression [the second of Eqs. (13.1) of Ref. 10] instead of that used above. The connection between this new

formula and (11) can be expressed by this summation relation

$$\sum_u \frac{(A+u)!}{u!\,(B-u)!\,(C-u)!\,(D+u)!\,(S+u+1)!} = \frac{1}{B!\,(B+D)!\,(S-A)!\,(S+C+1)!} \times \sum_u \frac{(A+u)!\,(B+D+u)!\,(S+C-A-u)!}{u!\,(C-u)!\,(D+u)!}. \tag{12}$$

This formula will be of use in carrying out the summations appearing in the process of deriving the expressions for isoscalar factors of CGC of R_5 to be dealt with in what follows.

IV. ON THE SEMISTRETCHED COUPLING SCHEMES

We consider multiplicity-free terms in the direct products of two irreducible representations of particular coupling schemes, called semistretched of the first and second kinds, as defined by Eqs. (27) and (28) of Ref. 6, respectively. The corresponding isoscalar factors we call standard semistretched of the first kind (s.s.i.f.'s I) and of the second kind (s.s.i.f.'s II), respectively. For the fully stretched coupling scheme the conditions (27) and (28) of Ref. 6 are satisfied simultaneously.

We suppose that the symmetry conditions (A.22) of Ref. 2 when applied to the s.s.i.f.'s give the s.s.i.f.'s as well. They remain s.s.i.f.'s after transposing the factors in the direct product according to this relation

$$\begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K\Lambda\rangle \\ I_1J_1 & I_2J_2 & IJ \end{bmatrix} = (-1)^{K-\Lambda-K_1+\Lambda_1-K_2+\Lambda_2+I_1+I_2-I+J_1+J_2-J} \times \begin{bmatrix} \langle K_2\Lambda_2\rangle & \langle K_1\Lambda_1\rangle & \langle K\Lambda\rangle \\ I_2J_2 & I_1J_1 & IJ \end{bmatrix} \tag{13}$$

as well as

$$\begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K\Lambda\rangle \\ I_1J_1 & I_2J_2 & IJ \end{bmatrix} = (-1)^{2\Lambda_1+2\Lambda_2+2\Lambda} \begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K\Lambda\rangle \\ J_1I_1 & J_2I_2 & JI \end{bmatrix}. \tag{14}$$

On the other hand, using the relations (21) of Ref. 6, with respect to the elements of the substitution group, we can obtain nonstandard semistretched i.f.'s (n.s.s.i.f.'s) of both kinds. The parameters of the n.s.s.i.f.'s I satisfy the condition

$$K_1 - \Lambda_1 + K_2 + \Lambda_2 = K - \Lambda \tag{15}$$

and those of the n.s.s.i.f.'s II,

$$\Lambda_1 + K_2 = \Lambda. \tag{16}$$

It is easy to prove that under four of the eight elements of the substitution group, generated by the elements (3) of Ref. 6, the s.s.i.f.'s I transform into n.s.s.i.f.'s I and s.s.i.f.'s II into n.s.s.i.f.'s II. Among these four elements only two are common. The other four elements do not change the character of s.s.i.f.'s.

States of any number of representations can be coupled by the semistretched coupling scheme. For example, in the case of three representations the states of the resulting representation are expressed in this way

$$\left| \begin{matrix} \langle K_1\Lambda_1, K_2\Lambda_2, K_3\Lambda_3, K_{12}\Lambda_{12}; K\Lambda\rangle \\ IMJN \end{matrix} \right\rangle = \sum_{\substack{I_iM_i \\ J_iN_i}} \begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K_3\Lambda_3\rangle & \langle K\Lambda\rangle \\ I_1M_1J_1N_1 & I_2M_2J_2N_2 & I_3M_3J_3N_3 & IMJN \end{bmatrix}_{\langle K_{12}\Lambda_{12}\rangle} \times \left| \begin{matrix} \langle K_1\Lambda_1\rangle \\ I_1M_1J_1N_1 \end{matrix} \right\rangle \left| \begin{matrix} \langle K_2\Lambda_2\rangle \\ I_2M_2J_2N_2 \end{matrix} \right\rangle \left| \begin{matrix} \langle K_3\Lambda_3\rangle \\ I_3M_3J_3N_3 \end{matrix} \right\rangle. \tag{17}$$

Here

$$\begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K_3\Lambda_3\rangle & \langle K\Lambda\rangle \\ I_1M_1J_1N_1 & I_2M_2J_2N_2 & I_3M_3J_3N_3 & IMJN \end{bmatrix}_{\langle K_{12}\Lambda_{12}\rangle} = \sum_{\substack{I_{12}M_{12} \\ J_{12}N_{12}}} \begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K_{12}\Lambda_{12}\rangle \\ I_1M_1J_1N_1 & I_2M_2J_2N_2 & I_{12}M_{12}J_{12}N_{12} \end{bmatrix} \begin{bmatrix} \langle K_{12}\Lambda_{12}\rangle & \langle K_3\Lambda_3\rangle & \langle K\Lambda\rangle \\ I_{12}M_{12}J_{12}N_{12} & I_3M_3J_3N_3 & IMJN \end{bmatrix} \tag{18}$$

is a generalized CGC of R_5 in analogy to the case of SU_2, as given by Eq. (21.7) of Ref. 10. In this particular case, the parameters labeling the repeated representation are not needed. The parameters labeling representations must satisfy the conditions

$$\sum_{i=1}^{3}(K_i + \Lambda_i) = K + \Lambda, \quad \sum_{i=1}^{2}(K_i + \Lambda_i) = K_{12} + \Lambda_{12} \tag{19}$$

for the semistretched coupling scheme of the first kind, and

$$\sum_{i=1}^{3} K_i = K, \quad K_1 + K_2 = K_{12} \tag{20}$$

for that of the second kind.

Equations (19) and (20) represent the generalization of (27) and (28) of Ref. 6. They may be easily generalized to the case of any number of representations

whose states are to be coupled by the semistretched coupling schemes.

The expression for the s.s.i.f.'s I is given by Eq. (29) of Ref. 6. It is worth mentioning that this formula, as well as (20) of Ref. 6, giving the expression for the stretched i.f., can be deduced more easily than in Ref. 6 by taking fixed values of some arbitrary parameters in (19) and the analogous formulas. For example, taking $i_1 = i_2 = 0$ in (19) of Ref. 6, we see that all other parameters, including summation parameters, are uniquely defined, and the question of using the summation formulas no longer arises.

V. THE SEMISTRETCHED ISOSCALAR FACTORS OF THE SECOND KIND

In this section we obtain the expressions for s.s.i.f.'s II. At first we obtain the formula

$$\begin{bmatrix} \langle\Lambda_1\Lambda_1\rangle & \langle\Lambda_2\Lambda_2\rangle & \langle\Lambda_1+\Lambda_2,\Lambda\rangle \\ I_1I_1 & I_2I_2 & I'J' \end{bmatrix}$$
$$= (-1)^{2\Lambda_1+2I_1}\frac{2\nabla(I'I_1I_2)\nabla(J'I_1I_2)}{\nabla(\Lambda\Lambda_1\Lambda_2)}\left(\frac{(4\Lambda_1+1)!\,(4\Lambda_2+1)!\,(2\Lambda+1)!}{(2\Lambda_1)!\,(2\Lambda_2)!\,(2\Lambda_1+2\Lambda_2-2\Lambda)!}\right.$$
$$\left.\times\frac{(2I_1+1)(2I_2+1)(2I'-2\Lambda)!\,(2\Lambda_1+2\Lambda_2-2I')!}{(2\Lambda_1-2I_1)!\,(2\Lambda_1+2I_1+2)!\,(2\Lambda_2-2I_2)!\,(2\Lambda_2+2I_2+2)!\,(2I'+1)!\,(2J'+1)!}\right)^{\frac{1}{2}}. \quad (21)$$

Here

$$I'+J'=\Lambda_1+\Lambda_2+\Lambda;$$

Eq. (21) is deduced by acting on the state of highest weight of the representation $\langle\Lambda_1+\Lambda_2,\Lambda)$ with the operator (8) and taking the product with the functions

$$\left|\begin{matrix}\langle\Lambda_1\Lambda_1\rangle \\ I_1I_1I_1, J'-I_2\end{matrix}\right\rangle,\quad \left|\begin{matrix}\langle\Lambda_2\Lambda_2\rangle \\ I_2, I'-I_1, I_2I_2\end{matrix}\right\rangle.$$

Trivial rearrangements give (21).

The next formula to be derived is

$$\begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K_1+K_2,\Lambda\rangle \\ I_1J_1 & I_2J_2 & I'J' \end{bmatrix}$$
$$= (-1)^{K_1+\Lambda_1-I_1-J_1}\frac{[(2I_1+1)(2J_1+1)(2I_2+1)(2J_2+1)]^{\frac{1}{2}}}{B[\langle K_1\Lambda_1\rangle I_1J_1]B[\langle K_2\Lambda_2\rangle I_2J_2]}$$
$$\times\frac{\nabla(I'I_1I_2)\nabla(J'J_1J_2)}{\nabla(K_1-\Lambda_1,I_1,J_1)\nabla(K_2-\Lambda_2,I_2,J_2)\nabla(\Lambda\Lambda_1\Lambda_2)}$$
$$\times\left(\frac{(2\Lambda+1)!\,(2K_1+2K_2-2I')!\,(2I'-2\Lambda)!}{(2\Lambda_1)!\,(2\Lambda_2)!\,(2K_1+2K_2-2\Lambda)!\,(2I'+1)!\,(2J'+1)!}\right)^{\frac{1}{2}},\quad I'+J'=K_1+K_2+\Lambda. \quad (22)$$

For its derivation we use (21) and a formula analogous to (19) of Ref. 6.

By the use of (9) for the s.s.i.f.'s II of the general form, we find the expression

$$\begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K_1+K_2,\Lambda\rangle \\ I_1J_1 & I_2J_2 & IJ \end{bmatrix} = \begin{bmatrix} I_1 & I_2 & I \\ I_1 & I-I_1 & I \end{bmatrix}^{-1}\begin{bmatrix} J_1 & J_2 & J \\ J-J_2 & J_2 & J \end{bmatrix}^{-1}$$
$$\times\sum Q[\langle K_1+K_2,\Lambda\rangle IJ,\alpha\beta]\frac{(2\beta)!}{(2i_1-2I_1)!\,(2j_2-2J_2)!}$$
$$\times\begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K_1+K_2,\Lambda\rangle \\ i_1j_1 & i_2j_2 & I'J' \end{bmatrix}\begin{bmatrix} i_1 & i_2 & I' \\ i_1 & I-i_1 & I' \end{bmatrix}\begin{bmatrix} j_1 & j_2 & J' \\ J'-j_2 & j_2 & J' \end{bmatrix}$$
$$\times\left\langle\begin{matrix}\langle K_1\Lambda_1\rangle \\ I_1I_1J_1, J-J_2\end{matrix}\right| F_{0-}^{\Lambda+\alpha-\beta-J}T_{--}^{2(i_1-I_1)}\left|\begin{matrix}\langle K_1\Lambda_1\rangle \\ i_1i_1j_1, J'-j_2\end{matrix}\right\rangle$$
$$\times\left\langle\begin{matrix}\langle K_2\Lambda_2\rangle \\ I_2, I-I_1, J_2J_2\end{matrix}\right| F_{-0}^{K_1+K_2-\alpha-\beta-I}T_{--}^{2(j_2-J_2)}\left|\begin{matrix}\langle K_2\Lambda_2\rangle \\ i_2, I'-i_1, j_2, j_2\end{matrix}\right\rangle. \quad (23)$$

The summation is carried out with respect to α, β, i_1, i_2, j_1, j_2, I', and J' satisfying the conditions $I' + J' = K_1 + K_2 + \Lambda$, $\alpha = J' - \Lambda$, and $\beta = i_1 + j_2 - I_1 - J_2$. These together with one summation parameter involved in Q, according to formula (23), give six linearly independent summation parameters.

After substituting the expressions for CGC's of SU_2 and those for submatrix elements into (23), we obtain a rather bulky expression. In order to eliminate the factors $2j_1 + 1$ and $2i_2 + 1$, which are awkward in the summation process, we use the relation

$$\sum_{i_{12}} \frac{(-1)^{i_{12}}(2i_{12}+1)(i_{12}+m_{12})!}{(i_{12}-m_{12})!\,\nabla^2(i_2 i_1 i_{12})\nabla^2(i_3 i i_{12})} = \left(\frac{(i_1 - i_2 + m_{12})!\,(i - i_3 + m_{12})!}{(2i+1)(2i_2)!\,(2i_3)!\,(i_1 + i_2 - m_{12})!\,(i + i_3 - m_{12})!}\right)^{\frac{1}{2}} \times \frac{(-1)^{i-i_3}}{\nabla(i_2 + i_3, i_1, i)} \begin{bmatrix} i_2 + i_3 & i_1 & i \\ i_2 - i_3 & m_{12} - i_2 & m_{12} - i_3 \end{bmatrix} \tag{24}$$

following from the equality

$$\sum_{i_{12}} \begin{bmatrix} i_1 & i_2 & i_{12} \\ m_{12} - i_2 & i_2 & m_{12} \end{bmatrix} \begin{bmatrix} i_{12} & i_3 & i \\ m_{12} & -i_3 & m_{12} - i_3 \end{bmatrix} \langle i_1 i_2 (i_{12}) i_3 i \mid i_3 i_2 (i_2 + i_3) i_1 i \rangle = \begin{bmatrix} i_3 & i_2 & i_2 + i_3 \\ -i_3 & i_2 & i_2 - i_3 \end{bmatrix} \begin{bmatrix} i_2 + i_3 & i_1 & i \\ i_2 - i_3 & m_{12} - i_2 & m_{12} - i_3 \end{bmatrix}. \tag{25}$$

For the expressions of the CGC's of SU_2 in (24), we select ones such that some of the summation formulas of Eq. (14.2)–(14.5) of Ref. 10 are applicable to the resulting expression. The result is

$$\begin{aligned}
&\begin{bmatrix} \langle K_1 \Lambda_1 \rangle & \langle K_2 \Lambda_2 \rangle & \langle K_1 + K_2, \Lambda \rangle \\ I_1 J_1 & I_2 J_2 & IJ \end{bmatrix} \\
&= \left(\frac{(2I_1+1)(2J_1+1)(2I_2+1)(2J_2+1)(2\Lambda)!}{(2\Lambda_1)!\,(2\Lambda_2)!}\right)^{\frac{1}{2}} \\
&\quad \times \frac{B[\langle K_1 + K_2, \Lambda \rangle IJ]\nabla(K_1 + K_2 - \Lambda, I, J)\nabla(I_1 I_2 I)\nabla(J_2 J_1 J)}{B[\langle K_1 \Lambda_1 \rangle I_1 J_1]B[\langle K_2 \Lambda_2 \rangle I_2 J_2]\nabla(K_1 - \Lambda_1, I_1, J_1)\nabla(K_2 - \Lambda_2, I_2, J_2)} \\
&\quad \times \sum_{u,v} (-1)^{K_1 - \Lambda_1 + K_2 - \Lambda_2 - J_1 - J_2 - I + u + v} \begin{bmatrix} \Lambda_1 & \Lambda_2 & \Lambda \\ u - \Lambda_1 & v - \Lambda_2 & u + v - \Lambda_1 - \Lambda_2 \end{bmatrix} \\
&\quad \times \left(\frac{(2\Lambda_1 - u)!\,(2\Lambda_2 - v)!\,(\Lambda - \Lambda_1 - \Lambda_2 + u + v)!}{u!\,v!\,(\Lambda_1 + \Lambda_2 + \Lambda - u - v)!}\right)^{\frac{1}{2}} \\
&\quad \times \frac{(K_1 - \Lambda_1 - I_1 + J_1 + u)!}{(I_1 + J_1 - K_1 + \Lambda_1 - u)!\,(K_1 - \Lambda_1 - I_1 + J_2 - J + u)!\,(K_1 - \Lambda_1 - I_1 + J_2 + J + u + 1)!} \\
&\quad \times \frac{(K_2 - \Lambda_2 + I_2 - J_2 + v)!}{(I_2 + J_2 - K_2 + \Lambda_2 - v)!\,(K_2 - \Lambda_2 + I_1 - J_2 - I + v)!\,(K_2 - \Lambda_2 + I_1 - J_2 + I + v + 1)!}.
\end{aligned} \tag{26}$$

Here one summation parameter is involved in the CGC of SU_2. Using the appropriate expression for this CGC, we can bring the sum to the form

$$\begin{aligned}
\sum_{u,v} \cdots &= \frac{(2\Lambda+1)^{\frac{1}{2}}}{\nabla(\Lambda\Lambda_1\Lambda_2)} \sum_{\alpha,\beta,z} \frac{(-1)^{\Lambda_1 + \Lambda_2 - \Lambda + I_1 + I_2 - I + \alpha + \beta + z}}{z!\,(\Lambda_1 + \Lambda_2 - \Lambda - z)!\,\alpha!\,(J_1 + J_2 - J - \alpha)!} \\
&\quad \times \frac{(2\Lambda_1 - z)!\,(\Lambda - \Lambda_1 + \Lambda_2 + z)!\,(2J_1 - \alpha)!}{(I_1 + J_1 - K_1 + \Lambda_1 - \alpha - z)!\,(J_1 + J_2 + J - \alpha + 1)!\,(I_2 + J_2 - K_2 - \Lambda_1 + \Lambda - \beta + z)!} \\
&\quad \times \frac{(I_1 + I_2 + J_1 + J_2 - K_1 - K_2 + \Lambda - \alpha - \beta)!}{\beta!\,(I_1 + I_2 - I - \beta)!\,(I_1 + I_2 + I - \beta + 1)!}.
\end{aligned} \tag{27}$$

Now let us consider the particular case of (26) and (27) when $\Lambda = \Lambda_1 + \Lambda_2$. Then the summation with respect to z falls off. Moreover, we then obtain an expression for the stretched i.f. which, according to formula (20) of Ref. 6, is proportional to the stretched $9j$ coefficient of SU_2. This gives us a new expression for the stretched $9j$ coefficient involving two summation parameters, as in Eq. (25.20) of Ref. 10. However, the

$$z!(\Lambda_1 + \Lambda_2 - z)!\alpha!(J_1 + J_2 - J - \alpha)!$$

expressions are quite different, and the transformation of one into the other is rather tedious. Using the new expression for the $9j$ coefficient, we conceal the summation with respect to α and β in (27) in the corresponding stretched $9j$ coefficient. Afterwards, we use Eq. (25.20) of Ref. 10 for this $9j$ coefficient and obtain the expression

$$\begin{bmatrix} \langle K_1\Lambda_1\rangle & \langle K_2\Lambda_2\rangle & \langle K_1+K_2,\Lambda\rangle \\ I_1J_1 & I_2J_2 & IJ \end{bmatrix}$$
$$= (-1)^{\Lambda_1+\Lambda_2-\Lambda}[(2I_1+1)(2J_1+1)(2I_2+1)(2J_2+1)(2\Lambda+1)]^{\frac{1}{2}}$$
$$\times \left(\frac{(2K_1-2\Lambda_1)!\,(2K_1+1)!\,(2K_1+2\Lambda_1+2)!\,(2K_2-2\Lambda_2)!\,(2K_2+1)!\,(2K_2+2\Lambda_2+2)!}{(2K_1+2K_2-2\Lambda)!\,(2K_1+2K_2+1)!\,(2K_1+2K_2+2\Lambda+2)!}\right)^{\frac{1}{2}}$$
$$\times \left[\begin{array}{c|ccc} K_1 & \Lambda_1 & I_1 & J_1 \\ K_2 & \Lambda_2 & I_2 & J_2 \\ K_1+K_2 & \Lambda & I & J \end{array}\right]. \tag{28}$$

Here

$$\left[\begin{array}{c|ccc} K_1 & \Lambda_1 & I_1 & J_1 \\ K_2 & \Lambda_2 & I_2 & J_2 \\ K_1+K_2 & \Lambda & I & J \end{array}\right]$$
$$= \frac{B[\langle K_1+K_2,\Lambda\rangle IJ]}{B[\langle K_1\Lambda_1\rangle I_1J_1]B[\langle K_2\Lambda_2\rangle I_2J_2]}$$
$$\times \frac{\nabla(K_1+K_2-\Lambda,I,J)(I+J-K_1-K_2+\Lambda)!}{\nabla(K_1-\Lambda_1,I_1,J_1)\nabla(K_2-\Lambda_2,I_2,J_2)\nabla(II_1I_2)\nabla(JJ_1J_2)\nabla(\Lambda\Lambda_1\Lambda_2)}$$
$$\times \left(\frac{(2K_1+2K_2-2\Lambda)!\,(2\Lambda)!\,(2K_1+2K_2+1)!\,(2K_1+2K_2+2\Lambda+2)!}{(2K_1-2\Lambda_1)!\,(2\Lambda_1)!\,(2K_1+1)!\,(2K_1+2\Lambda_1+2)!\,(2K_2-2\Lambda_2)!\,(2\Lambda_2)!\,(2K_2+1)!\,(2K_2+2\Lambda_2+2)!}\right)^{\frac{1}{2}}$$
$$\times \sum_{x,y,z} \frac{(-1)^{x+y+z}(2I_1-x)!\,(I-I_1+I_2+x)!\,(2J_1-y)!}{x!\,(I_1+I_2-I-x)!\,y!\,(J_1+J_2-J-y)!\,z!\,(\Lambda_1+\Lambda_2-\Lambda-z)!}$$
$$\times \frac{(J-J_1+J_2+y)!\,(2\Lambda_1-z)!\,(\Lambda-\Lambda_1+\Lambda_2+z)!}{(I_1+J_1-K_1+\Lambda_1-x-y-z)!\,(I+J+\Lambda-\Lambda_1-I_1-J_1-K_2+x+y+z)!}. \tag{29}$$

This quantity, which we call $11j$ coefficient and which does not belong to the class of $3nj$ coefficients, is more symmetric than the i.f. itself. Permutation of the last three columns does not change its value nor the phase. Transposition of the first two rows gives the phase factor

$$(-1)^{\Lambda_1+\Lambda_2-\Lambda+I_1+I_2-I+J_1+J_2-J}. \tag{30a}$$

Application of the elements of the substitution group allows us to deduce other forms of the expressions for the s.s.i.f.'s as well as for the n.s.s.i.f.'s. For example, using Eqs. (21b) and (21c) of Ref. 6 and applying (A.22) of Ref. 2, we obtain expressions for the s.s.i.f.'s which are less symmetric than (28), but are very useful for practical purposes. By comparing the two expressions for the same i.f., we deduce this property of the $11j$ coefficient:

$$\left[\begin{array}{c|ccc} K_1 & I_1 & J_1 & \Lambda_1 \\ K_2 & I_2 & J_2 & \Lambda_2 \\ K_1+K_2 & I_3 & J_3 & \Lambda_3 \end{array}\right]$$
$$= (-1)^{I_2+J_2+\Lambda_2-K_2}\left[\begin{array}{c|ccc} -K_1-K_2-2 & I_3 & J_3 & \Lambda_3 \\ K_2 & I_2 & J_2 & \Lambda_2 \\ -K_2-2 & I_1 & J_1 & \Lambda_1 \end{array}\right]. \tag{30b}$$

By the way, it is worth mentioning that mirror reflection substitutions ($X \to -X-1$, $X = I_i, J_i, \Lambda_i$, $i = 1, 2, 3$) change only the phase factor of the $11j$ coefficient.

VI. ISOSCALAR FACTOR COUPLING TWO SYMMETRIC REPRESENTATIONS

For this section there is left to consider the coupling procedure of two symmetrized vectors as well as that of the symmetrized vector with the symmetrized spinor, the case of two symmetrized spinors having been considered in Eq. (26) of Ref. 6. The direct products under consideration are multiplicity free, as is quite easy to prove by graphical methods similar to those used by Speiser[11] for SU_3 or by the algebraic methods of Bucella and Cattani.[12]

In the case of the direct product of two symmetrized vectors $\langle\Lambda_1\Lambda_1\rangle$ and $\langle\Lambda_2\Lambda_2\rangle$, reduced to the representations $\langle K\Lambda\rangle$, the parameters satisfy the conditions

$$K-\Lambda\geq 0,\quad \Lambda_1+\Lambda_2-K\geq 0,$$
$$\Lambda_1-\Lambda_2+\Lambda\geq 0,\quad \Lambda_2-\Lambda_1+\Lambda\geq 0, \tag{31}$$

constituting the triangular region. In the case of the product of a symmetrized vector $\langle\lambda\lambda\rangle$ with a symmetrized spinor $\langle k0\rangle$ when reduced to $\langle K\Lambda\rangle$, we have

$$2\lambda-2\Lambda\geq 0,\qquad 2K-2\lambda\geq 0,$$
$$K+\Lambda-k\geq 0,\quad k-K+\Lambda\geq 0, \tag{32}$$

the parameters constituting a quadrangular region which in particular cases degenerates into a triangular one. The linear combinations of parameters in (31) and (32) must be integers.

The expressions for i.f.'s under consideration can be found by the methods of this paper or by those of Ref. 6. However, the more difficult task is the normalization problem. We consider in detail particular cases of i.f.'s for demonstrating the derivation of expressions and normalization procedures. The general cases can be derived from the special ones by the methods of Sec. 5, in a way similar to that used for Eq. (28), the corresponding expressions being too bulky to be given here.

We consider the special case of coupling two symmetrized vectors in which $I'+J'=K+\Lambda$. Then we obtain

$$\begin{bmatrix}\langle\Lambda_1\Lambda_1\rangle & \langle\Lambda_2\Lambda_2\rangle & \langle K\Lambda\rangle\\ I_1I_1 & I_2I_2 & I'J'\end{bmatrix}$$
$$= N(\Lambda_1\Lambda_2,K\Lambda)(-1)^{\Lambda_1-\Lambda_2+\Lambda+I_1-I_2+I}$$
$$\times[(2I_1+1)(2I_2+1)(2\Lambda_1-2I_1)!\,(2\Lambda_1+2I_1+2)!\,(2\Lambda_2-2I_2)!\,(2\Lambda_2+2I_2+2)!]^{\frac{1}{2}}$$
$$\times\nabla(I'I_1I_2)\nabla(J'I_1I_2)\left(\frac{(2K-2I')!\,(2I'-2\Lambda)!}{(2I'+1)!\,(2J'+1)!}\right)^{\frac{1}{2}}$$
$$\times\sum_{u,v,z}\frac{(-1)^{u+v+z}(I_1+I_2+I'-2\Lambda+u+v-z)!}{u!\,(2\Lambda_1-2I_1-u)!\,(2I_1+u+1)!\,(2\Lambda_1-u+1)!\,v!\,(2\Lambda_2-2I_2-v)!}$$
$$\times\frac{(2\Lambda_1+2\Lambda_2-I_1-I_2-I'-u-v+z)!}{(2I_2+v+1)!\,(2\Lambda_2-v+1)!\,z!\,(K-\Lambda-z)!\,(2I'-2\Lambda-z)!\,(K+\Lambda-2I'+z)!}. \tag{33}$$

Here

$$N(\Lambda_1\Lambda_2,K\Lambda)=N_0(\Lambda_1\Lambda_2,K\Lambda)(2\Lambda_1+1)!\,(2\Lambda_2+1)!\,(K-\Lambda)!$$
$$\times\left(\frac{(2K-2\Lambda+1)(K+\Lambda-2\Lambda_1+1)(K+\Lambda-2\Lambda_1+2\Lambda_2+3)(2\Lambda_1+2\Lambda_2-2K)!}{2(2\Lambda_1+2\Lambda_2-K-\Lambda+1)(4\Lambda_1+1)!\,(4\Lambda_2+3)!\,(2\Lambda_1+2\Lambda_2-2\Lambda+1)!}\right)^{\frac{1}{2}}. \tag{34}$$

The expression (33) is deduced with the help of recurrence relations in a way similar to that used in Ref. 6. This means that the basis of the representation $\langle\Lambda_1\Lambda_1\rangle$ is to be constructed from two bases. These last ones are taken to be $\langle\Lambda_1,\Lambda_1-k\rangle$ and $\langle k0\rangle$. Afterwards, the resultant representation is to be constructed by the coupling scheme

$$\langle\Lambda_1,\Lambda_1-k\rangle,\ \langle\Lambda_2\Lambda_2\rangle\langle k0\rangle(\langle\Lambda_2,\Lambda_2-k\rangle)\langle K\Lambda\rangle.$$

Here k satisfies the condition $2k=2\Lambda_1+2\Lambda_2-K-\Lambda$. Hence, all couplings are stretched or semi-stretched.

The normalization factor $N_0(\Lambda_1\Lambda_2,K\Lambda)$ is the quantity reciprocal to the element of the recoupling matrix of R_5 given below in (38). In order to find it, we rearrange the sum on the right-hand side of (33). At first we reveal the fact that the $11j$ coefficient appears as defined by Eq. (29):

$$\left[\begin{array}{c|ccc}-\frac{1}{2}(K+\Lambda+3) & -I_1-1 & -I_2-1 & I'-\frac{1}{2}(K+\Lambda+1)\\ -\frac{1}{2}(K+\Lambda+3) & -I_1-1 & -I_2-1 & I'-\frac{1}{2}(K+\Lambda+1)\\ -K-\Lambda-3 & -2\Lambda_1-2 & -2\Lambda_2-2 & -K+\Lambda-1\end{array}\right]. \tag{35}$$

In (29) we apply the substitutions

$$K, \Lambda, I, J \to -K-2, -\Lambda-1, I, -J-1,$$
$$K_1, \Lambda_1, I_1, J_1 \to -K_1-2, -\Lambda_1-1, I_1, -J_1-1, \tag{36}$$

and obtain a new expression for the $11j$ coefficient. Then, we express the $11j$ coefficient (35) with the help of this new formula. Comparing this new expression with the old one, we replace the sum in (33) by

$$\frac{(-1)^{2\Lambda-2I'}(2\Lambda_1+2\Lambda_2-2\Lambda+1)!\,(I_1+I_2-I')!}{(2\Lambda_1+1)!\,(2\Lambda_2+1)!\,(K-\Lambda)!\,(2K-2I')!\,(2\Lambda_1+2I_1+2)!\,(2\Lambda_2+2I_2+2)!}$$
$$\times \sum_{x,y,z} \frac{(-1)^{x+y+z}(4\Lambda_1+2-x)!\,(4\Lambda_2+2-y)!\,(2K-2\Lambda-z)!}{x!\,(2\Lambda_1-2I_1-x)!\,(2\Lambda_1-x+1)!\,y!\,(2\Lambda_2-2I_2-y)!\,(2\Lambda_2-y+1)!}$$
$$\times \frac{(2\Lambda_1+2\Lambda_2-2\Lambda-I_1-I_2+I'-x-y-z)!}{z!\,(K-\Lambda-z)!\,(2I'-2\Lambda-z)!\,(2\Lambda_1+2\Lambda_2-2\Lambda-x-y-z+1)!}. \tag{37}$$

For finding the recoupling matrix mentioned above, we use the relation

$$\langle\langle\Lambda_1, \Lambda_1-k\rangle\langle k0\rangle(\langle\Lambda_1\Lambda_1\rangle)\langle\Lambda_2\Lambda_2\rangle\langle K\Lambda\rangle \,|\, \langle\Lambda_1, \Lambda_1-k\rangle\langle\Lambda_2\Lambda_2\rangle\langle k0\rangle(\langle\Lambda_2, \Lambda_2-k\rangle)\langle K\Lambda\rangle\rangle$$
$$\times \begin{bmatrix} \langle\Lambda_1, \Lambda_1-k\rangle & \langle\Lambda_2, \Lambda_2-k\rangle & \langle K\Lambda\rangle \\ \Lambda_1, \Lambda_1-k & \Lambda_2-k, \Lambda_2 & \tfrac{1}{2}(K+\Lambda), \tfrac{1}{2}(K+\Lambda) \end{bmatrix}$$
$$= \sum_{i+j=k} \begin{bmatrix} \langle\Lambda_1, \Lambda_1-k\rangle & \langle k0\rangle & \langle\Lambda_1\Lambda_1\rangle \\ \Lambda_1, \Lambda_1-k & ij & \Lambda_1-i, \Lambda_1-i \end{bmatrix}$$
$$\times \begin{bmatrix} \langle\Lambda_1\Lambda_1\rangle & \langle\Lambda_2\Lambda_2\rangle & \langle K\Lambda\rangle \\ \Lambda_1-i, \Lambda_1-i & \Lambda_2-j, \Lambda_2-j & \tfrac{1}{2}(K+\Lambda), \cdots \end{bmatrix} \begin{bmatrix} \langle\Lambda_2\Lambda_2\rangle & \langle k0\rangle & \langle\Lambda_2, \Lambda_2-k\rangle \\ \Lambda_2-j, \Lambda_2-j & ij & \Lambda_2-k, \Lambda_2 \end{bmatrix}$$
$$\times \langle\Lambda\, i(\Lambda_1-i)\Lambda_2-j\, \tfrac{1}{2}(K+\Lambda) \,|\, \Lambda_1, \Lambda_2-j\, i(\Lambda_2-k)\, \tfrac{1}{2}(K+\Lambda)\rangle$$
$$\times \langle\Lambda_1-k\, j(\Lambda_1-i)\Lambda_2-j\, \tfrac{1}{2}(K+\Lambda) \,|\, \Lambda_1-k, \Lambda_2-j\, j(\Lambda_2)\, \tfrac{1}{2}(K+\Lambda)\rangle. \tag{38}$$

All i.f.'s which appear here, as well as the SU_2 recoupling matrix, are known, except for the normalizing factor $N_0(\Lambda_1\Lambda_2, K\Lambda)$ of the second i.f. on the right-hand side. We divide both sides by this factor. Then on the left-hand side we obtain the square of the matrix element under consideration. Afterwards we apply the third or fifth of formulas (13.1) of Ref. 10 in order to simplify the expression on the right-hand side of (38). This allows us to carry out the summation with respect to i to obtain

$$N_0(\Lambda_1\Lambda_2, K\Lambda) = \Bigg(\frac{2(4\Lambda_1+1)!}{(K+\Lambda-2\Lambda_1+1)(K+\Lambda-2\Lambda_1+2\Lambda_2+3)}$$
$$\times \frac{(4\Lambda_2+3)!\,(2\Lambda_1+2\Lambda_2-K-\Lambda+1)}{(K+\Lambda+1)!\,(2\Lambda_1+2\Lambda_2-2K)!\,(2\Lambda_1+2\Lambda_2-2\Lambda+1)!}\Bigg)^{\frac{1}{2}}$$
$$\times \Bigg(\sum_{x,y} \frac{(-1)^{x+y}(4\Lambda_1+2-x)!\,(K+\Lambda-2\Lambda_1+x)!\,(4\Lambda_2+2-y)!}{x!\,(2\Lambda_1-x+1)!\,y!\,(2\Lambda_2-y+1)!\,(2\Lambda_1+2\Lambda_2-2K-x-y)!}$$
$$\times \frac{(K+\Lambda-2\Lambda_2+y)!\,(2\Lambda_1+2\Lambda_2-K-\Lambda-x-y)!}{(2\Lambda_1+2\Lambda_2-2\Lambda-x-y+1)!\,(2K+2\Lambda-2\Lambda_1-2\Lambda_2+x+y+1)!}\Bigg)^{-\frac{1}{2}}. \tag{39}$$

It is easy to see that the sum involved here corresponds to the nonstandard $11j$ coefficient

$$\left[\begin{array}{c|ccc} K+\Lambda+1 & 2\Lambda_1+1 & 2\Lambda_2+1 & -K+\Lambda-1 \\ -\tfrac{1}{2}(K+\Lambda+3) & \tfrac{1}{2}(K+\Lambda+1) & \tfrac{1}{2}(K+\Lambda+1) & \tfrac{1}{2}(K+\Lambda+1) \\ \tfrac{1}{2}(K+\Lambda+1) & \tfrac{1}{2}(K+\Lambda+1) & \tfrac{1}{2}(K+\Lambda+1) & -\tfrac{1}{2}(K+\Lambda+3) \end{array}\right], \tag{40a}$$

which in absolute value (only the square of our matrix element is needed) is equal to the $11j$ coefficient

$$\left[\begin{array}{c|ccc} -\tfrac{1}{2}(K+\Lambda+3) & -\tfrac{1}{2}(K+\Lambda+3) & -\tfrac{1}{2}(K+\Lambda+3) & -\tfrac{1}{2}(K+\Lambda+3) \\ -\tfrac{1}{2}(K+\Lambda+3) & -\tfrac{1}{2}(K+\Lambda+3) & -\tfrac{1}{2}(K+\Lambda+3) & -\tfrac{1}{2}(K+\Lambda+3) \\ -K-\Lambda-3 & 2\Lambda_1+1 & 2\Lambda_2+1 & -K+\Lambda-1 \end{array}\right]. \tag{40b}$$

The last $11j$ coefficient can be calculated from the expression

$$\begin{bmatrix} \Lambda & \Lambda & \Lambda & \Lambda \\ \Lambda & \Lambda & \Lambda & \Lambda \\ 2\Lambda & I & J & L \end{bmatrix}$$

$$= \frac{(-1)^p (p-I)!\,(p-J)!\,(p-L)!}{2(2\Lambda+1)(2\Lambda+1)!^3\,(I+J+L-p)!\,(p-I-J)!\,(p-I-L)!\,(p-J-L)!\,(p+1)!}$$
$$\times [(2\Lambda-I)!\,(2\Lambda+I+1)!\,(2\Lambda-J)!\,(2\Lambda+J+1)!\,(2\Lambda-L)!\,(2\Lambda+L+1)!]^{\frac{1}{2}}$$
$$\times \left(\frac{(I+J+L-2\Lambda)!\,(2\Lambda+L-I-J)!\,(2\Lambda-L-I+J)!}{(2\Lambda+L+I-J+1)!\,(2\Lambda+L-I+J+1)!} \right.$$
$$\left. \times \frac{(2\Lambda-L+I-J)!\,(2\Lambda+L+I+J+2)!}{(2\Lambda-L+I+J+1)!} \right)^{\frac{1}{2}}. \quad (41)$$

Here

$$p = \Lambda + \tfrac{1}{2}(I+J+L)$$

is integer; otherwise, (41) vanishes. This follows from the symmetry of the $11j$ coefficient.

Equation (41) is deduced with the help of recurrence formula (20) for i.f.'s of Ref. 2. In the general case it has 12 terms. However, four of them disappear because of the symmetry of the representations to be coupled. We avoid two terms by specializing the values of parameters of types M and N, and two more disappear because the corresponding parameters of SU_2 have maximal values. Of the four terms left, two vanish because of certain symmetry conditions. Hence, we are left with a recurrence formula involving only two i.f.'s, which allows us to obtain (41). The last one is a particular case of (29), which, however, we have not succeeded in deducing by the methods of Sec. 5.

After performing all the operations needed for the normalization factor, we find the expression

$$N(\Lambda_1\Lambda_2, K\Lambda) = \frac{(2\Lambda_1+1)!\,(2\Lambda_2+1)!\,(K-\Lambda)!\,(\Lambda_1+\Lambda_2+\Lambda+1)!}{2(K+\Lambda+1)!\,(\Lambda_1+\Lambda_2-\Lambda)!\,\nabla(\Lambda\Lambda_1\Lambda_2)}$$
$$\times \left(\frac{(2K-2\Lambda+1)(2\Lambda+1)!\,(2K+2)!\,(2K+2\Lambda+3)!}{(2\Lambda_1+2\Lambda_2-2\Lambda+1)!\,(2\Lambda_1+2\Lambda_2+2\Lambda+3)!} \right.$$
$$\left. \times \frac{(\Lambda_1+\Lambda_2-K)!\,(K-\Lambda_1+\Lambda_2)!\,(K+\Lambda_1-\Lambda_2)!}{(2K+2\Lambda_1-2\Lambda_2+1)!\,(2K-2\Lambda_1+2\Lambda_2+1)!\,(K+\Lambda_1+\Lambda_2+2)!} \right)^{\frac{1}{2}}. \quad (42)$$

It is more simple to obtain the i.f.'s in coupling $\langle\lambda\lambda\rangle$ and $\langle k0\rangle$. It is easy to find the relation

$$\begin{bmatrix} \langle K\Lambda\rangle & \langle K0\rangle & \langle\lambda\lambda\rangle \\ ij & ij & 00 \end{bmatrix} = (-1)^{K-\Lambda-i+j} \begin{bmatrix} K+\Lambda+1 & K-\Lambda & 2\lambda+1 \\ j-i & i-j & 0 \end{bmatrix}, \quad (43)$$

which can be brought to a form useful for our purpose by the use of Eq. (A.22) of Ref. 2. To obtain (43), we use a relation similar to (38) and find

$$\begin{bmatrix} \langle K\Lambda\rangle & \langle K-\Lambda,0\rangle & \langle\lambda\lambda\rangle \\ ij & i-a,j-a & 00 \end{bmatrix} = N \begin{bmatrix} \langle K-\Lambda,0\rangle & \langle 2a,0\rangle & \langle k0\rangle \\ i-a,j-a & aa & ij \end{bmatrix}$$
$$\times \begin{bmatrix} \langle K\Lambda\rangle & \langle K-\Lambda,0\rangle & \langle\lambda\lambda\rangle \\ ij & i-a,j-a & aa \end{bmatrix} \begin{bmatrix} \langle\lambda\lambda\rangle & \langle 2a,0\rangle & \langle\lambda\lambda\rangle \\ aa & aa & 00 \end{bmatrix} (-1)^{2k},$$
$$2a = k - K + \Lambda. \quad (44)$$

The recoupling matrices of SU_2 give the factor $(-1)^{2k}$ only. The last i.f. on the right-hand side is unity in view of the restrictions imposed on the parameters of SU_2 and of the normalization condition. The second i.f. is an n.s.s.i.f. I. It is expressed in terms of a triple stretched $9j$ coefficient by the methods of Ref. 6. This last one reduces to the CGC of SU_2 on the right-hand side of (43). Normalization is evident.

The general expression can be obtained by using the weight raising operator $(T_{++})^{2I'}$. The corresponding expression is rather too extensive to be given explicitly.

We give here two more special cases:

$$\begin{bmatrix} \langle \lambda\lambda\rangle & \langle k0\rangle & \langle K\Lambda\rangle \\ ii & i, k-i & 0, K-\Lambda \end{bmatrix} = (-1)^{2\lambda+K+\Lambda+k+2i} \begin{bmatrix} K-\Lambda & 2\lambda+1 & K+\Lambda+1 \\ k-2i & 2i+1 & k+1 \end{bmatrix}; \tag{45}$$

$$\begin{bmatrix} \langle \Lambda\Lambda\rangle & \langle 2b, 0\rangle & \langle \Lambda\Lambda\rangle \\ II & ij & I'I' \end{bmatrix} = (-1)^{2\Lambda+2b+2I'+1} \left(\frac{(2\Lambda+1)(2\Lambda+2)(2I+1)}{(2b+1)(2I'+1)} \right)^{\frac{1}{2}} \times \begin{Bmatrix} \Lambda - \frac{1}{2}(I-I'+i) & \Lambda - \frac{1}{2}(I'-I+j) & b \\ \Lambda + \frac{1}{2}(j+I-I'+1) & \Lambda + \frac{1}{2}(i+I'-I+1) & I+I'+\frac{1}{2} \end{Bmatrix}. \tag{46}$$

As we see, the first one is expressed in terms of a CGC of SU_2, and the other in terms of a $6j$ coefficient of SU_2.

VII. RECOUPLING MATRICES OF R_5

The recoupling matrices for the semistretched schemes have rather simple expressions. In the case of the first kind [conditions (19)], we have

$$\langle\langle K_1\Lambda_1\rangle\langle K_2\Lambda_2\rangle(\langle K_{12}\Lambda_{12}\rangle)\langle K_3\Lambda_3\rangle\langle K\Lambda\rangle||\langle K_3\Lambda_3\rangle\langle K_2\Lambda_2\rangle(\langle K_{32}\Lambda_{32}\rangle)\langle K_1\Lambda_1\rangle\langle K\Lambda\rangle\rangle$$
$$= \langle K_1-\Lambda_1, K_2-\Lambda_2(K_{12}-\Lambda_{12})K_3-\Lambda_3; K-\Lambda||K_3-\Lambda_3, K_2-\Lambda_2(K_{32}-\Lambda_{32})K_1-\Lambda_1; K-\Lambda\rangle. \tag{47}$$

The recoupling matrix of SU_2 on the right-hand side can be expressed in terms of a $6j$ coefficient. Equation (47) can be proved in this way.

As is well known, the recoupling matrix does not depend on the parameters labeling basis functions. On the other hand, its matrix elements are equal to the scalar product of the coupled states of the form (18). They are calculated by using the condition of orthogonality of the functions to be coupled. We take the resulting basis function, satisfying the conditions

$$I+J=K+\Lambda, \quad M=I, \quad N=J. \tag{48}$$

We then use the inequalities

$$I_i+J_i \le K_i+\Lambda_i, \quad I_{ij}+J_{ij} \le K_{ij}+\Lambda_{ij} \tag{49}$$

as well as the triangle conditions. The representations of the subgroup SU_2 are coupled by the stretched CGC's which are equal to unity. The i.f.'s are to be obtained from Eq. (29) of Ref. 6, in which the $9j$ coefficient becomes double stretched and the corresponding i.f. becomes a CGC of SU_2 of the form

$$\begin{bmatrix} K_1-\Lambda_1 & K_2-\Lambda_2 & K_{12}-\Lambda_{12} \\ I_1-J_1 & I_2-J_2 & I_{12}-J_{12} \end{bmatrix}. \tag{50}$$

Hence, only CGC's of SU_2 of the type (50) are left in the expression for the recoupling matrix. After summation with respect to I_i-J_i (under the condition of fixed values of I_i+J_i), we obtain the quantity isomorphic to the recoupling matrix of SU_2.

A similar result is obtained in the case of three representations coupled by the semistretched coupling scheme of the second kind. Then we have

$$\langle\langle K_1\Lambda_1\rangle\langle K_2\Lambda_2\rangle(\langle K_1+K_2, \Lambda_{12}\rangle)\langle K_3\Lambda_3\rangle\langle K_1+K_2+K_3, \Lambda\rangle|$$
$$|\langle K_3\Lambda_3\rangle\langle K_2\Lambda_2\rangle(\langle K_2+K_3, \Lambda_{32}\rangle)\langle K_1\Lambda_1\rangle\langle K_1+K_2+K_3, \Lambda\rangle\rangle$$
$$= \langle \Lambda_1\Lambda_2(\Lambda_{12})\Lambda_3\Lambda \,|\, \Lambda_3\Lambda_2(\Lambda_{32})\Lambda_1\Lambda\rangle. \tag{51}$$

To demonstrate this formula, we take the basis function of the resulting representation satisfying the conditions

$$I=M=K_1+K_2+K_3, \quad J=N=\Lambda. \tag{52a}$$

In such a case the expansion (18) involves i.f.'s equal to unity (from the normalization condition) when

$$I_i=K_i, \quad J_i=\Lambda_i, \quad I_{ij}=K_i+K_j, \quad J_{ij}=\Lambda_{ij} \tag{52b}$$

and vanishes otherwise. Moreover, all the CGC's of one of the subgroups SU_2 are equal to unity. Consequently, the recoupling matrix of R_5 as a scalar product of functions is equal to the recoupling matrix of the second subgroup SU_2 in this case.

Similar results are obtained in the case of an arbitrary number of basis functions coupled by the semistretched coupling schemes. By using the elements of the substitution group of parameters, we can widen the class of recoupling matrices of R_5 expressible in terms of $3nj$ coefficients of SU_2. Some of the couplings can become nonstandard; however, the corresponding $3nj$ coefficients remain standard quantities of the theory of angular momentum considered in Ref. 10. This means that the parameters which are the linear

combinations of parameters of R_5 satisfy the standard triangular conditions as required by the theory.

APPENDIX: A NEW EXPRESSION FOR THE 9j COEFFICIENT

The semistretched i.f. of the first kind may be treated in the same way as has been done with that of the second kind in Sec. 5. In such a case only one of four parameters i_1, i_2, j_1, and j_2 are linearly independent, as follows from Eqs. (49) and (50). Consequently, the number of summation parameters reduces to four. Comparing the expression obtained with Eq. (29) of Ref. 6, we deduce the following expression for the 9j coefficient:

$$\begin{Bmatrix} l_1 & l_2 & l \\ i_1 & i_2 & i \\ j_1 & j_2 & j \end{Bmatrix} = \frac{\nabla(j_2 j_1 j)\nabla(i_1 i_2 i)}{(2l+1)^{\frac{1}{2}}\nabla(i_1 j_1 l_1)\nabla(j_2 i_2 l_2)\nabla(jli)}$$
$$\times \sum_{m_1,m_2,u} \left(\frac{(l+m_1+m_2)!\,(l_1+m_1)!\,(l_2-m_2)!}{(l-m_1-m_2)!\,(l_1-m_1)!\,(l_2+m_2)!}\right)^{\frac{1}{2}} \begin{bmatrix} l_1 & l_2 & l \\ m_1 & m_2 & m_1+m_2 \end{bmatrix}$$
$$\times \frac{(-1)^{j-i+m_1+m_2}(i_1+j_1-m_1)!\,(i_2+j_2+m_2)!}{(j_1-i_1+m_1)!\,(i_2-j_2-m_2)!\,(i_1+j_2-i+m_2)!\,(i_1+j_2-j-m_1)!}$$
$$\times \frac{(i_1+j_2-i+m_2+u)!\,(l-i+j+u)!\,(2i-u)!}{(i_1+j_2+i+m_2+1)!\,u!\,(l+i-j-u)!\,(j-i+m_1+m_2+u)!}. \tag{A1}$$

For CGC's of SU_2 we use the second of Eq. (13.1) of Ref. 10 and renumerate the summation parameters in such a way that m_1 is involved in six factorials. The corresponding sum is isomorphic with the expression of CGC's of SU_2. Applying the symmetry properties of CGC's, we bring the expression to a form in which the summation with respect to one of the parameters can be carried out with the help of one of the Eqs. (14.2)–(14.5) of Ref. 10. This gives us the expression for the 9j coefficient:

$$\begin{Bmatrix} l_1 & l_2 & l \\ i_1 & i_2 & i \\ j_1 & j_2 & j \end{Bmatrix} = \frac{\nabla(i_1 i_2 i)\nabla(j_2 j_1 j)\nabla(lij)}{\nabla(i_1 j_1 l_1)\nabla(j_2 i_2 l_2)\nabla(l l_1 l_2)}$$
$$\times \sum_{x,y,z} \frac{(-1)^{i_1+i_2-i+j_1+j_2-j+x+y+z}(l-l_1+l_2+x)!\,(2l_1-x)!}{x!\,(l_1+l_2-l-x)!\,(l_1-i_1+j_1-x)!\,(l-l_1-i_2+j_2+x+y)!}$$
$$\times \frac{(i_1+j_1-l_1+x)!\,(l-j+i_1+i_2-y-z)!\,(l_2-i_2+j_2+y)!\,(2i_2-y)!}{y!\,(i_1+i_2-i-y)!\,(i_2-j_2+l_2-y)!\,(i_1+i_2+i-y+1)!\,(i_1-l_1-j_2+j+x+y)!}$$
$$\times \frac{(j_1-j_2+j+z)!\,(i+j-l+z)!}{z!\,(j_1+j_2-j-z)!\,(l+i-j-z)!\,(2j+z+1)!}, \tag{A2}$$

involving only three summation parameters as compared with Eq. (25.5) of Ref. 10 which involves four summation parameters.

From (55) it is easy to obtain particular cases of the 9j coefficient. Unfortunately, the symmetry of the expression (55) is rather poor. However, it satisfies the combined symmetry properties, as is easily proved by substitutions making some of the parameters negative [cf. Eqs. (24.19)–(24.31) of Ref. 10].

[1] R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[2] K. T. Hecht, Nucl. Phys. **63**, 177 (1965).
[3] B. T. Sharp and S. C. Pieper, J. Math. Phys. **9**, 663 (1968).
[4] N. Kemmer, D. L. Pursey, and S. A. Williams, J. Math. Phys. **9**, 1224 (1968).
[5] S. A. Williams and D. L. Pursey, J. Math. Phys. **9**, 1230 (1968).
[6] S. J. Ališauskas and A. P. Jucys, J. Math. Phys. **10**, 2227 (1969).
[7] S. J. Ališauskas, Z. B. Rudzikas, and A. P. Jucys, Dokl. Akad. Nauk SSSR **172**, 58 (1967); S. J. Ališauskas and A. P. Jucys, Lietuvos Fiz. Rinkinys **7**, 541 (1967); J. Math. Phys. **8**, 2250 (1967).
[8] R. M. Asherova and Yu. F. Smirnov, Nucl. Phys. **B4**, 399 (1968).
[9] S. J. Ališauskas, Lietuvos Fiz. Rinkinys **9**, 641 (1969).
[10] A. P. Jucys and A. A. Bandzaitis, *The Theory of Angular Momentum in Quantum Mechanics* (Mintis, Vilnius, Lithuanian SSR, 1965).
[11] D. R. Speiser, Helv. Phys. Acta **38**, 73 (1965).
[12] F. Buccella and U. Cattani, Nuovo Cimento Suppl. **6**, 1105 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Horizons and Analytic Extensions in Static Two-Dimensional Space-Times*†

BRENDAN B. GODFREY‡
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540

(Received 27 May 1970)

Necessary and sufficient conditions are derived for the analytic extendability of a static two-dimensional space–time. For all allowed cases, explicit analytic extensions are determined together with their corresponding Penrose–Carter diagrams. Extensions are classified and further discussed in terms of these diagrams, with special consideration given to the question of bifurcate Killing horizons. The application of these results to four-dimensional relativistic space–times is illustrated with specific examples.

1. INTRODUCTION

An important step towards the understanding of any new solution of the Einstein field equations is the determination of its maximal analytic extension. The need for such extensions arises particularly often in the study of static solutions, and in the following manner.

To make best use of the time invariance of static metrics in simplifying the field equations, we choose coordinates which cast the timelike Killing vector into the form

$$\xi^{\mu}_{(t)} = (0, 0, 0, 1). \tag{1}$$

In the event that the norm of this Killing vector should vanish on some hypersurface, defining a Killing horizon,[1] a coordinate singularity occurs. We can cite as well-known examples the Schwarzschild,[2,3,4] Reissner–Nordstrøm,[5,6] and Taub–NUT[7,8,9] metrics in their usual forms. In each case a maximal analytic extension has been found.

It should be emphasized that the three examples are really quite special. In the first place, all are Petrov type D, a highly restricted class of metrics.[10] Secondly, and of more importance to us at the moment, the analytic extensions are nearly all of essentially the same sort, following the form employed by M. Kruskal for the Schwarzschild solution.[4] Only in a few instances, such as $e = m$ for the Reissner–Nordstrøm solution,[6] does the form of the analytic extension fall into a second category. This limitation arises, not from any lack of originality on the part of various authors, but from the basic character of the examples themselves.

Recently, the author has made a study of static, axisymmetric, vacuum metrics, called Weyl metrics for brevity.[11] Among these solutions are many that exhibit a Killing horizon, yet admit an analytic extension distinctly different from those of the three previous examples. Others fail even to admit an analytic extension, although possessing a horizon nonsingular in the usual sense. These findings strongly recommend a systematic investigation of Killing horizons from the viewpoint of the analytic extensions they necessitate.

The seemingly most direct approach to the study of horizons, to analyze carefully the general static space–time, has one decisive drawback: The general static space–time is unknown, or at least so poorly known that little detailed information can be gleaned from it. Even treating the more specialized general Weyl solution, which is in principle explicitly known,[11] is of limited value. From a technical standpoint, an analytic extension can be determined only from a detailed knowledge of the space–time's geodesics, a knowledge but rarely available for Weyl metrics. It is true that the geodesic equations often can be integrated approximately near the horizon, but this is not entirely satisfactory either. Moreover, even if it were technically feasible explicitly to construct analytic extensions for many Weyl metrics, the welter of details would obscure the fundamental features we wish to understand.

For these reasons we subject to study the simplest possible metric, the general static, two-dimensional space–time. We call it a space–time only because its signature is $(+-)$; it satisfies no particular field equations.

In this limited domain we are able to establish necessary and sufficient conditions for analytic extendability. And when these conditions are met, we give explicitly the analytic extension together with its Penrose–Carter diagram.[6,12] Surveying the results, we see how special a class Kruskal extensions actually are. This point is made sharper by a brief examination of bifurcate Killing horizons[13] as they occur in static two-dimensional space–times. This

sort of horizon, consisting of four branches meeting at a true point (in two dimensions), is shown to be unique to Kruskal extensions.

The investigation is of primary value, we feel, to the extent that it provides a more intuitive understanding of horizons and analytic extensions by abstracting their basic features from the complicating details which so often abound. Nonetheless, it is fair to ask what specific application the calculations have to four-dimensional, general relativistic space–times.

Results seem to be directly utilizable only for static Petrov type D metrics. These decompose in a natural manner into two subspaces, one of which has signature $(+-)$.[10,14] The analytic extension is performed with regard to that subspace only. Extension of the Schwarzschild solution, for instance, involves only r and t. We use special cases of the lesser known C-metric[10,14] to provide additional examples in Sec. 5.

Application to Weyl metrics in general is more difficult. However, it appears that the added problems are only technical in nature. (Reference 11 examines this in detail.) It is not unreasonable to suppose the same to be true of more general static space–times.

2. EXTENSION THEOREM

Any static 2-space of signature $(+\ -)$ can be coordinatized so that

$$ds^2 = \phi^{-1}\,dx^2 - \phi\,dt^2, \qquad x_1 < x < x_2, \quad -\infty < t < \infty. \tag{2}$$

The connected open domain (x_1, x_2) is defined by requiring (i) that $\phi(x)$ be analytic and positive valued on the domain, and (ii) that the domain be not a proper subset of any connected open domain also satisfying condition (i). We assume that $\phi(x)$ itself admits no analytic extension. If it does, perform the extension and redefine the domain (x_1, x_2) as necessary. The question at issue is whether there exists some analytic metric of which Eq. (2) is but a restricted part.

It is evident that the metric can be extended only from one or both of the surfaces $x = x_i$, $i = 1, 2$.[15] Moreover, extension is impossible even at $x = x_i$ if the surface is singular or is located at infinity. We employ the usual definition for a singular surface: The curvature scalar, in this case

$$R = d^2\phi/dx^2, \tag{3}$$

is unbounded in the limit as the surface is approached by all geodesics intersecting it.[16] A surface is at infinity if no geodesic can intersect it at a finite affine distance.

The geodesics are easily determined by quadrature,

$$s = \pm\int [E^2 - \epsilon\phi(x)]^{-\frac{1}{2}}\,dx, \tag{4}$$

$$t = E\int \phi(x)^{-1}\,ds, \tag{5}$$

E and ϵ are constants of the motion. The geodesic is timelike, null, or spacelike as $\epsilon = 1$, 0, or -1. The affine parameter is s.

Now suppose $x_i = \pm\infty$. Additionally, suppose $\phi(x)/x^2$ is bounded as $x \to x_i$. Then Eq. (4) indicates that $s \to \pm\infty$ for all geodesics approaching x_i. So instead suppose $\phi(x)/x^2$ is unbounded as $x \to x_i$. However, now the scalar curvature is unbounded. In either case no extension is possible at x_i. This is the first step in the proof of our principal result.

Theorem: The metric (1) is analytically extendable from the surface $x = x_i$ $(i = 1, 2)$ if and only if x_i is finite and $\phi(x)$ is analytic there.

We next construct an analytic extension from the surface $x = x_i$ (x_i henceforth assumed finite) valid whenever $\phi(x)$ is analytic there. Introduce the semi-geodesic coordinate system

$$r = \int (1 + \phi)^{\frac{1}{2}}\phi^{-1}\,dx + t, \tag{6}$$

$$\tau = \int (1 + \phi)^{-\frac{1}{2}}\phi^{-1}\,dx + t. \tag{7}$$

The transformed line element

$$ds^2 = dr^2 - (1 + \phi)\,d\tau^2 \tag{8}$$

is the extension we seek, provided $\phi(x)$ is analytic in r and τ at $x = x_i$.

The original coordinate x is given implicitly in terms of r and τ by

$$r - \tau = \int (1 + \phi)^{-\frac{1}{2}}\,dx. \tag{9}$$

Therefore, $(r - \tau)$ is an analytic function of x at x_i whenever ϕ is. Moreover, $d(r - \tau)/dx \neq 0$ at x_i. These two facts imply that x is an analytic function of $(r - \tau)$.[17] Consequently, so is ϕ. Thus Eq. (8) is an analytic extension of Eq. (2), and sufficiency for the theorem is proven.

Proof of necessity for the theorem rests upon the fact that lines of constant τ in line element (8) are geodesics, the affine parameter being simply r. Because the scalar curvature Eq. (3) completely characterizes the space–time locally, it must be an analytic function of the affine parameter along any

geodesic if the metric itself is to be analytic along the geodesic.[18,19] An equivalent, but perhaps intuitively more obvious, statement is that the separation between nearby members of a family of geodesics, proportional to $(1 + \phi)^{\frac{1}{2}}$ in Eq. (8), must be an analytic function of distance along the geodesics if the metric is analytic there. Viewed in either way this argument requires that ϕ be an analytic function of x at x_i if an analytic extension is to be possible there.

3. KILLING HORIZONS

According to the extension theorem, analytic extension through a Killing horizon at x_i is possible only if ϕ is expressible as

$$\phi(x) = x^n f(x). \tag{10}$$

Here n is a positive integer. The function $f(x)$ is analytic and nonvanishing at x_i, which for notational convenience has been set equal to zero. The preceding section presented a coordinate patch adequate to cover one branch of the horizon at a time. It often proves valuable to be able to treat both the past and future branches simultaneously. The appropriate coordinate transformation is derived below.

Introduce a null coordinate system

$$ds^2 = 2\psi(u, w)\, du\, dw. \tag{11}$$

Equations connecting line elements (2) and (11) are

$$\begin{aligned} x_u^2/\phi - t_u^2\phi &= 0, \\ x_w^2/\phi - t_w^2\phi &= 0, \\ x_u x_w/\phi - t_u t_w \phi &= \psi. \end{aligned} \tag{12}$$

These may be rewritten in the equivalent forms

$$\begin{aligned} t_u &= x_w/\phi, \\ t_w &= -x_u/\phi, \end{aligned} \tag{13}$$

and

$$\psi = 2x_u x_w/\phi. \tag{14}$$

The integrability condition on (13), $t_{uw} = t_{wu}$, is seen to imply

$$t_{uw} = 0; \tag{15}$$

i.e.,

$$t = \xi(u) - \eta(w). \tag{16}$$

In terms of the arbitrary functions ξ and η, $x(u, w)$ is implicitly defined as

$$\int [\phi(x)]^{-1}\, dx = \xi(u) + \eta(w). \tag{17}$$

It is necessary only to require that

$$\psi = 2\phi\xi_u\eta_w \tag{18}$$

be both analytic in u and w and nonvanishing at the horizon.

Designate the past and future horizon branches, respectively, by $w = 0$ and $u = 0$. We must require that $\phi \cdot \xi_u$ therefore be analytic and nonvanishing at $u = 0$, with an analogous statement for $w = 0$. But $x_u = \phi \cdot \xi_u$, and so we will demand that x be an analytic function of u and w at the horizon. The conditions placed on ξ and η by this requirement may be determined by formally expanding x as a power series in u and w and formally solving for the coefficients using Eq. (17).

Instead, we explore the consequences of the appealing choice

$$\begin{aligned} \xi(u) &= \int [\phi(u)]^{-1}\, du, \\ \eta(w) &= \int [\phi(w)]^{-1}\, dw. \end{aligned} \tag{19}$$

The substitution of Eqs. (10) and (19) into Eq. (17) yields

$$\begin{aligned} a \ln(x) + x^{1-n}P(x) \\ = a \ln(u) + u^{1-n}P(u) + a \ln(w) + w^{1-n}P(w), \end{aligned} \tag{20}$$

where P is a power series, [$P(0) \neq 0$ for $n \neq 1$] and a is a constant ($a \neq 0$ for $n = 1$).

It is evident from Eq. (20) that, for w bounded away from zero, $x \sim u$ as $u \to 0$, except when $n = 1$. In that case, irrespective of whether w is bounded away from zero, $x \sim u \cdot w$ as $u \to 0$. In either case, if we write

$$z = x/(uw), \tag{21}$$

we are assured that z does not vanish on the future horizon branch. With this substitution, Eq. (20) becomes

$$\begin{aligned} F(u, w, z) = (u\,w\,z)^{n-1}a \ln(z) + P(u\,w\,z) \\ - (w\,z)^{n-1}P(u) - (u\,z)^{n-1}P(w) = 0. \end{aligned} \tag{22}$$

At $u = 0$, the newly defined function F is analytic in all its arguments. Moreover, F_z does not vanish there:

$$\begin{aligned} (F_z)_{u=0} &= (1 - n)w^{n-1}z^{n-2}P(0), \quad n \neq 1, \\ (F_z)_{u=0} &= a/z, \qquad\qquad\qquad n = 1. \end{aligned} \tag{23}$$

Consequently, z is defined implicitly by Eq. (22) as an analytic function of u and w at $u = 0$.[17] A similar statement holds for $w = 0$, and also for $u = w = 0$ if $n = 1$. For $n \neq 1$, however, the argument fails at $u = w = 0$.

It is evident from Eq. (21) that x is analytic under the same conditions. Since substitution of the limiting expression for x cited above into Eq. (18) indicates that ψ is neither zero nor infinite at $x = 0$, we have the desired result, that Eqs. (11), (16), (17), and (19) represent an analytic extension of Eq. (2) at the Killing horizon $x = 0$. Indeed, it represents simultaneously an extension for all horizons of (2).

Once again, it should be noted that the extension fails for $u = w = 0$ unless $n = 1$. However, examination of the geodesic equations reveals that the point in question lies at an infinite affine distance, so that the failure of the extension there is to be expected. This feature of $n \neq 1$ horizons is discussed in more detail in the following section.

4. THE PENROSE–CARTER DIAGRAMS

Penrose–Carter conformal diagrams for the two-dimensional Killing horizons are of two basic types, Fig. 1 for n odd and Fig. 2 for n even. The former consists of a pair of static (I) and a pair of dynamic (II) blocks, while the latter is an infinite chain of static blocks (two sets, I and II).

The mathematical complexity of the extension just given makes these block diagrams particularly useful in illustrating the general structure of a metric without requiring that the algebra actually be performed. Starting with the diagram for any horizon of the metric, one attaches the diagram for the next horizon at the appropriate shaded corner of the first diagram. This "building block" construction is continued until all extendable horizons of the metric are included. Finally, the remaining free shaded corners are identified as singularities, regions at infinite distance, etc., as the case may be. This was, in fact, the program carried out to obtain the diagrams for the examples of the next section. In the remainder of this section, we focus upon one particularly interesting aspect of these diagrams.

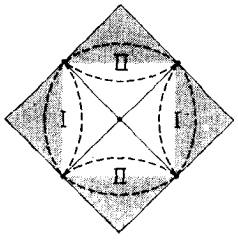

FIG. 1. Penrose–Carter diagram for line-element (11), n odd. Solid dark lines represent horizons. Broken lines are typical Killing vector orbits (lines of constant x). Dark points, not covered by the coordinate patch, are at infinite affine distance. Note, however, that this last statement does not apply to the center point for $n = 1$. The shaded areas, not necessarily covered by the coordinates (19), are shown to facilitate visualization of where this set of blocks can attach to others. The nature of the boundaries of the shaded regions (light lines) is determined by the behavior of $\phi(x)$ there. Region I is static, and region II dynamic.

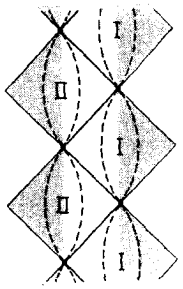

FIG. 2. Penrose–Carter diagram for line element (11), n even. Solid dark lines represent horizons. Broken lines are typical Killing vector orbits (lines of constant x). Dark points, not covered by the coordinate patch, are at infinite affine distance. The shaded areas, not necessarily covered by the coordinates (19), are shown to facilitate visualization of where this set of blocks can attach to others. The nature of the boundaries of the shaded regions (light lines) is determined by the behavior of $\phi(x)$ there. Regions I and II, though not identical, are both static.

In each diagram the vertices marked by heavy black dots represent spacelike or timelike infinity, as perhaps might be expected from the usual properties of conformal diagrams. In particular, the "origin" in Fig. 1 is at infinity and so, as previously noted, is not covered by the coordinate patch. This point serves to separate the four horizon branches, as the figure indicates.

An exception to this statement occurs in the case $n = 1$, for which the origin is perfectly accessible in finite time. Hence, the horizon branches actually intersect there, forming what Robert Boyer termed a bifurcate Killing horizon.[13] We see that, at least within the realm of this investigation, the bifurcate Killing horizon can occur only for $n = 1$. Let us further inquire into this situation.

An $n = 1$ coordinate singularity in the positive-definite analog to line element (2)

$$ds^2 = \phi^{-1}\, dx^2 + \phi\, dt^2 \tag{24}$$

is covered (to first order ignoring constant factors) by the quasirectangular coordinates

$$\rho \simeq x^{\frac{1}{2}} \sin(t/2),$$
$$\sigma \simeq x^{\frac{1}{2}} \cos(t/2). \tag{25}$$

This is essentially the transformation from polar to rectangular coordinates, in which the $n = 1$ coordinate singularity of (24) is the origin. There is, however, no question of carrying out an extension for any other value of n: The point in question now lies at infinity.

In the same approximation an $n = 1$ horizon is covered by

$$\rho = (u - w)/2 \simeq x^{\frac{1}{2}} \sinh(t/2),$$
$$\sigma = (u + w)/2 \simeq x^{\frac{1}{2}} \cosh(t/2). \tag{26}$$

No similar representation exists for $n > 1$. This we would expect, since such a representation would cover the origin. The correspondence between the remarks of this paragraph and those of the preceding one are fairly clear. That bifurcate Killing horizons

fail to occur for $n \neq 1$ is, from this standpoint, attributable to the absence of analytic extensions in the analogous positive-definite 2-spaces.

Incidently, Eq. (26) indicates the relationship between the present formulation (for $n = 1$) and the general $n = 1$ extensions derived by various authors.[5,14] It also suggests why these other methods are not readily applicable for $n > 1$.

5. APPLICATIONS

The analysis presented above finds immediate applications to static Petrov type D metrics, which naturally decompose into two parts, one transformable to Eq. (2), the other to Eq. (24).[10,14] Examples of $n = 1$[2,4,5,14] and static $n = 2$[6] are adequately covered in the literature. We therefore restrict ourselves to dynamic $n = 2$ and $n = 3$ cases. There are no known instances of $n > 3$ among static Petrov type D metrics. Neither are there any with n not an integer but a nonintegral real number exceeding two. For such a case, the space–time would not be analytically extendable at x_i, although the curvature invariants would be regular there.

The two examples are special cases of the C-metric [10,14]:

$$ds^2 = (x + y)^{-2}\{f(x)^{-1}\,dx^2 + f(x)\,d\phi^2 + [-f(-y)]^{-1}\,dy^2 - [-f(-y)]\,dt^2\},$$
$$f(z) \equiv z^3 + az + b. \qquad (27)$$

We are interested only in the $y - t$ part of the line element, the $x - \phi$ portion being easily extended where necessary. If we remain away from $x + y = \infty$, which is singular, and $x + y = 0$, which is at infinity, the effect of the multiplicative factor $(x + y)^{-2}$ can be ignored. What remains is precisely Eq. (2), with

$$\phi(y) = y^3 + ay - b. \qquad (28)$$

Penrose–Carter diagrams of the two examples are immediately obtainable as follows.

Let constants a and b be chosen so that

$$\phi(y) = (y - y_1)(y - y_2)^2, \quad y > y_1 > y_2 > 0. \qquad (29)$$

The block diagram for the horizon at y_1 is Fig. 1, while that at y_2 is the dynamic analog of Fig. 2 (i.e., rotated through 90°) for $n = 2$. The diagrams are joined through region II of Fig. 1. Figure 3 is, strictly speaking, not the maximal extension, because it shows all the identical $n = 1$ blocks attached at both ends to the same two $n = 2$ dynamic chains. In fact, each $n = 1$ block attaches at one end or the other to a new chain. The resulting network is difficult to imagine, much less to draw. No causality violations arise from the identifications made in Fig. 3.

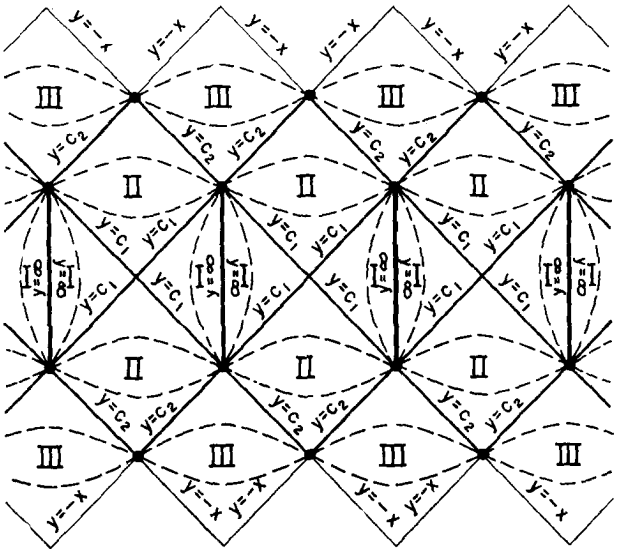

FIG. 3. Penrose–Carter diagram for constant $x - \phi$ slice ($x > 0$) of the C-metric. Parameters are chosen such that an $n = 1$ horizon occurs at $y = y_1$, and an $n = 2$ horizon at $y = y_2$. The wide black lines, $y = \infty$, denote singularities. The surfaces $y = -x$ are regular and at null infinity. Dark points are also at infinity. Dashed lines are Killing vector orbits, indicating that region I is static, but II and III dynamic.

Next let $a = b = 0$ so that an $n = 3$ horizon occurs at $y = 0$. The result is Fig. 4. In both Figs. 3 and 4 the behavior of the diagrams near the singularities and at null infinity is easily added by considering the previously neglected effect of $(x + y)^{-2}$. Both examples illustrate quite well the building block fashion in which Penrose–Carter diagrams can be constructed.

Reference 11 contains applications of our results to Weyl metrics. There all values of n occur. Moreover, cases arise where C^2 extensions are possible although

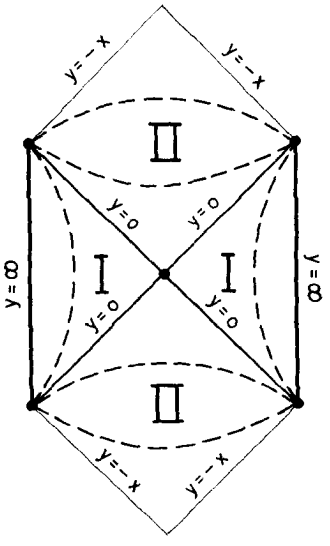

FIG. 4. Penrose–Carter diagram for a constant $x - \phi$ slice ($x > 0$) of the C-metric. Parameters are chosen ($a = b = 0$) such that an $n = 3$ horizon occurs at $y = 0$. The wide black lines, $y = \infty$, are singular surfaces. The surface $y = -x$ is regular and at null infinity. Dark points are also at infinity. Killing vector orbits (dashed lines) indicate that region I is static, and II dynamic.

analytic extensions are not. These correspond generally to Eq. (10) but with n any real number greater than two, rather than only an integer. Minor modification of the procedures given in this paper suffice to provide such extensions.

ACKNOWLEDGMENTS

I am indebted to John Wheeler, my thesis advisor, for constant help and encouragement in the preparation of this report. I am also grateful to James York for numerous discussions and much good advice. Charles Misner and Karel Kuchar provided valuable comments.

* Based on part of the author's doctoral thesis submitted to Princeton University, 1970.

† This work has been assisted in part by NSF Grant GP 7669.

‡ National Science Foundation Predoctoral Fellow, 1967–1970. Now at Air Force Weapons Laboratory, Kirtland AFB, N.M. 87117.

[1] B. Carter, J. Math. Phys. **10,** 70 (1969).

[2] D. Finkelstein, Phys. Rev. **110,** 965 (1958).

[3] C. Fronsdal, Phys. Rev. **116,** 778 (1959).

[4] M. Kruskal, Phys. Rev. **119,** 1743 (1960).

[5] J. Graves and D. Brill, Phys. Rev. **120,** 1507 (1960).

[6] B. Carter, Phys. Letters **21,** 423 (1966).

[7] C. Misner, J. Math. Phys. **4,** 924 (1963).

[8] C. Misner, in *Relativity Theory and Astrophysics*, J. Ehlers, Ed. (American Mathematical Society, Providence, 1967), Vol. 1.

[9] C. Misner and A. Taub, Zh. Eksp. Teor. Fiz. **55,** 223 (1968) [Sov. Phys. JETP **28,** 122 (1969)].

[10] J. Ehlers and W. Kundt, *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (Wiley, New York, 1960).

[11] B. Godfrey, *Static, Axisymmetric, Vacuum Metrics*, unpublished doctoral dissertation, Princeton University, 1970 .

[12] R. Penrose, *Relativity Groups and Topology*, C. DeWitt and B. DeWitt, Eds. (Gordon and Breach, New York, 1963).

[13] R. Boyer, Proc. Roy. Soc. (London) **311A,** 245 (1969).

[14] M. Walker, *Some Global Aspects of Space-Time Structures*, unpublished doctoral dissertation, Birkbeck College, London, 1969 .

[15] This statement is not rigorously true. As pointed out by Robert Geroch (e.g., lecture given at the International School of Physics "Enrico Fermi," July 1969), any space–time can be analytically extended: Make a cut in the space–time, perform extensions from the resulting free surfaces, and finally remove from the manifold the branch points introduced. Consequently, there is no unique maximal analytic extension. However, such extensions, in a sense admittedly difficult to define, seem artificial, and, in any case, will not be further considered here. With this limitation the footnoted sentence is correct. The author is indebted to Robert Geroch and to an anonymous referee for a discussion of this point.

[16] R. Geroch, J. Math. Phys. **9,** 450 (1968).

[17] E. Goursat, *A Course in Mathematical Analysis* (Dover, New York, 1959), Vol. 1.

[18] This is equivalent to the requirement that Fermi normal coordinates be constructable along any geodesic in an analytic metric. See G. Szekeres, Publ. Math. Debrecen **7,** 285 (1960).

[19] We can require equivalently that the invariant quantity $\xi^{\mu}_{(t)}\xi_{(t)\mu}$, equalling ϕ, be an analytic function of the affine parameter. Theoretical justification is suggested in part by R. Geroch, Commun. Math. Phys. **13,** 180 (1969), or Ref. 14. Whether the requirement generalizes to other Killing vectors in other space–times is an interesting question.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Theorem on Invariant Amplitudes*

RALPH F. GUERTIN†‡
Department of Physics, Yale University, New Haven, Connecticut 06520
and
Department of Physics and Lawrence Radiation Laboratory
University of California, Berkeley, California 94720

(Received 6 August 1970)

It is shown that any basis of covariant polynomials for a two-particle scattering process yields invariant amplitudes free of kinematical singularities, provided (a) the total number of basis polynomials equals the number of spin space components of the scattering amplitude and (b) the polynomials of each of the two parity signatures are separately linearly independent at all points where three of the particle 4-momenta are linearly independent. This result allows one to directly identify good basis sets without going through the very tedious algebra involved in comparing them to the sets of Hepp and Williams. The latter are not useful for practical applications because the spinor indices belonging to different particles are coupled and these sets do not transform into themselves under the relevant discrete symmetry operations.

I. INTRODUCTION

The description of a scattering process in terms of invariant amplitudes is useful because they have simple analytic properties and can trivially satisfy the various symmetry requirements. For this purpose, one must find a covariant basis that exhibits the appropriate symmetries and, moreover, ensures a decomposition into invariant amplitudes free of kinematical singularities. In this paper we prove a theorem that enables one to check by inspection whether a given basis is a satisfactory one.

Hepp[1] has rigorously proven that any set of holomorphic covariant functions $M(k)$ can be written in the form

$$M(k) = \sum_i A_i(s) Y^i(k),$$

where the "standard covariants" $Y^i(k)$ are matrices on the spinor indices and polynomials in the 4-vectors $k = \{k_1, \cdots, k_n\}$. The invariant amplitudes $A_i(s)$, which are functions of the independent invariants $s = \{s_1, \cdots, s_m\}$ formed from the k_i, are holomorphic except on the s-space image of k-space singularities of the functions $M(k)$. In general, a minimal set of standard covariants, i.e., a set whose number is just the number of values that the spinor indices take on, does not exist. But for the case of scattering amplitudes describing two incoming and two outgoing particles, and subject to the mass shell and four-momentum conservation constraints, Hepp proved that one can find a minimal set of $Y^i(k)$ having positive and negative signature under parity and whatever other discrete symmetry operations transform the particular process under consideration into itself.

To find the relevant standard covariants for any given $M(k)$, one must find a set of polynomials such that all covariant polynomials with the same spinor index types can be expressed in terms of them for all allowed values of the complex 4-vectors k. Hepp's proof is, in effect, a proof that such a basis exists and that any such basis gives a holomorphic decomposition (one free of kinematical singularities) into invariant amplitudes of the given holomorphic functions $M(k)$. Any other form for physical scattering functions is equivalent to the $2j + 1$ spinor form; thus Hepp has given a rigorous justification of the prescription originally developed by Hearn[2] for perturbation theory. Recently Scadron and Jones[3] have found some of the relations needed to apply this method to two-particle processes with arbitrary spins, and many other relations have been found by the author of this paper.[4]

Several examples of sets of covariants that give a holomorphic decomposition have been given by Hepp and, independently, by Williams,[5] but these do not transform into themselves under the discrete symmetry operations. To find a satisfactory basis, it is sufficient to show how any one of these sets can be written as linear combinations of covariants having definite signature under P, C, T, and exchange symmetry, with coefficients that are polynomials in the invariants. For those processes in which two of the four particles are spinless, this procedure is not too tedious and all the required covariants have been found.[4,6,7]

When more than two spinning particles are involved, the algebraic problems become rapidly unmanageable as the spins increase.[3,4,7,8] Thus, although the prescription of Hearn and Hepp solves the problem in principle, it does not solve it in practice.

The main aim of this work is to exhibit, in concrete terms, the origin of the difficulty in obtaining holomorphic invariant amplitudes and to develop a simple criterion that allows one to verify directly whether any given basis gives amplitudes that are free of kinematical singularities. The criterion is essentially this: Any minimal set of covariant polynomials is a good set if all the covariants of positive parity signature are linearly independent, and all the covariants of negative parity signature are linearly independent at all points where three of the particle momenta are linearly independent.

The core of the argument is a detailed description of the constraints that Lorentz covariance and analyticity impose in neighborhoods of points where only two of the particle 4-momenta are linearly independent. Once the nature of the constraints is clearly understood, the essential requirements on the standard covariants will be easy to see and the proof of the theorem will be straightforward.

Aside from the Hearn–Hepp procedure, the only good criterion that has been previously given for justifying the absence of kinematical singularities in a set of invariant amplitudes is that of Williams.[5] However, Williams' arguments were for a particular set of covariants that has no simple relation to the discrete symmetries.

One reason that the standard covariants given by Hepp and Williams have not found practical application is that all spinor indices with the same properties under proper Lorentz transformations are coupled with Clebsch–Gordan coefficients.[9] These couplings join together parts having different discrete symmetry properties, and also make the results impracticable for substitution into the unitarity relations. Any useful basis must avoid coupling spinor indices belonging to different particles, and this does not naturally emerge if one starts from the sets of Hepp or Williams. By our simple criterion, any proposed basis can be checked directly without comparing it to all the other polynomials having the same spinor index types or to the sets of Hepp or Williams.

To apply our criterion, the linear independence of the basis sets of each parity signature must be checked. This problem, although not trivial, is simpler than the one involved in previous methods. A paper dealing with procedures for checking the linear independence of basis sets is being prepared.

The general background material needed for this work is scattered in various places and some of it is unpublished. Many of the essential results are obscured by the abstruse mathematical forms of the original papers, and there is no satisfactory account of the general situation anywhere in the literature. A secondary aim of this paper is therefore to provide a systematic simple description of the previous results that form the basis of the present work. These include a theorem of Stapp,[10,11] which states that scattering functions are covariant under proper complex Lorentz transformations at all points where they are regular, and some basic properties of sets of complex 4-vectors.[12] We also discuss and extend a theorem of Hepp[1] and Williams[5] that allows one to express a multisheeted invariant function of 4-vectors as a function of invariants formed from them.

We confine our discussion to those cases in which all four particles are massive, since Zwanziger[13] has found all the standard covariants having definite parity signature when one or two are massless.

II. BASIC PROPERTIES OF TWO-PARTICLE SCATTERING AMPLITUDES

A. The S Matrix

A brief review will establish the notation and summarize some well-known properties. We consider a scattering process in which particles 1 and 2 are incoming, and particles 3 and 4 are outgoing. To each particle one may assign a wavefunction $\psi_{\alpha_i}{}^{i}(k_i)$ with the norm

$$\sum_{\alpha_i}\int \frac{d\mathbf{k}_i}{k_i^0}\,\psi_{\alpha_i}{}^{i*}(k_i)\psi_{\alpha_i}{}^{i}(k_i) = 1, \quad i = 1,\cdots,4, \quad (1)$$

where k_i is the 4-momentum and α_i the spin quantum number. The particle type is specified by $t_i = (m_i, j_i, q_i)$, where m_i is its mass, j_i is its spin, and q_i stands for all additive internal quantum numbers. It is convenient to introduce the notation $K_i = (k_i, t_i)$.

The relationship between measurements on initial and final particles is expressed by the S-matrix elements

$$\underset{\alpha_b;\,\alpha_a}{S(K_b;\,K_a)},$$

where the subscripts a and b indicate incoming and outgoing particles, respectively. The probability that the outgoing particles will be in the state specified by ψ_b if the incoming particles were in the state specified by ψ_a is $|S(\psi_b;\,\psi_a)|^2$, where

$$S(\psi_b;\,\psi_a) = \prod_{i=1}^{4}\int \frac{d\mathbf{k}_i}{k_i^0}\sum_{\alpha_i} \psi_{\alpha_3}{}^{3*}(k_3)\psi_{\alpha_4}{}^{4*}(k_4)\underset{\alpha_b;\,\alpha_a}{S(K_b;\,K_a)} \times \psi_{\alpha_1}{}^{1}(k_1)\psi_{\alpha_2}{}^{2}(k_2). \quad (2)$$

If the spin indices are in "canonical" form, i.e., if α_i is specified with respect to the 3 axis in the rest frame of particle i, the invariance of probabilities

under a simultaneous proper orthochronous Lorentz transformation of both incoming and outgoing states leads to the relation[14–16]

$$S(\Lambda K_a;\ \Lambda K_a)_{\alpha_b;\,\dot\alpha_a} = D^{(j_b)}(A(k_b))_{\alpha_b}{}^{\alpha_b'} \times D^{(j_a)}(A(k_a)^*)_{\dot\alpha_a}{}^{\dot\alpha_a'} S(K_b;\ K_a)_{\alpha_b';\,\dot\alpha_a'}. \quad (3)$$

The notation is such that

$$D^{(j_b)}(A(k_b))_{\alpha_b}{}^{\alpha_b'} = \prod_{i=3}^{4} D^{(j_i)}(A(k_i))_{\alpha_i}{}^{\alpha_i'}$$

and

$$D^{(j_a)}(A(k_a)^*)_{\dot\alpha_a}{}^{\dot\alpha_a'} = \prod_{i=1}^{2} D^{(j_i)}(A(k_i)^*)_{\dot\alpha_i}{}^{\dot\alpha_i'}.$$

According to the spinor calculus convention introduced in Appendix A, each of the raised indices on the transformation matrices is summed with the corresponding S-matrix index, and dots are introduced over incoming spin indices to indicate that they transform like the complex conjugate of outgoing ones. Here $\Lambda = \Lambda(A, A^*) \in L_+^\uparrow$, where, as explained in Appendix A, A is a matrix in $SL(2, C)$ and $A(k_i)$, defined by (A11), is the matrix in $SU(2)$ corresponding to the "Wigner rotation" in (A10). We have adopted the convention $\Lambda K_i = (\Lambda k_i, t_i)$.

B. The M Functions

The fact that the spin transformation matrices in (3) depend on the 4-momenta of the particles leads to "kinematical" singularities in the 4-momenta when one analytically continues such an expression out of the physical region of the original process. This, in turn, leads to somewhat complicated crossing relations between the amplitudes for the various physical processes that are connected by analytic continuation. To simplify the crossing, analytic, and Lorentz transformation properties, it is convenient to introduce the spinor amplitudes, or M functions.[14–17]

The M functions with lower undotted spinor indices assigned to outgoing particles and lower dotted indices to incoming particles are defined by

$$M(K_b;\ K_a)_{\alpha_b;\,\dot\alpha_a} = D^{(j_b)}(\mathcal{B}(k_b))_{\alpha_b}{}^{\alpha_b'} \times D^{(j_a)}(\mathcal{B}^*(k_a))_{\dot\alpha_a}{}^{\dot\alpha_a'} S(K_b;\ K_a)_{\alpha_b';\,\dot\alpha_a'}, \quad (4)$$

where $\mathcal{B}(k_i)$, defined in (A9), is the Hermitian matrix in $SL(2, C)$ corresponding to the Hermitian "boost" $L(k_i)$ in (A8) that carries $\bar k_i = (m_i, 0)$ into k_i. Then, using $\Lambda = \Lambda(A, A^*)$, we easily verify the following simple covariance property from (3), (4), and (A11):

$$M(\Lambda K_b;\ \Lambda K_a)_{\alpha_b;\,\dot\alpha_a} = D^{(j_b)}(A)_{\alpha_b}{}^{\alpha_b'} D^{(j_a)}(A^*)_{\dot\alpha_a}{}^{\dot\alpha_a'} M(K_b;\ K_a)_{\alpha_b';\,\dot\alpha_a'}. \quad (5)$$

One can use instead the M functions with all lower undotted indices. It may be verified from (5), (A4), and

$$M(K_b;\ K_a)_{\alpha_b;\,\alpha_a} = D^{(j_a)}(\sigma \cdot k_a C^{-1}/m_a)_{\alpha_a}{}^{\alpha_a'} M(K_b;\ K_a)_{\alpha_b;\,\dot\alpha_a'} \quad (6)$$

that

$$M(\Lambda K_b;\ \Lambda K_a)_{\alpha_b;\,\alpha_a} = D^{(j_b)}(A)_{\alpha_b}{}^{\alpha_b'} D^{(j_a)}(A)_{\alpha_a}{}^{\alpha_a'} M(K_b;\ K_a)_{\alpha_b';\,\alpha_a'}. \quad (7)$$

The choice of index type has no effect on the physics or on the analytic properties. The matrix that transforms a particular spinor index from one type to another is a holomorphic function of that particle's 4-momentum.

For a given process, $M(K_b; K_a)$, where we suppress the spinor indices, is expressed as a sum of a "no-scattering" part $M^{ns}(K_b; K_a)$ and a "connected part"[18]

$$M^0(K_b;\ K_a) = i\delta^4(k_1 + k_2 - k_3 - k_4) M^c(K_b;\ K_a). \quad (8)$$

Equation (8) defines M^c functions that are free of conservation delta functions. To avoid specifying spinor index types, one may write their physical region covariance properties, of which (5) and (7) are particular examples, in the form

$$M^c(\Lambda(A, A^*)K) = \Lambda_s(A, A^*) M^c(K), \quad (9)$$

where $K = (K_b; K_a)$. The M^c functions are said to be $L_+^\uparrow$ covariant in the physical region of the process $t_1 + t_2 \to t_3 + t_4$. If all four particles are spinless, then $\Lambda_s(A, A^*) \equiv I$ on the right-hand side of (9) and the single M^c function is said to be $L_+^\uparrow$ invariant in the given physical region.

III. PROPERTIES IN COMPLEX FOUR-MOMENTA SPACE

A. Stapp's Theorem

Let $\mathcal{K}_3$ denote the set of points $k = \{k_1, k_2, k_3, k_4\}$ in complex 4-vector space that satisfy the four mass-shell constraints $k_i \cdot k_i = (m_i)^2 > 0$ and the four conservation-law constraints $k_1 + k_2 = k_3 + k_4$. The M^c functions defined by (8) are assumed in theory to be analytic functions over $\mathcal{K}_3$ except for dynamical singularities.[10,14,19] The set $\mathcal{K}_3$ is a subset of the space of four complex 4-vectors. The notion of analyticity on such a subset is a standard mathematical concept.[20]

As usual we use the term holomorphic to designate the property of being analytic and single valued. A domain is a connected open set, and, for our purposes, it is sufficient to regard the "domain of holomorphy" of a function as a union of sheets, a sheet being the maximum domain whose points map one-to-one onto points in $\mathcal{K}_3$.[21] The sheets may overlap; in fact, the boundaries of the sheets are somewhat arbitrary, and any point on the domain of holomorphy lies on the interior of some sheet. One does not include poles and branch points on any sheet, even though it is customary in physics to speak of a pole as lying on a particular sheet when it lies on the boundary of that sheet.

For a given scattering process there is an M^c function corresponding to each combination of values of the spinor indices. Following Stapp,[10,11] we define the *domain of regularity* $\mathcal{R}$ of this set of M^c functions to be the intersection of their domains of holomorphy; i.e., $\mathcal{R}$ is the largest (multisheeted) domain to which all of these M^c functions can be simultaneously analytically continued. The M^c functions are known to be $L_+^\uparrow$ covariant [see Eq. (9)] at points lying in the physical region for which they were originally defined, but what can be said about covariance at other points on $\mathcal{R}$? Before stating Stapp's theorem, which gives a precise answer to our question, we require a few definitions.

A set of tensor-valued functions $F(k)$ is said to be $\mathfrak{L}_+$ covariant on a set of points $\mathcal{S} \in \mathcal{K}_3$ if

$$F(\Lambda(A, B)k) = \Lambda_s(A, B)F(k), \tag{10}$$

whenever k and $\Lambda(A, B)k$ are in $\mathcal{S}$. As explained in Appendix A, the matrices A and B are in $SL(2, C)$ and $\Lambda(A, B) \in \mathfrak{L}_+$, the group of proper complex Lorentz transformations. The notation on the right-hand side of (10) indicates that the matrix $D^{(j_i)}(A)$ acts from the left on the ith spinor index of $F(k)$ if it is a lower undotted one, whereas the matrix $D^{(j_i)}(B)$ acts on it from the left if it is a lower dotted one.

The $\mathfrak{L}_+$ orbit, $\mathfrak{L}_+k$, of any point $k \in \mathcal{K}_3$ is the set of all points $\Lambda k = \{\Lambda k_1, \cdots, \Lambda k_4\}$ obtained by letting $\Lambda \in \mathfrak{L}_+$ take on all possible values.

Stapp's Theorem[10,11,22]*:* Let the M^c functions for a given process be $L_+^\uparrow$ covariant and holomorphic on a real open connected set in $\mathcal{K}_3$ corresponding to physical points. Then the domain of regularity $\mathcal{R}$ of these M^c functions has the following properties:

S1. $\mathcal{R}$ is a union of $\mathfrak{L}_+$-invariant sheets $\mathfrak{U}_a$; i.e., if the image in $\mathcal{K}_3$ of a sheet $\mathfrak{U}_a$ contains a point k, then this image contains all points on $\mathfrak{L}_+k$.

S2. The M^c functions are $\mathfrak{L}_+$ covariant on each sheet $\mathfrak{U}_a$. That is, if the image in $\mathcal{K}_3$ of a sheet $\mathfrak{U}_a$ contains a point k, then

$$M^c(\Lambda(A, B)K) = \Lambda_s(A, B)M^c(K), \tag{11}$$

for each $\Lambda(A, B) \in \mathfrak{L}_+$.

S3. Any bounded connected set of physical points on $\mathcal{R}$ is contained in some single $\mathfrak{L}_+$-invariant sheet $\mathfrak{U}_a$.[23]

For a process involving four spinless particles, one may replace $\mathfrak{L}_+$ covariant by $\mathfrak{L}_+$ invariant in statement S2. A function $F(k)$ is $\mathfrak{L}_+$ invariant on a set of points $\mathcal{S}$ in $\mathcal{K}_3$ if it satisfies (10) with $\Lambda_s(A, B) \equiv I$ whenever k and $\Lambda(A, B)k$ are in $\mathcal{S}$.

B. $\mathfrak{L}_+$ Orbits in $\mathcal{K}_3$

We intend to investigate the constraints that $\mathfrak{L}_+$ covariance, as specified by Stapp's theorem, imposes at certain points on the domain of regularity of the M^c functions. But first we need some properties of points k in $\mathcal{K}_3$ and their $\mathfrak{L}_+$ orbits.

At any point $k = \{k_1, \cdots, k_4\}$ in $\mathcal{K}_3$ we may define the scalar invariants $k_i \cdot k_j$. For the case of an arbitrary number of particles, one should also consider the pseudoscalar invariants formed by contracting the completely antisymmetric tensor $\epsilon^{\mu\nu\lambda\rho}$ with the 4-momenta. The invariants taken together are then referred to as $\mathfrak{L}_+$ invariants, since they are invariant under any $\Lambda \in \mathfrak{L}_+$, while the scalars alone are referred to as $\mathfrak{L}$ invariants, since they are invariant under any $\Lambda \in \mathfrak{L} = \mathfrak{L}_+ \cup \mathfrak{L}_-$, where $\mathfrak{L}_-$ is the set of improper Lorentz transformations. In the case under consideration, the pseudoscalars vanish identically because 4-momentum conservation allows at most three of the momenta to be linearly independent at any point. Consequently, it makes no difference whether we refer to two distinct points as having the same $\mathfrak{L}_+$ invariants or the same $\mathfrak{L}$ invariants, and we will use the former of these two terms.

All points on the same $\mathfrak{L}_+$ orbit have the same $\mathfrak{L}_+$ invariants. But one cannot always specify orbits by the values of their invariants, since two distinct points with the same $\mathfrak{L}_+$ invariants do not necessarily lie on the same $\mathfrak{L}_+$ orbit, as will be discussed below.

Let n be the number of linearly independent vectors at the point $k = \{k_1, \cdots, k_4\}$, where the vectors are ordered so that the first n are linearly independent. Because of 4-momentum conservation, $n \leq 3$, and, because the mass shell condition prevents the 4-momenta from vanishing identically, $n \geq 1$. Let us

define the Gram determinant

$$G(k_1, k_2, k_3) = \det \begin{pmatrix} (m_1)^2 & k_1 \cdot k_2 & k_1 \cdot k_3 \\ k_2 \cdot k_1 & (m_2)^2 & k_2 \cdot k_3 \\ k_3 \cdot k_1 & k_3 \cdot k_2 & (m_3)^2 \end{pmatrix}, \quad (12)$$

and let r be the rank of this determinant at the point k. Hall and Wightman[12] gave the following relationship between the rank of the Gram determinant at any point and the number of linearly independent 4-vectors at that point.

r	n
3	3
2	2 or 3
1	1 or 2

The possibility of having $n > r$ is, as will also be seen below, a consequence of the fact that one can have complex lightlike vectors in the space orthogonal to the first n vectors when $n < 3$.

From the considerations of Hall and Wightman regarding the properties of complex 4-vectors, we can make the following remarks about points and orbits in $\mathcal{K}_3$.

H1. (a) At any $r = n = 3$, 2, or 1 point, one can write

$$k_i = \sum_{j=1}^{r} \mathcal{A}_{ij} k_j, \quad \text{for} \quad i = r+1, \cdots, 3, \quad (13)$$

where the $\mathcal{A}_{ij}$ are finite scalar coefficients. (Recall that the first $r = n$ vectors are linearly independent at the given point and that k_4 is globally determined by 4-momentum conservation.)

(b) If k and k' are any two $r = n$ points with the same $\mathfrak{L}_+$ invariants, they lie on the same $\mathfrak{L}_+$ orbit.

H2. (a) There exist $r = 2$, $n = 3$ points with the same $\mathfrak{L}_+$ invariants as any given $r = n = 2$ point. For example, consider the point k determined by (13) with $r = 2$. In the space orthogonal to the two linearly independent vectors k_1 and k_2, one may define unit vectors $\hat{e}^1$ and $\hat{e}^2$ such that $\hat{e}^i \cdot \hat{e}^j = -\delta_{ij}$. Then define

$$\omega_\pm = \hat{e}^1 \pm i\hat{e}^2. \quad (14)$$

It follows that $\omega_\pm \cdot k_1 = \omega_\pm \cdot k_2 = \omega_\pm \cdot \omega_\pm = 0$. Now consider two points $k^{(+)}$ and $k^{(-)}$ such that $k^{(\pm)} = \{k_1, k_2, k_3^{(\pm)}, k_4^{(\pm)}\}$, where

$$k_3^{(\pm)} = k_3 + C\omega_\pm = \sum_{i=1}^{2} \mathcal{A}_{3i} k_i + C\omega_\pm. \quad (15)$$

Here k_1, k_2, and k_3 are the same as for the $r = n = 2$ point in (13), while $C \neq 0$ is an arbitrary real or complex number. The points $k^{(+)}$ and $k^{(-)}$ are two distinct $n = 3$ points with the same $\mathfrak{L}_+$ invariants as the $r = n = 2$ point obtained by putting $C = 0$ in (15). They are related by an improper Lorentz transformation that changes ω_+ into ω_-, while leaving k_1 and k_2 the same.

(b) The points $k^{(+)}$ and $k^{(-)}$ determined by (15) and the $r = n = 2$ point obtained by putting $C = 0$ in that equation all lie on different $\mathfrak{L}_+$ orbits. Any other $r = 2$, $n = 3$ point with the same $\mathfrak{L}_+$ invariants lies on either the $\mathfrak{L}_+$ orbit of $k^{(+)}$ or the $\mathfrak{L}_+$ orbit of $k^{(-)}$. Any point on one of these two orbits is related to any point on the other by means of an improper Lorentz transformation. As a consequence of H1(b), we may state that, for any set of values of the $\mathfrak{L}_+$ invariants for which the rank of the Gram determinant is 2, there exist three different $\mathfrak{L}_+$ orbits.

(c) Consider the limit $C \to 0$ in (15), which yields an $r = n = 2$ point with the same $\mathfrak{L}_+$ invariants. This means that any neighborhood of an $r = n = 2$ point contains points of every $r = 2$, $n = 3$ orbit with the same $\mathfrak{L}_+$ invariants.

H3. (a) Similar remarks enable one to construct an infinite number of $r = 1$, $n = 2$ orbits with limit points on a given $r = n = 1$ orbit. In this case, however, any two $r = 1$, $n = 2$ points related by an improper Lorentz transformation lie on the same $\mathfrak{L}_+$ orbit.

(b) The occurrence of $r = 1$ points in $\mathcal{K}_3$ is possible only if the sum of some of the masses equals the sum of the others. This follows from the conservation of energy and the fact that every $r = n = 1$ orbit contains a point of the form $k_i = (\pm m_i, 0)$ for all $k_i \in k$.

H4. We define the little group $\mathfrak{G}_+(k)$ of a point k to be the set of proper complex Lorentz transformations that leave k invariant; i.e., $\Lambda \in \mathfrak{G}_+(k) \to \det \Lambda = 1$ and $\Lambda k = k$. At any point k with $n = 3$, the only matrix in $\mathfrak{G}_+(k)$ is the unit matrix. However, if $r = n \leq 2$, $\mathfrak{G}_+(k)$ is an infinite set.

C. The I_+-Saturated Kernel of the Domain of Regularity

We emphasize the fact that to a given point $k \in \mathcal{K}_3$ there can correspond many points on the domain of regularity $\mathcal{R}$ of the M^c functions for a given process, although at most one point on any sheet $\mathcal{U}_a \subset \mathcal{R}$. In the remainder of this paper, when we speak of a point k with certain values of r and n lying on $\mathcal{R}$, we actually mean a point on $\mathcal{R}$ whose image k in $\mathcal{K}_3$ has these values of r and n.

Thus Stapp's theorem in Part A of this section says essentially that (a) $\mathcal{R}$ is a union of $\mathcal{L}_+$ orbits, i.e., if a point k lies on $\mathcal{R}$, there is an $\mathcal{L}_+$ orbit $\mathcal{L}_+k$ that lies on $\mathcal{R}$ and contains the given point, and (b) the M^c functions are $\mathcal{L}_+$ covariant on $\mathcal{R}$, i.e., the M^c functions at any two points k and $\Lambda(A, B)k$ on the same $\mathcal{L}_+$ orbit on $\mathcal{R}$ are related by (11). [If all four particles are spinless, the M^c function is $\mathcal{L}_+$ invariant on $\mathcal{R}$, i.e., $\Lambda_s(A, B) \equiv I$ in (11).]

If an $r = n < 3$ point k' lies on $\mathcal{R}$, then every $r \neq n$ orbit for which it is a limit point lies on $\mathcal{R}$. This is because there is some full neighborhood $\mathcal{N}$ in $\mathcal{R}$ of the point such that there is a one-to-one mapping between points in $\mathcal{N}$ and points in a neighborhood $N(k') \in \mathcal{K}_3$—according to H.2(c) and H.3(a), $N(k')$ contains points of every $r \neq n$ orbit in $\mathcal{K}_3$ for which the $r = n$ point $k' \in N(k')$ is a limit point. By Stapp's theorem $\mathcal{R}$ must contain the full $\mathcal{L}_+$ orbit of any point in $\mathcal{N}$.

If an $r \neq n$ point lies on $\mathcal{R}$, the $r = n$ limit points of its $\mathcal{L}_+$ orbit do not necessarily lie on $\mathcal{R}$.[1] The *I_+-saturated kernel* $\mathcal{R}^{(+)}$ of $\mathcal{R}$ is the subset obtained by deleting from $\mathcal{R}$ all $r \neq n$ orbits whose $r = n$ limit points do not lie on $\mathcal{R}$.[1,5,11] All physical points on $\mathcal{R}$ lie on $\mathcal{R}^{(+)}$ because their image in $\mathcal{K}_3$ is real and the construction in (14) and (15) shows that $r \neq n$ points in $\mathcal{K}_3$ are always complex.

In the remainder of this paper the symbol $\mathcal{U}_a^{(+)}$ designates the set $\mathcal{U}_a \cap \mathcal{R}^{(+)}$, where $\mathcal{U}_a$ is some $\mathcal{L}_+$-invariant sheet on $\mathcal{R}$; i.e., $\mathcal{U}_a^{(+)}$ is the set obtained by deleting from $\mathcal{U}_a$ all $r \neq n$ orbits whose $r = n$ limit points do not lie on $\mathcal{R}$. We refer to $\mathcal{U}_a^{(+)}$ as the I_+-saturated kernel of the $\mathcal{L}_+$-invariant sheet $\mathcal{U}_a$.[25]

D. Kinematical Restrictions

The number of M^c functions for a given process is the same as the number of different combinations of values of the spinor indices. This is given by

$$N = \prod_{i=1}^{4} (2j_i + 1). \tag{16}$$

Equation (16) gives the number of independent scattering experiments at a fixed physical value of the 4-momenta, at least on a dense subset of the physical points. Of course, for the M^c functions under consideration discrete symmetries can lead to a relation between the results of various experiments, so that the number that are independently determined are less than the number in (16). Such restrictions will be ignored in this section and will be the concern of the next one.

We will show that, at any point on $\mathcal{R}^{(+)}$ at which the rank of the Gram determinant is less than 3, $\mathcal{L}_+$ covariance leads to linear relationships between the M^c functions; i.e., there are kinematical restrictions at such a point.[26] For any $r = n = 2$ point k on $\mathcal{R}$, this statement follows by letting $\Lambda(\bar{A}, \bar{B})$ be a matrix in the little group, $\mathcal{G}_+(k)$, defined in remark H4. Then the $\mathcal{L}_+$ covariance relation (11) becomes

$$M^c(K) = \Lambda_s(\bar{A}, \bar{B}) M^c(K). \tag{17}$$

We also find kinematical restrictions at any $r = 2$, $n = 3$ point lying on $\mathcal{R}^{(+)}$. The trivial extension of our results to $r = 1$ points will not be needed in this paper.

We restrict our attention to a single point on any given orbit on $\mathcal{R}^{(+)}$, since the number of M^c functions whose values are independent is the same at all points on the orbit. Any $r = n = 2$ orbit on $\mathcal{R}$ contains a point whose image $k \in \mathcal{K}_3$ is such that each of the vectors $k_i \in k$ has no components along the 1 and 2 axes. Then, from (A1),

$$\sigma \cdot k_i = \begin{pmatrix} (k_i)^0 + (k_i)^3 & 0 \\ 0 & (k_i)^0 - (k_i)^3 \end{pmatrix}, \quad i = 1, \cdots, 4. \tag{18}$$

Let us now make the following choice for the matrices $\bar{A}$ and $\bar{B} \in SL(2, C)$, where λ is any complex parameter:

$$\bar{A} = \begin{pmatrix} \exp(-\lambda/2) & 0 \\ 0 & \exp(\lambda/2) \end{pmatrix}, \tag{19a}$$

$$\bar{B} = \begin{pmatrix} \exp(\lambda/2) & 0 \\ 0 & \exp(-\lambda/2) \end{pmatrix}. \tag{19b}$$

From (A2), it follows that $\Lambda(\bar{A}, \bar{B})k_i = k_i$ for all $i = 1, \cdots, 4$, when k is a point in $\mathcal{K}_3$ of the form in (18), so that $\Lambda(\bar{A}, \bar{B})$ belongs to the little group $\mathcal{G}_+(k)$.

Because of (A3), we have

$$D^{(j_i)}(\bar{A}) = \exp(-\lambda J_3^{(j_i)}), \tag{20a}$$

$$D^{(j_i)}(\bar{B}) = \exp(\lambda J_3^{(j_i)}). \tag{20b}$$

Suppose that the particles are ordered so that the first l have lower undotted spinor indices and the remaining $4 - l$ have lower dotted ones, the assignment of index types to individual particles being completely arbitrary. Then, because of (17) and (20), we have, at any point k of the form (18) lying on $\mathcal{R}$,

$$M^c_{(\alpha)(\dot{\beta})}(K) = \exp\left[-\lambda\left(\sum_{s=1}^{l} \alpha_s - \sum_{t=l+1}^{4} \beta_t\right)\right] M^c_{(\alpha)(\dot{\beta})}(K), \tag{21}$$

where $(\alpha) = \alpha_1 \cdots \alpha_l$, and $(\dot{\beta}) = \dot{\beta}_{l+1} \cdots \dot{\beta}_4$, are not to be confused with the outgoing and incoming spinor indices of the preceding section. Equation (21) requires that

$$M^c_{(\alpha)(\dot{\beta})}(K) = 0 \quad \text{if} \quad \sum_s \alpha_s - \sum_t \beta_t \neq 0. \tag{22}$$

Therefore, the M^c functions have "kinematical zeros" at any such point. At other $r = n = 2$ points the relationship among the values of the M^c functions, as given by (17), will be more complicated, but the number of such linear relationships will be the same as the number of "zeros" in (22). It is convenient to continue to use the term "kinematical zeros" to refer to the restrictions at these latter points.

Now consider any $r = 2$, $n = 3$ orbit on $\mathcal{R}^{(+)}$ for which the $r = n = 2$ point k on $\mathcal{R}$ that we have just considered is a limit point. There is a point on the orbit whose image $k^{(+)}$ or $k^{(-)}$ in $\mathcal{K}_3$ is such that each vector $k_i^{(\pm)} \in k^{(\pm)}$ has the form $k_i^{(\pm)} = k_i + C_i\omega_\pm$, where k_i is given by (18), C_i is a real or complex number, and $\omega_\pm = \hat{e}^1 \pm i\hat{e}^2$ is a complex lightlike vector in the space orthogonal to the k_i. Choosing the real spacelike vectors $\hat{e}^1$ and $\hat{e}^2$ to be parallel to the 1 and 2 axes respectively, we have from (18) and (A1)

$$\sigma \cdot k_i^{(\pm)} = \begin{pmatrix} (k_i)^0 + (k_i)^3 & (1 \pm 1)C_i \\ (1 \mp 1)C_i & (k_i)^0 - (k_i)^3 \end{pmatrix}. \quad (23)$$

Using $\Lambda(\bar{A}, \bar{B})$ defined by (19) and (A2), but restricting ourselves to real values of λ, we find that

$$\lim_{\lambda \to \pm\infty} \Lambda(\bar{A}, \bar{B})k_i^{(\pm)} = k_i, \quad i = 1, \cdots, 4, \quad (24)$$

with k_i given by (18). Since the M^c functions are continuous at the point k on $\mathcal{R}$, we have, using (11), (20), (24), and the obvious notation $K_i^{(\pm)} = (k_i^{(\pm)}, t_i)$,

$$\lim_{\lambda \to \pm\infty} M^c_{(\alpha)(\dot{\beta})}(\Lambda(\bar{A}, \bar{B})K^{(\pm)}) = M^c_{(\alpha)(\dot{\beta})}(K) = \lim_{\lambda \to \pm\infty} \exp\left[-\lambda\left(\sum_s \alpha_s - \sum_t \beta_t\right)\right] M^c_{(\alpha)(\dot{\beta})}(K^{(\pm)}). \quad (25)$$

In particular, (25) implies that

$$M^c_{(\alpha)(\dot{\beta})}(K^{(\pm)}) = M^c_{(\alpha)(\dot{\beta})}(K) \quad \text{if} \quad \sum_s \alpha_s = \sum_t \beta_t. \quad (26)$$

Furthermore, for the limit in (25) to be consistent with the result in (22) for the $r = n = 2$ point k on $\mathcal{R}$, we must have

$$M^c_{(\alpha)(\dot{\beta})}(K^{(+)}) = 0 \quad \text{if} \quad \sum_s \alpha_s < \sum_t \beta_t, \quad (27a)$$

$$M^c_{(\alpha)(\dot{\beta})}(K^{(-)}) = 0 \quad \text{if} \quad \sum_s \alpha_s > \sum_t \beta_t. \quad (27b)$$

However, there is no kinematical restriction on the values of the M^c functions with $\sum_s \alpha_s > \sum_t \beta_t$ at a point on $\mathcal{R}^{(+)}$ whose image in $\mathcal{K}_3$ is $k^{(+)}$ and no such restriction on those with $\sum_s \alpha_s < \sum_t \beta_t$ at a point whose image is $k^{(-)}$. Either of Eqs. (27) is therefore sufficient to determine the number of kinematical zeros at an $r = 2$, $n = 3$ point on $\mathcal{R}^{(+)}$, this number being exactly half the number given by (22) for an $r = n = 2$ point on $\mathcal{R}$.[27]

By adding up the number of M^c functions not restricted by (22) or (27), we get the number of such functions whose values are free of kinematical constraints at any $r = 2$ point on $\mathcal{R}^{(+)}$. The result is given in Table I, which is actually valid for M^c functions with any number of particles, but for which at most four have spin.[4]

TABLE I. Number of independent M^c functions at an $r = 2$ point. These results are valid for the case when at most four particles have spins, although the total number of particles can be arbitrary, and are valid on the I_+-saturated kernel of the domain of regularity. Here r = rank of Gram determinant. We take $j_1 + j_2$ and $j_3 + j_4$ both to be integers, with $j_1 + j_2 \geq j_3 + j_4$ and $j_1 \geq j_2, j_3 \geq j_4$.

	Number of linearly independent 4-vectors, n	
	$n = 2$	$n = 3$
Case I $j_1 - j_2 \geq j_3 + j_4$	$(2j_2 + 1)(2j_3 + 1)(2j_4 + 1)$	$(j_1 + 1)(2j_2 + 1)(2j_3 + 1)(2j_4 + 1)$
Case II $j_3 + j_4 \geq j_1 - j_2$ $j_1 - j_2 \geq j_3 - j_4$	$(2j_2 + 1)(2j_3 + 1)(2j_4 + 1)$ $-\frac{1}{2}(j_2 + j_3 + j_4 - j_1)(j_2 + j_3 + j_4 - j_1 + 1)$ $\times (j_2 + j_3 + j_4 - j_1 + 2)$	$(j_1 + 1)(2j_2 + 1)(2j_3 + 1)(2j_4 + 1)$ $-\frac{1}{2}(j_2 + j_3 + j_4 - j_1)(j_2 + j_3 + j_4 - j_1 + 1)$ $\times (j_2 + j_3 + j_4 - j_1 + 2)$
Case III $j_3 - j_4 \geq j_1 - j_2$	$(2j_4 + 1)[(2j_2 + 1)(2j_3 + 1)$ $- \frac{1}{3}j_4(j_4 + 1) - (j_2 + j_3 - j_4 - j_1)$ $\times (j_2 + j_3 + j_4 - j_1 + 1)]$	$(2j_4 + 1)[(j_1 + 1)(2j_2 + 1)(2j_3 + 1)$ $- \frac{2}{3}j_4(j_4 + 1) - \frac{1}{2}(j_2 + j_3 - j_4 - j_1)$ $\times (j_2 + j_3 + j_4 - j_1 + 1)]$

IV. DISCRETE SYMMETRIES

A. Identical Particles, *PCT*, *P*, *T*, and *C*

In this section we determine the number of independent M^c functions at various points on $\mathcal{R}^{(+)}$ if the functions have definite signature under the discrete symmetry operations. The main object is to establish the results at points where the number of linearly independent momenta n is 3, but the rank of the Gram determinant r is 2. The understanding derived from this discussion leads to the results of the following section.

Invariance under *PCT* and the connection between spin and statistics are both consequences of the basic principles of *S*-matrix theory[19] and of field theory.[24,28] In terms of the connected parts of the S matrix in canonical form, the *PCT* identity reads

$$PCT: \quad \underset{\alpha_b;\,\dot\alpha_a}{S^c(K_b;\,K_a)} = \eta_{PCT}\prod_{i=1}^{4} D^{(j_i)}(C)_{\alpha_i}{}^{\alpha_i'}\,\underset{\alpha_a';\,\dot\alpha_b'}{S^c(\bar K_a;\,\bar K_b)}, \tag{28}$$

where η_{PCT} is a phase factor. Here $\bar K_i = (k_i, \bar t_i)$, with $\bar t_i = (m_i, j_i, -q_i)$ indicating an antiparticle. The spin–statistics connection states that exchanging the order of the momentum variables and spin components of any two identical initial or any two identical final particles of spin j changes the sign of the scattering function by $(-1)^{2j}$.

The assumption that transition probabilities are invariant under a change of direction of all spatial components of the 4-momenta leads to the relation

$$P: \quad \underset{\alpha_b;\,\dot\alpha_a}{S^c(K_b;\,K_a)} = \eta_p \underset{\alpha_b;\,\dot\alpha_a}{S^c(\tilde K_b;\,\tilde K_a)} \tag{29}$$

in the physical region of the process $t_1 + t_2 \to t_3 + t_4$. Here $\tilde K_i = (\tilde k_i, t_i)$, with $\tilde k_i = (k_i^0, -\mathbf{k}_i)$, and $\eta_p = \pm 1$, the "process intrinsic parity," is the product of the "particle intrinsic parities" of the particles occurring in the process.[29]

It is straightforward to show that if transition probabilities are invariant under time reversal, which involves exchanging initial and final states and changing the sign of all three vectors, one has for physical points

$$T: \quad \underset{\alpha_b;\,\dot\alpha_a}{S^c(K_b;\,K_a)} = \eta_T D^{(j_a)}(C)_{\alpha_a}{}^{\alpha_a'} D^{(j_b)}(C)_{\dot\alpha_b}{}^{\dot\alpha_b'}\,\underset{\alpha_a';\,\dot\alpha_b'}{S^c(\tilde K_a;\,\tilde K_b)}, \tag{30}$$

where $\eta_T = +1$ is required for an elastic process.

It may happen that *PT* is a symmetry, even if *P* and *T* are not. Then the *PCT* identity (28) requires charge-conjugation invariance,

$$C: \quad \underset{\alpha_b;\,\dot\alpha_a}{S^c(K_b;\,K_a)} = \eta_C \underset{\alpha_b;\,\dot\alpha_a}{S^c(\bar K_b;\,\bar K_a)}. \tag{31}$$

The relevant symmetry relations for M^c functions with spinor indices of the types introduced in (4) and (6) follow easily from (28)–(31) and are given in Table II. If a given symmetry is valid for a certain physical process, it is valid for the analytically continued functions and, consequently, for the processes related by crossing.[30]

B. Functions with Definite Parity Signature

Even when spatial inversion is not a symmetry of the process under consideration, one can find it useful to decompose the M^c functions for two incoming and two outgoing particles into parts having positive and negative parity signature. In the physical region of the process $t_1 + t_2 \to t_3 + t_4$, let us define the functions

$$\underset{\alpha_b;\,\alpha_a}{M_\epsilon^c(K_b;\,K_a)} = \tfrac{1}{2}\Big[\underset{\alpha_b;\,\alpha_a}{M^c(K_b;\,K_a)} + \epsilon D^{(j_b)}\Big(\frac{\sigma\cdot k_b}{m_b}\Big)_{\alpha_b\alpha_b'} \times D^{(j_a)}\Big(\frac{\sigma\cdot k_a}{m_a}\Big)_{\alpha_a\alpha_a'}\underset{\alpha_b';\,\alpha_a'}{M^c(\tilde K_b;\,\tilde K_a)}\Big], \tag{32}$$

for $\epsilon = \pm 1$. With the aid of (A12) we find that

$$\underset{\alpha_b;\,\alpha_a}{M_\epsilon^c(K_b;\,K_a)} = \epsilon D^{(j_b)}\Big(\frac{\sigma\cdot k_b}{m_b}\Big)_{\alpha_b\alpha_b'} D^{(j_a)}\Big(\frac{\sigma\cdot k_a}{m_a}\Big)_{\alpha_a\alpha_a'}\underset{\alpha_b';\,\alpha_a'}{M_\epsilon^c(\tilde K_b;\,\tilde K_a)}, \tag{33}$$

which indicates that the functions defined by (32) have definite parity signature (see Table II).

One may analytically continue the functions M_+^c and M_-^c defined by (32) over all of $\mathcal{R}^{(+)}$, and the following decomposition is valid at all points k on that domain:

$$\underset{\alpha_b;\,\alpha_a}{M^c(K_b;\,K_a)} = \underset{\alpha_b;\,\alpha_a}{M_+^c(K_b;\,K_a)} + \underset{\alpha_b;\,\alpha_a}{M_-^c(K_b;\,K_a)}. \tag{34}$$

By comparing (33) and (34) with Table II, we see that if spatial inversion symmetry is valid with $\eta_p = \pm 1$, then $M_\mp^c \equiv 0$.

According to the remarks in Sec. IIIB, the points k and $\tilde k$ in the analytically continued relation (33) lie on the same $\mathfrak{L}_+$ orbit on $\mathcal{R}^{(+)}$ unless they are $r = 2$, $n = 3$ points. If they are $r = 2$, $n = 3$ points, then remark H2(b) shows that they lie on two different $\mathfrak{L}_+$ orbits on $\mathcal{R}^{(+)}$ having the same $r = n = 2$ limit points. It is well known that $\mathfrak{L}_+$ covariance leads to

TABLE II. Effect of invariance under various discrete symmetry operations on the M^c functions. N_a = number of fermions in state a.

Symmetry	$M^c(K_b; K_a)$ $\alpha_b; \alpha_a$	$M^c(K_b; K_a)$ $\alpha_b; \dot{\alpha}_a$
PCT	$= (-1)^{N_a}\eta_{PCT}M^c(\bar{K}_a; \bar{K}_b)$ $\alpha_a; \alpha_b$	$= (-1)^{N_a}\eta_{PCT}M^c(\bar{K}_a; \bar{K}_b)$ $\dot{\alpha}_a; \alpha_b$ $= \eta_{PCT}D^{(j_a)}\left(C\frac{\tilde{\sigma}\cdot k_a}{m_a}\right)^{\alpha_a'}_{\alpha_a} D^{(j_b)}\left(\frac{\sigma\cdot k_b}{m_b}C\right)^{\alpha_b'}_{\alpha_b} M^c(\bar{K}_a; \bar{K}_b)$ $\alpha_a'; \dot{\alpha}_b'$
P	$\dot{\alpha}_b; \dot{\alpha}_a$ $= \eta_P M^c(\tilde{K}_b; \tilde{K}_a)$ $= \eta_P D^{(j_b)}\left(\frac{\sigma\cdot k_b}{m_b}\right)_{\alpha_b\alpha_b'} D^{(j_a)}\left(\frac{\sigma\cdot k_a}{m_a}\right)_{\alpha_a\alpha_a'} M^c(\tilde{K}_b; \tilde{K}_a)$ $\alpha_b'; \alpha_a'$	$\dot{\alpha}_b; \alpha_a$ $= (-1)^{N_a}\eta_P M^c(\tilde{K}_b; \tilde{K}_a)$ $= \eta_P D^{(j_b)}\left(\frac{\sigma\cdot k_b}{m_b}\right)_{\alpha_b\alpha_b'} D^{(j_a)}\left(\frac{\sigma^T\cdot k_a}{m_a}\right)_{\alpha_a\alpha_a'} M^c(\tilde{K}_b; \tilde{K}_a)$ $\alpha_b'; \dot{\alpha}_a'$
T	$\dot{\alpha}_a; \dot{\alpha}_b$ $= (-1)^{N_a}\eta_T M^c(\tilde{K}_a; \tilde{K}_b)$ $= (-1)^{N_a}\eta_T D^{(j_a)}\left(\frac{\sigma\cdot k_a}{m_a}\right)_{\alpha_a\alpha_a'} D^{(j_b)}\left(\frac{\sigma\cdot k_b}{m_b}\right)_{\alpha_b\alpha_b'} M^c(\tilde{K}_a; \tilde{K}_b)$ $\alpha_a'; \alpha_b'$	$= \eta_T D^{(j_a)}(C)_{\alpha_a}{}^{\alpha_a'} D^{(j_b)}(C)_{\alpha_b}{}^{\alpha_b'} M^c(\tilde{K}_a; \tilde{K}_b)$ $\alpha_a'; \dot{\alpha}_b'$
PT	$= (-1)^{N_a}\eta_{PT}M^c(K_a; K_b)$ $\alpha_a; \alpha_b$	$= (-1)^{N_a}\eta_{PT}M^c(K_a; K_b)$ $\dot{\alpha}_a; \alpha_b$ $= \eta_{PT}D^{(j_a)}\left(C\frac{\tilde{\sigma}\cdot k_a}{m_a}\right)^{\alpha_a'}_{\alpha_a} D^{(j_b)}\left(\frac{\sigma\cdot k_b}{m_b}C\right)^{\alpha_b'}_{\alpha_b} M^c(K_a; K_b)$ $\alpha_a'; \dot{\alpha}_b'$
C	$= \eta_c M^c(\bar{K}_b; \bar{K}_a)$ $\alpha_b; \alpha_a$	$= \eta_c M^c(\bar{K}_b; \bar{K}_a)$ $\alpha_b; \dot{\alpha}_a$

linear relations between the functions having the same parity signature at all $r = n$ points. We will review this result in order to extend it to $r = 2$, $n = 3$ points.

Because of the $D^{(j_i)}(\sigma \cdot k_i/m_i)$ matrices that act on the $M_\epsilon{}^c$ functions under the parity operation defined by (33), it is much simpler to use the connected parts of the S-matrix elements to count the number of independent functions. The S^c functions have kinematical singularities arising from the "boost" matrices that relate them to the M^c functions, as in (4). These singularities, unlike those of the M^c functions, are not Lorentz invariant, so we can always find points on any orbit at which the S^c functions are analytic if the M^c functions are holomorphic.[31]

By application of the proper boost matrices to (32) and (33), we obtain

$$S_\pm{}^c(K_b; K_a) = \tfrac{1}{2}[S^c(K_b; K_a) \pm S^c(\tilde{K}_b; \tilde{K}_a)], \quad (35)$$
$$\alpha_b; \dot{\alpha}_a \qquad \alpha_b; \dot{\alpha}_a \qquad \alpha_b; \dot{\alpha}_a$$

where

$$S_\epsilon{}^c(K_b; K_a) = \epsilon S_\epsilon{}^c(\tilde{K}_b; \tilde{K}_a). \quad (36)$$
$$\alpha_b; \dot{\alpha}_a \qquad \alpha_b; \dot{\alpha}_a$$

To see the restrictions at $r = n = 3$ or $r = n = 2$ points on $\mathfrak{R}$, it is best to choose a point $k = \{k_b; k_a\}$ on a given orbit such that all spatial components are normal to the 2 axis; then a rotation of π about this axis carries $\tilde{k} \in \mathfrak{K}_3$ into k. Since the matrix $A(k_i)$ in (A11) corresponding to the Wigner rotation is equal to A when A is unitary and since the required rotation matrix for spin j_i is now simply the matrix $D^{(j_i)}(C)$ in (A5), Eq. (3) gives us

$$S_\epsilon{}^c(K_b; K_a) = \epsilon(-1)^{\sum_{i=1}^4 (j_i - \alpha_i)} S_\epsilon{}^c(K_b; K_a). \quad (37)$$
$$\alpha_b; \dot{\alpha}_a \qquad\qquad -\alpha_b; -\dot{\alpha}_a$$

Equation (37) allows us to conclude that at any $r = 3$ point on the domain of regularity of the M^c functions, disregarding possible restrictions due to other symmetry operations, the number of $M_\epsilon{}^c$ functions whose values are independent is

$$N_\epsilon = \tfrac{1}{2}\prod_{i=1}^4 (2j_i + 1) \quad (38)$$

if there are fermions involved in the process. On the other hand, the number is

$$N_\epsilon = \tfrac{1}{2}\left(\prod_{i=1}^4 (2j_i + 1) + \epsilon(-1)^{\sum_{i=1}^4 j_i}\right) \quad (39)$$

if all the particles are bosons.

In order to obtain the restrictions at $r = n = 2$ points on $\mathcal{R}$, we note that, in terms of the $S_\epsilon{}^c$ functions, (22) becomes

$$S_\epsilon{}^c(K_b; K_a) = 0 \quad \text{if} \quad \alpha_1 + \alpha_2 \neq \alpha_3 + \alpha_4. \quad (40)$$
$$\alpha_b; \dot{\alpha}_a$$

Then (37) restricts the number of $M_\epsilon{}^c$ functions whose values are independent to exactly half the number allowed by (40) alone, if some of the particles are fermions, and to that number plus $\frac{1}{2}\epsilon(-1)^{\sum_{i=1}^4 j_i}$, if all the particles are bosons. The number allowed by (40) alone has already been given in Table I.

From remark H2(b) in Sec. III, we know that an $r = 2, n = 3$ point is related to its spatial inverse only by an improper Lorentz transformation. However, on $\mathcal{R}^{(+)}$ the limit (25) must be valid for the $M_\epsilon{}^c$ functions also and we do get restrictions at $r = 2, n = 3$ points. In particular, in place of (26) we have

$$M_\epsilon{}^c(K_b; K_a) = M_\epsilon{}^c(K_b{}^{(\pm)}; K_a{}^{(\pm)}), \quad \text{for} \quad \sum_{i=1}^4 \alpha_i = 0, \quad (41)$$
$$\alpha_b; \alpha_a \qquad\qquad \alpha_b; \alpha_a$$

if $k^{(\pm)}$, as defined by (23), is an $r = 2, n = 3$ point on $\mathcal{R}^{(+)}$, and k is an $r = n = 2$ limit point of $\mathcal{L}_+ k^{(\pm)}$ of the form (18).

Similarly, in place of (27) we get

$$M_\epsilon{}^c(K_b{}^{(+)}; K_a{}^{(+)}) = 0 \quad \text{if} \quad \sum_{i=1}^4 \alpha_i < 0, \quad (42a)$$
$$\alpha_b; \alpha_a$$

$$M_\epsilon{}^c(K_b{}^{(-)}; K_a{}^{(-)}) = 0 \quad \text{if} \quad \sum_{i=1}^4 \alpha_i > 0, \quad (42b)$$
$$\alpha_b; \alpha_a$$

but there are no restrictions on the components not accounted for by (41) or (42).

By inspection of (41), we see that at the $r = 2$, $n = 3$ point under consideration, the $M_\epsilon{}^c$ functions satisfying $\sum_i \alpha_i = 0$ are subject to the same restrictions as they are subject to at the $r = n = 2$ limit point of the orbit. Therefore, the number of independent functions with $\sum_i \alpha_i = 0$ is the same as the number mentioned in the sentence following (40). Adding to this the number of functions not restricted by either (41) or (42), we find that the number of $M_\epsilon{}^c$ functions whose values are not subject to any linear restriction among themselves is exactly the same at any $r = 2, n = 3$ point on $\mathcal{R}^{(+)}$ as the number at any $r = 3$ point on $\mathcal{R}$—this number is given by either (38) or (39).

C. Restrictions in Special Cases

Invariance of a scattering process under any discrete symmetry other than spatial inversion can restrict the number of independent M^c functions only in special cases in which the point in momentum space resulting from the symmetry operation lies on the same $\mathcal{L}_+$ orbit as the original point. We will consider the restrictions for $r = 3$ points by working in the physical region of a particular c.m. system with the 1 and 3 axes orientated as in Fig. 1 and with the 2 axis pointing out of the paper. The restrictions thus obtained are easily extended to all $r = 3$ points on $\mathcal{R}$ and also to the points on $\mathcal{R}^{(+)}$ with $r < 3$.

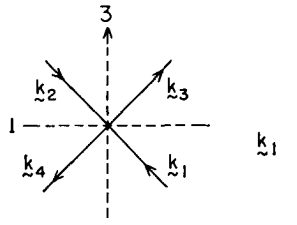

FIG. 1. Special reference frame used for obtaining discrete symmetry restrictions. The 2 axis points out of the paper.

For example, for a scattering process of the form $t_1 + t_1 \to t_3 + t_3$,the exchange symmetry for identical particles leads to a restriction when one simultaneously exchanges the orders of the initial particles among themselves and of the final particles among themselves. In the reference system of Fig. 1, a rotation of π about the 2 axis carries the transformed momenta back into the original orientation and we have

$$E: \quad S^c(K_3, K_4; K_1, K_2) = (-1)^{\sum_{i=1}^4 (j_i + \alpha_i)} S^c(K_3, K_4; K_1, K_2),$$
$$\alpha_3, \alpha_4; \dot{\alpha}_1, \dot{\alpha}_2 \qquad\qquad -\alpha_4, -\alpha_3; -\dot{\alpha}_2, -\dot{\alpha}_1$$
$$\text{for} \quad t_1 = t_2 \quad \text{and} \quad t_3 = t_4. \quad (43)$$

The functions for the crossed process $t_1 + \bar{t}_3 \to \bar{t}_1 + t_3$ have the same number of independent components as those allowed by (43). This restriction could also have been obtained by applying the *PCT* relation (28) in the new channel. For a process of the form $t + t \to t + t$ with $t = \bar{t}$, the *PCT* relation and the symmetry under the exchange of identical particles simultaneously lead to restrictions in the same channel. In the reference frame of Fig. 1, a rotation of π about the 3 axis carries the point on the right-hand side of (28) into that on the left and we obtain

$$PCT: \quad S^c(K_3, K_4; K_1, K_2) = S^c(K_3, K_4; K_1, K_2),$$
$$\alpha_3, \alpha_4; \dot{\alpha}_1, \dot{\alpha}_2 \qquad\qquad -\alpha_1, -\alpha_2; -\dot{\alpha}_3, -\dot{\alpha}_4$$
$$\text{for} \quad t_1 = t_2 = t_3 = t_4 = t = \bar{t}. \quad (44)$$

Time-reversal invariance restricts the number of independent components only in an elastic two-particle process. In the reference frame of Fig. 1, the required exchange of the 4-momenta is brought about

TABLE III. Number of independent $M_\epsilon{}^c$ functions having definite signature under T and PT for the process $t_1 + t_2 \to t_1 + t_2$. These results hold at any point on the I_+-saturated kernel of the domain of regularity where there are three linearly independent 4-momenta. Exceptional cases in which exchange symmetry or PTC (or both) must give a restriction are: (1) $t_1 = t_2$; (2) $t_1 = \bar{t}_2$; (3) both $t_1 = \bar{t}_1$ and $t_2 = \bar{t}_2$. These exceptions are either given directly by the processes in Tables IV and V or related to them through crossing.

(a) If at least one of the incoming particles is a fermion

$\epsilon = P$	T	PT	Number of terms
+	+	+	$\frac{1}{4}(2j_1+1)(2j_2+1)(4j_1j_2+2j_1+2j_2+3)$
+	−	−	$\frac{1}{4}(2j_1+1)(2j_2+1)(4j_1j_2+2j_1+2j_2-1)$
−	+	−	$\frac{1}{4}(2j_1+1)^2(2j_2+1)^2$
−	−	+	$\frac{1}{4}(2j_1+1)^2(2j_2+1)^2$

(b) If both particles are bosons

$\epsilon = P$	T	PT	Number of terms
+	+	+	$\frac{1}{4}[(2j_1+1)(2j_2+1)(4j_1j_2+2j_1+2j_2+3)+1]$
+	−	−	$\frac{1}{4}[(2j_1+1)(2j_2+1)(4j_1j_2+2j_1+2j_2-1)+1]$
−	+	−	$\frac{1}{4}[(2j_1+1)^2(2j_2+1)^2-1]$
−	−	+	$\frac{1}{4}[(2j_1+1)^2(2j_2+1)^2-1]$

by a rotation of π about the 1 axis and (30) becomes, with $\eta_T = 1$,

$$T: \quad S^c(K_3, K_4; K_1, K_2)\,{}^{\alpha_3, \alpha_4; \dot{\alpha}_1, \dot{\alpha}_2} = (-1)^{\alpha_1+\alpha_2-\alpha_3-\alpha_4} S^c(K_3, K_4; K_1, K_2)\,{}^{\alpha_1, \alpha_2; \dot{\alpha}_3, \dot{\alpha}_4} \quad \text{for} \quad t_1 = t_3 \quad \text{and} \quad t_2 = t_4. \quad (45)$$

Equations (37) and (43)–(45) are sufficient to find all possible restrictions at $r = 3$ points on $\mathcal{R}$. In Tables III, IV, and V we have listed the number of independent $M_\epsilon{}^c$ functions having definite signature under the relevant symmetry operations for those cases in which spatial inversion is not the only possible symmetry that can give a restriction. In Table IV we have allowed only those terms satisfying $(-1)^{2(j_1+j_3)} = 1$ when taking (43) into account. This is because the additive quantum numbers independent of the Lorentz group must be the same for t_1 and t_3 in a process of the form $t_1 + t_1 \to t_3 + t_3$, and in nature all strongly interacting particles with half odd-integer spin apparently have odd baryon number, while those with integer spin have even baryon number.

Of course, when spatial inversion is a symmetry for all processes in nature, only terms with $\epsilon = P = +1$ are nonvanishing in any of the tables,[29] and when time reversal is a symmetry for the elastic processes in Tables III and V, only terms with $T = +1$ are nonvanishing. If neither P nor T is a symmetry, but

TABLE IV. Number of independent $M_\epsilon{}^c$ functions for the process $t_1 + t_1 \to t_3 + t_3$ with $t_1 \neq t_3$. These results hold at any point on the I_+-saturated kernel of the domain of regularity where there are three linearly independent 4-momenta. Since the additive quantum numbers that are independent of the Lorentz group must be the same for t_1 and t_3, such a process apparently occurs in nature only when t_1 and t_3 are both fermions or both bosons. The table is valid even when $t_1 = \bar{t}_1$.

(a) When particles t_1 and t_3 are fermions

$\epsilon = P$	Number of terms
+	$\frac{1}{4}(2j_1+1)(2j_3+1)(4j_1j_3+2j_1+2j_3+3)$
−	$\frac{1}{4}(2j_1+1)^2(2j_3+1)^2$

(b) When t_1 and t_3 are bosons

$\epsilon = P$	Number of terms
+	$(2j_1j_3+j_1+j_3+1)^2$
−	$(2j_1j_3+j_1+j_3)(2j_1j_3+j_1+j_3+1)$

TABLE V. Number of independent $M_\epsilon{}^c$ functions having definite signature under T and PT for the process $t + t \to t + t$. These results hold at any point on the I_+-saturated kernel of the domain of regularity where there are three linearly independent 4-momenta.

(a) When particle t is a fermion

$\epsilon = P$	T	PT	Number of terms
+	+	+	$\frac{1}{8}(2j+1)(8j^3+12j^2+18j+7)$
+	−	−	$\frac{1}{8}(2j+1)(8j^3+12j^2+2j-1)$
−	+	−	$\frac{1}{8}(2j+1)^4$
−	−	+	$\frac{1}{8}(2j+1)^4$

(b) When particle t is a boson. If, in addition, $t = \bar{t}$, then only terms with $PT = +$ are allowed by the PCT theorem

$\epsilon = P$	T	PT	Number of terms
+	+	+	$2j^4+4j^3+6j^2+4j+1$
+	−	−	$2j^2(j+1)^2$
−	+	−	$j(j+1)(2j^2+2j+1)$
−	−	+	$j(j+1)(2j^2+2j+1)$

PT is, then only terms with $PT = +1$ are nonvanishing in Tables III and V.[32]

Any process not accounted for in Tables III, IV, and V for which a discrete symmetry other than spatial inversion can restrict the number of independent M^c functions is related through crossing to one in the table. For example, in the process $t_1 + \bar{t}_1 \to t_2 + \bar{t}_2$ charge-conjugation invariance can yield a restriction, but this is exactly the same restriction given by PT symmetry for the process $t_1 + t_2 \to t_1 + t_2$.

Although Tables III, IV, and V were derived for $r = 3$ points, they also tell us how many independent $M_\epsilon{}^c$ functions having the given symmetry properties there are at $r = 2$, $n = 3$ points on $\mathcal{R}^{(+)}$. We have seen at the end of Part B of this section that $\mathfrak{L}_+$ covariance allows the number of independent $M_\epsilon{}^c$ functions to be the same at an $r = 2$, $n = 3$ point on $\mathcal{R}^{(+)}$ as at the $r = 3$ points, the latter of which form a dense subdomain of $\mathcal{R}^{(+)}$.

The tables derived from the considerations of this section are important for the application of the criteria to be presented in the next one. In checking any set of covariant polynomials to see whether they give invariant amplitudes free of kinematical singularities, it is useful to know in advance how many of any discrete symmetry signature one should have.

V. INVARIANT AMPLITUDES

A. Extension of the Theorem of Hepp and Williams

We have been considering the M^c functions for two incoming and two outgoing particles with spin as functions of their 4-momenta on the mass shell. The above discussion of kinematical constraints on functions having definite parity signature, together with a result of Hepp and Williams, will enable us to describe the analytic properties in terms of $\mathfrak{L}_+$ invariants. Because of the mass shell conditions and 4-momenta conservation, one can form only two independent $\mathfrak{L}_+$ invariants. These can be taken to be two of the three linearly related Mandelstam invariants:

$$s = (k_1 + k_2)^2, \qquad t = (k_1 - k_3)^2,$$

and

$$u = (k_1 - k_4)^2.$$

Let us first consider a process for which all four particles are spinless—that is, one for which the M^c function is $\mathfrak{L}_+$ invariant,

$$M^c(K) = M^c(\Lambda K), \tag{46}$$

for any point k on the domain of regularity $\mathcal{R}$.

According to a theorem of Hepp[1] and Williams[5], any such function can be expressed as a holomorphic function of the independent $\mathfrak{L}_+$ invariants on the image of the I_+-saturated kernel $\mathfrak{U}_a{}^{(+)}$ of each sheet $\mathfrak{U}_a \subset \mathcal{R}$:

$$M^c(K) = A(s(k), t(k)) \equiv A(s, t). \tag{47}$$

Thus (47) defines an analytic function A over the invariants s and t. The domain of regularity of this function is the image of $\mathcal{R}^{(+)}$.[33]

Actually the restriction to the image of the I_+-saturated kernel is not necessary—the domain of regularity of the function A defined by (47) is the image of the full domain of regularity of M^c. This extension of the result of Hepp and Williams arises from the fact that $\mathcal{R}$ is automatically I_+ saturated; i.e., $\mathcal{R}^{(+)} = \mathcal{R}$.

Lemma 1: Let the domain of regularity of the function F be a domain $\mathcal{R}(F)$ lying over $\mathcal{K}_3$. Suppose

the function F is $\mathfrak{L}_+$ invariant on $\mathfrak{R}(F)$. Then $\mathfrak{R}(F)$ is I_+ saturated, i.e., the $r = n$ limit points of every $r \neq n$ orbit on $\mathfrak{R}(F)$ also lie on $\mathfrak{R}(F)$.

The proof of the above lemma, due to Stapp and this author, is given in Appendix B.

B. Standard Covariants

We would like to generalize the preceding considerations of this section to the cases in which one or more of the four particles has spin. That is, we would like to be able to express the $M_\epsilon{}^c$ functions for any process with two incoming and two outgoing particles in terms of invariant functions that are holomorphic in the $\mathfrak{L}_+$ invariants s and t everywhere on the image of the subset $\mathfrak{U}_a{}^{(+)}$ of each $\mathfrak{L}_+$-invariant sheet $\mathfrak{U}_a \subset \mathfrak{R}$. The spin dependence and, consequently, all the kinematical properties of the $M_\epsilon{}^c$ functions will be accounted for by polynomials in the 4-momenta referred to as "standard covariants." First we will precisely define the latter.

Definition: Consider the M^c functions for a scattering process involving two incoming particles of spins j_1 and j_2 and two outgoing particles of spins j_3 and j_4. A set of spinor functions $Y_{+1}{}^{(g)}(k)$, for $g = 1, \cdots, \bar{N}_+$, and $Y_{-1}{}^{(g)}(k)$, for $g = 1, \cdots, \bar{N}_-$, where $\bar{N}_+ + \bar{N}_- = \prod_{i=1}^{4} (2j_i + 1)$, is said to be a set of *standard covariants* for this process (and the processes related by crossing) if they satisfy the following five properties:

SC1. They are polynomials in the four momenta for the process, subject to the mass shell and 4-momentum conservation constraints. (They are therefore holomorphic everywhere.)

SC2. They are $\mathfrak{L}_+$ covariant; i.e.,

$$Y_{\substack{\epsilon \\ \alpha_b;\,\alpha_a}}{}^{(g)}(\Lambda(A, B)k) = D^{(j_b)}(A)_{\alpha_b}{}^{\alpha_b'} D^{(j_a)}(A)_{\alpha_a}{}^{\alpha_a'} Y_{\substack{\epsilon \\ \alpha_b';\,\alpha_a'}}{}^{(g)}(k). \quad (48)$$

SC3. They have definite signature under the spatial inversion operation:

$$Y_{\substack{\epsilon \\ \alpha_b;\,\alpha_a}}{}^{(g)}(k) = \epsilon D^{(j_b)}\left(\frac{\sigma \cdot k_b}{m_b}\right)_{\alpha_b\alpha_b'} D^{(j_a)}\left(\frac{\sigma \cdot k_a}{m_a}\right)_{\alpha_a\alpha_a'} Y_{\substack{\epsilon \\ \alpha_b';\,\alpha_a'}}{}^{(g)}(\tilde{k}). \quad (49)$$

SC4. The functions of each of the two parity signatures are separately linearly independent at all $n = 3$ points. That is, if we form the functions

$$\Gamma_{\substack{\epsilon \\ \alpha_b;\,\alpha_a}}(k) = \sum_{g=1}^{\bar{N}_\epsilon} \gamma_\epsilon{}^{(g)} Y_{\substack{\epsilon \\ \alpha_b;\,\alpha_a}}{}^{(g)}(k), \quad (50)$$

where the $\gamma_\epsilon{}^{(g)}$ are real or complex numbers, then, at any $n = 3$ point $k = \{k_b; k_a\}$, the only solution to the equations

$$\Gamma_{\substack{\epsilon \\ \alpha_b;\,\alpha_a}}(k) = 0$$

for all values of $\{\alpha_b; \alpha_a\}$ is $\gamma_\epsilon{}^{(g)} = 0$, for all $g = 1, \cdots, \bar{N}_\epsilon$.

SC5. If the number of M^c functions whose values are independent is restricted by any discrete symmetry other than parity, then each of the spinor functions $Y_\epsilon{}^{(g)}(k)$ has definite signature $+1$ or -1 under this operation (the form of the symmetry operation is the same as that for the M^c functions in Table II).

The above properties of the standard covariants lead to the following lemma, the proof of which is given in Appendix C:

Lemma 2: Consider a set of standard covariants for the M^c functions describing a process with two incoming and two outgoing particles. Then:

L1. At any $r = 3$ point, the standard covariants of parity signature $+1$ are linearly independent of those of parity signature -1.

L2. For each value of ϵ, $\bar{N}_\epsilon = N_\epsilon$ as given by (38) if there are some fermions involved in the process, or $\bar{N}_\epsilon = N_\epsilon$ as given by (39) if all the particles are bosons.

L3. In those cases in which property SC5 holds, the number of standard covariants having a given signature under any of the applicable symmetry operations is in agreement with the number in Tables III, IV, and V.

The choice of a set of standard covariants for any process is by no means unique. If we have found a set $Y_\epsilon{}^{(g)}(k)$, for $g = 1, \cdots, N_\epsilon$, satisfying the required properties, and if we can write

$$Y_\epsilon{}^{(g)}(k) = \sum_{g'=1}^{N_\epsilon} f_\epsilon{}^{gg'}(s, t)\, \bar{Y}_\epsilon{}^{(g')}(k), \quad (51)$$

such that the coefficients $f_\epsilon{}^{gg'}(s, t)$ are globally holomorphic functions of the Mandelstam invariants with $\det((f_\epsilon{}^{gg'}(s, t)))$ nowhere zero, then the $Y_\epsilon{}^{(g)}(k)$ also form a set of standard covariants.

C. Invariant Amplitudes for Scattering Functions with Spin

We are now ready to consider the possibility of expanding the $M_\epsilon{}^c$ functions for a given process in terms of standard covariants having the same discrete

symmetry signatures, using the properties of the standard covariants in the definition and lemma of Part B of this section. Our results are expressed by the following theorem.

Theorem: Consider the M^c functions describing a process with two incoming particles and two outgoing particles. On the I_+-saturated kernel $\mathcal{R}^{(+)}$ of the domain of regularity $\mathcal{R}$ of the M^c functions, one may write the following global decompositions:

$$M_\epsilon{}^c = \sum_{g=1}^{N_\epsilon} A_\epsilon{}^{(g)} Y_\epsilon{}^{(g)}, \tag{52}$$

where the $Y_\epsilon{}^{(g)}$ are any set of standard covariants for the process. The "invariant amplitudes" $A_\epsilon{}^{(g)}$, for $g = 1, \cdots, N_\epsilon$, are holomorphic functions of the Mandelstam invariants s and t on the image of the I_+-saturated kernel $\mathcal{U}_a{}^{(+)}$ of each $\mathfrak{L}_+$-invariant sheet $\mathcal{U}_a \subset \mathcal{R}$.

If the $M_\epsilon{}^c$ functions for a particular value of ϵ are identically zero, the above result is trivial with vanishing $A_\epsilon{}^{(g)}$. In our proof of the theorem for nontrivial cases we will first completely ignore those cases in which the number of independent $M_\epsilon{}^c$-function components is less than the number in (38) or (39) at $n = 3$ points on $\mathcal{R}^{(+)}$. The necessary modifications for the exceptional cases will be easy to make.

For our proof we will first try writing on the subset $\mathcal{U}_a{}^{(+)}$ of a particular sheet $\mathcal{U}_a \subset \mathcal{R}$

$$\underset{\alpha_b;\,\alpha_a}{M_\epsilon{}^c(K)} = \sum_{g=1}^{N_\epsilon} \mathcal{A}_\epsilon{}^{(g)}(k) \underset{\alpha_b;\,\alpha_a}{Y_\epsilon{}^{(g)}(k)}, \tag{53}$$

which corresponds to (52), except that we regard the invariant amplitudes $\mathcal{A}_\epsilon{}^{(g)}$ as functions of the 4-momenta for the present. We will show that (53) is invertible; that is, we will solve for the amplitudes $\mathcal{A}_\epsilon{}^{(g)}$ in terms of the $M_\epsilon{}^c$ functions and show that this does not introduce any singularities not present in the $M_\epsilon{}^c$ functions themselves. Finally, the theorem of Hepp and Williams will allow us to express the analytic properties of the invariant amplitudes in terms of $\mathfrak{L}_+$ invariants. Recall that, according to the last paragraph, we are ignoring the exceptional cases for the present.

Consider the scalars

$$\underset{\alpha_b;\,\alpha_a}{\mathcal{Y}_{\epsilon',\epsilon}^{(g',g)}(k)} = \underset{\alpha_b;\,\alpha_a}{Y_{\epsilon'}{}^{(g')}(k) Y_\epsilon{}^{(g)}(k)}. \tag{54}$$

Each of the above invariant functions is holomorphic everywhere on $\mathcal{K}_3$ and the determinant formed from them, $\det(\mathcal{Y}_{\epsilon',\epsilon}^{(g',g)})$, cannot vanish at any $r = 3$ point. The only way the determinant at such a point could vanish would be for some of the standard covariants to be linearly dependent, contradicting statements SC4 and L1. Since the standard covariants are not all linearly independent at an $r = 2$ point, $\det(\mathcal{Y}_{\epsilon,\epsilon}^{(g',g)})$ must vanish at such a point.

The determinant just introduced consists of four blocks, the upper left-hand one having components of the form $\mathcal{Y}_{+1,+1}^{(g',g)}$, the lower right-hand one $\mathcal{Y}_{-1,-1}^{(g',g)}$, the upper right-hand one $\mathcal{Y}_{+1,-1}^{(g',g)}$, and the lower left-hand one $\mathcal{Y}_{-1,+1}^{(g',g)}$. However, it follows from (49) that $\mathcal{Y}_{\epsilon,-\epsilon}^{(g',g)} \equiv 0$, since one has an invariant function of three independent 4-vectors that has negative signature under spatial inversion, and such a function vanishes identically. Consequently, the determinant is factorizable:

$$\det(\mathcal{Y}_{\epsilon',\epsilon}^{(g',g)}) = [\det(\mathcal{Y}_{+1,+1}^{(g',g)})][\det(\mathcal{Y}_{-1,-1}^{(g',g)})], \tag{55}$$

and neither $\det(\mathcal{Y}_{+1,+1}^{(g',g)})$ nor $\det(\mathcal{Y}_{-1,-1}^{(g',g)})$ can vanish at any $r = 3$ point. Since (55) must vanish at $r = 2$ points, we must have

$$\det(\mathcal{Y}_{\epsilon,\epsilon}^{(g',g)}) \propto G^N(k_1, k_2, k_3),$$

where k_1, k_2, and k_3 are any three of the momenta and N is some integer.

The $\mathfrak{L}_+$-invariant functions

$$\underset{\alpha_b;\,\alpha_a}{\mathcal{M}_{\epsilon',\epsilon}^{(g)}(k)} = \underset{\alpha_b;\,\alpha_a}{M_{\epsilon'}{}^c(k) Y_\epsilon{}^{(g)}(k)}, \quad g = 1, \cdots, N_\epsilon, \tag{56}$$

are holomorphic everywhere on the particular domain $\mathcal{U}_a{}^{(+)}$ under consideration with $\mathcal{M}_{\epsilon,-\epsilon}^{(g)} \equiv 0$. From (53), (54), and (56) we obtain the set of equations

$$\sum_{g'} \mathcal{A}_\epsilon{}^{(g')}(k) \mathcal{Y}_{\epsilon,\epsilon}^{(g',g)}(k) = \mathcal{M}_{\epsilon,\epsilon}^{(g)}(k), \quad g = 1, \cdots, N_\epsilon. \tag{57}$$

At least at $r = 3$ points, where $\det(\mathcal{Y}_{\epsilon,\epsilon}^{(g',g)})$ cannot vanish, (57) is soluble for the invariant functions $\mathcal{A}_\epsilon{}^{(g)}$ in terms of the $\mathcal{M}_{\epsilon,\epsilon}^{(g)}$. The solution of (57) has the form

$$\mathcal{A}_\epsilon{}^{(g)}(k) = \frac{\mathcal{C}_\epsilon{}^{(g)}(k)}{G^N(k_1, k_2, k_3)}, \quad g = 1, \cdots, N_\epsilon, \tag{58}$$

where each $\mathcal{C}_\epsilon{}^{(g)}$ is some combination of the $\mathcal{Y}_{\epsilon,\epsilon}^{(g',g)}$ and the $\mathcal{M}_{\epsilon,\epsilon}^{(g)}$.

The numerator $\mathcal{C}_\epsilon{}^{(g)}$ on the right-hand side of (58) is holomorphic on the given domain $\mathcal{U}_a{}^{(+)}$. The only possible singularities of the invariant amplitudes on the left-hand side of (58) on this domain $\mathcal{U}_a{}^{(+)}$ are poles

at points where the rank of the Gram determinant is less than three.

Let us look at (53) once more. We have already seen that the invariant amplitudes on the right-hand side of the equation are holomorphic on the image in $\mathcal{K}_3$ of the $r = 3$ points on the domain $\mathcal{U}_a^{(+)}$. Since these $r = 3$ points from a dense subdomain of $\mathcal{U}_a^{(+)}$, as we approach any $r < 3$ point on $\mathcal{U}_a^{(+)}$ the limit of the right-hand side of (53) must exist and be equal to the value of the left-hand side at the given point. If the limit point is an $r = 2$, $n = 3$ point, the standard covariants are all linearly independent at the point and there can be no cancelling singularities in the invariant amplitudes; i.e., the limit at such a point must exist for each invariant amplitude in (58) separately and not just for the right-hand side of (53) as a whole.

The above considerations show that there exists some neighborhood of each $r = 2$, $n = 3$ point on the domain $\mathcal{U}_a^{(+)}$ under consideration such that $\mathcal{C}_\epsilon^{(g)}$ in (58) has the form $\mathcal{C}_\epsilon^{(g)}(k) = \mathcal{X}_\epsilon^{(g)}(k)G^N$, where $\mathcal{X}_\epsilon^{(g)}$ is holomorphic in the given neighborhood. From (58), $\mathcal{A}_\epsilon^{(g)}(k) = \mathcal{X}_\epsilon^{(g)}(k)$ is holomorphic in that neighborhood.

The $\mathfrak{L}_+$-invariant sheets whose union is $\mathcal{R}$ can be chosen to overlap, and the particular sheet $\mathcal{U}_a$ for which the decomposition (53) was carried out was arbitrary. Thus, the above procedure defines a unique set of functions $\mathcal{A}_\epsilon^{(g)}$, for $g = 1, \cdots, N_\epsilon$. The domain of regularity of each function $\mathcal{A}_\epsilon^{(g)}$ is a domain $\mathcal{R}(\mathcal{A}_\epsilon^{(g)})$ over $\mathcal{K}_3$, and $\mathcal{A}_\epsilon^{(g)}$ is $\mathfrak{L}_+$ invariant on $\mathcal{R}(\mathcal{A}_\epsilon^{(g)})$. We have seen that each domain $\mathcal{R}(\mathcal{A}_\epsilon^{(g)})$ contains all $n = 3$ points on $\mathcal{R}^{(+)}$. Because of Lemma 1 each domain $\mathcal{R}(\mathcal{A}_\epsilon^{(g)})$ also contains all $r = n = 2$ points on $\mathcal{R}$.

Because of the theorem of Hepp and Williams, we may express each invariant amplitude as a function of $\mathfrak{L}_+$ invariants on the image of the I_+-saturated kernel $\mathcal{U}_a^{(+)}$ of each sheet $\mathcal{U}_a \subset \mathcal{R}$:

$$\mathcal{A}_\epsilon^{(g)}(k) = A_\epsilon^{(g)}(s(k), t(k)) \equiv A_\epsilon^{(g)}(s, t), \qquad g = 1, \cdots, N_\epsilon, \quad (59)$$

with the possible exception of $r = 1$ points on the domain $\mathcal{U}_a^{(+)}$. But, at $r = 1$ points, $k_i \cdot k_j = \pm m_i m_j$ for all values of i and j. Consequently, such points are isolated in the space of the $\mathfrak{L}_+$ invariants. It is well known that an analytic function of several complex variables cannot have isolated singularities.[34]

Equation (59) therefore defines functions $A_\epsilon^{(g)}$, each of whose domain of regularity is a domain over the space of the Mandelstam invariants s and t. This domain is the image of all points on $\mathcal{R}^{(+)}$, so the proof of the theorem is completed—for those cases in which the number of independent M_ϵ^c functions is given by (38) or (39).[35]

In the exceptional cases in which properties SC5 and L3 are applicable, the scalars $\mathcal{Y}_{\epsilon,\epsilon}^{(g',g)}$ in (54) formed from standard covariants that have opposite signatures under any of the applicable discrete symmetries vanish identically. Then the determinant in (55) splits up further; that is, det $(\mathcal{Y}_{\epsilon,\epsilon}^{(g',g)})$, for each value of ϵ, can itself be written as a product of smaller determinants. Furthermore, when the M_ϵ^c functions have definite signature under the symmetries in question, the scalars in (56) involving standard covariants with different symmetry properties also vanish identically. The solution of (53) proceeds very much as before, except that now only the standard covariants having the correct symmetry properties need be used in the expansion, and the set of equations to be solved is of smaller order. Since the standard covariants continue to be linearly independent at $r = 2$, $n = 3$ points, which was the crucial factor in our previous proof, we have no singularities in the invariant amplitudes at $r < 3$ points on $\mathcal{R}^{(+)}$.

We have seen in Sec. IVB that $\mathfrak{L}_+$ covariance requires the various M_ϵ^c functions to satisfy certain linear relations at all points on $\mathcal{R}^{(+)}$. Because of the properties of the standard covariants, the decomposition on the right-hand side of (52) automatically satisfies these relations. Therefore, there is no point at which some linear combination of the invariant amplitudes must vanish in order for this decomposition to satisfy the required kinematical constraints—i.e., the invariant amplitudes are free of "kinematical zeros."

VI. SUMMARY AND DISCUSSION

The main results of this paper are as follows.

(a) If the domain of regularity $\mathcal{R}$ of the scattering functions M^c for a $2 \to 2$ process contains a point at which only two of the external 4-momenta are linearly independent then, as is well known, $\mathcal{R}$ contains also points having the same scalar invariants but with three linearly independent momenta. At any of these points $\mathfrak{L}_+$ covariance requires the number of linearly independent components of M^c to be less than the dimensionality of the spin space.

(b) Let $\mathcal{P}$ be the parity operation for the M^c functions and let $M_\epsilon^c = \frac{1}{2}(M^c + \epsilon \mathcal{P} M^c)$. The number of linearly independent components of the functions M_ϵ^c is the same at all points on their domain of regularity at which there are three linearly independent momenta. This result continues to hold if one imposes additional discrete symmetry requirements.

(c) If one expresses the individual functions M_+^c and M_-^c as sums of covariant polynomials times invariant functions, then these invariant functions will be holomorphic in the Mandelstam invariants s and t except at the image of the singularities in 4-momenta space of the corresponding M_ϵ^c functions, provided (i) the total number of basis polynomials equals the dimensionality of the spin space and (ii) the basis polynomials for each of the two parity signatures are separately linearly independent at all points at which the number of linearly independent momenta is three.

Our result allows the awkward comparison to the basis sets of Hepp and Williams to be avoided. It reduces the problem to the essential one of the linear independence properties of the proposed basis set.

In another paper we intend to discuss the problem of checking the linear independence of the polynomials of each parity signature. Several theorems that greatly simplify the practical procedure will be given, together with many practical applications.

ACKNOWLEDGMENTS

The author is indebted to Professor C. M. Sommerfield, Professor D. N. Williams, and Dr. H. P. Stapp for many helpful discussions. The comments of Dr. Stapp have been invaluable in preparing the final version of the paper. The author first became interested in the subject studied here while at the University of Nijmegen, with support from F.O.M. of the Netherlands and the Netherlands-America Foundation, and wishes to thank Professor J. J. de Swart for his encouragement. He is grateful to Professor S. Mandelstam and Professor C. Zemach for the hospitality of the Berkeley Physics Department and to Professor G. F. Chew for the hospitality of the Lawrence Radiation Laboratory.

APPENDIX A: SPINOR CALCULUS AND LORENTZ TRANSFORMATIONS

This appendix serves to clarify the notation of the main part of the paper and other sources should be consulted for more complete details.[4,5,14–17,24,36] We use the superscript T to indicate the transpose of any matrix and † to indicate the Hermitian conjugate. The Pauli matrices are

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{A1}$$

and the convention for any real or complex 4-vector z is that $z^\mu = (z^0, \mathbf{z})$.

Any two matrices A and $B \in SL(2, C)$, the group of two-dimensional unimodular matrices, define a

$$\Lambda(A, B) \in \mathfrak{L}_+,$$

the group of proper complex Lorentz transformations, through the relation

$$\sigma \cdot \Lambda(A, B)z = A\sigma \cdot zB^{\mathrm{T}}. \tag{A2}$$

In particular, if $B = A^*$, $\Lambda(A, A^*) \in L_+^\uparrow$ is a real proper orthochronous Lorentz transformation.

The full complex Lorentz group is $\mathfrak{L} = \mathfrak{L}_+ \cup \mathfrak{L}_-$, where any $\Lambda_I \in \mathfrak{L}_-$ is an improper Lorentz transformation with $\det \Lambda_I = -1$. In contrast to the real Lorentz group, which has four components because the unit matrix $I \in L_+^\uparrow$ and the simultaneous reflection of all four coordinate axes $-I \in L_+^\downarrow$ are not related by any continuous transformation, the complex Lorentz group has only two components, $\mathfrak{L}_+$ and $\mathfrak{L}_-$, because I and $-I$ are connected by a continuous path in $\mathfrak{L}_+$.

By the usual methods, one obtains a $2j + 1$ by $2j + 1$ matrix $D^{(j)}(A)$ corresponding to any $A \in SL(2, C)$. There always exist real parameters $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$ such that one can write $D^{(j)}(A) = D^{(j)}(H)D^{(j)}(V)$ with

$$D^{(j)}(V) = \exp(-i\boldsymbol{\phi} \cdot \mathbf{J}^{(j)}), \tag{A3a}$$

$$D^{(j)}(H) = \exp(\boldsymbol{\lambda} \cdot \mathbf{J}^{(j)}). \tag{A3b}$$

The $J^{(j)}$ are the familiar generators of rotations.

The generalization of (A2) to arbitrary integer or half-odd integer j is then

$$D^{(j)}(\sigma \cdot \Lambda(A, B)z) = D^{(j)}(A)D^{(j)}(\sigma \cdot z)D^{(j)}(B^{\mathrm{T}}), \tag{A4a}$$

or, more specifically,

$$D^{(j)}(\sigma \cdot \Lambda(A, B)z)_{\alpha\dot{\beta}} = D^{(j)}(A)_\alpha{}^{\alpha'} D^{(j)}(B)_{\dot{\beta}}{}^{\dot{\beta}'} D^{(j)}(\sigma \cdot z)_{\alpha'\dot{\beta}'}, \tag{A4b}$$

which serves to clarify the meaning of lower dotted and lower undotted spinor indices. In particular, for real Lorentz transformations a dotted spinor index transforms like the complex conjugate of an undotted one.

The matrix $D^{(j)}(C)$ is defined by

$$D^{(j)}(C^{-1})^{\alpha\beta} = D^{(j)}(C^{-1})^{\dot{\alpha}\dot{\beta}} = (-1)^{j-\alpha}\delta^{\alpha,-\beta}, \tag{A5a}$$

$$D^{(j)}(C)_{\alpha\beta} = D^{(j)}(C)_{\dot{\alpha}\dot{\beta}} = (-1)^{2j} D^{(j)}(C^{-1})^{\alpha\beta}. \tag{A5b}$$

An important property is

$$D^{(j)}(C)D^{(j)}(A)D^{(j)}(C^{-1}) = D^{(j)}(A^{\mathrm{T}^{-1}}). \tag{A6}$$

$D^{(j)}(C^{-1})$, acting from the left on any lower spinor index, turns it into an upper spinor index. Upper

undotted spinor indices are therefore acted on from the left by $D^{(j)}(A^{T^{-1}})$ and upper dotted ones are acted on from the left by $D^{(j)}(B^{T^{-1}})$. Contraction of an upper undotted index with a lower undotted one, or of the corresponding dotted indices with each other, yields a scalar.

In addition to (A6) one has

$$D^{(j)}(C)D^{(j)}(\sigma^T \cdot z)D^{(j)}(C^{-1}) = D^{(j)}(\bar{\sigma} \cdot z), \quad \text{(A7)}$$

where $\bar{\sigma}_\mu = \sigma^\mu = (\sigma_0, -\boldsymbol{\sigma})$.

Let k be any real 4-vector on the mass shell; that is, $k^2 = m^2$, where $m \neq 0$ is the mass of the particle under consideration. Its rest-frame value is $\bar{k} = (m, 0)$ and the "boost" $L(k)$ is the Hermitian matrix in $L_+^\uparrow$ defined by

$$k = L(k)\bar{k}. \quad \text{(A8)}$$

One may write $L(k) = \Lambda(\mathcal{B}(k), \mathcal{B}^*(k))$, where the Hermitian matrix $\mathcal{B}(k) \in SL(2, C)$ is given by

$$\begin{aligned}\mathcal{B}(k) &= (\sigma \cdot k/m)^{\frac{1}{2}} \\ &= [2m(m + k^0)]^{-\frac{1}{2}}[m + k^0 + \boldsymbol{\sigma} \cdot \mathbf{k}]. \quad \text{(A9)}\end{aligned}$$

Corresponding to any $\Lambda \in L_+^\uparrow$ and any real 4-momentum k on the mass shell, one may define the "Wigner rotation"

$$R(k, \Lambda) = L^{-1}(\Lambda k)\Lambda L(k), \quad \text{(A10)}$$

which is well known to those familiar with the unitary representations of the inhomogeneous Lorentz group. One may write $R(k, \Lambda) = \Lambda(A(k), A(k)^*)$, where $A(k) \in SU(2)$ is given by

$$A(k) = \mathcal{B}^{-1}(\Lambda k)A\mathcal{B}(k). \quad \text{(A11)}$$

The following relation is valid for any complex value of k on the mass shell:

$$D^{(j)}(\sigma \cdot k/m)D^{(j)}(\bar{\sigma} \cdot k/m) = I. \quad \text{(A12)}$$

APPENDIX B: PROOF OF LEMMA 1

The details given here are due to Stapp and this author. Some definitions we make use of are the following.

$\mathcal{K}_l$: The points $k = \{k_1, \cdots, k_{l+1}\}$ in complex 4-vector space subject to the conditions $k_i \cdot k_i = (m_i)^2 > 0$; $i = 1, \cdots, l+1$, and $\sum_{i=1}^{l+1} \epsilon_i k_i = 0$, where $\epsilon_i = \pm 1$.

I_+: The mapping that takes sets in k space to their images $S(k)$ in the space of the $\mathcal{L}_+$ invariants; i.e., $S(k) = I_+(k) = (I(k), P(k))$, where $I(k)$ is the set of all inner products formed from the k_i and $P(k)$ is the set of all pseudoscalars formed from them.

$\mathcal{M}_{l+}$: The space of $\mathcal{L}_+$ invariants corresponding to the points of $\mathcal{K}_l$; i.e., $\mathcal{M}_{l+} \equiv I_+(\mathcal{K}_l)$.

To prove Lemma 1, we need the following lemma.

Williams' lemma[37] (open mappings from $\mathcal{K}_3$ to $\mathcal{M}_{3+}$): The I_+ image of a neighborhood of a point $k \in \mathcal{K}_3$ is a neighborhood of $S(k) = I_+(k)$ in $\mathcal{M}_{3+}$; i.e., the map I_+: $\mathcal{K}_3 \to \mathcal{M}_{3+}$ is open.

Proof of Lemma 1: According to remark H1(b) in Sec. IIIB, there is a one-to-one mapping between orbits in $\mathcal{K}_3$ and points $S \in \mathcal{M}_{3+} \equiv I_+(\mathcal{K}_3)$. Thus $\mathcal{F}(S)$ defined by $\mathcal{F}(S(k)) = F(k)$ is uniquely defined for all $r = 3$ points k lying on $\mathcal{R}(F)$. The set of points $r \leq 2$ is a set of codimension 1 in invariant space, since it is defined by $G(k) = G(S(k)) \equiv G(S) = 0$. If any $r = n$ point $\bar{k}$ lies on $\mathcal{R}(F)$, then $\mathcal{F}(S) \equiv \mathcal{F}(S(k))$ is bounded at $\bar{S} = S(\bar{k})$, and in fact in a full neighborhood of $\bar{S}$ by virtue of Williams' lemma on open mappings. But if $\mathcal{F}(S)$ is single valued on $G \neq 0$ and bounded in a neighborhood of $\bar{S}$, then it is holomorphic at $\bar{S}$.[38]

Since $\mathcal{F}(S)$ is holomorphic at $\bar{S}$, every $r = n$ limit point of the orbit $\mathcal{L}_+\bar{k}$ and every other $r \neq n$ orbit having these limit points lies on $\mathcal{R}(F)$. This is because any such point whose image in $\mathcal{K}_3$ is k satisfies $S(k) = \bar{S}$ and at any such point $F(k) = \mathcal{F}(S(k))$ is a holomorphic function of a holomorphic function. Of course, the value of $F(k)$ at all the above points with $S(k) = \bar{S}$ must be defined to be $\mathcal{F}(\bar{S})$. QED

Lemma 1 cannot be extended to $\mathcal{L}_+$-invariant functions over $\mathcal{K}_l$ with $l > 3$. That is, the domain of regularity $\mathcal{R}(H)$ of an $\mathcal{L}_+$-invariant function H over $\mathcal{K}_l$ with $l > 3$ does not necessarily contain the $r = n$ limit points of an $r \neq n$ orbit lying on $\mathcal{R}(H)$. This is because Williams' lemma on open mappings is not valid for neighborhoods of $r \neq n$ points in $\mathcal{K}_l$ with $l > 3$.[37]

APPENDIX C: PROOF OF LEMMA 2

To prove statement L1 of the lemma, we will first assume that at a given $r = 3$ point the standard covariants of signature $+1$ are not linearly independent of those of signature -1, and we will then show that this assumption is not consistent with property SC4 of the standard covariants. That is, we assume that for some $r = 3$ point k there exists a set of nonzero γ_+'s and γ_-'s such that, for the functions defined by (50),

$$\Gamma_+(k) + \Gamma_-(k) = 0, \quad \text{(C1)}$$
$$\alpha_b; \alpha_a \qquad \alpha_b; \alpha_a$$

for all choices of $\{\alpha_b; \alpha_a\}$.

But, since any $r = 3$ point k lies on the same $\mathcal{L}_+$ orbit as the point $\tilde{k}$ obtained by spatial inversion, there exists some $A_k \in SL(2, C)$ such that (48) and (50) give

$$\underset{\alpha_b;\,\alpha_a}{\Gamma_\epsilon(k)} = D^{(j_b)}(A_k)_{\alpha_b}^{\alpha_b'} D^{(j_a)}(A_k)_{\alpha_a}^{\alpha_a'} \underset{\alpha_b';\,\alpha_a'}{\Gamma_\epsilon(\tilde{k})}. \quad \text{(C2)}$$

After substituting (C2) into (C1) and multiplying each spinor index in the result from the left by $D^{(j_i)}(A_k^{-1})$, we get

$$\underset{\alpha_b;\,\alpha_a}{\Gamma_+(\tilde{k})} + \underset{\alpha_b;\,\alpha_a}{\Gamma_-(\tilde{k})} = 0. \quad \text{(C3)}$$

Thus, if (C1) is valid for all $\{\alpha_b;\,\alpha_a\}$ at any $r = 3$ point, (C3) is also true.

Because of (49) and (50),

$$\underset{\alpha_b;\,\alpha_a}{\Gamma_\epsilon(k)} = \epsilon D^{(j_b)}\left(\frac{\sigma \cdot k_b}{m_b}\right)_{\alpha_b\alpha_b'} D^{(j_a)}\left(\frac{\sigma \cdot k_a}{m_a}\right)_{\alpha_a\alpha_a'} \underset{\alpha_b';\,\alpha_a'}{\Gamma(\tilde{k})}. \quad \text{(C4)}$$

If we multiply each spinor index in (C3) from the left by $D^{(j_i)}(\sigma \cdot k_i/m_i)$ and then make use of (C4), we get

$$\underset{\alpha_b;\,\alpha_a}{\Gamma_+(k)} - \underset{\alpha_b;\,\alpha_a}{\Gamma_-(k)} = 0. \quad \text{(C5)}$$

The consistency of (C1) and (C5) requires that

$$\underset{\alpha_b;\,\alpha_a}{\Gamma_\epsilon(k)} = 0, \quad \text{(C6)}$$

for both $\epsilon = +1$ and $\epsilon = -1$, for all choices of $\{\alpha_b;\,\alpha_a\}$. But, as mentioned after (50) in property SC4, the linear independence of the standard covariants of the same parity signature at any $n = 3$ point means that (C6) cannot be true there for nonzero $\gamma_\epsilon^{(g)}$. Consequently, (C1) cannot be true and statement L1 of the lemma is valid.[39]

We now consider statement L2 of the lemma. First note that each component

$$\underset{\alpha_b;\,\alpha_a}{Y_\epsilon^{(g)}(k)},$$

for a fixed value of g, but different values of $\{\alpha_b;\,\alpha_a\}$, is actually a different function. However, as was the case for the M_ϵ^c functions in Sec. IV, (49) means that at most N_ϵ of their values, where N_ϵ is given by (38) or (39), whichever is appropriate, can actually be chosen independently at any $n = 3$ point. Thus, the number of standard covariants $Y_\epsilon^{(g)}(k)$ that are linearly independent for a given ϵ cannot be greater than N_ϵ; i.e., $\bar{N}_\epsilon \leq N_\epsilon$. Since $\bar{N}_+ + \bar{N}_- = \prod_{i=1}^4 (2j_i + 1) = N_+ + N_-$, one must have $\bar{N}_\epsilon = N_\epsilon$.

Finally, consider statement L3 of the lemma. If, for example, $m_1 = m_3$, $j_1 = j_3$, $m_2 = m_4$, and $j_2 = j_4$, we require that standard covariants have definite signature under the simultaneous exchanges

$$(k_1, \alpha_1) \leftrightarrow (k_3, \alpha_3)$$

and

$$(k_2, \alpha_2) \leftrightarrow (k_4, \alpha_4),$$

which is the same as the PT operation for the M^c functions in Table II. Then the same considerations that led to Table III tell us how many linearly independent covariants at most can have a particular signature under this operation, and considerations such as those in the proof of L2 show that this equals the actual number of such standard covariants. Note that, by choosing our covariants to have definite PT signature, we automatically assure that they have definite signature under T. The above is easily extended to the case $m_1 = m_2, j_1 = j_2, m_3 = m_4$, and $j_3 = j_4$, when the covariants are chosen to have definite signature under the simultaneous exchanges $(k_1, \alpha_1) \leftrightarrow (k_2, \alpha_2)$ and $(k_3, \alpha_3) \leftrightarrow (k_4, \alpha_4)$, and to the case of all equal masses, when definite signatures under both types of exchanges mentioned in this paragraph are chosen.

* This work was supported in part by an NSF Contract administered by Yale University, in part by a grant from the Air Force Office of Scientific Research, University of California, Berkeley, and in part by the U.S. Atomic Energy Commission.

† This paper is based on a Ph.D. thesis from Yale University.

‡ Present address: Physics Department, Rice University, Houston, Texas 77001.

[1] K. Hepp, Helv. Phys. Acta **36**, 355 (1963); **37**, 55 (1964).
[2] A. C. Hearn, Nuovo Cimento **21**, 333 (1961).
[3] M. D. Scadron and H. F. Jones, Phys. Rev. **173**, 1734 (1968).
[4] R. F. Guertin, dissertation, Yale University, 1969 (unpublished).
[5] D. N. Williams, Lawrence Radiation Laboratory Report UCRL-11113, September 1963 (unpublished).
[6] D. N. Williams, Department of Physics, University of Michigan, Ann Arbor, Michigan, private communication, has found these standard covariants in the same form as given in Ref. 4. The derivation given by the author of this paper in Ref. 4 is not the original one and does not start from the results of Hepp. Most of the results of Williams and this author have also been found by K. Hardenberg, K. H. Mütter, and W. R. Theis, Nuovo Cimento **62A**, 385 (1969).
[7] G. Fox, Phys. Rev. **157**, 1493 (1967), couples the spin indices of the two particles with Clebsch–Gordan coefficients, whereas in Refs. 4 and 6 they are left uncoupled.
[8] H. Kleinert and K. H. Mütter, Nuovo Cimento **63A**, 657 (1969).
[9] One basis is given by K. Hepp (Appendix B of the second paper in Ref. 1) in which spinor indices belonging to different particles are left uncoupled and in which the discrete symmetries are displayed. He lists all 16 standard covariants for four spin-$\frac{1}{2}$ particles.
[10] H. P. Stapp, Lawrence Radiation Laboratory Report UCRL-10843, August 1963 (unpublished).
[11] P. Minkowski, D. N. Williams, and R. Seiler, in *Lectures in Theoretical Physics*, W. E. Brittin and A. O. Barut, Eds. (University of Colorado Press, Boulder, Col., 1965), Vol. VIIA, p. 173.
[12] D. Hall and A. S. Wightman, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **31**, No. 5 (1957).
[13] D. Zwanziger, Ref. 11, p. 190.
[14] H. P. Stapp, in *High Energy Physics and Elementary Particles* (International Atomic Energy Agency, Vienna, 1965), Vol. I, p. 3.
[15] H. Joos, Fortschr. Physik **10**, 65 (1962).
[16] A. O. Barut, *The Theory of the Scattering Matrix* (Macmillan, New York, 1967).
[17] J. R. Taylor, J. Math. Phys. **7**, 181 (1966).

[18] This is a special case of the cluster decomposition law discussed in Refs. 14, 17, and 19. See also J. R. Taylor, Phys. Rev. **142**, 1236 (1966). A clear and simple discussion is given by R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U.P., Cambridge, 1966).

[19] H. P. Stapp, J. Math. Phys. **9**, 1548 (1968).

[20] Reference 10, pp. 100–04; Ref. 19, p. 1553; H. P. Stapp, Lawrence Radiation Laboratory Report UCRL-18769, February 1969 (unpublished), pp. 3–9; C. Chandler and H. P. Stapp, J. Math. Phys. **10**, 826 (1969), p. 829 and Appendix B.

[21] Rigorous definitions are to be found in Refs. 5 and 11. See also A. S. Wightman, in *Dispersion Relations and Elementary Particles*, C. de Witt and R. Omnes, Eds. (Wiley, New York, 1960), p. 227.

[22] The theorem, as given here, is valid for M^c functions describing the interaction of $l+1$ particles when $\mathfrak{K}_3$ is replaced by $\mathfrak{K}_l$ ($\mathfrak{K}_l$ is defined in Appendix B). The M^c functions are "connected parts" (see Ref. 18) that are free of 4-momentum conservation delta functions.

[23] In its most general form, Stapp's theorem (Refs. 10 and 11) guarantees $\mathfrak{L}_+$ covariance on the entire multisheeted domain of regularity of a set of tensor-valued functions originally defined to be holomorphic and either $L_+^{\uparrow}$ covariant or $\mathfrak{L}_+$ covariant on some arbitrarily small real or complex domain that does not have to be on the mass shell. It is a generalization of earlier work showing that functions originally defined to be holomorphic and $L_+^{\uparrow}$ covariant on the so-called "future tube," such as the Wightman functions in axiomatic field theory, have a single-sheeted $\mathfrak{L}_+$-covariant extension to the "extended tube," which is the union of the $\mathfrak{L}_+$ orbits of all points on the "future tube." See A. S. Wightman, J. Indian Math. Soc. **24**, 625 (1960), and R. Jost, in *Lectures on Field Theory and the Many-Body Problem*, E. R. Caianiello, Ed. (Academic, New York, 1961), p. 127. See also Ref. 24.

[24] R. F. Streater and A. S. Wightman, *PCT, Spin, and Statistics, and All That* (Benjamin, New York, 1964).

[25] The usual definition of the I_+-saturated kernel $\mathfrak{U}_a^{(+)}$ of an $\mathfrak{L}_+$-invariant sheet $\mathfrak{U}_a$ on the domain of regularity $\mathfrak{R}$ of the M^c functions for a given process is: $\mathfrak{U}_a^{(+)}$ is the largest subset of $\mathfrak{U}_a$ such that if $\mathfrak{U}_a^{(+)}$ contains a point, then $\mathfrak{U}_a^{(+)}$ contains every point whose image in $\mathfrak{K}_3$ has the same $\mathfrak{L}_+$ invariants (see Refs. 1, 5, and 11). However, for a function defined over $\mathfrak{K}_3$ our definition of the I_+-saturated kernel of an $\mathfrak{L}_+$-invariant sheet $\mathfrak{U}_a$ is equivalent to the usual one. This is a consequence of Lemma 4 of Ref. 5.

[26] A general discussion of linear relationships for physical values of the 4-momenta using helicity amplitudes is given by J. Daboul, "Linear Symmetries of Scattering Amplitudes," Temple University, preprint, November 1969.

[27] We will not discuss the restrictions that Lorentz covariance places on the derivatives of the M^c functions at $r=2$ points. The considerations here will prove to be sufficient for specifying the properties of the "standard covariants" introduced in Sec. V when the decomposition of the M^c functions into invariant amplitudes is considered.

[28] R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence, R.I., 1965).

[29] H. P. Stapp, Phys. Rev. **128**, 1963 (1962). Analyticity and the superposition principle require $\eta_\rho = \pm 1$. The cluster decomposition law then requires $\eta_\rho = +1$ for an elastic process. The "particle intrinsic parities" can be chosen to be real if the only conservation laws are additive ones. This means that $\eta_\rho = +1$ for a process of the form $t_1 + t_1 \to t_3 + t_3$.

[30] For a discussion of the crossing properties, see Refs. 17 and 19.

[31] Note that, because of the kinematical branch points in the "boost" matrices, the mapping from the M^c functions to the S^c functions at any point on the domain of regularity of the former is not single valued. This fact is not of any importance for the application we have in mind here.

[32] We call attention to the fact that in Table III the total number of terms is $(2j_1+1)^2(2j_2+1)^2$, in agreement with (16), but the total number is less in Tables IV and V, because S-matrix theory and field theory forbid the occurrence of terms with the wrong signature under exchange symmetry and under *PCT*.

[33] In its more general form, the theorem of Hepp and Williams says that one can express any multisheeted invariant function of any number of 4-vectors (regardless of whether or not mass-shell constraints are present) as a holomorphic function of the independent $\mathfrak{L}_+$ invariants on the image of the I_+-saturated kernel of its domain of regularity. However, with more than four functionally independent vectors (e.g., a scattering amplitude with more than five particles), it is not possible to use the same set of $\mathfrak{L}_+$ invariants globally; in other words, one must use "local" $\mathfrak{L}_+$ invariant coordinates. The result of Hepp and Williams is a generalization to arbitrary domains of a theorem of Hall and Wightman (Ref. 12) that enables one to express an $\mathfrak{L}$-invariant single-sheeted function regular everywhere on the "extended tube" as a function of "local" $\mathfrak{L}$-invariant coordinates. (See also Ref. 23.)

[34] R. C. Gunning and H. Rossi, *Analytic Functions of Several Complex Variables* (Prentice-Hall, Englewood Cliffs, N.J., 1965), p. 21.

[35] Our theorem generalizes the approach of Williams (Ref. 5). Williams' arguments were for his particular basis and depended on a certain physically irrelevant decomposition of the scattering function into "tensor" and "pseudotensor" parts.

[36] A. S. Wightman, in *Dispersion Relations and Elementary Particles*, C. de Witt and R. Omnes, Eds. (Wiley, New York, 1960), p. 159.

[37] Lemma 5 and Appendix IV of Ref. 5.

[38] Reference 34, p. 19.

[39] It is worth noting that if k is an $r=2$, $n=3$ point, then $\tilde{k}$ lies on a different $\mathfrak{L}_+$ orbit, according to H2(b), and (C2) is not valid for any choice of $A_k \in SL(2,C)$. In this case it is possible to satisfy (C1) with

$$\Gamma_{+\,\alpha_b;\alpha_a}(k) \neq 0 \text{ and } \Gamma_{-\,\alpha_b;\alpha_a}(k) \neq 0,$$

in contrast to the result (C6) for an $r=3$ point. The standard covariants of signature $+1$ are not linearly independent of those of signature -1 at any $r=2$, $n=3$ point; at such a point the total number of linearly independent standard covariants in the two sets is the same as the number given in Table I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Current-Stress Tensor Commutators

STANLEY DESER*
Brandeis University, Waltham, Massachusetts 02154
AND
L. KENT MORRISON†
Kansas State University, Manhattan, Kansas 66502

(Received 20 August 1970)

The model independent components of current-stress tensor equal time commutators are derived for arbitrary spin 0 and 1 systems.

I. INTRODUCTION

Equal time commutators (ETC) of currents with stress tensor components have been discussed in connection with the structure of current–current ETC.[1,2,3] As expected physically from their significance as local generators, some of their components will be model independent. In this article we shall derive these for systems of spin 0 and 1 by considering the response to variations of external vector and gravitational fields. We shall obtain the following model independent ETC between the conserved electromagnetic (or Yang–Mills) current and the symmetric stress tensor[4]:

$$i[j^0(x,t), T^{00}(x',t)] = \partial_i[j^i(x,t)\delta^3(x-x')], \quad (1a)$$

$$i[j^0(x,t), T^{0i}(x',t)] = \partial_i[j^0(x,t)\delta^3(x-x')], \quad (1b)$$

$$i[j^i(x,t), T^{00}(x',t)] = -j^0(x,t)\partial_i'\delta^3(x-x') - \partial_0 j^i(x,t)\delta^3(x-x'). \quad (1c)$$

(These equations obey the consistency requirement $[\partial_\mu j^\mu, T^{00'}] = 0$.) We start from the following general form of the action in the presence of an external gravitational field, valid[5] for systems of spin 0 and 1:

$$I = \int \left(\sum_a \pi_a\dot{\phi}_a + N\theta_0^0(\phi, \pi, g_{ij}) + {}^3g^{ij}N_i\theta_j^0(\phi, \pi, g_{ij})\right)d^4x. \quad (2)$$

Here (π_a, ϕ_a) represent the canonical variables, θ_μ^0 are functions of π_a, ϕ_a, and g_{ij} only while $N \equiv (-g^{00})^{-\frac{1}{2}}$, $N_i \equiv g_{0i}$, and ${}^3g^{ij}$ is the matrix inverse of g_{ij}. This form is known[6] to yield the model-independent ETC for $[T^{00}, T^{00'}]$ (Schwinger–Dirac condition) as well as those of $[T^{0\mu}, T^{0\nu'}]$. We neglect operator ordering and other quantum problems such as higher Schwinger terms,[7] as the model independent parts we seek are already present at the classical level. While fermion systems do not take the form in Eq. (2), it may be shown[7,8] that they also satisfy Eq. (1).

II. CURRENT-STRESS TENSOR COMMUTATORS

The usual first-order action for low spin fields in flat space has the well-known form

$$I = \int [\pi_a\dot{\phi}_a - \mathcal{H}(\pi_a, \phi_a)]d^4x,$$

in terms of the unconstrained canonical degrees of freedom (π_a, ϕ_a), with $\mathcal{H}$ independent of time derivatives for general interactions (including derivative coupling).

With minimal coupling to gravitation, the generally covariant form of Eq. (2) emerges with an appropriate choice of transformation behavior of (π_a, ϕ_a). The crucial point is that this form fixes the explicit dependence on $g_{0\mu}$ to be through the linear (N, N_i) factors. This is a general feature of lower spin and no longer holds for higher spin fields, where elimination of constrained variables introduces further $g_{\mu\nu}$ dependence, which in fact produces grave consistency problems there.[9] Likewise we note that the time derivative enters only in the kinetic $\pi\dot{\phi}$ term.

To derive the desired commutators, we use the definitions of the current and stress tensor as the response of a system to variations of external vector and metric sources:

$$\delta_A I = \int j^\mu \delta A_\mu d^4x, \quad (3a)$$

$$\delta_g I = \tfrac{1}{2}\int \mathcal{T}^{\mu\nu}\delta g_{\mu\nu}d^4x, \quad (3b)$$

or, in terms of matrix elements,

$$\frac{\delta\langle a \mid b\rangle}{\delta A_\mu(x)} = i\,\langle a|\, j^\mu(x)\,|b\rangle, \quad (4a)$$

$$\frac{\delta\langle a \mid b\rangle}{\delta g_{\mu\nu}(x)} = 2i\,\langle a|\, \mathcal{T}^{\mu\nu}(x)\,|b\rangle. \quad (4b)$$

Second mixed variations, which are independent of

order, then read

$$\delta_A\delta_g\langle a \mid b\rangle = \delta_g\delta_A\langle a \mid b\rangle$$
$$= \frac{i}{2}\iint d^4x d^4x' \delta A_\sigma(x)\delta g_{\mu\nu}(x')$$
$$\times \langle a\mid \left(\frac{\delta \mathfrak{T}^{\mu\nu}(x')}{\delta A_\sigma(x)} + i[\mathfrak{T}^{\mu\nu}(x')j^\sigma(x)]_+\right)\mid b\rangle$$
$$= i\iint d^4x d^4x' \delta A_\sigma(x)\delta g_{\mu\nu}(x')$$
$$\times \langle a\mid \left(\frac{\delta j^\sigma(x)}{\delta g_{\mu\nu}(x')} + \frac{i}{2}[j^\sigma(x)\mathfrak{T}^{\mu\nu}(x')]_+\right)\mid b\rangle, \tag{5}$$

so that the reciprocity relation

$$2\frac{\delta j^\sigma(x)}{\delta g_{\mu\nu}(x')} = \frac{\delta \mathfrak{T}^{\mu\nu}(x')}{\delta A_\sigma(x)} \tag{6}$$

holds.

We consider now the general forms of the desired ETC implied by gauge invariance of the theory and conservation of the corresponding current.[10] Both still hold in the presence of $g_{\mu\nu}(x)$, since the quantity $\partial_\mu j^\mu(x)$ is a general coordinate scalar when j^μ is a vector density. We take in Eq. (5) the special case $\delta A_\sigma = \partial_\sigma\lambda(x)$, for which the left-hand side then vanishes, and integrate by parts on ∂_σ to obtain the flat space relation[11]

$$i[j^0(x), T^{\mu\nu}(x')]\delta(x^0 - x^{0'}) = -2\partial_\sigma\left[\frac{\delta j^\sigma(x)}{\delta g_{\mu\nu}(x')}\right]_\eta \tag{7}$$

since the gauge equation holds for arbitrary $\delta g_{\mu\nu}$.

We now apply the general result, Eq. (7), to the action given by Eq. (2) in the presence of an external vector field A_μ, where we shall set $A_\mu = 0$ and $g_{\mu\nu} = \eta_{\mu\nu}$ after varying. For minimal coupling it is clear that, since ∂_0 only appears in the metric-independent kinetic term, $\pi\partial_0\phi$, $j^0 \sim \pi\phi$ is metric independent. Thus $\delta j^0/\delta g_{\mu\nu} \equiv 0$, in agreement with time locality requirements, for otherwise the terms $\partial_0(\delta j^0/\delta g'_{\mu\nu})$ would involve $\partial_0\delta(x^0 - x^{0'})$, inconsistent with the time locality of the left-hand side of Eq. (7). Thus, we have the stronger form

$$i[j^0(x), T^{\mu\nu}(x')]\delta(x^0 - x^{0'}) = -2\partial_i\left[\frac{\delta j^i(x)}{\delta g_{\mu\nu}(x')}\right]_\eta. \tag{8}$$

The dependence of j^i on $g_{0\mu}$ which we now require may also be given quite explicitly. Differentiating Eq. (2) with respect to A_i gives[12]

$$j^i = \int\left(N'\frac{\delta\theta_0^{0'}}{\delta A_i} + {}^3g^{lm}N'_l\frac{\delta\theta_m^{0'}}{\delta A_i}\right)d^4x'.$$

For the $\delta/\delta g_{00}$ derivative, we find (always at flat space) that

$$\left.\frac{\delta j^i}{\delta g'_{00}}\right|_\eta = -\tfrac{1}{2}\left(\frac{\delta\theta_0^{0'}}{\delta A_i}\right)_\eta = -\tfrac{1}{2}\delta^4(x - x')j^i(x)\big|_\eta, \tag{9}$$

where the last equality follows from the previous equation at $g = \eta$. Likewise we may obtain

$$\left.\frac{\delta j^i}{\delta g_{0k}}\right|_\eta = \tfrac{1}{2}\delta^{kl}\left(\frac{\delta\theta_l^{0'}}{\delta A_i}\right)_\eta.$$

To evaluate $\delta\theta_l^{0'}/\delta A_i$, we recall that, at flat space where $\theta_i^0 = T^0{}_i$, the momentum or spatial translation density always has the simple form (for $A_\mu = 0$)

$$T^0{}_l = -\sum_a \pi_a\partial_l\phi_a + \partial_k D^k{}_l, \tag{10}$$

where $D^k{}_l$ is a *neutral* function of (π, ϕ) and does not involve derivatives. Now, in the presence of A_μ, the divergence term does not contribute to $\delta T^{0i}/\delta A_j$ since it is neutral {$\partial_k D^{ki}$ does *not* become $(\partial_k - ieA_k)D^{ki}$ and there is no possible ambiguity involving

$$[(\partial_\mu - ieA_\mu), (\partial_\nu - ieA_\nu)] \sim F_{\mu\nu}$$

since D has *no* derivatives}. The only contribution comes from the $\pi\partial\phi$ part, which yields

$$\frac{\delta T^{0'}{}_l}{\delta A_i} = 2\left(\frac{\delta j^i}{\delta g_{0l}}\right)_\eta$$
$$= ie\sum_a \pi_a\phi_a\delta_{il}\delta^4(x - x')$$
$$= -\delta_{il}\delta^4(x - x')j^0(x). \tag{11}$$

Insertion of Eqs. (9) and (11) into (8) now yields the first two of our results, (1a) and (1b).

The ETC's $[j^0, T^{ij}]$ are not model independent; thus, $\delta j_i/\delta g_{kl}$ involves the explicit g_{kl} dependence of θ_0^0, namely

$$\frac{\delta}{\delta g_{kl}}\frac{\delta\theta_0^0}{\delta A_i} \sim \frac{\delta T^{kl}}{\delta A_i},$$

which depends on any gradient coupling present.

As for the remaining ETC $[j^i, T^{\mu\nu}]$, only $[j^i, T^{00}]$ is model independent. Using the Jacobi identity on $[M^{0i}, T^{00}, j^0]$ in terms of the symmetric $T^{\mu\nu}$ yields $[j^i, T^{00}]$ in terms of the $[j^0, T^{0\mu}]$ ETC, and the result is Eq. (1c). Note that, in this process, the symmetric $T^{\mu\nu}$ *must* be used (or else M^{0i} is not the Lorentz generator). It is thus essential to have the $[j^0, T^{0\mu}]$ ETC for the symmetric rather than the canonical stress tensor in order to derive[13] Eq. (1c).

Finally, we remark that our results follow equally well for, say, Yang–Mills currents.

* Supported in part by U.S.A.F. OAR.

† Address for 1970–71: Indiana Univ.-Purdue Univ., Indianapolis, Indiana 46205.

[1] D. J. Gross and R. Jackiw, Phys. Rev. **163**, 1688 (1967).
[2] J. C. Katzin and W. B. Rolnick, Phys. Rev. **182**, 1403 (1969).
[3] R. Jackiw, Phys. Rev. **175**, 2058 (1968).
[4] Equation (1a) was first given by J. Schwinger, Nuovo Cimento **30**, 278 (1963) and independently by Gross and Jackiw.[1]
[5] D. Boulware and S. Deser, J. Math. Phys. **8**, 1468 (1967); Phys. Rev. **151**, 1278 (1966). In dealing with charged systems, the appropriate charge matrices are understood.

[6] S. Deser and L. K. Morrison, J. Math. Phys. **11**, 596 (1970).

[7] For these, see D. N. Levin, thesis, Harvard University, 1970, and Ref. 8.

[8] For explicit calculations, see, for example, H. Genz and J. Katz, UCRL, Report No. UCRL-19438, 1969.

[9] See J. Schwinger, Phys. Rev. **130**, 406, 800 (1963) for spin $\frac{3}{2}$. Spin 2 is treated by C. Aragone, Nuovo Cimento **64A**, 841 (1969), and C. Aragone and S. Deser, Nuovo Cimento (in press).

[10] The converse route is also possible: Although the stress tensor is no longer conserved in an external A_μ field (which also acts as a preferred vector destroying general coordinate invariance), its nonconservation is model independent. For an arbitrary system, one has (S. Deser, Nuovo Cimento, to be published),

$$\mathcal{C}^{\mu\nu}{}_{;\nu} = -F^{\mu}{}_{\nu} j^{\nu}$$

or, at $g = \eta$, $\Delta^\mu \equiv T^{\mu\nu}{}_{,\nu} = F_\nu{}^\mu j^\nu$. This gives $\delta\Delta^\mu/\delta A_\lambda$ explicitly at $A^\mu = 0$. One may then make use of the easily derived relation

$$-i[T^{0\nu}(x), j^\lambda(x')]\delta(x^0 - x^{0'}) = \partial_\mu\left(\frac{\delta T^{\mu\nu}}{\delta A'_\lambda}\right) - \frac{\delta\Delta^\nu}{\delta A'_\lambda}$$

together with reciprocity to derive the ETC of Eqs. (1). In this treatment, $[j^i, T^{00}]$ follows directly *without* use of the Jacobi identity, in contrast to the derivation in text. Using $\delta T^{\mu\nu}/\delta A'_0 = 2\delta j^{0'}/\delta g_{\mu\nu} = 0$ and the above form for Δ^μ, immediately yields Eqs. (1a) and (1b). Likewise, from $\delta T^{0\nu}/\delta A'_k = 2\delta j^{k'}/\delta g_{0\nu} = -\delta^4(x - x')[\delta^{\nu 0} j^k + \delta^{\nu k} j^0]$ and Δ^0, we obtain Eq. (1c).

[11] Equation (7) follows equivalently from Schwinger's relation.

$$-i\int d^4x'[A(x), \delta\mathfrak{L}(x')]\delta(x^0 - x^{0'}) = \partial_0\delta' A(x) - \delta' B(x),$$

valid whenever the relation $\partial_0 A = B$ is maintained for arbitrary variations of the external field, indicated by δ'. In particular, take $A = j^0$, $B = -\partial_i j^i$, and $\delta' = \delta_\sigma$, and interchange ∂_i and δ', which, unlike a $\partial_0 - \delta'$ interchange, *is* valid.

[12] Had we used, for spin 0, the "new improved" stress tensor, $\theta_{\mu\nu}$ [C. G. Callan, Jr., S. Coleman, and R. Jackiw, Ann. Phys. (N.Y.) **59**, 42 (1970)], corresponding to a nonminimal gravitational coupling $R\phi^*\phi$, our results would have been unaltered (in agreement with direct computation) because the additional term remains A_μ independent, just like a mass term (despite the presence of second derivatives in R).

[13] Jackiw[3] showed that Eqs. (1a) and (1b) hold for the canonical stress tensor. From our point of view, they may be derived by the type of argument in the discussion of Eq. (10) in text.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Random Walks with Nonnearest Neighbor Transitions. I. Analytic 1-D Theory for Next-Nearest Neighbor and Exponentially Distributed Steps*

KATJA LAKATOS-LINDENBERG AND KURT E. SHULER
Department of Chemistry, University of California, San Diego, La Jolla, California 92037

(Received 17 July 1970)

We present here exact analytic results for random walks on one-dimensional lattices with nonnearest neighbor transitions. After deriving the generating function for such lattices with and without boundaries, we have calculated a number of moment properties (mean first passage times to absorption, mean recurrence times and their dispersion, mean excursion from the origin, etc.) for random walks with next-nearest neighbor transitions and for random walks with exponentially distributed step length. In the latter case, variation of one of the parameters permits us to cover the whole range of step lengths from nearest neighbor transitions to steps of any finite length l. Since we have obtained explicit expressions for the generating function for these walks, any additional desired moment properties can readily be calculated. Among the interesting results of this study are: (1) The moment results for random walks with next-nearest neighbor transitions differ from the analogous nearest neighbor results at most by a factor of $O(1)$; (2) the one-dimensional moment results for walks with arbitrary step length differ from the analogous one-dimensional results for walks with nearest neighbor transitions by several orders of magnitude; (3) the mean time to absorption for a random walker with equal probabilities for steps of arbitrary length in *one dimension* agrees to within a factor of $O(1)$ with the mean time for absorption for a random walker with nearest neighbor steps in *three dimensions*; (4) the mean time to absorption for a random walker with equal probabilities for steps of arbitrary lengths is independent of the dimensionality of the lattice.

I. INTRODUCTION

Previous work on random walks on lattices has been limited to *nearest neighbor* transitions. The modern techniques in this field have been developed by Montroll and his co-workers[1–5] and have been applied by them to a number of problems in 1-D, 2-D, and 3-D lattices.

There is, however, a large class of physical problems which cannot be correctly described by random walks with transitions only between nearest neighbor states. Some examples are: rotational relaxation with transitions between rotational levels with $|\Delta J| > 1$, vibrational relaxation with transitions between vibrational levels with $|\Delta v| > 1$, stochastic theory of chromatography with a variable absorption length along the column, surface diffusion, and exciton diffusion. In general, stochastic processes where the transitions in the appropriate space (coordinate, energy, momentum, etc.) are due to strong interactions should be described by random walks involving transitions of longer "range" than the usual nearest neighbor case. It is clear, therefore, that it would be very useful to

[6] S. Deser and L. K. Morrison, J. Math. Phys. **11**, 596 (1970).

[7] For these, see D. N. Levin, thesis, Harvard University, 1970, and Ref. 8.

[8] For explicit calculations, see, for example, H. Genz and J. Katz, UCRL, Report No. UCRL-19438, 1969.

[9] See J. Schwinger, Phys. Rev. **130**, 406, 800 (1963) for spin $\frac{3}{2}$. Spin 2 is treated by C. Aragone, Nuovo Cimento **64A**, 841 (1969), and C. Aragone and S. Deser, Nuovo Cimento (in press).

[10] The converse route is also possible: Although the stress tensor is no longer conserved in an external A_μ field (which also acts as a preferred vector destroying general coordinate invariance), its nonconservation is model independent. For an arbitrary system, one has (S. Deser, Nuovo Cimento, to be published),

$$\mathcal{C}^{\mu\nu}{}_{;\nu} = -F^{\mu}{}_{\nu} j^{\nu}$$

or, at $g = \eta$, $\Delta^\mu \equiv T^{\mu\nu}{}_{,\nu} = F_\nu{}^\mu j^\nu$. This gives $\delta\Delta^\mu/\delta A_\lambda$ explicitly at $A^\mu = 0$. One may then make use of the easily derived relation

$$-i[T^{0\nu}(x), j^\lambda(x')]\delta(x^0 - x^{0'}) = \partial_\mu\left(\frac{\delta T^{\mu\nu}}{\delta A'_\lambda}\right) - \frac{\delta\Delta^\nu}{\delta A'_\lambda}$$

together with reciprocity to derive the ETC of Eqs. (1). In this treatment, $[j^i, T^{00}]$ follows directly *without* use of the Jacobi identity, in contrast to the derivation in text. Using $\delta T^{\mu\nu}/\delta A'_0 = 2\delta j^{0'}/\delta g_{\mu\nu} = 0$ and the above form for Δ^μ, immediately yields Eqs. (1a) and (1b). Likewise, from $\delta T^{0\nu}/\delta A'_k = 2\delta j^{k'}/\delta g_{0\nu} = -\delta^4(x - x')[\delta^{\nu 0} j^k + \delta^{\nu k} j^0]$ and Δ^0, we obtain Eq. (1c).

[11] Equation (7) follows equivalently from Schwinger's relation.

$$-i\int d^4x'[A(x), \delta\mathfrak{L}(x')]\delta(x^0 - x^{0'}) = \partial_0\delta' A(x) - \delta' B(x),$$

valid whenever the relation $\partial_0 A = B$ is maintained for arbitrary variations of the external field, indicated by δ'. In particular, take $A = j^0$, $B = -\partial_i j^i$, and $\delta' = \delta_\sigma$, and interchange ∂_i and δ', which, unlike a $\partial_0 - \delta'$ interchange, *is* valid.

[12] Had we used, for spin 0, the "new improved" stress tensor, $\theta_{\mu\nu}$ [C. G. Callan, Jr., S. Coleman, and R. Jackiw, Ann. Phys. (N.Y.) **59**, 42 (1970)], corresponding to a nonminimal gravitational coupling $R\phi^*\phi$, our results would have been unaltered (in agreement with direct computation) because the additional term remains A_μ independent, just like a mass term (despite the presence of second derivatives in R).

[13] Jackiw[3] showed that Eqs. (1a) and (1b) hold for the canonical stress tensor. From our point of view, they may be derived by the type of argument in the discussion of Eq. (10) in text.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Random Walks with Nonnearest Neighbor Transitions. I. Analytic 1-D Theory for Next-Nearest Neighbor and Exponentially Distributed Steps*

KATJA LAKATOS-LINDENBERG AND KURT E. SHULER
Department of Chemistry, University of California, San Diego, La Jolla, California 92037

(Received 17 July 1970)

We present here exact analytic results for random walks on one-dimensional lattices with nonnearest neighbor transitions. After deriving the generating function for such lattices with and without boundaries, we have calculated a number of moment properties (mean first passage times to absorption, mean recurrence times and their dispersion, mean excursion from the origin, etc.) for random walks with next-nearest neighbor transitions and for random walks with exponentially distributed step length. In the latter case, variation of one of the parameters permits us to cover the whole range of step lengths from nearest neighbor transitions to steps of any finite length l. Since we have obtained explicit expressions for the generating function for these walks, any additional desired moment properties can readily be calculated. Among the interesting results of this study are: (1) The moment results for random walks with next-nearest neighbor transitions differ from the analogous nearest neighbor results at most by a factor of $O(1)$; (2) the one-dimensional moment results for walks with arbitrary step length differ from the analogous one-dimensional results for walks with nearest neighbor transitions by several orders of magnitude; (3) the mean time to absorption for a random walker with equal probabilities for steps of arbitrary length in *one dimension* agrees to within a factor of $O(1)$ with the mean time for absorption for a random walker with nearest neighbor steps in *three dimensions*; (4) the mean time to absorption for a random walker with equal probabilities for steps of arbitrary lengths is independent of the dimensionality of the lattice.

I. INTRODUCTION

Previous work on random walks on lattices has been limited to *nearest neighbor* transitions. The modern techniques in this field have been developed by Montroll and his co-workers[1–5] and have been applied by them to a number of problems in 1-D, 2-D, and 3-D lattices.

There is, however, a large class of physical problems which cannot be correctly described by random walks with transitions only between nearest neighbor states. Some examples are: rotational relaxation with transitions between rotational levels with $|\Delta J| > 1$, vibrational relaxation with transitions between vibrational levels with $|\Delta v| > 1$, stochastic theory of chromatography with a variable absorption length along the column, surface diffusion, and exciton diffusion. In general, stochastic processes where the transitions in the appropriate space (coordinate, energy, momentum, etc.) are due to strong interactions should be described by random walks involving transitions of longer "range" than the usual nearest neighbor case. It is clear, therefore, that it would be very useful to

extend the present theory of random walks to step-lengths of arbitrary size.

We present here exact analytic results for random walks with next-nearest neighbor transitions and with exponentially distributed step lengths for one-dimensional lattices with and without boundaries, using the generating function method developed by Montroll.[1] The emphasis in this paper is on development of techniques and methodology rather than on applications. We believe that these results can readily be extended to two and three dimensions and we plan to present such results in a subsequent paper.

II. SYMMETRIC NEXT-NEAREST NEIGHBOR RANDOM WALK

A. Generating Function on Perfect Infinite Lattice and on Perfect N-ring

In this and subsequent sections we shall use the following notation:

$p(l_i, l_j)$ is the probability that the random walker makes a step from site l_i to site l_j.

$P_n(l - l_0)$ is the probability that the random walker starting at site $l = l_0$ is at site l after the nth step on an infinite perfect lattice, i.e., a lattice without traps and boundaries.

The probabilities $p(l_i, l_j)$ and $P_n(l - l_0)$ have the properties

$$\sum_{l_i} p(l_i, l_j) = \sum_{l_j} p(l_i, l_j) = 1, \tag{1}$$

$$\sum_{l} P_n(l - l_0) = 1. \tag{2}$$

The generating function $G(z, l - l_0)$ is defined as

$$G(z, l - l_0) \equiv \sum_{n=0}^{\infty} z^n P_n(l - l_0). \tag{3}$$

We consider only random walks for which, on a perfect infinite lattice, $p(l_i, l_j)$ depends only on the difference $(l_i - l_j)$. For a walker starting at the origin $l_0 \equiv 0$, the probability $P_n(l)$ is the coefficient of $e^{il\phi}$ in the expression[1]

$$\lambda^n(\phi) \equiv \left(\sum_{k=-\infty}^{\infty} p(k) e^{ik\phi}\right)^n, \quad k = l_i - l_j, \tag{4}$$

i.e.,

$$P_n(l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \lambda^n(\phi) e^{-il\phi}\, d\phi. \tag{5}$$

The generating function $G(z, l)$ is then

$$G(z, l) = \frac{1}{2\pi} \sum_{n=0}^{\infty} z^n \int_{-\pi}^{\pi} \lambda^n(\phi) e^{-il\phi}\, d\phi = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-il\phi}}{1 - z\lambda(\phi)}\, d\phi. \tag{6}$$

For the next-nearest neighbor symmetric random walk we write

$$\begin{aligned} p(l_i - l_j) &= \tfrac{1}{2}(1 - \alpha), && \text{for } |l_i - l_j| = 1, \\ &= \tfrac{1}{2}\alpha, && \text{for } |l_i - l_j| = 2, \\ &= 0, && \text{otherwise,} \end{aligned} \tag{7}$$

where $0 \leq \alpha \leq 1$. Using Eqs. (3)–(7) leads to the generating function

$$G(z, l, \alpha) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-il\phi}}{1 - z[(1 - \alpha)\cos\phi + \alpha \cos 2\phi]}\, d\phi. \tag{8}$$

The evaluation of this integral, which is detailed in Appendix A, yields

$$G(z, l, \alpha) = \frac{2}{z\alpha(s_4 - s_3)} \left(\frac{s_3^{|l|+1}}{(s_3 - s_1)(s_3 - s_2)} - \frac{s_4^{|l|+1}}{(s_4 - s_1)(s_4 - s_2)} \right), \tag{9}$$

where

$$\begin{aligned} s_1 &= ((\alpha - 1)/4\alpha)[1 + (1 + y_-)^{\frac{1}{2}}][1 + (1 + y_+)^{\frac{1}{2}}], \\ s_2 &= ((\alpha - 1)/4\alpha)[1 + (1 + y_-)^{\frac{1}{2}}][1 - (1 + y_+)^{\frac{1}{2}}], \\ s_3 &= ((\alpha - 1)/4\alpha)[1 - (1 + y_-)^{\frac{1}{2}}][1 + (1 + y_+)^{\frac{1}{2}}], \\ s_4 &= ((\alpha - 1)/4\alpha)[1 - (1 + y_-)^{\frac{1}{2}}][1 - (1 + y_+)^{\frac{1}{2}}], \end{aligned} \tag{10}$$

and where

$$y_{\pm} = \frac{4\alpha}{z(1 - \alpha)^2} \{(1 - z\alpha) \pm [(1 - z)(1 + z - 2\alpha z)]^{\frac{1}{2}}\}. \tag{11}$$

The generating function $G(z, l, \alpha)$ of Eq. (9) is particularly useful for random walks on an infinite lattice. For random walks involving absorbing boundaries it is convenient to work on a closed ring of N lattice sites. For such an "N-ring," the generating function $G_N(z, l, \alpha)$ is[1]

$$G_N(z, l, \alpha) = \sum_{m=-\infty}^{\infty} G(z, l + mN, \alpha). \tag{12}$$

From Eq. (9) it then follows that

$$G_N(z, l, \alpha) = \frac{2}{z\alpha(s_4 - s_3)} \left[\frac{s_3}{(s_3 - s_1)(s_3 - s_2)} \left(\frac{s_3^{|l|} + s_3^{N-|l|}}{1 - s_3^N} \right) - \frac{s_4}{(s_4 - s_1)(s_4 - s_2)} \left(\frac{s_4^{|l|} + s_4^{N-|l|}}{1 - s_4^N} \right) \right]. \tag{13}$$

The symbols $|l|$ and $N - |l|$ in Eq. (13) label the same site on an N-ring.

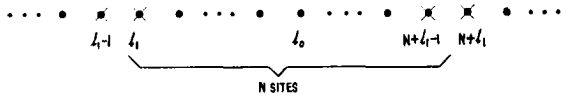

FIG. 1(a). One-dimensional chain. In a nearest and next-nearest neighbor random walk, the mean first passage time for a walker starting at site $l = l_0$ to cross the boundary sites l_1 and $N + l_1 - 1$ is equal to the mean time for the walker to be absorbed at any one of four trap sites l_1, $l_1 - 1$, $N + l_1 - 1$, and $N + l_1$.

B. Mean First Passage Times

We are now in a position to calculate various specific results for a random walk on a one-dimensional lattice with and without boundaries for our model of nearest and next-nearest neighbor step lengths. We will first calculate the mean time required (i.e., the average number of steps) for a random walker starting at lattice site l_0 to cross for the first time the boundaries l_1 and $(N + l_1 - 1)$ of an interval containing N sites [see Fig. 1(a)]. Since the random walker can make steps to next-nearest sites, the "crossing" of the boundary points l_1 and $(N + l_1 - 1)$ can take place either by stepping on these points or by crossing over these points and stepping on the points $(l_1 - 1)$ and $(N + l_1)$. Figure 1(b) shows the equivalence of the "N-ring" formulation of this problem with that on an open lattice. The problem of mean first passage times as formulated above is completely equivalent to the calculation of a "mean time to absorption" or a "mean time to trapping." We shall discuss mean first passage times throughout this paper in terms of mean time to trapping.

Let $Q_n(l - l_0)$ be the probability that a walker starting at site $l = l_0$ reaches point l on the nth step in the presence of the absorbing points $l_1 - 1$ and l_1 on an open lattice [see Figs. 2(a) and 2(b)]. The probability that the walker has not been trapped by the nth step is

$$\sum_{l \neq l_1, l_1 - 1} Q_n(l - l_0).$$

The probability that he is trapped on or before the nth step is

$$\left(1 - \sum_{l \neq l_1, l_1 - 1} Q_n(l - l_0)\right).$$

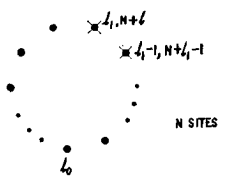

FIG. 1(b). One-dimensional N-ring. The mean absorption time of Fig. 1(a) is equal to the mean absorption time of a walker starting at site l_0 with traps at sites l_1 and $l_1 - 1$.

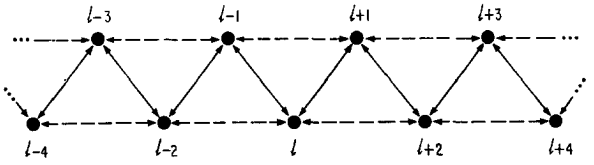

FIG. 2(a). Schematic of nearest and next-nearest neighbor stepping probabilities on an infinite one-dimensional chain. The solid lines connect nearest neighbor sites between which the walker can step with probability $\frac{1}{2}(1 - \alpha)$ in the directions indicated by the arrows. The broken lines connect next-nearest neighbor sites between which the walker can step with probability $\frac{1}{2}\alpha$.

The probability that he is trapped *on* the nth step is then

$$\left(1 - \sum_{l \neq l_1, l_1 - 1} Q_n(l - l_0)\right) - \left(1 - \sum_{l \neq l_1, l_1 - 1} Q_{n-1}(l - l_0)\right).$$

The average number of steps required for trapping is

$${}_2\bar{n}_\alpha(l_0, l_1) = \sum_{n=1}^{\infty} n \sum_{l \neq l_1, l_1 - 1} [Q_{n-1}(l - l_0) - Q_n(l - l_0)], \tag{14}$$

where the subscript 2 denotes the two trapping sites. In terms of the generating function

$$F(z, l - l_0, \alpha) \equiv \sum_{n=0}^{\infty} z^n Q_n(l - l_0), \tag{15}$$

we have

$${}_2\bar{n}_\alpha(l_0, l_1) = -\frac{\partial}{\partial z}\left((1 - z) \sum_{l \neq l_1, l_1 - 1} F(z, l - l_0, \alpha)\right)_{z=1}. \tag{16}$$

Since

$$\sum_l Q_n(l - l_0) = 1, \tag{17}$$

it follows that

$$\sum_l F(z, l - l_0, \alpha) = \sum_{n=0}^{\infty} z^n \sum_l Q_n(l - l_0) = 1/(1 - z). \tag{18}$$

Equation (16) can thus be rewritten as

$${}_2\bar{n}_\alpha(l_0, l_1) = \frac{\partial}{\partial z}\{(1 - z)[F(z, l_1 - l_0, \alpha) + F(z, l_1 - 1 - l_0, \alpha)]\}_{z=1}. \tag{19}$$

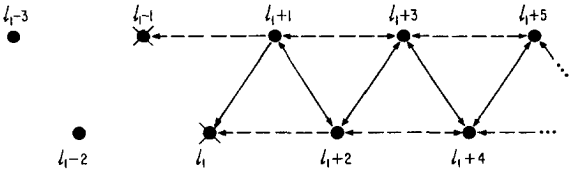

FIG. 2(b). Schematic of nearest and next-nearest neighbor stepping probabilities on an infinite one-dimensional chain with traps at sites l_1 and $l_1 - 1$. The starting point of the walker is at any site $l > l_1$.

On an N-ring [Fig. 1(b)] we have

$$ {}_2\bar{n}_\alpha(l_0, l_1) = \frac{\partial}{\partial z}\{(1-z)[F_N(z, l_1 - l_0, \alpha) + F_N(z, l_1 - 1 - l_0, \alpha)]\}_{z=1}, \quad (20) $$

where $F_N(z, l - l_0, \alpha)$ is the generating function for walks on an N-ring with two trapping sites. In Appendix B we derive the general relation between the generating function $F_N(z, l - l_0, \alpha)$ and the generating function $G_N(z, l, \alpha)$ of Eq. (13). For the special case of $l = l_1$ and $l = l_1 - 1$ required for the evaluation of Eq. (20) this relation becomes

$$ F_N(z, l_1 - l_0, \alpha) + F_N(z, l_1 - 1 - l_0, \alpha) = \frac{G_N(z, l_1 - l_0, \alpha) + G_N(z, l_1 - 1 - l_0, \alpha)}{(1-z)[G_N(z, 0, \alpha) + G_N(z, 1, \alpha)]}. \quad (21) $$

Using the result (13) for $G_N(z, l, \alpha)$ or its expansion in powers of $(1 - z)$ given by

$$ G_N(z, l, \alpha) \sim \frac{1}{N(1-z)} + \frac{N^2 - 6N|l| + 6l^2}{6N(1+3\alpha)} - \frac{(15\alpha+1)}{6N(1+3\alpha)^2} + \frac{2\alpha(X^{|l|} + X^{N-|l|})[(1+3\alpha)(1-\alpha)]^{\frac{1}{2}}}{(1-X^N)(1+3\alpha)^2(1-\alpha)} + o(1-z), \quad (22) $$

where

$$ X = -(1/2\alpha)\{1 + \alpha - [(1+3\alpha)(1-\alpha)]^{\frac{1}{2}}\}, \quad (23) $$

we then find from Eq. (20) that

$$ {}_2\bar{n}_\alpha(l_0, l_1) = \frac{(|l_1 - l_0 - 1|)(N - |l_1 - l_0|)}{(1+3\alpha)} + \alpha N \frac{[(1+3\alpha)(1-\alpha)]^{\frac{1}{2}}}{(1+3\alpha)^2(1-\alpha)} \frac{(1+X)}{(1-X^N)} \times (1 + X^{N-1} - X^{|l_1-l_0-1|} - X^{N-|l_1-l_0|}). \quad (24) $$

Averaging Eq. (24) over the starting point $l = l_0$ yields

$$ {}_2\langle n\rangle_\alpha \equiv \frac{1}{N-2}\sum_{l_0 \neq l_1, l_1 - 1} {}_2\bar{n}_\alpha(l_0, l_1) = \frac{N^3}{6(1+3\alpha)(N-2)} \times \left(1 - \frac{3[(1+3\alpha)(1-\alpha)]^{\frac{1}{2}}}{N(1+3\alpha)} \frac{(1+X^N)}{(1-X^N)} + \frac{2(1-3\alpha)}{N^2(1+3\alpha)}\right). \quad (25) $$

In the limit $N \to \infty$, we have

$$ \lim_{N\to\infty} {}_2\langle n\rangle_\alpha \to \frac{N^2}{6(1+3\alpha)}. \quad (26) $$

The above results should be compared with the corresponding results for nearest neighbor random walks on an N-ring with one trapping point[1,6]:

$$ {}_1\bar{n}(l_0, l_1) = (|l_1 - l_0|)(N - |l_1 - l_0|), \quad (27) $$

$$ {}_1\langle n\rangle \equiv \frac{1}{N-1}\sum_{l_0 \neq l_1} {}_1\bar{n}(l_0, l_1) = \frac{N(N+1)}{6}, \quad (28) $$

and

$$ \lim_{N\to\infty} {}_1\langle n\rangle \to N^2/6. \quad (29) $$

For $\alpha = 0$, which corresponds to the pure nearest neighbor random walk, Eqs. (24)–(26) reduce to Eqs. (27)–(29) as they should.

The case $\alpha = 1$, which corresponds to a random walk with next-nearest neighbor transitions only, must be considered separately. The generating function $G(z, l, \alpha)$ on an infinite lattice given by Eq. (9) continuously approaches, as $\alpha \to 1$, that of a nearest neighbor random walk on an infinite lattice[1] in which only every other lattice site is available to the walker:

$$ \lim_{\alpha\to 1} G(z, l, \alpha) \to \frac{1}{(1-z^2)^{\frac{1}{2}}}\left(\frac{1-(1-z^2)^{\frac{1}{2}}}{z}\right)^{|l/2|}, \quad l \text{ even}, $$
$$ \to 0, \quad l \text{ odd}. \quad (30) $$

The generating function $G_N(z, l, \alpha)$ on an N-ring given by Eqs. (13) and (22) is discontinuous at $\alpha = 1$. For $0 \leq \alpha < 1$, the walker has N sites available. At $\alpha = 1$, two situations can occur. If N is even, the walker can only step on $N/2$ of the sites, thus performing effectively a nearest neighbor random walk on a ring which is half as long as the original N-ring. If N is odd, all the sites are available to the walker, but half of them can only be reached after making a full circuit on the other half. From Eq. (26) which has no discontinuity as $\alpha \to 1$, we find

$$ {}_2\langle n\rangle_{\alpha=1} \to N^2/24. \quad (31) $$

The mean first passage times for trapping for nearest and for next-nearest neighbor random walks thus differ at most by a factor of 4. This result is intuitively obvious [e.g., see Eq. (29)] in that for a finite interval of N lattice sites, the walker has only $N/2$ sites available in the limiting case of the pure next-nearest neighbor random walk.

In the above development we have discussed mean first passage times where the events to be counted were the arrival of the random walker at a boundary point or his crossing over the boundary point. We will now discuss the problem of mean first passage time for the case where the random walker *must*

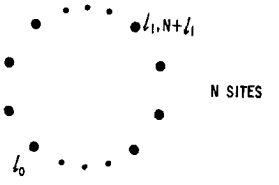

FIG. 3. One-dimensional N-ring. The walker begins his walk at $l = l_0$ and the trap is at site $l_1 = N + l_1$.

step on the boundary point. In nearest neighbor random walks there is no need to make this distinction; there is, however, a need to make such a distinction for random walks involving transitions to next nearest sites. It is intuitively clear even before making the detailed calculation that the mean first passage time for stepping on a given lattice site will be considerably longer than that for stepping or crossing that lattice site for finite N. To develop these results we use the N-ring shown in Fig. 3, with the random walker starting at $l = l_0$ and arriving at $l = l_1$ for the first time. (The mean first passage time for stepping on a particular lattice site on an infinite open lattice is clearly infinite.) The mean recurrence time is just a special case of the above with $l_1 = l_0$.

Let $D_n(l - l_0)$ be the probability that a walker starting at site $l = l_0$ reaches point l on the nth step for the first time on an "open" lattice. The probability $D_n(l - l_0)$ can be related to the probability $P_n(l - l_0)$ of Sec. IIA by[1]

$$P_n(l - l_0) = \sum_{j=1}^{n} D_j(l - l_0)P_{n-j}(0), \tag{32}$$

where $P_{n-j}(0)$ is the probability that a walker starting at any lattice site l returns to the same site l after $n - j$ steps. Equation (32) expresses the fact that a walker can arrive at site l after n steps by arriving there for the first time after j steps and returning to site l in $n - j$ steps. We define the generating function $E(z, l - l_0, \alpha)$ by

$$E(z, l - l_0, \alpha) \equiv \sum_{n=1}^{\infty} z^n D_n(l - l_0). \tag{33}$$

Multiplying Eq. (33) by z^n and summing from $n = 1$ to $n = \infty$ gives, where δ_{l,l_0} is the Kronecker delta,

$$\begin{aligned}\sum_{n=1}^{\infty} z^n P_n(l - l_0) &= G(z, l - l_0, \alpha) - \delta_{l,l_0} \\ &= \sum_{n=1}^{\infty}\sum_{j=1}^{n} z^j D_j(l - l_0) z^{n-j} P_{n-j}(0) \\ &= \sum_{j=1}^{\infty}\sum_{m=0}^{\infty} z^j D_j(l - l_0) z^m P_m(0) \\ &= E(z, l - l_0, \alpha) G(z, 0, \alpha). \end{aligned} \tag{34}$$

Solving for $E(z, l - l_0, \alpha)$ yields

$$E(z, l - l_0, \alpha) = \frac{G(z, l - l_0, \alpha) - \delta_{l,l_0}}{G(z, 0, \alpha)}. \tag{35}$$

On an N-ring

$$E_N(z, l - l_0, \alpha) = \frac{G_N(z, l - l_0, \alpha) - \delta_{l,l_0}}{G_N(z, 0, \alpha)}. \tag{36}$$

The mean first passage time for arriving at site $l = l_1$ is

$${}_1\bar{n}_\alpha(l_0, l_1) = \frac{\partial}{\partial z} E_N(z, l_1 - l_0, \alpha)\Big|_{z=1}. \tag{37}$$

For $l_1 \neq l_0$ we have

$${}_1\bar{n}_\alpha(l_0, l_1) = \frac{\partial}{\partial z} \frac{G_N(z, l_1 - l_0, \alpha)}{G_N(z, 0, \alpha)}\Bigg|_{z=1}, \tag{38}$$

where the subscript 1 denotes the one trapping site. For $l_1 = l_0$ we have

$${}_1\bar{n}_\alpha(l_0, l_0) = -\frac{\partial}{\partial z} \frac{1}{G_N(z, 0, \alpha)}\Bigg|_{z=1}, \tag{39}$$

where ${}_1\bar{n}_\alpha(l_0, l_0)$ is the mean recurrence time. Using $G_N(z, l - l_0, \alpha)$ of Eq. (22), we obtain

$$\begin{aligned}{}_1\bar{n}_\alpha(l_0, l_1) = {} & \frac{(|l_1 - l_0|)(N - |l_1 - l_0|)}{(1 + 3\alpha)} \\ & + \frac{2N\alpha[(1 + 3\alpha)(1 - \alpha)]^{\frac{1}{2}}}{(1 + 3\alpha)^2(1 - \alpha)} \\ & \times \frac{(1 + X^N - X^{|l_1 - l_0|} - X^{N-|l_1-l_0|})}{(1 - X^N)}, \end{aligned} \tag{40}$$

where X is defined in Eq. (23). Averaging Eq. (40) over starting points $l = l_0$ of the walker yields

$$\begin{aligned}{}_1\langle n\rangle_\alpha &\equiv \frac{1}{N - 1}\sum_{l_0 \neq l_1} {}_1\bar{n}_\alpha(l_0, l_1) \\ &= \frac{N^3}{6(1 + 3\alpha)(N - 1)} \\ &\quad \times \left(1 + \frac{12[(1 + 3\alpha)(1 - \alpha)]^{\frac{1}{2}}(1 + X^N)}{N(1 - \alpha)(1 + 3\alpha)(1 - X^N)} \right. \\ &\quad \left. - \frac{15\alpha + 1}{N^2(1 + 3\alpha)}\right). \end{aligned} \tag{41}$$

In the limit as $N \to \infty$, we have

$$\lim_{N\to\infty} {}_1\langle n\rangle_\alpha \to \frac{N^2}{6(1 + 3\alpha)}, \tag{42}$$

which is *asymptotically* identical with ${}_2\langle n\rangle_\alpha$ of Eq. (26). The corresponding results for the mean first passage times for the pure nearest neighbor random walk are given by Eqs. (27)–(29).

As we mentioned earlier, we would expect for finite N that the mean first passage time for stepping on a particular lattice site is longer than that for stepping on or crossing that lattice site. In the former case, on an N-ring, the walker could circle the ring several

times in either or both directions before actually stepping on the particular site, whereas in the case of stepping on *or* crossing the site the walker cannot circumnavigate the ring and is clearly trapped more quickly. From Eqs. (25) and (41) it follows that

$$ {}_1\langle n\rangle_\alpha - {}_2\langle n\rangle_\alpha = Ng(\alpha) + f(\alpha) > 0, \tag{43}$$

for $\alpha \neq 0$. For $\alpha = 0$, which corresponds to the pure nearest neighbor random walk, $g(\alpha) = f(\alpha) = 0$, and ${}_1\langle n\rangle_0 = {}_2\langle n\rangle_0$ as should be the case.

For the mean recurrence time we find from Eq. (39)

$$ {}_1\bar{n}_\alpha(l_0, l_0) = N, \quad 0 \leq \alpha < 1. \tag{44}$$

This result is a special case of a general theorem on recurrence times in Markoff chains.[1] As applied to random walks, it states that the mean recurrence time of a random walker is equal to the total number of sites available to the walker, independently of the dimensionality of the walk and of the form of the stepping probability function $p(l_i - l_j)$. For the pure next-nearest neighbor random walk with $\alpha = 1$ we find ${}_1\bar{n}_\alpha(l_0, l_0) = N/2$ for N even and ${}_1\bar{n}_\alpha(l_0, l_0) = N$ for N odd. The dispersion in the mean recurrence time does depend very sensitively on the details of the random walk. We will evaluate this quantity for exponentially distributed steps in Sec. III.

Analogous calculations can be carried out for lattices with one or two reflecting boundaries. We will, however, postpone such calculations to a subsequent paper where we plan to discuss some physical applications.

C. Excursion from the Origin

We now direct our attention to the "excursion" from his point of origin of our random walker on a perfect infinite lattice after a given number of steps, i.e., to the moments of $P_n(l - l_0)$ with respect to $(l - l_0)$. With no loss of generality we can take the starting point of the walker to be $l_0 \equiv 0$.

The moments of $P_n(l)$ can be obtained from the stepping probabilities $p(l_i - l_j)$ by inverting Eq. (5) [1]:

$$ \lambda^n(\phi) = \sum_{l=-\infty}^{\infty} P_n(l) e^{il\phi}, \tag{45}$$

where $\lambda^n(\phi)$ is defined in Eq. (4). For the moments of $P_n(l)$ we then obtain

$$ \sum_{l=-\infty}^{\infty} l^m P_n(l) = (-i)^m \left(\frac{\partial^m}{\partial \phi^m} \lambda^n(\phi) \right)_{\phi=0}. \tag{46}$$

It is clear that for a *symmetric* random walk, i.e., $P_n(l) = P_n(-l)$, all the odd moments are zero. Thus the mean distance from the origin ($m = 1$) is zero.

We will therefore consider a general nearest and next-nearest neighbor random walk using the following stepping probabilities:

$$\begin{aligned} p(l_i - l_j) &= \beta, \quad \text{for} \quad l_i - l_j = 2, \\ &= \gamma, \quad \text{for} \quad l_i - l_j = 1, \\ &= \epsilon, \quad \text{for} \quad l_i - l_j = -1, \\ &= \eta, \quad \text{for} \quad l_i - l_j = -2, \\ &= 0, \quad \text{otherwise}, \end{aligned} \tag{47}$$

where

$$ \beta + \gamma + \epsilon + \eta = 1 \tag{48}$$

and where β, γ, ϵ, and η are ≥ 0. Equation (4) then becomes

$$ \lambda^n(\phi) = (\beta e^{2i\phi} + \gamma e^{i\phi} + \epsilon e^{-i\phi} + \eta e^{-2i\phi})^n, \tag{49}$$

and, using Eqs. (46) and (48), we obtain

$$ \langle l\rangle \equiv \sum_{l=-\infty}^{\infty} l P_n(l) = n[(2\beta + \gamma) - (2\eta + \epsilon)], \tag{50}$$

$$\begin{aligned} \langle l^2\rangle &\equiv \sum_{l=-\infty}^{\infty} l^2 P_n(l) \\ &= n[1 + 3(\beta + \eta)] \\ &\quad + n(n-1)[(2\beta + \gamma) - (2\eta + \epsilon)]^2 \end{aligned} \tag{51}$$

and, for the dispersion,

$$\begin{aligned} &\langle l^2\rangle - \langle l\rangle^2 \\ &\quad = n\{1 + 3(\beta + \eta) - [(2\beta + \gamma) - (2n + \epsilon)]^2\}. \end{aligned} \tag{52}$$

For a symmetric random walk, where $\beta = \eta = \alpha/2$ and $\gamma = \epsilon = (1 - \alpha)/2$, we have

$$ \langle l\rangle = 0, \tag{53}$$

$$ \langle l^2\rangle = n(1 + 3\alpha). \tag{54}$$

The corresponding results for a nearest neighbor random walk,[1] with $p(1) = p$, $p(-1) = q = 1 - p$, and $p(l_i - l_j) = 0$ otherwise, are

$$ \langle l\rangle_1 = n(p - q), \tag{55}$$

$$ \langle l^2\rangle_1 = n + n(n-1)(p - q)^2, \tag{56}$$

$$ \langle l^2\rangle_1 - \langle l\rangle_1^2 = n[1 - (p - q)^2], \tag{57}$$

where the subscript 1 denotes the nearest neighbor results. For a symmetric nearest neighbor random walk ($p = q = \frac{1}{2}$)

$$ \langle l\rangle_1 = 0, \tag{58}$$

$$ \langle l^2\rangle_1 = n. \tag{59}$$

Equation (50) shows that it is possible to have a zero average displacement from the origin even for an asymmetric next-nearest neighbor random walk, i.e., for $\beta + \gamma \neq \epsilon + n$. For instance, $\langle l\rangle = 0$ for $\beta = \frac{3}{16}$,

$\gamma = \frac{4}{16}$, $\epsilon = \frac{8}{16}$, and $\eta = \frac{1}{16}$. What we wish to emphasize here is that the mean distance from the origin is not determined solely by the total probabilities of stepping right or left, but depends on the individual probabilities for steps to nearest and next-nearest neighbor sites.

The dispersion $\langle l^2 \rangle - \langle l \rangle^2$ also depends sensitively on the value of each of the stepping probability parameters in a next-nearest neighbor random walk [see Eqs. (52) and (57)]. The dispersion becomes greater when the probabilities of reaching next-nearest neighbors (β and η) increase. In a symmetric random walk [Eqs. (54), (59)] the dispersion in the next-nearest neighbor random walk is greater than that for the nearest neighbor random walk, with the ratio $\langle l^2 \rangle / \langle l^2 \rangle_1$ having a maximum value of 4 for $\alpha = 1$. This is due to the fact that, in a symmetric, pure, next-nearest neighbor random walk, the walker, after n steps, can be at most twice as far from the origin as in the pure nearest neighbor random walk.

A comparison of the various results presented above for nearest neighbor walks and for nearest *plus* next-nearest neighbor walks shows that the effect of the next-nearest neighbor step is not very large. In general, most of the mean properties, such as mean absorption times, mean excursion, etc., for the one and two step walks differ at most by a ratio of $(1 + 3\alpha)$ which has a maximum value of 4 when $\alpha = 1$. Certainly, there are no order of magnitude changes. However prosaic, this result is of real interest. It should lay to rest, for instance, once and for all, the ideas expressed by some authors, including one of the present ones,[7] that next-nearest neighbor transitions due to variational anharmonicity of oscillators could sufficiently improve random walk calculations of the rate of dissociation of diatomic molecules to make them agree with experiments. There are a number of other examples in the literature where the possibility of next-nearest neighbor transitions has been invoked as a possible means for order magnitude changes in random walk calculations. It is now clear that this assumption is not correct.

If next-nearest neighbor transitions do not produce any dramatic effects, the question of course presents itself as to the effect of *very long jumps* in random walk calculations. Some years ago, Shuler and Weiss[8] showed via computer calculations that order of magnitude effects can be obtained in mean first passage times by the use of exponentially distributed jump probabilities of the form $p(l) = ce^{-\alpha l}$. We were unable at that time to effect an analytic solution of our random walk problem. The powerful new methods developed by Montroll[1-4] since that time now permits one to carry out such an analytical treatment for random walks with very long steps. The next section is devoted to developing the analytical result for such random walks using the above exponentially distributed jump probabilities.

III. RANDOM WALK WITH EXPONENTIALLY DISTRIBUTED STEP LENGTHS

A. Generating Function on Perfect Infinite Lattice and on Perfect N-Ring

On a perfect infinite lattice we use the stepping probabilities $p(l_i, l_j)$ for the random walker:

$$\begin{aligned} p(l_i - l_j) &= c_1 e^{-(l_i - l_j)a}, \quad l_i > l_j, \\ &= c_2 e^{-(l_j - l_i)a}, \quad l_i < l_j, \\ &= 0, \qquad\qquad l_i = l_j, \end{aligned} \tag{60}$$

with a, b, c_1, $c_2 > 0$. This stepping probability function describes an asymmetric random walk if $a \neq b$ and/or $c_1 \neq c_2$. The parameter a determines the relative probabilities that the walker takes steps of different lengths to the right, e.g.,

$$\frac{p(l_i - l_j + 1)}{p(l_i - l_j)} = e^{-a}, \quad l_i > l_j. \tag{61}$$

The parameter b determines the same relative probabilities for steps to the left. The parameters c_1 and c_2 in combination with a and b determine whether the walker will step preferentially to his left or to his right, e.g.,

$$\frac{p(1)}{p(-1)} = \frac{c_1}{c_2} e^{-(a-b)}. \tag{62}$$

The constants a, b, c_1, and c_2 are related by the normalization of the $p(l_i - l_j)$. For $k = l_i - l_j$ we find

$$\sum_{k=-\infty}^{\infty} p(k) = \frac{c_1}{e^a - 1} + \frac{c_2}{e^b - 1} = 1. \tag{63}$$

When $a = b$ and $c_1 = c_2 = c$, we have a symmetric random walk with exponentially distributed steps. We then write

$$\begin{aligned} p(l_i - l_j) &= ce^{-|l_i - l_j|a}, \quad |l_i - l_j| > 0, \\ &= 0, \qquad\qquad l_i = l_j, \end{aligned} \tag{64}$$

with the normalization relation

$$c = (e^a - 1)/2. \tag{65}$$

The choice $p(0) = 0$ is completely arbitrary and any other choice could have been made subject to the normalization condition (1).

The nearest neighbor random walk results will be recovered in the limit as $a \to \infty$ since for that limit a nonzero stepping probability $p(l_i - l_j)$ is obtained

only for $|l_i - l_j| = 1$. In the limit as $a \to 0$, $p(l_i - l_j) \to (1/N)$ (for a finite lattice with N sites) and the random walker can make steps of any length $|l_j - l_i|$ with equal probability $1/N$. Adjustment of the parameters a and b thus enable one to study random walks with a very broad range of step probabilities from nearest neighbor to equal probability of step length of any size.

To obtain the generating function for the stepping probabilities given by Eq. (60), we need to evaluate the quantity $\lambda(\phi)$:

$$\begin{aligned}\lambda(\phi) &= \sum_{k=-\infty}^{\infty} p(k)e^{ik\phi} \\ &= c_1 \sum_{k=1}^{\infty} e^{-ka}e^{ik\phi} + c_2 \sum_{k=-1}^{-\infty} e^{kb}e^{ik\phi} \\ &= c_1 \sum_{k=1}^{\infty} e^{-ka}e^{ik\phi} + c_2 \sum_{k=1}^{\infty} e^{-kb}e^{-k\phi} \\ &= \frac{c_1}{e^{(a-i\phi)} - 1} + \frac{c_2}{e^{(b+i\phi)} - 1}. \end{aligned} \tag{66}$$

The generating function is then

$$G(z, l, a, b, c_1, c_2) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-il\phi} \left(1 - \frac{zc_1}{e^{(a-i\phi)} - 1} - \frac{zc_2}{e^{(b+i\phi)} - 1}\right)^{-1} d\phi. \tag{67}$$

The evaluation of this integral, which is detailed in Appendix A, yields

$$\begin{aligned} G(z, l, a, b, c_1, c_2) &= (y_1)^l \frac{[A + z(c_1 - c_2)D]}{B}, \quad l \geq 1, \\ &= (y_2)^{-l} \frac{[A + z(c_2 - c_1)D]}{B}, \quad l \leq -1, \\ &= \frac{[A + (2 + zc_1 + zc_2)D]}{B}, \quad l = 0, \end{aligned} \tag{68}$$

where

$$A = (e^{a+b} - 1)(zc_1 + zc_2) + 2e^{a+b}z^2c_1c_2 - z^2c_1^2 - z^2c_2^2, \tag{69}$$

$$D = [(e^{a+b} + 1 + zc_1 + zc_2)^2 - 4e^{a+b}(1 + zc_1)(1 + zc_2)]^{\frac{1}{2}}, \tag{70}$$

$$B = 2(1 + zc_1)(1 + zc_2)D, \tag{71}$$

$$y_1 = \frac{e^{a+b} + 1 + z(c_1 + c_2) - D}{2e^a(1 + zc_2)}, \tag{72}$$

$$y_2 = \frac{e^{a+b} + 1 + z(c_1 + c_2) - D}{2e^b(1 + zc_1)} \tag{73}$$

and where the constants c_1, c_2, a, and b are related by Eq. (63). For the symmetric random walk the generating function reduces to

$$\begin{aligned} G(z, l, a) &= \frac{(x)^{|l|}z(e^{2a} - 1)}{[2 + z(e^a - 1)]D_0}, \quad |l| \geq 1, \\ &= \frac{z(e^{2a} - 1) + 2D_0}{[2 + z(e^a - 1)]D_0}, \quad l = 0, \end{aligned} \tag{74}$$

where

$$D_0 = \{(e^a + 1)[e^a + 1 + z(e^a - 1)](1 - z)\}^{\frac{1}{2}} \tag{75}$$

and

$$x = \frac{e^{2a} + 1 + z(e^a - 1) - (e^a - 1)D_0}{e^a[2 + z(e^a - 1)]}. \tag{76}$$

The generating functions of Eqs. (68) and (74) generate random walks on a perfect infinite lattice. The corresponding generating functions for walks on a perfect N-ring can be evaluated, as in Sec. II, from Eq. (12). Before doing this, we will work out the corresponding stepping probabilities on an N-ring, which we will denote by $p_N(l_i - l_j)$. Since an N-ring can be thought of as an infinite chain wrapped up in such a way that sites $\cdots l - 2N,\ l - N,\ l,\ l + N,\ l + 2N \cdots$ coincide for each l, the $p_N(l_i - l_j)$ are related to the $p(l_i - l_j)$ of Eq. (60) by an equation similar to Eq. (12). If we denote $l_i - l_j$ by k, we can write

$$p_N(k) = \sum_{n=-\infty}^{\infty} p(k + nN), \quad -N < k < N. \tag{77}$$

Suppose $0 < k < N$. Then

$$\begin{aligned} p_N(k) &= \sum_{n=0}^{\infty} p(k + nN) + \sum_{n=1}^{\infty} p(k - nN) \\ &= c_1e^{-ka} \sum_{n=0}^{\infty} (e^{-Na})^n + c_2e^{kb} \sum_{n=1}^{\infty} (e^{-Nb})^n \\ &= \left(\frac{c_1}{1 - e^{-Na}}\right)e^{-ka} + \left(\frac{c_2}{1 - e^{-Nb}}\right)e^{-(N-k)b}. \end{aligned} \tag{78}$$

For $-N < k < 0$ we have

$$\begin{aligned} p_N(k) &= \sum_{n=1}^{\infty} p(k + nN) + \sum_{n=0}^{\infty} p(k - nN) \\ &= c_1e^{-ka} \sum_{n=1}^{\infty} (e^{-Na})^n + c_2e^{kb} \sum_{n=0}^{\infty} (e^{-Nb})^n \\ &= \left(\frac{c_1}{1 - e^{-Na}}\right)e^{-(N+k)a} + \left(\frac{c_2}{1 - e^{-Nb}}\right)e^{kb}. \end{aligned} \tag{79}$$

If $k = 0$ (or, equivalently, $k = N$ and $k = -N$), we have

$$\begin{aligned} p_N(k) &= \sum_{n=-\infty}^{\infty} p(nN) \\ &= c_1 \sum_{n=1}^{\infty} (e^{-Na})^n + c_2 \sum_{n=1}^{\infty} (e^{-Nb})^n \\ &= \frac{c_1e^{-Na}}{1 - e^{-Na}} + \frac{c_2e^{-Nb}}{1 - e^{-Nb}}. \end{aligned} \tag{80}$$

The combination of exponential terms in Eqs. (78)–(80) reflects the fact that on an N-ring the walker can get from one site to another site a distance k away either by making a step of length $|k|$ in one direction or by a step of length $N - |k|$ in the opposite direction. For a symmetric random walk [i.e., $a = b$, $c_1 = c_2 = (e^a - 1)/2$], the N-ring stepping probabilities become

$$p_N(k) = \frac{(e^a - 1)}{2(1 - e^{-Na})}(e^{-|k|a} + e^{-(N-|k|)a}), \quad 0 < |k| < N,$$

$$p_N(-N) = p_N(0) = p_N(N) = \frac{(e^a - 1)e^{-Na}}{(1 - e^{-Na})}. \quad (81)$$

In the limit $a \to 0$, $b \to 0$ (i.e., $a \ll 1/N$, $b \ll 1/N$) the walker can, with equal probability, step on any of the N points of the ring at each step. It is easily seen from Eqs. (78)–(80) and the normalization condition that

$$\lim_{\substack{a\to 0 \\ b\to 0}} p_N(k) = 1/N, \quad 0 \leq |k| < N. \quad (82)$$

The generating function on an N-ring obtained by substituting Eq. (68) into Eq. (12) is

$$\begin{aligned} G_N(z, l, a, b, c_1, c_2) &= \left(\frac{A + z(c_1 - c_2)D}{B}\right)\frac{y_1^l}{1 - y_1^N} + \left(\frac{A + z(c_2 - c_1)D}{B}\right)\frac{y_2^{N-l}}{1 - y_2^N}, \quad l \geq 1, \\ &= \left(\frac{A + z(c_1 - c_2)D}{B}\right)\frac{y_1^{N+l}}{1 - y_1^N} + \left(\frac{A + z(c_2 - c_1)D}{B}\right)\frac{y_2^{-l}}{1 - y_2^N}, \quad l \leq -1, \\ &= \left(\frac{A + z(c_1 - c_2)D}{B}\right)\frac{y_1^N}{1 - y_1^N} + \left(\frac{A + z(c_2 - c_1)D}{B}\right)\frac{y_2^N}{1 - y_2^N} + \left(\frac{A + (2 + zc_1 + zc_2)D}{B}\right), \quad l = 0, \quad (83) \end{aligned}$$

where A, B, D, y_1, and y_2 are defined in Eqs. (69)–(73). For a symmetric random walk, the generating function on an N-ring becomes

$$\begin{aligned} G_N(z, l, a) &= \left(\frac{z(e^{2a} - 1)}{[2 + z(e^a - 1)]D_0}\right)\frac{(x^{|l|} + x^{N-|l|})}{(1 - x^N)}, \quad |l| \geq 1, \\ &= \left(\frac{z(e^{2a} - 1)}{[2 + z(e^a - 1)]D_0}\right)\frac{(1 + x^N)}{(1 - x^N)} + \frac{2}{[2 + z(e^a - 1)]}, \quad l = 0, \quad (84) \end{aligned}$$

where D_0 and x are defined in Eqs. (75) and (76).

It is easily checked that the generating functions given in this section reduce to the corresponding generating function for the nearest neighbor random walk in the limit $a \to \infty$, $b \to \infty$.

B. Mean First Passage Times

We will first calculate the mean time required for a random walker with exponentially distributed step lengths starting at lattice site $l = l_0$ to cross for the first time the boundaries l_1 and $(N + l_1)$ of an interval containing $N - 1$ sites [see Fig. 4(a)]. Since the random walker can make steps of any length, the "crossing" of the boundary points l_1 and $(N + l_1)$ can take place either by stepping on these points or by crossing these points and stepping on the points $l_1 - k$ and $N + l_1 + k$ for all $k \geq 1$. The problem of the mean first passage time formulated above is completely equivalent to the problem of mean absorption time or mean time for trapping with absorbing boundaries at l_1 and $(N + l_1)$ in that we define absorbing boundaries at l_1 and $(N + l_1)$ by placing traps at all sites $l_1 - k$ and $N + l_1 + k$ with $k \geq 0$ [see Fig. 4(b)].

In Sec. IIB, where we dealt with the same problem for a nearest and next-nearest neighbor random walk, we formulated the problem of two absorbing boundaries in terms of a walk on an N-ring containing two trapping sites. In the present case, where we allow the walker to take steps of any length and therefore need an infinite number of traps to define an absorbing boundary, we cannot follow the same N-ring formulation. We will thus work on an open lattice. We will use the N-ring formulation later, when we consider the mean first passage time for stepping on a particular lattice site.

Let $Q_n(l - l_0)$ be the probability that a walker starting at site $l = l_0$ reach point l on the nth step in the presence of the absorbing points $l_1 - k$ and $N + l_1 + k$, $k \geq 0$. The probability that the walker has

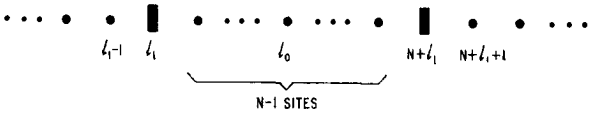

FIG. 4(a). One-dimensional infinite chain with absorbing boundaries at sites l_1 and $N + l_1$. The walker starts at site l_0, with $l_1 < l_0 < N + l_1$.

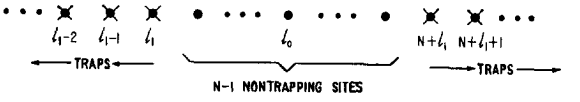

FIG. 4(b). For a random walk with exponentially distributed step lengths, the absorbing boundaries of Fig. 4(a) can be constructed by placing traps at sites $l_1 - k$ and $N + l_1 + k$ with $k = 0, 1, \cdots, \infty$.

not been trapped by the nth step is

$$\sum_{l=l_1+1}^{N+l_1-1} Q_n(l-l_0).$$

The probability that he is trapped on or before the nth step is

$$\left(1-\sum_{l=l_1+1}^{N+l_1-1} Q_n(l-l_0)\right).$$

The probability that he is trapped *on* the nth step is

$$\left(1-\sum_{l=l_1+1}^{N+l_1-1} Q_n(l-l_0)\right)-\left(1-\sum_{l=l_1+1}^{N+l_1-1} Q_{n-1}(l-l_0)\right).$$

The average number of steps required for trapping is then

$${}_\infty\bar{n}_{a,b,c_1,c_2}(l_0,l_1) = \sum_{n=0}^{\infty} n \sum_{l=l_1+1}^{N+l_1-1} [Q_{n-1}(l-l_0)-Q_n(l-l_0)], \quad (85)$$

where the subscript ∞ denotes the infinite number of trapping sites. In terms of the generating function

$$F(z,l-l_0,a,b,c_1,c_2) \equiv \sum_{n=0}^{\infty} z^n Q_n(l-l_0), \quad (86)$$

we have

$${}_\infty\bar{n}_{a,b,c_1,c_2}(l_0,l_1) = -\frac{\partial}{\partial z}\left((1-z)\sum_{l=l_1+1}^{N+l_1-1} F(z,l-l_0,a,b,c_1,c_2)\right)_{z=1}. \quad (87)$$

We have evaluated Eq. (87) only for the symmetric random walk [i.e., $a=b$, $c_1=c_2=(e^a-1)/2$]. We can thus simplify the notation and write

$${}_\infty\bar{n}_a(l_0,l_1) = -\frac{\partial}{\partial z}\left((1-z)\sum_{l=l_1+1}^{N+l_1-1} F(z,l-l_0,a)\right)_{z=1}. \quad (88)$$

Following the same steps here as we did between Eqs. (16) and (19), we can rewrite the sum in Eq. (88) over nontrapping sites as a sum over trapping sites. Since

$$\sum_{l=-\infty}^{\infty} F(z,l-l_0,a) = \frac{1}{1-z}, \quad (89)$$

we have

$${}_\infty\bar{n}_a(l_0,l_1) = \frac{\partial}{\partial z}\left[(1-z)\left(\sum_{l=-\infty}^{l_1} F(z,l-l_0,a) + \sum_{l=l_1+N}^{\infty} F(z,l-l_0,a)\right)\right]_{z=1}. \quad (90)$$

In Appendix B we derive the relation between the combination of generating functions $F(z,l-l_0,a)$ appearing in Eq. (90) and the generating function $G(z,l,a)$ for a perfect infinite lattice of Eq. (74). The result is

$$\sum_{l=-\infty}^{l_1} F(z,l-l_0,a) + \sum_{l=l_1+N}^{\infty} F(z,l-l_0,a) = \frac{1}{1-z} - \sum_{l=l_1+1}^{N+l_1-1} F(z,l-l_0,a) = \frac{[G(z,l_0-l_1,a)+G(z,N-(l_0-l_1),a)](1+x)}{(1-z)\{[(1-xe^{-a})-x^N(x-e^{-a})]G(z,0,a)+[(1+e^{-a})+x^{N-1}(1+2x-e^{-a})]G(z,1,a)\}}, \quad (91)$$

where x is defined in Eq. (76). The expression corresponding to Eq. (91) in the presence of only one absorbing boundary is

$$\frac{1}{1-z} - \lim_{N\to\infty}\sum_{l=l_1+1}^{N+l_1-1} F(z,l-l_0,a) = \sum_{l=-\infty}^{l_1} F(z,l-l_0,a) = \frac{[G(z,l_1-l_0,a)+G(z,l_1-1-l_0,a)]}{(1-z)[(1-xe^{-a})G(z,0,a)+(1+e^{-a})G(z,1,a)]}, \quad (92)$$

where we have used the fact that $x<1$ and

$$\lim_{N\to\infty} x^N = 0.$$

Substituting Eq. (91) into Eq. (88) or into Eq. (90), we find

$${}_\infty\bar{n}_a(l_0,l_1) = \frac{(1-e^{-a})^2}{(1+e^{-a})}(l_0-l_1)[N-(l_0-l_1)] + e^{-a}\left[\left(N\frac{1-e^{-a}}{1+e^{-a}}\right)+1\right]. \quad (93)$$

Averaging Eq. (90) over the starting point l_0 yields

$${}_\infty\langle n\rangle_a \equiv \frac{1}{N-1}\sum_{l_0=l_1+1}^{N+l_1-1} {}_\infty\bar{n}_a(l_0,l_1) = \frac{(1-e^{-a})^2}{(1+e^{-a})}\frac{N(N+1)}{6} + \frac{e^{-a}(1-e^{-a})}{(1+e^{-a})}N + e^{-a}. \quad (94)$$

In the limit $N\to\infty$, for $a \gg 1/N$,

$${}_\infty\langle n\rangle_a \xrightarrow[N\to\infty]{} \frac{(1-e^{-a})^2}{(1+e^{-a})}\frac{N^2}{6}. \quad (95)$$

The corresponding results for nearest neighbor random walks are given in Eqs. (27)–(29). It can readily be verified that Eqs. (92)–(95) approach the nearest neighbor random walk results in the limit $a \to \infty$. Thus, for instance,

$$ {}_{\infty}\langle n\rangle = {}_{1}\langle n\rangle = N(N+1)/6,$$
$$\lim_{N\to\infty} {}_{\infty}\langle n\rangle = {}_{1}\langle n\rangle \to N^2/6, \quad \text{etc.}$$

In the limit $a \to 0$ (i.e., $a \ll 1/N$), when all lattice sites (whether or not they are traps) are equally likely to be reached at each step, we have

$$\lim_{a\to 0} {}_{\infty}\bar{n}_a(l_0, l_1) = \lim_{a\to 0} {}_{\infty}\langle n\rangle_a \approx 1 + Na/2 \approx 1, \quad (96)$$

and the walker, in the extreme case $a = 0$, is absorbed on the average after the first step. This results from the fact that there are an infinite number of trapping sites $k = 0, 1, \cdots, \infty$ and only a finite number N of nontrapping sites, so that the probability that the walker steps on a nontrapping site on his first step is a set of measure zero. It is evident from a comparison of the above results for the random walk with exponentially distributed steps with the nearest neighbor random walk results that there are now indeed order of magnitude differences. Thus the ratio of the mean first passage times for the cases $a \to 0$ and $a \to \infty$ is of order N^2 with $0 < N \leq \infty$.

The above discussion was concerned with mean first passage times for a random walker to *arrive at or cross* the boundaries of a finite interval of lattice sites. The derivation of the first passage time for *stepping on* a particular lattice site on an N-ring has already been given in Sec. II, Eqs. (32)–(38). For a random walk with exponentially distributed step lengths one obtains

$${}_{1}\langle n\rangle_a = \frac{(1-e^{-a})^2}{(1+e^{-a})}\frac{N(N+1)}{6} + \frac{2Ne^{-a}}{(1+e^{-a})}. \quad (97)$$

In the limit $N \to \infty$, for $a \gg 1/N$,

$${}_{1}\langle n\rangle_a \xrightarrow[N\to\infty]{} \frac{(1-e^{-a})^2}{(1+e^{-a})}\frac{N^2}{6}. \quad (98)$$

In the limit $a \to 0$, when all lattice sites on the N-ring can be reached with equal probability at each step, we find

$$\lim_{a\to 0} {}_{1}\langle n\rangle_a \approx N(1 - a/2) \approx N. \quad (99)$$

The mean recurrence time ($l_0 = l_1$) for a random walk on an N-ring is given by [see Eq. (39)]

$$\bar{n}_a(l_0, l_0) = -\frac{\partial}{\partial z}\left(\frac{1}{G_N(z, 0, a)}\right)_{z=1}. \quad (100)$$

Using Eq. (89) for $G_N(z, 0, a)$, we obtain

$$\bar{n}_a(l_0, l_0) = N, \quad (101)$$

in agreement with remarks following Eq. (44).

We have also evaluated the dispersion $\Delta n_a(l_0, l_0)$ in the mean recurrence time for our random walker on an N-ring with symmetric exponentially distributed stepping probabilities. We define

$$\Delta n_a(l_0, l_0) \equiv [\overline{n_a^2}(l_0, l_0) - \bar{n}_a^2(l_0, l_0)]^{\frac{1}{2}}. \quad (102)$$

The use of Eqs. (32)–(36) and of the identity $\overline{n^2} = \overline{n(n-1)} + \bar{n}$ leads to

$$\Delta^2 n_a(l_0, l_0) = \left(\frac{\partial^2}{\partial z^2}\frac{1}{G_N(z, 0, a)} - \frac{\partial}{\partial z}\frac{1}{G_N(z, 0, a)} - \bar{n}_a^2(l_0, l_0)\right)_{z=1}. \quad (103)$$

Using $G_N(z, 0, a)$ of Eq. (84), we then obtain

$$\Delta^2 n_a(l_0, l_0) = \frac{N^3}{3}\frac{(1-e^{-a})^2}{(1+e^{-a})} - N^2\frac{(1-2e^{-a})}{(1+e^{-a})} + \frac{N}{3}\frac{(2-4e^{-a}-e^{-2a})}{(1+e^{-a})}. \quad (104)$$

In the nearest neighbor random walk limit $a \to \infty$, with N fixed, we find

$$\lim_{a\to\infty}\frac{\Delta n_a(l_0, l_0)}{\bar{n}_a(l_0, l_0)} \to \left(\frac{N(N-1)(N-2)}{3N^2}\right)^{\frac{1}{2}}, \quad (105)$$

in agreement with the nearest neighbor random walk result.[1] In the limit as N becomes very large, this reduces to

$$\lim_{\substack{a\to\infty \\ N\to\text{large}}}\frac{\Delta n_a(l_0, l_0)}{\bar{n}_a(l_0, l_0)} \to \left(\frac{N}{3}\right)^{\frac{1}{2}}. \quad (106)$$

In the limit $a \to 0$ corresponding to equal probabilities for jumps of any length and for fixed N we find

$$\lim_{a\to 0}\frac{\Delta n_a(l_0, l_0)}{\bar{n}_a(l_0, l_0)} \to \left((1-a) - \frac{1}{N}(1-a)\right)^{\frac{1}{2}}, \quad (107)$$

which, in the limit as $N \to \infty$, reduces to

$$\lim_{\substack{a\to 0 \\ N\to\infty}}\frac{\Delta n_a(l_0, l_0)}{\bar{n}_a(l_0, l_0)} \to (1-a)^{\frac{1}{2}}. \quad (108)$$

These results show that the maximum dispersion in the mean recurrence time occurs for the nearest neighbor random walk and that the dispersion decreases, though not monotonically, as the step length increases, i.e., as $a \to 0$.

C. Excursion from the Origin

To consider the excursion of the random walker with exponentially distributed step lengths from his point of origin after n steps, we use the equations already given in Sec. IIC:

$$\langle l^m\rangle_{\exp} \equiv \sum_{l=-\infty}^{\infty} l^m P_n(l) = (-i)^m \left(\frac{\partial^m}{\partial \phi^m}\,\lambda^n(\phi)\right)_{\phi=0}, \tag{109}$$

where the subscript exp denotes the exponentially distributed step length. The quantity $\lambda(\phi)$ for this case is obtained from Eq. (66) as

$$\lambda^n(\phi) = \left(\frac{c_1}{e^{(a-i\phi)}-1} + \frac{c_2}{e^{(b+i\phi)}-1}\right)^n, \tag{110}$$

where [see Eq. (63)]

$$\frac{c_1}{e^a-1} + \frac{c_2}{e^b-1} = 1, \tag{111}$$

and we find for $\langle l\rangle$, the mean distance from the origin after n steps,

$$\langle l\rangle_{\exp} \equiv \sum_{l=-\infty}^{\infty} l P_n(l) = n\left(\frac{c_1 e^a}{(e^a-1)^2} - \frac{c_2 e^b}{(e^b-1)^2}\right). \tag{112}$$

The corresponding result for a nearest neighbor random walk,[1] with $p(1) = p$, $p(-1) = q = 1 - p$, and $p(l_i - l_j) = 0$ otherwise, is

$$\langle l\rangle_1 = n(p - q). \tag{113}$$

In the exponential random walk, the total probability for a walker to step to his right on a given step is $c_1/(e^a - 1)$; the total probability that he steps to his left is $c_2/(e^b - 1)$. Making the correspondence

$$\frac{c_1}{(e^a-1)} \leftrightarrow p, \quad \frac{c_2}{(e^b-1)} \leftrightarrow q \tag{114}$$

between the exponential and nearest neighbor random walks, we can rewrite Eq. (112) as

$$\langle l\rangle_{\exp} = n\left(p\,\frac{e^a}{(e^a-1)} - q\,\frac{e^b}{(e^b-1)}\right). \tag{115}$$

In the limit $a \to \infty$, $b \to \infty$ the results for the "exponential" random walk thus again approach the nearest neighbor random walk results. In the limit as $a, b \ll 1$, Eq. (115) becomes

$$\langle l\rangle_{\exp} = n\left(p\,\frac{(1+a)}{a} - q\,\frac{(1+b)}{b}\right). \tag{116}$$

If, for the sake of simplicity, we now take $a = b$, Eq. (116) reduces to

$$\lim_{a=b\ll 1} \langle l\rangle_{\exp} = n(1 + a/a)(p - q) \simeq n/a(p - q). \tag{117}$$

A comparison of Eqs. (117) and (113) shows that the mean excursion from the origin for a random walker with exponentially distributed step lengths is larger than that for a nearest neighbor random walk by the factor $1/a$ for small a. It is evident that we are again dealing with order of magnitude differences between two random walks.

From Eq. (109) we obtain the dispersion in the excursion from the origin for the exponential random walk:

$$\langle l^2\rangle_{\exp} - \langle l\rangle^2_{\exp} = n\left[\frac{c_1 e^a(e^a+1)}{(e^a-1)^3} + \frac{c e_2^{\,b}(e^b+1)}{(e^b-1)^3} - \left(\frac{c_1 e^a}{(e^a-1)^2} - \frac{c_2 e^b}{(e^b-1)^2}\right)\right]. \tag{118}$$

For a symmetric random walk Eq. (118) reduces to

$$\langle l^2\rangle_{\exp} - \langle l\rangle^2_{\exp} = \frac{n(1+e^{-a})}{(1-e^{-a})^2}. \tag{119}$$

When $a \to \infty$, this result approaches the nearest neighbor random walk result $\langle l^2\rangle_1 - \langle l\rangle_1^2 = n$ of Eqs. (57), (58). As a becomes very small, i.e., $a \ll 1$, the dispersion increases with a as

$$\lim_{a\to 0} \langle l^2\rangle_{\exp} - \langle l\rangle^2_{\exp} \to n\left(\frac{2-a}{a^2}\right), \tag{120}$$

and becomes infinite for $a = 0$ when the walker can reach any point on the lattice with equal probability.

IV. CONCLUSIONS

We discuss below some interesting consequences and some as yet unsolved problems which arise from the results as presented in Sec. III.

We found [Eq. (101)] that for a random walker with exponentially distributed step lengths the mean first passage time to reach a particular lattice site on a one-dimensional N-ring as $a \to 0$ [$p(k) = 1/N$] is given by

$$\lim_{a\to 0} {}_1\langle n\rangle_a = N. \tag{121}$$

It is easy to calculate the generating function for the above random walk with $a = 0$ on a D-dimensional lattice. One then finds that *for N fixed and finite the result (121) is independent of the dimensionality of the lattice.* One can convince oneself, by "topological" reasoning, that this result must be correct. Since the random walker can go with equal probability $1/N$ to *any* point on the D-dimensional lattice, i.e., *each* point on a D-dimensional lattice of N lattice points is equally accessible in one step independent of D, the dimensionality cannot influence the result for ${}_1\langle n\rangle_{a=0}$ for fixed N. This is in distinction to the nearest

neighbor random walk, where the number of lattice sites which are accessible on each step increases sharply with dimensionality. It is therefore clear that appropriately chosen properties of 3-D random walks with exponentially distributed jump probabilities can, *in the limit as* $a = 0$, be calculated from the 1-D model.

Montroll[1,2] has evaluated the mean first passage time to a particular lattice site in one-, two-, and three-dimensional lattices with periodic boundary conditions for nearest neighbor random walks ($a \to \infty$). He finds the following leading terms for large N:

$$ {}_1\langle n\rangle_{\substack{a\to\infty \\ 1\text{-D}}} = k_1 N^2 + O(N), \tag{122} $$

$$ {}_1\langle n\rangle_{\substack{a\to\infty \\ 2\text{-D}}} = k_2 N \log N + O(N), \tag{123} $$

$$ {}_1\langle n\rangle_{\substack{a\to\infty \\ 3\text{-D}}} = k_3 N + O(1), \tag{124} $$

where the k's are constants which depend on the details of the lattice structure. It is very interesting to note that there is agreement, to within a constant of $O(1)$, between the very long step result [$a = 0$, $p(k) = 1/N$] in one dimension, Eq. (121), and the nearest neighbor transition result in three dimensions [Eq. (124)] for the mean first passage time ${}_1\langle n\rangle$. We believe that the comparison of other, appropriately chosen properties of 1-D random walks with exponentially distributed step length and $a = 0$, with 3-D nearest neighbor random walk results, will show analogous agreement to within $O(1)$.

The above result appears to have a close connection with recent theoretical work on phase transitions in 1-D systems with an infinite range of forces.[9] It was found in these studies that phase transitions which occurred in real 3-D systems presumably through pair interactions (i.e., nearest neighbor interactions), such as gas–liquid critical points (van der Waal's condensation) and magnetic and other order–disorder transitions in solids, could be obtained mathematically in 1-D systems in the limit of infinite range forces. One-dimensional models with finite and certainly with short-range forces do not show any phase transitions. Speaking generally, and very loosely, it would appear that the "connectivity" of long range 1-D models is of the same order as the "connectivity" of short range 3-D systems and that 3-D properties which depend strongly on this "connectivity" can be modeled by 1-D systems with the appropriate long-range interactions (step length, forces, etc.).

The above results are displayed schematically in Figs. 5(a) and 5(b). A study of these figures immediately raises the following questions:

(1) What is the form of ${}_1\langle n\rangle_a^{3\text{-D}}$ and of analogous properties (moments) of random walks, as a function of a, $0 < a < \infty$, in three dimensions? We know the results for $a = 0$ and for $a = \infty$ as shown in Fig. 5(a), but we do not know as yet how to connect the two extremum points. If it should turn out that ${}_1\langle n\rangle_a^{3\text{-D}}$ is essentially independent of a, i.e., of the step length, it would have important implications in the random walk interpretation of a number of experimental observations in physics and chemistry.

(2) What is the form of ${}_1\langle n\rangle_a$ and analogous moments in higher dimensions, $D > 3$, as a function of a and D for $a \neq 0$? In other words, how does one fill in the empty expanse of Fig. 5(b)? We note from Fig. 5(b) that for nearest neighbor random walks ${}_1\langle n\rangle$ decreases with increasing dimension for $D = 1, 2, 3$. How does it behave for $D > 3$? If it were to decrease with higher D, one would have the very surprising and unlikely result that, for $D > 3$, the mean first passage time for absorption is shorter for nearest neighbor walks than for walks with steps of arbitrary length. It is also possible that ${}_1\langle n\rangle_{\text{n.n.}}$ becomes independent of D for $D > 3$ and remains of $O(N)$. If this were the case, higher-dimensional results for nearest neighbor walks could then be deduced from the 3-D results or the 1-D results of random walks with exponentially distributed step length with $a = 0$. This question is of importance in the use of random walk models in the study of the unimolecular decay of polyatomic molecules, the study of the rate of formation of polymers from monomers and for many other problems.

The generating functions derived in Sec. II and III allow one to evaluate many quantities in addition to the ones we have calculated here. Other typical quantities of interest are, for example, the number of distinct lattice sites visited after n steps, random walk properties in the presence of reflecting boundaries, random walks in the presence of reflecting boundaries and traps, etc. We will direct ourselves to the questions posed above and to the general dependence of random walk results on the "connectivity" of the system in subsequent papers.

ACKNOWLEDGMENTS

The authors would like to thank Dr. D. Bedeaux for many helpful and stimulating discussions. We also wish to thank Miss K. Champion for her skill and devotion in the preparation of this manuscript.

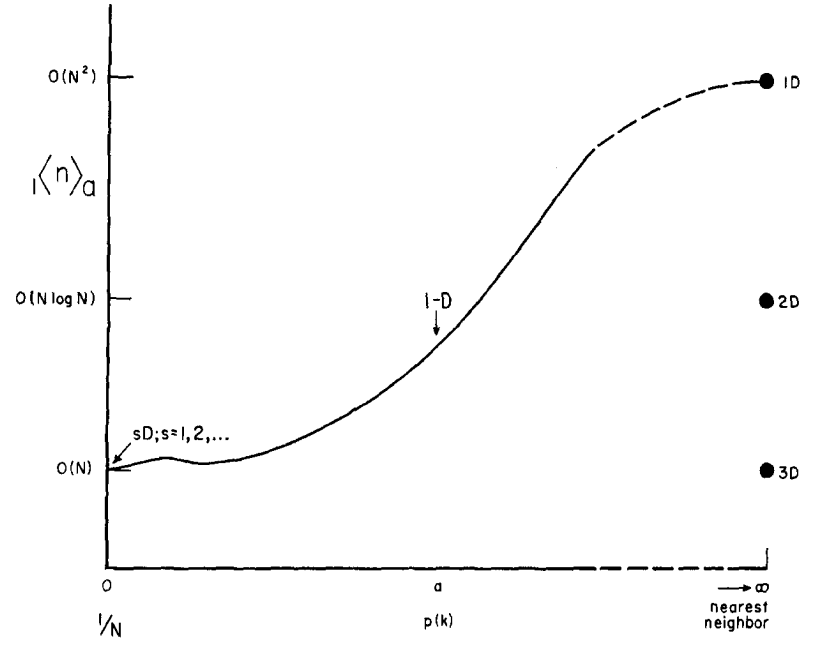

FIG. 5(a). Schematic of the dependence of the mean first passage time to a particular lattice site on a (step length) in various dimensions.

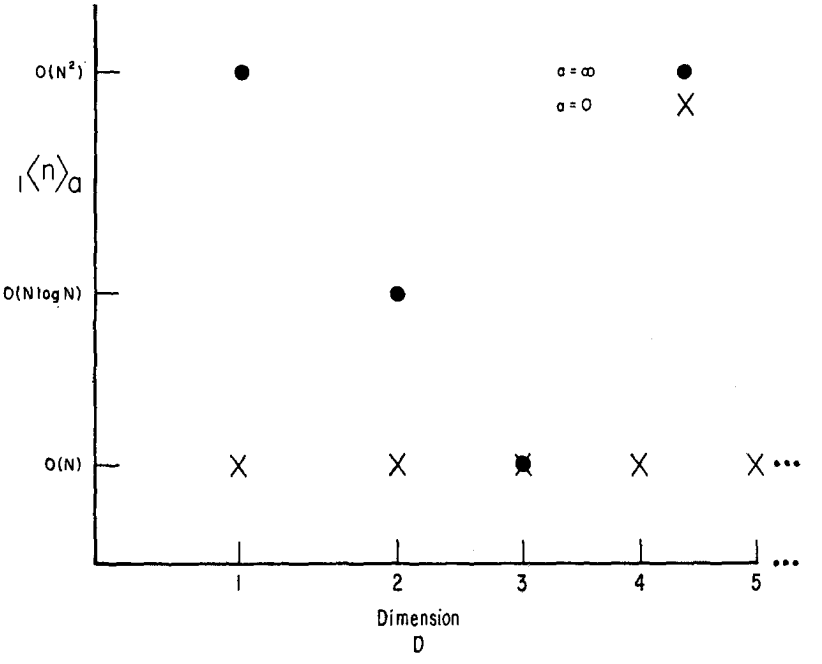

FIG. 5(b). Schematic of the dependence of the mean first passage time to a particular lattice site on dimension for nearest neighbor random walks ($a = \infty$) and for random walks with $p(k) = 1/N$, $a = 0$.

APPENDIX A: EVALUATION OF GENERATING FUNCTIONS

1. Symmetric Next-Nearest Neighbor Random Walk

The generating function for symmetric next-nearest neighbor random walks can be evaluated by complex contour integration. We rewrite Eq. (8) as follows:

$$G(z, l, \alpha) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-il\phi}\, d\phi}{1 - (z/2)[(1-\alpha)(e^{i\phi} + e^{-i\phi}) + \alpha(e^{2i\phi} + e^{-2i\phi})]}. \tag{A1}$$

In a symmetric one-dimensional random walk, all results can only depend on the absolute distance $|l|$ of the walker from his point of origin. We therefore perform the integration (A1) only for $l \geq 0$. With the change of variables

$$s = e^{-i\phi} \tag{A2}$$

(the corresponding change of variables for $l \leq 0$ is $s = e^{+i\phi}$), Eq. (A1) becomes

$$G(z, l, \alpha) = \frac{i}{\pi z \alpha} \oint \frac{s^{l+1}\, ds}{\{s^4 + [(1-\alpha)/\alpha]s^3 - (2s^2/z\alpha) + [(1-\alpha)/\alpha]s + 1\}}, \tag{A3}$$

where the integral is over the unit circle. The zeros of the quartic polynomial in the denominator of the integral are easily found because of the symmetry of the polynomial coefficients: If s_i is a zero of the polynomial, then so is $1/s_i$. The factorization of the quartic polynomial is then reduced to the solution of two quadratic equations in two unknowns. We write

$$\begin{aligned} &s^4 + [(1-\alpha)/\alpha]s^3 - (2s^2/z\alpha) + [(1-\alpha)/\alpha]s + 1 \\ &\quad = (s - s_1)(s - s_2)(s - 1/s_1)(s - 1/s_2) \\ &\quad = s^4 + 1 - (s^3 + s)(s_1 + s_2 + 1/s_1 + 1/s_2) \\ &\qquad + s^2(s_1 s_2 + s_1/s_2 + s_2/s_1 + 1/s_1 s_2 + 2). \end{aligned} \tag{A4}$$

Equating coefficients of powers of s gives the equations

$$\begin{aligned} s_1 + s_2 + 1/s_1 + 1/s_2 &= -(1-\alpha)/\alpha, \\ s_1 s_2 + s_1/s_2 + s_2/s_1 + 1/s_1 s_2 + 2 &= -2/z\alpha. \end{aligned} \tag{A5}$$

Solving these two quadratic equations in two unknowns yields the results of Eq. (10) with $s_3 = 1/s_2$ and $s_4 = 1/s_1$.

To determine which of the poles are within the unit circle, we examine them in detail only in the neighborhood of $z = 1$ since in Sec. II we only use the generating function for values of z close to 1. Expanding the s_i in powers of $(1 - z)$, we find

$$\begin{aligned} s_1 \sim\ & -(1/2\alpha)\{1 + \alpha + [(1+3\alpha)(1-\alpha)]^{\frac{1}{2}}\} \\ & + O((1-z)/(1-\alpha)) > 1 \\ & \text{for}\quad 0 \leq \alpha < 1, \quad (1-z)/(1-\alpha) \ll 1. \end{aligned} \tag{A6}$$

Hence s_1 is outside the unit circle for all values of α in the range of interest for z close to 1. Since $s_4 = 1/s_1$, s_4 is then within the unit circle. For s_2 we have

$$s_2 \sim 1 + [2(1-z)/(1+3\alpha)]^{\frac{1}{2}} > 1. \tag{A7}$$

Hence s_2 is outside the unit circle for all values of α. Since $s_3 = 1/s_2$, s_3 is within the unit circle. Evaluating the residues at the poles s_3 and s_4 then directly yields Eq. (9) for $G(z, l, \alpha)$.

2. Random Walks with Exponentially Distributed Step Lengths

The generating function given by Eq. (67) can be evaluated by complex contour integration. By simple algebraic manipulation we can write

$$G(z, l, a, b, c_1, c_2) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{(e^a e^{-i\phi} - 1)(e^b e^{i\phi} - 1)e^{-il\phi}}{(e^a e^{-i\phi} - 1)(e^b e^{i\phi} - 1) - zc_1(e^b e^{i\phi} - 1) - zc_2(e^a e^{-i\phi} - 1)}\, d\phi. \tag{A8}$$

For $l \geq 0$, we make the change of variables

$$s = e^{-i\phi}, \tag{A9}$$

and Eq. (A8) becomes

$$G(z, l \geq 0, a, b, c_1, c_2) = \frac{i}{2\pi} \oint \frac{s^l(e^{a+b} - se^a - s^{-1}e^b + 1)}{e^a(1 + zc_2)s^2 - (e^{a+b} + 1 + zc_1 + zc_2)s + e^b(1 + zc_1)}\, ds, \tag{A10}$$

where the integral is over the unit circle. Solving the simple quadratic polynomial in the denominator of the integrand, we then have

$$\begin{aligned} G(z, l \geq 0, a, b, c_1, c_2) = & \frac{i(e^{a+b} + 1)}{2\pi e^a(1 + zc_2)} \oint \frac{s^l\, ds}{(s - s_+)(s - s_-)} \\ & - \frac{ie^a}{2\pi e^a(1 + zc_2)} \oint \frac{s^{l+1}\, ds}{(s - s_+)(s - s_-)} - \frac{ie^b}{2\pi e^a(1 + zc_2)} \oint \frac{s^{l-1}\, ds}{(s - s_+)(s - s_-)}, \end{aligned} \tag{A11}$$

where

$$s_\pm = \frac{e^{a+b} + 1 + zc_1 + zc_2 \pm [(e^{a+b} + 1 + zc_1 + zc_2)^2 - 4e^{a+b}(1 + zc_1)(1 + zc_2)]^{\frac{1}{2}}}{2e^a(1 + zc_2)}. \tag{A12}$$

For $l \leq 0$, we make the change of variables

$$s = e^{+i\phi} \tag{A13}$$

in Eq. (A8). The result is similar to Eq. (A11), with the interchanges $a \leftrightarrow b$, $c_1 \leftrightarrow c_2$, and $l \leftrightarrow -l$:

$$G(z, l \leq 0, a, b, c_1, c_2) = \frac{i(e^{a+b}+1)}{2\pi e^b(1+zc_1)} \oint \frac{s^{-l}\,ds}{(s-r_+)(s-r_-)} - \frac{ie^b}{2\pi e^b(1+zc_1)} \oint \frac{s^{-l+1}\,ds}{(s-r_+)(s-r_-)} - \frac{ie^a}{2\pi e^b(1+zc_1)} \oint \frac{s^{-l-1}\,ds}{(s-r_+)(s-r_-)}, \quad \text{(A14)}$$

where

$$r_\pm = \frac{e^{a+b}+1+zc_1+zc_2 \pm [(e^{a+b}+1+zc_1+zc_2)^2 - 4e^{a+b}(1+zc_1)(1+zc_2)]^{\frac{1}{2}}}{2e^b(1+zc_1)}. \quad \text{(A15)}$$

One can determine algebraically which of the poles s_+, s_-, r_+, and r_- are within the unit circle, but it is simpler to do it on physical grounds using the following relations between the roots:

$$s_+ r_- = s_- r_+ = 1. \quad \text{(A16)}$$

If both s_+ and s_- were outside of the unit circle, then Eq. (A11) would yield $G(z, l > 0) = 0$ and the walker could never walk to the right of his point of origin, so that this situation is immediately ruled out for all positive finite values of the parameters a, b, c_1, and c_2. Both s_+ and s_- cannot be within the unit circle because if this were so, then, from Eq. (A16), r_+ and r_- would both be outside of the unit circle. This would yield $G(z, l < 0) = 0$ and the walker could never reach any point to the left of his point of origin. We thus conclude that one of s_+ and s_- is inside the unit circle and the other is outside, and similarly for r_+ and r_-.

It can be seen from Eq. (A12) that $s_+ > s_-$ for all values of the parameters a, b, c_1, and c_2 with the relation between the parameters given by Eq. (63). We write

$$s_+ - s_- = \frac{[(e^{a+b}+1+zc_1+zc_2)^2 - 4e^{a+b}(1+zc_1)(1+zc_2)]^{\frac{1}{2}}}{e^a(1+zc_2)}. \quad \text{(A17)}$$

The denominator in this expression is always positive; the quantity inside the radical is positive for all values of z of interest as can be seen by rewriting it as follows:

$$(e^{a+b}+1+zc_1+zc_2)^2 - 4e^{a+b}(1+zc_1)(1+zc_2) = \left([e^a(e^b-z)-(1-z)] - \frac{zc_2}{e^b-1}[2e^{a+b}-e^a-e^b]\right)^2 + \frac{4c_2 z(1-z)e^b(e^a-1)(e^{a+b}-1)}{e^b-1} \geq 0 \quad \text{for} \quad 0 \leq z \leq 1, \quad \text{(A18)}$$

where we have used Eq. (63) to eliminate c_1. Hence we conclude that the poles s_- and consequently r_- [see Eq. (A16)] are always the ones within the unit circle. Incidentally, we have also shown that s_+, s_-, r_+, and r_- are always real. The integrals in Eqs. (A11) and (A14) can now be easily evaluated. For $l > 0$, we need only the residue of each integral at $s = s_-$; for $l < 0$ we need only the residues at $s = r_-$; at $l = 0$, an additional pole within the unit circle appears at $s = 0$ in the third integral of both Eqs. (A11) and (A14), and so an additional residue must be added. Equation (68) for the generating function then immediately follows, with $s_- \equiv y_1$ and $s_+ \equiv y_2$.

APPENDIX B: GENERATING FUNCTIONS IN THE PRESENCE OF ABSORBING BOUNDARIES

1. Symmetric Next-Nearest Neighbor Random Walks-Absorbing Boundary

The generating function for a random walk in the presence of absorbing boundaries can be related to the generating function for a random walk on a perfect lattice [Eqs. (9) and (13)] via the difference equations satisfied by the $P_n(l)$ and the $Q_n(l)$. Consider a perfect one-dimensional infinite chain (without traps) and a random walker who begins his walk at $l = l_s$. The probability $P_n(l - l_s)$ that the walker

is at site l after n steps satisfies the difference equation [see Fig. 2(a)]

$$\begin{aligned} P_n(l-l_s) &= \tfrac{1}{2}\alpha P_{n-1}(l-l_s+2) \\ &\quad + \tfrac{1}{2}(1-\alpha)P_{n-1}(l-l_s+1) \\ &\quad + \tfrac{1}{2}(1-\alpha)P_{n-1}(l-l_s-1) \\ &\quad + \tfrac{1}{2}\alpha P_{n-1}(l-l_s-2), \end{aligned} \tag{B1}$$

with the initial condition

$$P_0(l-l_s) = \delta_{l,l_s}, \tag{B2}$$

and the normalization

$$\sum_{l=-\infty}^{\infty} P_n(l-l_s) = 1. \tag{B3}$$

Multiplying Eq. (B1) by z^n and summing from $n=1$ to $n=\infty$ then gives the difference equation for the generating function $G(z, l-l_s, \alpha)$:

$$\begin{aligned} &G(z,l-l_s,\alpha) - \tfrac{1}{2}\alpha z G(z,l-l_s+2,\alpha) \\ &\quad - \tfrac{1}{2}(1-\alpha)zG(z,l-l_s+1,\alpha) \\ &\quad - \tfrac{1}{2}(1-\alpha)zG(z,l-l_s-1,\alpha) \\ &\quad - \tfrac{1}{2}\alpha z G(z,l-l_s-2,\alpha) = \delta_{l,l_s} \end{aligned} \tag{B4}$$

or

$$\mathfrak{L}G(z,l-l_s,\alpha) = \delta_{l,l_s}, \tag{B5}$$

where Eq. (B4) defines the operator $\mathfrak{L}$. The generating function for walks on an N-ring satisfies the same difference equation:

$$\begin{aligned} &\mathfrak{L}G_N(z,l-l_s,\alpha) \\ &\quad \equiv G_N(z,l-l_s,\alpha) - \tfrac{1}{2}\alpha z G_N(z,l-l_s+2,\alpha) \\ &\quad - \tfrac{1}{2}(1-\alpha)zG_N(z,l-l_s+1,\alpha) \\ &\quad - \tfrac{1}{2}(1-\alpha)zG_N(z,l-l_s-1,\alpha) \\ &\quad - \tfrac{1}{2}\alpha z G_N(z,l-l_s-2,\alpha) = \delta_{l,l_s}. \end{aligned} \tag{B6}$$

Now consider a random walker starting at $l=l_0$ with trapping sites at $l=l_1$ and $l=l_1-1$. For the sake of simplicity in the equations that follow, we choose $l_1=0$ and include the obvious l_1 dependence only at the end of the calculation. The choice $l_0 > l_1$ does not affect the generality of the calculation. The probability $Q_n(l-l_0)$ that the walker reaches point l on the nth step in the presence of the trapping sites satisfies the following set of difference equations [see Fig. 2(b)]:

$$\begin{aligned} Q_n(l-l_0) &= \tfrac{1}{2}\alpha Q_{n-1}(l-l_0+2) \\ &\quad + \tfrac{1}{2}(1-\alpha)Q_{n-1}(l-l_0+1) \\ &\quad + \tfrac{1}{2}(1-\alpha)Q_{n-1}(l-l_0-1) \\ &\quad + \tfrac{1}{2}\alpha Q_{n-1}(l-l_0-2), \quad l \geq 3, \end{aligned} \tag{B7}$$

and, since the walker cannot escape from sites $l=0$ and $l=-1$,

$$\begin{aligned} Q_n(2-l_0) &= \tfrac{1}{2}\alpha Q_{n-1}(4-l_0) \\ &\quad + \tfrac{1}{2}(1-\alpha)Q_{n-1}(3-l_0) \\ &\quad + \tfrac{1}{2}(1-\alpha)Q_{n-1}(1-l_0), \\ Q_n(1-l_0) &= \tfrac{1}{2}\alpha Q_{n-1}(3-l_0) \\ &\quad + \tfrac{1}{2}(1-\alpha)Q_{n-1}(2-l_0), \\ Q_n(-l_0) &= \tfrac{1}{2}\alpha Q_{n-1}(2-l_0) \\ &\quad + \tfrac{1}{2}(1-\alpha)Q_{n-1}(1-l_0) \\ &\quad + Q_{n-1}(-l_0), \\ Q_n(-1-l_0) &= \tfrac{1}{2}\alpha Q_{n-1}(1-l_0) + Q_{n-1}(-1-l_0), \\ Q_n(l-l_0) &= 0, \quad l \leq -2. \end{aligned} \tag{B8}$$

The initial condition and normalization are

$$Q_0(l-l_0) = \delta_{l,l_0}, \tag{B9}$$

$$\sum_l Q_n(l-l_0) = 1. \tag{B10}$$

Multiplying Eqs. (B7) and (B8) by z^n and summing from $n=1$ to $n=\infty$ yields the following difference equations for the generating function $F(z, l-l_0, \alpha)$ defined in Eq. (15):

$$\mathfrak{L}F(z,l-l_0,\alpha) = H(z,l-l_0,\alpha). \tag{B11}$$

The operator $\mathfrak{L}$ is defined in Eq. (B4) and (B5), and

$$\begin{aligned} &H(z,l-l_0,\alpha) \\ &\quad = \delta_{l,l_0} - F(z,-l_0,\alpha) \\ &\qquad \times [\tfrac{1}{2}\alpha z(\delta_{l,2}+\delta_{l,-2}) \\ &\qquad + \tfrac{1}{2}(1-\alpha)z(\delta_{l,1}+\delta_{l,-1}) - z\delta_{l,0}] \\ &\qquad - F(z,-1-l_0,\alpha)[\tfrac{1}{2}\alpha z(\delta_{l,1}+\delta_{l,-3}) \\ &\qquad + \tfrac{1}{2}(1-\alpha)z(\delta_{l,0}+\delta_{l,-2}) - z\delta_{l,-1}]. \end{aligned} \tag{B12}$$

Comparison of Eqs. (B5) and (B11) shows that $G(z, l-l_s, \alpha)$ is the Green's function required for the solution of the inhomogeneous equation satisfied by $F(z, l-l_0, \alpha)$:

$$F(z,l-l_0,\alpha) = \sum_{l'} G(z,l-l',\alpha)H(z,l'-l_0,\alpha). \tag{B13}$$

Substituting Eq. (B12) into Eq. (B13) yields

$$\begin{aligned} &F(z,l-l_0,\alpha) \\ &\quad = G(z,l-l_0,\alpha) - (z/2)F(z,-l_0,\alpha) \\ &\qquad \times [\alpha G(z,l+2,\alpha) + (1-\alpha)G(z,l+1,\alpha) \\ &\qquad + (1-\alpha)G(z,l-1,\alpha) + \alpha G(z,l-2,\alpha) \\ &\qquad - 2G(z,l,\alpha)] - (z/2)F(z,-1-l_0,\alpha) \\ &\qquad \times [G(z,l+3,\alpha) + (1-\alpha)G(z,l+2,\alpha) \\ &\qquad + (1-\alpha)G(z,l,\alpha) + \alpha G(z,l-1,\alpha) \\ &\qquad - 2G(z,l+1,\alpha)]. \end{aligned} \tag{B14}$$

Via Eq. (B4), this expression simplifies further:

$$F(z, l - l_0, \alpha) = G(z, l - l_0, \alpha) - F(z, -l_0, \alpha) \times [(1 - z)G(z, l, \alpha) - \delta_{l,0}] - F(z, -1 - l_0, \alpha)[(1 - z)G(z, l + 1, \alpha) - \delta_{l,-1}]. \tag{B15}$$

Letting $l = 0$ and $l = -1$ in Eq. (B15), solving the two resulting equations for $F(z, -l_0, \alpha)$ and $F(z, -1 - l_0, \alpha)$ in terms of the $G(z, l, \alpha)$, and substituting back into Eq. (B15) finally yields

$$\begin{aligned} F(z, l - l_0, \alpha) &= G(z, l - l_0, \alpha) \\ &+ \frac{[G(z, l_0, \alpha)G(z, 0, \alpha) - G(z, l_0 + 1, \alpha)G(z, 1, \alpha)]}{[G^2(z, 1, \alpha) - G^2(z, 0, \alpha)]} \\ &\times \left(G(z, l, \alpha) - \frac{(\delta_{l,0})}{(1 - z)}\right) \\ &- \frac{[G(z, l_0, \alpha)G(z, 1, \alpha) - G(z, l_0 + 1, \alpha)G(z, 0, \alpha)]}{[G^2(z, 1, \alpha) - G^2(z, 0, \alpha)]} \\ &\times \left(G(z, l + 1, \alpha) - \frac{(\delta_{l,-1})}{(1 - z)}\right). \end{aligned} \tag{B16}$$

In particular, for the first passage time evaluated in Eq. (20) we need

$$F(z, -l_0, \alpha) + F(z, -1 - l_0, \alpha) = \frac{[G(z, -l_0, \alpha) + G(z, -1 - l_0, \alpha)]}{[1 - z][G(z, 0, \alpha) + G(z, 1, \alpha)]}. \tag{B17}$$

Equation (21) immediately follows for an N-ring with traps located at $l = l_1$ and $l = l_1 - 1$, with l_1 arbitrary.

2. Symmetric Random Walks with Exponentially Distributed Step Lengths-Absorbing Boundary

Consider a perfect one-dimensional infinite chain (without traps) and a random walker who begins his walk at $l = l_s$. The probability $P_n(l - l_s)$ that the walker is at site l after n steps satisfies the difference equation

$$P_n(l - l_s) = \tfrac{1}{2}(e^a - 1) \sum_{l'=-\infty}^{\infty}{}' e^{-|l-l'|a} P_{n-1}(l' - l_s), \tag{B18}$$

where the prime on the sum indicates the omission of the $l' = l$ term. The $P_n(l - l_s)$ satisfy an initial condition and a normalization condition given by

$$P_0(l - l_s) = \delta_{l,l_s}, \tag{B19}$$

$$\sum_{l=-\infty}^{\infty} P_n(l - l_s) = 1. \tag{B20}$$

Multiplying Eq. (B18) by z^n and summing from $n = 1$ to ∞ then gives the difference equation for the generating function $G(z, l - l_s, a)$:

$$G(z, l - l_s, a) - (z/2)(e^a - 1) \times \sum_{l'=-\infty}^{\infty}{}' e^{-|l-l'|a} G(z, l' - l_s, a) = \delta_{l,l_s}, \tag{B21}$$

or

$$\mathfrak{L}G(z, l - l_s, a) = \delta_{l,l_s}, \tag{B22}$$

where Eq. (B21) defines the operator $\mathfrak{L}$. Next, consider a random walk on an infinite chain with trapping sites at $l = l_1 - k$ and $l = N + l_1 + k$ for all $k \geq 0$. For the sake of simplicity in the equations that follow, we choose $l_1 = 0$ and include the obvious l_1 dependence at the end of the calculation. The walker starts his walk at $l = l_0$, with $1 \leq l_0 \leq N - 1$. The probability $Q_n(l - l_0)$ that the walker reach point l on the nth step in the presence of the trapping sites satisfies the following difference equation:

$$\begin{aligned} Q_n(l - l_0) &= \tfrac{1}{2}(e^a - 1) \sum_{l'=1}^{N-1} e^{-|l-l'|a} Q_{n-1}(l' - l_0), \\ &\qquad 1 \leq l \leq N - 1, \\ &= \tfrac{1}{2}(e^a - 1) \sum_{l'=1}^{N-1} e^{-|l-l'|a} Q_{n-1}(l' - l_0) \\ &\quad + Q_{n-1}(l - l_0), \quad l \leq 0 \text{ and } l \geq N, \end{aligned} \tag{B23}$$

where we have accounted for the fact that the walker cannot escape from the trapping sites. The initial and normalization conditions are

$$Q_0(l - l_0) = \delta_{l,l_0}, \tag{B24}$$

$$\sum_{l=-\infty}^{\infty} Q_n(l - l_0) = 1. \tag{B25}$$

Multiplying Eq. (B23) by z^n and summing from $n = 1$ to $n = \infty$ yields the following difference equation for the generating function $F(z, l - l_0, a)$ defined in Eq. (86):

$$\mathfrak{L}F(z, l - l_0, a) = H(z, l - l_0, a), \tag{B26}$$

where the operator $\mathfrak{L}$ is defined in Eqs. (B21) and (B22) and

$$\begin{aligned} H(z, l - l_0, a) &= \delta_{l,l_0} - (z/2)(e^a - 1) \sum_{l'=-\infty}^{0} e^{-|l-l'|a} F(z, l' - l_0, a) \\ &\quad - (z/2)(e^a - 1) \sum_{l'=N}^{\infty} e^{-|l-l'|a} F(z, l' - l_0, a), \\ &\qquad 1 \leq l \leq N - 1, \\ &= -(z/2)(e^a - 1) \sum_{l'=-\infty}^{0}{}' e^{-|l-l'|a} F(z, l' - l_0, a) \\ &\quad - (z/2)(e^a - 1) \sum_{l'=N}^{\infty}{}' e^{-|l-l'|a} F(z, l' - l_0, a) \\ &\quad + zF(z, l - l_0, a), \quad l \leq 0 \text{ and } l \geq N. \end{aligned} \tag{B27}$$

Equations (B22) and (B26) show that $G(z, l - l_s, a)$ is the Green's function required for the solution of the inhomogeneous equation satisfied by $F(z, l - l_0, a)$:

$$F(z, l - l_0, a) = \sum_{l'=-\infty}^{\infty} G(z, l - l', a)H(z, l' - l_0, a). \tag{B28}$$

Substituting Eq. (B27) into Eq. (B28) yields

$$\begin{aligned} F(z, l - l_0, a) &= G(z, l - l_0, a) \\ &\quad - (z/2)(e^a - 1) \sum_{\substack{k=-\infty \\ k \neq m}}^{\infty} \sum_{m=-\infty}^{0} G(z, l - k, a)e^{-|k-m|a} \\ &\quad \times F(z, m - l_0, a) \\ &\quad - (z/2)(e^a - 1) \sum_{\substack{k=-\infty \\ k \neq m}}^{\infty} \sum_{m=N}^{\infty} G(z, l - k, a)e^{-|k-m|a} \\ &\quad \times F(z, m - l_0, a) \\ &\quad + z \sum_{m=-\infty}^{0} G(z, l - m, a)F(z, m - l_0, a) \\ &\quad + z \sum_{m=N}^{\infty} G(z, l - m, a)F(z, m - l_0, a). \end{aligned} \tag{B29}$$

Equation (B21) allows us to eliminate the sums over the variable k, since

$$-(z/2)(e^a - 1) \sum_{k=-\infty}^{\infty}{}' G(z, l - k, a)e^{-|k-m|a} = \delta_{l,m} - G(z, l - m, a). \tag{B30}$$

Equation (B29) then becomes

$$\begin{aligned} F(z, l - l_0, a) &= G(z, l - l_0, a) \\ &\quad + \sum_{m=-\infty}^{0} F(z, m - l_0, a) \\ &\quad \times [(z - 1)G(z, l - m, a) + \delta_{l,m}] \\ &\quad + \sum_{m=N}^{\infty} F(z, m - l_0, a) \\ &\quad \times [(z - 1)G(z, l - m, a) + \delta_{l,m}]. \end{aligned} \tag{B31}$$

We now have the choice to evaluate the sum of Eq. (B31) over all nontrapping sites [Eq. (88)] or over all trapping sites [Eq. (90)]. These sums are, of course, related via the conservation of walkers expressed by Eq. (89):

$$\sum_{l=-\infty}^{0} F(z, l - l_0, a) + \sum_{l=N}^{\infty} F(z, l - l_0, a) = 1/(1 - z) - \sum_{l=1}^{N-1} F(z, l - l_0, a). \tag{B32}$$

Although there are fewer nontrapping sites ($1 \leq l \leq N - 1$) than there are trapping sites ($l \leq 0$, $l \geq N$), it turns out to be computationally more convenient to work in terms of the latter.

If l is a trapping site, Eq. (B31) can be rewritten:

$$\begin{aligned} F(z, l - l_0, a)G(z, 0, a) &= G(z, l - l_0, a)/(1 - z) \\ &\quad - \sum_{\substack{m=-\infty \\ m \neq l}}^{0} G(z, l - m, a)F(z, m - l_0, a) \\ &\quad - \sum_{\substack{m=N \\ m \neq l}}^{\infty} G(z, l - m, a)F(z, m - l_0, a), \\ &\qquad l \leq 0 \text{ and } l \geq N. \end{aligned} \tag{B33}$$

Using Eq. (74) written in the form

$$G(z, l, a) = f(z, a)x^{|l|}, \quad |l| \geq 1, \tag{B34}$$

where

$$f(z, a) = \frac{z(e^{2a} - 1)}{[2 + z(e^a - 1)]D_0}, \tag{B35}$$

and where x and D_0 are defined in Eqs. (75) and (76), we can further rewrite Eq. (B33):

$$\begin{aligned} F(z, l - l_0, a)G(z, 0, a) &= [f(z, a)/(1 - z)]x^{|l-l_0|} \\ &\quad - f(z, a) \sum_{\substack{m=-\infty \\ l \neq m}}^{0} F(z, m - l_0 a)x^{|l-m|} \\ &\quad - f(z, a) \sum_{\substack{m=N \\ l \neq m}}^{\infty} F(z, m - l_0, a)x^{|l-m|}, \\ &\qquad l \leq 0 \text{ and } l \geq N. \end{aligned} \tag{B36}$$

This equation can now be summed over trapping sites l since the right side involves only trivial sums of the form $\sum_l x^{\pm l}$. The result is

$$\begin{aligned} &[G(z, 0, a) + 2f(z, a)x/(1 - x)] \\ &\quad \times \left(\sum_{l=-\infty}^{0} F(z, l - l_0, a) + \sum_{l=N}^{\infty} F(z, l - l_0, a) \right) \\ &= [f(z, a)/(1 - z)](x^{l_0} + x^{N-l_0})/(1 - x) \\ &\quad + f(z, a)[(1 - x^N)/(1 - x)] \\ &\quad \times [F(z, -l_0, a) + F(z, N - l_0, a)] \\ &\quad + f(z, a)[(x - x^N)/(1 - x)] \\ &\quad \times \sum_{m=-\infty}^{1} F(z, m - l_0, a)x^{-m} \\ &\quad + f(z, a)x^{-N}[(x - x^N)/(1 - x)] \\ &\quad \times \sum_{m=N+1}^{\infty} F(z, m - l_0, a)x^m. \end{aligned} \tag{B37}$$

The last two sums in this equation can be expressed in a simpler form by noting that Eq. (B36) gives these sums in terms of $F(z, -l_0, a)$ and $F(z, N - l_0, a)$, which are the generating functions for stepping on the first traps adjacent to either end of the nontrapping

interval:

$$G(z,0,a)F(z,-l_0,a) = [f(z,a)/(1-z)]x^{l_0} - f(z,a)\sum_{m=-\infty}^{1} F(z,m-l_0,a)x^{-m} - f(z,a)\sum_{m=N+1}^{\infty} F(z,m-l_0,a)x^{m} - f(z,a)x^{N}F(z,N-l_0,a), \quad \text{(B38)}$$

$$G(z,0,a)F(z,N-l_0,a) = [f(z,a)/(1-z)]x^{N-l_0} - f(z,a)x^{N}F(z,-l_0,a) - f(z,a)x^{N}\sum_{m=-\infty}^{1} F(z,m-l_0,a)x^{-m} - f(z,a)x^{-N}\sum_{m=N+1}^{\infty} F(z,m-l_0,a)x^{m}. \quad \text{(B39)}$$

Solving for

$$\sum_{m=-\infty}^{1} F(z,m-l_0,a)x^{-m}$$

and

$$\sum_{m=N+1}^{\infty} F(z,m-l_0,a)x^{m}$$

in terms of $F(z,-l_0,a)$ and $F(z,N-l_0,a)$ and substituting the result into Eq. (B37) yields

$$[G(z,0,a)+2f(z,a)x/(1-x)] \times \left(\sum_{l=-\infty}^{0} F(z,l-l_0,a)+\sum_{l=N}^{\infty} F(z,l-l_0,a)\right) = \frac{f(z,a)x^{l_0}+x^{N-l_0}(1+x)}{(1-z)(1-x)(1+x^N)} + \frac{(x-x^N)}{(1+x^N)(1-x)} \times [f(z,a)-G(z,0,a)] \times [F(z,-l_0,a)+F(z,N-l_0,a)]. \quad \text{(B40)}$$

It now only remains to eliminate $[F(z,-l_0,a)+F(z,N-l_0,a)]$ from the right side of Eq. (B40) since all other quantities on the right side are known. This can be done by noting that through Eqs. (B26) and (B27) we can get an independent relation between $[F(z,-l_0,a)+F(z,N-l_0,a)]$ and the sums appearing on the left of Eq. (B40); Eqs. (B26) and (B27) can be rewritten in the form

$$(1-z)F(z,l-l_0,a) = (z/2)(e^{a}-1)\sum_{l'=1}^{N-1} e^{-(l'-l)a}F(z,l'-l_0,a), \quad l \le 0,$$
$$= (z/2)(e^{a}-1)\sum_{l'=1}^{N-1} e^{-(l-l')a}F(z,l'-l_0,a), \quad l \ge N. \quad \text{(B41)}$$

Hence we have

$$(1-z)\sum_{l=-\infty}^{0} F(z,l-l_0,a) = \frac{z}{2}\frac{(1-e^{a})}{(1-e^{-a})}\sum_{l'=1}^{N-1} e^{-l'a}F(z,l'-l_0,a), \quad \text{(B42)}$$

$$(1-z)F(z,-l_0,a) = (z/2)(1-e^{a})\sum_{l'=1}^{N-1} e^{-l'a}F(z,l'-l_0,a), \quad \text{(B43)}$$

$$(1-z)\sum_{l=N}^{\infty} F(z,l-l_0,a) = \frac{z}{2}\frac{(1-e^{a})}{(1-e^{-a})}e^{-Na}\sum_{l'=1}^{N-1} e^{l'a}F(z,l'-l_0,a), \quad \text{(B44)}$$

$$(1-z)F(z,N-l_0,a) = (z/2)(1-e^{a})e^{-Na}\sum_{l'=1}^{N-1} e^{l'a}F(z,l'-l_0,a). \quad \text{(B45)}$$

Equations (B42)–(B45) directly give

$$[F(z,-l_0,a)+F(z,N-l_0,a)] = (1-e^{-a})\sum_{l'=-\infty}^{0} F(z,l-l_0,a)+\sum_{l'=N}^{\infty} F(z,l-l_0,a). \quad \text{(B46)}$$

Combining Eqs. (B40) and (B46) then finally yields

$$\sum_{l=-\infty}^{0} F(z,l-l_0,a)+\sum_{l=N}^{\infty} F(z,l-l_0,a) = \frac{f(z,a)(x^{l_0}+x^{N-l_0})(1+x)}{[1-z]\{[(1-xe^{-a})-x^{N}(x-e^{-a})]G(z,0,a)+[(1+e^{-a})+x^{N-1}(1+2x-e^{-a})]xf(z,a)\}} = \frac{[G(z,l_0,a)+G(z,N-l_0,a)](1+x)}{[1-z]\{[(1-xe^{-a})-x^{N}(x-e^{-a})]G(z,0,a)+[(1+e^{-a})+x^{N-1}(1+2x-e^{-a})]G(z,1,a)\}}. \quad \text{(B47)}$$

Equation (91) immediately follows upon defining the absorbing boundaries to begin at $l=l_1$ and $l=N-l_1$ for arbitrary l_1.

* This research was supported by the National Science Foundation under Grant GP 10536.

[1] E. W. Montroll, Proc. Symp. Appl. Math. **16**, 193 (1964).
[2] E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).
[3] E. W. Montroll, J. Phys. Soc. Japan Suppl. **26**, 6 (1969).
[4] E. W. Montroll, J. Math. Phys. **10**, 753 (1969).
[5] R. A. Elliott, K. Lakatos, and R. S. Knox, J. Stat. Phys. **1**, 253 (1969).
[6] In the nearest neighbor random walk on an N-ring, one trapping point only is required for absorption of the walker with probability 1.
[7] E. W. Montroll and K. E. Shuler, Advan. Chem. Phys. **1**, 361 (1958).
[8] K. E. Shuler and G. H. Weiss, J. Chem. Phys. **38**, 505 (1963).
[9] See, e.g., M. Kac, G. E. Uhlenbeck, P. C. Hemmer, J. Math. Phys. **4**, 216, 229 (1963); M. Kac and E. Helfand, *ibid.*, 1078 (1963); M. Kac, in *Statistical Physics, Phase Transitions and Superfluidity* (Gordon and Breach, New York, 1968), Vol. 1, p. 245 et seq.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Geometric Approach to Invariance Groups and Solution of Partial Differential Systems*

B. KENT HARRISON† AND FRANK B. ESTABROOK
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

(Received 16 July 1970)

Methods are discussed for discovery of physically or mathematically special families of exact solutions of systems of partial differential equations. Such systems are described geometrically using equivalent sets of differential forms, and the theory derived for obtaining the generators of their invariance groups—vector fields in the space of forms. These *isovectors* then lead naturally to all the special solutions discussed, and it appears that other special ansätze must similarly be capable of geometric description. Application is made to the one-dimensional heat equation, the vacuum Maxwell equations, the Korteweg–de Vries equation, one-dimensional compressible fluid dynamics, the Lambropoulos equation, and the cylindrically symmetric Einstein–Maxwell equations.

I. INTRODUCTION

Exact solutions of the sets of coupled nonlinear partial differential equations of physics (fluid physics, general relativity, etc.) are of great value for the physical insight they can give into extreme cases not susceptible to numerical or approximate treatment. The search for such solutions is, however, not a systematic process, but depends in practice on intuitive skill and physical or geometrical analogy for the divination of successful ansätze. In first attempt at a more systematic methodology, we have considered, as a *practical* process of applied mathematics, the use of the generators of the invariance group of a given set of equations. Our method for finding these appears to be new; it is based on Cartan's geometric formulation of partial differential equations in the language of exterior differential forms, and introduces the fields of *isovectors* which generate the invariance transformations. With these, we indeed find certain specializations of the set considered, which may be soluble exactly; in particular we refer to classes of so-called similarity solutions and to algebraically special solutions based on invariant subspaces. Other methods for systematic discovery of exact solutions must also have close connections to this underlying group structure.

In the following, we sketch the method and illustrate use of it by application to a number of simple examples, including the vacuum Maxwell equations, the one-dimensional heat equation, and some simple nonlinear cases from fluid dynamics; and also we relate it to "solution generation" procedures which were previously applied by one of the authors to a specialization of the general relativistic Einstein–Maxwell equations. We trust that its usefulness in other such nonlinear contexts will become evident.

The basic work on invariance groups of partial differential equations has been well established by its originators, especially Lie.[1] It has been used recently by Bluman and Cole[2] in discussing very general similarity solutions of the one-dimensional heat equation. We present here what we believe to be a more useful, intuitive, and concise formulation in the language of differential forms. Cartan's geometric theory of partial differential equations[3] has been expounded by Slebodzinski,[4] to whose terminology we remain close, and by Choquet–Bruhat[5]; a careful survey and bibliography of modern work and formulations is due to Hermann.[6] We extend their geometrical picture to emphasize the role of the *isovector fields*, the generators of geometric transformations with suitable algebraic invariance properties. These isovector fields generate the invariance group, or *isogroup*, and in turn lead to varieties of special solutions. It seems to us that such an approach to partial differential equations is so deep that all successful ansätze for special solutions must, in it, take their most transparent form, as algebraic criteria on geometric objects in general spaces.

II. EXTERIOR DIFFERENTIAL FORMS

We first recapitulate the notation and basic identities of the calculus of exterior differential forms. In a differentiable manifold of n dimensions, we regard a p-form ω *either* as an index-free notation for a p-vector (or completely antisymmetric covariant pth-rank tensor) $\omega_{\mu_1\cdots\mu_p}$ *or* as the value (or scalar magnitude) of this p-vector when contracted with a p-fold infinitesimal (simple, contravariant, completely antisymmetric) tensor of extension describing an elementary cell of a p-dimensional submanifold,

$$\omega_{\mu_1\cdots\mu_p}\, d\xi^{[\mu_1}_{1} \cdots d\xi^{\mu_p]}_{p}.^{7}$$

Modern expositions equivalently treat 1-forms as elements of the cotangent space at a point of the

manifold, i.e., as mappings from the tangent space into the reals.[5,6,8]

At each point of the manifold, the forms have a Grassmann algebra (with noncommutative antisymmetrized outer multiplication denoted by $\wedge$) over a 2^n-dimensional linear vector space. Over the manifold we may use the three operations of exterior differentiation, d, of contraction with a vector field V (a contravariant vector V^μ), $V \lrcorner \omega$, and of Lie derivative with respect to V, $\pounds_V \omega$. These give forms of rank $p+1$, $p-1$, and p, respectively. The following results are to be found in the references, or may easily be written from the equivalent operations of the tensor calculus.[7]

Let ω be a p-form, σ a q-form, f a 0-form, c a constant, and V and W vector fields. Then

$$\begin{aligned} \omega \wedge \sigma &= (-1)^{pq} \sigma \wedge \omega, \\ d(\omega \wedge \sigma) &= d\omega \wedge \sigma + (-1)^p \omega \wedge d\sigma, \\ ddw &= 0, \\ dc &= 0, \end{aligned} \tag{1}$$

$$\begin{aligned} (V \lrcorner + W \lrcorner)\omega &= V \lrcorner \omega + W \lrcorner \omega, \\ (fV) \lrcorner \omega &= f(V \lrcorner \omega), \\ V \lrcorner (\omega \wedge \sigma) &= (V \lrcorner \omega) \wedge \sigma \\ &\quad + (-1)^p \omega \wedge (V \lrcorner \sigma), \end{aligned} \tag{2}$$

$$\begin{aligned} \underset{V}{\pounds}\, \omega &= V \lrcorner d\omega + d(V \lrcorner \omega), \\ \underset{V}{\pounds}\, f &= V \lrcorner df, \\ \underset{V}{\pounds}\, d\omega &= d\left(\underset{V}{\pounds}\, \omega\right), \\ \underset{V}{\pounds}\, (\omega \wedge \sigma) &= \left(\underset{V}{\pounds}\, \omega\right) \wedge \sigma + \omega \wedge \left(\underset{V}{\pounds}\, \sigma\right), \\ \underset{V}{\pounds}\, (W \lrcorner \omega) &= [V, W] \lrcorner \omega + W \lrcorner \left(\underset{V}{\pounds}\, \omega\right). \end{aligned} \tag{3}$$

$[V, W]$ is the commutator, or Lie product, of these vectors: It is the contravariant vector field $\pounds_V W^\mu$. In the following, we will usually omit the exterior product operator $\wedge$.

III. CARTAN'S THEORY

Cartan[3–6] has discussed criteria for the equivalence of a given set of partial differential equations (in $n-p$ dependent variables and p independent variables) with a closed set of differential forms on an n-dimensional manifold. The latter set is the basis of a differential ideal of the Grassmann algebra of forms on the manifold. Any set *algebraically equivalent* to the set generates the same ideal, and could as well serve as the basis. An *integral manifold* of the ideal is a subspace, the extension elements of which *annul*—give zero values to—all forms in the ideal, or the *section*[7] of which with the ideal gives forms, in the subspace, identically zero.

In tensor notation, in an n-dimensional manifold spanned by x^μ, a p-dimensional subspace is described by n parametric equations $x^\mu = x^\mu(y^A)$, $A = 1, \cdots, p$. The y^A can be adopted as coordinates in the subspace. We can apply the *section operators* ${}_A\lambda^\mu = \partial x^\mu / \partial y^A$ to covariant quantities, for example, vectors, ϕ_μ, obtaining *sectioned* covariant quantities $\tilde{\phi}_A \equiv {}_A\lambda^\mu \phi_\mu$, etc., in the subspace, i.e., quantities which transform properly under transformations $y^A = y^A(\bar{y}^A)$. The sectioning of forms is equivalently defined. If sectioning a form ϕ gives a form $\tilde{\phi}$ which is identically zero, any extension element of the subspace must annul ϕ, and conversely.

If an integral manifold of dimensionality p exists, and if, in it, p of the variables may vary freely (i.e., may also be adopted as the coordinates in the manifold), it amounts to a functional specification of the other $n-p$ variables—the former are then called independent, and the latter dependent, variables. Such an integral manifold in the n-space can geometrically represent a *solution* of the original set of partial differential equations. In fact, denoting the $n-p$ variables by z^i, $i = 1, \cdots, n-p$, and the p variables as x^A, $A = 1, \cdots, p$, one must be able to reobtain the original partial differential equations by substituting, into the set of forms in n dimensions, the relations *sectioning* those forms into the p-manifold:

$$d\tilde{x}^A = dx^A, \quad d\tilde{z}^i = \frac{\partial z^i}{\partial x^A} dx^A, \tag{4}$$

and then requiring the coefficients of the forms dx^A, $dx^A \wedge dx^B$, etc., to vanish. By this procedure we impose the independence of the variables x^A on the set of forms.

The integral manifolds of a differential ideal are classified by Cartan as (1) general or (2) singular. The singular manifolds are poorly understood as a class, and must be investigated by ad hoc methods. The general manifolds are those which can be obtained by a step-by-step procedure of Cauchy–Kowalewski integrations, beginning with one-dimensional integral manifolds and giving a hierarchy, or chain, of such, of every dimensionality up to the maximum (or genus) g. In a beautiful, systematic, theoretical development, Cartan gives the local algebraic existence criteria for these general solutions; in particular, if, for a dimensionality $p \leq g$, these criteria show the existence of general integral manifolds in which the variables x^A and their differentials dx^A can freely be chosen, the ideal is said to be *in involution* with respect to the x^A.[3,4]

When one imposes independence of the variables x^A on a set of forms, if it is not already in involution with respect to these variables, he perforce searches for a class of singular integral manifolds. If they exist, it is possible by prolongation—the introduction of more variables and more forms—to derive a new set which is in involution with the required set x^A.[3,4,7] In this way one discovers variables, in terms of which the desired solutions are obtainable from appropriate Cauchy data. We, however, have been more concerned with the converse problem: given a set of physically motivated partial differential equations, to write an equivalent set of forms which *is* in involution with respect to the independent variables (i.e., a set whose general integral manifolds are the desired solutions). In all the cases below this has been achieved, as we have verified by detailed application of Cartan's criteria. (There is, however, one delicate point. If the ideal of forms does not have the additional property of being *complete*,[3] it can happen that the ideal may be augmented with some additional forms, giving a larger ideal that still leads to all the solutions of the same set of partial differential equations.) Indeed we believe involutory sets to be achievable for the partial differential equations of any well-set physical theory. All this may well be the mathematical basis for the otherwise somewhat mysterious need for introduction of certain sets of "extra" variables in physical theories (such as in general relativity, where we refer to the Riemann tensor components) whose partial differential equations would seem otherwise strangely incomplete. Further insight into this remark comes also from consideration of the structure of Maxwell's equations, as we will discuss later.

With the involutory property, then, we regard the differential ideals as precisely equivalent to—or giving a geometric form to—the structures of the equivalent sets of partial differential equations immediately derivable from them. Both the general and singular solutions of a set of partial differential equations must partake of the invariance group of the collection of geometric objects described by the equivalent ideal, and it is such groups and their generators that we investigate and use to discover special solutions.

IV. THE ISOGROUP

We write a set of forms in n-space as ω_i, $i = 1, \cdots$; the ith form ω_i is taken to be of degree p_i. We then require that the ideal be invariant under the continuous dragging transformation (or mapping, or diffeomorphism) generated by the vector field V, by writing

$$\underset{V}{\pounds}\,\omega_i = \sum_j \lambda_i^j \omega_j. \tag{4}$$

The sum includes only forms ω_j for which $p_j \leq p_i$ (thus including ω_i itself). λ_i^j is an arbitrary form of degree $p_i - p_j$. V is a contravariant vector field in the space of n dimensions, and is to be considered as a function of all n variables (i.e., $n - p$ "dependent" and p "independent").

A set of Lie increments $\epsilon\,\pounds_V\,\omega_i$ represents simultaneous infinitesimal changes in all the objects ω_i equivalent to the "active" coordinate change, or point transformation,

$$'x^\mu = x^\mu + \epsilon V^\mu \quad (\epsilon \text{ infinitesimal}), \tag{5}$$

of the background coordinatized n-manifold. Requiring these increments of the set of ω_i to produce *no change* in any of the integral manifolds (which clearly preserves the form of the original system of partial differential equations) is thus the same as having the equivalent active point transformation take all solutions simultaneously into other solutions. Clearly, the most general way to achieve the former is with Eq. (4), which asserts that the increments of ω_i are in the ideal, and conversely; in Cartan's terminology, the ω_i and the transformed ω_i are algebraically equivalent sets and the ideal is unchanged.

To explicitly find the invariance group, we expand Eq. (4) in terms of basis p_i-forms (products of p_i 1-forms, usually the dx^j). This is done with the aid of the identities for forms (1)–(3). We then equate coefficients of the basis p_i-forms to zero and algebraically eliminate the components of the λ_i^j. This yields a system of *linear* partial differential equations for the components of the V, which can usually be solved in a straightforward fashion.

We denote a solution V an *isovector* of the ideal and, finding all such, consider the transformation group, or *isogroup*, which they generate, and its Lie algebra. If the N, say, distinct isovector fields are labeled by a subscript A, $B = 1, 2, \cdots, N$, the structure constants of the group follow from

$$[V_A, V_B] = C_{AB}^C V_C.$$

One may ask: Cannot this all be done in indicial notation without ever introducing forms? The answer is, of course, yes; in fact, invariance groups have been computed traditionally in essentially this way.[1,2] However, it seems to us in practice that the use of forms usually enables calculation to be carried out more quickly and easily and that recognition of the n-dimensional geometric structures involved can be of great heuristic value in discovering special classes of solutions.

V. SPECIAL SOLUTION SETS

We now mention several ways of using the isovectors to generate additional forms which, when adjoined to the original set, give an augmented ideal whose solutions are in some sense *special*: A subset of the general solutions of the original set. It can happen that such a restricted problem is soluble in exact terms, whereas the general problem is not. It can also happen that a suitably restricted problem may have considerable physical interest by itself. It seems that any way of finding suitable restrictions—i.e., discovering augmented ideals which still have nonempty sets of solutions—must closely depend on similar uses of the isogroup.

(1) Invariant subspaces (or varieties) of the isogroup are described by degeneration of the rank of the matrix of V^{μ}_{A}. Here $\mu = 1, \cdots, n$ are labels of coordinates in the space of forms, and $A = 1, \cdots, N$ are labels of the isovectors. These varieties are taken into themselves by the transformations of the group[9]; hence any integral manifold specialized to be immersed in such a variety continues to be so immersed when transformed—there then being a set of such special solutions. They annul an augmented ideal including the 0-forms of the invariant variety and their exterior derivatives. For reasons that will become apparent when we discuss Maxwell's equations, we can call such solutions "algebraically special."

(2) A generalization of this is to search for all other sets of forms which, when adjoined to the original ideal, give an augmented ideal which (a) is still in involution with respect to the desired independent variables (this, however, may be achievable by prolongation) and (b) has an isogroup which includes all the original isovectors and so is at least as large as the original isogroup (an exception to this last could be made, if physically a subgroup of the original isogroup were deemed most significant and alone were required to be preserved). How to achieve this in general, by systematic search, is not clear to us. We find an example of such a process, however, in the discussion of the vacuum Maxwell equations.

One possible way to achieve the above is clear if the Lie algebra of the isogroup (or one of its subgroups) has an ideal—i.e., if the isogroup has an invariant subgroup. The ideal of forms then may be augmented by contractions of some or all of them with *all* the isovectors in the ideal of the isogroup, and we will show that the resulting augmented ideal of differential forms will have an isogroup certainly not smaller, and even perhaps larger, than the original isogroup (or subgroup). The subset of solutions thus singled out may have quite interesting additional physical symmetries. Let V_A represent the isovectors from the ideal of the isogroup and V be a general isovector. From the original ideal of forms, we choose a closed set ω_i such that $\pounds_V \omega_i$ is in the ideal ω_i (this may be the entire original ideal), and augment the ideal with the contracted forms $\omega_{Ai} \equiv V_A \lrcorner \omega_i$. The augmented ideal is still closed, since from Eq. (3)

$$d\omega_{Ai} = \pounds_{V_A} \omega_i - V_A \lrcorner (d\omega_i); \tag{6}$$

the first term on the right is in the ideal of ω_i, since V_A *is* an isovector, and the second term is clearly in the augmented ideal. To prove our assertion about the isogroup, then, consider [using Eq. (3)]

$$\pounds_V \omega_{Ai} = \pounds_V V_A \lrcorner \omega_i = [V, V_A] \lrcorner \omega_i + V_A \lrcorner \pounds_V \omega_i; \tag{7}$$

since V_A is an ideal, $[V, V_A]$ belongs to V_A; by hypothesis $\pounds_V \omega_i$ is in ω_i, so that the second term on the right side belongs to ω_{Ai} and so that any V will still be an isovector of the augmented set.

(3) If we are willing to search for a class of solutions no longer invariant under the isogroup, we can, from a single isovector V, taken from the general solution of the linear equations (4), obtain a set of forms

$$\sigma_i = V \lrcorner \omega_i.$$

(It should be noted that, since such a V is a superposition of the generators, it contains N arbitrarily chosen parameters.) Then

$$\begin{aligned} \pounds_V \sigma_i &= \pounds_V V \lrcorner \omega_i = V \lrcorner \pounds_V \omega_i \\ &= V \lrcorner \sum_j \lambda^j_i \omega_j \\ &= \sum_j [V \lrcorner \lambda^j_i] \omega_j + \sum_j (-1)^{p_i - p_j} \lambda^j_i \sigma_j . \end{aligned}$$

Thus, since the ideal ω_i is invariant under V, the *augmented* ideal $\{\omega_i, \sigma_i\}$ is also invariant under this particular V (but not under the rest of the isovectors). This suggests that we may annul the augmented ideal $\{\omega_i, \sigma_i\}$ to find (a nonempty) class of special solutions of the original equations. By imposing independent variables one recovers the original set of partial differential equations from the ω_i, plus an auxiliary set from the σ_i. Of course, one can too severely limit the solutions by imposing arbitrary auxiliary conditions; but by imposing just the set found using one particular, but arbitrary, isovector in the above manner, we find the most general so-called "similarity" solutions of the original set. They functionally

depend on the ratios of the N parameters in V. This technique is especially appropriate when one is also dealing with boundary conditions invariant under a V.

(4) Finally, this last method suggests a sometimes useful generalization of the isovectors themselves. In the typical case we have considered, the isovectors are first found from the original set of forms and then are contracted with these forms to get new forms. However, these new forms *could* have been included as part of the original set, although their exact expression was not known since V has not yet been found. Thus we can augment the original set of forms with contracted forms constructed with a generalized isovector—one which itself preserves only the *augmented* ideal:

$$\underset{V}{\pounds}\,\omega_i = \sum_j \lambda_i^j \omega_i + \sum_j \mu_i^j V \lrcorner\, \omega_j. \tag{8}$$

The μ_i^j are forms of degree $p_i - p_j + 1$. It appears possible to make a variety of restrictive choices of the μ_i^j; in fact, leaving them too free can result in V being undetermined. We are not convinced that any class of solutions defined in this way will have much intrinsic physical interest, but if boundaries are also involved, this procedure may be appropriate. Indeed, we will show later that certain recently found "nonclassical similarity solutions" discussed by Bluman and Cole can be constructed in exactly this way.

VI. EXAMPLES

A. The One-Dimensional Heat Equation

We have

$$\psi_{xx} = \psi_t. \tag{9}$$

Here and in the rest of the section, lower case Latin subscripts denote partial differentiation, as does subscript ψ.

To convert this to a set of differential forms, we first reduce it to a first-order set by defining new variables: $u = \psi_t$ and $y = \psi_x$. The given equation is now $y_x = u$.

The set to be annulled has one 1-form and two 2-forms:

$$\begin{aligned} \alpha &= d\psi - u\,dt - y\,dx, \\ d\alpha &= -du\,dt - dy\,dx, \\ \beta &= u\,dx\,dt - dy\,dt. \end{aligned} \tag{10}$$

The number of variables $n = 5$: t, x, ψ, u, and y. By taking the outer derivative of β, the derived form is found to be in the ideal of α, $d\alpha$, and β, so that the set is closed.

The set is in involution with respect to x and t, and gives back the original partial differential equations if we impose independence of these variables, requiring the sectioned forms to be annulled by elements of a 2-manifold in which dx and dt are independent forms. By substituting $du = u_x\,dx + u_t\,dt$, etc., into α, this gives $u = \psi_t$ and $y = \psi_x$. Next, if we section $d\alpha$, we have

$$\begin{aligned} d\tilde{\alpha} &= -(u_t\,dt + u_x\,dx)\,dt - (y_t\,dt + y_x\,dx)\,dx \\ &= (y_t - u_x)\,dx\,dt, \end{aligned}$$

so that annulling gives $y_t = u_x$, the correct integrability condition. Finally, sectioning β yields

$$\tilde{\beta} = (u - y_x)\,dx\,dt,$$

so that annulling gives $u = y_x$ as above.

We now find the isogroup by requiring that the Lie derivatives of α, $d\alpha$, and β with respect to a vector field V be in the ideal of α, $d\alpha$, and β. The general isovector V is to be considered a function of all five variables. We give the calculation in detail to demonstrate the method. We consider first

$$\underset{V}{\pounds}\,\alpha = \lambda\alpha. \tag{11}$$

No other term on the right-hand side is possible since α is the only 1-form. We treat this with a trick (which can be used when there is a single 1-form). Write

$$F = V \lrcorner\, \alpha. \tag{12}$$

From Eq. (3) we have

$$\begin{aligned} &\underset{V}{\pounds}\,\alpha = V \lrcorner\, d\alpha + d(V \lrcorner\, \alpha), \\ &\therefore \quad V \lrcorner\, d\alpha = \lambda\alpha - dF. \end{aligned} \tag{13}$$

Expanding on the basis of 1-forms $d\psi$, dt, du, dx, and dy, we have

$$\begin{aligned} -V^u\,dt + V^t\,du - V^y\,dx + V^x\,dy \\ = \lambda(d\psi - u\,dt - y\,dx) - F_t\,dt \\ - F_x\,dx - F_u\,du - F_y\,dy \\ - F_\psi\,d\psi. \end{aligned}$$

From Eq. (12) we get

$$F = V^\psi - uV^t - yV^x.$$

We equate coefficients of each basis form to zero, eliminate λ, and solve for the V^i:

$$\begin{aligned} V^x &= -F_y, & V^t &= -F_u, \\ V^u &= F_t + uF_\psi, & V^u &= F_x + yF_\psi, \\ & & V^\psi &= F - uF_u - yF_y. \end{aligned} \tag{14}$$

If we take the exterior derivative of Eq. (11), we get

$$\underset{V}{\pounds}\,d\alpha = (d\lambda)\alpha + \lambda\,d\alpha; \tag{15}$$

thus $\pounds_V\,d\alpha$ is already in the ideal. We need only

consider $\pounds_V \beta$ to complete the calculation. We put

$$\pounds_V \beta = \xi\beta + \omega\alpha - \zeta\, d\alpha, \tag{16}$$

where ξ, ζ, and ω are arbitrary 0-, 0-, and 1-forms, respectively. We may require that ω have no $d\psi$ term without loss of generality. Expansion of Eq. (16) yields

$$\begin{aligned} V^u\, dx\, dt + u\, dV^x\, dt + u\, dx\, dV^t - dV^y\, dt - dy\, dV^t \\ = \xi(u\, dx\, dt - dy\, dt) + \zeta(du\, dt + dy\, dx) \\ + (A\, dt + B\, dx + C\, du + D\, dy) \\ \times (d\psi - u\, dt - y\, dx). \end{aligned} \tag{17}$$

ξ, ζ, A, B, C, and D are arbitrary and are to be eliminated. The expression dV^i, $i = x, t, y$, is here just an abbreviation for $V^i_t\, dt + V^i_x\, dx + V^i_\psi\, d\psi + V^i_u\, du + V^i_y\, dy$. The equations obtained by equating coefficients of all 2-forms to zero are

$$\begin{aligned} V^u + u(V^x_x + V^t_t) - V_x = \xi u + Ay - Bu, \\ uV^x_u = V^y_u = \zeta - Cu, \\ uV^x_y - V^y_y - V^t_t = -\xi - Du, \\ uV^t_u = Cy,\ uV^t_y + V^t_x = -\zeta + Dy, \\ V^t_u = 0,\ -uV^x_\psi + V^y_\psi = A, \\ uV^t_\psi = B,\quad 0 = C,\quad -V^t_\psi = D. \end{aligned} \tag{18}$$

Solutions of Eqs. (14) and (18) now yields

$$\begin{aligned} V^t &= 2k_6t^2 + 2k_4t + k_1, \\ V^x &= 2k_6tx + k_4x - 2k_5t + k_2, \\ V^u &= (-\tfrac{1}{2}k_6x^2 + k_5x - 5k_6t + k_3 - 2k_4)u \\ &\quad + 2(k_5 - k_6x)y - k_6\psi + g_t, \\ V^y &= (-\tfrac{1}{2}k_6x^2 + k_5x - 3k_6t + k_3 - k_4)y \\ &\quad + (k_5 - k_6x)\psi + g_x, \\ V^\psi &= (-\tfrac{1}{2}k_6x^2 + k_5x - k_6t + k_3)\psi + g, \end{aligned} \tag{19}$$

where $g = g(t, x)$ satisfies $g_{xx} = g_t$. The k_i, $i = 1, \cdots, 6$, are constants.

We have six parameters and one (somewhat) arbitrary function. If we set all but one of these equal to zero, the remaining one describes an independent generator of the invariance group. There are thus seven separate generators (the last one actually comprising an infinite family of generators), which we summarize in Table I. Rows 1–6 are characterized by k_{1-6}, row 7 by g. A description of each type of transformation is provided where feasible. It should be noted that this particular group has been known for many years.[2]

Generators 1–4, and7 are in fact obvious from direct inspection of the heat equation; 5 and 6, however, might not have been anticipated. Generators 3 and 7 result from the linearity of the equation: It is an advantage of the present approach that even were the linearity not apparent, it would become so through the isogroup structure.

Similarity solutions of the heat equation are found by taking one particular (but arbitrary) generator V of the isogroup (omitting No. 7, however) and defining new forms:

$$F = V \lrcorner \alpha,\quad \theta = V \lrcorner d\alpha,\quad \sigma = V \lrcorner \beta. \tag{20}$$

We ask that the solution 2-manifolds annul these forms as well as α, $d\alpha$, and β.

The adjoined 0-form F is the same F appearing in Eq. (12) (but now for our particular V). Thus the similarity solution 2-manifold is described by setting

$$F = V^\psi - V^t\psi_t - V^x\psi_x = 0, \tag{21}$$

where we have substituted $u = \psi_t$ and $y = \psi_x$, relations true on that manifold. As might be expected, annulling θ merely yields $\partial F/\partial t = \partial F/\partial x = 0$. Finally, annulling σ yields the known equations $u_x = y_t$ and $y_x = u$.

From Eq. (19), u and y do not appear in V^t, V^x, and V^ψ, so that we have a quasilinear partial differential equation for ψ as a function of x and t, Eq. (21), to be solved in addition to the original heat equation. The general solution is of the form $\psi = H(x, t)G(\eta)$, where $\eta(x, t)$ is a "similarity variable" and $G(\eta)$ is arbitrary. H and η involve the $n - 1 = 5$ ratios of

TABLE I. Isogroup of $\psi_{xx} = \psi_t$ ($u = \psi_t$, $y = \psi_x$).

No.	V^t	V^x	V^ψ	V^u	V^y	Description
1	1	0	0	0	0	time translation
2	0	1	0	0	0	space translation
3	0	0	ψ	u	y	ψ scale change
4	$2t$	x	0	$-2u$	$-y$	t, x scale change
5	0	$-2t$	$x\psi$	$2y + xu$	$\psi + xy$	Galilean transformation
6	$2t^2$	$2xt$	$-(\frac{1}{2}x^2 + t)\psi$	$-\frac{1}{2}x^2u - 5tu - 2xy - \psi$	$-\frac{1}{2}x^2y - 3ty - x\psi$	
7	0	0	g	g_t	g_x	addition of arbitrary solution

the parameters in V. Substitution into the original heat equation results in an equation for G. Explicit forms for η, H, and G are given by Bluman and Cole, in a recent careful discussion of similarity solutions of this equation.[2]

Finally, as an example of similarity solution by generalization of isovectors, we can modify Eq. (16) by writing

$$\underset{V}{\pounds}\,\beta = -\zeta\, d\alpha + \xi\beta + \omega\alpha + \tau\theta. \tag{22}$$

We have added an additional term $\tau\theta = \tau V \lrcorner\, d\alpha$, where τ is an arbitrary 1-form, and wish to solve this equation for V—now a generalized isovector. Equation (11) for α we preserve unchanged. We use Eqs. (14), as before, for the V^i in terms of F. Elimination of ξ, ζ, ω, and τ produces *one* equation for F:

$$\begin{aligned} 0 = {} & F_{uu}(F_x + yF_\psi + uF_y)^2 - 2F_u(F_x + yF_\psi + uF_y) \\ & \times (uF_{uy} + yF_{u\psi} + F_{xu}) \\ & + F_u^2(-F_t + F_{xx} + 2uF_{xy} + 2yF_{x\psi} + 2uyF_{y\psi} \\ & + u^2F_{yy} + y^2F_{\psi\psi}). \end{aligned} \tag{23}$$

From *any* solution of this equation, we can find the five V^i. We must realize, however, that in general V^t, V^x, and V^ψ now will be functions of u and y and that Eq. (21) will now give a nonlinear equation for ψ.

Here we will content ourselves by further arbitrarily requiring V^t, V^x, and V^ψ to be independent of u and y. This requirement enables us to restrict F and to break up Eq. (23) into parts. If we write $V^x/V^t = A$ and $V^\psi/V^t = B$, then we can show that $A = A(x, t)$, $B = C(x, t)\psi + D(x, t)$, and

$$\begin{aligned} & A_t + 2AA_x - A_{xx} = -2C_x, \\ & C_t - C_{xx} + 2CA_x = 0, \\ & D_t - D_{xx} + 2DA_x = 0. \end{aligned} \tag{24}$$

These equations were obtained by Bluman and Cole in another manner, which they denoted a "nonclassical" approach to similarity solutions. We have obtained such solutions here by a more geometrical method. Any solution of these equations, coupled with the differential equations $dx/dt = A$ and $d\psi/dt = B$, yields a generalized similarity solution. The "classical" case is the special case $A_{xx} = 0$.

B. The Vacuum Maxwell Equations

As for so many other considerations of theoretical physics, these seem to furnish the illustration *par excellence* for our present method. We are presented with a set of eight partial differential equations in $n = 10$ variables ($n - p$ = six dependent, E_x, E_y, E_z, B_x, B_y, B_z; p = four independent, t, x, y, z). We first label these in a straightforward way as x^4, x^5, x^6, x^7, x^8, x^9 and x^0, x^1, x^2, x^3, and clearly can write a set of eight four-forms (in a general 10-dimensional manifold)

$$\begin{aligned} & dx^4\,dx^1\,dx^0\,dx^2 + dx^6\,dx^3\,dx^0\,dx^2 + dx^8\,dx^3\,dx^1\,dx^2, \\ & dx^5\,dx^2\,dx^0\,dx^3 + dx^4\,dx^1\,dx^0\,dx^3 + dx^9\,dx^1\,dx^2\,dx^3, \\ & dx^6\,dx^3\,dx^0\,dx^1 + dx^5\,dx^2\,dx^0\,dx^1 + dx^7\,dx^2\,dx^3\,dx^1, \\ & dx^7\,dx^1\,dx^0\,dx^2 + dx^9\,dx^3\,dx^0\,dx^2 - dx^5\,dx^3\,dx^1\,dx^2, \\ & dx^8\,dx^2\,dx^0\,dx^3 + dx^7\,dx^1\,dx^0\,dx^3 - dx^6\,dx^1\,dx^2\,dx^3, \\ & dx^9\,dx^3\,dx^0\,dx^1 + dx^8\,dx^2\,dx^0\,dx^1 - dx^4\,dx^2\,dx^3\,dx^1, \\ & dx^7\,dx^2\,dx^3\,dx^0 + dx^8\,dx^3\,dx^1\,dx^0 + dx^9\,dx^1\,dx^2\,dx^0, \\ & dx^4\,dx^2\,dx^3\,dx^0 + dx^5\,dx^3\,dx^1\,dx^0 + dx^6\,dx^1\,dx^2\,dx^0, \end{aligned} \tag{25}$$

from which, by imposing independence of x^0, x^1, x^2, and x^3, we can immediately derive the original partial differential set. This set of forms, however, is not in involution according to Cartan's criteria, so that the desired solutions are singular integral manifolds.

It is interesting that the first six of these 4-forms, however, do form an involutory system with respect to x^0, x^1, x^2, and x^3. The isogroup of this system has 10 parameters in addition to the infinite part expressing the linearity. The corresponding 10 isovectors generate transformations of time translation, space–time scale change, three-dimensional rigid spatial translation and rotation, and scale change and duality rotation of the **E** and **B** fields.

Adjoining the last two forms to the set does not eliminate the possibility of any solution—as arbitrary additional forms might. Rather, a subset of solutions is selected—viz., the singular solutions describing vacuum Maxwell fields—and we will next see in detail how much broader is the invariance group of this more restricted set. The last two forms are thus in an explicit sense *compatible* with the first six—this is an example of discovery of special subfamilies of solutions, described by an augmented set of forms. This particular example is, we realize, already well known as the fact that the vanishing of the divergences of **E** and **B** need only be required on a spacelike surface: The curl equations then ensure it throughout.

The set (25) can obviously be systematically completed and generalized to $n = 12$ by cyclically adjoining terms in two new variables, say, $E_t = x^{10}$ and $B_t = x^{11}$, e.g., to the first form add

$$dx^{10}\,dx^3\,dx^1\,dx^0,$$

to the fourth $dx^{11}\,dx^3\,dx^1\,dx^0$, to the seventh

$$dx^{10}\,dx^1\,dx^2\,dx^3,$$

etc. The resulting set is in involution, and in fact expresses Dirac's electron theory, for $m_0 = 0$ and no applied field.[10] The electron mass and applied field terms can be included as coefficients of another term $dx^0\,dx^1\,dx^2\,dx^3$ adjoined to each form, without affecting the closure and involutory properties. We hope to treat more such general sets in a future paper.

To find the invariance group of the Maxwell set (25), we find it more convenient to consider an ideal based on the two 3-forms in 10 dimensions

$$\begin{aligned}\alpha \equiv\ & dx^4\,dx^1\,dx^0 + dx^5\,dx^2\,dx^0 + dx^6\,dx^3\,dx^0 \\ & + dx^7\,dx^2\,dx^3 + dx^8\,dx^3\,dx^1 + dx^9\,dx^1\,dx^2, \\ \beta \equiv\ & dx^7\,dx^1\,dx^0 + dx^8\,dx^2\,dx^0 + dx^9\,dx^3\,dx^0 \\ & - dx^4\,dx^2\,dx^3 - dx^5\,dx^3\,dx^1 - dx^6\,dx^1\,dx^2.\end{aligned} \tag{26}$$

The eight 4-forms first considered are in the ideal of α and β. Moreover, α and β *are* in involution with respect to x^0, x^1, x^2, and x^3, and so it is their *general* integral 4-manifolds which give vacuum Maxwell fields. Form equations equivalent to α and β have been given frequently in recent literature,[11] but it has not perhaps been fully emphasized that these are best regarded as geometrical constructs in 10-dimensional space. We are *not* using moving orthogonal frames in a Minkowski 4-space; indeed this last only emerges as a part of the group structure which we find for solutions, and is not given *a priori*.

To simplify the algebraic labor, we take advantage of the permutation symmetry of α and β in dx^1, dx^2, and dx^3 to introduce a complex 3-vector notation (but with no presumption of any sort of 3-metric, the dot product simply being notation for a sum of products from 1 to 3). We define the following quantities:

$$\begin{aligned}\mathbf{h} &= \mathbf{E} + i\mathbf{B}, \\ \gamma &= \alpha + i\beta, \\ \mathbf{r} &= \hat{\mathbf{i}}x + \hat{\mathbf{j}}y + \hat{\mathbf{k}}z,\end{aligned} \tag{27}$$

and will work with complex 3-vector differential forms, using all the usual identities of vector-dyadic analysis, but suitably modifying them to take account of the noncommutativity of differential forms. For example, for such forms we have

$$d\mathbf{r} \times d\mathbf{r} = 2(\hat{\mathbf{i}}\,dy\,dz + \hat{\mathbf{j}}\,dz\,dx + \hat{\mathbf{k}}\,dx\,dy),$$

instead of zero.

Bearing these considerations in mind, we can write the complex 3-form γ as

$$\gamma = d\mathbf{h} \cdot (d\mathbf{r}\,dt - \tfrac{1}{2}i\,d\mathbf{r} \times d\mathbf{r}). \tag{28}$$

The 4-dimensional integral manifolds which annul this are solutions of the Maxwell partial differential equations for vacuum. The isovector equation is

$$\underset{V}{\pounds}\,\gamma = \mu\gamma + \nu\gamma^*, \tag{29}$$

where the star represents a complex conjugate and μ and ν are unspecified complex 0-forms. The variables are t, $\mathbf{r}$, $\mathbf{h}$, and $\mathbf{h}^*$. We write the isovector components as $T = V^t$, $\mathbf{R} = V^{\mathbf{r}}$, and $\mathbf{H} = V^{\mathbf{h}}$. $\mathbf{R}$ and T are real, $\mathbf{H}$ complex. Then expansion gives

$$\begin{aligned}& d\mathbf{H} \cdot (d\mathbf{r}\,dt - \tfrac{1}{2}i\,d\mathbf{r} \times d\mathbf{r}) \\ & \quad + d\mathbf{h} \cdot (d\mathbf{R}\,dt + d\mathbf{r}\,dT - i\,d\mathbf{R} \times d\mathbf{r}) \\ & = \mu\,d\mathbf{h} \cdot (d\mathbf{r}\,dt - \tfrac{1}{2}i\,d\mathbf{r} \times d\mathbf{r}) \\ & \quad + \nu\,d\mathbf{h}^* \cdot (d\mathbf{r}\,dt + \tfrac{1}{2}i\,d\mathbf{r} \times d\mathbf{r}).\end{aligned} \tag{30}$$

In order to use this equation, we need several algebraic identities involving vectors, dyadics, and differential forms, which are presented here without proof. In the following, $\mathbf{A}$ is a vector 0-form, $\mathbf{M}$ is a dyadic 0-form, $\boldsymbol{\omega}$ is a vector 1-form, $\mathbf{m} = \frac{1}{2}\mathbf{M} \times \mathbf{I}$ (the antisymmetric part of $\mathbf{M}$ expressed as an axial vector), and $\mathbf{I}$ is the unit dyadic:

$$\begin{aligned}2\boldsymbol{\omega}(\mathbf{A} \cdot \boldsymbol{\omega}) &= \mathbf{A} \times (\boldsymbol{\omega} \times \boldsymbol{\omega}), \\ (\boldsymbol{\omega} \cdot \mathbf{M}) \times \boldsymbol{\omega} &= \tfrac{1}{2}[\mathbf{I}(\operatorname{Tr}\mathbf{M}) - \mathbf{M}] \cdot (\boldsymbol{\omega} \times \boldsymbol{\omega}), \\ \boldsymbol{\omega} \cdot \mathbf{M} \cdot \boldsymbol{\omega} &= \mathbf{m} \cdot \boldsymbol{\omega} \times \boldsymbol{\omega}, \\ \boldsymbol{\omega} \cdot \mathbf{M} \cdot \boldsymbol{\omega} \times \boldsymbol{\omega} &= 2(\operatorname{Tr}\mathbf{M})\omega_1\omega_2\omega_3.\end{aligned} \tag{31}$$

$\boldsymbol{\nabla}$ will represent the operator $\hat{\mathbf{u}}_i \partial/\partial x^i$; $\boldsymbol{\nabla}_{\mathbf{h}}$ will represent $\hat{\mathbf{u}}_i \partial/\partial h^i$. The $\hat{\mathbf{u}}_i$ are the orthonormal basis $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$.

We consider first the $t\mathbf{hh}^*$ terms, which are

$$d\mathbf{h} \cdot (d\mathbf{h}^* \cdot \boldsymbol{\nabla}_{\mathbf{h}^*}\mathbf{R}) = 0$$

or

$$d\mathbf{h}^* \cdot \boldsymbol{\nabla}_{\mathbf{h}^*}\mathbf{R} \cdot d\mathbf{h} = 0.$$

We may remove the $d\mathbf{h}^*$ and the $d\mathbf{h}$—the coefficients are separately equal to zero—to obtain

$$\boldsymbol{\nabla}_{\mathbf{h}^*}\mathbf{R} = 0$$

and

$$\boldsymbol{\nabla}_{\mathbf{h}}\mathbf{R} = 0 \tag{32}$$

by a complex conjugation. Thus $\mathbf{R} = \mathbf{R}(t, \mathbf{r})$ only. The $\mathbf{rhh}^*$ terms now are

$$(d\mathbf{h} \cdot d\mathbf{r})(d\mathbf{h}^* \cdot \boldsymbol{\nabla}_{\mathbf{h}^*}T) = 0,$$

so that

$$\boldsymbol{\nabla}_{\mathbf{h}^*}T = \boldsymbol{\nabla}_{\mathbf{h}}T = 0. \tag{33}$$

Thus $T = T(t, \mathbf{r})$. With these restrictions, the $t\mathbf{hh}$ and $\mathbf{rhh}$ terms vanish.

The $\mathbf{h}^*\mathbf{r}t$ and $\mathbf{h}^*\mathbf{rr}$ terms yield

$$\boldsymbol{\nabla}_{\mathbf{h}^*}\mathbf{H} = \nu\mathbf{I} = -\nu\mathbf{I}. \tag{34}$$

Thus $v = 0$ and $\mathbf{H} = \mathbf{H}(t, \mathbf{r}, \mathbf{h})$. The **rrr** term is

$$-\tfrac{1}{2}i\, d\mathbf{r} \cdot \nabla\mathbf{H} \cdot d\mathbf{r} \times d\mathbf{r} = 0,$$

which yields

$$\nabla \cdot \mathbf{H} = 0. \tag{35}$$

The **rr***t* terms, after dropping dt, are

$$d\mathbf{r} \cdot \nabla\mathbf{H} \cdot d\mathbf{r} - \tfrac{1}{2}i\mathbf{H}_t\, d\mathbf{r} \times d\mathbf{r} = 0$$

or

$$\tfrac{1}{2}\nabla \times \mathbf{H} \cdot d\mathbf{r} \times d\mathbf{r} - \tfrac{1}{2}i\mathbf{H}_t \cdot d\mathbf{r} \times d\mathbf{r} = 0,$$

so that

$$\nabla \times \mathbf{H} - i\mathbf{H}_t = 0. \tag{36}$$

The *t***rh** terms yield

$$\nabla_{\mathbf{h}}\mathbf{H} + \mathbf{R}\nabla + \mathbf{I}T_t + i\mathbf{R}_t \times \mathbf{I} = \mu\mathbf{I}, \tag{37}$$

and the **rr***h* terms give

$$\nabla_{\mathbf{h}}\mathbf{H} + i\nabla T \times \mathbf{I} - \nabla\mathbf{R} + (\nabla \cdot \mathbf{R})\mathbf{I} = \mu\mathbf{I}. \tag{38}$$

Subtraction of these equations results in

$$\mathbf{R}\nabla + \nabla\mathbf{R} + i(\mathbf{R}_t - \nabla T) \times \mathbf{I} + \mathbf{I}(T_t - \nabla \cdot \mathbf{R}) = 0. \tag{39}$$

Take real and imaginary parts:

$$\mathbf{R}_t = \nabla T, \tag{40}$$

$$\mathbf{R}\nabla + \nabla\mathbf{R} = (\nabla \cdot \mathbf{R} - T_t)\mathbf{I}. \tag{41}$$

The trace of the second equation gives

$$\nabla \cdot \mathbf{R} = 3T_t. \tag{42}$$

These equations can be recognized as conformal Killing equations in a 4-space with 3-1 signature; the solution depends on 15 parameters and is

$$\begin{aligned}\mathbf{R} &= \mathbf{a}t^2 + \mathbf{b}t + \mathbf{c} + \mathbf{d} \times \mathbf{r} + (2et + f)\mathbf{r} \\ &\quad + 2(\mathbf{a} \cdot \mathbf{r})\mathbf{r} - \mathbf{a}r^2, \\ T &= \mathbf{r} \cdot (2\mathbf{a}t + \mathbf{b}) + e(r^2 + t^2) + ft + g;\end{aligned} \tag{43}$$

a, **b**, **c**, **d**, e, f, and g are real constants.

It is now easy to integrate for **H**; the result is

$$\mathbf{H} = \mathbf{h} \times \mathbf{u} + \mathbf{h}w + \mathbf{I}(t, \mathbf{r}), \tag{44}$$

where

$$\mathbf{u} = 2\mathbf{r} \times \mathbf{a} - \mathbf{d} - i(2\mathbf{a}t + \mathbf{b} + 2e\mathbf{r})$$

and

$$w = \mu - 2(2\mathbf{a} \cdot \mathbf{r} + 2et + f).$$

[The **h**-curl of Eq. (37) shows that $\mu = \mu(t, \mathbf{r})$ only.] Substitution into Eqs. (35) and (36) shows that

$$\nabla \times \mathbf{I} - i\mathbf{I}_t = 0, \quad \nabla \cdot \mathbf{I} = 0, \tag{45}$$

and μ is a constant (complex).

The complete group is thus characterized by 17 parameters and six functions (the **I**). We summarize the group as follows: Eqs. (45), with Table II.

TABLE II. Isogroup of the vacuum Maxwell equations. Meaning of generators corresponding to parameters in Eqs. (43) and (44).

No.	Parameters	Description
1	g	time translation
2–4	**c**	space translation
5–7	**b**	Lorentz transformation
8–10	**d**	space rotation
11	f	space–time scale change
12–15	e, **a**	conformal transformations
16	Re μ	field scale change
17	Im μ	duality rotation
18	**I**	addition of arbitrary solution

The coordinate transformations may be written in four-dimensional language as

$$\begin{aligned}V^\mu = \lambda^\mu + Kx^\mu + \eta^{\mu\alpha}\omega_{\alpha\nu}x^\nu + 2(\alpha_\tau x^\tau)x^\mu \\ - \alpha^\mu(\eta_{\tau\nu}x^\tau x^\nu),\end{aligned} \tag{46}$$

where $\omega_{\alpha\nu} = -\omega_{\nu\alpha}$ and the $\eta_{\tau\nu}$ is the Minkowski metric.

The matrix V^μ_A, $\mu = 1, \cdots, 10$, $A = 1, \cdots, 17$, is in general of rank 10: The transformation group is multiply transitive over the 10-space. The only degeneracy occurs on the eight-dimensional submanifold described by

$$\mathbf{E} \cdot \mathbf{E} - \mathbf{B} \cdot \mathbf{B} = 0 \quad \text{and} \quad \mathbf{E} \cdot \mathbf{B} = 0. \tag{47}$$

That this is an invariant variety may be directly verified by taking Lie derivatives with respect to the isovectors. It is interesting that when the Minkowski interpretation is made (i.e., that **E** and **B** are components of a bivector $F_{\mu\nu}$ in a *metric* 4-space spanned by x^0, x^1, x^2, x^3), these conditions appear as algebraic specializations at a point, apparently not dependent on the field equations.

C. The Korteweg–de Vries Equation[12]

The equation is

$$u_t + uu_x + \epsilon u_{xxx} = 0; \tag{48}$$

ϵ is a constant (usually small).

We write $z = u_x$, $w = u_{xx}$, and $y = u_{xt}$. Then the forms we need are

$$\begin{aligned}\alpha &= dz - w\,dx - y\,dt, \\ d\alpha &= -dw\,dx - dy\,dt, \\ \beta &= (du - z\,dx)\,dt, \\ \gamma &= du\,dx + uz\,dt\,dx - \epsilon\,dw\,dt.\end{aligned} \tag{49}$$

The variables are t, x, u, z, w, and y. The invariance group is given in Table III.

It is of considerable interest to compare this group with the isogroup of the same equation with $\epsilon = 0$:

TABLE III. Invariance group of the Korteweg–de Vries equation.

No.	V^t	V^x	V^u	V^z	V^w	V^y	Type
1	1	0	0	0	0	0	time translation
2	0	1	0	0	0	0	space translation
3	$3t$	x	$-2u$	$-3z$	$-4w$	$-6y$	x, t scale change
4	0	t	1	0	0	$-w$	Galilean transformation

$u_t + uu_x = 0$. This latter isogroup is given by

$$V^u = f(u, x - ut),$$
$$V^x - uV^t = tf(u, x - ut) + g(u, x - ut). \quad (50)$$

V^t is a completely arbitrary function of t, x, and u; f and g are arbitrary functions of their arguments. The inclusion of the ϵ term drastically reduces the isogroup of the equation.

Choose an arbitrary isovector V. We may without loss of generality adjust the ratio of generators 1, 2, 3, and 4 in V to be $-3t_0 : -(x_0 + 2t_0U) : 1 : 2U$, where t_0, x_0, and U are constants. With this choice, we have $V^t = 3(t - t_0)$, $V^x = x - x_0 + 2U(t - t_0)$, and $V^u = -2(u - U)$. To obtain a similarity solution, we annul the form $V \lrcorner \beta$ and obtain one new equation:

$$V^u = V^x u_x + V^t u_t. \quad (51)$$

Standard integration yields

$$u - U = T^{-\frac{2}{3}}F(\eta), \quad \text{where} \quad \eta = T^{-\frac{1}{3}}(X - UT),$$

and

$$X = x - x_0, \quad T = t - t_0.$$

Substitution into the Korteweg–de Vries equation itself now shows that F satisfies

$$\epsilon F''' + F'(F - \tfrac{1}{3}\eta) - \tfrac{2}{3}F = 0. \quad (52)$$

Any F satisfying this equation yields a general similarity solution containing the three parameters U, x_0, and t_0. A trivial solution is $F = \eta$, which leads to $u = X/T$.

We did not explicitly include $d\beta$ and $d\gamma$ in writing the closed set of forms to be annulled, since these 3-forms are quickly seen to be in the ideal. This, however, raises the question whether *closed* 2-forms, say $d\zeta$ and $d\xi$, algebraically equivalent to β and γ, cannot be found, in terms of which the set of forms would even more obviously be closed. This question is also of interest because of the contact it makes with the discovery of so-called conservation laws of the original partial differential equation. From Stokes theorem we have $\oint \zeta = \int d\zeta$, where the first integral is over a closed curve bounding the 2-manifold of the second integral (remember, we are in a six-dimensional space). Under the second integral we may place the section $\widetilde{d\zeta}$ of the 2-form with the manifold, and if the 2-manifold is such as to annul the 2-form—i.e., a solution of the partial differential equation—this last vanishes. We are left with the relation $\oint \tilde{\zeta} = 0$ in the solution manifold. By setting $\tilde{\zeta} = X\,dt - T\,dx$—where X is zero at the ends of some interval in line integration around a rectangular path in x, t space—it follows that the quantity $I \equiv \int T\,dx$ will be time independent. We have not searched for all such possible forms ζ; we have, however, found those specialized to be of the form $\zeta = \tilde{\zeta} = X\,dt - T\,dx$, by writing

$$d\zeta = \mu\gamma + \lambda\,d\alpha + \sigma\beta + \Lambda\alpha,$$

where μ, λ, and σ are unknown 0-forms and Λ a 1-form, and integrating the resulting linear partial differential equations. There are five distinct solutions, which may be linearly superimposed:

A. $X = \epsilon w + \frac{1}{2}u^2$,
$T = u$,
$d\zeta = -\gamma + u\beta$;

B. $X = -\epsilon uw + \frac{1}{2}\epsilon z^2 - \frac{1}{3}u^3$,
$T = -\frac{1}{2}u^2$,
$d\zeta = u\gamma - (\epsilon w + u^2)\beta + \epsilon z\alpha\,dt$;

C. $X = yu - \frac{1}{2}\epsilon w^2 - \frac{1}{2}wu^2 - (1/8\epsilon)u^4$,
$T = -wu - \frac{1}{2}z^2 - (1/6\epsilon)u^3$,
$d\zeta = [w + (1/2\epsilon)u^2]\gamma + [y - wu - (1/2\epsilon)u^3]\beta + z\alpha\,dx - u\,d\alpha$;

D. $X = \epsilon wx - \epsilon wut + \frac{1}{2}\epsilon z^2 t - \epsilon z + \frac{1}{2}xu^2 - \frac{1}{3}tu^3$,
$T = xu - \frac{1}{2}tu^2$,
$d\zeta = (tu - x)\gamma + (-\epsilon wt + xu - tu^2)\beta + \epsilon(zt - 1)\alpha\,dt$;

E. $X = \dfrac{\partial\phi}{\partial t} + y\dfrac{\partial\phi}{\partial z}$,
$T = -\dfrac{\partial\phi}{\partial x} - w\dfrac{\partial\phi}{\partial z}$,
$d\zeta = -d\left(\dfrac{\partial\phi}{\partial z}\alpha\right)$;

ϕ is an arbitrary function of t, x, and z. The first three of these correspond directly to the first three conservation laws in the second of Refs. 12; the fourth is apparently of less interest because the independent variables x and t occur in X. We see this process as a

generalization of a search for traditional first integrals—0-forms whose exterior derivatives are in the differential ideal—to higher order.

D. One-Dimensional Compressible Fluid Dynamics

We consider isentropic flow, so that the pressure P is a function of density only: $P = P(\rho)$. With the usual definition of sound speed, $c = (dP/d\rho)^{\frac{1}{2}}$, the equations are

$$\rho_t + \rho u_x + u\rho_x = 0,$$
$$\rho u_t + \rho u u_x + c^2\rho_x = 0.$$

We write these equations as the 2-forms

$$\alpha = d\rho\, dx - \rho\, du\, dt - u\, d\rho\, dt,$$
$$\beta = \rho\, du\, dx - \rho u\, du\, dt - c^2\, d\rho\, dt. \qquad (53)$$

The variables are t, x, ρ, and u. The isogroup is given in Table IV. F and G are functions of ρ and u and obey (subscripts denote derivatives)

$$G_u = F_\rho \quad \text{and} \quad \rho^2 G_\rho = c^2 F_u, \qquad (54)$$

yielding

$$c^2 G_{uu} - \rho^2 G_{\rho\rho} = 0.$$

One notes the interchange of dependent and independent variables—with consequent linearization. In fact, we have exactly performed a hodograph transformation. The hodograph equations result from imposing independence of ρ and u in Eqs. (53):

$$x_u + \rho t_\rho - u t_u = 0,$$
$$\rho x_\rho - \rho u t_\rho + c^2 t_u = 0.$$

Substitution of $x = uF/\rho - G$, $t = F/\rho$, where F and G are functions of u and ρ, yields again Eqs. (54).

E. The Lambropoulos Equation

The equation is

$$P_{xy} + axP_x + byP_y + cxyP + P_t = 0; \qquad (55)$$

a, b, and c are constants. This equation has been recently discussed by at least five authors.[13] We define $U = P_x$ and $Q = P_y$, and consider six variables t, x, y, P, U, and Q. We write the differential forms to be annulled as

$$\theta = (dP - U\, dx - Q\, dy)\, dt,$$
$$d\theta = -(dU\, dx + dQ\, dy)\, dt, \qquad (56)$$

and

$$\phi = [dU\, dt - dP\, dy + (axU + byQ + cxyP)\, dy\, dt]\, dx.$$

The isogroup turns out to have the following structure:

$$V^x = f + xg,$$
$$V^y = l + ym,$$
$$V^P = r + Ps,$$
$$V^u = U(s - g) + r_x + Ps_x, \qquad (57)$$
$$V^Q = Q(s - m) + r_y + Ps_y,$$
$$V^t = \int (g + m)\, dt + \delta,$$

where f, g, l, and m are functions of t, δ is a constant, and r and s are functions of t, x, and y. These functions obey the equations (a prime denotes differentiation with respect to t)

$$\begin{aligned}
&(g + m)'' - [(a + b)^2 - 4c](g + m) = 0,\\
&(g - m)' = (a - b)(g + m),\\
&f'' + (b - a)f' + (c - ab)f = 0,\\
&l'' + (a - b)l' + (c - ab)l = 0,\\
&r_{xy} + axr_x + byr_y + cxyr + r_t = 0,\\
&s_x = l' + ym' - bl - by(g + m),\\
&s_y = f' + xg' - af - ax(g + m),\\
&s_t = 2(ab - c)xy(g + m) + (ab - c)(xl + yf)\\
&\quad - ax(l' + ym') - by(f' + xg')\\
&\quad + a(g + m) - g'.
\end{aligned} \qquad (58)$$

The addition of an arbitrary solution is given by r. To detail the other transformations, we write

$$\lambda_{1,2} = -\tfrac{1}{2}(a + b) \pm (\tfrac{1}{4}(a + b)^2 - c)^{\frac{1}{2}},$$

and note that

$$\lambda_1 + \lambda_2 = -(a + b) \quad \text{and} \quad \lambda_1\lambda_2 = c.$$

We then have two cases. Note that in each case we have nine parameters: α, β, γ, δ, μ, ν, ξ, η, and k.

Case 1: $\lambda_1 \neq \lambda_2$:

$$\begin{aligned}
g &= \gamma + \alpha(a + \lambda_1)e^{(\lambda_1 - \lambda_2)t} + \beta(b + \lambda_1)e^{-(\lambda_1 - \lambda_2)t},\\
m &= -\gamma + \alpha(b + \lambda_1)e^{(\lambda_1 - \lambda_2)t} + \beta(a + \lambda_1)e^{-(\lambda_1 - \lambda_2)t},\\
f &= \mu e^{(a+\lambda_1)t} + \nu e^{-(b+\lambda_1)t},\\
l &= \xi e^{(b+\lambda_1)t} + \eta e^{-(a+\lambda_1)t},\\
s &= (\lambda_1 - \lambda_2)(\alpha\lambda_1 e^{(\lambda_1 - \lambda_2)t} + \beta\lambda_2 e^{-(\lambda_1 - \lambda_2)t})xy\\
&\quad + (\xi\lambda_1 e^{(b+\lambda_1)t} + \eta\lambda_2 e^{-(a+\lambda_1)t})x\\
&\quad + (\mu\lambda_1 e^{(a+\lambda_1)t} + \nu\lambda_2 e^{-(b+\lambda_1)t})y\\
&\quad - \alpha\lambda_1 e^{(\lambda_1 - \lambda_2)t} + \beta\lambda_2 e^{-(\lambda_1 - \lambda_2)t} + k,
\end{aligned} \qquad (59)$$

TABLE IV. Isogroup for one-dimensional compressible fluid flow.

No.	V^t	V^x	V^ρ	V^u	Type
1	t	x	0	0	t, x scale change
2	0	t	0	1	Galilean transformation
3	$\frac{F}{\rho}$	$\frac{uF}{\rho} - G$	0	0	Hodograph transformation

so that

$$V^t = \delta + \alpha e^{(\lambda_1-\lambda_2)t} - \beta e^{-(\lambda_1-\lambda_2)t}.$$

Case 2: $\lambda_1 = \lambda_2 \equiv \lambda$:

$$\begin{aligned} g &= 2\alpha t + \beta + (a-b)(\alpha t^2 + \beta t) + \gamma, \\ m &= 2\alpha t + \beta - (a-b)(\alpha t^2 + \beta t) - \gamma, \\ f &= (\mu + \nu t)e^{(a+\lambda)t}, \\ e &= (\xi + \eta t)e^{-(a+\lambda)t}, \\ s &= 2(2\lambda\alpha t + \lambda\beta + \alpha)xy + x(\eta + \lambda\xi + \lambda\eta t)e^{-(a+\lambda)t} \\ &\quad + y(\nu + \lambda\mu + \lambda\nu t)e^{(a+\lambda)t} \\ &\quad - 2(\lambda\alpha t + \lambda\beta + \alpha)t + k, \end{aligned} \tag{60}$$

so that

$$V^t = 2\alpha t^2 + 2\beta t + \delta.$$

We summarize the significance of the parameters in Table V. There are no invariant varieties.

For similarity solutions, set $r = 0$ and adjoin the forms contracted with a general isovector. Only one equation now results:

$$V^P = V^t P_t + V^u P_y + V^x P_x, \tag{61}$$

which is quasilinear. The standard integration procedure gives

$$\frac{dx}{dt} = \frac{f + xg}{V^t}, \qquad \frac{dy}{dt} = \frac{l + ym}{V^t}, \qquad \frac{dP}{dt} = P\frac{s}{V^t}. \tag{62}$$

Similarity variables are

$$w = -L + x/M,$$

and

$$z = -H + y/Q, \tag{63}$$

TABLE V. Invariance group parameters for the Lambropoulos equation.

No.	Parameter	Type
1	α	time-dependent scale change in x, y
2	β	time-dependent scale change in x, y
3	γ	constant scale change in x, y
4	δ	time translation
5	ξ	time-dependent translation in y
6	η	time-dependent translation in y
7	μ	time-dependent translation in x
8	ν	time-dependent translation in x
9	k	scale change in P
10	r	addition of arbitrary solution

where

$$\begin{aligned} M &= \exp\left(\int (V^t)^{-1} g\, dt\right), \\ Q &= \exp\left(\int (V^t)^{-1} m\, dt\right), \\ (QM &= V^t), \\ L &= \int (MV^t)^{-1} f\, dt, \\ H &= \int (QV^t)^{-1} l\, dt. \end{aligned} \tag{64}$$

Also,

$$P = F(w, z)\exp(Awz + Bz + Dw + E), \tag{65}$$

where

$$\begin{aligned} A &= m - bV^t, \\ B &= L(g - aV^t) + M^{-1}f, \\ D &= H(m - bV^t) + Q^{-1}l, \\ E &= \int dt[LH(m - bV^t) \\ &\quad + L(V^tH'' + V^{t\prime}H' + mH') \\ &\quad + H(V^tL'' + V^{t\prime}L' + gL') \\ &\quad + a + (V^t)^{-1}(K - g)], \end{aligned} \tag{66}$$

where

$$\begin{aligned} K &= k + \gamma - a\delta, \quad \text{Case 1}, \\ &= k + \gamma + \beta - a\delta, \quad \text{Case 2}. \end{aligned}$$

The equation for F is

$$F_{wz} - CwF_w + \Gamma F = 0, \tag{67}$$

where

$$\begin{aligned} C &= 2\gamma - (a-b)\delta, \\ \Gamma &= Gwz + nw + pz + q, \end{aligned}$$

and

$$G = \gamma^2 + \gamma\delta(b-a) + \delta^2(c - ab).$$

The slight asymmetry in this equation is caused by the slight asymmetry in A above. p and n are integration constants resulting from the L and H quadratures. q is a combination of constants p, n, K, the isogroup parameters, and a, b, and c.

A readily soluble subcase (given here for Case 1) is obtained by setting

$$\gamma = \delta = n = p = q = 0.$$

Then $F = F_1(w) + F_2(z)$. One has a six-parameter (α, β, ξ, η, μ, ν), two-arbitrary-function (F_1 and F_2) solution of the original equation. In this case, one has

$$\begin{aligned} H &= \frac{V^t(Q^{-1}l)'}{\alpha\beta(\lambda_1 - \lambda_2)^2}, \\ L &= \frac{V^t(M^{-1}f)'}{\alpha\beta(\lambda_1 - \lambda_2)^2}, \\ k &= -\alpha\beta(\lambda_1 - \lambda_2)^2 LH + (V^t)^{-1}fl, \end{aligned}$$

and k is guaranteed to be constant.

TABLE VI. Isogroup of the Einstein–Maxwell cylindrical wave equations.

No.	t	r	U	C	A	B	F	G	Type
1	1	0	0	0	0	0	0	0	time translation
2	t	r	0	0	$-A$	$-B$	$-F$	$-G$	r, t scale change
3	0	0	0	1	0	0	0	0	potential translation
4	0	0	1	C	0	0	F	G	U, C scale change
5	0	0	C	$\frac{1}{2}(C^2 - e^{2U})$	F	G	$CF - Ae^{2U}$	$CG - Be^{2U}$	

The two special solutions reported by Neuringer[13] are special similarity solutions for which $f = l = 0$, $\beta = 0$, $\alpha = -\delta \neq 0$, and $K =$ either 0 or $-2\delta(\lambda_1 - \lambda_2)$.

F. Einstein–Maxwell Cylindrical Wave Equations

Cylindrical wave equations for combined gravitational and electromagnetic fields have been presented before by one of the authors[14] in connection with solution generation methods in general relativity:

$$U_{rr} + (1/r)U_r - U_{tt} = e^{-2U}(C_t^2 - C_r^2),$$
$$C_{rr} + (1/r)C_r - C_{tt} = 2(C_rU_r - C_tU_t). \quad (68)$$

e^{2U} is a metric coefficient (gravitational potential); C is an electromagnetic potential.

If we write $A = U_t$, $B = U_r$, $F = C_t$, $G = C_r$, and include U and C in our set of variables, we obtain the following differential forms:

$$\begin{aligned} \alpha &= dU - A\,dt - B\,dr, \\ \beta &= dC - F\,dt - G\,dr, \\ d\alpha &= -dA\,dt - dB\,dr, \\ d\beta &= -dF\,dt - dG\,dr, \\ \gamma &= dB\,dt + dA\,dr - [e^{-2U}(F^2 - G^2) - B/r]\,dr\,dt, \\ \delta &= dG\,dt + dF\,dr - [2(BG - AF) - G/r]\,dr\,dt. \end{aligned} \quad (69)$$

The isogroup has five generators, as shown in Table VI.

We see from this table that the isogroup breaks naturally into two subgroups, which respectively transform only the original independent (t, r) or dependent variables (U, C). The first subgroup is quite trivial; the second is a multiply transitive group on (U, C) space which turns out to be integrable. Integration of $dU/d\tau = V^U$ and $dC/d\tau = V^C$ yields finite transformations which can be put in the form

$$C' = \xi + \frac{\zeta(C - \eta)}{(C - \eta)^2 + V^2},$$
$$V' = \frac{\zeta V}{(C - \eta)^2 + V^2}, \quad (70)$$

where $V = e^U$, $V' = e^{U'}$ and ξ, η, and ζ are constants. These equations generate new, physically distinct, solutions from old; they are implicit in the previous work, where they were found by a different approach.[14] A discrete transformation, $C' = -C$ and $V' = V$, was also presented there, and is indeed obvious from Eqs. (68). We do not find it with our present approach because it cannot be obtained by a continuous transformation from zero.

We list here, for completeness, 1-forms ζ whose exterior derivative $d\zeta$ is in the ideal (analogous to the treatment for the Korteweg–de Vries equation). We write $\zeta = R\,dt - T\,dr$ and give four solutions. The $d\zeta$ are not listed since they are rather long.

$$\begin{aligned} &\text{A.} \quad R = 2r(FGe^{-2U} + AB), \\ &\qquad\quad T = -re^{-2U}(F^2 + G^2) - r(A^2 + B^2); \\ &\text{B.} \quad R = rGe^{-2U}, \\ &\qquad\quad T = -rFe^{-2U}; \\ &\text{C.} \quad R = r(e^{-2U}CG + B), \\ &\qquad\quad T = -r(e^{-2U}CF + A); \\ &\text{D.} \quad R = r[(C^2e^{-2U} - 1)G + 2BC], \\ &\qquad\quad T = -r[(C^2e^{-2U} - 1)F + 2AC]. \end{aligned} \quad (71)$$

For each of these there is an integral conservation law for Eqs. (68).

ACKNOWLEDGMENTS

We wish to thank E. D. Fackerell for his interest and encouragement and for his help with the recent literature of similarity transformations. H. D. Wahlquist has been closely associated with us during all phases of the work, and has made many significant contributions of insight and criticism which are difficult to acknowledge adequately.

* Sponsored by the National Aeronautics and Space Administration under Contract No. NAS7-100.

† NAS–NRC Senior Resident Research Associate, supported by NASA. Present address: Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84601.

[1] L. V. Ovsjannikov, *Gruppovye svostva differentsialny uravneni* (U.S.S.R. Academy of Science, Novosibirsk, 1962) (*Group Properties of Differential Equations*, G. W. Bluman, trans., 1967).

[2] G. W. Bluman and J. D. Cole, J. Math. Mech. **18**, 1025 (1969).

[3] É. Cartan, *Les systèmes différentials extérieurs et leurs applications géométriques* (Hermann, Paris, 1945).
[4] W. Slebodzinski, *Formes extérieures et leurs applications* (Pánstwowe Wydawnictwo Naukowe, Warsaw, 1954, 1963), Vols. I and II.
[5] Y. Choquet-Bruhat, *Géométrie différentielle et systèmes extérieurs* (Dunod, Paris, 1968).
[6] R. Hermann, *Differential Geometry and the Calculus of Variations* (Academic, New York, 1968); Advan. Math. **1**, 265 (1965).
[7] J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed.; J. A. Schouten and W. v.d.Kulk, *Pfaff's Problem and Its Generalizations* (Clarendon, Oxford, 1949).
[8] N. J. Hicks, *Notes on Differential Geometry* (Van Nostrand, Princeton, N.J., 1965).
[9] L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961), pp. 67 ff.
[10] Compare J. Frenkel, *Wave Mechanics: Advanced General Theory* (Oxford U.P., Oxford, 1934), pp. 264–67.
[11] Cf., e.g., C. W. Misner and J. A. Wheeler, Ann. Phys. **2**, 525 (1957).
[12] Cf., e.g., R. M. Miura, C. S. Gardner, M. D. Kruskal, and C. H. Su, J. Math. Phys. **9**, 1202, 1204 (1968); **10**, 536 (1969); **11**, 952 (1970).
[13] P. Lambropoulos, J. Math. Phys. **8**, 2167 (1967); J. Neuringer, *ibid.* **10**, 250 (1969); M. E. Goldstein, *ibid.* **11**, 667 (1970); M. Kolsrud, *ibid.*, 829 (1970); R. Wilcox, *ibid.*, 1235 (1970).
[14] B. K. Harrison, Phys. Rev. **138**, B488 (1965); J. Math. Phys. **9**, 1744 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

A Class of Type [4] Perfect Fluid Space-Times*

M. OLESON
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada

(Received 21 September 1970)

The existence of gravitational fields of Petrov–Penrose type [4] in the presence of a perfect fluid is established. In particular, the general type [4] solution of the Einstein field equations with a perfect fluid as source is obtained subject to the restriction that the repeated principal null congruence of the Weyl tensor is geodesic. The line element is expressed in terms of four arbitrary functions of a single variable, and in general admits no Killing vectors. The fluid flow is irrotational, but has nonzero shear, expansion, and acceleration. The physically reasonable requirement $0 < p \leq A$, where p is the fluid pressure and A the rest energy density, is imposed, and restricts the domain of validity of the solutions to a certain extent. In addition, it is shown that the stronger condition $0 < p \leq A/3$ excludes certain of the above solutions and further restricts the domain of validity of the remaining ones.

1. INTRODUCTION

At any point P in space–time, the Weyl conform tensor (assumed nonzero) defines a set of four directions (possibly coincident) along the null cone at P which satisfy[1]

$$k_{[a}C_{b]cd[e}k_{f]}k^ck^d = 0,$$

called the gravitational principal null directions (p.n.d.'s). A gravitational field is said to be null or of Petrov–Penrose type [4] in some region if all the p.n.d.'s coincide in that region. The resulting p.n.d. then satisfies[1]

$$C_{abcd}k^d = 0. \tag{1.1}$$

Gravitational fields of this type have been interpreted as a simple form of gravitational radiation propagating along the congruence of null curves to which the repeated p.n.d. is tangent.[2–4] For a source-free field (i.e., zero Ricci tensor) the Goldberg–Sachs theorem[5] asserts that this congruence, called the *principal null congruence (p.n.c.) of the type* [4] *gravitational field*, is geodesic and shear-free.

In this paper we consider type [4] gravitational fields with a perfect fluid as source. Such fields have been considered by Kundt and Trümper[2] and by Szekeres,[3] who derived some general relationships between the p.n.c. and the fluid congruence, but obtained no explicit solutions.

The Einstein field equations with a perfect fluid as source read[6]

$$R_{ab} - \tfrac{1}{2}g_{ab}R = -[(A + p)u_au_b - pg_{ab}], \qquad u^au_a = +1, \tag{1.2}$$

where u_a is the fluid velocity, A the energy density, and p the pressure of the fluid, which we take to satisfy the inequalities

$$0 < p \leq A. \tag{1.3}$$

The signature of the metric tensor g_{ab} is $(- - - +)$.

We will show that, for type [4] solutions of (1.2), the p.n.c. cannot be geodesic and shear-free, as in the vacuum case. We therefore consider type [4] solutions for which this congruence is geodesic, but has nonzero

[3] É. Cartan, *Les systèmes différentials extérieurs et leurs applications géométriques* (Hermann, Paris, 1945).
[4] W. Slebodzinski, *Formes extérieures et leurs applications* (Pánstwowe Wydawnictwo Naukowe, Warsaw, 1954, 1963), Vols. I and II.
[5] Y. Choquet-Bruhat, *Géométrie différentielle et systèmes extérieurs* (Dunod, Paris, 1968).
[6] R. Hermann, *Differential Geometry and the Calculus of Variations* (Academic, New York, 1968); Advan. Math. **1**, 265 (1965).
[7] J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed.; J. A. Schouten and W. v.d.Kulk, *Pfaff's Problem and Its Generalizations* (Clarendon, Oxford, 1949).
[8] N. J. Hicks, *Notes on Differential Geometry* (Van Nostrand, Princeton, N.J., 1965).
[9] L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961), pp. 67 ff.
[10] Compare J. Frenkel, *Wave Mechanics: Advanced General Theory* (Oxford U.P., Oxford, 1934), pp. 264–67.
[11] Cf., e.g., C. W. Misner and J. A. Wheeler, Ann. Phys. **2**, 525 (1957).
[12] Cf., e.g., R. M. Miura, C. S. Gardner, M. D. Kruskal, and C. H. Su, J. Math. Phys. **9**, 1202, 1204 (1968); **10**, 536 (1969); **11**, 952 (1970).
[13] P. Lambropoulos, J. Math. Phys. **8**, 2167 (1967); J. Neuringer, *ibid.* **10**, 250 (1969); M. E. Goldstein, *ibid.* **11**, 667 (1970); M. Kolsrud, *ibid.*, 829 (1970); R. Wilcox, *ibid.*, 1235 (1970).
[14] B. K. Harrison, Phys. Rev. **138**, B488 (1965); J. Math. Phys. **9**, 1744 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

A Class of Type [4] Perfect Fluid Space-Times*

M. OLESON
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada

(Received 21 September 1970)

The existence of gravitational fields of Petrov–Penrose type [4] in the presence of a perfect fluid is established. In particular, the general type [4] solution of the Einstein field equations with a perfect fluid as source is obtained subject to the restriction that the repeated principal null congruence of the Weyl tensor is geodesic. The line element is expressed in terms of four arbitrary functions of a single variable, and in general admits no Killing vectors. The fluid flow is irrotational, but has nonzero shear, expansion, and acceleration. The physically reasonable requirement $0 < p \leq A$, where p is the fluid pressure and A the rest energy density, is imposed, and restricts the domain of validity of the solutions to a certain extent. In addition, it is shown that the stronger condition $0 < p \leq A/3$ excludes certain of the above solutions and further restricts the domain of validity of the remaining ones.

1. INTRODUCTION

At any point P in space–time, the Weyl conform tensor (assumed nonzero) defines a set of four directions (possibly coincident) along the null cone at P which satisfy[1]

$$k_{[a}C_{b]cd[e}k_{f]}k^ck^d = 0,$$

called the gravitational principal null directions (p.n.d.'s). A gravitational field is said to be null or of Petrov–Penrose type [4] in some region if all the p.n.d.'s coincide in that region. The resulting p.n.d. then satisfies[1]

$$C_{abcd}k^d = 0. \tag{1.1}$$

Gravitational fields of this type have been interpreted as a simple form of gravitational radiation propagating along the congruence of null curves to which the repeated p.n.d. is tangent.[2–4] For a source-free field (i.e., zero Ricci tensor) the Goldberg–Sachs theorem[5] asserts that this congruence, called the *principal null congruence (p.n.c.) of the type* [4] *gravitational field*, is geodesic and shear-free.

In this paper we consider type [4] gravitational fields with a perfect fluid as source. Such fields have been considered by Kundt and Trümper[2] and by Szekeres,[3] who derived some general relationships between the p.n.c. and the fluid congruence, but obtained no explicit solutions.

The Einstein field equations with a perfect fluid as source read[6]

$$R_{ab} - \tfrac{1}{2}g_{ab}R = -[(A + p)u_au_b - pg_{ab}], \qquad u^au_a = +1, \tag{1.2}$$

where u_a is the fluid velocity, A the energy density, and p the pressure of the fluid, which we take to satisfy the inequalities

$$0 < p \leq A. \tag{1.3}$$

The signature of the metric tensor g_{ab} is $(- - - +)$.

We will show that, for type [4] solutions of (1.2), the p.n.c. cannot be geodesic and shear-free, as in the vacuum case. We therefore consider type [4] solutions for which this congruence is geodesic, but has nonzero

shear. Section 2 contains a discussion of the well-known properties of congruences of null geodesics. In Sec. 3 the discussion is specialized to the case under consideration, and the dependence of the expansion and shear of the p.n.c. on an affine parameter is derived. The general line element is given in Sec. 4, and some of its properties discussed in Sec. 5.

2. PROPERTIES OF NULL GEODESIC CONGRUENCES

In discussing the properties of a congruence of null geodesics, it is convenient to use a null tetrad[7] (k^a, n^a, m^a, $\bar{m}^a$), where k^a and n^a are real null vector fields, m^a is a complex null vector field, and $\bar{m}^a$ is the complex conjugate of m^a. The orthogonality properties of these vectors at any point are

$$k^a m_a = k^a \bar{m}_a = n^a m_a = n^a \bar{m}_a = 0,$$
$$k^a n_a = 1, \quad m^a \bar{m}_a = -1. \tag{2.1}$$

It follows that the metric tensor g_{ab} can be expressed in terms of the null tetrad as

$$g_{ab} = k_a n_b + k_b n_a - m_a \bar{m}_b - m_b \bar{m}_a. \tag{2.2}$$

The real vector field k^a is chosen to be tangent to the given congruence of null geodesics. The remaining freedom in the choice of null tetrad is described by[7]

$$\hat{k}^a = Rk^a, \quad \hat{m}^a = e^{iS}(m^a - R\bar{T}k^a),$$
$$\hat{n}^a = R^{-1}n^a - Tm^a - \bar{T}\bar{m}^a + RT\bar{T}k^a, \tag{2.3}$$

where R and S are arbitrary real functions and T is complex.

The geometric properties of the null congruence are then described by the complex scalars

$$\rho = k_{a;b}m^a\bar{m}^b, \quad \sigma = k_{a;b}m^a m^b, \tag{2.4}$$

which, under (2.3), transform according to

$$\hat{\rho} = R\rho, \quad \hat{\sigma} = Re^{2iS}\sigma. \tag{2.5}$$

The real part of ρ is called the *expansion*, the imaginary part the *twist*, while the absolute value of σ is called the *shear*. By means of a tetrad transformation (2.3) with $S = T = 0$,we can ensure that the tangent field k^a corresponds to an affine parameter on the congruence. We then have the following expressions for the geometric scalars[7]:

$$\rho + \bar{\rho} = -k^a{}_{;a},$$
$$(\rho - \bar{\rho})^2 = -2k_{[a;b]}k^{a;b}, \tag{2.6}$$
$$\sigma\bar{\sigma} = \tfrac{1}{2}[k_{(a;b)}k^{a;b} - \tfrac{1}{2}(k^a{}_{;a})^2].$$

These quantities have a simple geometric interpretation. Intuitively speaking, if $\sigma \neq 0$, a small circle will be deformed into an ellipse as one follows the null geodesics. On the other hand, if $\sigma = 0$ and $\rho + \bar{\rho} \neq 0$, the radius of a small circle will change along the congruence. Finally the vanishing of $\rho - \bar{\rho}$ is a necessary and sufficient condition for the null geodesics to generate null hypersurfaces.[8]

Equations governing the behavior of ρ and σ along the null geodesics may be derived by contracting the Ricci identities

$$k_{a;bc} - k_{a;cb} = k_d R^d{}_{abc},$$

with $m^a k^b \bar{m}^c$ and $m^a k^b m^c$, and substituting from (2.6). One obtains, after specializing the tetrad so that[9]

$$m_{a;b}\bar{m}^a k^b = 0, \tag{2.7}$$

the following equations[10]:

$$D\rho = \rho^2 + \sigma\bar{\sigma} + \Phi_{00}, \tag{2.8}$$

$$D\sigma = (\rho + \bar{\rho})\sigma + \Psi_0, \tag{2.9}$$

where $D = k^a\partial/\partial x^a$ denotes differentiation along the congruence, and

$$\Phi_{00} = -\tfrac{1}{2}R_{ab}k^a k^b, \tag{2.10}$$

$$\Psi_0 = -C_{abcd}k^a m^b k^c m^d. \tag{2.11}$$

3. PROPERTIES OF THE NULL GEODESICS

We apply the results of the preceding section to the p.n.c. (assumed to be geodesic) of the type [4] perfect fluid space–time. First, Eq. (1.1) implies[11]

$$\Psi_0 = 0. \tag{3.1}$$

Furthermore, it is a consequence of the field equation (1.2) and the Bianchi identities (as verified in the Appendix using the Newman–Penrose[12] formalism) that the p.n.c. satisfies the restrictions

$$\rho = \bar{\rho}, \tag{3.2}$$

$$\Phi_{00} = 3\sigma\bar{\sigma}, \tag{3.3}$$

which are invariant under the tetrad transformation (2.3). Equations (2.8) and (2.9) thus reduce to

$$D\rho = \rho^2 + 4\sigma\bar{\sigma}, \tag{3.4}$$

$$D\sigma = 2\rho\sigma. \tag{3.5}$$

In addition, by virtue of the field equations (1.2), Eq. (3.3) can be written

$$(A + p)(k^a u_a)^2 = 6\sigma\bar{\sigma}, \tag{3.6}$$

which implies $\sigma \neq 0$ in the presence of a fluid. Furthermore, Eq. (3.4) implies $\rho \neq 0$.

On taking Eq. (3.2) into account, we can state the following theorem.

Theorem 3.1: If the principal null congruence of a type [4] gravitational field with perfect fluid as source is geodesic, then this congruence generates null hypersurfaces, but has nonzero shear and expansion.

The coordinate system is now specialized so that x^1, x^2, x^4 are constant along the null geodesics of the p.n.c. and $x^3 = r$ is an affine parameter along this congruence. Then the tangent field k^a is given by

$$k^a = \delta^a_3, \tag{3.7}$$

and Eqs. (3.4) and (3.5) may be integrated to yield the r dependence of ρ and σ. There are two distinct cases to be considered.

Case I: $\rho^2 - 4\sigma\bar{\sigma} = 0$:

$$\rho = -1/(2r), \quad \sigma = -1/(4r). \tag{3.8}$$

Case II: $\rho^2 - 4\sigma\bar{\sigma} \neq 0$:

$$\rho = -r/(r^2 - k^2), \quad \sigma = k/[2(r^2 - k^2)], \tag{3.9}$$

with $k = k(x^1, x^2, x^4)$.

We have used a coordinate transformation of the form $r' = r + f(x^1, x^2, x^4)$ to eliminate the "constant" of integration and a tetrad transformation (2.3) with $R = 1$, $T = 0$, and $\partial S/\partial r = 0$ [so that (2.7) is preserved], to set $\sigma = \bar{\sigma}$ [see Eq. (2.5)].

4. THE LINE ELEMENTS

We complete the construction of the coordinate system as follows. Let the null hypersurfaces generated by the p.n.c. (see Theorem 3.1) be given by $v =$ constant, and let ξ and η be parameters labeling distinct geodesics in these hypersurfaces. Then the quantities $(x^1, x^2, x^3, x^4) \equiv (\xi, \eta, r, v)$, where r is the affine parameter along the p.n.c. (introduced in the preceding section), form a coordinate system which is closely related to that of Robinson and Trautman.[13]

It is convenient to use the tetrad freedom (2.3) with $R = 1$, $S = 0$, to set $m^a u_a = 0$ so that u_a is of the form

$$u^a = (1/2B)k^a + Bn^a, \tag{4.1}$$

where

$$B = u^a k_a. \tag{4.2}$$

The simple r-dependence of ρ and σ [see Eqs. (3.8) and (3.9)] now enables one to integrate the field equations in Newman–Penrose form, in a manner similar to that used by Newman and Tambourino,[14] to obtain the r dependence of the metric tensor, the fluid velocity, energy density, and pressure, together with certain equations relating the various "constants" of integration that arise. One finds, in agreement with a result of Kundt and Trümper,[2] that the fluid congruence is hypersurface orthogonal. Furthermore, the coordinates can be specialized so that the corresponding family of hypersurfaces is given by $r =$ const. It follows, by virtue of (3.7) and (4.2), that

$$u_a = B\delta^3_a. \tag{4.3}$$

In the above procedure, the function $k(\xi, \eta, v)$ of Eq. (3.9) is transformed to be a positive constant[15] k. Class II necessarily divides into two subclasses IIa and IIb depending on whether $\rho^2 - 4\sigma\bar{\sigma} < 0$ or $\rho^2 - 4\sigma\bar{\sigma} > 0$. The affine parameter r can assume the values

$$0 < r < \infty, \tag{4.4}$$

$$0 \leq r < k, \tag{4.5}$$

$$0 < k < r < \infty \tag{4.6}$$

in classes I, IIa, and IIb, respectively. The line elements, representing the general solution to the problem under consideration, are given below.

Class I: $\rho^2 - 4\sigma\bar{\sigma} = 0$:

$$ds^2 = -r^{\frac{3}{2}}\left(d\xi - 2r^{-\frac{1}{2}}\frac{\partial H}{\partial \xi}\,dv\right)^2 - r^{\frac{1}{2}}\left(d\eta + 2r^{\frac{1}{2}}\frac{\partial H}{\partial \eta}\,dv\right)^2 + 2H\,dr\,dv - B^{-2}H^2\,dv^2,$$

where

$$B^{-2} = 4[a^2(v)r^{\frac{1}{2}} + b^2 r^{\frac{3}{2}}], \quad b = \text{const},$$
$$H = L_1(\eta, v)\cos a(v)\xi + L_2(\eta, v)\sin a(v)\xi,$$

with $a(v) > 0$ as a nonconstant function. The functions $L_\alpha(\eta, v)$, $\alpha = 1, 2$, which are not permitted to vanish simultaneously, are each expressed in terms of two arbitrary functions $E_\alpha(v)$ and $F_\alpha(v)$ by[16]

$$L_\alpha(\eta, v) = E_\alpha(v)\cos[b\eta + F_\alpha(v)], \quad b \neq 0,$$
$$= E_\alpha(v)\eta + F_\alpha(v), \quad b = 0.$$

The fluid pressure and energy density assume the form

$$p = (\tfrac{3}{4})r^{-\frac{3}{2}}[a^2(v) - 7b^2 r],$$
$$A - p = 12b^2 r^{-\frac{1}{2}}. \tag{4.7}$$

In deriving this solution we have imposed the condition $p \leq A$ [see (1.3)]. If $b = 0$, the requirement $p > 0$ is also satisfied for all permissible values (4.4) of r with $a(v)$ essentially[17] arbitrary. However, if $b \neq 0$, it is necessary to restrict the function $a(v)$ and the domain (4.4) of r according to $7b^2\beta \leq a^2(v)$, $r < \beta$, where β is an arbitrarily assigned positive constant.

Class IIa: $\rho^2 - 4\sigma\bar{\sigma} < 0$:

$$ds^2 = -(k^2 - r^2)R^{-1}\left(d\xi - Rk^{-1}\frac{\partial H}{\partial \xi}\,dv\right)^2 - (k^2 - r^2)R\left(d\eta + R^{-1}k^{-1}\frac{\partial H}{\partial \eta}\,dv\right)^2 + 2H\,dr\,dv - B^{-2}H^2\,dv^2,$$

where

$$B^{-2} = 2(k^2 - r^2)^{\frac{1}{2}}[k^{-1}a^2(v) - b^2(r + k)],$$
$$b = \text{const},$$
$$H = L_1(\eta, v)\cos a(v)\xi + L_2(\eta, v)\sin a(v)\xi,$$
$$R(r) = (k - r)^{\frac{1}{2}}(k + r)^{-\frac{1}{2}},$$

with $a(v) > 0$ as a nonconstant function. The functions $L_\alpha(\eta, v)$, $\alpha = 1, 2$, which are not permitted to vanish simultaneously, are each expressed in terms of two arbitrary functions $E_\alpha(v)$ and $F_\alpha(v)$ according to[16]

$$L_\alpha(\eta, v) = \begin{cases} E_\alpha(v)\cos\{[a^2(v) - 2b^2k^2]^{\frac{1}{2}}\eta + F_\alpha(v)\}, & a^2(v) > 2b^2k^2, \\ E_\alpha(v)\cosh\{[2b^2k^2 - a^2(v)]^{\frac{1}{2}}\eta + F_\alpha(v)\}, & a^2(v) < 2b^2k^2. \end{cases}$$

The pressure and energy density are given as

$$p = (\tfrac{3}{2})(k^2 - r^2)^{-\frac{3}{2}}[b^2(4r^3 - 5k^2r - k^3) + ka^2(v)],$$
$$A - p = 12b^2r(k^2 - r^2)^{-\frac{1}{2}}. \tag{4.8}$$

In the case $a^2(v) > 2b^2k^2$, the requirement $p > 0$ holds for all permissible values (4.5) of r provided that $a(v)$ satisfies the additional restriction $a^2(v) > [1 + (\frac{5}{3})^{\frac{3}{2}}]k^2b^2$. If $b = 0$, $a(v)$ is essentially[17] arbitrary. In the case $a^2(v) < 2b^2k^2$ ($\Rightarrow b \neq 0$) it is necessary to restrict the function $a(v)$ and the domain of r according to $a^2(v) \geq \alpha k^2b^2$, $r < \beta k$, where, for fixed $1 < \alpha < 2$, $\beta < 1$ is the unique positive number satisfying[18] $4\beta^3 - 5\beta + \alpha - 1 = 0$.

Class IIb: $\rho^2 - 4\sigma\bar{\sigma} > 0$:

$$ds^2 = -(r^2 - k^2)R^{-1}\left(d\xi + k^{-1}R\frac{\partial H}{\partial \xi}\,dv\right)^2 - (r^2 - k^2)R\left(d\eta - k^{-1}R^{-1}\frac{\partial H}{\partial \eta}\,dv\right)^2 + 2H\,dr\,dv - B^{-2}H^2\,dv^2,$$

where

$$B^{-2} = 2(r^2 - k^2)^{\frac{1}{2}}[b^2(r + k) - \epsilon k^{-1}a^2(v)],$$
$$b = \text{const},$$
$$R(r) = (r - k)^{\frac{1}{2}}(r + k)^{-\frac{1}{2}},$$

with $a(v) > 0$ as a nonconstant function. The nonvanishing function $H(\xi, \eta, v)$ is determined in terms of arbitrary functions $E_\alpha(v)$, $F_\alpha(v)$, $\alpha = 1, 2$, its explicit form depending on $\epsilon = \pm 1$ according to[16] the following.

(i) $\epsilon = +1$:

$$H = L_1(\eta, v)\cos a(v)\xi + L_2(\eta, v)\sin a(v)\xi,$$
$$L_\alpha(\eta, v) = E_\alpha(v)\cos\{[2b^2k^2 - a^2(v)]^{\frac{1}{2}}\eta + F_\alpha(v)\}, \quad a^2(v) < 2b^2k^2.$$

(ii) $\epsilon = -1$:

$$H = L_1(\eta, v)\cosh a(v)\xi + L_2(\eta, v)\sinh a(v)\xi,$$
$$L_\alpha(\eta, v) = E_\alpha(v)\cos\{[a^2(v) + 2b^2k^2]^{\frac{1}{2}}\eta + F_\alpha(v)\}.$$

For corresponding ϵ, the pressure and energy density are

$$p = (\tfrac{3}{2})(r^2 - k^2)^{-\frac{3}{2}}[b^2(-4r^3 + 5k^2r + k^3) - \epsilon ka^2(v)],$$
$$A - p = 12b^2r(r^2 - k^2)^{-\frac{1}{2}}. \tag{4.9}$$

In the case $\epsilon = -1$, the requirement $p > 0$ holds with $b \neq 0$, provided we restrict the function $a(v)$ and the domain (4.6) of r according to $\alpha k^2b^2 \leq a^2(v)$ and $r < \beta k$, where, for a fixed arbitrary positive number α, we have a unique $\beta > 1$ satisfying[19] $4\beta^3 - 5\beta - 1 - \alpha = 0$. There are, however, no restrictions on $a(v)$ or the domain of r if $b = 0$.[17]

In the case $\epsilon = +1$ ($\Rightarrow b \neq 0$), the function $a(v$ and the domain (4.6) of r are restricted by the inequalities $[a(v)/b]^2 \leq \alpha k^2$, $r < k\beta$, where $\alpha < 2$ determines $\beta > 1$ as the unique real number satisfying $4\beta^3 - 5\beta + \alpha - 1 = 0$.

In connection with these restrictions, we should note that some authors[20] impose, instead of (1.3), the stronger condition $0 < p \leq \frac{1}{3}A$ on the fluid pressure. Solutions with $b = 0$ are then immediately excluded from our line elements, but all others with suitable functions $a(v)$ survive. For example, Class I will then require that the function $a(v)$ and the domain (4.4) of r be restricted by the inequalities $0 < 7b^2\beta \leq a^2(v) \leq 15b^2\alpha$, $0 < \alpha \leq r < \beta$, where α and β are positive constants which satisfy $1 < \beta/\alpha < \frac{15}{7}$ but are otherwise arbitrary.

The Killing equations were solved for the above line elements. It was found that, provided $a'(v) \neq 0$, (which we have assumed) the only metrics which admit a nontrivial group of isometries are contained in the subcase of Class I for which $b = 0$. For example, if

$$H = \cos a(v)\xi \Rightarrow b = 0,$$

this metric admits a one-parameter group of isometries, unless $a(v)$ satisfies

$$(a/a')' = -\tfrac{4}{3},$$

in which case there are two independent Killing vector fields, this being the maximum number possible. As a

second example, if

$$H = \eta \cos a(v)\xi \Rightarrow b = 0,$$

there are no Killing vector fields unless $a(v)$ satisfies

$$(a/a')' = -1,$$

in which case there is one. Finally, we note that if $a'(v) = 0$, the line elements of all three classes are conformally flat (see Sec. 5) and admit a six-parameter group of isometries.

5. PROPERTIES OF THE WEYL TENSOR AND FLUID CONGRUENCE

For type [4] gravitational fields, the Weyl conform tensor defined by

$$C_{abcd} = R_{abcd} + g_{a[c}R_{d]b} + R_{a[c}g_{d]b} - (R/3)g_{a[c}g_{d]b},$$

with $R_{ab} = R^{c}{}_{abc}$, can be written in terms of the null tetrad of Sec. 2 as

$$C_{abcd} = -4(CV_{ab}V_{cd} + \bar{C}\bar{V}_{ab}\bar{V}_{cd}), \tag{5.1}$$

where

$$V_{ab} = k_{[a}m_{b]},$$

and C is a scalar, in general complex.[21]

For the class of solutions considered here, it is a consequence of the field equations and the Bianchi identities that the function C is proportional to the shear of the p.n.c. and to the shear scalar $(\sigma_{ab}\sigma^{ab})^{\frac{1}{2}}$ of the fluid congruence, the latter being defined by[22]

$$\sigma_{ab} = u_{(a;b)} - \dot{u}_{(a}u_{b)} - (\theta/3)(g_{ab} - u_a u_b), \tag{5.2}$$

where

$$\theta = u^{a}{}_{;a}, \quad \dot{u}_a = u_{a;b}u^b \tag{5.3}$$

are the expansion and acceleration of the fluid, respectively. The explicit formula, derived in the Appendix, is

$$2C^2 = 3\sigma^2 B^{-6}\sigma_{ab}\sigma^{ab}, \tag{5.4}$$

so that C is real. When the expressions of Sec. 4 are substituted into (5.4), one obtains[23]

$$C = a(v)a'(v)F(r)/H, \tag{5.5}$$

where

$$F(r) = \begin{cases} r^{-\frac{1}{2}}, & \text{Class I,} \\ (k^2 - r^2)^{-\frac{1}{2}}, & \text{Class IIa,} \\ \epsilon(r^2 - k^2)^{-\frac{1}{2}}, & \text{Class IIb.} \end{cases} \tag{5.6}$$

Thus the Weyl tensor is nonzero if and only if the function $a(v)$ in the line elements is nonconstant (as assumed). In addition, the vectors k^a and m^a, determining the null bivector V_{ab}, are given by (3.7) and

$$m^a = 2^{-\frac{1}{2}}|F(r)|\,[f(r)\delta^a_1 + if(r)^{-1}\delta^a_2],$$

where $F(r)$ is defined by (5.6) and $f(r)$ by

$$f(r) = \begin{cases} r^{-\frac{1}{4}}, & \text{Class I,} \\ (k - r)^{\frac{1}{4}}(k + r)^{-\frac{1}{4}}, & \text{Class IIa,} \\ (r - k)^{\frac{1}{4}}(r + k)^{-\frac{1}{4}}, & \text{Class IIb.} \end{cases}$$

It is shown in the Appendix that V_{ab} is covariantly constant along the p.n.c. so that

$$C_{abcd;e}k^e = -4\frac{\partial C}{\partial r}(V_{ab}V_{cd} + \bar{V}_{ab}\bar{V}_{cd}).$$

The behavior of C_{abcd} along the p.n.c. is thus determined solely by the function $F(r)$ of Eq. (5.6) which gives the r dependence of C. Note[24] that, for the Robinson–Trautman type [4] *vacuum* solutions [whose repeated p.n.c. is hypersurface generating (and hence geodesic), shear-free, but expanding], the function C of (5.1) varies as r^{-1}.

We remarked earlier that the fluid congruence is necessarily hypersurface orthogonal. One also finds that it is further restricted by having the shear scalar related to the acceleration scalar according to

$$\dot{u}^a\dot{u}_a = -(\tfrac{3}{2})\sigma_{ab}\sigma^{ab}. \tag{5.7}$$

Finally, the expansion θ is calculated to be

$$\theta = -3B^{-2}\frac{\partial B}{\partial r} + H^{-1}\frac{\partial B}{\partial v}. \tag{5.8}$$

From Eq. (5.4), (5.7), and (5.8) we thus conclude that, *for the class of type [4] perfect fluid solutions under consideration, the fluid has nonzero shear, acceleration, and expansion.* For the conformally flat solutions [$a'(v) = 0$], the fluid has zero shear and acceleration but nonzero expansion, and satisfies an equation of state of the form $p = f(A)$. These solutions therefore belong to the Friedman class of perfect fluid space-times.[25]

We conclude by discussing the question as to whether a region of one of the type [4] perfect fluid solutions under consideration can be suitably joined to a vacuum region of space-time. It is customary to require the Lichnerowicz junction condition to hold across a fluid–vacuum interface Σ. This implies[26] that the pressure vanishes on Σ and that the fluid velocity u^a is tangent to Σ. We thus investigate the existence of a hypersurface on which $p = 0$, such that

$$\left[u^a\frac{\partial p}{\partial x^a}\right]_{p=0} = 0. \tag{5.9}$$

From Eqs. (4.7)–(4.9) it follows immediately that hypersurfaces on which p vanishes exist if and only if the constant b is nonzero. However, a straightforward calculation shows that for all the line elements with

$b \neq 0$ [and $a(v)$ nonconstant] the condition (5.9) is violated. We thus conclude that *none of the present class of type [4] perfect fluid solutions can be joined to a vacuum solution with the Lichnerowicz junction condition being satisfied across the fluid vacuum interface.*

ACKNOWLEDGMENTS

I am indebted to J. Wainwright for valuable discussions and suggestions throughout the course of this work. For making possible a computer check of some results, I would like to thank J. Bolce and R. Hurdal of the University of Waterloo Computing Center.

APPENDIX: DERIVATION OF MAIN RESULTS

In deriving Eqs. (3.2) and (3.3), it is not necessary to introduce a coordinate system explicitly. It is convenient, however, to use the tetrad freedom (2.3) with $S = T = 0$ to set $u^a k_a = 2^{-\frac{1}{2}}$, which means that the tangent field k^a no longer corresponds to an affine parameter (as in Sec. 3). Equation (4.1) simplifies to

$$u^a = 2^{-\frac{1}{2}}(k^a + n^a). \tag{A1}$$

By means of the field Eqs. (1.2), and (A1), the tetrad components[12] $\Phi_{\alpha\beta}$, $\alpha, \beta = 0, 1, 2$, of the Ricci tensor become

$$\begin{aligned} \Phi_{00} &= 2\Phi_{11} = \Phi_{22} = (\tfrac{1}{4})(A + p) > 0, \\ \Phi_{01} &= \Phi_{02} = \Phi_{12} = 0, \\ \Lambda &= R/24 = (A - 3p)/24. \end{aligned} \tag{A2}$$

Since k^a is tangent to the repeated p.n.c., which is assumed geodesic, we have

$$\kappa = 0, \Psi_0 = \Psi_1 = \Psi_2 = \Psi_3 = 0, \Psi_4 \neq 0, \tag{A3}$$

where Ψ_α, $\alpha = 0, 1, \cdots, 4$, are the tetrad components of the Weyl tensor.[12] We use the Ricci and Bianchi identities in Newman–Penrose form (Ref. 1, pp. 348–51), with the former labeled $I_1, I_2, \cdots, I_{18}$ and the latter $II_1, II_2, \cdots, II_{11}$. With the specializations (A2) and (A3), the Bianchi identities II_1–II_6 simplify to

$$-\delta\Phi_{00} = (\bar{\pi} - 2\bar{\alpha} - 2\beta)\Phi_{00}, \tag{A4}$$

$$0 = (\bar{\lambda} - \sigma)\Phi_{00} \Rightarrow \sigma = \bar{\lambda}, \tag{A5}$$

$$(D - \Delta)\Phi_{00} = (\bar{\mu} - 2\mu - 2\gamma - 2\bar{\gamma} + 2\bar{\rho} - \rho)\Phi_{00}, \tag{A6}$$

$$-\delta\Phi_{00} = (-\bar{\nu} + \tau + 2\bar{\pi})\Phi_{00}, \tag{A7}$$

$$\delta\Phi_{00} = (-2\bar{\nu} + \bar{\pi} + 2\tau)\Phi_{00}, \tag{A8}$$

$$\begin{aligned} (D - \Delta)\Phi_{00} &= 3\sigma\Psi_4 \\ &+ (2\bar{\mu} - \mu + \bar{\rho} - 2\rho - 2\epsilon - 2\bar{\epsilon})\Phi_{00}. \end{aligned} \tag{A9}$$

Equations (A4), (A7), and (A8) immediately imply

$$\alpha + \bar{\beta} = \pi - \nu + \bar{\tau} = 0. \tag{A10}$$

Since $(D - \Delta)\Phi_{00}$ is real, (A6) entails

$$\rho - \bar{\rho} + \mu - \bar{\mu} = 0, \tag{A11}$$

and, by means of (A5), (A10), and (A11), the imaginary part of the Ricci identity I_{12} reads

$$(\rho - \bar{\rho})(\rho - \bar{\mu} + \epsilon + \bar{\epsilon} - \gamma - \bar{\gamma}) = 0. \tag{A12}$$

Furthermore, by virtue of (A5) and (A10), the expression

$$D\sigma + (-D\bar{\lambda} + \delta\bar{\pi}) + (\Delta\bar{\lambda} - \delta\bar{\nu}) + (-\Delta\sigma + \delta\tau)$$

vanishes identically. When it is evaluated using the Ricci identities I_2, I_7, I_{10}, and I_{16}, one obtains for Ψ_4 the expression

$$\Psi_4 = 2\bar{\sigma}(\rho - \bar{\mu} + \epsilon + \bar{\epsilon} - \gamma - \bar{\gamma}). \tag{A13}$$

On the other hand, Eqs. (A6) and (A9) imply

$$0 = 3\sigma\Psi_4 - 2(\rho - \bar{\mu} + \epsilon + \bar{\epsilon} - \gamma - \bar{\gamma})\Phi_{00}, \tag{A14}$$

which with Eqs. (A12) and (A13) yields $(\rho - \bar{\rho})\Psi_4 = 0$ and $(\Phi_{00} - 3\sigma\bar{\sigma})\Psi_4 = 0$, verifying (3.2) and (3.3).

To proceed further, we substitute (A1) into the definition (5.2) of σ_{ab} and introduce the expressions for the covariant derivatives $k_{a;b}$, $n_{a;b}$ in terms of the spin coefficients to obtain[27]

$$\sigma_{ab}\sigma^{ab} = (\tfrac{1}{8})(\rho - \bar{\mu} + \epsilon + \bar{\epsilon} - \gamma - \bar{\gamma})^2,$$

which when combined with (A13) yields

$$\Psi_4^2 = 12\bar{\sigma}^2\sigma_{ab}\sigma^{ab}.$$

If we use (2.3) with $R = 1$, $T = 0$ to set $\sigma = \bar{\sigma}$, and introduce an affine parameter, we obtain the required equation (5.4).

Retaining the nonaffine parameter, we next obtain further simplification of the spin coefficients. Equation II_7 immediately implies $\epsilon = \bar{\epsilon}$, while (A7) with (3.3) gives

$$2\delta\sigma = -\bar{\pi}\sigma. \tag{A15}$$

A linear combination of I_4, I_5, I_{11}, I_{13}, I_{15}, and I_{18} yields an expression for $\delta(\rho - \bar{\mu} + \epsilon + \bar{\epsilon} - \gamma - \bar{\gamma})$, which with (A13) and (A15) enables us to evaluate $\delta\Psi_4$. On comparison with II_8 we obtain

$$\bar{\pi} = 8\beta, \tag{A16}$$

which when substituted with (A15) in I_{11} and I_{13} implies

$$\delta\rho = \delta\mu = 0. \tag{A17}$$

The commutator[28] $(\delta D - D\delta)$ is applied to ρ and the operator δ is applied to I_1 yielding two expressions for $\delta D\rho$ [simplified by (A17)], which when equated necessitate

$$\alpha = \beta = \pi = 0, \Rightarrow \delta\epsilon = \delta\sigma = 0. \quad \text{(A18)}$$

To summarize, we have the following restrictions on the spin coefficients:

$$\kappa = \alpha = \beta = \pi = 0,$$
$$\rho - \bar{\rho} = \mu - \bar{\mu} = \sigma - \bar{\sigma} = \lambda - \bar{\lambda} = 0, \quad \text{(A19)}$$
$$\epsilon - \bar{\epsilon} = \sigma - \bar{\lambda} = \nu - \bar{\tau} = 0. \quad \text{(A20)}$$

Before completing the integration of the field Eqs. in Newman–Penrose form, it is essential to use the transformation (2.3) with $S = T = 0$ to re-introduce an affine parameter (characterized by $\epsilon + \bar{\epsilon} = 0$) along the p.n.c. We thus require that

$$(\epsilon \hat{+} \bar{\epsilon}) = R(\epsilon + \bar{\epsilon}) + DR = 0. \quad \text{(A21)}$$

Such a transformation will in general not preserve $\alpha = \bar{\beta} = 0$ since

$$(\alpha \hat{+} \bar{\beta}) = \delta \ln R. \quad \text{(A22)}$$

However, because we have $\delta(\epsilon + \bar{\epsilon}) = 0$, we can solve (A21) for R, subject to $\delta R = 0$, and thus achieve $\epsilon + \bar{\epsilon} = 0$ while preserving $\bar{\alpha} = \beta = 0$.[29] If we write $R = 2^{\frac{1}{2}}B$, Eq. (A1) assumes the form (4.1). The conditions (A19) are left unchanged while (A20) becomes

$$\epsilon = 0, \quad \sigma = 2B^2\bar{\lambda}, \quad \tau = 2B^2\bar{\nu}. \quad \text{(A23)}$$

With the aid of (3.3), the Ricci identities I_1 and I_2 now reduce to (3.4) and (3.5), and the integration can be carried out as described following Eq. (4.2).[30]

Finally, note that, as an immediate consequence of their definitions,[12] the spin coefficients $\kappa = \pi = \epsilon = 0$ immediately imply that the complex null bivector V_{ab} of Eq. (5.1) satisfies $V_{ab;c}k^c = 0$, as mentioned in Sec. 5.

* This work was supported by the National Research Council of Canada.

[1] See, for example, F. A. E. Pirani, *Brandeis Summer Institute, 1964* (Prentice-Hall, Englewood Cliffs, N.J., 1965), pp. 320–23.

[2] W. Kundt and M. Trümper, Akad. Wiss. Mainz **12**, 968 (1962).

[3] P. Szekeres, J. Math. Phys. **7**, 751 (1966).

[4] R. Penrose, Ann. Phys. **10**, 171 (1960).

[5] The Goldberg–Sachs theorem is stated, for example, in Ref. 1, p. 345. For a brief discussion of the terminology relevant to the theorem see Sec. 2 of the present paper.

[6] The units are chosen so that the gravitational constant is $(8\pi)^{-1}$ and the speed of light in vacuo is 1.

[7] See, for example, Ref. 1, pp. 337–42. We will use ";", "( )", and "[ ]" to denote covariant differentiation, symmetrization, and skew-symmetrization, respectively.

[8] For further details see, for example, R. Penrose, *Battelle Rencontres*, C. DeWitt and J. Wheeler, Eds. (Benjamin, New York, 1968), pp. 165–67.

[9] This can be achieved using the tetrad transformation (2.3) with $R = 1$, $T = 0$.

[10] For a spinor derivation see Ref. 8.

[11] The condition $\Psi_0 = 0$ is necessary and sufficient in order that k^a be parallel to a p.n.d. (not necessarily repeated) of C_{abcd}. For a more detailed explanation see Ref. 12.

[12] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[13] I. Robinson and A. Trautman, Proc. Roy. Soc. London **A265**, 463 (1961/62).

[14] E. T. Newman and L. A. Tambourino, J. Math. Phys. **3**, 902 (1962). Once the simplifications of the spin coefficients given by (A19) and (A23) have been achieved, the integration procedure follows very closely the method of the above paper. Thus, the details are omitted.

[15] One can in fact set k equal to 1.

[16] The freedom $v \to f(v)$ in the choice of v may be used to fix one of the functions $E_1(v)$, $E_2(v)$, $F_1(v)$, and $F_2(v)$.

[17] Aside from being a nonconstant positive function, $a(v)$ is arbitrary.

[18] One finds, more strongly, that the cubic equation requires $\beta < \frac{1}{2}(2^{\frac{1}{2}} - 1)$.

[19] If $1 < \beta < \frac{1}{2}(2^{\frac{1}{2}} + 1)$, we can, however, take $a(v)$ essentially arbitrary.

[20] See, for example, Ref. 3, p. 758.

[21] See Ref. 3, p. 753.

[22] As defined in J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).

[23] The question of sign in obtaining (5.5) from (5.4) was resolved by calculating C directly from the Eq. (A13) using the values of the spin coefficients generated by the integration of the Newman–Penrose equations.

[24] See Ref. 13. For a general summary of the behavior of the Weyl tensor in *vacuo* along a repeated principal null congruence, see Ref. 1, p. 364.

[25] See, for example, L. C. Shepley and A. H. Taub, Commun. Math. Phys. **5**, 237 (1967).

[26] A. Lichnerowicz, *Théories relativites de la gravitation et de l'électromagnétisme* (Masson, Paris, 1955), pp. 61–64.

[27] See p. 339 of Ref. 1. Note that $k_{a;b}$ must be modified slightly since we are not using an affine parameter.

[28] The complete set of commutators is listed in Ref. 12, p. 570.

[29] With the simplifications (A19) and (A20) we could use the Ricci equations I_2, I_7, and I_8 to obtain explicitly a suitable function R to satisfy (A21), with (A22) vanishing in view of (A17) and (A18). One finds that $R = \sigma\mu^2$ for $\mu \neq 0$ and $R = \sigma^{\frac{1}{3}}$ for $\mu = 0$.

[30] To preserve (A19) and (A23), the function R is now restricted by $DR = \delta R = 0$. We note that the remaining freedom in the choice of tetrad was used to simplify an integration constant in $\gamma + \bar{\gamma}$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Finite Transformations in Various Representations of $SU(3)$

S. P. ROSEN
Purdue University, Lafayette, Indiana

(Received 20 August 1970)

A projection operator expansion is used to obtain the explicit forms of finite transformations in the triplet, octet, and decuplet representations of $SU(3)$. The projection operators are obtained from the characteristic equations for the matrices in these representations. These characteristic equations are developed from known properties of the λ_i, F_i, and decuplet matrices; where multiple eigenvalues appear, it is shown that the relevant matrices also satisfy reduced equations in which no eigenvalue appears twice or more. General features of the method are discussed.

1. INTRODUCTION

The eightfold way[1] and related particle symmetries are usually discussed in terms of infinitesimal $SU(3)$ transformations. These transformations are easy to use, and they provide us with enough mathematical apparatus to derive most consequences of the theory. As a result, the art of transforming quantities through finite angles in $SU(3)$ space has been somewhat neglected. Since there are occasions where such transformations cannot be avoided, we describe in this paper a general method for handling them. The method can be applied to all representations of $SU(3)$, but for practical reasons we restrict ourselves to the triplet, octet, and decuplet.

Our approach to the problem is akin to an eigenfunction expansion: If we can expand the quantity we wish to transform in terms of the eigenfunctions of the transformation operator, then we can carry out the transformation simply by replacing the operator by its appropriate eigenvalues. Now it often happens that, while we know the eigenvalues of an operator, we do not know the eigenfunctions. To avoid the labor of computing them explicitly, we shall therefore make our expansion in terms of a complete set of projection operators, one for each eigenvalue. As we shall see, these projection operators depend only upon the original transformation operator and its eigenvalues.

To obtain the projection operators, we consider an n-dimensional vector space spanned by vectors

$$\mathbf{a} = (a_1, a_2, \cdots, a_n). \tag{1.1}$$

Finite transformations in this space take the form

$$\mathbf{a} \to \mathbf{a}' = e^{i\theta A}\mathbf{a}, \tag{1.2}$$

where A is an $n \times n$ Hermitian matrix and θ is a parameter. By the Cayley–Hamilton theorem,[2] A must obey the characteristic equation

$$\Omega \equiv \prod_{k=1}^{n} (A - \alpha_k) = 0, \tag{1.3}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_n$ are its eigenvalues. When the eigenvalues are all distinct, we define the matrices

$$P_l(A) = \prod_{k \neq l} \left(\frac{A - \alpha_k}{\alpha_l - \alpha_k}\right), \quad l = 1, 2, \cdots, n, \tag{1.4}$$

and observe that, as a result of Eq. (1.3),

$$P_l(A)A = AP_l(A) = \alpha_l P_l(A), \quad \text{no sum on } l,$$
$$P_l(A)f(A) = f(A)P_l(A) = f(\alpha_l)P_l(A), \tag{1.5}$$

where f is any function of A. In particular, when we take f to be $P_m(A)$, we find that

$$P_m(A)P_l(A) = \delta_{ml}P_l(A). \tag{1.6}$$

We can also use the partial fraction decomposition of Ω^{-1} [see Eq. (1.3)] as a formal identity to prove that

$$\sum_{l=1}^{n} P_l(A) = I, \tag{1.7}$$

where I is the unit $n \times n$ matrix. Thus the $P_l(A)$ form a complete set of projection operators, and by virtue of Eqs. (1.5) and (1.7) we obtain the desired expansion for $\mathbf{a}'$ in Eq. (1.2):

$$\mathbf{a}' = e^{i\theta A}\mathbf{a} = \sum_{l=1}^{n} e^{i\theta\alpha_l} P_l(A)\mathbf{a}. \tag{1.8}$$

When the eigenvalues of A are not all distinct, we cannot use this method as it stands because the denominator in Eq. (1.4) will have zero as a factor. However, we can modify it quite easily if A satisfies not only the full Eq. (1.3), which contains repeated factors, but also a reduced characteristic equation with no repeated factors. Let us therefore suppose that A satisfies the equation

$$\prod_{k=1}^{m} (A - \alpha_k) = 0, \tag{1.9}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_m$, $m < n$, are the distinct eigenvalues. We can then define projection operators exactly as in Eq. (1.4), except that k and l both run from 1 to

m instead of 1 to n. The completeness relation of Eq. (1.9) is modified in exactly the same way, and so the expansion of Eq. (1.8) becomes

$$\mathbf{a}' = e^{i\theta A}\mathbf{a} = \sum_{l=1}^{m} e^{i\theta\alpha_l} P_l(A)\mathbf{a}. \tag{1.10}$$

In the next three sections of the paper, we study the characteristic equations for the three-, eight-, and ten-dimensional representations of $SU(3)$, and we show that, whenever a repeated root occurs, the appropriate matrix satisfies a reduced equation. To carry out this program, we shall need certain properties of the $SU(3)$ matrices and of the spaces in which they operate; since many of the properties have already been discussed by different authors,[3-7] we shall review them without giving proofs.

In the fifth section of the paper we consider finite transformations. From the results of the previous sections, we give explicit formulas for the triplet, octet, and decuplet representations of $SU(3)$.

2. CHARACTERISTIC EQUATIONS FOR THE TRIPLET

The three-dimensional representation of $SU(3)$ consists of matrices λ_i, $i = 1, 2, \cdots, 8$, which obey the standard commutation and anticommutation rules[1]

$$\begin{aligned}[\lambda_i, \lambda_j] &= 2if_{ijk}\lambda_k, \\ \{\lambda_i, \lambda_j\} &= \tfrac{4}{3}\delta_{ij}I + 2d_{ijk}\lambda_k.\end{aligned} \tag{2.1}$$

These matrices are traceless and normalized so that

$$\mathrm{Tr}\,(\lambda_i) = 0, \quad \mathrm{Tr}\,(\lambda_i\lambda_j) = 2\delta_{ij}. \tag{2.2}$$

Each of them satisfies the identity

$$M^3 - \tfrac{1}{2}\,\mathrm{Tr}\,(M^2)M - \tfrac{1}{3}\,\mathrm{Tr}\,(M^3)I \equiv 0, \tag{2.3}$$

as can be shown with the aid of the Cayley–Hamilton theorem.[2] We shall use this result to study the characteristic equations of the λ_i, but, before we do so, we must learn something about the geometry of the octet.

The octet space of $SU(3)$ is spanned by real vectors $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_8)$. As is well known, we can construct two invariants and a dual vector from $\boldsymbol{\pi}$:

$$X = \pi_i\pi_i, \quad Y = d_{ijk}\pi_i\pi_j\pi_k, \tag{2.4}$$

$$\Pi_i = d_{ijk}\pi_j\pi_k. \tag{2.5}$$

The invariant X is the square of the norm of $\boldsymbol{\pi}$, and it is never negative; the invariant Y, on the other hand, is the inner product of $\boldsymbol{\pi}$ and its dual vector $\boldsymbol{\Pi}$,

$$Y = \pi_i\Pi_i, \tag{2.6}$$

and it can be positive, negative, or zero. To determine the properties of $\boldsymbol{\Pi}$, we use a property of the d-coefficients,[3,5]

$$\begin{aligned}d_{ij\alpha}d_{kl\alpha} + d_{ik\alpha}d_{jl\alpha} + d_{il\alpha}d_{jk\alpha} \\ = \tfrac{1}{3}(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}),\end{aligned} \tag{2.7}$$

which can be derived from Eq. (2.3). We then find that the norm of $\boldsymbol{\Pi}$ is proportional to X,

$$\Pi_i\Pi_i = \tfrac{1}{3}X^2, \tag{2.8}$$

and that its dual is a linear combination of $\boldsymbol{\pi}$ and itself,

$$\begin{aligned}d_{ijk}\pi_j\Pi_k &= \tfrac{1}{3}X\pi_i, \\ d_{ijk}\Pi_j\Pi_k &= \tfrac{2}{3}Y\pi_i - \tfrac{1}{3}X\Pi_i.\end{aligned} \tag{2.9}$$

Because the vector $\boldsymbol{\Pi} + \mu\boldsymbol{\pi}$ must have a nonnegative norm for all values of the parameter μ, the value of the invariant Y is always bounded by

$$X^3 - 3Y^2 \geq 0. \tag{2.10}$$

When Y vanishes, the vector $\boldsymbol{\pi}$ is orthogonal to its dual [Eq. (2.7)], and it is called an "s-vector" by Michel and Radicatti.[6] When Y reaches its maximum absolute value, $\boldsymbol{\pi}$ is parallel to its dual,

$$Y = \pm(X^3/3)^{\frac{1}{2}}, \quad \Pi_i = \pm(X/3)^{\frac{1}{2}}\pi_i, \tag{2.11}$$

and it is known as a "q-vector."[6] The significance of this classification will become apparent below.

We now define the traceless matrix

$$(\lambda\cdot\pi) = \sum_i \lambda_i\pi_i. \tag{2.12}$$

Since it satisfies the identity of Eq. (2.3), its characteristic equation is

$$(\lambda\cdot\pi)^3 - X(\lambda\cdot\pi) - \tfrac{2}{3}YI = 0 \tag{2.13}$$

and its eigenvalues satisfy

$$\alpha^3 - X\alpha - \tfrac{2}{3}YI = 0. \tag{2.14}$$

This equation will have three real roots as long as the condition of Eq. (2.10) is fulfilled. They are

$$\begin{aligned}\alpha_k = 2(X/3)^{\frac{1}{2}}\cos\tfrac{1}{3}(\varphi + 2k\pi), \quad k = 1, 2, 3, \\ \cos\varphi = 3^{\frac{1}{2}}Y/(X)^{\frac{3}{2}}.\end{aligned} \tag{2.15}$$

When Y is zero, the roots are

$$\alpha_1 = -\sqrt{X}, \quad \alpha_2 = 0, \quad \alpha_3 = +\sqrt{X}. \tag{2.16}$$

If we assume that $\boldsymbol{\pi}$ is a unit vector ($X = 1$), then Eq. (2.16) implies that the operator $\frac{1}{2}(\lambda\cdot\pi)$ has the same eigenvalue spectrum as does the isospin matrix $\frac{1}{2}\lambda_3$ in the standard representation.[1] In other words, when π is an s-vector, $\frac{1}{2}(\lambda\cdot\pi)$ is an element of an $SU(2)$ subalgebra.[6]

When the absolute value of Y is at a maximum, two of the three roots coincide with one another. The roots are

$$[\mp(X/3)^{\frac{1}{2}}, \mp(X/3)^{\frac{1}{2}}, \pm 2(X/3)^{\frac{1}{2}}] \quad \text{for} \quad Y = \pm(X^3/3)^{\frac{1}{2}}. \tag{2.17}$$

For $\boldsymbol{\pi}$ a unit vector with negative Y, the matrix $(\lambda \cdot \pi)$ has the same spectrum as λ_8 in the standard representation.[1] Therefore $(\lambda \cdot \pi)$ behaves as a hypercharge-type operator whenever $\boldsymbol{\pi}$ is a q-vector.[6]

In general $|Y|$ lies somewhere between zero and $(X^3/3)^{\frac{1}{2}}$ and $\boldsymbol{\pi}$ is neither an s-vector nor a q-vector. Consequently, the matrix $(\lambda \cdot \pi)$ has three distinct eigenvalues, but it is not an isospin-type operator. It is always possible, however, to find linear combinations of $\boldsymbol{\pi}$ and $\boldsymbol{\Pi}$ which do fall into the s- and q-vector categories; therefore, we can always convert an arbitrary matrix $(\lambda \cdot \pi)$ into an isospin or a hypercharge operator.

It remains for us to show that when $Y = \pm(X^3/3)^{\frac{1}{2}}$, the matrix $(\lambda \cdot \pi)$ satisfies a quadratic equation as well as the cubic one of Eq. (2.13). To do this, we multiply the second equation of (2.1) by $\pi_i \pi_j$ and sum over the indices i and j. We obtain

$$(\lambda \cdot \pi)^2 = \tfrac{2}{3} XI + (\lambda \cdot \Pi).$$

From Eq. (2.11) we can rewrite this as

$$(\lambda \cdot \pi)^2 = \tfrac{2}{3} XI \pm (X/3)^{\frac{1}{2}}(\lambda \cdot \pi), \quad Y = \pm(X^3/3)^{\frac{1}{2}},$$

or

$$[(\lambda \cdot \pi) \pm (X/3)^{\frac{1}{2}}][(\lambda \cdot \pi) \mp 2(X/3)^{\frac{1}{2}}] = 0, \qquad Y = \pm(X^3/3)^{\frac{1}{2}}. \tag{2.18}$$

This is exactly the equation we would expect from Eq. (2.17) if we were to drop one of the repeated roots.

Up to now we have been using the standard representation (3). We now turn to the conjugate representation $(\bar{3})$, whose matrices are related to those of the (3) representation by

$$\bar{\lambda}_i = -\lambda_i^{\mathrm{T}}, \tag{2.19}$$

where T denotes transpose. The $\bar{\lambda}_i$ satisfy the first equation of Eq. (2.1) and the conditions of Eqs. (2.2) and (2.3); but instead of the second equation of Eq. (2.1) they obey

$$\{\bar{\lambda}_i, \bar{\lambda}_j\} = \tfrac{4}{3}\delta_{ij} - 2d_{ijk}\bar{\lambda}_k. \tag{2.20}$$

Because of this, the representations λ_i and $\bar{\lambda}_i$ cannot be related to one another by unitary transformations, and hence they are not equivalent representations of $SU(3)$.

To obtain the characteristic equation for $(\bar{\lambda} \cdot \pi)$, we merely replace $(\lambda \cdot \pi)$ by $(-\bar{\lambda} \cdot \pi)$ in Eq. (2.13). The characteristic roots are therefore opposite in sign from those of $(\lambda \cdot \pi)$, but the multiplicities remain the same. In other words, the properties of the $(\bar{3})$ representation can be obtained directly from those of the (3) by the replacement

$$(\lambda \cdot \pi) \to -(\bar{\lambda} \cdot \pi), \qquad \alpha_k \to -\bar{\alpha}_k \tag{2.21}$$

in Eqs. (2.13) and (2.15)–(2.18).

3. THE OCTET

From the f and d coefficients of Eq. (2.1) we can construct eight-dimensional matrices

$$(F_i)_{mn} = -if_{imn}, \quad (D_i)_{mn} = d_{imn}, \tag{3.1}$$

which obey the commutation rules

$$[F_i, F_j] = if_{ijk}F_k, \quad [F_i, D_j] = if_{ijk}D_k. \tag{3.2}$$

The F_i form an eight-dimensional representation of $SU(3)$ and the D_i are operators transforming as an octet with respect to the group. Because D_i and F_i are respectively symmetric and antisymmetric matrices, we can rewrite the second equation of (3.2) as[3]

$$F_k D_l + F_l D_k = D_l F_k + D_k F_l = d_{klm}F_m. \tag{3.3}$$

From Eq. (2.7) and various Jacobi identities for λ-matrices, we can show that[3,5]

$$\{F_i, F_j\} + 3\{D_i, D_j\} = 2\delta_{ij}I, \tag{3.4}$$

$$\{F_i, F_j\} = \delta_{ij}I + 3d_{ijk}D_k - R_{ij},$$
$$(R_{ij})_{mn} = \delta_{im}\delta_{jn} + \delta_{in}\delta_{jm}. \tag{3.5}$$

Equations (3.3), (3.4), and (3.5) are the basic equations that we need to examine the characteristic equations for the octet matrices.

A. The General Equation

We begin with some preliminary results. Multiplying Eq. (3.3) by $\pi_k \pi_l$ and summing, we obtain

$$2(F \cdot \pi)(D \cdot \pi) = 2(D \cdot \pi)(F \cdot \pi) = (F \cdot \Pi), \tag{3.6}$$

where the notation $(A \cdot B)$ represents $\sum_k A_k B_k$. Since the quantities in Eq. (3.6) are matrices, we can operate with them on the column vector $\boldsymbol{\pi}$:

$$2(D \cdot \pi)_{\alpha\beta}(F \cdot \pi)_{\beta\gamma}\pi_\gamma = (F \cdot \Pi)_{\alpha\beta}\pi_\beta.$$

But, because F_k is totally antisymmetric,

$$(F \cdot \pi)_{\beta\gamma}\pi_\gamma = -if_{k\beta\gamma}\pi_k\pi_\gamma = 0,$$

and so

$$(F \cdot \Pi)_{\alpha\beta}\pi_\beta = if_{\alpha\rho\beta}\Pi_\rho\pi_\beta = 0. \tag{3.7}$$

From this result and Eq. (3.2) we find that

$$[F \cdot \pi, F \cdot \Pi] = 0. \tag{3.8}$$

We may now multiply Eq. (3.5) by $\pi_i \Pi_j$ and use Eqs. (3.8) and (2.9) to write

$$2(F \cdot \pi)(F \cdot \Pi) = Y + X(D \cdot \pi) + R_{ij}\pi_i \Pi_j. \quad (3.9)$$

To eliminate the last term of Eq. (3.9), we multiply by $(F \cdot \pi)$ and use the antisymmetry of F_k together with Eq. (3.7); then with the aid of Eq. (3.6) we obtain

$$[4(F \cdot \pi)^2 - X](F \cdot \pi)(D \cdot \pi) = Y(F \cdot \pi). \quad (3.10)$$

To eliminate $(D \cdot \pi)$ from this equation, we note that from Eq. (3.4) we have

$$(F \cdot \pi)^2 + 3(D \cdot \pi)^2 = XI. \quad (3.11)$$

We therefore square both sides of Eq. (3.9) and find that[6]

$$(F \cdot \pi)^2\{[4(F \cdot \pi)^2 - X]^2[X - (F \cdot \pi)^2] - 3Y^2\} = 0. \quad (3.12)$$

Equation (3.12) is of degree eight in the (8×8) matrix $(F \cdot \pi)$, and hence it must be the characteristic equation that we would obtain from the Cayley–Hamilton theorem. We note that zero is a repeated eigenvalue and that the other six eigenvalues may or may not be distinct from one another. Thus we must show that $(F \cdot \pi)$ satisfies reduced equations if we are to develop projection operators along the lines described at the beginning of the paper. To pursue this question, we shall examine separately the three cases in which π is an s-vector, a q-vector, or neither.

B. Equation for s-Vectors

When π is an s-vector ($Y = 0$), the characteristic equation (3.12) has two single roots $(\pm\sqrt{X})$, and three double ones $(0, \pm\frac{1}{2}\sqrt{X})$. We therefore expect $(F \cdot \pi)$ to obey the reduced equation

$$(F \cdot \pi)[4(F \cdot \pi)^2 - X][X - (F \cdot \pi)^2] = 0. \quad (3.13)$$

To prove that it does indeed obey Eq. (3.13), we multiply Eq. (3.10) by $(D \cdot \pi)$ and use the fact that $Y = 0$ to write

$$(F \cdot \pi)[4(F \cdot \pi)^2 - X](D \cdot \pi)^2 = 0.$$

Equation (3.13) follows from this and Eq. (3.11).

C. Equation for q-Vectors

When π is a q-vector ($3Y^2 = X^3$), Eq. (3.12) has zero as a quadruple root, and $\pm(3^{\frac{1}{2}}X/2)$ as double roots. To derive the reduced equation in this case, we go back to Eq. (3.5) and multiply by $\pi_i\pi_j$ to obtain

$$2(F \cdot \pi)^2 = X + 3D \cdot \Pi - R_{ij}\pi_i\pi_j. \quad (3.14)$$

We eliminate the last term by multiplying by $(F \cdot \pi)$, and then with the aid of Eq. (2.11) we find that

$$2(F \cdot \pi)^3 = X(F \cdot \pi) + (3Y/X)(F \cdot \pi)(D \cdot \pi).$$

Using Eq. (3.6) and (2.11) again, we can rewrite this as[6]

$$(F \cdot \pi)[4(F \cdot \pi)^2 - 3X] = 0. \quad (3.15)$$

This is exactly the reduced equation we require since it contains each of the three distinct eigenvalues once and only once.

D. Equation for General Vectors

When π is neither an s-vector nor a q-vector, equation (3.12) has zero as a double root. To determine the other roots, we note that since the inequality of Eq. (2.10) holds, we can define

$$\sin\psi = (3Y^2/X^3)^{\frac{1}{2}}, \quad -\pi/2 < \psi < \pi/2. \quad (3.16)$$

The remaining roots can then be expressed as

$$\alpha_k = \cos\tfrac{1}{3}[\psi + (k-1)\pi], \quad k = 1, 2, 3, 4, 5, 6, \quad (3.17)$$

and they are all distinct from one another. Therefore, in this case we need only eliminate one of the $(F \cdot \pi)$ factors of Eq. (3.12).

To do this, we write

$$\begin{aligned}(F \cdot \pi)[4(F \cdot \pi)^2 - X]^2(D \cdot \pi)^2 &\equiv \{(F \cdot \pi)[4(F \cdot \pi)^2 - X](D \cdot \pi)\} \\ &\quad \times \{[4(F \cdot \pi)^2 - X](D \cdot \pi)\} \\ &= Y(F \cdot \pi)\{[4(F \cdot \pi)^2 - X](D \cdot \pi)\} \\ &= Y^2(F \cdot \pi), \end{aligned} \quad (3.18)$$

where we have used Eq. (3.10) twice. From Eq. (3.18) and (3.11), we obtain the required equation, namely

$$(F \cdot \pi)\{[4(F \cdot \pi)^2 - X]^2[X - (F \cdot \pi)^2] - 3Y^2\} = 0. \quad (3.19)$$

This completes our discussion of the octet matrices.

4. THE TEN-DIMENSIONAL REPRESENTATION

Besides the standard commutation rules

$$[B_i, B_j] = if_{ijk}B_k, \quad (4.1)$$

the matrices B_i of the ten-dimensional representation also satisfy[8]

$$d_{ijk}B_iB_j = \tfrac{3}{2}B_k \quad (4.2)$$

and

$$\sum_{i=1}^{8} B_iB_i = 6I, \quad (4.3)$$

where I is the 10×10 unit matrix. We can use these properties to construct symmetric second- and third-rank tensors

$$T_{ij} = \{B_i, B_j\} - \tfrac{9}{5}d_{ijk}B_k - \tfrac{3}{2}\delta_{ij}I, \tag{4.4}$$

$$\begin{aligned} T_{ijk} = {} & \sum_{\text{all permutations}} B_iB_jB_k \\ & - \tfrac{15}{7}(\delta_{ij}B_k + \delta_{jk}B_i + \delta_{ki}B_j) \\ & - \tfrac{27}{7}(d_{ij\rho}B_\rho B_k + d_{jk\rho}B_\rho B_i + d_{ki\rho}B_\rho B_j) \\ & + \tfrac{162}{35}d_{ijk}I \end{aligned} \tag{4.5}$$

which are traceless:

$$T_{\alpha\alpha} = d_{i\alpha\beta}T_{\alpha\beta} = d_{i\alpha\beta}T_{\alpha\beta k} = T_{\alpha\alpha k} = 0. \tag{4.6}$$

Since there are 27 independent T_{ij} and 64 independent T_{ijk}, these tensors combine with the nine matrices B_i and I to form a complete set in the space of 10×10 Hermitian matrices, and any member of this space can be written as a linear combination of them. In particular, the fourth-rank tensor

$$S_{ijkl} = T_{ijk}B_l + T_{jkl}B_i + T_{kli}B_j + T_{lij}B_k \tag{4.6'}$$

can be expressed as[9]

$$\begin{aligned} S_{ijkl} = {} & \{\tfrac{5}{7}(d_{ij\alpha}T_{\alpha kl} + d_{kl\alpha}T_{\alpha ij}) + \tfrac{38}{49}(\delta_{ij}T_{kl} + \delta_{kl}T_{ij}) \\ & - \tfrac{12}{49}d_{ij\alpha}d_{kl\beta}T_{\alpha\beta} + \text{cyclic permutations of } j, k, l\}. \end{aligned} \tag{4.7}$$

From this result we shall deduce the characteristic equation for $(B \cdot \pi) = \sum_{i=1}^{8} B_i\pi_i$.

A. The General Equation

Multiplying both sides of Eq. (4.7) by $\pi_i\pi_j\pi_k\pi_l$ and summing over the indices, we obtain an equation for $(B \cdot \pi)$ and $(B \cdot \Pi)$:

$$\begin{aligned} 3(B \cdot \Pi)^2 + [3X - 12(B \cdot \pi)^2](B \cdot \Pi) + 4(B \cdot \pi)^4 \\ - 4X(B \cdot \pi)^2 + 6Y(B \cdot \pi) = 0. \end{aligned} \tag{4.8}$$

Next we multiply Eq. (4.7) by $\pi_i\pi_j\pi_k\Pi_l$, and after some tedious algebra we obtain another equation:

$$\begin{aligned} -6(B \cdot \pi)(B \cdot \Pi)^2 + [4(B \cdot \pi)^3 - X(B \cdot \pi) + 3Y] \\ \times (B \cdot \Pi) - 2X(B \cdot \pi)^3 - 2Y(B \cdot \pi)^2 + 2X^2(B \cdot \pi) \\ = 0. \end{aligned} \tag{4.9}$$

To eliminate $(B \cdot \Pi)^2$, we multiply Eq. (4.8) by $2(B \cdot \pi)$ and add the resultant to Eq. (4.9):

$$\begin{aligned} [20(B \cdot \pi)^3 - 5X(B \cdot \pi) - 3Y](B \cdot \Pi) \\ = 2(B \cdot \pi)[4(B \cdot \pi)^4 - 5X(B \cdot \pi)^2 \\ + 5Y(B \cdot \pi) + X^2]. \end{aligned} \tag{4.10}$$

We can now eliminate $(B \cdot \Pi)$ altogether by multiplying Eq. (4.8) by $[20(B \cdot \pi)^3 - 5X(B \cdot \pi) - 3Y]^2$ and making use of Eq. (4.10). After some lengthy manipulation we obtain a characteristic equation of the tenth degree:

$$\begin{aligned} (B \cdot \pi)[4(B \cdot \pi)^3 - 3(B \cdot \pi) + \cos\psi] \\ \times [4(B \cdot \pi)^3 - 3(B \cdot \pi) - \cos\psi] \\ \times [4(B \cdot \pi)^3 - 9(B \cdot \pi) - 3^{\frac{3}{2}}\sin\psi] = 0, \end{aligned} \tag{4.11}$$

where we have set $X = 1$ and $3^{\frac{1}{2}}Y = \sin\psi$ [see Eq. (3.16)].

When $|Y|$ lies between zero and its maximum value, Eq. (4.11) has ten distinct roots. Seven of them, namely

$$\alpha_0 = 0, \quad \alpha_k = \cos\tfrac{1}{3}[\psi + (k-1)\pi], \quad k = 1, 2, 3, 4, 5, 6, \tag{4.12}$$

are the same as in the eight-dimensional representation [see Eqs. (3.17) and (3.19)], but the remaining three are new:

$$\alpha_k = 3^{\frac{1}{2}}\cos\tfrac{1}{6}[2\psi + (4k-1)\pi], \quad k = 7, 8, 9. \tag{4.13}$$

When Y is zero, there are three double roots $(0, \pm\frac{1}{2})$ and four single roots $(\pm 1, \pm\frac{3}{2})$; this spectrum is exactly the same as that of T_3 in the tenfold representation. When $Y = -1/\sqrt{3}$, the eigenvalues of $Q = (2/\sqrt{3})(B \cdot \pi)$ run from $+1$ to -2 in unit steps, and their multiplicities decrease from fourfold for $(+1)$ to singlefold for (-2), again in unit steps. This spectrum matches exactly the spectra of hypercharge and electric charge.[1]

We now show that when $\boldsymbol{\pi}$ is either a q-vector or an s-vector, $(B \cdot \pi)$ obeys an appropriate reduced equation.

B. Equations for s-Vectors and q-Vectors

When Y is zero, Eq. (4.10) becomes

$$\begin{aligned} 5(B \cdot \pi)[4(B \cdot \pi)^2 - X](B \cdot \Pi) \\ = 2(B \cdot \pi)[4(B \cdot \pi)^2 - X][(B \cdot \pi)^2 - X], \end{aligned} \tag{4.14}$$

and $(B \cdot \Pi)$ can be eliminated from Eqs. (4.8) and (4.9) to yield

$$\begin{aligned} 15(B \cdot \pi)(B \cdot \Pi)^2 = 2(B \cdot \pi)[(B \cdot \pi)^2 - X] \\ \times [2(B \cdot \pi)^2 - 3X]. \end{aligned} \tag{4.15}$$

We now multiply Eq. (4.15) by $[4(B \cdot \pi)^2 - X]$ and use Eq. (4.14) twice. In this way we obtain an equation, namely

$$\begin{aligned} (B \cdot \pi)[4(B \cdot \pi)^2 - X][(B \cdot \pi)^2 - X] \\ \times [4(B \cdot \pi)^2 - 9] = 0, \end{aligned} \tag{4.16}$$

in which each eigenvalue appears once and only once.

When $Y = -1/\sqrt{3}$ and $X = 1$, we find from Eq. (2.11) that $\Pi_i = (-1/\sqrt{3})\pi_i$. Equation (4.8) becomes

$$(B \cdot \pi)[4(B \cdot \pi)^3 + (4\sqrt{3})(B \cdot \pi)^2 - 3(B \cdot \pi) - 3\sqrt{3}] = 0 \quad (4.17)$$

or, in terms of $Q = (2/\sqrt{3})(B \cdot \pi)$,

$$Q(Q + 2)(Q^2 - 1) = 0. \quad (4.18)$$

Again, the distinct eigenvalues appear only once in the equation. Thus, in both the s-vector and the q-vector cases, $(B \cdot \pi)$ obeys a reduced equation.

C. The Conjugate Decuplet

It is not difficult to show that the matrix $\bar{B}_i$, obtained by taking the negative transpose of B_i, obeys the standard $SU(3)$ commutation rules. However, because

$$\bar{B}_i = -B_i^{\mathrm{T}}, \quad (4.19)$$

the relation of Eq. (4.2) becomes[8]

$$d_{ijk}\bar{B}_i\bar{B}_j = -\tfrac{3}{2}\bar{B}_k, \quad (4.20)$$

and so the $\bar{B}_i$ give us a representation which is not equivalent to the original B_i. The characteristic equation for this conjugate decuplet is obtained by replacing $(B \cdot \pi)$ by $-(\bar{B} \cdot \pi)$ everywhere in the preceding discussion. As a result, the signs of the eigenvalues change, but the multiplicities remain the same.

5. FINITE TRANSFORMATIONS

We denote the infinitesimal generators of $SU(3)$ by $\hat{F}_i$, and any matrix representation of them by M_i. The operators and the matrices then satisfy the commutation rules

$$[\hat{F}_i, \hat{F}_j] = if_{ijk}\hat{F}_k \quad (5.1)$$

and

$$[M_i, M_j] = if_{ijk}M_k, \quad (5.2)$$

respectively.

A tensor operator Ψ_α is said to transform according to the representation M_i if

$$[\hat{F}_i, \Psi_\rho] = \Psi_\tau(M_i)_{\tau\rho}. \quad (5.3)$$

From this and the relation

$$e^{i\theta(\hat{F}\cdot\pi)}\Psi_\rho e^{-i\theta(\hat{F}\cdot\pi)} = \Psi_\rho + i\theta[\hat{F}\cdot\pi, \Psi_\rho] + \frac{(i\theta)^2}{2!}[\hat{F}\cdot\pi, [\hat{F}\cdot\pi, \Psi_\rho]] + \frac{(i\theta)^3}{3!}[\hat{F}\cdot\pi, [\hat{F}\cdot\pi, [\hat{F}\cdot\pi, \Psi_\rho]]] + \cdots,$$

$$(\hat{F}\cdot\pi) = \sum_{k=1}^{8}\hat{F}_k\pi_k, \quad (5.4)$$

we conclude that

$$e^{i\theta(\hat{F}\cdot\pi)}\Psi_\rho e^{-i\theta(\hat{F}\cdot\pi)} = \Psi_\tau(e^{i\theta M\cdot\pi})_{\tau\rho}. \quad (5.5)$$

If α_k, $k = 1, 2, \cdots, m$, denote the distinct eigenvalues of $M \cdot \pi = \sum_{k=1}^{8} M_i\pi_i$ and if $P_l(M \cdot \pi)$ is defined as in Eqs. (1.4) and (1.5), we can rewrite Eq. (5.5) as

$$e^{i\theta(\hat{F}\cdot\pi)}\Psi_\rho e^{-i\theta(\hat{F}\cdot\pi)} = \sum_l \Psi_\tau[P_l(M \cdot \pi)]_{\tau\rho}e^{i\theta\alpha_l}. \quad (5.6)$$

This is our basic formula for finite transformations. We now apply it to the special cases of the octet, triplet, and decuplet; for convenience we shall assume from now on that $\boldsymbol{\pi}$ is a unit vector,

$$X = \sum_{i=1}^{8}\pi_i\pi_i = 1. \quad (5.7)$$

A. The Triplet

In the three-dimensional representation (3) the matrices M_i are equal to $\frac{1}{2}(\lambda_i)$ and the transformation operator is $e^{i(\theta/2)(\lambda\pi)}$. The projection operator for the eigenvalue α_k of $(\lambda \cdot \pi)$ is, in general,

$$P_k = [\alpha_k(\lambda \cdot \pi)^2 + \alpha_k^2(\lambda \cdot \pi) + 2Y/3]/2(\alpha_k + Y), \quad (5.8)$$

where α_k is given by Eq. (2.15). When $\boldsymbol{\pi}$ is an s-vector, the eigenvalues become $+1, 0, -1$ and the appropriate projection operators are

$$P_{\pm 1} = \tfrac{1}{2}[(\lambda \cdot \pi)^2 \pm (\lambda \cdot \pi)], \quad P_0 = [1 - (\lambda \cdot \pi)^2], \quad \text{for } Y = 0. \quad (5.9)$$

When π is a q-vector, there are only two distinct eigenvectors, $1/\sqrt{3}$ and $-2/\sqrt{3}$, and the projection operators are

$$P_{1/\sqrt{3}} = \tfrac{1}{3}[3^{\frac{1}{2}}(\lambda \cdot \pi) + 2],$$
$$P_{-2/\sqrt{3}} = \tfrac{1}{3}[-3^{\frac{1}{2}}(\lambda \cdot \pi) + 1], \quad Y = -1/\sqrt{3}. \quad (5.10)$$

From Eqs. (5.5) and (5.6) we can express the transformation operator as

$$e^{i(\theta/2)(\lambda\cdot\pi)} = \sum_k P_k(\lambda \cdot \pi)e^{i\theta\alpha_k/2} = A + B(\lambda \cdot \pi) + C(\lambda \cdot \pi)^2, \quad (5.11)$$

where the coefficients A, B, and C are

$$A = \frac{2Y}{3}\sum_k \frac{e^{i\theta\alpha_k/2}}{2(\alpha_k + Y)}, \quad B = \sum_k \frac{\alpha_k^2 e^{i\theta\alpha_k/2}}{2(\alpha_k + Y)}, \quad C = \sum_k \frac{\alpha_k e^{i\theta\alpha_k/2}}{2(\alpha_k + Y)} \quad (5.12)$$

in the general case,

$$A = 1, \quad B = i\sin\theta/2, \quad C = (\cos\theta/2 - 1) \quad (5.13)$$

in the s-vector case, and

$$A = \tfrac{1}{3}[2e^{i\theta/2\sqrt{3}} + e^{-i\theta/\sqrt{3}}],$$
$$B = (1/\sqrt{3})[e^{i\theta/2\sqrt{3}} - e^{-i\theta/\sqrt{3}}], \qquad (5.14)$$
$$C = 0$$

in the q-vector case. Given the relations

$$(\lambda \cdot \pi)^2 = (\lambda \cdot \Pi) + \tfrac{2}{3}I,$$
$$\alpha/(\alpha + Y) = 2/(3\alpha^2 - 1), \qquad (5.15)$$

which follow from Eqs. (2.1), (2.14), and (5.7), we see that the general result of Eqs. (5.11) and (5.12) is identical to the expression for $e^{i(a\cdot\lambda)}$ derived from a different point of view by Macfarlane, Sudbery, and Weisz.[5] Equation (5.13) is the same as the expression for $e^{i(\theta/2)\mathbf{J}\cdot\mathbf{n}}$ in the three-dimensional representation of $SU(2)$[10]; this is not surprising in view of the fact that $(\lambda \cdot \pi)$ belongs to an $SU(2)$ subgroup of $SU(3)$ whenever $\boldsymbol{\pi}$ is an s-vector. The q-vector case of Eq. (5.14) is the only one in which the transformation operator is linear in $(\lambda \cdot \pi)$.

For the conjugate three-dimensional representation $(\bar{3})$, the eigenvalues of $(\bar{\lambda} \cdot \pi)$ are opposite in sign from those of $(\lambda \cdot \pi)$, and the corresponding projection operators can be obtained by applying Eq. (2.21) to Eq. (5.8). The expression for $e^{i(\theta/2)(\bar{\lambda}\cdot\pi)}$ then becomes

$$e^{i(\theta/2)(\bar{\lambda}\cdot\pi)} = \bar{A} + \bar{B}(\bar{\lambda} \cdot \pi) + \bar{C}(\bar{\lambda} \cdot \pi)^2, \quad (5.16)$$

where

$$\bar{A} = \frac{-2Y}{3}\sum_k \frac{e^{i\theta\bar{\alpha}_k/2}}{2(\bar{\alpha}_k - Y)}, \quad \bar{B} = \sum_k \frac{e^{i\theta\bar{\alpha}_k/2}(\bar{\alpha}_k)^2}{2(\bar{\alpha}_k - Y)},$$
$$\bar{C} = \sum_k \frac{e^{i\theta\bar{\alpha}_k/2}\bar{\alpha}_k}{2(\bar{\alpha}_k - Y)} \qquad (5.17)$$

in general,

$$\bar{A} = 1, \quad \bar{B} = i \sin \theta/2, \quad \bar{C} = (\cos \theta/2 - 1) \quad (5.18)$$

in the s-vector case, and

$$\bar{A} = \tfrac{1}{3}[2e^{-i\theta/2\sqrt{3}} + e^{i\theta/\sqrt{3}}],$$
$$\bar{B} = (1/\sqrt{3})[e^{i\theta/\sqrt{3}} - e^{-i\theta/2\sqrt{3}}], \qquad (5.19)$$
$$\bar{C} = 0$$

in the q-vector case. As a check on these results, we observe that they can also be obtained by taking the transpose of Eq. (5.11), replacing θ by $(-\theta)$, and noting that $\bar{\alpha}_k = -\alpha_k$.

B. The Octet

In the case of the octet representation the matrices M_i are equal to the F_i of Eq. (3.1) and the transformation operator is $e^{i\theta(F\cdot\pi)}$. The eigenvalues of $(F \cdot \pi)$ are

$$\alpha_0 = 0,$$
$$\alpha_k = \cos \tfrac{1}{3}[\psi + (k-1)\pi], \quad k = 1, \cdots, 6, \quad (5.20)$$

[see Eqs. (3.16) and (3.17)] and the projection operators are

$$P_0 = \frac{1}{\cos^2 \psi}[\cos^2 \psi - 9(F \cdot \pi)^2 + 24(F \cdot \pi)^4 - 16(F \cdot \pi)^6], \qquad (5.21)$$

$$P_i = \frac{(F \cdot \pi)[4\alpha_i(F \cdot \pi)^2 + 4\alpha_i^2(F \cdot \pi) + \cos \psi][4(F \cdot \pi)^3 - 3(F \cdot \pi) + \cos \psi]}{6 \cos \psi \alpha_i(2\alpha_i + \cos \psi)}, \quad i = 1, 3, 5, \quad (5.22)$$

$$P_j = \frac{-(F \cdot \pi)[4\alpha_j(F \cdot \pi)^2 + 4\alpha_j^2(F \cdot \pi) - \cos \psi][4(F \cdot \pi)^3 - 3(F \cdot \pi) - \cos \psi]}{6 \cos \psi \alpha_j(2\alpha_j - \cos \psi)}, \quad j = 2, 4, 6. \quad (5.23)$$

To expand $e^{i\theta F\cdot\pi}$ in terms of the P_l, we use the fact that [see Eq. (3.17)]

$$\alpha_k = -\alpha_{k+3}. \qquad (5.24)$$

We then find that

$$e^{i\theta F\cdot\pi} = 1 + \sum_{n=1}^{6} A_n(F \cdot \pi)^n, \qquad (5.25)$$

where

$$A_1 = \sum_{i=1,3,5} \frac{(i \sin \theta\alpha_i) \cos \psi}{3\alpha_i(2\alpha_i + \cos \psi)},$$

$$A_2 = \sum_{i=1,3,5} \left(\frac{-3}{\cos^2 \psi} + \frac{(4\alpha_i^2 - 3) \cos \theta\alpha_i}{3\alpha_i(2\alpha_i + \cos \psi)}\right),$$

$$A_3 = \sum_{i=1,3,5} \frac{4(\cos \psi - 3\alpha_i)(i \sin \theta\alpha_i)}{3 \cos \psi(2\alpha_i + \cos \psi)},$$

$$A_4 = \sum_{i=1,3,5} \left(\frac{8}{\cos^2 \psi} + \frac{4(\cos \psi - 3\alpha_i) \cos \theta\alpha_i}{3(\cos \psi)\alpha_i(2\alpha_i + \cos \psi)}\right),$$

$$A_5 = \sum_{i=1,3,5} \frac{16\alpha_i(i \sin \theta\alpha_i)}{3 \cos \psi(2\alpha_i + \cos \psi)},$$

$$A_6 = \sum_{i=1,3,5} \frac{-16}{3 \cos \psi}\left(\frac{1}{\cos \psi} - \frac{\cos \theta\alpha_i}{2\alpha_i + \cos \psi}\right). \qquad (5.26)$$

When $\boldsymbol{\pi}$ is an s-vector, $(F \cdot \pi)$ has five eigenvalues instead of seven,

$$\alpha_k = 0, \pm\tfrac{1}{2}, \pm 1, \qquad (5.27)$$

and the corresponding projection operators are

$$\begin{aligned} P_0 &= [4(F\cdot\pi)^2 - 1][(F\cdot\pi)^2 - 1], \\ P_{\pm\frac{1}{2}} &= \mp\tfrac{4}{3}[1 \pm 2(F\cdot\pi)][(F\cdot\pi)^3 - (F\cdot\pi)], \\ P_{\pm 1} &= \pm\tfrac{1}{6}[1 \pm (F\cdot\pi)][4(F\cdot\pi)^3 - (F\cdot\pi)]. \end{aligned} \tag{5.28}$$

The coefficients A_n of Eq. (5.25) in this case are

$$\begin{aligned} A_1 &= (i/3)(8\sin\theta/2 - \sin\theta), \\ A_2 &= (\tfrac{16}{3}\cos\theta/2 - \tfrac{1}{3}\cos\theta - 5), \\ A_3 &= (i/3)(4\sin\theta - 8\sin\theta/2), \\ A_4 &= (4 + \tfrac{4}{3}\cos\theta - \tfrac{16}{3}\cos\theta/2), \\ A_5 &= A_6 = 0. \end{aligned} \tag{5.29}$$

When $\boldsymbol{\pi}$ is a q-vector, the eigenvalues are

$$\alpha_k = 0, \quad \pm 3^{\frac{1}{2}}/2, \tag{5.30}$$

and the projection operators become

$$\begin{aligned} P_0 &= \tfrac{1}{3}[3 - 4(F\cdot\pi)^2], \\ P_\pm &= \pm(1/\sqrt{3}[1 \pm (2/\sqrt{3})(F\cdot\pi)](F\cdot\pi). \end{aligned} \tag{5.31}$$

The corresponding coefficients in the expression of Eq. (5.25) for $e^{i\theta(F\cdot\pi)}$ are

$$\begin{aligned} A_1 &= (2i/\sqrt{3})\sin(\theta\sqrt{3}/2), \\ A_2 &= \tfrac{4}{3}(\cos(\theta\sqrt{3}/2) - 1), \\ A_3 &= A_4 = A_5 = A_6 = 0. \end{aligned} \tag{5.32}$$

As in the s-vector case of the triplet representation [see Eq. (5.18)], so here the coefficients A_1 and A_2 are such that the matrix $(2/\sqrt{3})(F\cdot\pi)$ behaves as a member of a three-dimensional representation of an $SU(2)$ subgroup[10] of $SU(3)$.

C. The Decuplet

The matrices representing the decuplet are given by the B_i of Eqs. (4.1)–(4.3). In general they have ten distinct eigenvalues [see Eqs. (4.12) and (4.13)], and hence there are ten projection operators. In terms of the quantities

$$\begin{aligned} L &= 4(B\cdot\pi)^3 - 3(B\cdot\pi) + \cos\psi, \\ M &= 4(B\cdot\pi)^3 - 3(B\cdot\pi) - \cos\psi, \\ N &= 4(B\cdot\pi)^3 - 9(B\cdot\pi) - (3\sqrt{3})\sin\psi, \\ R_i &= 4\alpha_i(B\cdot\pi)^2 + 4\alpha_i^2(B\cdot\pi), \end{aligned} \tag{5.33}$$

they are

$$P_0 = LMN/(3^{\frac{3}{2}}\sin\psi\cos^2\psi),$$

$$P_i = \frac{(B\cdot\pi)(R_i + \cos\psi)NL}{6\alpha_i\cos\psi(4\alpha_i^3 - \alpha_i)[4\alpha_i^3 - 9\alpha_i - (3\sqrt{3})\sin\psi]}, \quad i = 1, 3, 5,$$

$$P_j = \frac{(B\cdot\pi)(\cos\psi - R_j)MN}{6\alpha_j\cos\psi(4\alpha_j^3 - \alpha_j)[4\alpha_j^3 - 9\alpha_j - (3\sqrt{3})\sin\psi]}, \quad j = 2, 4, 6,$$

$$P_k = \frac{(B\cdot\pi)[R_k + (3\sqrt{3})\sin\psi]LM}{3\alpha_k(4\alpha_k^3 - 3\alpha_k)(4\alpha_k^3 - 3\alpha_k + \cos\psi)(4\alpha_k^3 - 3\alpha_k - \cos\psi)}, \quad k = 7, 8, 9. \tag{5.34}$$

When $\boldsymbol{\pi}$ is an s-vector, there are seven distinct eigenvalues, namely $0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}$, and the corresponding projection operators are

$$\begin{aligned} P_0 &= (-\tfrac{16}{9})[(B\cdot\pi)^2 - \tfrac{1}{4}][(B\cdot\pi)^2 - 1] \times [(B\cdot\pi)^2 - \tfrac{9}{4}], \\ P_{\pm\frac{1}{2}} &= \tfrac{4}{3}(B\cdot\pi \pm \tfrac{1}{2})(B\cdot\pi)[(B\cdot\pi)^2 - 1] \times [(B\cdot\pi)^2 - \tfrac{9}{4}], \\ P_{\pm 1} &= (-\tfrac{8}{15})(B\cdot\pi \pm 1)(B\cdot\pi)[(B\cdot\pi)^2 - \tfrac{1}{4}] \times [(B\cdot\pi)^2 - \tfrac{9}{4}], \\ P_{\pm\frac{3}{2}} &= \tfrac{4}{45}(B\cdot\pi \pm \tfrac{3}{2})(B\cdot\pi)[(B\cdot\pi)^2 - \tfrac{1}{4}] \times [(B\cdot\pi)^2 - 1]. \end{aligned} \tag{5.35}$$

In the case of a q-vector, the only distinct eigenvalues of $(B\cdot\pi)$ are $0, \pm\frac{1}{2}\sqrt{3}, -\sqrt{3}$, and the appropriate projection operators are

$$\begin{aligned} P_0 &= (-4/3\sqrt{3})[(B\cdot\pi)^2 - \tfrac{3}{4}][(B\cdot\pi) + \sqrt{3}], \\ P_{\pm\frac{1}{2}\sqrt{3}} &= (4/9\sqrt{3})(B\cdot\pi \pm \tfrac{1}{2}\sqrt{3})(B\cdot\pi)[(B\cdot\pi) + \sqrt{3}], \\ P_{-\sqrt{3}} &= (-4/9\sqrt{3})(B\cdot\pi)[(B\cdot\pi)^2 - \tfrac{3}{4}]. \end{aligned} \tag{5.36}$$

From the general expansion

$$e^{i\theta(B\cdot\pi)} = \sum_l P_l(B\cdot\pi)e^{i\theta\alpha_l}, \tag{5.37}$$

we can express the transformation matrix as a polynomial in $(B\cdot\pi)$:

$$e^{i\theta(B\cdot\pi)} = 1 + \sum_n A_n(\theta)(B\cdot\pi)^n. \tag{5.38}$$

In general the polynomial is of the ninth degree [see Eq. (5.34)], but it reduces to the sixth degree when $\boldsymbol{\pi}$ is an s-vector and to the third degree when π is a q-vector;

$$\begin{aligned} A_n = 0 \quad \text{for} \quad & n \geq 10, \quad \text{in general}, \\ & n \geq 7, \quad \boldsymbol{\pi} \text{ an } s\text{-vector}, \\ & n \geq 4, \quad \boldsymbol{\pi} \text{ a } q\text{-vector}. \end{aligned} \tag{5.39}$$

The specific formulas for the A_n can be obtained by comparing Eqs. (5.37) and (5.38); they are very complicated and so we shall not quote them here.

The conjugate ten-dimensional representation $(\overline{10})$ is described by the $\bar{B}_i$ matrices of Eqs. (4.19) and

(4.20). Its eigenvalues are opposite in sign from those of the (10) representation. In all cases, the appropriate projection operators for the $(\overline{10})$ can be obtained from those of the (10) by means of the substitution

$$(B \cdot \pi) \to -(\bar{B} \cdot \pi),$$
$$\alpha_k \to -\bar{\alpha}_k, \tag{5.40}$$

where $\bar{\alpha}_k = -\alpha_k$.

6. DISCUSSION

We have now completed the task of finding explicit forms for finite transformations in the triplet, octet, and decuplet representations of $SU(3)$. The projection operator method we have used is a general one which can, in principle, be applied to all representations. In practice the main problem associated with the method is that of finding the characteristic equations in the various representations.

We believe that the device we used to obtain the characteristic equation of the ten-dimensional representation can be extended to all other triangular representations[9]; however, because we have to consider tensors of higher and higher rank, the complexity of the device probably increases very quickly. As far as other types of representation are concerned, we are not sure what general tricks are available.

Throughout our analysis we have separated the s- and q-vector cases from the general one. In both instances multiple eigenvalues appear, and we have to turn from the general characteristic equation to the reduced one in which no root appears more than once. It is possible, however, to pass from the general case to the s- and q-vector by means of careful limiting procedures. The rules are that if α is a single root of $M \cdot \pi$ and β is an n-tuple root, then $(M \cdot \pi - \beta)^n/(\alpha - \beta)^n$ should be replaced by $(M \cdot \pi - \beta)/(\alpha - \beta)$ and $(M \cdot \pi - \beta)^{n-1}/(\beta - \beta)^{n-1}$ by $1/n$. The reason for using $1/n$ in the second case is that there are formally n projection operators for β when it becomes an n-tuple root. The validity of these rules is easy to demonstrate as long as the reduced characteristic equation is satisfied.

Finally we note that the projection operator method can be used for many other groups besides $SU(3)$. In fact the s-vector case of $(\lambda \cdot \pi)$ [see Eqs. (5.11) and (5.13)] and the q-vector case of $(F \cdot \pi)$ [see Eqs. (5.25) and (5.32)] are simple examples of its use for $SU(2)$. The method is probably a good one to start with for any group, but it may not be the best in all cases.

Note added in proof: Y. Lehrer-Ilamed [Proc. Cambridge Phil. Soc. **60,** 61 (1964)] has given a general formula expressing any function $f(A)$ as a polynomial in the matrix A with coefficients determined by the eigenvalues of A. He applied the formula to finite $SU(2)$ transformations. The author is grateful to Dr. Y. Dothan for drawing his attention to the work of Lehrer-Ilamed. For other methods for dealing with finite $SU(3)$ transformations, see D. F. Holland [J. Math. Phys. **10,** 531 (1969)] and references contained therein.

[1] See, for example, M. Gell-Mann, and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, 1964).
[2] S. MacLane and G. Birkoff, *Algebra* (Macmillan, New York, 1967), p. 316.
[3] D. S. Carlstone, S. P. Rosen, and S. Pakvasa, Phys. Rev. **174,** 1877 (1968).
[4] L. M. Kaplan and M. Resnikoff, J. Math. Phys. **8,** 2194 (1967); V. I. Ogievetskii and I. V. Polyvarinov, Yad. Fiz. **4,** 853 (1966) [Sov. J. Nucl. Phys. **4,** 605 (1967)].
[5] A. J. Macfarlane, A. Sudbery, and P. H. Weisz, Commun. Math. Phys. **11,** 77 (1968).
[6] L. Michel and L. A. Radicatti, *The Geometry of the Octet* (to be published).
[7] R. E. Cutkosky, Ann. Phys. (N.Y.) **23,** 415 (1963). See especially the appendix written in cooperation with Dr. P. Tarjanne.
[8] S. P. Rosen, J. Math. Phys. **5,** 289 (1964). The result follows from Eqs. (6) and (26) of this paper together with the definition $B^r_\rho = (\bar{\lambda}_i B_i)_{\rho r}$.
[9] S. P. Rosen (to be published).
[10] See, for example, A. Messiah, *Quantum Mechanics,* Vol. II (North-Holland, Amsterdam, 1962), p. 578.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Consequences of a Field-Source Identity in a Spin-2 Meson Theory*

K. RAMAN†
Physics Department, Brown University, Providence, Rhode Island 02912

(Received 24 August 1970)

We discuss the restrictions imposed on a spin-2 meson theory when the source is related to the stress tensor. We consider a spin-2 meson theory with a field-source identity, which postulates that the traceless part of the stress tensor is proportional to the irreducible part of the spin-2 meson field. We examine the consequences of imposing the field-source identity and Lorentz covariance on the spin-2 meson theory. We show that the field-source identity determines the parts of the singular terms in the stress-tensor equal-time commutation relations which are the most singular in the coupling strength g. It further requires that some of the singular terms be q-numbers. We obtain the constraints imposed on the field-dependence of the source $J_{\mu\nu}(x)$ by the field-source identity and the Lorentz covariance conditions. We show that as a consequence of some of these, the interaction term $gJ_{\mu\nu}(x)$ must be nonlinear in the coupling strength g.

1. INTRODUCTION

In an earlier paper we have discussed the general noncommutation and field-dependence requirements on the source of a spin-2 field when the source is assumed to be divergenceless. The physically most interesting divergenceless second-rank symmetric tensor is the stress tensor, and accordingly it is of interest to examine the restrictions on the theory when the source of the spin-2 field is related to the stress tensor.

We have elsewhere[1] discussed how the hypothesis of the tensor-meson dominance[2] of the matrix elements of the stress tensor may be expressed in a field theory by a field-source identity in a manner analogous to the field-current identity[3] that has been postulated for expressing the hypothesis of vector-meson dominance[4] of the matrix elements of the electromagnetic current. For a theory with one neutral spin-2 meson, the field-source identity takes the form

$$\Theta_{\mu\nu}(x) = (m^2/g)[U_{\mu\nu}(x) - \eta_{\mu\nu}u(x)], \qquad (1.1)$$

where $\Theta_{\mu\nu}(x)$ is the stress tensor, $u(x)$ is the trace $U_\sigma{}^\sigma$, and $U_{\mu\nu}(x)$ is the renormalized field operator for a neutral spin-2 field with mass m. The coupling strength g is defined by taking the source of the spin-2 field to be $gJ_{\mu\nu}$, with a suitable normalization of the source $J_{\mu\nu}$, as discussed in Ref. 1.

In a spin-2 meson theory with a field-source identity, the source of the spin-2 field is also closely related to the stress tensor; the relation between their matrix elements has been discussed in Ref. 1. In this paper, we examine what restrictions are imposed on such a theory by the requirements of Lorentz covariance, in addition to the general field dependence and noncommutation requirements discussed earlier.[5] The restrictions imposed by Lorentz covariance on the spin-2 meson theory are taken into account by considering the equal-time commutation relations (ETCR's) among the components of the stress tensor.[6,7] The general form of the stress-tensor ETCR consistent with Poincaré invariance and locality has been discussed by Boulware and Deser.[7,8]

We obtain the following results:

(1) The field-source identity determines the parts of the singular terms in the stress-tensor ETCR's which are c-numbers and which are the most singular in the coupling constant g.

(2) Because of the field-source identity, it is necessary that some of the singular terms in the stress-tensor ETCR's be q-numbers.

(3) The field-source identity and the stress-tensor ETCR's impose explicit constraints on the nature of the field dependence of the source density $J_{\mu\nu}(x)$ in the field equations for the neutral spin-2 meson.

(4) As a consequence of some of these constraints, the interaction term $gJ_{\mu\nu}(x)$ for the spin-2 field must be nonlinear in the coupling strength g.

The first result is analogous to the result following from the field-current identity in a vector-meson theory, where the identity determines the singular terms in the ETCR's among the vector currents.[9]

The most important restrictions on the theory are those imposed by Lorentz covariance, expressed by the stress-tensor ETCR's stated in Sec. 2. The constraints on the field dependence of the source which follow from the covariance conditions and the field-source identity are discussed in Sec. 2.

On requiring the Jacobi identity to hold for the double commutators of the source $J_{\mu\nu}$ and the dynamical variables of the spin-2 field, strong restrictions are imposed on the operator nature of the singular

terms in the stress-tensor ETCR's. These are discussed in Sec. 3. In Sec. 4, we show that the restrictions following from the field-source identity and the Lorentz covariance conditions imply that the source $gJ_{\mu\nu}$ of the spin-2 field must be nonlinear in the coupling strength g. This is interesting because it shows the essentially nonlinear nature of a theory in which the source is related to the stress tensor. A familiar example of this is theory of the gravitational field.[10]

In Sec. 5 we verify that the source $J_{\mu\nu}$ of the spin-2 field obeys certain requirements, which follow from the recent work of Orzalesi, Sucher, and Woo,[11] and Divgi and Woo.[12] In Sec. 6, we briefly summarize our conclusions.

2. CONSTRAINTS IMPOSED BY COVARIANCE CONDITIONS AND FIELD-SOURCE IDENTITY

The conditions imposed on a field theory by Lorentz covariance, as expressed by the ETCR's among the components of the stress tensor, are the following[7,8]:

$$i[\Theta_{00}(x), \Theta_{00}(y)]_{x_0=y_0} = [\Theta_{0k}(x) + \Theta_{0k}(y)] \times \partial_k\delta(x-y) + \omega_{00,00}(x, y), \tag{2.1}$$

$$i[\Theta_{00}(x), \Theta_{0k}(y)]_{x_0=y_0} = [\Theta_{kl}(x) + \Theta_{00}(y)\eta_{kl}] \times \partial_l\delta(x-y) + \omega_{00,0k}(x, y), \tag{2.2}$$

$$i[\Theta_{00}(x), \Theta_{kl}(y)]_{x_0=y_0} = [-\partial_0\Theta_{kl}(x) + \Theta_{0k}(y)\partial_l + \Theta_{0l}(y)\partial_k] \times \delta(x-y) + \omega_{00,kl}(x, y), \tag{2.3}$$

$$i[\Theta_{0k}(x), \Theta_{0l}(y)]_{x_0=y_0} = [\Theta_{0l}(x)\partial_k + \Theta_{0k}(y)\partial_l] \times \delta(x-y) + \omega_{0k,0l}(x, y), \tag{2.4}$$

$$i[\Theta_{0k}(x), \Theta_{mn}(y)]_{x_0=y_0} = [\Theta_{mn}(x)\eta_{kl} - \Theta_{ml}(x)\eta_{nk} - \Theta_{nl}(y)\eta_{mk}] \times \partial_l\delta(x-y) + \omega_{0k,mn}(x, y). \tag{2.5}$$

Here, the operators $\omega_{00,00}$, etc., are subject to the following constraints, as a result of the commutation relations (CR's) among the generators of the Poincaré group:

(a) $$\int d^3x\,\omega\cdots(x, y) = 0 \tag{2.6}$$

for the singular term $\omega\cdots$ occurring in each of the equations (2.1)–(2.5),

(b) $$\langle\omega_{00,00}\rangle = 0 = \langle\omega_{00,kl}\rangle = \langle\omega_{0k,0l}\rangle, \tag{2.7}$$

(c) $$\omega_{00,00}(x, y) = -\omega_{00,00}(y, x),$$
$$\omega_{0k,0m}(x, y) = -\omega_{0m,0k}(y, x), \tag{2.8}$$

(d) $$\int d^3x x_k\omega_{00,00}(x, y) = 0,$$
$$\int d^3x x_k\omega_{00,0m}(x, y) = 0,$$
$$\int d^3x x_k\omega_{00,mn}(x, y) = 0, \tag{2.9}$$
$$\int d^3x[x_k\omega_{0l,0m}(x, y) - x_l\omega_{0k,0m}(x, y)] = 0,$$
$$\int d^3x[x_k\omega_{0l,mn}(x, y) - x_l\omega_{0k,mn}(x, y)] = 0.$$

Consider a theory with only one neutral tensor meson, with the quantum numbers of the vacuum. The field-source identity

$$\Theta_{\mu\nu}(x) = \frac{m^2}{g}[U_{\mu\nu}(x) - \eta_{\mu\nu}u(x)], \tag{2.10}$$

together with the ETCR's (2.1)–(2.5) and the canonical commutation relations for the spin-2 field, impose constraints on the structure of the interaction.

To obtain these, we first write down the equations describing a spin-2 field interacting with a divergenceless source $J_{\mu\nu}(x)$:

$$\partial_\rho\Pi^\rho{}_{\mu\nu} - \tfrac{1}{2}(\partial_\mu\Pi_\nu + \partial_\nu\Pi_\mu) - \tfrac{1}{2}\eta_{\mu\nu}\partial_\lambda(\tilde{\Pi}_\lambda - \Pi_\lambda) = \tfrac{1}{2}m^2(U_{\mu\nu} - \eta_{\mu\nu}u) - \tfrac{1}{2}gJ_{\mu\nu}, \tag{2.11}$$

$$(\Pi^\nu{}_\lambda{}^\mu - \tfrac{1}{2}\delta_\lambda{}^\mu\tilde{\Pi}^\nu) + (\Pi^\mu{}_\lambda{}^\nu - \tfrac{1}{2}\delta_\lambda{}^\nu\tilde{\Pi}^\nu) - \eta^{\mu\nu}\Pi_\lambda = \partial_\lambda W^{\mu\nu} - \tfrac{1}{2}\delta_\lambda{}^\mu(\partial_\sigma W^{\nu\sigma}) - \tfrac{1}{2}\delta_\lambda{}^\nu(\partial_\sigma W^{\mu\sigma}), \tag{2.12}$$

$$\partial^\mu J_{\mu\nu} = 0. \tag{2.13}$$

These lead to the following equations:

$$\partial^\mu(U_{\mu\nu} - \eta_{\mu\nu}u) = 0, \tag{2.14}$$

$$\Pi^\lambda{}_{\mu\nu} = \tfrac{1}{2}(\partial_\mu U_{\lambda\nu} + \partial_\nu U_{\lambda\mu} - \partial_\lambda U_{\mu\nu}), \tag{2.15}$$

$$u = \frac{-g}{3m^2}j, \quad j \equiv J^\sigma{}_\sigma, \tag{2.16}$$

$$(\Box^2 + m^2)U_{\mu\nu} - (\partial_\mu\partial_\nu + m^2\eta_{\mu\nu})u = gJ_{\mu\nu}. \tag{2.17}$$

In the above $\eta_{\mu\nu}$ denotes the Minkowski metric, $(1, -1, -1, -1)$, and $u \equiv U^\sigma{}_\sigma$.

Choosing the dynamical variables to be

$$U_{kl}{}^T = U_{kl} - \tfrac{1}{3}\eta_{kl}U_{mm}, \tag{2.18}$$

and the momentum π_{kl} conjugate to $U_{kl}{}^T$,

$$\pi_{kl} = -2\Pi^0{}_{ij} + 2(\nabla^2 - \tfrac{3}{2}m^2)^{-1}(\partial_k\partial_l - \tfrac{1}{2}m^2\eta_{kl})\Pi^0{}_{mm}, \tag{2.19}$$

we may express the other components of $U_{\mu\nu}$ as follows:

$$U_{mm} = (\tfrac{2}{3}\nabla^2 - m^2)^{-1}[(\partial\partial U^T) - gJ_{00}], \quad (2.20)$$

$$U_{k0} = -\frac{1}{m^2}(\partial\pi)_k + \frac{2}{3m^4}\partial_k(\partial\partial\pi) + \frac{g}{m^2}\left(J_{0k} - \frac{2}{3m^2}\partial_k\partial_m J_{0m}\right), \quad (2.21)$$

$$U_{00} = (\tfrac{2}{3}\nabla^2 - m^2)^{-1}[(\partial\partial U^T) - gJ_{00}] - \frac{g}{3m^2}j. \quad (2.22)$$

Here j denotes the trace $J_\sigma{}^\sigma$, and we have used the notation

$$(\partial\pi)_k \equiv \partial^l\pi_{kl}, \quad (\partial\partial\pi) \equiv \partial^k\partial^l\pi_{kl}, \quad (\partial\partial U^T) \equiv \partial^k\partial^l U_{kl}{}^T. \quad (2.23)$$

We substitute the field-source identity in the relations (2.1)–(2.5) and re-express the results in terms of the dynamical variables. The relation (2.2), for instance, leads to the following relation:

$$\begin{aligned} &i[D^{-1}\partial\partial U^T(\mathbf{x}), D^{-1}J_{00}(\mathbf{y})] + i[D^{-1}J_{00}(\mathbf{x}), D^{-1}\partial\partial U^T(\mathbf{y})] \\ &\quad - gi[D^{-1}J_{00}(\mathbf{x}), D^{-1}J_{00}(\mathbf{y})] \\ &= (1/m^4)[\tilde\pi_k(x) + \tilde\pi_k(y)]\partial^k\delta(x-y) \\ &\quad - (g/m^4)[\tilde\jmath_k(x) + \tilde\jmath_k(y)]\partial^k\delta(x-y) \\ &\quad - (g/m^4)\omega_{00,00}(x,y). \end{aligned} \quad (2.24)$$

Here D denotes the operator

$$D = (\tfrac{2}{3}\nabla^2 - m^2), \quad (2.25)$$

and we have introduced the notation

$$\tilde\pi_k \equiv (\partial\pi)_k - (2/3m^2)\partial_k(\partial\partial\pi), \quad (2.26)$$

$$\tilde\jmath_k \equiv J_{0k} - (2/3m^2)\partial_k\partial_l J_{0l}. \quad (2.27)$$

When the source density $J_{\mu\nu}(x)$ is nonsingular in the coupling strength g, then (2.24) implies the following relation:

$$\begin{aligned} &i[\partial\partial U^T(x), J_{00}{}^{(0)}(y)] + i[J_{00}{}^{(0)}(x), \partial\partial U^T(y)] \\ &\quad = D(x)D(y)\{(1/m^4)[\tilde\pi_k(x) + \tilde\pi_k(y)]\partial_k\delta(x-y) \\ &\qquad - (1/m^4)\bar\omega_{00,00}(x,y)\}. \end{aligned} \quad \text{(a)}$$

Here, $\bar\omega_{00,00}(x,y)$ denotes the coefficient in possible terms in $\omega_{00,00}(x,y)$ of order $1/g$:

$$\omega_{00,00} = (1/g)\bar\omega_{00,00} + \omega^{(0)}_{00,00} + O(g), \quad (2.28)$$

and[13]

$$J^{(0)}_{\mu\nu}(x) = [J_{\mu\nu}(x)]_{g=0}. \quad (2.29)$$

The relations obtained similarly from (2.2)–(2.5) are the following:

$$\omega_{00,0k}{}^S(x,y) = -(2/3g^2)\nabla^2\partial_k\delta(x-y), \quad \text{(b1)}$$

$$\begin{aligned} &i[\partial\partial U_T, \tilde\jmath^{(0)}_k] + i[J^{(0)}_{00}, \tilde\pi_k] \\ &\quad = D(x)\{U_{kl}{}^T(x)\partial_l\delta(x-y) \\ &\qquad + \tfrac{1}{3}[D^{-1}\partial\partial U_T(x)]\partial_k\delta(x-y) \\ &\qquad + [D^{-1}\partial\partial U_T(y)]\partial_k\delta(x-y) \\ &\qquad + (1/m^2)\bar\omega_{00,0k}(x,y)\}, \end{aligned} \quad \text{(b2)}$$

$$\begin{aligned} &i[\partial\partial U^T(\mathbf{x}), j^{(0)}(\mathbf{y})] \\ &\quad = (1/m^2)D(x)(\{[\tilde\pi_k(x) - \tilde\pi_k(y)]\partial_k \\ &\qquad - (1/m^2)D\partial\partial\pi(x)\}\delta(x-y) \\ &\qquad - \{\bar\omega_{00,ll}(x,y) - \bar\omega_{00,00}(x,y)\}), \end{aligned} \quad \text{(c1)}$$

$$\begin{aligned} &i[J^{(0)}_{00}(\mathbf{x}), U_{mn}{}^T(\mathbf{y})] \\ &\quad = (1/m^2)D(x)\{\pi_{mn}(x) - (1/m^2)[\partial_m\tilde\pi_n(x) \\ &\qquad + \partial_n\tilde\pi_m(x) - \tfrac{2}{3}\eta_{mn}\partial_l\tilde\pi_l(x) \\ &\qquad - \tilde\pi_m(y)\partial_n - \tilde\pi_n(y)\partial_m + \tfrac{2}{3}\eta_{mn}\tilde\pi_l(y)\partial_l \\ &\qquad + (2/3m^2)(\partial_m\partial_n - \tfrac{1}{3}\eta_{mn}\nabla^2)\partial\partial\pi(x)]\} \\ &\qquad \times \delta(x-y) - (1/m^2)D(x)\bar\omega_{00,mn}{}^T(x,y), \end{aligned} \quad \text{(c2)}$$

$$i[\tilde\jmath^{(0)}_k(\mathbf{x}), \tilde\pi_l(\mathbf{y})] + i[\tilde\pi_k(\mathbf{x}), \tilde\jmath^{(0)}_l(\mathbf{y})] = [\tilde\pi_l(x)\partial_k + \tilde\pi_k(y)\partial_l]\delta(x-y) - \bar\omega_{0k,0m}(x,y), \quad \text{(d)}$$

$$\begin{aligned} &\omega_{0k,mn}{}^s(x,y) \\ &\quad = (m^2/g^2)[\tfrac{1}{2}(\eta_{km}\partial_n + \eta_{kn}\partial_m - \tfrac{2}{3}\eta_{mn}\partial_k)\delta(x-y) \\ &\qquad - (2/3m^2)\partial_m\partial_n\partial_k\delta(x-y)], \end{aligned} \quad \text{(e1)}$$

$$\begin{aligned} &i[\tilde\jmath^{(0)}_k(\mathbf{x}), \partial\partial U^T(\mathbf{y})] + i[\tilde\pi_k(\mathbf{x}), J_{00}(\mathbf{y})] \\ &\quad - (1/3m^2)[\tilde\pi_k(\mathbf{x}), Dj(\mathbf{y})] \\ &= D(y)(\{D^{-1}\partial\partial U^T(x)\partial_k \\ &\qquad - 2[U_{kl}{}^T(y) + \tfrac{1}{3}\eta_{kl}D^{-1}\partial\partial U^T(y)]\partial_l\}\delta(x-y) \\ &\qquad + (1/m^2)\bar\omega_{0k,ll}(x,y)), \end{aligned} \quad \text{(e2)}$$

and

$$\begin{aligned} &i[\tilde\jmath^{(0)}_k(\mathbf{x}), U_{mn}{}^T(\mathbf{y})] \\ &\quad = U_{mn}{}^T(x)\partial_k\delta(x-y) \\ &\qquad - \{\eta_{nk}[U_{ml}{}^T(y) + \tfrac{1}{3}\eta_{ml}D^{-1}\partial\partial U^T(y)] \\ &\qquad + \eta_{mk}[U_{nl}{}^T(y) + \tfrac{1}{3}\eta_{nl}D^{-1}\partial\partial U^T(y)] \\ &\qquad - \tfrac{2}{3}\eta_{mn}[U_{kl}{}^T(y) + \tfrac{1}{3}\eta_{kl}D^{-1}\partial\partial U^T(y)]\} \\ &\qquad \times \partial_l\delta(x-y) + (1/m^2)\bar\omega_{0k,mn}{}^T(x,y). \end{aligned} \quad \text{(e3)}$$

The notation used in the above equations is as follows. $\omega_{00,0k}{}^s$ is the part of $\omega_{00,0k}$ which is the most singular in g, while $(1/g)\bar\omega_{00,0k}$ is the part which behaves as g^{-1}. Similarly for $\omega_{0k,mn}{}^s$ and $\bar\omega_{0k,mn}$. Writing explicitly, we have

$$\omega_{00,0k} = (1/g^2)[g^2\omega_{00,0k}{}^s] + (1/g)\bar\omega_{00,0k} + \text{a part regular in } g, \text{ etc.} \quad (2.30)$$

In obtaining the relations following from (2.2) and (2.5), we have used the canonical commutation relations for the spin-2 field:

$$i[U_{kl}{}^T(x), \pi_{mn}(y)]_{x_0=y_0} = \delta_{kl,mn}{}^T\delta(x-y), \quad (2.31)$$

where

$$\delta_{kl,mn}{}^T \equiv \tfrac{1}{3}(\eta_{km}\eta_{ln} + \eta_{kn}\eta_{lm} - \tfrac{2}{3}\eta_{kl}\eta_{mn}). \quad (2.32)$$

The results (b1) and (e1) tell us that in the singular terms $\omega_{00,0k}$ and $\omega_{0k,mn}$ in the stress-tensor ETCR's (2.2) and (2.5), the parts most singular in g are c numbers. We note that the trace (in m, n) of this c-number part of $\omega_{0k,mn}$ is equal to the corresponding

c-number part of $\omega_{00,0k}$:

$$\omega_{0k,mm}{}^{S}(x, y) = \omega_{00,0k}{}^{S}(x, y). \tag{2.33}$$

This is a consequence of (2.2) and (2.5) and the Lorentz transformation properties of $\langle 0|\,[\theta_{\mu\nu}(x), \theta_{\lambda\sigma}(y)]\,|0\rangle$.

The relations (a), (b2), (c1), (c2), (d), (e2), and (e3) express constraints on the structure of the g-independent parts of the source $J_{\mu\nu}$. For instance, the relations (a), (c1), and (c2) imply that $J_{00}(x)$ and $j(x)$ must depend on the canonical momentum variable $\pi_{kl}(x)$. These constraints are in addition to the restrictions imposed because $J_{\mu\nu}$ has zero divergence; the latter have been discussed in Ref. 1.

3. CONSTRAINTS IMPOSED BY THE JACOBI IDENTITY

In the last section, we have obtained a number of constraints on the source density $J_{\mu\nu}$ in the form of relations for the commutators of components of $J_{\mu\nu}$ with the dynamical variables $U_{mn}{}^{T}$ and π_{mn}. Since the dynamical variables themselves satisfy the canonical commutation relations (CCR's), the Jacobi identity for double commutators will impose consistency conditions. We examine these in this section.

There are three independent constraints arising from Jacobi identities applied to the commutator constraints obtained in Sec. 3. First, consider the relations (b2), (c2), and (e3), and the CCR's. The CCR's are given by (2.31) and the vanishing of $[U_{mn}{}^{T}(x), U_{kl}{}^{T}(y)]$ and $[\pi_k(x), \pi_l(y)]$ at equal times.

From (e3), we may obtain an expression for the equal-time commutator (ETC)

$$i[\partial\partial U^{T}, \vartheta_k^{(0)}]. \tag{3.1}$$

Using this together with (b2) gives an expression for the ETC

$$i[J_{00}^{(0)}, \tilde{\pi}_k]. \tag{3.2}$$

The relation (c2) gives an expression for the ETC

$$i[J_{00}^{(0)}, U_{mn}{}^{T}], \tag{3.3}$$

while the CCR gives the ETC

$$i[\tilde{\pi}_k, U_{mn}{}^{T}], \tag{3.4}$$

which is a c number.

The Jacobi identity for the triple commutators of the operators $\tilde{\pi}_k$, $U_{mn}{}^{T}$, and J_{00} then leads to the following relation:

$$[U_{mn}{}^{T}(\mathbf{z}), \{D(x)\bar{\omega}_{00,0k}(\mathbf{x}, \mathbf{y}) + \partial_l\partial_j\bar{\omega}_{0k,lj}{}^{T}(\mathbf{y}, \mathbf{x})\}] + [\tilde{\pi}_k(y), D(x)\bar{\omega}_{00,mn}{}^{T}(\mathbf{z}, \mathbf{x})] = 0. \tag{f}$$

This gives a constraint on the operator character of the singular terms $\omega_{00,0k}$, $\omega_{0k,lj}{}^{T}$, and $\omega_{00,mn}{}^{T}(z, x)$ occurring in the stress-tensor ETCR. Thus (f) is consistent with the two functions

$$\bar{\omega}_{00,mn}{}^{T}(z, x) \tag{3.5}$$

and

$$\{D(x)\bar{\omega}_{00,0k}(x, y) + \partial_l\partial_j\bar{\omega}_{0k,lj}{}^{T}(y, x)\}, \tag{3.6}$$

both being c numbers. In a model in which this is true, we would have the additional restriction that

$$\bar{\omega}_{00,mn}{}^{T}(z, x) = 0, \tag{3.7}$$

since the vacuum expectation value (VEV) of $\omega_{00,mn}$ must vanish, which follows from (2.3). We stress that (f) does not necessarily require (3.5) and (3.6) to be c numbers; however, the latter may be true in a particular model.

In a manner similar to the above, we may start with the relations (b2), (c1), and (e2), and the CCR; the Jacobi identity for the operators

$$j(y), \quad \tilde{\pi}_n(x), \quad \text{and} \quad \partial\partial U^{T}(z) \tag{3.8}$$

then leads to the constraint

$$[\tilde{\pi}_k(\mathbf{x}), D(\mathbf{z})D(\mathbf{y})\{\bar{\omega}_{00,ll}(\mathbf{z}, \mathbf{y}) - \bar{\omega}_{00,00}(\mathbf{z}, \mathbf{y})\}] - 3m^2[\partial\partial U_T(\mathbf{z}), D(\mathbf{y}) \times \{\bar{\omega}_{0k,mm}(\mathbf{y}, \mathbf{x}) + \bar{\omega}_{00,0k}(\mathbf{y}, \mathbf{x})\}] = 0. \tag{g}$$

Again, the constraint (g) would be consistent with the quantities

$$\{\bar{\omega}_{00,ll}(z, x) - \bar{\omega}_{00,00}(z, x)\} \tag{3.9}$$

and

$$\{\bar{\omega}_{0k,mm}(x, y) + \bar{\omega}_{00,0k}(y, x)\} \tag{3.10}$$

being c numbers. In a model in which this is true, (3.9) would further imply that

$$\bar{\omega}_{00,ll}(z, x) = \bar{\omega}_{00,00}(z, x). \tag{3.11}$$

Note that each term in (3.11) must be a q number or zero, since it has a vanishing VEV.

Finally, we consider the relations (d) and (e3) and the CCR. Using these and the Jacobi identity for the operators

$$\vartheta_m, \quad \tilde{\pi}_n, \quad \text{and} \quad U_{kl}{}^{T}, \tag{3.12}$$

we obtain the following constraint:

$$\begin{aligned} &i[U_{mn}{}^{T}(\mathbf{y}), \bar{\omega}_{0k,0l}(\mathbf{x}, \mathbf{y})] + (i/m^2)[\tilde{\pi}_l(\mathbf{z}), \bar{\omega}_{0k,mn}{}^{T}(\mathbf{x}, \mathbf{y})] \\ &\quad + (i/m^2)[\tilde{\pi}_k(\mathbf{x}), \bar{\omega}_{0l,mn}{}^{T}(\mathbf{z}, \mathbf{y})] \\ &= -F_{lmn}(y)\delta(x - y)\partial_k\delta(x - z) \\ &\quad - F_{kmn}(y)\delta(z - y)\partial_l\delta(x - z) \\ &\quad + \eta_{kr}F_{lmn}(z)\delta(z - x)\partial_r\delta(x - y) \\ &\quad + \eta_{lr}F_{kmn}(x)\delta(x - z)\partial_r\delta(z - y) \\ &\quad - \Big(\eta_{nk}F_{lmr}(z) + \eta_{mk}F_{lnr}(z) - \tfrac{2}{3}\eta_{mn}F_{lkr}(z) \\ &\quad - \frac{4}{9m^2}\delta_{mn,kr}{}^{T}\frac{\partial}{\partial z_l}\nabla_z^2\Big)\delta(z - y)\partial_r\delta(x - y) \\ &\quad - \Big(\eta_{nl}F_{kmr}(x) + \eta_{ml}F_{knr}(x) - \tfrac{2}{3}\eta_{mn}F_{klr}(x) \\ &\quad - \frac{4}{9m^2}\delta_{mn,lr}{}^{T}\frac{\partial}{\partial x_k}\nabla_x^2\Big)\delta(x - y)\partial_r\delta(z - y). \end{aligned} \tag{h}$$

Here we have used the notation

$$F_{lmn}(x) = \delta_{jl,mn}{}^{T}\frac{\partial}{\partial x_j} - \frac{2}{3m^2}\frac{\partial}{\partial x_l} \times \left(\frac{\partial^2}{\partial x_m \partial x_n} - \tfrac{1}{3}\eta_{mn}\nabla_x^2\right). \quad (3.13)$$

In contrast to the constraints (f) and (g), the constraint (h) requires that at least one of the singular terms $\omega_{0k,0l}$ and $\omega_{0k,mn}{}^{T}$ must have a q-number part.

We note that a simple set of assumptions about the operator character of the singular terms in the stress-tensor ETCR's that would be consistent with the constraints (f), (g), and (h) (which in turn follow from the covariance conditions and the field-source identity) are that $\bar{\omega}_{00,0k}$ and $\bar{\omega}_{0k,mn}{}^{T}$ are c numbers, $\bar{\omega}_{00,mn}{}^{T} = 0$, while $\bar{\omega}_{00,ll} = \bar{\omega}_{00,00}$ are q numbers (or zero) and $\bar{\omega}_{0k,0l}$ is a nonvanishing q number.

These are statements only about the parts of the singular terms that are proportional to $1/\bar{g}$. The terms proportional to $1/g^2$ in $\omega_{00,0k}$ and $\bar{\omega}_{0k,mn}$ are given by (b1) and (e1), while $\bar{\omega}_{00,mn}$ and $\bar{\omega}_{0k,0l}$ have no terms proportional to $1/g^2$. Therefore, the simple assumptions above, if true for $\bar{\omega}_{00,0k}$, etc., are also true for the parts of $\bar{\omega}_{00,0k}$, $\bar{\omega}_{0k,mn}$, etc., that are singular in g. The considerations of this paper do not give any constraints on the parts that are regular in g (apart from the constraints on the VEV's of $\omega_{00,0k}$, $\omega_{00,mn}$, etc., that follow directly from the stress-tensor ETCR's).

4. NONLINEARITY OF THE SOURCE IN THE COUPLING STRENGTH

We now examine the additional relations that would follow from the stress-tensor ETCR's (2.1)–(2.5) and the field-source identity if the source term $gJ_{\mu\nu}$ were linear in the coupling constant g, that is, if $J_{\mu\nu}{}^{(0)}(x) = J_{\mu\nu}(x)$.

For instance, consider the relation (2.24) following from (2.1) and the field-source identity. When $J_{\mu\nu}(x) = J_{\mu\nu}{}^{(0)}(x)$, then the terms of order g in (2.24) lead to the relation

$$i[D^{-1}J_{00}(x), D^{-1}J_{00}(y)] = (1/m^4)[\mathfrak{J}_k(x) + \mathfrak{J}_k(y)]\partial_k\delta(x-y) + (1/m^4)\omega_{00,00}{}^{(0)}(x,y), \quad (4.1)$$

where $\omega_{00,00}{}^{(0)}$ is the part of $\omega_{00,00}$ that is independent of g.

Similarly, we obtain from (2.1)–(2.5) and the field-source identity the following relations [if we assume $J_{\mu\nu}(x) = J_{\mu\nu}{}^{(0)}(x)$]:

$$i[D^{-1}J_{00}, \mathfrak{J}_k] = [\tfrac{1}{3}D^{-1}J_{00}(x) + D^{-1}J_{00}(y)]\partial_k\delta(x-y) - (1/m^2)\omega_{00,0k}{}^{(0)}(x,y), \quad (4.2)$$

$$i[D^{-1}J_{00}, j] = (1/m^2)\{[\mathfrak{J}_k(x) - \mathfrak{J}_k(y)]\partial_k + \tfrac{1}{3}\partial_0 j(x) + m^{-2}D\partial_l J_{0l}(x)\}\delta(x-y) - (1/m^2)[\omega_{00,ll}{}^{(0)}(x,y) - \omega_{00,00}{}^{(0)}(x,y)], \quad (4.3)$$

$$0 = \{[\mathfrak{J}_m(y)\partial_n + \mathfrak{J}_n(y)\partial_m - \tfrac{2}{3}\eta_{mn}\mathfrak{J}_k(y)\partial_k] - [\partial_m J_{0n}(x) + \partial_n J_{0m}(x) - \tfrac{2}{3}\eta_{mn}\partial_k J_{0k}(x)] + (2/3m^2)(\partial_m\partial_n - \tfrac{1}{3}\eta_{mn}\nabla^2)\partial_l J_{0l}(x)\}\delta(x-y) + \omega_{00,mn}{}^{(0)T}(x,y), \quad (4.4)$$

$$i[\mathfrak{J}_k(\mathbf{x}), \mathfrak{J}_l(\mathbf{y})] = -[\mathfrak{J}_l(x)\partial_k + \mathfrak{J}_k(y)\partial_l]\delta(x-y) - \omega_{0k,0l}{}^{(0)}(x,y), \quad (4.5)$$

$$-i[\mathfrak{J}_k, D^{-1}J_{00}] + (i/m^2)[\mathfrak{J}_k, j] = -[D^{-1}J_{00}(x) - (1/m^2)j(x)]\partial_k\delta(x-y) + 2[D^{-1}J_{00}(y) - (1/m^2)j(y)]\partial_k\delta(x-y) + (1/m^2)\omega_{0k,ll}{}^{(0)}(x,y). \quad (4.6)$$

$$0 = \tfrac{2}{3}\delta_{mn,kl}{}^{T}[D^{-1}J_{00}(y) - (1/m^2)j(y)]\partial_l\delta(x-y) + (1/m^2)\omega_{0k,mn}{}^{(0)T}(x,y). \quad (4.7)$$

In the above equations, $\omega_{00,mn}{}^{(0)T}$ is the traceless part of $\omega_{00,mn}{}^{(0)}$ (with respect to the indices m and n), and $\omega_{00,mn}{}^{(0)}$ is the g-independent part of $\omega_{00,mn}$, etc.

The relations (4.4) and (4.7) directly give strong constraints on the singular terms $\omega_{00,mn}{}^{(0)T}$ and $\omega_{0k,mn}{}^{(0)T}$. The Jacobi identities for triple commutators, e.g., those involving $\mathfrak{J}_k$, j, and J_{00}, give further restrictions on the singular terms.

Consider (4.7). This directly gives the g-independent, traceless (in m, n) part of the singular term $\omega_{0k,mn}$ in the stress-tensor ETCR (2.5). It is seen that this expression is consistent with the general requirements on $\omega_{0k,mn}$.

Next we consider (4.4). The properties (2.6) and (2.9) required of $\omega_{00,mn}(x,y)$ require that

$$\int d^3x\,\omega_{00,mn}{}^{(0)T}(x,y) = 0 \quad (4.8)$$

and

$$\int d^3x\,x^k\omega_{00,mn}{}^{(0)T}(x,y) = 0. \quad (4.9)$$

(4.8) and (4.9) require that

$$[\partial_m J_{0n}(x) + \partial_n J_{0m}(x) - \tfrac{2}{3}\eta_{mn}\partial_k J_{0k}(x)] - (2/3m^2)(\partial_m\partial_n - \tfrac{1}{3}\eta_{mn}\nabla^2)\partial_l J_{0l}(x) = 0, \quad (4.10)$$

while (4.9) and (4.7) require the following:

$$-[\eta_{kn}\mathfrak{J}_m(y) + \eta_{km}\mathfrak{J}_n(y) - \tfrac{2}{3}\eta_{mn}\mathfrak{J}_k(y)] + y_k\{[\partial_m J_{0n}(y) + \partial_n J_{0m}(y) - \tfrac{2}{3}\eta_{mn}\partial_l J_{0l}(y)] - (2/3m^2)(\partial_m\partial_n - \tfrac{1}{3}\eta_{mn}\nabla^2)\partial_l J_0(y)\} = 0. \quad (4.11)$$

(4.11) implies that

$$\mathfrak{F}_m(0) = 0. \tag{4.12}$$

As the origin of y can be chosen arbitrarily, this implies

$$\mathfrak{F}_m(y) = 0. \tag{4.13}$$

Recalling (2.27), we see that (4.10) and (4.13) would require that

$$J_{0k}(x) = 0. \tag{4.14}$$

This shows that the assumption that $J_{\mu\nu}(x)$ is independent of g implies that $J_{\mu\nu}(x) = 0$ in a Lorentz-covariant theory with a field-source identity. Therefore, such a theory with a nonzero interaction requires that the source term $gJ_{\mu\nu}(x)$ be nonlinear in g.

5. VERIFICATION OF SOME PROPERTIES OF THE SOURCE $J_{\mu\nu}$

Recently, results have been obtained by Orzalesi, Sucher, and Woo,[11] and by Divgi and Woo,[12] which state that if $S_{\mu\nu}(x)$ is a symmetric, local second-rank tensor operator with zero divergence, obeying certain additional conditions,[14] then the space integrals of $S_{\mu 0}$ and $(x_\mu S_{\nu 0} - x_\nu S_{\mu 0})$ will be proportional to the generators P_μ and $M_{\mu\nu}$ of the Poincaré group:

$$\int d^3x S_{\mu 0}(x) = CP_\mu, \tag{5.1}$$

$$\int d^3x(x_\mu S_{\nu 0} - x_\nu S_{\mu 0}) = CM_{\mu\nu}. \tag{5.2}$$

The source tensor $J_{\mu\nu}(x)$ of the spin-2 field $U_{\mu\nu}(x)$ has the properties required of the operator $S_{\mu\nu}(x)$, as noted in Ref. 1. We have pointed out there that this may be used for fixing the normalization of $J_{\mu\nu}(x)$, by requiring that

$$\int d^3x J_{\mu 0}(x) = P_0. \tag{5.3}$$

The results of Refs. 11 and 12 then imply that we should be able to write $J_{\mu 0}(x)$ in the form

$$J_{\mu 0}(x) = \Theta_{\mu 0}(x) + B_{\mu 0}(x), \tag{5.4}$$

such that $B_{\mu 0}(x)$ obeys the conditions

$$\int d^3x B_{\mu 0}(x) = 0, \tag{5.5}$$

$$\int d^3x[x_\mu B_{\nu 0}(x) - x_\nu B_{\mu 0}(x)] = 0. \tag{5.6}$$

We shall now check that the properties of the source $J_{\mu\nu}$ are consistent with this requirement. Using the field-source identity (2.10) in Eq. (2.17) for $U_{\mu\nu}$, we obtain the following:

$$(\Box^2 + m^2)\Theta_{\mu\nu} = m^2\left(J_{\mu\nu} - \frac{1}{3m^2}(\partial_\mu\partial_\nu - \Box^2\eta_{\mu\nu})j\right), \tag{5.7}$$

which leads to

$$(\Box^2 + m^2)\Theta(x) = (\Box^2 + m^2)j(x), \tag{5.8}$$

where Θ denotes the trace Θ^σ_σ. This implies

$$j(x) = \Theta(x) + \chi(x), \tag{5.9}$$

where $\chi(x)$ obeys the equation

$$(\Box^2 + m^2)\chi(x) = 0. \tag{5.10}$$

We therefore obtain from (5.7) the equation

$$J_{\mu\nu} = \Theta_{\mu\nu} + B_{\mu\nu}, \tag{5.11}$$

where

$$B_{\mu\nu} = (\Box^2/m^2)\Theta_{\mu\nu} + (1/3m^2)(\partial_\mu\partial_\nu - \Box^2\eta_{\mu\nu})(\Theta + \chi). \tag{5.12}$$

We must now check whether $B_{\mu 0}$ has the properties (5.5) and (5.6). We first write

$$\int d^3x B_{i0}(x) = \frac{\partial_0^2}{m^2}\int d^3x\Theta_{i0}(x) - \frac{1}{m^2}\int d^3x\nabla^2\Theta_{i0}(x) + \frac{1}{3m^2}\partial_0\int d^3x\partial_i[\Theta(x) + \chi(x)]. \tag{5.13}$$

The first term is of the form $(1/m^2)\partial_0^2 P_i$ and vanishes (by momentum conservation). The remaining terms vanish by Gauss's theorem if we assume the vanishing of the surface integrals at infinity. We thus obtain

$$\int d^3x B_{i0}(x) = 0. \tag{5.14}$$

Similarly, we write

$$\int d^3x B_{00}(x) = (\partial_0^2/m^2)P_0 - (1/m^2)\int d^3x\nabla^2 \times \{\Theta_{00}(x) - \tfrac{1}{3}[\Theta(x) + \chi(x)]\} = 0. \tag{5.15}$$

Therefore, the requirement (5.5) is satisfied. We next write

$$\begin{aligned}\int d^3x(x_iB_{j0} - x_jB_{i0}) &= \partial_0^2\int d^3x(x_i\Theta_{j0} - x_j\Theta_{i0}) \\ &\quad + \int d^3x(x_i\nabla^2\Theta_{j0} - x_j\nabla^2\Theta_{i0}) \\ &\quad + \tfrac{1}{3}\partial_0\int d^3x(x_i\partial_j - x_j\partial_i)(\Theta + \chi) \\ &= \partial_0^2 M_{ij} = 0,\end{aligned} \tag{5.16}$$

where we have integrated by parts, set surface terms equal to zero, and have used angular momentum conservation.

We finally write

$$\int d^3x(x_iB_{00} - x_0B_{i0})$$
$$= \partial_0^2 M_{i0} - \int d^3x x_i \nabla^2 \Theta_{00} + x_0 \int d^3x \nabla^2 \Theta_{i0}$$
$$+ \tfrac{1}{3}\int d^3x x_i \nabla^2(\Theta + \chi) - \tfrac{1}{3}x_0\partial_0 \int d^3x \partial_i(\Theta + \chi)$$
$$= \partial_0^2 M_{i0} = -[P_0, [P_0, M_{i0}]] = i[P_0, P_i] = 0. \tag{5.17}$$

This shows that $J_{\mu 0}$ can be written in the form (5.4), with $B_{\mu 0}$ satisfying the conditions (5.5) and (5.6), so that our model satisfies the requirements of Refs. 11 and 12.

In concluding, we note that there are several tensors in a spin-2 theory such that their space integrals may be used for constructing the generators of the Poincaré group. For instance, from the equation

$$gJ_{\mu\nu} = (\Box^2 + m^2)(U_{\mu\nu} - \eta_{\mu\nu}u) + (\eta_{\mu\nu}\Box^2 - \partial_\mu\partial_\nu)u$$

and the requirement that the space integrals of $J_{\mu 0}$ and $(x_\mu J_{\nu 0} - x_\nu J_{\mu 0})$ should be proportional to the generators of the Poincaré group, it follows that one may also use the tensor

$$(\Box^2 + m^2)(U_{\mu\nu} - \eta_{\mu\nu}u) \tag{5.18}$$

instead of $J_{\mu\nu}$ to construct the generators P_μ and $M_{\mu\nu}$. Our hypothesis that $(U_{\mu\nu} - \eta_{\mu\nu}u)$ is proportional to the stress tensor is consistent with this, as this hypothesis implies that the space integrals of (5.18) will also be proportional to P_μ and $M_{\mu\nu}$. A more general form of the field-source identity would be the following:

$$\Theta_{\mu\nu} = (a + b\Box^2)(U_{\mu\nu} - \eta_{\mu\nu}u) + c(\eta_{\mu\nu}\Box^2 - \partial_\mu\partial_\nu)u. \tag{5.19}$$

We have postulated the identity in the form (2.10) because of its simplicity and because it involves no arbitrary parameters.

6. CONCLUDING REMARKS

In this paper, we have obtained the constraints imposed by a field-source identity and Lorentz covariance on a spin-2 meson theory. The results obtained here indicate the nature of the restrictions on a theory in which the spin-2 field and its source are closely related to the stress tensor. Although we have not exhibited a model theory satisfying all the field-dependence requirements, the latter are given explicitly and do not appear to lead to any inconsistency, which lends support to the existence of such field theories.

As discussed in Ref. 1, the field-source identity may be extended in a straightforward way to take into account the two neutral spin-2 mesons f and f'. The constraints imposed on such theories by Lorentz covariance are a straightforward extension of the results obtained here.

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

† Address during 1970–71: Physics Department, Wesleyan University, Middletown, Connecticut 06457.

[1] K. Raman, Phys. Rev., D **2**, 1577 (1970).

[2] E.g., see H. Pagels, unpublished report, University of North Carolina; P. G. O. Freund, Phys. Letters **2**, 136 (1962); D. H. Sharp and W. G. Wagner, Phys. Rev. **131**, 1226 (1963); S. H. Patil and Y. P. Yao, *ibid.* **153**, 1455 (1966); W. Królikowski, Phys. Letters **24B**, 305 (1967); R. Delbourgo, A. Salam, and J. Strathdee, Nuovo Cimento **49A**, 593 (1967).

[3] N. M. Kroll, T. D. Lee, and B. Zumino, Phys. Rev. **157**, 1376 (1967).

[4] M. Gell-Mann and F. Zachariasen, Phys. Rev. **124**, 953 (1961); J. J. Sakurai, Ann. Phys. (N.Y.) **11**, 1 (1960).

[5] K. Raman, J. Math. Phys **11**, 2935 (1970).

[6] P. A. M. Dirac, Rev. Mod. Phys. **34**, 1 (1962); J. Schwinger, Phys. Rev. **127**, 324 (1962); **130**, 406, 800 (1963).

[7] D. G. Boulware and S. Deser, J. Math. Phys. **8**, 1468 (1967).

[8] These ETCR's are valid for fields with spin ≤ 1 and not in general for fields with higher spin. Also, there are other indications that there may be inconsistencies in treating fields with higher spin as fundamental fields. Our point of view here is the following. We shall assume that a consistent theory contains only spins 0, $\frac{1}{2}$, and 1 as fundamental fields, so that the stress tensor, for instance, is constructed only from the operators of these fields. We regard the spin-2 fields introduced here as phenomenological fields. We hope to discuss further work along these lines elsewhere. (For a discussion of phenomenological fields and effective actions, see B. Zumino, Lectures at the Brandeis University Summer Institute, 1970, CERN preprint.)

[9] T. D. Lee, S. Weinberg, and B. Zumino, Phys. Rev. Letters **18**, 1029 (1967).

[10] For example, see S. N. Gupta, Rev. Mod. Phys. **29**, 334 (1957).

[11] C. Orzalesi, J. Sucher, and C. H. Woo, Phys. Rev. Letters **21**, 1550 (1960).

[12] D. R. Divgi and C. H. Woo, Phys. Rev. Letters **23**, 1510 (1969).

[13] Note that, as discussed in Ref. 1, in a theory with a field-source identity, the g^{-1} dependence on the right-hand side of this identity would arise from the derivative coupling part of the interaction Lagrangian $\mathfrak{L}_{\text{int}}$ (for the interaction between the gravitational field and the spin-2 field). It is this that gives rise to terms of $O(g^{-2})$ in the stress-tensor ETCR, although the source $J_{\nu\mu}(x)$, which arises from the nonderivative part of $\mathfrak{L}_{\text{int}}$, is assumed to be nonsingular in g. We have pointed out in Ref. 1 that a theory with a field-source identity is possible only when there is a nonzero interaction, so that $g \neq 0$.

[14] For a statement of the conditions, we refer the reader to Ref. 12.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Singularities in Forward Multiparticle Scattering Amplitudes and the S-Matrix Interpretation of Higher Virial Coefficients

ROGER DASHEN*
Institute for Advanced Study, Princeton, New Jersey 08540
AND
SHANG-KENG MA†‡
Institute for Advanced Study, Princeton, New Jersey 08540
and
Department of Physics and Institute for Pure and Applied Physical Sciences, University of California San Diego, La Jolla, California 92037

(Received 8 October 1970)

The role of singular forward multiparticle scattering amplitudes in S-matrix formulas for the higher virial coefficients is studied in detail. It is shown that by means of a simple limiting process one can give a precise, unambiguous meaning to the traces over on-shell scattering amplitudes which appear in our previous formulas for the virial coefficients. The general arguments, in which we maintain a reasonable level of rigor, are supported by explicit calculations for the third and fourth virial coefficients. Also, the angular momentum expansion used in earlier work is shown to converge.

1. INTRODUCTION

It is intuitively clear that the thermodynamic properties of a dilute interacting gas can be understood in terms of the collisions involving a small number of particles.[1,2] The concept of the scattering matrix, which gives a complete description of the collision process, therefore enters into statistical mechanics.

By way of the virial expansion, the role of the S matrix in statistical mechanics was examined in a previous paper,[3] hereafter referred to as I. A simple prescription for calculating the virial coefficients in terms of S-matrix elements was derived.

This prescription, stripped of details, reads

$$b_n \sim \int d\epsilon e^{-\beta\epsilon} \operatorname{Im} \operatorname{Tr} S^{-1} \frac{\overleftrightarrow{\partial}}{\partial \epsilon} S, \qquad (1.1)$$

and thus relates the n-particle S matrix at c.m. energy ϵ to the coefficients b_n from which the virial series can be obtained readily. A generalized form of (1.1) gives a basis for extrapolating toward a statistical mechanics of relativistic gases. The qualitative and physical aspects of (1.1) and some simple applications were discussed.

It was also pointed out in I that an annoying problem remains in (1.1) even in its simplest nonrelativistic form. The problem is as follows. One wants (1.1) to depend only on strictly on-shell scattering amplitudes. In our original derivation $\operatorname{Tr} S^{-1}(\overleftrightarrow{\partial} S/\partial\epsilon)$ is supposed to be understood as limit as $E \to \epsilon + i0$ of $\operatorname{Tr} S^{-1}(E)[\overleftrightarrow{\partial} S(E)/\partial E]$ where E is a complex off-shell energy whose on-shell value is ϵ. One can easily see from the discussion in I that $\operatorname{Tr} S^{-1}(\overleftrightarrow{\partial} S/\partial\epsilon)$ as defined this way will depend only on on-shell quantities provided that the scattering amplitude or T matrix does not become infinite anywhere on the energy shell. If the T matrix does blow up somewhere on the energy shell, a special discussion is required. These remarks are relevant for physics because, while the two-body T matrix is always finite in the physical scattering region, for three or more bodies the on-shell T matrix has poles in the physical region. A typical diagram with this property is shown in Fig. 1. The origin and nature of these singularities are discussed below, here we simply note that they indicate a possible difficulty in a program directed towards expressing all the b_n in terms of on-shell amplitudes. In I we suggested a method for handling these singular pieces of multiparticle amplitudes and concluded that, in principle, everything was all right and b_n could indeed be written in terms of on-shell amplitudes. This analysis was, however, not rigorous. In a later paper,[4] hereafter called II, we studied b_3 in considerable detail and came to the same conclusion for this particular case. Unfortunately, the more rigorous methods of II were cumbersome and not easily generalizable to arbitrary n. The purpose of the present paper is to complete the program. We will show that in spite of the poles in the multiparticle amplitudes, all the b_n can in fact be calculated from on-shell quantities. Also we give a recipe for computing b_n which is, in principle, very simple and elegant. In practice, however, our present formulas will probably not be overly useful. This leads us to remark that although we have now shown that all the b_n are determined by on-shell scattering data, there is still a lot of work to be done on the formalism. What we feel is needed is a hard study of the qualitative properties of scattering amplitudes involving several

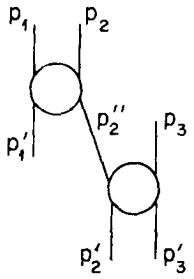

FIG. 1. A typical diagram for the scattering amplitude which becomes singular for forward scattering, where $\mathbf{p}_i = \mathbf{p}_i'$ and the intermediate state energy denominator vanishes, as a consequence of momentum conservation.

particles, with the possible goal of finding a rearrangement of the virial series which would be more useful in practice (e.g., more rapidly convergent in certain cases). This more difficult problem will only be touched upon in the present paper.

Let us now be more specific.

First we have to understand where the above mentioned singularities come from and what they look like. Consider as an illustration the on-shell scattering amplitude for three particles of equal mass. Let the initial momenta be $\mathbf{p}_1'$, $\mathbf{p}_2'$, and $\mathbf{p}_3'$ and the final momenta be $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$, where $\sum \mathbf{p}_i = \sum \mathbf{p}_i'$ and $\sum p_i^2 = \sum p_i'^2$ since we are on shell. This amplitude blows up when $p_1^2 + p_2^2 = p_1'^2 + (\mathbf{p}_1' - \mathbf{p}_1 - \mathbf{p}_2)^2$. This singularity, which has the analytic form of a pole, is a reflection of the fact that one way to make a three-particle event is to have two successive two-particle collisions $2' + 3' \to 2'' + 3$, followed by $2'' + 1' \to 2 + 1$, with $\mathbf{p}_2'' = \mathbf{p}_2 + \mathbf{p}_1 - \mathbf{p}_1'$. (See Fig. 1.) In a kinematic configuration where the relation $p_1^2 + p_2^2 = p_1'^2 + (\mathbf{p}_1' - \mathbf{p}_1 - \mathbf{p}_2)^2$ holds, energy can be conserved in each two-body event separately. When this happens the two collisions may occur arbitrarily far apart in time and the S-matrix as conventionally defined blows up. In particular, we note that in the forward direction where $\mathbf{p}_1 = \mathbf{p}_1'$ and $\mathbf{p}_2 = \mathbf{p}_2'$, one is sitting right on top of the pole. Unfortunately (1.1) contains, among other things, integrals over forward scattering amplitudes. This does not, of course, imply a divergence in b_3. It simply means that one must carefully specify the on-shell interpretation of (1.1).

We have discovered a simple modification of (1.2) that makes the on-shell nature of the S-matrix expressions for b_n explicit. It is to write

$$b_n \sim \lim_{\theta \to 0} \int d\epsilon e^{-\beta\epsilon} \operatorname{Im} \operatorname{Tr} \left(S^+(\epsilon) \frac{\overleftrightarrow{\partial}}{\partial \epsilon} S(\epsilon) e^{i\theta J_3} \right), \quad (1.2)$$

where $e^{i\theta J_3}$ is a rotation around the third axis in the c.m. frame and ϵ is now explicitly the on-shell energy. The rule is to do the trace and integration over ϵ before taking the limit as $\theta \to 0$. Because of rotational invariance, which is assumed in the derivation of (1.2), the direction of the third axis is entirely arbitrary. The reason that S can immediately be placed on shell in (1.2) but not in (1.1) is that with the rotation $e^{i\theta J_3}$ inserted, the singular forward amplitudes are avoided.

It should *not* be thought that (1.1) is wrong or fails to give an onshell formula for b_n. If one is careful with the contour of the ϵ-integration in (1.1), this formula gives results identical to (1.2). As stated above, the latter is explicitly on-shell. Notice that a forward amplitude in (1.1) will be replaced by a limit from a nonforward direction when one goes to (1.2). Thus, one might anticipate that, upon carefully doing the ϵ integration over the forward amplitudes in (1.1), a new set of terms related to on-shell scattering arbitrarily close to, but not exactly in, the forward direction will arise. This is indeed the case. It is, however, not a simple task to recognize these terms. This is why the on-shell nature of (1.1) is not transparent.

The method of treating (1.1) suggested in I and worked out in detail for b_3 in II was to take the trace for each fixed angular momentum J and then to sum over J. For a given J, the T matrix is sufficiently regular on the energy shell so that none of the difficulties described above appear. One then has to verify that the sum over J converges and gives the right answer. For b_3 this was done in II. The angular momentum method is very closely related to the $\theta \to 0$ limit in (1.2). The precise connection is explained in the text.

In previous work on the connection between the b_n and the S matrix, it appears either that the existence of singular forward scattering amplitudes has been ignored or else that the treatment has been restricted to a fixed angular momentum J, ignoring questions about the convergence of the sum over J.

We are aware that a reasonable amount of mathematical care and rigor is needed to firmly establish an equation like (1.2). For this reason we have taken time to give rigorous justification for those steps in the derivation of (1.1) which might appear to be particularly questionable. In an appendix we use path integral techniques to prove that b_n can be written as the limit of a trace where, as in (1.2), the final state is rotated relative to the initial state. Armed with this result, we can obtain (1.2) in exactly the same way that (1.1) was obtained earlier. The steps in the latter calculation are all well defined if one remembers to first establish the formula in a finite volume and then take an infinite volume limit. We are careful to show that the limit as $\theta \to 0$ can be interchanged with a limit of infinite volume. As a byproduct we are able to prove that the angular momentum series of b_n converges. While there is room for improvement, we

feel that the level of rigor maintained here should satisfy most readers.

In addition to obtaining the general results described above, we illustrate the situation with several specific calculations for b_3 and b_4. For b_3, we compute the contribution of all singular pieces of $S^\dagger \overleftrightarrow{(\partial S/\partial \epsilon)}$. We do this both by using the explicitly on-shell small angle limit (1.2) and by carefully doing the ϵ integration in (1.1). As expected, the two methods agree and also agree with the angular momentum sum worked out in II. The next coefficient b_4 brings in some qualitatively new effects. Again we calculate the contributions of singular amplitudes using both (1.1) and (1.2). Again both methods agree. We consider these explicit computations as strong evidence that our basic approach is indeed correct.

Due to the already extreme length of this paper, we have chosen to omit any discussion of applications. We would, however, like to state the following fact without proof or supporting discussion. It is that the forward singularities in $S^\dagger(\partial S/\partial \epsilon)$ have the interesting property that they give the leading contribution to b_n at low temperatures. As mentioned above, these singularities are the result of two or more successive, well separated collisions each involving fewer than n particles. Therefore, the pieces of b_n which are largest for small temperatures depend only on the scatterings of fewer than n particles. In particular, the largest piece of b_n comes from successive two-body collisions and can be expressed as a function of the two-body scattering lengths above.[5] This might turn out to have some useful consequences.

The paper is organized as follows. In the next section we review some basic facts about the virial coefficients and state some results about the $\theta \to 0$ limit. Our basic formulas and results are also stated there. Derivations given in I are, however, not repeated. Section 3 is basically a prelude to Secs. 4 and 5, where the explicit computations for b_3 and b_4 are discussed. In the case of b_4, most of the detailed work is relegated to appendices.

We will restrict ourselves to nonrelativistic particles of a single species. The extension to relativistic situations discussed in I is straightforward. For simplicity, we will also assume that there are no bound states, and we will take account of Fermi or Bose statistics only when it is important to do so.

2. BACKGROUND AND SOME BASIC RESULTS

This section is intended to provide a fairly rigorous foundation for our later discussions of the virial series from the point of view of scattering theory. We will really not be exclusively concerned with scattering theory here. Rather, we begin by reviewing some of what is known about the virial expansion in the more usual U-function formalism. We then state a few simple but basic results which are necessary to justify the $\theta \to 0$ limit described in the Introduction. The proofs are relegated to appendices. Then we review our previous results concerning the connection between b_n and the S matrix and, finally, state our main result.

A. Definitions

We are interested in a nonrelativistic gas of volume V, temperature β^{-1}, chemical potential μ, and pressure P. The power series expansion for the grand potential $\Omega = -PV$ is

$$\Omega = \Omega_0 - V\beta^{-1}\lambda^{-3}\sum_{n=2}^{\infty} b_n(V)Z^n, \tag{2.1}$$

$$\lambda^{-1} = (m/2\pi\beta)^{\frac{1}{2}}, \tag{2.2}$$

where $Z = e^{\beta\mu}$, $\Omega_0 = -V\beta^{-1}\lambda^{-3}Z$ is the grand potential for an ideal gas, and m is the mass of a gas molecule. In (2.1), the coefficients depend on V: We are only interested in

$$b_n \equiv \lim_{V\to\infty} b_n(V). \tag{2.3}$$

With suitable restrictions on the interactions, this limit can be rigorously shown to exist.[6] We consider only cases where (2.3) makes sense, and we will not concern ourselves further with questions having to do with interchanging the limit $V \to \infty$ and the sum over n in (2.1). This latter point has been extensively discussed in the literature.[6] We pause only to say that the ability to interchange the sum and limit essentially defines the gas phase.

Our object of study, then, is b_n as defined by (2.1) and (2.3). We recall that b_n depends on the dynamics of a finite number (n) of particles in an infinite volume. This is the basic reason why b_n can be expressed in terms of on-shell scattering amplitudes.

The usual expression for b_n involves a function

$$\langle \mathbf{x}_1 \cdots \mathbf{x}_n | \, U_n \, | \mathbf{y}_1 \cdots \mathbf{y}_n \rangle \equiv \text{connected part of } \{\langle \mathbf{x}_1 \cdots \mathbf{x}_n | \, e^{-\beta H_n} \, | \mathbf{y}_1 \cdots \mathbf{y}_n \rangle\}, \tag{2.4}$$

where H_n is the n-body Hamiltonian and "connected part of" means to subtract out products of the U_m for $m < n$ in the usual way. The algorithm for constructing U_n is well known so we need not go into the details, except to remark that in perturbation theory language U_n is simply the sum of all connected diagrams.

It is convenient to factor the c.m. dependence out of U_n. We define

$$\langle \mathbf{x}_1 \cdots \mathbf{x}_n | \, U_n \, | \mathbf{y}_1 \cdots \mathbf{y}_n \rangle = n^{\frac{3}{2}} \lambda^{-3} \exp\left[-(2nm/\beta)(\mathbf{X} - \mathbf{Y})^2\right] \times \langle \{\mathbf{x}\} | \, U_n \, | \{\mathbf{y}\} \rangle_{\text{c.m.}}, \quad (2.5)$$

where $n\mathbf{X} = \sum \mathbf{x}_i$, $n\mathbf{Y} = \sum \mathbf{y}_i$ and the matrix element with the "c.m." subscript is independent of $\mathbf{X}$ and $\mathbf{Y}$. The labels $\{\mathbf{x}\}$ and $\{\mathbf{y}\}$ on the states are supposed to stand for any pairs of $3n - 3$ coordinates specifying the relative distances between particles in the c.m. frame. It is easiest to think of $\{x\}$ as the full set of coordinates $\mathbf{x}_i \cdots \mathbf{x}_n$ subject to the constraint $\sum \mathbf{x}_i = 0$, which obviously leaves $3n - 3$ variables. The object b_n is then given by

$$b_n = (n^{\frac{3}{2}}/n!) \int d\mu_x \langle \{\mathbf{x}\} | \, U_n \, | \{\mathbf{x}\} \rangle_{\text{c.m.}}, \quad (2.6)$$

where $d\mu_x$ is a $3n - 3$ dimensional integration in the $3n$ dimensional space of the $\mathbf{x}$'s. Because the integrand in (2.6) is translationally invariant, one has a certain amount of freedom in defining this integration. For our purposes it is convenient to set

$$d\mu_x = \delta^3(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n) d^3\mathbf{x}_1 d^3\mathbf{x}_2 \cdots d^3\mathbf{x}_n. \quad (2.7)$$

The integration runs between $+\infty$ and $-\infty$ for all the $\mathbf{x}$'s.

We will also need the expression for b_n in terms of momentum space variables. We define

$$\langle \mathbf{p} | \, U_n \, | \mathbf{q} \rangle_{\text{c.m.}} \equiv \int d\mu_x \, d\mu_y e^{-i(\sum \mathbf{p}_i \cdot \mathbf{x}_i - \sum \mathbf{q}_i \cdot \mathbf{y}_i)} \langle \{\mathbf{x}\} | \, U_n \, | \{\mathbf{y}\} \rangle_{\text{c.m.}}, \quad (2.8)$$

where $\mathbf{p}$ and $\mathbf{q}$ stand for the variables $\mathbf{p}_1 \cdots \mathbf{p}_n$ and $\mathbf{q}_1 \cdots \mathbf{q}_n$, *subject to the c.m. constraints* $\sum \mathbf{p}_i = 0$ *and* $\sum \mathbf{q}_i = 0$. In many future formulas this c.m. constraint will be implicit. It is important to keep this in mind. Clearly, one can write

$$b_n = (n^{\frac{3}{2}}/n!) \int d\mathbf{p} \langle \mathbf{p} | \, U_n \, | \mathbf{p} \rangle_{\text{c.m.}}, \quad (2.9)$$

where

$$d\mathbf{p} = (2\pi)^3 \delta^3(\mathbf{p}_1 + \mathbf{p}_2 + \cdots + \mathbf{p}_n) \times \left(\frac{d^3p_1}{(2\pi)^3} \frac{d^3p_2}{(2\pi)^3} \cdots \frac{d^3p_n}{(2\pi)^3} \right). \quad (2.10)$$

Equations (2.6) and (2.9) are neatly summarized by

$$b_n = (n^{\frac{3}{2}}/n!) \, \text{Tr} \, (U_n)_{\text{c.m.}}, \quad (2.11)$$

where the meaning of the trace over c.m. variables should be obvious.

B. The Small Angle and Infinite Volume Limits

We define

$$F_n(\theta) = \text{Tr} \, (U_n e^{i\theta J_3})_{\text{c.m.}} = \int \langle \{\mathbf{x}\} | \, U_n \, | \{R_\theta \mathbf{x}\} \rangle_{\text{c.m.}} \, d\mu_x = \int \langle \mathbf{p} | \, U_n \, | R_\theta \mathbf{p} \rangle_{\text{c.m.}} \, d\mathbf{p}, \quad (2.12)$$

where $e^{i\theta J_3}$ is a rotation about the third axis in the center of mass system and $\{R_\theta \mathbf{x}\}$ and $R_\theta \mathbf{p}$ are the correspondingly rotated coordinates and momenta. In all cases of physical interest, U_n is rotationally invariant so that $e^{i\theta J_3}$ can be replaced by a similar rotation about any axis without changing $F_n(\theta)$.

In Appendix A, it is shown that for potentials which fall off rapidly at large separations between particles, $F_n(\theta)$ is continuous at $\theta = 0$, i.e.,

$$b_n = (n^{\frac{3}{2}}/n!) F_n(0) = \lim_{\theta \to 0} (n^{\frac{3}{2}}/n!) F_n(\theta). \quad (2.13)$$

Specifically, we prove (2.13) for potentials of strictly finite range or which fall off exponentially with interparticle distance. Actually, (2.13) probably holds for all potentials of interest to us, i.e., those potentials which are well enough behaved to construct a scattering theory of the usual kind. This will become clear when the reader sees how we use this result.

Equation (2.13) is not too surprising. From a physicist's point of view, the only possible worry might be that $F_n(\theta)$ is defined as a trace over a system which is allowed to occupy an infinite volume. If the volume were finite, (2.13) would be essentially trivial. For this and other reasons to become apparent later, it is useful to record a result about the infinite volume limit.

Consider an artificial system confined to a sphere in the center of mass system defined by $\sum_i x_i^2 < \Lambda^2$ (remember $\sum_i x_i = 0$). The boundary condition is to be that the wavefunction vanishes at $\sum x_i^2 = \Lambda^2$. Let $U_n(\Lambda)$ be the new U function and define

$$\mathcal{F}_n(\theta, \Lambda) = \text{Tr} \, [U_n(\Lambda) e^{i\theta J_3}]_{\text{c.m.}} = \int_{\sum x_1^2 < \Lambda^2} \langle \{\mathbf{x}\} | \, U_n(\Lambda) \, | \{R_\theta \mathbf{x}\} \rangle \, d\mu_x. \quad (2.14)$$

For potentials of the type described above, it is shown in Appendix B that

$$\lim_{\Lambda \to \infty} \mathcal{F}_n(\theta, \Lambda) = F_n(\theta)$$

and

$$\lim_{\Lambda \to \infty} \mathcal{F}_n(0, \Lambda) = F_n(0), \quad (2.15)$$

or

$$b_n = \lim_{\theta\to 0}\left(\lim_{\Lambda\to\infty}(n^{\frac{3}{2}}/n!)\mathcal{F}_n(\theta,\Lambda)\right) = \lim_{\Lambda\to\infty}\left(\lim_{\theta\to 0}(n^{\frac{3}{2}}/n!)\mathcal{F}_n(\theta,\Lambda)\right).$$

This result is useful because it is sometimes convenient to manipulate expressions for F_n with a finite Λ, taking the limit as $\Lambda \to \infty$ at the end of the calculation. According to (2.15), we can always freely interchange the limits $\theta \to 0$ and $\Lambda \to \infty$.

C. The Angular Momentum Series for b_n

As mentioned in the Introduction, it is often useful to compute b_n in an angular momentum basis. That is, one writes

$$b_n = \sum_J b_{Jn} \tag{2.16}$$

where b_{Jn} is that part of b_n which comes from states whose c.m. angular momentum is J. We now turn to the properties of the b_{Jn} and the series (2.16). Among other things, we can use (2.13) to prove that (2.16) converges, a result which makes our previous work on b_3 (see II) completely rigorous.

Suppose we want to break the trace of a rotationally symmetrical operator A up into pieces coming from states of fixed J. This may be done by means of the projection operator

$$P_J = (8\pi^2)^{-1}(2J+1)\sum_M \int dR \mathfrak{D}^{J*}_{MM}(R)U(R), \tag{2.17}$$

where $U(R)$ is the Hilbert space transformation corresponding to a rotation R, $\mathfrak{D}^J_{MM}(R)$ are the usual rotation matrices for spin J, and the integration is over the rotation group. By definition, one has

$$\operatorname{Tr} A = \sum_J \operatorname{Tr}_J A,$$
$$\operatorname{Tr}_J A = \operatorname{Tr} AP_J. \tag{2.18}$$

If A is rotationally invariant so that $U(R)AU^{-1}(R) = A$, the above formulas can be considerably simplified. Consider

$$\operatorname{Tr}_J A = (8\pi^2)^{-1}(2J+1)\int dR \sum_m \mathfrak{D}^{J*}_{mm}(R)\operatorname{Tr}[AU(R)] \tag{2.19}$$

for rotationally symmetric A. We can parametrize the rotations R by a unit vector $\mathbf{e}$ and an angle θ, corresponding to a rotation through an angle θ around the $\mathbf{e}$ axis. Now $\sum_m \mathfrak{D}^{J*}_{mm}(R)$ and $\operatorname{Tr}[AU(R)]$ are themselves rotationally symmetric and can therefore depend only on θ, not on $\mathbf{e}$. Hence we can do the $\mathbf{e}$ integration in (2.19) trivially leaving only a one-dimensional integration over θ. All we have to do is compute the volume element dR in terms of $\mathbf{e}$ and θ, and then compute $\sum_m \mathfrak{D}^J_{mm}$ as a function of θ. The calculation is standard and the answer is

$$\operatorname{Tr}_J A = (2\pi)^{-1}(2J+1) \times \int_{-\pi}^{\pi} d\theta(1-\cos\theta)\chi_J(\theta)\operatorname{Tr}(Ae^{i\theta J_3}), \tag{2.20}$$

where

$$\chi_J(\theta) = \sum_{m=-J}^{J} e^{i\theta m}. \tag{2.21}$$

Obviously, we could have replaced $e^{i\theta J_3}$ in (2.20) by a rotation around any axis, since $\operatorname{Tr}(Ae^{i\theta \mathbf{e}\cdot\mathbf{J}})$ is independent of $\mathbf{e}$ for $\mathbf{e}^2 = 1$ and A rotationally symmetric.

Let us now apply (2.20) to the virial coefficients. From (2.10) and (2.12) we have

$$b_{Jn} = (n^{\frac{3}{2}}/2\pi n!)(2J+1)\int_{-\pi}^{\pi} d\theta(1-\cos\theta)\chi_J(\theta)F_n(\theta). \tag{2.22}$$

The completeness relation

$$(2\pi)^{-1}\sum_J (2J+1)(1-\cos\theta)\chi_J(\theta) = \delta(\theta), \tag{2.23}$$

plus the fact proven above that $F_n(\theta)$ is continuous at $\theta = 0$, proves the convergence of (2.16), according to the usual theorems on orthogonal series. Note also that, according to (2.15), the sum over J can be interchanged with the limit of infinite volume ($\Lambda \to \infty$).

Having shown that $\sum_J b_{Jn}$ converges in the mathematical sense, it is still desirable to have some idea about rate at which (2.17) converges. One expects to be able to neglect the b_{Jn} in (2.16) for $J \gg \bar{J}$, where $\bar{J}$ is a typical angular momentum in the states $|\{\mathbf{x}\}\rangle_{\text{c.m.}}$ in (2.6). Let $\bar{p}$ be a typical momentum and d be a typical distance between particles, then

$$\bar{J} \sim n\bar{p}d. \tag{2.24}$$

To estimate d, we have to consider two effects. The first is the range of the interaction, call it a. If two particles are separated by a distance less than a, they will certainly contribute to b_n. The other effect is that particle 1, for example, can interact with particle 2 and then propagate freely over to particle 3 where it interacts again. The relevant distance here is the distance over which particle two can propagate. From the well known imaginary-time analogy, we know that its distance can be estimated from the single particle diffusion equation

$$\frac{\partial\psi}{\partial\beta} = -\frac{1}{2m}\nabla^2\psi.$$

Over a time β, a particle can diffuse a distance $(m/\beta)^{\frac{1}{2}} \sim \lambda$, the thermal wavelength. Hence the second length is λ and combining the two effects, we get

$$d \sim a + \lambda. \tag{2.25}$$

Of course, $\bar{p}$ is of order λ^{-1}, which finally gives

$$J \sim n(a/\lambda + 1). \tag{2.26}$$

Note that $\bar{J}$ increases with n. Also $\bar{J}$ does not tend to zero as $\lambda \to \infty$, i.e., at zero temperature. This, as will be seen later, is a consequence of the singular amplitudes in the on-shell scattering formulas. For $n = 2$, the second effect mentioned above does not exist, hence

$$\bar{J}_2 \sim 2(a/\lambda), \tag{2.27}$$

as one would expect from the Beth–Uhlenbeck expression for b_2 in terms of two-body phase shifts. Note that intuition obtained from looking only at the two-body case is misleading here.

D. Off-Shell S-Matrix Formulas

In I it was shown that

$$\mathrm{Tr}(U_n)_{\mathrm{c.m.}} = \frac{1}{4\pi i}\int e^{-\beta E}\,\mathrm{Tr}\left(AS_n^{-1}(E)\frac{\overleftrightarrow{\partial}}{\partial E}S_n(E)\right)_{\mathrm{c,c.m.}} dE, \tag{2.28}$$

where A is a symmetrization or antisymmetrization operator explained in I, and the off-shell S matrix is defined by

$$\begin{aligned} S(E) &= 1 + [(E + i\eta - H_0)^{-1} \\ &\quad - (E - i\eta - H_0)]T(E), \\ S^{-1}(E) &= 1 - [(E + i\eta - H_0)^{-1} \\ &\quad - (E - i\eta - H_0)^{-1}]T^{\dagger}(E), \\ T(E) &= V + V(E + i\eta - H)V, \\ T^{\dagger}(E) &= V + V(E - i\eta - H)V. \end{aligned} \tag{2.29}$$

H_0 is a free particle Hamiltonian

$$V = H - H_0.$$

The object η is a positive infinitesimal which serves to define the integration contour in (2.28). (We use E rather than ϵ for the energy to remind ourselves that we are off-shell. This convention will not be followed in later sections.) The subscript "c.m." means the S matrix with the center of mass motion factored out. Specifically, the H in (2.29) is the total Hamiltonian minus the kinetic energy of the c.m. Thus,

$$\left(AS^{-1}\frac{\overleftrightarrow{\partial}S}{\partial E}\right)_{\mathrm{c.m.}}$$

contains no overall momentum conservation delta functions. The subscript "c" means to take the connected part. The algorithm for obtaining

$$\left(AS_n^{-1}(E)\frac{\overleftrightarrow{\partial}}{\partial E}S_n(E)\right)_{\mathrm{c}}$$

from $S_n^{-1}[\overleftrightarrow{\partial}S_n(E)/\partial E]$ itself is identical to the algorithm for obtaining U_n from the W functions. For example, for $n = 3$ and Maxwell–Boltzmann statistics such that $A = 1$, one has

$$\begin{aligned} \left(S_3^{-1}\frac{\overleftrightarrow{\partial}}{\partial E}S_3\right)_{\mathrm{c}} &= S_3^{-1}\frac{\partial}{\partial E}S_3 - S_2^{-1}(1,2)\frac{\overleftrightarrow{\partial}}{\partial E}S_2(1,2) \\ &\quad - S_2^{-1}(1,3)\frac{\overleftrightarrow{\partial}}{\partial E}S_2(1,3) \\ &\quad - S_2^{-1}(2,3)\frac{\overleftrightarrow{\partial}}{\partial E}S_2(2,3), \end{aligned} \tag{2.30}$$

where $S_2(i,j)$ is the two-particle S matrix for particles i and j, it is obtained by setting $V = V_2(i,j)$ in (2.29). In general, U_n is obtained from W_n ($W_n \sim e^{-\beta H_n}$) by subtracting off terms of the form $e^{-\beta H'}$, where H' is equal to H_n minus the interaction terms connecting two or more clusters of particles, and then adding in certain exchange terms. Correspondingly

$$\left(AS_n^{-1}\frac{\overleftrightarrow{\partial}S_n}{\partial E}\right)_{\mathrm{c}}$$

is obtained from $S_n^{-1}(\overleftrightarrow{\partial}S_n/\partial E)$ by subtracting off terms of the form $S'^{\dagger}\overleftrightarrow{\partial}S'/\partial E$ where S' is defined by (2.29), with $H = H'$, and then adding exchange terms. In perturbation theory language $[AS_n^{-1}(\overleftrightarrow{\partial}S_n/\partial E)]_{\mathrm{c}}$ is just the sum of all *connected* diagrams contributing to $AS_n^{-1}(\overleftrightarrow{\partial}S_n/\partial E)$. It is easiest to think of $[AS_n^{-1}(\partial S_n/\partial E)]_{\mathrm{c}}$ in this perturbation theoretic sense, but its definition is in no way dependent on perturbation theory.

The derivation of (2.28) given in I involved the cyclic invariance property [Tr (ABC) = Tr (CAB)] of the trace. To do things properly, one should use a finite volume so that all traces are convergent and well defined. The limit of infinite volume is then taken after (2.28) has been established for a finite volume. We are implicitly assuming, then, that

$$\left(AS_n^{-1}\frac{\overleftrightarrow{\partial}}{\partial E}S_n\right)_{\mathrm{c}}$$

has a sensible limit as the volume tends to infinity. This is certainly true to any finite order in perturbation theory and with a little thinking the reader will easily see, in fact, that this must be the case if scattering theory is to make any sense at all. For $n = 3$, the proofs that the Fadeev equations have reasonable solutions provide a rigorous justification. No doubt, similar rigorous results could be obtained for $n > 3$. We are, of course, assuming that the potentials fall of fast enough at large separations between particles so that the S matrix as usually defined exists. Thus, Coulomb interactions for example are not allowed in the present discussion.

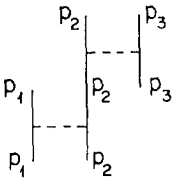

FIG. 2. A lowest-order singular diagram. The dashed lines denote two-body potentials.

Since

$$(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1} \sim -2\pi i\delta(E - H_0)$$

for infinitesimal η, it appears that (2.28) depends only on on-shell matrix elements of T. Upon closer inspection this is not so obvious. If the relevant matrix elements of $T(E)$ are themselves singular as E approaches the energy shell, one cannot simply interpret $(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1}$ as a delta function, but must carefully investigate the manner in which the contour in (2.28) winds its way around the singularities of the integrand. The T-matrix elements are singular, but nevertheless (2.28) does depend only on on-shell scattering quantities. This is the main point of the present paper. We remind the reader that singular T-matrix elements are in no way an indication of infinities in b_n. The only difficulty one encounters is that these singularities in T make it harder to show that b_n can be obtained from strictly on-shell quantities. The small angle limit is designed to accomplish the latter.

Before giving explicit on-shell formulas, it is well to look at a simple example of a singularity in T such as that pictured in Fig. 2. Let $n = 3$, and suppose that particles 1 and 2 interact with a potential V_{12}, particles 2 and 3 interact through V_{23}, and that there is no direct interaction between 1 and 3. Exchange will be ignored, so A is set equal to 1. Keeping only terms of order V_{12}, V_{23}, and $V_{12}V_{23}$, the three-body T matrix is

$$T_3(E) = V_{12} + V_{23} + V_{12}(E + i\eta - H_0)V_{23} + V_{23}(E + i\eta - H_0)V_{12} + O((V_{12})^2, (V_{23})^2). \quad (2.31)$$

To compute

$$\left(S_3^{-1}\frac{\overleftrightarrow{\partial}}{\partial E}S_3\right)_c$$

we need the two-body T matrices which to this order are

$$T_2(1, 2) = V_{12} + O[(V_{12})^2],$$
$$T_2(2, 3) = V_{23} + O[(V_{23})^2], \quad (2.32)$$
$$T_2(1, 3) = 0.$$

Using (2.30), one readily verifies that

$$\left(S_3^{-1}(E)\frac{\overleftrightarrow{\partial}}{\partial E}S_3(E)\right)_c = \frac{1}{2}\frac{\partial}{\partial E}\{[(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1}] \times (V_{12}[(E + i\eta - H_0)^{-1} + (E - i\eta - H_0)^{-1}]V_{23} + V_{23}[(E + i\eta - H_0)^{-1} + (E - i\eta - H_0)^{-1}]V_{12})\} + O((V_{12})^2, (V_{23})^2), \quad (2.33)$$

and that

$$\langle \mathbf{p}| \left(S_3^{-1}(E)\frac{\overleftrightarrow{\partial}}{\partial E}S_3(E)\right)_c |\mathbf{p}\rangle_{\text{c.m}} = \tilde{V}_{12}(0)\tilde{V}_{23}(0)\frac{\partial}{\partial E} \times \{[E + i\eta - \epsilon(\mathbf{p})]^{-1} - [E - i\eta - \epsilon(\mathbf{p})]^{-1}\} \times \{[E + i\eta - \epsilon(\mathbf{p})]^{-1} + [E - i\eta - \epsilon(\mathbf{p})]^{-1}\}, \quad (2.34)$$

where $\epsilon(\mathbf{p}) = \sum p_i^2/2m$, and the $\tilde{V}$'s are the Fourier transforms of the V's. Note that, according to our c.m. convention, there is no overall momentum conservation delta function. Furthermore, by definition

$$\int e^{-\beta E}\,\mathrm{Tr}\left(S_3^{-1}(E)\frac{\overleftrightarrow{\partial}}{\partial E}S_3(E)\right)_{\text{c.c.m.}} dE = \beta\tilde{V}_{12}(0)\tilde{V}_{23}(0)\int dE e^{-\beta E}\int d\mathbf{p} \times \{[E + i\eta - \epsilon(\mathbf{p})]^{-1} - [E - i\eta - \epsilon(\mathbf{p})]^{-1}\} \times \{[E + i\eta - \epsilon(\mathbf{p})]^{-1} + [E - i\eta - \epsilon(\mathbf{p})]^{-1}\} + O((V_{12})^2, (V_{13})^2), \quad (2.35)$$

where $d\mathbf{p}$ is defined by (2.10). It is clear that in (2.35) we cannot set

$$(E + i\eta - \epsilon)^{-1} - (E - i\eta - \epsilon)^{-1} = -2\pi i\delta(E - \epsilon)$$

without getting into trouble. Rather one must set

$$[(E + i\eta - \epsilon)^{-1} - (E - i\eta - \epsilon)^{-1}] \times [(E + i\eta - \epsilon)^{-1} + (E - i\eta - \epsilon)^{-1}] = 2\pi i\frac{\partial}{\partial E}\delta(E - \epsilon), \quad (2.36)$$

with the result that

$$\int e^{-\beta E}\,\mathrm{Tr}\left(S_3^{-1}(E)\frac{\overleftrightarrow{\partial}}{\partial E}S_3(E)\right)_{\text{c.c.m.}} dE = 2\pi i\beta^2\tilde{V}_{12}(0)\tilde{V}_{23}(0)\int e^{-\beta\epsilon(\mathbf{p})}\,d\mathbf{p} + O((V_{12})^2, (V_{23})^2). \quad (2.37)$$

This example should give ample illustration of why one must be careful about interpreting

$$(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1}$$

as $-2\pi i\delta(E - H_0)$. It also illustrates some other general points.

(i) The $(E + i\eta - \epsilon)^{-1} + (E - i\eta - \epsilon)^{-1}$ singularity in $T_3(E)$ does not cause the trace or b_3 to be singular. It only requires that we be careful with the contour of integration.

(ii) Even though one cannot put

$$(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1} = -2\pi i\delta(E - H_0)$$

in (2.33), the trace in (2.35) still depends, to our present accuracy, only on on-shell quantities. This follows from the fact that if terms of order $(V_{12})^2$ and $(V_{23})^2$ are neglected, $\tilde{V}_{12}$ and $\tilde{V}_{23}$ are the on-shell two-body T matrices. As we shall see, an analogous but less trivial result is true in higher orders.

(iii) One can easily see that the reason that the above singularity in T_3 coincides with the singularity in $(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1}$ is that we took a forward matrix element of $[S_3^{-1}(\overleftrightarrow{\partial} S_3/\partial E)]_c$. For a general matrix element

$$\langle \mathbf{p}| \left(S_3^{-1} \frac{\overleftrightarrow{\partial}}{\partial E} S_3\right)_c |\mathbf{p}'\rangle_{\text{c.m.}},$$

the two singularities do not coincide and

$$(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1}$$

can be replaced by $-2\pi i\delta(E - H_0)$. This is why the small angle limit is useful.

Finally we record a useful variant of (2.28). By integrating by parts, one readily verifies that

$$\text{Tr}\,(U_n)_{\text{c.m.}} = \frac{\beta}{2\pi} \int e^{-\beta E}\, \text{Im Tr}\,[A \ln S_n(E)]_{\text{c,c.m.}}\, dE, \tag{2.38}$$

where the subscripts c and c.m. have the same meaning as before.

E. On-Shell S-Matrix Formulas

Using the same methods as were employed in obtaining (2.28), one can show that

$$\text{Tr}\,(U_n e^{i\theta J_3})_{\text{c.m.}} \equiv F_n(\theta) = \frac{1}{4\pi i} \int e^{-\beta E}\, \text{Tr}\left(AS^{-1}(E) \frac{\overleftrightarrow{\partial}}{\partial E} S(E) e^{i\theta J_3}\right)_{\text{c,c.m.}} dE, \tag{2.39}$$

where, as before, $e^{i\theta J_3}$ is a rotation around the third axis in the c.m. The derivation of (2.39), which will be left to the reader, depends on the assumption of rotational invariance so that $e^{i\theta J_3} S e^{-i\theta J_3} = S$, etc.

As is the case with (2.28), a proper derivation of (2.39) involves starting in a finite volume, establishing the formula and then taking the limit of infinite volume. Rotational invariance can be preserved by taking the volume to be a sphere in the c.m. frame. As discussed above, the limit of infinite volume can be shown to be interchangeable with the limit as $\theta \to 0$. Thus we may assume an infinite volume and a finite θ. The limit as $\theta \to 0$ will be taken later.

The advantage of (2.39) is that it is easy to get on to the energy shell. Since the trace of

$$\left(AS^{-1} \frac{\overleftrightarrow{\partial}}{\partial E} S e^{i\theta J_3}\right)_c$$

involves only nonforward matrix elements of

$$\left(AS^{-1} \frac{\overleftrightarrow{\partial}}{\partial E} S\right)_c,$$

the singularities in

$$(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1}$$

no longer coincide with singularities of the T matrix. Therefore, in the absence of bound states,

$$(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1}$$

can be set equal to $-2\pi i\delta(E - H_0)$. The qualifying remark about bound states is necessary. This was, however, thoroughly discussed in I and need not be gone into here. Henceforth, we assume that there are no bound states.

To give an explicitly on-shell version of (2.39), it is convenient to define some notations. Let

$$|\mathbf{p}\rangle \equiv |\mathbf{p}_1 \cdots \mathbf{p}_n\rangle_{(\sum \mathbf{p}_i = 0)} = \epsilon^{\frac{1}{2}}(2m\epsilon)^{-\frac{3}{4}(n-1)} |\epsilon\xi\rangle, \tag{2.40}$$

where $\epsilon = \sum p_i^2/2m$ is the energy and ξ stands for some $3n - 4$ dimensionless variables which, along with ϵ, specify the c.m. momenta of the n particles. A particular choice for the set ξ might be the vectors

$$\mathbf{y}_i = \mathbf{p}_i(2m\epsilon)^{-\frac{1}{2}}, \tag{2.41}$$

which are not all independent but satisfy

$$\sum \mathbf{y}_i = 0, \quad \sum \mathbf{y}_i^2 = 1. \tag{2.42}$$

Let the volume element be

$$d\mathbf{p} = \epsilon^{-1}(2m\epsilon)^{\frac{3}{2}(n-1)}\, d\epsilon\, d\xi, \tag{2.43}$$

where $d\mathbf{p}$ is defined by (2.10). For the particular choice (2.42), $d\xi$ is easily shown to be

$$d\xi = (2\pi)^3 \delta(\sum \mathbf{y}_i)\delta(\sum \mathbf{y}_i^2 - 1) \times d^3\mathbf{y}_1/(2\pi)^3 \cdots d^3\mathbf{y}_n/(2\pi)^3. \tag{2.44}$$

Finally, let

$$\mathcal{S}(\xi, \epsilon, \xi') = \delta(\xi - \xi') - 2\pi i \langle \epsilon\xi | T(\epsilon + i0) | \epsilon\xi' \rangle, \quad (2.45)$$

and

$$\mathcal{S}(\xi, \epsilon, R_\theta \xi') = \delta(\xi - R_\theta \xi') - 2\pi i \langle \epsilon\xi | T(\epsilon + i0) e^{i\theta J_3} | \epsilon\xi' \rangle, \quad (2.46)$$

in an obvious notation. Note that we are defining *on-shell* S-matrix elements for a given energy ϵ. With the above normalization, $\mathcal{S}$ is dimensionless.

Now setting

$$(E + i\eta - H_0)^{-1} - (E - i\eta - H_0)^{-1} = -2\pi i \delta(E - H_0)$$

in (2.39), which is allowed for $\theta \neq 0$, and doing some algebra identical to that done in I, one finds

$$F_n(\theta) = \frac{1}{4\pi i} \int_0^\infty e^{-\beta\epsilon} \operatorname{Tr} \left(A\mathcal{S}^\dagger(\epsilon) \frac{\overleftrightarrow{\partial}}{\partial\epsilon} \mathcal{S}(\epsilon) e^{i\theta J_3} \right)_c d\epsilon, \quad (2.47)$$

where

$$\operatorname{Tr} \left(A\mathcal{S}^\dagger(\epsilon) \frac{\overleftrightarrow{\partial}}{\partial\epsilon} \mathcal{S}(\epsilon) e^{i\theta J_3} \right)_c = \int d\xi\, d\xi'\, d\xi'' \left(A_{\xi\xi'} \mathcal{S}^\dagger(\xi', \epsilon, \xi'') \frac{\overleftrightarrow{\partial}}{\partial\epsilon} \mathcal{S}(\xi'', \epsilon, R_\theta \xi) \right)_c, \quad (2.48)$$

$$\mathcal{S}^\dagger(\xi, \epsilon, \xi') = [\mathcal{S}(\xi', \epsilon, \xi)]^*, \quad (2.49)$$

and

$A_{\xi\xi'}$ are the matrix elements of the exchange operator defined in I. (2.50)

The subscript "c" means the connected part as before and $\mathcal{S}^{-1}$ has been replaced by $\mathcal{S}^\dagger$, since S is unitary on-shell. Finally, collecting everything together, the virial coefficient is given by

$$b_n = \lim_{\theta \to 0} \frac{n^{\frac{3}{2}}}{4\pi i n!} \int e^{-\beta\epsilon} \operatorname{Tr} \left(A\mathcal{S}^\dagger(\epsilon) \frac{\overleftrightarrow{\partial}}{\partial\epsilon} \mathcal{S}(\epsilon) e^{i\theta J_3} \right)_c d\epsilon, \quad (2.51)$$

which is an explicitly on-shell formula. Equation (2.51) is our main result. Its use will be illustrated in the following sections.

On-shell formulas for the terms b_{Jn} in the angular momentum sum are easily obtained by inserting (2.47) into (2.22). Also the analog of (2.38)

$$b_n = \lim_{\theta \to 0} (n^{\frac{3}{2}}\beta / 2\pi n!) \int e^{-\beta\epsilon} \operatorname{Im} \operatorname{Tr} [A e^{i\theta J_3} \ln \mathcal{S}(\epsilon)]_c \, d\epsilon \quad (2.52)$$

is sometimes useful.

3. INTERIM SUMMARY AND NOTATION

Equations (2.51) and (2.52) give explicit formulas for b_n in terms of on-shell scattering amplitudes. *The rule is always to compute the trace and do the integral over ϵ before taking the limit as $\theta \to 0$.* If one proceeds in this order the calculation is well defined at every step. Basically, what is happening is that forward n-body scattering amplitudes are being defined by a particularly simple limiting procedure. The forward amplitudes thus defined do blow up, but the integrated trace remains finite and does give the correct b_n. As mentioned before, because of rotational invariance the rotation, $e^{i\theta J_3}$ can be replaced by a rotation through an angle θ about any axis.

In the following sections, we compute the contributions to b_3 and b_4 which come from pieces of $[\mathcal{S}^{-1}(\overleftrightarrow{\partial}\mathcal{S}/\partial\epsilon)]$ which blow up in the forward direction. This provides an explicit demonstration of how (2.51) and (2.52) work.

Of course, one must get the same b_n from (2.28). As mentioned above, in this equation one must be exceedingly careful about the integration contour. Consequently, (2.28) is not a *manifestly* on-shell expression. Nevertheless the final expressions one obtains from (2.28) are on-shell and do agree with (2.51). Below, we show this directly for b_3 and b_4. This is accomplished simply by doing the E integration carefully and applying some identities from scattering theory. The angle θ is never introduced here. To us, this is a most convincing demonstration of the correctness of the whole scheme.

Actually, we have found that in practice neither (2.51) nor (2.52) are particularly convenient. This is because the states $|\epsilon\xi\rangle$ tend to be complicated. The most convenient expression for b_n seems to be

$$b_n = \lim_{\theta \to 0} \frac{n^{\frac{3}{2}}\beta}{2\pi n!} \int e^{-\beta\epsilon} \operatorname{Im} \operatorname{Tr} [A \ln S_n(\epsilon + i0) e^{i\theta J_3}]_c \, d\epsilon, \quad (3.1)$$

which, as it stands, contains the full off-shell S matrix. However, with $\theta \neq 0$ we know that we can write

$$S_n(\epsilon + i0) = 1 - 2\pi i \delta(H_0 - \epsilon) T_n(\epsilon + i0), \quad (3.2)$$

so that everything ends up on-shell with an energy ϵ. We use ϵ instead of E to remind us of this fact. It is easiest to work out the traces in the $|\mathbf{p}\rangle$ representation defined in (2.10) and (2.11). To do this, one expands

$$\ln S_n(\epsilon + i0) = -2\pi i \delta(\epsilon - H_0) T_n(\epsilon + i0) + (2\pi)^2 \delta(\epsilon - H_0) T_n(\epsilon + i0) \delta(\epsilon - H_0) T_n(\epsilon + i0) + \cdots \quad (3.3)$$

to obtain

$$b_n = -\lim_{\theta\to 0} \frac{n^2\beta}{n!} \mathrm{Re}\left(\int d\mathbf{p} e^{-\beta\epsilon(\mathbf{p})} \langle \mathbf{p}| AT_n[\epsilon(\mathbf{p}) + i0] |\mathbf{p}'\rangle - 2\pi i \int d\mathbf{p}\, d\mathbf{p}'' e^{-\beta\epsilon(\mathbf{p})}\delta[\epsilon(\mathbf{p}) - \epsilon(\mathbf{p}'')] \times \langle \mathbf{p}| AT_n[\epsilon(\mathbf{p}) + i0] |\mathbf{p}''\rangle \times \langle \mathbf{p}''| T_n[\epsilon(\mathbf{p}'') + i0] |\mathbf{p}'\rangle + \cdots\right)_c, \quad (3.4)$$

where the state $|\mathbf{p}'\rangle$ is defined by

$$|\mathbf{p}'\rangle = e^{i\theta J_3} |\mathbf{p}\rangle \quad (3.5)$$

and the energy is

$$\epsilon(\mathbf{p}) = (1/2m) \sum_i \mathbf{p}_i^2. \quad (3.6)$$

Note that $\epsilon(\mathbf{p}') = \epsilon(\mathbf{p})$, so that (3.4) is on-shell.

Our calculations in the following sections will be based on (3.1) or some variant of it. The following conventions will be observed.

(i) The symbol $\mathbf{p}$ without a subscript will be reserved for the set $\mathbf{p}_1 \cdots \mathbf{p}_n$ momenta in the center of mass $\sum \mathbf{p}_i = 0$. An individual particle's momenta will always carry a subscript.

(ii) The subscript "c.m." will be suppressed [as it is in (3.1)–(3.4)] with the understanding that we are always working in the c.m. system.

(iii) The mass of all particles will be set equal to $\frac{1}{2}$ so that the energy is $\epsilon(\mathbf{p}) = \sum \mathbf{p}_i^2$. The symbol ϵ with no argument will also be used as the energy variable in a T matrix, i.e., $T(\epsilon)$. In this case it need not be the on-shell value of energy.

(iv) When doing algebra and combinatories, the rotation $e^{i\theta J_3}$ will usually be suppressed. Since, by rotational invariance, $e^{i\theta J_3}$ commutes with all operators of interest, this factor can be reinstated inside the trace at any stage in the calculation. It is best to think of the symbol "Tr" as being equipped, either explicitly or implicitly, with a rotation $e^{i\theta J_3}$. The difference between (2.28) and (2.51), for example, is then just the difference between setting $\theta = 0$ before or after doing the integration over energy. Furthermore, in explicit calculations of matrix elements, the state $e^{i\theta J_3} |\mathbf{p}\rangle$ will usually be denoted simply as $|\mathbf{p}'\rangle$ as in (3.4) and (3.5). When this is done the meaning should be clear from the context.

(v) Exchange will generally be ignored and the exchange operator suppressed.

The general matrix element

$$\langle \mathbf{p}| T(\epsilon) |\mathbf{p}'\rangle \quad (3.7)$$

will be called "on-shell" if $\epsilon(\mathbf{p}) = \epsilon = \epsilon(\mathbf{p}')$, and called "forward" if $\mathbf{p} = \mathbf{p}'$. Sometimes $|\mathbf{p}'\rangle$ will be $e^{i\theta J_3} |\mathbf{p}\rangle$ so that $\epsilon(\mathbf{p})$ is automatically equal to $\epsilon(\mathbf{p}')$. We shall often encounter T matrices involving only two of the n particles, say particles i and j. We shall denote them by

$$\langle \mathbf{p}| T_{ij}(\epsilon) |\mathbf{p}'\rangle, \quad (3.8)$$

where $\mathbf{p}_k = \mathbf{p}_k'$ for $k \neq i, j$ and $\mathbf{p}_i + \mathbf{p}_j = \mathbf{p}_i' + \mathbf{p}_j'$ by momentum conservation. Momentum delta functions are always left out of T-matrix elements. The explicit meaning of (3.8) is

$$\langle \mathbf{p}| T_{ij}(\epsilon) |\mathbf{p}'\rangle = \langle \mathbf{p}_i\mathbf{p}_j| T\left(\epsilon - \sum_{k\neq i,j} \mathbf{p}_k^2\right) |\mathbf{p}_i'\mathbf{p}_j'\rangle, \quad (3.9)$$

where T is the two-body T matrix and $\epsilon - \sum_{k\neq i,j} \mathbf{p}_k^2$ is the two-body energy. Invariance under the product of time reversal and parity requires that

$$\langle \mathbf{p}_i\mathbf{p}_j| T(\omega) |\mathbf{p}_i'\mathbf{p}_j'\rangle = \langle \mathbf{p}_i'\mathbf{p}_j'| T(\omega) |\mathbf{p}_i\mathbf{p}_j\rangle. \quad (3.10)$$

We will find this result very useful. Another useful fact is that

$$\langle \mathbf{p}_i\mathbf{p}_j| T(\omega) |\mathbf{p}_i'\mathbf{p}_j'\rangle = \langle \mathbf{p}_i + \mathbf{Q}, \mathbf{p}_j + \mathbf{Q}| T(\omega') |\mathbf{p}_i' + \mathbf{Q}, \mathbf{p}_j' + \mathbf{Q}\rangle \quad (3.11)$$

for any $\mathbf{Q}$, where $\omega' = 2Q^2 + 2\mathbf{Q}\cdot(\mathbf{p}_i + \mathbf{p}_j) + \omega$.

Also, to save writing we introduce the conventional Green's functions

$$G(z) = (z - H)^{-1}, \qquad G_0(z) = (z - H_0)^{-1}. \quad (3.12)$$

Calculations based on (2.38) which start out with $\theta = 0$ will be called "off-shell." (Even though we know that they end up giving on-shell results.) Calculations which start with $\theta \neq 0$ are manifestly on-shell and hence are called "on-shell." Often the on-shell and off-shell calculations can be done at the same time provided we agree to call

$$G_0(\epsilon + i\eta) - G_0(\epsilon - i\eta) = \text{off-shell} - 2\pi i\delta(\epsilon - H_0),$$
$$G_0(\epsilon + i\eta) + G_0(\epsilon - i\eta) = \text{off-shell } 2P/\epsilon - H_0, \quad (3.13)$$

and remember to keep η fixed and finite until the end of the calculation. In connection with doing both kinds of calculations at once, we remind the reader of convention (iv) above.

In the following section we will be breaking $(\ln S)_c$ up into

$$(\ln S)_c = (\ln S)_c|_{\text{sing}} + (\ln S)_c|_{\text{nonsing}}, \quad (3.14)$$

where $(\ln S)_c|_{\text{sing}}$ contains all pieces of $(\ln S)_c$ which are singular in the forward direction on the energy shell. This splitting of $(\ln S)_c$ into two pieces is not unique, but this will not be important. What we will do is to take a specific division of $(\ln S)_c$ and compute only the contribution of $(\ln S)_c|_{\text{sing}}$ to b_n. The remaining contribution from $(\ln S)_c|_{\text{nonsing}}$ will be a strictly on-shell object no matter how we choose to compute its trace. Therefore, when we have shown how to use the $\theta \to 0$ limit to compute the trace of $(\ln S)_c|_{\text{sing}}$ we will have shown that all of b_n can be expressed in terms of on-shell quantities. Similarly, when we have shown that the "on" and "off" shell methods are equivalent for $(\ln S)_c|_{\text{sing}}$, we will have shown that they are equivalent for all of $(\ln S)_c$.

4. SINGULAR THREE-BODY TERMS

The contribution of singular diagrams to b_3 was studied in detail using an angular momentum representation in II. Here we shall obtain the same results via the small angle limit.

First let us write the logarithm appearing in (3.1) as a series

$$\operatorname{Tr}(\ln S)_c = -\sum_{m=1}^{\infty} (1/m) \operatorname{Tr}[2\pi i\delta(\epsilon - H_0)T(\epsilon + i0)]_c^m. \tag{4.1}$$

For simplicity the factor $e^{i\theta J_3}$ will be suppressed until needed. In counting the diagrams representing the terms in (4.1), we must keep each distinct diagram only once. Taking advantage of the invariance of the trace under cyclic permutations of operators, the factor m^{-1} may be conveniently eliminated by adopting the convention that two diagrams are the same if they can be obtained from each other by a cyclic permutation of T matrices.

The singular three-body diagrams are given in Fig. 3. We use a wavy line to denote the factor $(\epsilon - H_0 + i0)^{-1}$, which is part of a single three-body T matrix, in contrast to the factors of $2\pi i\delta(\epsilon - H_0)$ which join different T matrices and are denoted by solid lines.

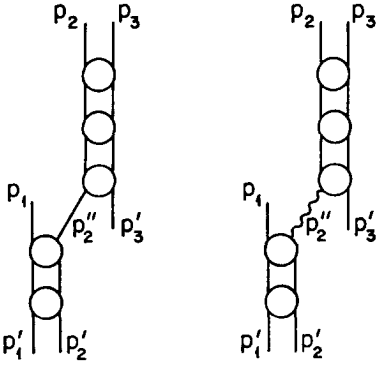

FIG. 3. Three-body singular diagrams. The circles (there can be any number of them joined by solid lines), denote two-body T matrices. The wavy line denotes a factor G_0. Notice that diagrams obtained by cyclically permuting the T matrices are identical by our convention.

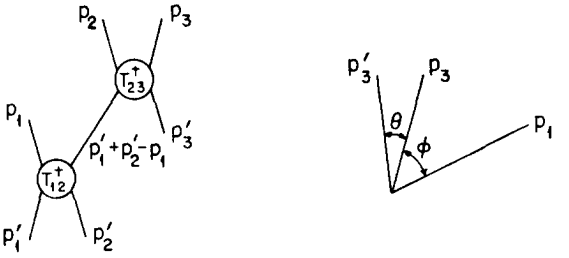

FIG. 4. Kinematical variables in evaluating the three-body singular diagrams.

These diagrams are easily summed (see II) giving[7]

$$(1/2\pi)\operatorname{Im}\operatorname{Tr}(\ln S)_c|_{\text{sing}} = -\operatorname{Re}\operatorname{Tr}\left(\delta(\epsilon - H_0)T_{12}^{\dagger}(\epsilon)\mathrm{P}\frac{1}{\epsilon - H_0}T_{23}^{\dagger}(\epsilon)\right), \tag{4.2}$$

where P denotes the principal part, T_{12} and T_{23} are two-body T matrices, and we have used the identity

$$\sum_{m=0}^{\infty}[2\pi i\delta(\epsilon - H_0)T_{ij}]^m = [1 - 2\pi i\delta(\epsilon - H_0)T_{ij}]^{-1} = 1 + 2\pi i\delta(\epsilon - H_0)T_{ij}^{\dagger}. \tag{4.3}$$

For our purpose of illustrating the small angle limit, it is not necessary to perform this sum. It will be clear as we proceed that summing is irrelevant except to simplify the writing.

The kinematic variables are depicted in Fig. 4. Substituting (4.2) into (3.1) and integrating over ϵ, we find

$$b_{3\,\text{sing}} = \lim_{\theta\to 0} -3^{\frac{3}{2}}\beta\int d\mathbf{p}e^{-\beta\epsilon(\mathbf{p})} \times \operatorname{Re}[\langle\mathbf{p}|\,T_{23}^{\dagger}(\epsilon)\,|\mathbf{p}''\rangle \times \langle\mathbf{p}''|\,T_{12}^{\dagger}(\epsilon)\,|\mathbf{p}'\rangle\mathrm{P}(p_2^2 - p_2''^2)^{-1}], \tag{4.4}$$

where $|\mathbf{p}'\rangle = e^{i\theta J_3}|\mathbf{p}\rangle$ as in (3.5), and the intermediate momenta $\mathbf{p}''$ are determined by momentum conservation, i.e.,

$$\begin{aligned}\mathbf{p}_1'' &= \mathbf{p}_1,\\ \mathbf{p}_3'' &= \mathbf{p}_3',\\ \mathbf{p}_2'' &= -\mathbf{p}_1'' - \mathbf{p}_3'' = -\mathbf{p}_1 - \mathbf{p}_3'.\end{aligned} \tag{4.5}$$

Note that whereas the integrand in (4.4) is a piece of the *on-shell* three-body amplitude, the two-body T matrices $T_{23}^{\dagger}$ and $T_{12}^{\dagger}$ are not on-shell except at the point $p_2''^2 = p_2^2$. It is convenient to evaluate (4.4) using the variables p_1^z, p_3^z and ν_1, ν_3, ϕ and φ defined by

$$\begin{aligned}p_1^x &= \nu_1\cos\varphi, \quad p_1^y = \nu_1\sin\varphi,\\ p_3^x &= \nu_3\cos(\varphi + \phi), \quad p_3^y = \nu_3\sin(\varphi + \phi).\end{aligned} \tag{4.6}$$

We do not need $\mathbf{p}_2$ because of the condition $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = 0$. Evidently ϕ is the angle between the projections of $\mathbf{p}_1$ and $\mathbf{p}_3$ onto the (x, y) plane. One easily

verifies that

$$d\mathbf{p} = (2\pi)^{-6} v_1 v_3 \, dv_1 \, dv_3 \, d\phi \, d\varphi \, dp_1^z \, dp_3^z. \quad (4.7)$$

By rotational invariance, the integrand is independent of φ, so that integrating over φ just gives a factor of 2π. Also, because the singular denominator

$$\begin{aligned} p_2^2 - p_2''^2 &= 2(\mathbf{p}_1 \cdot \mathbf{p}_3 - \mathbf{p}_1 \cdot \mathbf{p}_3') \\ &= 2v_1 v_3[\cos\phi - \cos(\phi + \theta)] \end{aligned} \quad (4.8)$$

is independent of p_1^z and p_3^z, it is convenient to do the integrations over the latter variables first. Defining

$$\begin{aligned} G(v_1, v_3, \phi, \theta) = &-(2\pi)^{-5} 3^{\frac{3}{2}} \beta \int dp_1^z \, dp_3^z e^{-\beta\epsilon(\mathbf{p})} \\ &\times \mathrm{Re}\,[\langle \mathbf{p}|\, T_{23}^\dagger \,|\mathbf{p}''\rangle \langle \mathbf{p}''|\, T_{12}^\dagger \,|\mathbf{p}'\rangle] \end{aligned} \quad (4.9)$$

one has

$$\begin{aligned} b_3 = \lim_{\theta \to 0} \int dv_1 \, dv_3 \, d\phi G(v_1, v_3, \phi, \theta) \\ \times \mathrm{P}[\cos\phi - \cos(\phi + \theta)]^{-1}. \end{aligned} \quad (4.10)$$

Note that, as (4.10) stands, the limit as $\theta \to 0$ cannot be brought inside the integral. This is because we have already set

$$\begin{aligned} (\epsilon + i\eta - H_0)^{-1} - (\epsilon - i\eta - H_0)^{-1} \\ = -2\pi i \delta(\epsilon - H_0) \end{aligned}$$

to get on shell. We can, however, transform the integrand to a form in which the limit and integral can be interchanged. First we charge variables from ϕ to $\phi' = \phi + \frac{1}{2}\theta$. Since the integrand is periodic in ϕ, the limits of ϕ' can be taken to be the same, $-\pi$ to π, as those of ϕ. Then one has

$$\begin{aligned} b_3 = \lim_{\theta \to 0} \int_0^\infty \int_0^\infty \int_{-\pi}^{\pi} dv_1 \, dv_3 \, d\phi' \\ \times \frac{G(v_1, v_3, \phi' - \frac{1}{2}\theta, \theta)}{2 \sin(\theta/2)} \mathrm{P} \frac{1}{\sin\phi'}. \end{aligned} \quad (4.11)$$

Next we note that, by rotational invariance, the rotation $e^{i\theta J_3}$ could have been replaced by $e^{-i\theta J_3}$ without in any way effecting the value of the trace. Thus we can add to (4.11) the same expression with θ replaced by $-\theta$ and divide by two to obtain

$$\begin{aligned} &b_{3\,\mathrm{sing}} \\ &= \lim_{\theta \to 0} \tfrac{1}{2} \int dv_1 \, dv_3 \, d\phi' \\ &\quad \times \frac{[G(v_1, v_3, \phi' - \theta/2, \theta) - G(v_1, v_3, \phi' + \theta/2, -\theta)]}{2 \sin\theta/2} \\ &\quad \times \mathrm{P} \frac{1}{\sin\phi'} \\ &= \tfrac{1}{2} \int dv_1 \, dv_3 \, d\phi \left(-\frac{1}{2} \frac{\partial G}{\partial \phi}(v_1, v_3, \phi, 0) \right. \\ &\quad \left. + \frac{\partial G}{\partial \theta}(v_1, v_3, \phi, 0) \right) \mathrm{P} \frac{1}{\sin\phi}, \end{aligned} \quad (4.12)$$

which is finite since G is finite as $\theta \to 0$. Since the integrand is periodic in ϕ', integration over the principle part of $(\sin\phi')^{-1}$ is well defined.

Now let us compute the derivatives of G which appear in (4.12). To this end, we display the general functional form of the T matrices by writing

$$\begin{aligned} \langle \mathbf{p}|\, T_{23}^\dagger(\epsilon) \,|\mathbf{p}''\rangle &= \langle \mathbf{p}_2 \mathbf{p}_3|\, T^\dagger(\epsilon - p_1^2) \,|\mathbf{p}'' \mathbf{p}_3'\rangle \\ &= F(\omega, \omega'', t, \epsilon'), \end{aligned} \quad (4.13)$$

where

$$\begin{aligned} \epsilon' &= \epsilon - p_1^2, \\ \omega &= (\mathbf{p}_3 - \mathbf{p}_2)^2 \\ &= (2p_3^z + p_1^z)^2 + 4v_3^2 + v_1^2 + 4v_1 v_3 \cos\phi, \\ \omega'' &= (\mathbf{p}_3' - \mathbf{p}_2'')^2 \\ &= (2p_3^z + p_1^z)^2 + 4v_3^2 + v_1^2 + 4v_1 v_3 \cos(\phi + \theta), \\ t &= (\mathbf{p}_3 - \mathbf{p}_3')^2 = 2v_3^2(1 - \cos\theta), \end{aligned} \quad (4.14)$$

and the total energy ϵ is

$$\begin{aligned} \epsilon(\mathbf{p}) &= \mathbf{p}_1^2 + \mathbf{p}_2^2 + \mathbf{p}_3^2 \\ &= 2[(p_1^z)^2 + (p_3^z)^2 + p_1^z p_3^z + v_1^2 + v_3^2 + v_1 v_3 \cos\phi]. \end{aligned} \quad (4.15)$$

According to (3.10), F is symmetric in ω and ω'', that is

$$F(\omega, \omega'', t, \epsilon) = F(\omega'', \omega, t, \epsilon). \quad (4.16)$$

It is now straightforward to compute the desired angular derivatives of $\langle \mathbf{p}|\, T_{23}^\dagger \,|\mathbf{p}''\rangle$. One finds

$$\begin{aligned} \left(-\frac{1}{2} \frac{\partial}{\partial\phi} + \frac{\partial}{\partial\theta} \right) F \Big|_{\theta=0} \\ = 2v_1 v_3 \sin\phi \left(\frac{\partial F}{\partial \omega} - \frac{\partial F}{\partial \omega''} + \frac{1}{2} \frac{\partial F}{\partial \epsilon'} \right) \Big|_{\substack{\omega=\omega'' \\ t=0}}, \end{aligned} \quad (4.17)$$

but according to (4.16)

$$\frac{\partial F}{\partial \omega} \Big|_{\omega=\omega''} = \frac{\partial F}{\partial \omega''} \Big|_{\omega=\omega''} \quad (4.18)$$

so that

$$\left(-\frac{1}{2} \frac{\partial}{\partial\phi} + \frac{\partial}{\partial\theta} \right) F \Big|_{\theta=0} = -v_1 v_3 \sin\phi \, \frac{\partial F}{\partial \epsilon} \Big|_{\substack{\omega=\omega'' \\ t=0}}, \quad (4.19)$$

and

$$\begin{aligned} \left(-\frac{1}{2} \frac{\partial}{\partial\phi} + \frac{\partial}{\partial\theta} \right) \langle \mathbf{p}|\, T_{23}^\dagger(\epsilon) \,|\mathbf{p}''\rangle \Big|_{\theta=0} \\ = -v_1 v_3 \sin\phi \, \langle \mathbf{p}|\, \frac{\partial}{\partial \epsilon} T_{23}^\dagger(\epsilon) \,|\mathbf{p}\rangle \Big|_{\epsilon=\epsilon(\mathbf{p})}. \end{aligned} \quad (4.20)$$

Obviously, the same result holds for the derivatives of $\langle \mathbf{p}''| T_{12}^{\dagger} |\mathbf{p}'\rangle$, and clearly

$$\left(\frac{1}{2}\frac{\partial}{\partial\phi} + 2\frac{\partial}{\partial\theta}\right)e^{-\beta\epsilon}\Big|_{\theta=0} = -v_1 v_3 \sin\phi \frac{\partial}{\partial\epsilon} e^{-\beta\epsilon}\Big|_{\epsilon=\epsilon(\mathbf{p})} \tag{4.21}$$

Then putting everything together, one has

$$\left(-\frac{1}{2}\frac{\partial}{\partial\phi} + \frac{\partial}{\partial\theta}\right)G\Big|_{\theta=0} = \frac{v_1 v_3 \sin\phi 3^{\frac{3}{2}}}{(2\pi)^5}\int dp_1^z\, dp_3^z \frac{\partial}{\partial\epsilon} \times [e^{-\beta\epsilon} \operatorname{Re}\langle \mathbf{p}| T_{23}^{\dagger}(\epsilon) |\mathbf{p}\rangle\langle \mathbf{p}| T_{12}^{\dagger}(\epsilon) |\mathbf{p}\rangle]|_{\epsilon=\epsilon(\mathbf{p})}, \tag{4.22}$$

and changing variables back to $\mathbf{p}$, the final result is

$$b_{3\,\mathrm{sing}} = \frac{3^{\frac{3}{2}}\beta}{2}\operatorname{Re}\int d\mathbf{p}\frac{\partial}{\partial\epsilon}[e^{-\beta\epsilon}\langle \mathbf{p}_2\mathbf{p}_3| T_{23}^{\dagger}(\epsilon) |\mathbf{p}_2\mathbf{p}_3\rangle \times \langle \mathbf{p}_1\mathbf{p}_2| T_{12}^{\dagger}(\epsilon) |\mathbf{p}_1\mathbf{p}_2\rangle]|_{\epsilon=\epsilon(\mathbf{p})}. \tag{4.23}$$

Thus we have an explicit example of how the small angle limit can be used to give well-defined expressions for b_n in terms of on-shell scattering amplitudes. Note that (4.23) does contain some off-shell derivatives of two-body amplitudes. These are, however, part of the *on-shell* three-body amplitude. This should not be surprising if one remembers that the on-shell three-body T matrix as constructed from, say, the Fadeev equations, depends on the complete off-shell two-body T matrix, not just its on-shell value. What counts in the present context is only that b_3 can be written in terms of on-shell three-body amplitudes. This is guaranteed by the fact that, for finite θ, our calculations are explicitly on-shell. The result in (4.23) is obtained by letting θ approach zero after calculating with finite θ and is therefore an on-shell quantity. The part $b_{3\,\mathrm{nonsing}}$ of b_3 which comes from the rest of the expansion of $\ln S$ is also explicitly on-shell. Thus we have obtained an expression for b_3 which contains only strictly on-shell three-body scattering amplitudes.

As stated above, we can also obtain (4.23) starting with (2.28), where $\theta = 0$ from the beginning. This calculation starts off-shell. If in (4.2) we substitute (3.13), the result is

$$(1/2\pi)\operatorname{Im}\operatorname{Tr}(\ln S)_c|_{\mathrm{sing}} = (1/4\pi)\operatorname{Im}\operatorname{Tr}\{[G_0(\epsilon + i\eta) - G_0(\epsilon - i\eta)]T_{12}^{\dagger}(\epsilon) \times [G_0(\epsilon + i\eta) + G_0(\epsilon - i\eta)T_{23}^{\dagger}(\epsilon)]\}. \tag{4.24}$$

Inserting the intermediate state to bring (4.24) into the form of (4.4) with the ϵ integration not yet done, one finds

$$b_{3\,\mathrm{sing}} = \frac{-3^{\frac{3}{2}}\beta}{4\pi}\operatorname{Im}\int d\epsilon e^{-\beta\epsilon}\int d\mathbf{p} \times \{[\epsilon - \epsilon(\mathbf{p}) + i\eta]^{-1} - [\epsilon - \epsilon(\mathbf{p}) - i\eta]^{-1}\} \times \{[\epsilon - \epsilon(\mathbf{p}) + i\eta]^{-1} + [\epsilon - \epsilon(\mathbf{p}) - i\eta]^{-1}\} \times \langle \mathbf{p}| T_{12}^{\dagger}(\epsilon) |\mathbf{p}\rangle\langle \mathbf{p}| T_{23}^{\dagger}(\epsilon) |\mathbf{p}\rangle, \tag{4.25}$$

since for $\theta = 0$, $\mathbf{p} = \mathbf{p}' = \mathbf{p}''$. Now we proceed as we did in (2.36) we obtain

$$b_{3\,\mathrm{sing}} = \frac{-3^{\frac{3}{2}}\beta}{2}\operatorname{Re}\int d\epsilon e^{-\beta\epsilon}\int d\mathbf{p}\delta'[\epsilon - \epsilon(\mathbf{p})] \times \langle \mathbf{p}| T_{12}^{\dagger}(\epsilon) |\mathbf{p}\rangle\langle \mathbf{p}| T_{23}^{\dagger}(\epsilon) |\mathbf{p}\rangle, \tag{4.26}$$

which is clearly the same as (4.23). The off-shell calculation is obviously much easier, but, having obtained (4.23) in this manner, we would then have to show that we are really dealing with an on-shell object. This would essentially amount to going through the whole $\theta \to 0$ business to show that the integrand in (4.23) was a certain limit of the on-shell three-body amplitude as one approaches the forward direction.

5. SINGULAR DIAGRAMS IN b_4

Due to algebraic complexities, it is very difficult to apply the same kind of detailed analysis on $b_{3\,\mathrm{sing}}$ to the singular diagrams of arbitrary number of particles. While already vast in number and complex in structure, such a detailed analysis can still be done for the four-body diagrams with a reasonable amount of labor. Although those with more than four particles are much more complicated, we feel that the four-body ones already contain most of the qualitative features of singular diagrams in general.

Our purpose here is not to evaluate the contribution of all four-body singular diagrams to b_4, but to demonstrate the small angle limit interpretation of the S-matrix formula for b_4 by evaluating representative singular diagrams and thereby exhibiting qualitative features which are absent in three-body diagrams.

Figure 5 shows typical four-body singular diagrams. The ones in Fig. 5(a) are made up of two nonsingular parts joined by a single energy denominator or δ-function. They are qualitatively the same as those three-body diagrams shown in Fig. 3. The ones in Figs. 5(b), and 5(c) are characterized by two intermediate δ-functions, or two energy denominators, or one each,

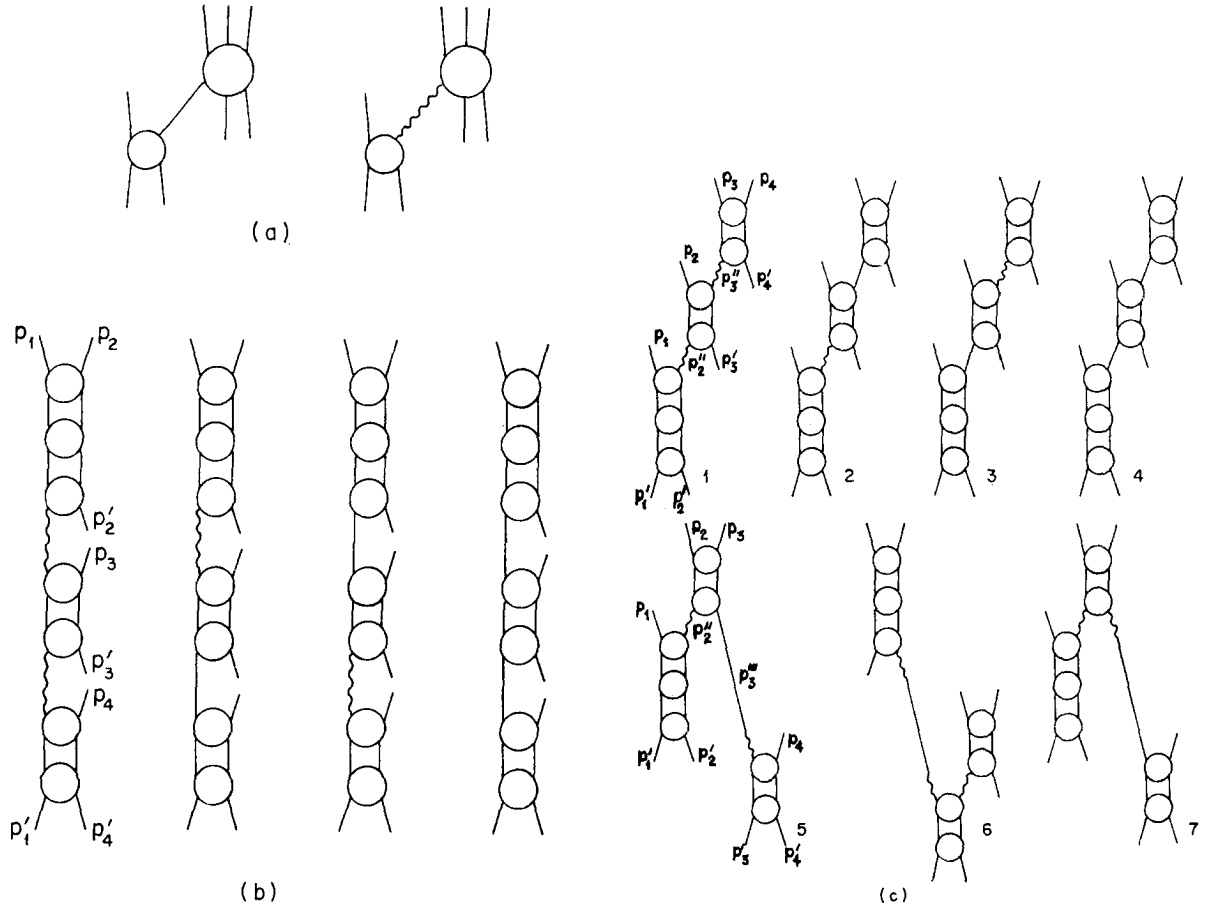

FIG. 5. Four-body singular diagrams without exchanges. (a) Diagrams with two nonsingular parts joined by one line. (b), (c) Diagrams with three nonsingular parts joined by two single lines.

joining three nonsingular parts. These are, roughly speaking, more singular than the three-body diagrams. There are also singular diagrams involving "exchanges" owing to the identity of particles. These singular diagrams will be discussed later in this section. We now proceed to study the diagrams in Fig. 5. For the clarity of discussion, we shall leave the part of algebra which is too complicated for a continuous presentation to the appendices, which may also be read independently.

We will not make explicit changes of variables in the integrals as was done in the calculation of $b_{3\,\mathrm{sing}}$. Rather, the actual changes to cylindrical coordinates will be left to the readers imagination. Also, we will generally ignore the z components of momenta. It is clear that the p^z played no real role in b_3, and we might as well have worked in an imaginary two dimensional world. We will make heavy use of the symmetry of various integrands with respect to rotations, interchange of particle labels and interchange of initial and final momenta in a T-matrix element. Equation (3.10) will be used often and should be kept in mind.

A. Diagrams a

The contribution of diagrams of the type shown in Fig. 5(a) can be summarized as

$$-\frac{1}{2\pi}\,\mathrm{Im}\,(\mathrm{Tr}\ln S)_a$$
$$=\mathrm{Re}\,\mathrm{Tr}\,\delta(\epsilon-H_0)\Lambda_2\mathrm{P}(1/\epsilon-H_0)\Lambda_3, \quad (5.1)$$

which resembles (4.2). Λ_2 and Λ_3 are nonsingular parts involving two and three particles, respectively. The small angle limit can be carried out the same way as was done in the last section, and will not be repeated here.

B. Diagram b

This subsection is a simplified version of Appendix C.

First let us sum the repeated two-body T matrices in Fig. 5(b) to obtain $T^{\dagger}$'s. Again, this is unnecessary

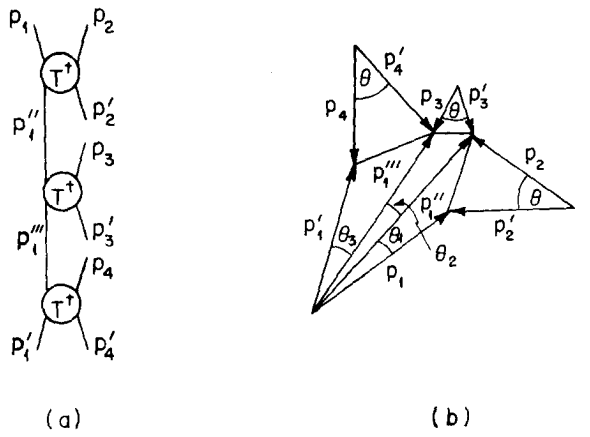

FIG. 6. Geometry of the momenta in Fig. 5(b).

except for a simpler notation. We have

$$-(1/2\pi)\,\mathrm{Im}\,(\mathrm{Tr}\ln S)_b$$
$$= \mathrm{Re}\,\mathrm{Tr}\,\delta(\epsilon - H_0)[T_{12}^\dagger G_0 T_{13}^\dagger G_0 T_{14}^\dagger + T_{12}^\dagger 2\pi i\delta(\epsilon - H_0)T_{13}^\dagger G_0 T_{14}^\dagger + \tfrac{1}{3}T_{12}^\dagger 2\pi i\delta(\epsilon - H_0)T_{13}^\dagger 2\pi i\delta(\epsilon - H_0)T_{14}^\dagger]. \quad (5.2)$$

The factor $\frac{1}{3}$ in the last factor is to compensate for the fact that the cyclic permutations of the last diagram are identical. Substituting (5.2) in (3.1), we find their contribution to b_4:

$$4^{-\frac{3}{2}}(b_4)_b = -\lim_{\theta\to 0}\beta\int d\mathbf{p}e^{-\beta\epsilon(\mathbf{p})} \times \mathrm{Re}\,\langle\mathbf{p}|\,T_{12}^\dagger\,|\mathbf{p}''\rangle\langle\mathbf{p}''|\,T_{13}^\dagger\,|\mathbf{p}'''\rangle\langle\mathbf{p}'''|\,T_{14}^\dagger\,|\mathbf{p}'\rangle \times [\mathrm{P}(p_1^2 - p_1''^2)^{-1}\mathrm{P}(p_1^2 - p_1'''^2)^{-1} - \tfrac{1}{3}\pi^2\delta(p_1^2 - p_1''^2)\delta(p_1^2 - p_1'''^2)]. \quad (5.3)$$

We shall ignore the z components of the momenta, since they are irrelevant here. Figure 6 shows the geometry of the vectors $\mathbf{p}_1$, $\mathbf{p}_1''$, and $\mathbf{p}_1'''$ and $\mathbf{p}_1'$. The T-matrix elements together with the Boltzmann factor is a function of p_1, p_1'', p_1''', θ_1, θ_2, and θ_3. Except for the total energy variable [see (C4)]

$$\epsilon(\mathbf{p}) = p_1^2 + p_2^2 + p_3^2 + p_4^2, \quad (5.4)$$

which is not symmetric in p_1^2, $p_1''^2$, and $p_1'''^2$ because of the p_1^2 term, the product of the three T-matrix elements is otherwise completely symmetric in $\mathbf{p}_1$, $\mathbf{p}_1''$, and $\mathbf{p}_1'''$. We may regard $\mathbf{p}_1$, $\mathbf{p}_1''$, $\mathbf{p}_1'''$ as the integration variables. For any function $u(p_1^2)$ which, in addition to its p_1^2 dependence explicitly indicated, is also a symmetric function of p_1^2, $p_1''^2$, and $p_1'''^2$, we have

$$u(p_1^2)[\mathrm{P}(p_1^2 - p_1''^2)^{-1}\mathrm{P}(p_1^2 - p_1'''^2)^{-1} - (\pi^2/3)\delta(p_1^2 - p_1'''^2)] = \tfrac{1}{3}[u(p_1^2)\mathrm{P}(p_1^2 - p_1''^2)^{-1}\mathrm{P}(p_1^2 - p_1'''^2)^{-1} + u(p_1''^2)\mathrm{P}(p_1''^2 - p_1^2)^{-1}\mathrm{P}(p_1''^2 - p_1'''^2)^{-1} + u(p_1'''^2)\mathrm{P}(p_1'''^2 - p_1^2)^{-1}\mathrm{P}(p_1'''^2 - p_1''^2)^{-1} - \pi^2\delta(p_1^2 - p_1''^2)\delta(p_1^2 - p_1'''^2)u(p_1^2)]. \quad (5.5)$$

Let $\overline{p_1^2} = \frac{1}{3}(p_1^2 + p_1''^2 + p_1'''^2)$ and expand u around $\overline{p_1^2}$. (5.5) reduces to

$$\frac{1}{6}\frac{\partial^2}{\partial(\overline{p_1^2})^2}u(\overline{p_1^2}) + O(p_1^2 - \overline{p_1^2}, p_1''^2 - \overline{p_1^2}, p_1'''^2 - \overline{p_1^2}). \quad (5.6)$$

It follows that as $\theta \to 0$, when p_1^2, $p_1''^2$, $p_1'''^2$, and $\overline{p_1^2}$ all become the same, we have

$$\frac{\partial}{\partial\overline{p_1^2}} \to \frac{\partial}{\partial\epsilon(\mathbf{p})}$$

and, for (5.3),

$$4^{-\frac{3}{2}}(b_4)_b = -\beta\int d\mathbf{p}\,\frac{1}{6}\frac{\partial^2}{\partial\epsilon^2}(e^{-\beta\epsilon}\langle\mathbf{p}|\,T_{14}^\dagger\,|\mathbf{p}\rangle\langle\mathbf{p}|\,T_{13}^\dagger(\epsilon)\,|\mathbf{p}\rangle \times \langle\mathbf{p}|\,T_{12}^\dagger(\epsilon)\,|\mathbf{p}\rangle)_{\epsilon=\epsilon(\mathbf{p})}. \quad (5.7)$$

This result can also be obtained from an off-shell calculation as shown in Appendix C.

The spirit of the above calculation is quite clear: One takes advantage of the symmetries and transforms the integrand into a form where the $\theta \to 0$ limit becomes obvious. In spite of the principle parts and δ functions, the expression (5.5) is effectively a smooth function of θ as shown by (5.6).

C. Diagrams c

Next we apply the small angle limit prescription to the evaluation of the diagrams in Fig. 5(c). These diagrams are different from those in Fig. 5(b) in that they do not possess symmetry under cyclic permutation and are more difficult to evaluate. This subsection is a summary of the steps leading to a clear small angle limit. The full detailed discussion is given in Appendix D, which is complete by itself.

Figure 5(c) gives, after being summed over repeated two-body T matrices,

$$-(1/2\pi)\,\mathrm{Im}\,(\mathrm{Tr}\ln S)_c$$
$$= \mathrm{Re}\,\mathrm{Tr}\,\delta(\epsilon - H_0)[T_{34}^\dagger G_0 T_{23}^\dagger G_0 T_{12}^\dagger \quad (1)$$
$$+ T_{34}^\dagger 2\pi i\delta(\epsilon - H_0)T_{23}^\dagger G_0 T_{12}^\dagger \quad (2)$$
$$+ T_{34}^\dagger G_0 T_{23}^\dagger 2\pi i\delta(\epsilon - H_0)T_{12}^\dagger \quad (3)$$
$$+ T_{34}^\dagger 2\pi i\delta(\epsilon - H_0)T_{23}^\dagger 2\pi i\delta(\epsilon - H_0)T_{12}^\dagger \quad (4)$$
$$+ T_{23}^\dagger G_0 T_{12}^\dagger G_0 T_{34}^\dagger \quad (5)$$
$$+ T_{12}^\dagger G_0 T_{34}^\dagger G_0 T_{23}^\dagger \quad (6)$$
$$+ T_{23}^\dagger G_0 T_{12}^\dagger 2\pi i\delta(\epsilon - H_0)T_{34}^\dagger], \quad (7) \quad (5.8)$$

where the labels (1)–(7) identify the terms with the diagrams in Fig. 5(c). Taking advantage of the symmetries under space inversion, time reversal, and interchanging particle labels, we can simplify (5.8)

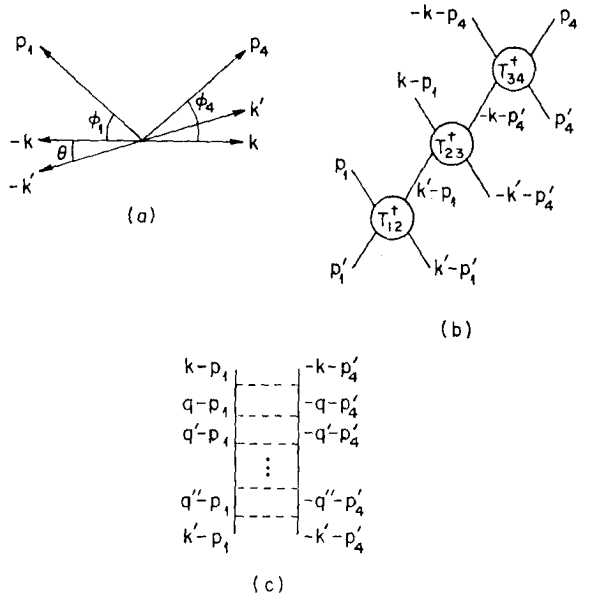

FIG. 7. (a) Definition of angles in evaluating Fig. 5(c). (b) The two-body T matrices in Fig. 5(c). (c) A term in the perturbation expansion for $T_{23}^{\dagger}$.

considerably and obtain

$$
\begin{aligned}
&-\frac{1}{2\pi}\,\mathrm{Im}\,(\mathrm{Tr}\ln S)_c \\
&= \mathrm{Re}\,\mathrm{Tr}\,\delta(\epsilon - H_0)\{T_{34}^{\dagger}(P/\epsilon - H_0)T_{23}^{\dagger}(P/\epsilon - H_0)T_{12}^{\dagger} \\
&\quad + T_{23}^{\dagger}(P/\epsilon - H_0)[T_{12}^{\dagger}(P/\epsilon - H_0)T_{34}^{\dagger} \\
&\quad + T_{34}^{\dagger}(P/\epsilon - H_0)T_{12}^{\dagger}] \\
&\quad + \pi^2 T_{23}^{\dagger}\delta(\epsilon - H_0)T_{12}^{\dagger}\delta(\epsilon - H_0)T_{34}^{\dagger}\}. \qquad (5.9)
\end{aligned}
$$

For a simple illustration of the main point of the calculation, let us regard the two-body T-matrix elements in (5.9) as constants. We substitute (5.9) in (3.1) and integrate over ϵ. A little algebra gives

$$
\begin{aligned}
(b_4)_b = &-4^{\frac{3}{2}}\beta\int d\mathbf{p}e^{-\beta\epsilon(\mathbf{p})}(T^{\dagger})^3 \\
&\times 2\mathrm{P}\frac{1}{2\mathbf{p}_1\cdot(\mathbf{k}'-\mathbf{k})}\,\mathrm{P}\frac{1}{2\mathbf{p}_4\cdot(\mathbf{k}-\mathbf{k}')}, \qquad (5.10)
\end{aligned}
$$

where the vectors $\mathbf{k}$ and $\mathbf{k}'$ are defined as

$$
\begin{aligned}
\mathbf{k} &= \mathbf{p}_1 + \mathbf{p}_2 = -\mathbf{p}_3 - \mathbf{p}_4, \\
\mathbf{k}' &= \mathbf{p}_1' + \mathbf{p}_3' = -\mathbf{p}_3' - \mathbf{p}_4'. \qquad (5.11)
\end{aligned}
$$

Equation (5.10) is analogous to (4.4). It has two denominators instead of one. It suggests that we can do the same as we did in Sec. 4 for each denominator. Let us define the angles ϕ_1 and ϕ_4 by Fig. 7(a). Then

$$
\begin{aligned}
p_2^2 &= p_1^2 + k^2 + p_1 k\cos\phi_1, \\
p_3^2 &= p_4^2 + k^2 + p_4 k\cos\phi_4, \qquad (5.12) \\
\epsilon(\mathbf{p}) &= 2(p_1^2 + p_4^2 + p_1 k\cos\phi_1 + p_4 k\cos\varphi_4).
\end{aligned}
$$

We then change the variables ϕ_1, ϕ_4 to $\phi_1 - \theta/2$, $\phi_4 + \theta/2$ so that we have

$$
\begin{aligned}
-2\mathbf{k}\cdot\mathbf{p}_1 &= 2kp_1\cos(\phi_1 - \theta/2) \equiv x^-, \\
-2\mathbf{k}'\cdot\mathbf{p}_1 &= 2kp_1\cos(\phi_1 + \theta/2) \equiv x^+, \\
2\mathbf{k}\cdot\mathbf{p}_4 &= 2kp_4\cos(\phi_4 + \theta/2) \equiv y^+, \qquad (5.13) \\
2\mathbf{k}'\cdot\mathbf{p}_4 &= 2kp_4\cos(\phi_4 - \theta/2) \equiv y^-.
\end{aligned}
$$

The total energy is, by (5.12),

$$
\epsilon(\mathbf{p}) = 2(p_1^2 + p_4^2 + k^2) + x^- + y^+. \qquad (5.14)
$$

Thus, the integrand of (5.10) is proportional to

$$
\begin{aligned}
&e^{-\beta x^-}\mathrm{P}(x^- - x^+)^{-1}e^{-\beta y^+}\mathrm{P}(y^+ - y^-)^{-1} \\
&\quad = \tfrac{1}{4}\mathrm{P}[e^{-\beta x^-} - e^{-\beta x^+}/(x^- - x^+)] \\
&\qquad \times \mathrm{P}[(e^{-\beta y^-} - e^{-\beta y^-})/(y^+ - y^-)], \qquad (5.15)
\end{aligned}
$$

where we have symmetrized this integrand by taking advantage of the fact that when $\phi_{1,4} \to -\phi_{1,4}$, $x^{\pm}, y^{\pm} \to x^{\mp}, y^{\mp}$. As $\theta \to 0$, we have $x^+, y^+ \to x^-, y^-$ and (5.15) becomes the second derivative of the Boltzmann factor with respect to $\epsilon(\mathbf{p})$. Equation (5.10) becomes

$$
(b_4)_b = -4^{\frac{3}{2}}\beta\int d\mathbf{p}\,\frac{1}{2}\frac{\partial^2}{\partial\epsilon(\mathbf{p})^2}e^{-\beta\epsilon(\mathbf{p})}(T^{\dagger})^3. \qquad (5.16)
$$

When the p dependence of the T-matrix elements are taken into account, the algebra becomes more complicated but the procedure is about the same. The calculation in Appendix D shows that

$$
\begin{aligned}
(b_4)_b = &-4^{\frac{3}{2}}\beta\int d\mathbf{p}\,\frac{1}{2}\left(\frac{\partial^2}{\partial\epsilon^2}e^{-\beta\epsilon}\langle\mathbf{p}|\,T_{34}^{\dagger}(\epsilon)\,|\mathbf{p}\rangle\right. \\
&\left.\times\langle\mathbf{p}|\,T_{23}^{\dagger}(\epsilon)\,|\mathbf{p}\rangle\langle\mathbf{p}|\,T_{12}^{\dagger}(\epsilon)\,|\mathbf{p}\rangle\right)_{\epsilon=\epsilon(\mathbf{p})}. \qquad (5.17)
\end{aligned}
$$

It is easily shown that (5.17) can be obtained by using off-shell forward amplitudes. [See (D34) and (D35).]

D. Singular Exchange Diagrams

For the sake of simplicity we have ignored the identity of particles. As was shown in detail in Paper I, it can be accounted for by introducing an antisymmetrization operator A for Fermions (a symmetrization operator for Bosons) into the trace. It is sufficient to discuss the Fermion case here. We have

$$
\begin{aligned}
&\mathrm{Tr}\,(\ln S)_c \to \mathrm{Tr}\,(A\ln S)_c, \\
&A = \sum_P (-)^P P, \qquad (5.18)
\end{aligned}
$$

where P permutes the particle labels, $(-)^P = \pm 1$ depending on whether P is even or odd, and the connected diagrams defined by the subscript c now include those joined by exchanges.

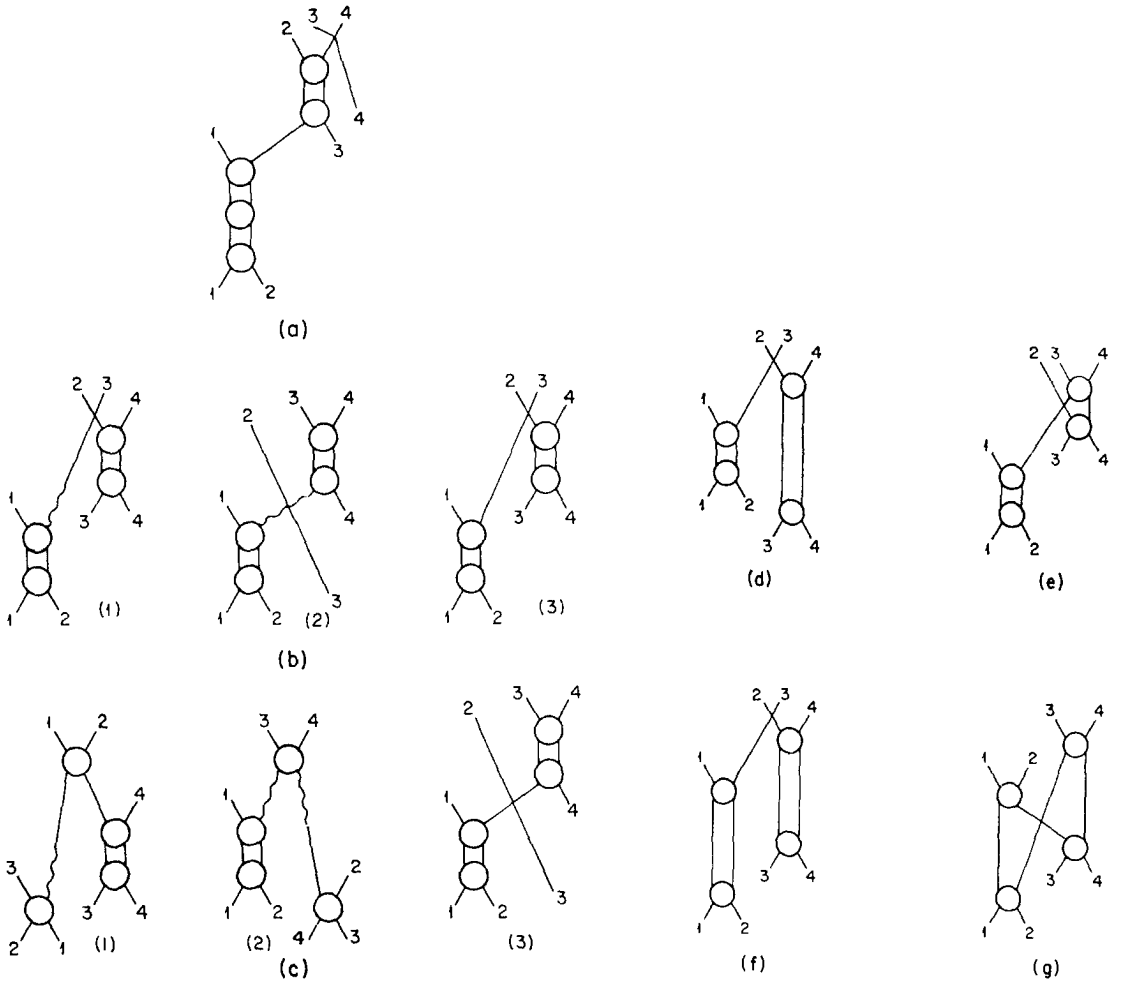

FIG. 8. Four-body singular diagrams with exchange. (a) Exchange unrelated to the singular behavior. (b) Singular behavior closely related to exchange. (c) These diagrams are already included in (b) since they are cyclic permutations of those in (b). In (3), we have inverted the order of the labels. (d) and (e) are equivalent (by cyclic permutation). (f) This is not a singular diagram. It is equivalent to (g).

As can be easily shown, for three-body diagrams, those involving exchange are nonsingular. This is not the case for four-body diagrams. Figure 8 shows the general singular four-body exchange diagrams that are qualitatively different from the diagrams we discussed so far. Those with exchanges within two-body T matrices will not be discussed since they correspond to modifications of two-body T matrices only.

Notice that we still keep our diagram convention that each diagram includes all cyclic permutations of its T matrices. For example, the diagrams in Fig. 8(c) are already included in 8(b) and must not be counted again.

In the diagrams shown in Fig. 8(a), the exchange involves particles 3 and 4, which play no role in the singular energy denominator. These diagrams can thus be included in the category of Fig. 5(a), which has been discussed already. The diagrams in Fig. 8(b) will be studied in detail shortly. They are the only four-body singular exchange diagrams which cannot be included in Fig. 5(a). The diagram in Fig. 8(d) may look like the third diagram in Fig. 8(b), but because of the order of the T matrices it is quite different, as one can write down in detail and see. In fact, it is already included in Fig. 8(e), which again falls in the category of Fig. 5(a). The one in Fig. 8(f) is not singular despite its appearance. It is equivalent to the one in 8(g).

The diagrams in Fig. 8(b) are either disconnected pieces joined by an exchange, or with the singular denominator (or δ function) involving the exchanged particle. Let us write down their contribution to $-(1/2\pi)\,\mathrm{Im}\,(\mathrm{Tr}\,A \ln S)_c$;

$$\begin{aligned}
&\mathrm{Re\ Tr}\,(-P_{23})\delta(\epsilon - H_0)T^{\dagger}_{34}G_0T^{\dagger}_{12}, \quad \text{from 1,}\\
&\mathrm{Re\ Tr}\,(-P_{23})\delta(\epsilon - H_0)T^{\dagger}_{24}G_0T^{\dagger}_{12}, \quad \text{from 2,}\\
&\mathrm{Re\ Tr}\,(-P_{23})\delta(\epsilon - H_0)[\tfrac{1}{2}T^{\dagger}_{34}2\pi i\delta(\epsilon - H_0)T^{\dagger}_{12}\\
&\qquad + \tfrac{1}{2}T^{\dagger}_{24}2\pi i\delta(\epsilon - H_0)T^{\dagger}_{12}], \quad \text{from 3,}
\end{aligned} \tag{5.19}$$

where we have summed the repeated two-body T matrices to obtain the $T^\dagger$'s, P_{23} exchanges 2 and 3, and we have written the contribution of Diagram 3 as the average of (b3) and (c3). Combining the terms in (5.19), we obtain

$$(\tfrac{1}{2}m)\,\mathrm{Im}\,(\mathrm{Tr}\,A \ln S)_{1+2+3} = A_1 + A_2,$$

$$A_1 = \mathrm{Re}\,\mathrm{Tr}\,P_{23}\delta(\epsilon - H_0)T_{34}^\dagger\,\mathrm{P}\frac{1}{\epsilon - H_0}\,T_{12}^\dagger, \quad (5.20)$$

$$A_2 = \mathrm{Re}\,\mathrm{Tr}\,P_{23}\delta(\epsilon - H_0)T_{24}^\dagger\,\mathrm{P}\frac{1}{\epsilon - H_0}\,T_{12}^\dagger.$$

Multiplying (5.20) by the Boltzmann factor and integrating over ϵ we have

$$b_{4x} = b_{41} + b_{42},$$

$$\begin{aligned} b_{41} &= -4^{\frac{3}{2}}\beta\int d\mathbf{p}e^{-\beta\epsilon}A_1\,d\epsilon \\ &= \lim_{\theta\to 0} 4^{\frac{3}{2}}\beta\,\mathrm{Re}\int d\mathbf{p} \\ &\quad\times e^{-\beta\epsilon(\mathbf{p})}\langle\mathbf{p}_2\mathbf{p}_4|\,T^\dagger\,|\mathbf{p}_3'\mathbf{p}_4'\rangle\langle\mathbf{p}_1\mathbf{p}_3|\,T^\dagger\,|\mathbf{p}_1'\mathbf{p}_2'\rangle \\ &\quad\times \mathrm{P}(p_2^2 - p_3^2)^{-1}(2\pi)^3\delta(\mathbf{p}_1 + \mathbf{p}_3 - \mathbf{p}_1' - \mathbf{p}_2'), \end{aligned} \quad (5.21)$$

$$\begin{aligned} b_{42} &= 4^{\frac{3}{2}}\beta\int d\mathbf{p}e^{-\beta\epsilon}A_2\,d\epsilon \\ &= \lim_{\theta\to 0} 4^{\frac{3}{2}}\beta\,\mathrm{Re}\int d\mathbf{p} \\ &\quad\times e^{-\beta\epsilon(\mathbf{p})}\langle\mathbf{p}_3\mathbf{p}_4|\,T^\dagger\,|\mathbf{p}_2''\mathbf{p}_4'\rangle\langle\mathbf{p}_1\mathbf{p}_2''|\,T^\dagger\,|\mathbf{p}_1'\mathbf{p}_2'\rangle \\ &\quad\times \mathrm{P}(p_2^2 - p_2''^2)^{-1}(2\pi)^3\delta(\mathbf{p}_2 - \mathbf{p}_3'). \end{aligned} \quad (5.22)$$

The quantities b_{41} and b_{42} are contributions from Fig. 8(b1) and 8(b2), respectively, with the intermediate G_0 replaced by the corresponding principle parts. The δ-function parts of the G_0 are removed by the diagram in Fig. 8(b).

Consider (5.21) first. The off-shell energy variables of the two-body T-matrix elements are evaluated at

$$\begin{aligned} \epsilon(\mathbf{p}) - p_1^2 - p_3^2 &= p_2^2 + p_4^2, \quad \text{for } T_{34}^\dagger, \\ \epsilon(\mathbf{p}) - p_3^2 - p_4^2 &= p_1^2 + p_2^2, \quad \text{for } T_{12}^\dagger. \end{aligned} \quad (5.23)$$

Equation (5.21) is unchanged if we interchange 1 and 4, 2 and 3. This corresponds to putting $T_{12}^\dagger$ higher than $T_{34}^\dagger$ in Fig. 8(b1). Under this interchange, the principle part in (5.21) changes sign. Therefore, we can write the T-matrix elements and the principle part as

$$\begin{aligned} &\tfrac{1}{2}[\langle\mathbf{p}_2\mathbf{p}_4|\,T^\dagger(p_2^2 + p_4^2)\,|\mathbf{p}_3'\mathbf{p}_4'\rangle\langle\mathbf{p}_1\mathbf{p}_3|\,T^\dagger(p_1^2 + p_2^2)\,|\mathbf{p}_1'\mathbf{p}_2'\rangle \\ &\quad - \langle\mathbf{p}_1\mathbf{p}_3|\,T^\dagger(\mathbf{p}_3^2 + p_1^2)\,|\mathbf{p}_1'\mathbf{p}_2'\rangle \\ &\quad\times\langle\mathbf{p}_2\mathbf{p}_4|\,T^\dagger(p_3^2 + p_4^2)\,|\mathbf{p}_3'\mathbf{p}_4'\rangle]\mathrm{P}(p_2^2 - p_3^2)^{-1}. \end{aligned} \quad (5.24)$$

In the limit as $\theta \to 0$, $p_2 \to p_3$, and (5.24) simply gives the derivative of the T-matrix elements with respect to their off-shell energy variables, i.e., (5.24) becomes

$$\frac{1}{2}\left(\frac{\partial}{\partial\epsilon}\langle\mathbf{p}|\,T_{34}^\dagger(\epsilon)\,|\mathbf{p}\rangle\langle\mathbf{p}|\,T_{12}^\dagger(\epsilon)\,|\mathbf{p}\rangle\right)_{\epsilon=\epsilon(\mathbf{p})}. \quad (5.25)$$

It follows that (5.21) gives

$$\begin{aligned} b_{41} &= 4^{\frac{3}{2}}\beta\,\mathrm{Re}\int d\mathbf{p}(2\pi)^3\delta(\mathbf{p}_2 - \mathbf{p}_3)e^{-\beta\epsilon(\mathbf{p})} \\ &\quad\times\frac{1}{2}\left(\frac{\partial}{\partial\epsilon}\langle\mathbf{p}|\,T_{34}^\dagger(\epsilon)\,|\mathbf{p}\rangle\langle\mathbf{p}|\,T_{12}^\dagger(\epsilon)\,|\mathbf{p}\rangle\right)_{\epsilon=\epsilon(\mathbf{p})}. \end{aligned} \quad (5.26)$$

We now turn to (5.22). We shall use the variables $\mathbf{p}_1$, $\mathbf{p}_4$, $\mathbf{k} = \mathbf{p}_1 + \mathbf{p}_2$ and the angles ϕ_1, ϕ_4 used in our discussion of Fig. 5(c). [See Fig. 7(a).]

For the denominator in (5.22), we have

$$\begin{aligned} p_2^2 - p_2''^2 &= (\mathbf{k} - \mathbf{p}_1)^2 - (\mathbf{k}' - \mathbf{p}_1)^2 \\ &= 2(\mathbf{k} - \mathbf{k}')\cdot\mathbf{p}_1 \\ &= x^- - x^+, \end{aligned} \quad (5.27)$$

where $x^\pm$ are defined in (5.13). The angular dependence of the T-matrix elements have been analyzed. [See (D20) and (D21).] We have

$$\begin{aligned} \langle\mathbf{p}_3\mathbf{p}_4|\,T^\dagger(p_3^2 + p_4^2)\,|\mathbf{p}_2''\mathbf{p}_4'\rangle &= f_{34}(p_3^2 + p_4^2;\, y^+, y^-), \\ \langle\mathbf{p}_1\mathbf{p}_2''|\,T^\dagger(p_1^2 + p_2^2)\,|\mathbf{p}_1'\mathbf{p}_2'\rangle &= f_{12}(p_1^2 + p_2^2;\, x^-, x^+). \end{aligned} \quad (5.28)$$

In view of the momentum δ function in (5.22), we have $p_3^2 = p_2^2$, so that

$$\begin{aligned} \epsilon(\mathbf{p}) &= p_1^2 + p_4^2 + 2p_2^2 \\ &= p_1^2 + p_4^2 + 2(p_1^2 + k^2) + 2x^-, \\ p_2^2 &= p_3^2 = p_1^2 + k^2 + x^-. \end{aligned} \quad (5.29)$$

The condition given by this δ function can also be written as

$$\mathbf{k} + \mathbf{k}' = \mathbf{p}_1 - \mathbf{p}_4', \quad (5.30)$$

since $\mathbf{p}_2 = \mathbf{k} - \mathbf{p}_1$, $\mathbf{p}_3' = \mathbf{k}' - \mathbf{p}_4'$. It follows that

$$\begin{aligned} p_4^2 &= (\mathbf{k} + \mathbf{k}' - \mathbf{p}_1)^2 \\ &= (\mathbf{k} + \mathbf{k}')^2 + p_1^2 - (x^- + x^+). \end{aligned} \quad (5.31)$$

The condition (5.30) is shown graphically in Fig. 9. The δ function in (5.22) is thus

$$\begin{aligned} &\delta(p_1\sin\phi_1 - p_4\sin\phi_4) \\ &\quad\times\delta(p_1\cos\phi_1 + p_4\cos\phi_4 + |\mathbf{k} + \mathbf{k}'|). \end{aligned} \quad (5.32)$$

If we change the signs of ϕ_1 and ϕ_4 simultaneously, (5.32), (5.31) are unchanged. The functions f_{34} and f_{12}

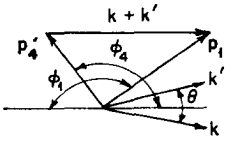

FIG. 9. Geometry of the condition $\mathbf{k} + \mathbf{k}' = \mathbf{p}_1 - \mathbf{p}_4'$.

are also unchanged if we hold their first arguments fixed. The only thing that is not invariant under this change of signs is the p_2^2 occurring in the energy variables in (5.28) and (5.29). Again, the effect of the principle part in (5.22), which has the form $P/(x^- - x^+)$ by (5.27), is to differentiate with respect to the energy variable in the $\theta \to 0$ limit. Now we must notice that there is a factor 2 in front of the x^- in $\epsilon(\mathbf{p})$ in (5.29). Thus, there appears a factor of 2 in differentiating the energy variable of the Boltzmann factor. Therefore, for the integrand of (5.22), we have

$$\delta(\mathbf{p}_2 - \mathbf{p}_3)\Big(\frac{1}{2}\frac{\partial}{\partial_\epsilon} e^{-\beta\epsilon} \langle \mathbf{p}|\, T_{34}^\dagger(\epsilon)\, |\mathbf{p}\rangle \langle \mathbf{p}|\, T_{12}^\dagger(\epsilon)\, |\mathbf{p}\rangle + \langle \mathbf{p}|\, T_{34}^\dagger\, |\mathbf{p}\rangle \langle \mathbf{p}|\, T_{12}^\dagger(\epsilon)\, |\mathbf{p}\rangle \frac{1}{2}\frac{\partial}{\partial\epsilon} e^{-\beta\epsilon}\Big)_{\epsilon=\epsilon(\mathbf{p})}, \quad (5.33)$$

where the last term takes care of the factor 2 we just mentioned. Combining (5.26) and (5.33), we have

$$\begin{aligned} b_{4x} &= b_{41} + b_{42} \\ &= 4^{\frac{3}{2}}\beta \,\mathrm{Re}\int d\mathbf{p}\,\delta(\mathbf{p}_2 - \mathbf{p}_3) \\ &\quad \times \Big(\frac{\partial}{\partial\epsilon} e^{-\beta\epsilon} \langle \mathbf{p}|\, T_{34}^\dagger(\epsilon)\, |\mathbf{p}\rangle \langle \mathbf{p}|\, T_{12}^\dagger(\epsilon)\, |\mathbf{p}\rangle\Big)_{\epsilon=\epsilon(\mathbf{p})}. \end{aligned} \quad (5.34)$$

Notice that the Boltzmann factor in (5.26) is not differentiated, while the one in (5.33) gives rise to two terms. The sum becomes a total derivative as (5.34) shows.

Finally, let us evaluate b_{4x} using off-shell amplitudes in the momentum representation. Consider A_1 in (5.20). Let us replace the δ function and the principle part by (2.36). Then we have

$$\begin{aligned} A_1' = \mathrm{Re}\int d\mathbf{p}\{-\tfrac{1}{2}\delta'[\epsilon - \epsilon(\mathbf{p})]\} \langle \mathbf{p}|\, T_{34}^\dagger(\epsilon)\, |\mathbf{p}\rangle \\ \times \langle \mathbf{p}|\, T_{12}^\dagger(\epsilon)\, |\mathbf{p}\rangle (2\pi)^3\delta(\mathbf{p}_2 - \mathbf{p}_3)e^{-\beta\epsilon}. \end{aligned} \quad (5.35)$$

We put a prime on A_1 because (5.35) is *not* the same as the A_1 in (5.20) where the trace is interpreted as the small angle limit (3.1). From (5.35), we obtain

$$\begin{aligned} b_{41}' &= 4^{\frac{3}{2}}\beta\int d\epsilon e^{-\beta\epsilon} A_1' \\ &= 4^{\frac{3}{2}}\beta \,\mathrm{Re}\int d\mathbf{p}\,\frac{1}{2}\Big(\frac{\partial}{\partial\epsilon} e^{-\beta\epsilon} \langle \mathbf{p}|\, T_{34}^\dagger(\epsilon)\, |\mathbf{p}\rangle \\ &\quad \times \langle \mathbf{p}|\, T_{12}^\dagger(\epsilon)\, |\mathbf{p}\rangle\Big)_{\epsilon=\epsilon(\mathbf{p})} (2\pi)^3\delta(\mathbf{p}_2 - \mathbf{p}_3). \end{aligned} \quad (5.36)$$

The same procedure shows that A_2' leads to the same result. The sum $b_{41}' + b_{42}' = 2b_{41}'$ is clearly the same as (5.34). We have thus observed the interesting fact that, in evaluating individual singular diagrams contributing to b_4, the on-shell $\theta \to 0$ method and the off-shell method do not always agree. Of course, the sum of all singular diagrams always comes out the same, as was shown above. Note that the failure of the two methods to agree when only a single diagram is taken into account is not an indication that there is something wrong with our general proof. We can obtain (3.1) from the rigorously proven (2.13) only when S is the complete S matrix not just some piece of it.

To summarize, what we have accomplished in this section is an analysis of the singularities in four-body amplitudes which are relevant for b_4. We have shown explicitly for representative diagrams that the small angle limit for defining the trace is well defined and gives the same results as do calculations using off-shell matrix elements. We feel that these results demonstrate sufficiently the validity of our general conclusions.

ACKNOWLEDGMENT

The helpful comments of Professor C. N. Yang are gratefully appreciated.

APPENDIX A

We wish to prove that (2.13) holds. Since

$$\langle\{\mathbf{x}\}|\, U_n\, |\{\mathbf{y}\}\rangle_{\mathrm{c.m.}}$$

is a continuous function of its argument, this evidently boils down to proving that $|\langle\{\mathbf{x}\}|\, U_n\, |\{R_\theta\mathbf{x}\}\rangle|$ is uniformly bounded (for sufficiently small θ) by an integrable function of $\{\mathbf{x}\}$.

First we consider Maxwell–Boltzmann statistics, where the U function is given by Feynman's path integral formula

$$\langle\{\mathbf{x}\}|\, U_n\, |\{\mathbf{y}\}\rangle_{\mathrm{c.m.}} = \int_{\{\mathbf{y}\}}^{\{\mathbf{x}\}} \exp\Big(-\frac{m}{2}\int_0^\beta \{\dot{\mathbf{z}}(u)\}^2\, du\Big) \times \mathfrak{V}_n(\{\mathbf{z}\})\mathfrak{D}\{\mathbf{z}(u)\} \quad (A1)$$

(Maxwell–Boltzmann statistics),

where, in Feynman's notation, the integration $\mathfrak{D}\{\mathbf{z}(u)\}$ is over all paths $\{\mathbf{z}(u)\}$ such that $\{\mathbf{z}(0)\} = \{\mathbf{y}\}$ and $\{\mathbf{z}(\beta)\} = \{\mathbf{x}\}$. $\mathfrak{V}_n(\{\mathbf{z}\})$ is closely related to the classical cluster function; it is

$$\begin{aligned} \mathfrak{V}_2 &= \exp\Big(-\int_0^\beta V_{12}(\{\mathbf{z}(u)\})\, du\Big) - 1 \\ \mathfrak{V}_3 &= \exp\Big(-\int_0^\beta V_{123}(\{\mathbf{z}(u)\})\, du\Big) \\ &\quad - \exp\Big(-\int_0^\beta V_{12}(\{\mathbf{z}(u)\})\, du\Big) \\ &\quad - \exp\Big(-\int_0^\beta V_{23}(\{\mathbf{z}(u)\})\, du\Big) \\ &\quad - \exp\Big(-\int_0^\beta V_{13}(\{\mathbf{z}(u)\})\, du\Big) + 2, \end{aligned} \quad (A2)$$

etc., where, for example, V_{12} is the potential energy of the two particles 1 and 2 and V_{123} is the total potential energy of the three indicated particles. The general scheme for constructing $\mathfrak{V}_n$ should be clear. Another, shorter notation for (A1) is

$$\langle\{\mathbf{x}\}|\, U_n\, |\{\mathbf{y}\}\rangle_{\text{c.m.}} = \int dP_{xy}\mathfrak{V}_n(\{z\}) \qquad \text{(A3)}$$

(Maxwell–Boltzmann statistics),

where dP_{xy} is the Weiner measure.[8]

If $-B_n$ is a lower bound on the total n-particle potential energy and if all potentials go to zero for large separations between particles, then it is easy to show that for any path[8]

$$\mathfrak{V}_n(\{z\}) \leq C_n e^{\beta B_n}, \qquad \text{(A4)}$$

where C_n is the number of terms in the expression for U_n as a sum of products of W functions, i.e., the number of terms on the right in (A2). Then, since the Weiner measure is positive,[8]

$$\int dP_{xy}\mathfrak{V}_n(\{z\}) \leq C_n e^{\beta B_n}\int dP_{xy}, \qquad \text{(A5)}$$

and it is useful to remember that

$$\int dP_{xy} = n^{-\frac{3}{2}}\lambda^{3-3n}\exp\left[-(m/2\beta)(\{\mathbf{x}\} - \{\mathbf{y}\})^2\right], \qquad \text{(A6)}$$

where $\{\mathbf{x}\}^2 = \sum \mathbf{x}_i^2$, etc.

In the above formulas and in what follows, it is important to keep in mind that we are working in the n-particle c.m. Thus, $\sum \mathbf{x}_i = 0$, $\sum \mathbf{y}_i = 0$ *and the paths* $\{\mathbf{z}(u)\}$ *satisfy* $\sum_i \mathbf{z}_i(u) = 0$ *for all* u.

Now let ρ_x be the radius in the c.m., i.e., $\rho_x = (\sum (\mathbf{x}_i)^2)^{\frac{1}{2}}$. We are interested in large ρ_x and, for reasons to become apparent later, we restrict $\{\mathbf{y}\}$ to lie in the region $\sum (\mathbf{x}_i - \mathbf{y}_i)^2 < 16(n^4)^{-1}\rho_x^2$. For $\{\mathbf{y}\} = \{R_\theta \mathbf{x}\}$ this can be accomplished by restricting θ to small angles which satisfy

$$1 - \cos\theta < (32n^4)^{-1}. \qquad \text{(A7)}$$

Having restricted θ in this way, let us divide the paths in (A3) into two classes. Class I is defined as all those paths for which $\sum_i [\mathbf{x}_i - \mathbf{z}_i(u)]^2 < (16n^4)^{-1}\rho_x^2$, and Class II contains all other paths. Then, in an obvious notation,

$$\int dP_{xy}\mathfrak{V}_n(\{z\}) = \int_{\text{I}} dP_{xy}\mathfrak{V}_n(\{z\}) + \int_{\text{II}} dP_{xy}\mathfrak{V}_n(\{z\}) \qquad \text{(A8)}$$

and

$$\int dP_{xy}\mathfrak{V}_n(\{z\}) \leq \int_{\text{I}} dP_{xy}\mathfrak{V}_n(\{z\}) + C_n e^{\beta B n}\int_{\text{II}} dP_{xy}, \qquad \text{(A9)}$$

where (A5) has been used. The two integrals on the right-hand side of (A9) will be treated separately.

For fixed $\{\mathbf{x}\}$, consider the function $\varphi(\{\mathbf{y}\}, \beta)$ defined by

$$\varphi(\{\mathbf{y}\}, \beta) = \int_{\text{II}} dP_{xy} = \int dP_{xy} - \int_{\text{I}} dP_{xy}, \qquad \text{(A10)}$$

where the β dependence on the right is implicitly contained in dP_{xy}. It satisfies the heat conduction equation

$$(1/2m)\nabla^2\varphi = (\partial/\partial\beta)\varphi, \qquad \text{(A11)}$$

where ∇^2 is the Laplacian in the $3n - 3$ dimensional space of the $\{\mathbf{y}\}$, and the boundary condition

$$\varphi(\{\mathbf{y}\}, 0) = 0. \qquad \text{(A12)}$$

To determine φ completely we need one more boundary condition. To this end, we note that $\int_{\text{I}} dP_{xy}$ is the Green's function for an imaginary time Schrödinger equation with an infinite potential barrier on the surface $\sum_i (\mathbf{x}_i - \mathbf{y}_i)^2 = (16n^4)^{-1}\rho_x^2$. Applying such a barrier just keeps all paths in a sphere of radius $(4n^2)^{-1}\rho_x$ centered at $\{\mathbf{x}\}$, i.e., it just picks out paths of class I. Evidently $\int_{\text{I}} dP_{xy}$ vanishes when $\{\mathbf{y}\}$ approaches the potential barrier and, combining (A6) and (A10), one has

$$\varphi(\{\mathbf{y}\}, \beta)\big|_{\sum(\mathbf{x}_i-\mathbf{y}_i)^2=(16n^4)^{-1}\rho_x^2} = n^{-\frac{3}{2}}\lambda^{3-3n}\exp\left(-\frac{m}{32\beta n^4}\rho_x^2\right). \qquad \text{(A13)}$$

The problem of finding φ has thus been reduced to that of solving a simple spherically symmetric heat conduction problem in $3N - 3$ dimensions. Actually, we do not even have to solve this problem. Think of φ as a temperature distribution which vanishes everywhere inside the sphere $\sum (\mathbf{y}_i - \mathbf{x}_i)^2 = (16n^4)^{-1}\rho_x^2$ at $\beta = 0$ and which has the temperature prescribed by (A13) on the sphere. Then it is obvious that for a given β, φ must then be less than the maximum of the boundary temperature for smaller values of β. For $\rho_x^2 > (16\beta n^4)[m(3n - 3)]^{-1}$, the boundary temperature is a monotonically increasing function of β so that

$$\int_{\text{II}} dP_{xy} < n^{-\frac{3}{2}}\lambda^{3-3n}\exp\left(-\frac{m}{32\beta n^4}\rho_x^2\right)\left(\frac{m}{32\beta n^4}\rho_x > \tfrac{3}{2}(n-1)\right). \qquad \text{(A14)}$$

Now let us tackle the other term in (A9). For paths of class I $\sum_i [\mathbf{x}_i - \mathbf{z}_i(u)]^2 < (16n^4)^{-1}\rho_x^2$, and in particular

$$|\mathbf{x}_i - \mathbf{z}_i(u)| < (4n^2)^{-1}\rho_x \qquad \text{(A15)}$$

for all $i = 1 \cdots n$ and $0 \leq u \leq \beta$. Think of $\mathbf{x}_i$, $i = 1, 2, \cdots n$, as a set of n points in ordinary three-dimensional space, and let S be the set of points contained in n spheres of radius $(2n^2)^{-1}\rho_x$ centered at the points $\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_n$. Evidently, the various three-dimensional trajectories described by $\mathbf{z}_1(u), \mathbf{z}_2(u), \cdots, \mathbf{z}_n(u)$ all lie in the set S. We now want to show that S is disconnected, that is that S breaks up into at least two pieces which have no point in common. The proof is by contradiction. If S were connected, the maximum distance between any two of the centers $\mathbf{x}_1 \cdots \mathbf{x}_n$ would be $(n-1)\rho_x/n^2$. But this is impossible since

$$\sum_j (\mathbf{x}_1 - \mathbf{x}_j)^2 = nx_1^2 + \rho_x^2 \tag{A16}$$

for $\sum_j \mathbf{x}_j = 0$, which implies that $(\mathbf{x}_j - \mathbf{x}_j)^2 \geq n^{-1}\rho_x^2$ for at least one j. Therefore S breaks up into two disconnected pieces S_1 and S_2. Next, since each trajectory $\mathbf{z}_i(u)$ remains in a sphere of radius $(4n^2)^{-1}\rho_x$ about its starting point $\mathbf{x}_i$, and the spheres which define S have twice this radius, it follows that

$$|\mathbf{z}_i(u) - \mathbf{z}_j(u)| > (2n^2)^{-1}\rho_x \tag{A17}$$

for $\mathbf{z}_i$ in S_1 and $\mathbf{z}_j$ in S_2. What we have shown, then, is that all paths in class I have the property that the particle trajectories split up into two or more clusters with no particle in one cluster getting closer than $(2n^2)^{-1}\rho_x$ to any particle in cluster two. For large ρ_x this implies that $\mathfrak{V}_n(\{\mathbf{z}\})$ is very small. To be more specific, let V_{int} be that part of the potential energy which connects particles in cluster one with those in cluster two. Furthermore, let $\bar{V}(\rho_x)$ be the maximum of $|V_{\text{int}}|$ over all possible decompositions of the n particles into clusters and over all configurations for a given clustering, subject to the condition that no particles from two different clusters are closer than $(2n^2)^{-1}\rho_x$. Then it is possible to show that

$$|\mathfrak{V}_n(\{\mathbf{z}\})| \leq |e^{\beta\bar{V}(\rho_x)} - 1|\, C_n e^{\beta B_n} \tag{A18}$$

for all paths in class I. Equation (A18) is a direct consequence of (A2) and the properties of the classical cluster functions. For $n = 3$, one easily sees that $\bar{V}(\rho_x)$ is the maximum of $|V_{ij}(\mathbf{x}_i - \mathbf{x}_j)|$ for $i \neq j = 1, 2, 3$, with $|\mathbf{x}_i - \mathbf{x}_j| > (2n^2)^{-1}\rho_x$. If all potentials have a strictly finite range, then $\bar{V}(\rho_x)$ is clearly identically zero for large enough ρ_x and, for potentials that fall off exponentially at large distances, $\bar{V}$ also falls off exponentially with ρ_x.

Finally, we have

$$\int_{\mathrm{I}} dP_{xy}\mathfrak{V}_n(\{z\}) \leq |e^{\beta\bar{V}(\rho_x)} - 1|\, C_n e^{\beta B_n} \int dP_{xy}$$
$$\leq |e^{\beta\bar{V}(\rho_x)} - 1|\, C_n e^{\beta B_n} \int dP_{xy}, \tag{A19}$$

and putting everything together gives

$$|\langle\{\mathbf{x}\}|\, U_n\, |\{R_\theta \mathbf{x}\}\rangle_{\text{c.m.}}|$$
$$< n^{-\frac{3}{2}}\lambda^{3-3n} C_n e^{\beta B_n}\left[\exp\left(-\frac{m}{32\beta n^4}\rho_x^2\right) + |e^{\beta\bar{V}(\rho_x)} - 1|\right] \tag{A20}$$

(Maxwell–Boltzmann statistics),

for ρ_x sufficiently large and θ satisfying (A7). Now, if the potentials fall off fast enough so that $|e^{\beta\bar{V}(\rho_x)} - 1|$ is integrable, which is certainly the case for finite range or exponentially falling potentials, then we have obtained an integrable bound on

$$\langle\{\mathbf{x}\}\,|\, U_n\{R_\theta \mathbf{x}\}\rangle_{\text{c.m.}},$$

and by the dominated convergence theorem, it is legal to bring the limit as $\theta \to 0$ inside the integral in (2.12).

It remains to take Bose and Fermi statistics into account. In this case one has to consider paths $\{\mathbf{z}\}$ starting at $\{\mathbf{x}\}$ but ending not at $\{R_\theta \mathbf{x}\}$ but at $\{PR_\theta \mathbf{x}\} = \{R_\theta P\mathbf{x}\}$, where P is some permutation of coordinates. Furthermore the integrand in (A3) is no longer just that given by (A2). Let us call the new integrand $\mathfrak{V}_n^{P'}$. For the moment we need to know nothing about $\mathfrak{V}_n^{P}$ other than that it is bounded by $C_n' e^{\beta B_n}$, where C_n' has the same significance as before. Now we divide the $(3n-3)$-dimensional space of $\{\mathbf{x}\}$'s into two regions. Region A is to be that where

$$\sum (P\mathbf{x}_i - \mathbf{x}_i)^2 \geq (32n^4)^{-1}\rho_x^2,$$

and Region B contains all $\{\mathbf{x}\}$ such that

$$\sum (P\mathbf{x}_i - \mathbf{x}_i)^2 < (32n^4)^{-1}\rho_x^2.$$

For θ so small that $|1 - \cos\theta| + |\sin\theta| < (128n^4)^{-1}$, it is easy to show that $\sum (PR_\theta \mathbf{x}_i - \mathbf{x}_i)^2 \geq (64n^4)^{-1}\rho_x^2$ in Region A and $\sum (PR_\theta \mathbf{x}_i - \mathbf{x}_i)^2 < (16n^4)^{-1}\rho_x^2$ in Region B. When $\{\mathbf{x}\}$ is in A, the path integral is trivially bounded by

$$\int dP^P \mathfrak{V}_n^P < C_n' e^{\beta B_n \lambda^{3-3n}} \exp\left[-(m/128n^4 B)\rho_x^2\right], \tag{A21}$$

which follows from (A5) and (A6). Region B can be treated as before. Dividing the paths into classes I and II as above, the previous bound holds for the Class II paths provided only that we replace C_n by C_n'. The class I paths again break up into two sets which always stay at least a distance $(2n^2)^{-1}\rho_x$ apart. Evidently, the permutation P cannot be interchanging the coordinates of two particles in different clusters. It is an elementary property of the U functions for Fermi or Bose statistics that when the particles are broken

up into two sets not connected by the permutation P, then the integrand $\mathfrak{V}_n^P$ satisfies the same bound (A19) for large distances between clusters as in the case of Maxwell–Boltzmann statistics, provided only that we again replace C_n by C_n'. Therefore, the treatment of Region B reduces to the case done above. One easily sees that (2.13) in the text is valid for Bose and Fermi statistics also.

APPENDIX B

We are concerned here with the limits of $\theta \to 0$ and with infinite volume. Only the case of Maxwell–Boltzmann statistics will be discussed. The extension to other statistics along the lines of Appendix A is trivial. Defining the sphere $\sum x_1^2 = \rho_x^2 < \Lambda^2$ as in the text, let

$$\langle\{\mathbf{x}\}|\ U_n\ |\{R_\theta \mathbf{x}\}\rangle_{\text{c.m.}}^{\Lambda} = \int_{\Lambda} dP_{R_\theta xx}\, \mathfrak{V}_n, \tag{B1}$$

where the subscript on the path integral means to include only those paths which stay inside a sphere of radius Λ. This is equivalent to putting an infinite potential barrier on the surface $\rho_x = \Lambda$, and enforces the boundary condition that the wave functions vanish on the sphere. We will not bother to prove that

$$\lim_{\theta\to 0}\lim_{\Lambda\to\infty}\int_{\Lambda} dP_{R_\theta xx}\,\mathfrak{V}_n = \lim_{\Lambda\to 0}\lim_{\theta\to 0}\int_{\Lambda} dP_{R_\theta xx}\,\mathfrak{V}_n = \int dP_{xx}\,\mathfrak{V}_n, \tag{B2}$$

since pointwise convergence should be obvious. One should, on the other hand, worry somewhat about bringing the limits inside the integral in (2.14). The proof that this can be done requires only a simple modification of the arguments in Appendix A. If $\rho_x \geq \Lambda$, then $\langle\{\mathbf{x}\}|\ U_n\ |\{R_\theta \mathbf{x}\}\rangle_{\text{c.m.}}^{\Lambda}$ is identically zero. (Remember that $\sum (R_\theta x_i)^2 = \sum x_1^2 = \rho_x^2$.) For $\rho_x < \Lambda$, breaks the paths into classes I_Λ and II_Λ which are the same as before, except that paths leaving the sphere are to be omitted. The contribution of the Class II_Λ paths is now bounded by

$$C_n e^{\beta B_n}\int_{\text{II}_\Lambda} dP_{xy} < C_n e^{\beta B_n}\int_{\text{II}} dP_{xy}, \tag{B3}$$

since II_Λ contains fewer paths than II. Similarly the I_Λ paths are easily seen to be bounded by

$$\int_{\text{I}_\Lambda} dP_{xy}\,\mathfrak{V}_n \leq |e^{\beta \mathfrak{V}(\rho_x)} - 1|\, C_n e^{\beta B_n}\int_{\text{I}_\Lambda} dP_{xy} \leq |e^{\beta \mathfrak{V}(\rho_x)} - 1|\, C_n e^{\beta B_n}\int_{\text{I}} dP_{xy}, \tag{B4}$$

since again I_Λ is smaller than I. With the same assumptions on the potentials as before, we thus have an integrable bound on $|\langle\{\mathbf{x}\}|\ U_n\ |\{R_\theta \mathbf{x}\}\rangle_{\text{c.m.}}^{\Lambda}|$, which is independent of both θ and Λ. Therefore the limits as $\theta \to 0$ and $\Lambda \to \infty$ can be brought under the integral over $\{\mathbf{x}\}$ in either order. Equation (2.28) may be thought of as the result one obtains by taking $\theta \to 0$ first and then $\Lambda \to \infty$. Equation (2.51) and most of our later work assumes the opposite order, $\Lambda \to \infty$ and then $\theta \to 0$.

APPENDIX C

This appendix is devoted to the task of evaluating the contribution of Fig. 5(b) to b_4 by using the small angle limit prescription.

We first sum over the repeated two-body T matrices and obtain

$$\begin{aligned} &-\frac{1}{2\pi}\,\text{Im}\,(\text{Tr}\ln S)_b \\ &= \text{Re Tr}\,\delta(\epsilon - H_0)[T_{12}^\dagger G_0 T_{13}^\dagger G_0 T_{14}^\dagger \\ &\quad + T_{12}^\dagger 2\pi i\delta(\epsilon - H_0) T_{13}^\dagger G_0 T_{14}^\dagger \\ &\quad + \tfrac{1}{3} T_{12}^\dagger 2\pi i\delta(\epsilon - H_0) T_{13}^\dagger 2\pi i\delta(\epsilon - H_0) T_{14}^\dagger]. \end{aligned} \tag{C1}$$

Here we remind ourselves that diagrams differing by a cyclic permutation of T matrices are considered as the same diagram in our convention. That is why we have the factor $\frac{1}{3}$ for the last term of (C1), to correct for over counting.

Substituting (C1) in (3.1) and performing the ϵ integral, we find

$$\begin{aligned} &4^{-\frac{3}{2}}(b_4)_b \\ &= -\lim_{\theta\to 0}\beta\int dp \\ &\quad \times e^{-\beta\epsilon(\mathbf{p})}\,\text{Re}\,\langle\mathbf{p}|\ T_{12}^\dagger\ |\mathbf{p}''\rangle\langle\mathbf{p}''|\ T_{13}^\dagger\ |\mathbf{p}'''\rangle\langle\mathbf{p}'''|\ T_{14}^\dagger\ |\mathbf{p}'\rangle \\ &\quad \times \text{P}(p_1^2 - p_1''^2)^{-1}\text{P}(p_1^2 - p_1'''^2)^{-1} \\ &\quad - \tfrac{1}{3}\pi^2\delta(p_1^2 - p_1''^2)\delta(p_1^2 - p_1'''^2)], \end{aligned} \tag{C2}$$

where the vectors $\mathbf{p}_1''$, $\mathbf{p}_1'''$ are defined in Fig. 6 and

$$|\mathbf{p}''\rangle = |\mathbf{p}_1''\mathbf{p}_2'\mathbf{p}_3\mathbf{p}_4\rangle, \quad |\mathbf{p}'''\rangle = |\mathbf{p}_1'''\mathbf{p}_2'\mathbf{p}_3'\mathbf{p}_4\rangle.$$

The algebra of evaluating (C2) is sufficiently complicated owing to the energy and momentum dependence of the T-matrix elements. We shall therefore divide it into three parts:

(1) The essential step is to use $\mathbf{p}_1$, $\mathbf{p}_1''$, $\mathbf{p}_1'''$ as independent variables of integration. To make this point transparent, we first ignore the structure of the T-matrix elements and regard them as constants to simplify the algebra.

(2) Then the general procedure of taking the $\theta \to 0$ limit is clear, and we put in the full energy momentum dependence T-matrix elements back.

(3) Finally, we compare the result with that obtained by using off-shell forward amplitudes.

Since the z components of the momenta do not play any role in our discussion, we shall ignore them.

1. Constant $T^\dagger$ Assumed

While particles 2, 3, and 4 appear on the same footing, particle 1 is clearly special. Roughly speaking, the effect of particles 2, 3, 4 is that of a "self-energy correction" to the propagation of particle 1.

Figure 6(b) shows how $\mathbf{p}_2$, $\mathbf{p}_3$, and $\mathbf{p}_4$ are related to the momentum transfers $\mathbf{p}_1 - \mathbf{p}_1''$, $\mathbf{p}_1'' - \mathbf{p}_1'''$, and $\mathbf{p}_1''' - \mathbf{p}_1'$. The angle θ together with the fact that $p_i^2 = p_i'^2$ completely determines $\mathbf{p}_2$, $\mathbf{p}_3$, and $\mathbf{p}_4$ once the momentum transfers are given. We have

$$\begin{aligned} p_2^2 &= (\mathbf{p}_1 - \mathbf{p}_1'')^2/4\sin^2(\theta/2), \\ p_3^2 &= (\mathbf{p}_1'' - \mathbf{p}_1''')^2/4\sin^2(\theta/2), \\ p_4^2 &= (\mathbf{p}_1''' - \mathbf{p}_1')^2/4\sin^2(\theta/2). \end{aligned} \tag{C3}$$

In terms of the angles θ_1, θ_2, and θ_3 defined in Fig. 6(b), we have

$$\epsilon(\mathbf{p}) = p_1^2 + p_2^2 + p_3^2 + p_4^2, \tag{C4}$$

$$\begin{aligned} p_2^2 + p_3^2 + p_4^2 &= [4\sin^2(\theta/2)]^{-1}[(p_1 - p_1'')^2 + (p_1'' - p_1''')^2 \\ &\quad + (p_1 - p_1''')^2 + 2p_1p_1''(1 - \cos\theta_1) \\ &\quad + 2p_1''p_1'''(1 - \cos\theta_2) + 2p_1'''p_1(1 - \cos\theta_3)], \end{aligned}$$

where $\theta_1 + \theta_2 + \theta_3 = \theta$. We can use p_1, p_1'', p_1''', θ_1, θ_2, and θ_3 as variables of integration. In (C2), the Boltzmann factor $e^{-\beta\epsilon(\mathbf{p})}$ is the only factor depending on θ_1, θ_2, and θ_3. It is clear from (C4) that, after integrating over θ_1, θ_2, and θ_3, the Boltzmann factor gives

$$e^{-\beta p_1^2} w(p_1, p_1'', p_1'''), \tag{C5}$$

where w is symmetric in p_1, p_1'', and p_1'''. The term with principal parts in (C2) is not symmetric in p_1, p_1'', and p_1'''. Let us symmetrize it and write the integrand as

$$\begin{aligned} w(p_1, p_1'', p_1''')\tfrac{1}{3}(&e^{-\beta p_1^2}\mathrm{P}(p_1^2 - p_1''^2)^{-1}\mathrm{P}(p_1^2 - p_1'''^2)^{-1} \\ &+ e^{-\beta p_1''^2}\mathrm{P}(p_1''^2 - p_1^2)^{-1}\mathrm{P}(p_1''^2 - p_1'''^2)^{-1} \\ &+ e^{-\beta p_1'''^2}\mathrm{P}(p_1'''^2 - p_1^2)^{-1}\mathrm{P}(p_1'''^2 - p_1''^2)^{-1} \\ &- \pi^2\delta(p_1^2 - p_1''^2)\delta(p_1''^2 - p_1'''^2)e^{-\beta p_1^2}). \end{aligned} \tag{C6}$$

In view of (C4), the Boltzmann factor falls off rapidly when θ is small, and for p_1, p_1'', and p_1''' not nearly equal. Thus, $w(p_1, p_1'', p_1''')$ is negligible except for $p_1 \approx p_1'' \approx p_1'''$. Let

$$\bar{p}_1^2 = \tfrac{1}{3}(p_1^2 + p_1''^2 + p_1'''^2), \tag{C7}$$

and expand the exponentials in (C6) around $\bar{p}_1^2$:

$$\begin{aligned} e^{-\beta p_1^2} = e^{-\beta\bar{p}_1^2} &+ (p_1^2 - \bar{p}_1^2)\frac{\partial}{\partial\bar{p}_1^2}e^{-\beta\bar{p}_1^2} \\ &+ \tfrac{1}{2}(p_1^2 - \bar{p}_1^2)^2\frac{\partial^2}{\partial(\bar{p}_1^2)^2}e^{-\beta\bar{p}_1^2} + O[(p_1^2 - \bar{p}_1^2)^3]. \end{aligned} \tag{C8}$$

Before substituting (C8) in (C7), we note the algebraic identity

$$\begin{aligned} a^m\mathrm{P}(a-b)^{-1}\mathrm{P}(a-c)^{-1} &+ b^m\mathrm{P}(b-c)^{-1}\mathrm{P}(b-a)^{-1} \\ &+ c^m\mathrm{P}(c-a)^{-1}\mathrm{P}(c-b)^{-1} \\ &- \pi^2\delta(a-b)\delta(b-c)a^m \\ &= 1, \quad \text{if} \quad m = 2, \\ &= 0, \quad \text{if} \quad m = 0, 1, 3, 4\cdots, \end{aligned} \tag{C9}$$

which can be verified easily. Now we substitute (C8) in (C6) and apply (C9) with $p_1^2 = a$, $p_1''^2 = b$, $p_1'''^2 = c$. We obtain from (C6)

$$w(p_1, p_1'', p_1''')\left(\frac{1}{3}\frac{1}{2}\frac{\partial^2}{\partial(\bar{p}_1^2)^2}e^{-\beta\bar{p}_1^2} + O(p_1^2 - \bar{p}_1^2, p_1''^2 - \bar{p}_1^2, p_1'''^2 - \bar{p}_1^2)\right). \tag{C10}$$

In the $\theta \to 0$ limit, the last term is negligible. Writing the first term as $\frac{1}{6}(\partial^2/\partial\epsilon(\mathbf{p})^2)e^{-\beta\epsilon(\mathbf{p})}$, we have

$$4^{-\frac{3}{2}}(b_4)_b = -\beta\int d\mathbf{p}\,\frac{1}{6}\frac{\partial^2}{\partial\epsilon(\mathbf{p})^2}e^{-\beta\epsilon(\mathbf{p})}(T^\dagger)^3, \tag{C11}$$

where $(T^\dagger)^3$ is the T-matrix elements whose $\mathbf{p}$ dependence we have ignored so far.

2. Structure of T Included

As is shown in Fig. 6(b), the vectors $\mathbf{p}_i$, $\mathbf{p}_i'$, $i = 2, 3, 4$, are completely determined by $\mathbf{p}_1$, $\mathbf{p}_1''$, and $\mathbf{p}'''$. The two-body T-matrix element

$$\langle\mathbf{p}_1\mathbf{p}_2|\, T^\dagger(\epsilon(\mathbf{p}) - p_3^2 - p_4^2)\,|\mathbf{p}_1''\mathbf{p}_2'\rangle$$

depends on p_1, p_1'', θ_1, θ and the two-body off-shell energy variable

$$\begin{aligned} \epsilon(\mathbf{p}) - p_3^2 - p_4^2 &= p_1^2 + p_2^2 \\ &\equiv p_1^2 + \nu(p_1, p_1'', \theta_1). \end{aligned} \tag{C12}$$

We may write

$$\begin{aligned} \langle\mathbf{p}_1\mathbf{p}_2|\, T^\dagger\,|\mathbf{p}_1''\mathbf{p}_2'\rangle &= f(p_1^2 + p_2^2, p_1, p_1'', \theta_1) \\ &\equiv f(p_1^2; p_1, p_1'', \theta_1). \end{aligned} \tag{C13}$$

It is easy to verify that f is symmetric in p_1 and p_1'' if the first argument is held fixed. Similarly, the second T-matrix element in (C2) depends on p_1'', p_1''', θ_2, θ

and the two-body off-shell energy variable

$$\epsilon(\mathbf{p}) - p_2^2 - p_4^2 = p_1^2 + p_3^2 = p_1^2 + \nu(p_1'', p_1''', \theta_2). \quad \text{(C14)}$$

We have

$$\langle \mathbf{p}_1'' \mathbf{p}_3 | \, T^\dagger \, | \mathbf{p}_1''' \mathbf{p}_3' \rangle = f(p^2; p_1'', p_1''', \theta_2). \quad \text{(C15)}$$

The same arguments show that

$$\langle \mathbf{p}_1''' \mathbf{p}_4 | \, T^\dagger \, | \mathbf{p}_1' \mathbf{p}_4' \rangle = f(p_1^2; p''', p, \theta_3). \quad \text{(C16)}$$

The product of the T-matrix elements and the Boltzmann factor, after being integrated over θ_1, θ_2, and θ_3, is a function

$$g(p_1^2; p_1^2, p_1''^2, p_1'''^2), \quad \text{(C17)}$$

which is symmetric in the last three variables, similar to (C5). Now we symmetrize the integrand of (C2) as we did in (C6). All previous arguments leading to (C10) hold, and we obtain, in place of the first term of (C10),

$$\frac{1}{6} \frac{\partial^2}{\partial (\bar{p}_1^2)^2} g(\bar{p}_1^2; p_1^2, p_1''^2, p_1'''^2). \quad \text{(C18)}$$

This derivative is equivalent to that taken with respect to the off-shell energy variable ϵ. Thus, (C11) is generalized to read

$$(b_4)_b = -4^{\frac{3}{2}} \beta \int d\mathbf{p} \, \frac{1}{6} \Big[\frac{\partial^2}{\partial \epsilon^2} \, e^{-\beta\epsilon} \langle \mathbf{p} | \, T_{12}^\dagger(\epsilon) \, | \mathbf{p} \rangle \times \langle \mathbf{p} | \, T_{13}^\dagger(\epsilon) \, | \mathbf{p} \rangle \langle \mathbf{p} | \, T_{14}^\dagger(\epsilon) \, | \mathbf{p} \rangle \Big]_{\epsilon = \epsilon(\mathbf{p})}. \quad \text{(C19)}$$

3. Evaluation Using Off-Shell Forward Amplitudes

This is easily done by making the replacement (3.13), and keeping the same η in (C1). A little algebra shows

$$-(1/2\pi) \operatorname{Im} (\operatorname{Tr} \ln S)_b = \tfrac{1}{3} \int d\mathbf{p} \operatorname{Re} \langle \mathbf{p} | \, G_0^3 - G_0^{\dagger 3} \, | \mathbf{p} \rangle \times \langle \mathbf{p} | \, T_{12}^\dagger \, | \mathbf{p} \rangle \langle \mathbf{p} | \, T_{13}^\dagger \, | \mathbf{p} \rangle \langle \mathbf{p} | \, T_{14}^\dagger \, | \mathbf{p} \rangle. \quad \text{(C20)}$$

Since

$$\langle \mathbf{p} | \, G_0^3 - G_0^{\dagger 3} \, | \mathbf{p} \rangle = \tfrac{1}{2} \delta''[\epsilon - \epsilon(\mathbf{p})], \quad \text{(C21)}$$

we have

$$(b_4)_b = 4^{\frac{3}{2}} \beta \int d\epsilon e^{-\beta\epsilon} (1/2\pi) \operatorname{Im} (\operatorname{Tr} \ln S)_b = -4^{\frac{3}{2}} \beta \int d\mathbf{p} \, \frac{1}{6} \Big[\frac{\partial^2}{\partial \epsilon} \, e^{-\beta\epsilon} \langle \mathbf{p} | \, T_{12}^\dagger \, | \mathbf{p} \rangle \langle \mathbf{p} | \, T_{13}^\dagger \, | \mathbf{p} \rangle \times \langle \mathbf{p} | \, T_{12}^\dagger \, | \mathbf{p} \rangle \Big]_{\epsilon = \epsilon(\mathbf{p})}, \quad \text{(C22)}$$

which is the same as (C19).

APPENDIX D

In this appendix we shall evaluate the contribution of Fig. 5(c) to b_4 using the small angle limit prescription. Let us first write down their contribution to the trace of the logarithm:

$$\begin{aligned} -(1/2\pi) \operatorname{Im} (\operatorname{Tr} \ln S)_c &= \operatorname{Re} \operatorname{Tr} \delta(\epsilon - H_0) \times (T_{32}^\dagger G_0 T_{23}^\dagger G_0 T_{12}^\dagger && (1) \\ &\quad + T_{32}^\dagger G_0 T_{23}^\dagger \Delta_0 T_{12}^\dagger && (2) \\ &\quad + T_{34}^\dagger \Delta_0 T_{23}^\dagger G_0 T_{12}^\dagger && (3) \\ &\quad + T_{34}^\dagger \Delta_0 T_{23}^\dagger \Delta_0 T_{12}^\dagger && (4) \\ &\quad + T_{23}^\dagger G_0 T_{12}^\dagger G_0 T_{34}^\dagger && (5) \\ &\quad + T_{12}^\dagger G_0 T_{34}^\dagger G_0 T_{23}^\dagger && (6) \\ &\quad + T_{23}^\dagger G_0 T_{12}^\dagger \Delta_0 T_{34}^\dagger), && (7) \quad \text{(D1)} \end{aligned}$$

where $\Delta_0 \equiv 2\pi i \delta(\epsilon - H_0)$, and the labels (1)–(7) identify the terms with diagrams in Fig. 5(c). We can simplify this expression by taking advantage of various symmetries such as time reversal, space inversion, and particle label permutation under which the trace is unchanged. We obtain

$$\begin{aligned} -\frac{1}{2\pi} \operatorname{Im} (\operatorname{Tr} \ln S)_c &= \operatorname{Re} \operatorname{Tr} \delta(\epsilon - H_0) \times \{ T_{34}^\dagger (P/\epsilon - H_0) T_{23}^\dagger (P/\epsilon - H_0) T_{12}^\dagger \\ &\quad + T_{23}^\dagger (P/\epsilon - H_0) [T_{12}^\dagger (P/\epsilon - H_0) T_{34}^\dagger \\ &\quad + T_{34}^\dagger (P/\epsilon - H_0) T_{12}^\dagger] \\ &\quad + \pi^2 T_{23}^\dagger \delta(\epsilon - H_0) T_{12}^\dagger \delta(\epsilon - H_0) T_{34}^\dagger \}. \quad \text{(D2)} \end{aligned}$$

For clarity, we shall again divide the work into three steps:

(1) Assume all the T-matrix elements in (D2) are constants, and factor them out. Then the algebra will be simple and the essential features of the approach will be clear.

(2) Then we put back in the T-matrix elements with their full energy and momentum dependence.

(3) Finally, we shall obtain the same result by using off-shell forward amplitudes.

These diagrams possess less symmetries than those in Fig. 5(b) and are more complicated. The structure is very different. It is therefore not surprising that the algebra here will appear quite different from that in Appendix C.

1. Constant T Assumed

Again, the z components of the momenta play no role in the discussion and will be ignored. The geometry is depicted in Fig. 7(a). The vector $\mathbf{k}$ is defined

as

$$\mathbf{k} = \mathbf{p}_1 + \mathbf{p}_2 = -\mathbf{p}_3 - \mathbf{p}_4. \tag{D3}$$

We shall regard $\mathbf{p}_1$, $\mathbf{p}_4$, and $\mathbf{k}$ as independent variables of integration. By geometry, we have

$$\begin{aligned} p_2^2 &= (\mathbf{k} - \mathbf{p}_1)^2 = p_1^2 + k^2 + 2kp_\perp \cos\phi_1, \\ p_3^2 &= (-\mathbf{k} - \mathbf{p}_4)^2 = p_4^2 + k^2 + 2kp_4 \cos\phi_4, \\ \epsilon(\mathbf{p}) &= 2(p_1^2 + p_4^2 + k^2 + kp_1 \cos\phi_1 + kp_4 \cos\phi_4). \end{aligned} \tag{D4}$$

We further define

$$\begin{aligned} d_1 &= 2(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{p}_1 = 2kp_1[\cos\phi_1 - \cos(\phi_1 + \theta)], \\ d_1' &= 2(\mathbf{k} - \mathbf{k}') \cdot \mathbf{p}_1' = 2kp_1[\cos\phi_1 - \cos(\phi_1 - \theta)], \\ d_4 &= 2(\mathbf{k} - \mathbf{k}') \cdot \mathbf{p}_4 = 2kp_4[\cos\phi_4 - \cos(\phi_4 - \theta)], \\ d_4' &= 2(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{p}_4' = 2kp_4[\cos\phi_4 - \cos(\phi_4 + \theta)]. \end{aligned} \tag{D5}$$

Substituting (D2) in (3.1) and integrating over ϵ, we obtain

$$(b_4)_c = -4^{\frac{3}{2}}\beta \int d\mathbf{p} A, \tag{D6}$$

$$\begin{aligned} A &= (T^\dagger)^3[\mathrm{P}(1/d_4')\mathrm{P}(1/d_1) + \mathrm{P}(1/d_4)\mathrm{P}(1/d_1)]e^{-\beta\epsilon(\mathbf{p})} \\ &= 2(T^\dagger)^3\mathrm{P}(1/d_4)\mathrm{P}(1/d_1)e^{-\beta\epsilon(\mathbf{p})}. \end{aligned} \tag{D7}$$

We have made use of the identity (C9) with $m = 0$, i.e.,

$$\begin{aligned} \mathrm{P}(x + y)^{-1}(\mathrm{P}x^{-1} + \mathrm{P}y^{-1}) + \pi^2\delta(x)\delta(y) \\ = \mathrm{P}x^{-1}\mathrm{P}y^{-1}, \end{aligned} \tag{D8}$$

and have changed ϕ_4 to $-\phi_4$ in d_4', in arriving at (D7). Let us make the change of variables

$$\begin{aligned} \phi_1 &\to \phi_1 - \theta/2, \\ \phi_4 &\to \phi_4 + \theta/2 \end{aligned} \tag{D9}$$

and define

$$\begin{aligned} x^\pm &= p_1 k \cos(\phi_1 \pm \theta/2), \\ y^\pm &= p_4 k \cos(\phi_4 \pm \theta/2). \end{aligned} \tag{D10}$$

Then

$$\begin{aligned} -2\mathbf{k} \cdot \mathbf{p}_1 &= x^-, \quad -2\mathbf{k}' \cdot \mathbf{p}_1 = x^+, \\ 2\mathbf{k} \cdot \mathbf{p}_4 &= y^+, \quad 2\mathbf{k}' \cdot \mathbf{p}_4 = y^-, \\ \epsilon(\mathbf{p}) &= 2(p_1^2 + p_4^2 + k^2) + x^- + y^+. \end{aligned} \tag{D11}$$

It then follows from (D7) that, under the ϕ_1, ϕ_4 integration,

$$\begin{aligned} A &= 2e^{-2\beta(p_1^2 + p_4^2 + k^2)} \\ &\quad \times e^{-\beta(x^- + y^+)}\mathrm{P}(x^- - x^+)^{-1}\mathrm{P}(y^+ - y^-)^{-1} \\ &= \tfrac{1}{2}e^{-2\beta(p_1^2 + p_4^2 + k^2)}\,\mathrm{P}\,\frac{e^{-\beta x^-} - e^{-\beta x^+}}{x^- - x^+}\,\mathrm{P}\,\frac{e^{-\beta y^+} - e^{-\beta y^-}}{y^+ - y^-} \\ &= \tfrac{1}{2}\beta^2 e^{-2\beta(p_1^2 + p_4^2 + k^2)}e^{-\beta(x^- + y^+)} \\ &\quad + O(x^- - y^+, y^+ - y^-). \end{aligned} \tag{D12}$$

where we have made use of the fact that $x^\pm$, $y^\pm \to x^\mp$, $y^\mp$ when we change ϕ_1, ϕ_4 to $-\phi_1$, $-\phi_4$. The small angle limit $\theta \to 0$ is now transparent. As $\theta \to 0$, $x^+ \to x^-$, $y^+ \to y^-$. Equation (D12) becomes, in this limit, simply

$$\frac{1}{2}\frac{\partial^2}{\partial\epsilon(\mathbf{p})^2}e^{-\beta\epsilon(\mathbf{p})}, \tag{D13}$$

which can be substituted in (D6) to obtain $(b_4)_c$. The essential feature of the above procedure is thus to write the integrand as a function of ϕ_1 and ϕ_4, which appear to be the convenient variables for the energy denominators, so that the $\theta \to 0$ limit becomes transparent.

2. Variable T-Matrix Elements

When one includes the energy and momentum dependence of the T-matrix elements, the algebra becomes very involved. We shall still use k, p_1, p_4, ϕ_1, and ϕ_4 as independent variables. The $\theta \to 0$ limit will be clear as a result of analyzing the ϕ_1, ϕ_4 dependence of the T-matrix elements. This ϕ_1, ϕ_4 dependence of T-matrix elements comes from (a) their dependence on the total energy $\epsilon(\mathbf{p})$ which depends on the angles via (D4), and (b) the angular dependence of the initial and final momenta specifying these matrix elements.

As will be shown, the effect of the singular denominators is to differentiate the T-matrix elements with respect to the angles. Owing to the symmetry properties of the T-matrix elements, these derivatives with respect to angles can be conveniently expressed as those with respect to $\epsilon(\mathbf{p})$. The angular dependence through (b) turns out to have no effect in the $\theta \to 0$ limit.

Let us examine the role of the energy variable $\epsilon(\mathbf{p})$ first. For the first and last term in (D2), $T_{12}^\dagger$, $T_{23}^\dagger$, $T_{34}^\dagger$, regarded as two-body T matrices, have their off-shell energy variables evaluated at

$$\begin{aligned} \epsilon(\mathbf{p}) &= p_3^2 - p_4^2 = p_1^2 + p_2^2 \quad \text{for } T_{12}^\dagger, \\ \epsilon(\mathbf{p}) - p_1^2 - p_4^2 &= p_2^2 + p_3^2 \quad \text{for } T_{23}^\dagger, \\ \epsilon(\mathbf{p}) - p_1^2 - p_2^2 &= p_3^2 + p_4^2 \quad \text{for } T_{34}^\dagger. \end{aligned} \tag{D14}$$

On the other hand, for the other two terms in (D2), while $T_{23}^\dagger$ and $T_{34}^\dagger$ still have the same energy variables given by (D14), $T_{12}^\dagger$ has its energy variable evaluated at

$$\epsilon(\mathbf{p}) - p_3''^2 - p_4^2 = p_1^2 + p_2^2 + (p_3^2 - p_3''^2). \tag{D15}$$

We shall do the problem in two steps. In Step I, we ignore the $p_3^2 - p_3''^2$ term in (D15) and complete the calculation. In Step II, we correct the error.

Step I: Similar to (D6), we obtain from (D2)

$$A = u(\phi_1, \phi_4)\mathrm{P}\frac{1}{d_4'}\,\mathrm{P}\frac{1}{d_1} + v(\phi_1, \phi_4)\mathrm{P}\frac{1}{d_4}\,\mathrm{P}\frac{1}{d_1}, \tag{D16}$$

where

$$\begin{aligned} u(\phi_1, \phi_4) &= \langle \mathbf{p}_3\mathbf{p}_4|\, T^\dagger\, |\mathbf{p}_3''\mathbf{p}_4'\rangle \langle \mathbf{p}_2\mathbf{p}_3''|\, T^\dagger\, |\mathbf{p}_2''\mathbf{p}_3'\rangle \\ &\quad \times \langle \mathbf{p}_1\mathbf{p}_2''|\, T^\dagger\, |\mathbf{p}_1'\mathbf{p}_2'\rangle\, e^{-\beta\epsilon(\mathbf{p})}, \\ v(\phi_1, \phi_4) &= \langle \mathbf{p}_2\mathbf{p}_3|\, T^\dagger\, |\mathbf{p}_2''\mathbf{p}_3'''\rangle \langle \mathbf{p}_1\mathbf{p}_2''|\, T^\dagger\, |\mathbf{p}_1'\mathbf{p}_2'\rangle \\ &\quad \times \langle \mathbf{p}_3'''\mathbf{p}_4|\, T^\dagger\, |\mathbf{p}_3'\mathbf{p}_4'\rangle e^{-\beta\epsilon(\mathbf{p})}. \end{aligned} \tag{D17}$$

Of course, besides ϕ_1 and ϕ_4, u and v depend on θ, p_1^2, p_4^2, and k^2, which are suppressed to simplify the notation. Since $d_4' \to d_4$ when we change ϕ_4 to $-\phi_4$, according to (D5) and (D6) can be written as

$$A = [u(\phi_1, -\phi_4) + v(\phi_1, \phi_4)]\mathrm{P}\frac{1}{d_1}\,\mathrm{P}\frac{1}{d_4}. \tag{D18}$$

Again, we make the change of variables (D9) and use the variables $x^\pm$, $y^\pm$ defined by (D11). We then have

$$\begin{aligned} A = [&u(\phi_1 - \theta/2, -\phi_4 - \theta/2) \\ &+ v(\phi_1 - \theta/2, \phi_4 + \theta/2)] \\ &\times \mathrm{P}\frac{1}{x^- - x^+}\,\mathrm{P}\frac{1}{y^+ - y^-}. \end{aligned} \tag{D19}$$

Before attempting the step corresponding to (D12), we have to examine the angular dependence of u and v.

Consider u, defined by (D17). The first matrix element [see Fig. 7(b)] is a function of $k^2, p_4^2, -\mathbf{k}\cdot\mathbf{p}_4' = -y^+$, $-\mathbf{k}\cdot\mathbf{p}_4' = -y^-$ and the energy variable $p_3^2 + p_4^2$. Here $-\mathbf{k}$ is the total momentum of the two particles. We can express the angular dependence of this matrix element as

$$\langle \mathbf{p}_3\mathbf{p}_4|\, T^\dagger\, |\mathbf{p}_3''\mathbf{p}_4'\rangle = f_{34}(p_3^2 + p_4^2;\, y^+,\, y^-). \tag{D20}$$

Since f_{34} is symmetric in the last two variables, it is unchanged under the change of variable $\phi_4 \to -\phi_4$ if the first argument is kept fixed.

Similarly, the third matrix element [see Fig. 7(b)] in u can be written as

$$\langle \mathbf{p}_1\mathbf{p}_2''|\, T^\dagger\, |\mathbf{p}_1'\mathbf{p}_2'\rangle = f_{12}(p_1^2 + p_2^2;\, x^+,\, x^-). \tag{D21}$$

We note that (D20) depends on ϕ_4 and (D21) on ϕ_1, but the second matrix element of u depends on both ϕ_1 and ϕ_4, and is more complicated. We shall parametrize it the following way. Label a general perturbation term for $T_{23}^\dagger$ as shown in Fig. 7(c). Written in terms of these labels, this term is

$$\begin{aligned} &\int \frac{d^3q}{(2\pi)^3}\frac{d^3q'}{(2\pi)^3}\cdots\frac{d^3q''}{(2\pi)^3} V(\mathbf{k} - \mathbf{q}) \\ &\quad \times V(\mathbf{q} - \mathbf{q}')\cdots V(\mathbf{q}'' - \mathbf{k}') \\ &\quad \times [p_2^2 + p_3^2 - 2q^2 - p_1^2 - p_4^2 - 2\mathbf{q}\cdot(\mathbf{p}_4' - \mathbf{p}_1)]^{-1} \\ &\quad \times [p_2^2 + p_3^2 - 2q^2 - p_1^2 - p_4^2 - 2\mathbf{q}'\cdot(\mathbf{p}_4' - \mathbf{p}_1)]^{-1} \\ &\quad \times \cdots \\ &\quad \times [p_2^2 + p_3^2 - 2q''^2 - p_1^2 - p_4^2 - 2\mathbf{q}''\cdot(\mathbf{p}_4' - \mathbf{p}_1)]^{-1}. \end{aligned} \tag{D22}$$

It is a function of

$$\begin{aligned} p_2^2 + p_3^2,\quad 2\mathbf{k}\cdot(\mathbf{p}_4' - \mathbf{p}_1) &= y^+ + x^-, \\ 2\mathbf{k}'\cdot(\mathbf{p}_4' - \mathbf{p}_1) &= y^- + x^+, \end{aligned}$$

and

$$(\mathbf{p}_4' - \mathbf{p}_1)^2 = p_4^2 + p_1^2 - 2p_1p_4\cos(\phi_1 - \phi_4).$$

Therefore, we have

$$\begin{aligned} &e^{-\beta\epsilon(\mathbf{p})}\langle \mathbf{p}_2\mathbf{p}_3''|\, T^\dagger\, |\mathbf{p}_2''\mathbf{p}_3'\rangle \\ &\quad = f_{23}(p_2^2 + p_3^2;\, x^- + y^+,\, x^+ + y^-,\, \cos(\phi_1 - \phi_4)), \end{aligned} \tag{D23}$$

where we have included the Boltzmann factor in f_{23}.

Equations (D20), (D21), and (D23) summarize the angular dependence of $u(\phi_1 - \theta/2, -\phi_4 - \theta/2)$. The same analysis applies to $v(\phi_1 - \theta/2, \phi_4 + \theta/2)$ as well. The result is

$$\langle \mathbf{p}_1\mathbf{p}_2''|\, T^\dagger\, |\mathbf{p}_1'\mathbf{p}_2'\rangle = f_{12}(p_1^2 + p_2^2;\, x^+,\, x^-), \tag{D24}$$

$$\langle \mathbf{p}_3'''\mathbf{p}_4|\, T^\dagger\, |\mathbf{p}_3'\mathbf{p}_4'\rangle = f_{34}(p_3^2 + p_4^2;\, y^-,\, y^+), \tag{D25}$$

$$\begin{aligned} &\langle \mathbf{p}_2\mathbf{p}_3|\, T^\dagger\, |\mathbf{p}_2''\mathbf{p}_3'''\rangle\, e^{-\beta\epsilon(\mathbf{p})} \\ &\quad = f_{23}(p_2^2 + p_3^2;\, x^- + y^+,\, x^+ + y^-,\, \cos(\phi_1 + \phi_4)). \end{aligned} \tag{D26}$$

These equations summarize the angular dependence of $v(\phi_1 - \theta/2, \phi_4 + \theta/2)$. The functions f_{12}, f_{23}, f_{34} are all symmetric in their second and third arguments by the space–time inversion symmetry of the T matrix. It is clear that

$$u(\phi_1 - \theta/2, -\phi_4 - \theta/2) + v(\phi_1 - \theta/2, \phi_4 + \theta/2)$$

is unchanged under the variable changes $\phi_1 \to -\phi_1$ and $\phi_4 \to -\phi_4$ if, in f_{ij}, the first arguments are held fixed. Thus the asymmetry in ϕ_1 and in ϕ_4 comes in only through the first arguments of these functions via

$$\begin{aligned} p_2^2 &= (\mathbf{k} - \mathbf{p}_1)^2 = k^2 + p_1^2 + x^-, \\ p_3^2 &= (-\mathbf{k} - \mathbf{p}_4)^2 = k^2 + p_4^2 + y^+. \end{aligned} \tag{D27}$$

We now substitute (D20)–(D26) in (D19) and obtain

$$\lim_{\theta\to 0} A = \lim_{\theta\to 0} 2f_{12}(x^-)f_{23}(x^- + y^+)f_{34}(y^+) \times \mathrm{P}(x^- - x^+)^{-1}\mathrm{P}(y^+ - y^-)^{-1} = \frac{1}{4}\frac{\partial^2}{\partial x \partial y} 2f_{12}(x)f_{23}(x+y)f_{34}(y), \quad \text{(D28)}$$

where only the $x^\pm$, $y^\pm$ dependence through the energy variables, i.e., through the first arguments of f_{12}, f_{23}, f_{34}, are kept, and x, y are the $\theta \to 0$ limits of $x^\pm$, $y^\pm$. We have thus completed Step I.

Step II: Now we proceed to include the fact that the $T_{12}^\dagger$ in the second and the third terms of (D2) has its energy variable given by (D15) instead of (D14), i.e., a difference of

$$p_3^2 - p_3'''^2 = y^+ - y^-. \quad \text{(D29)}$$

The term we have missed in A is

$$\begin{aligned}\Delta A &= T_{23}^\dagger(p_2^2 + p_3^2)[T_{12}^\dagger(p_1^2 + p_2^2 + p_3^2 - p_3'''^2) - T_{12}^\dagger(p_1^2 + p_2^2)]T_{34}^\dagger(p_3^2 + p_4^2) \times \mathrm{P}(d_1 + d_4)^{-1}\mathrm{P}(1/d_4)e^{-\beta\epsilon(\mathbf{p})} \\ &= f_{23}(x^- + y^+)[f_{12}(x^- + y^+ - y^-) - f_{12}(x^-)] \times f_{34}(y^+)\mathrm{P}(x^- + y^+ - x^+ - y^-)^{-1} \times \mathrm{P}(y^+ - y^-)^{-1}. \quad \text{(D30)}\end{aligned}$$

We have utilized our knowledge about the T-matrix elements summarized in (D20)–(D26) and have suppressed all irrelevant variables. Now we expand

$$f_{12}(x^- + y^+ - y^-) - f_{12}(x^-) = f_{12}'(x^-)(y^+ - y^-) + \tfrac{1}{2}f_{12}''(x^-)(y^+ - y^-)^2 + O[(y^+ - y^-)^3]. \quad \text{(D31)}$$

Substituting (D31) in (D30), the $\theta \to 0$ limit can be taken readily. We find

$$\lim_{\theta\to 0}\Delta A = \tfrac{1}{2}(f_{12}''f_{23}f_{34} + f_{12}f_{23}f_{34}'' + f_{12}'f_{23}f_{34}' + f_{12}'f_{23}'f_{34} + f_{12}f_{23}'f_{34}'). \quad \text{(D32)}$$

Step II is then completed.

It is now a matter of counting terms to show that the sum of (D32) and (D28) is

$$\lim_{\theta\to 0}(A + \Delta A) = \frac{1}{2}\left[\frac{\partial^2}{\partial\epsilon^2} e^{-\beta\epsilon}\langle\mathbf{p}|\, T_{34}^\dagger(\epsilon)\,|\mathbf{p}\rangle\langle\mathbf{p}|\, T_{23}^\dagger\,|\mathbf{p}\rangle \times \langle\mathbf{p}|\, T_{12}^\dagger(\epsilon)\,|\mathbf{p}\rangle\right]_{\epsilon=\epsilon(\mathbf{p})},$$

$$(b_4)_c = -\lim_{\theta\to 0} 4^{\frac{3}{2}}\beta\int d\mathbf{p}(A + \Delta A). \quad \text{(D33)}$$

3. Evaluation by Using Off-Shell Amplitudes

This is done easily by the replacement (3.13) and keeps the same η throughout in (D1). After a little algebra, (D1) becomes

$$\begin{aligned}-(1/2\pi)\,\mathrm{Im}\,(\mathrm{Tr}\ln S)_c &= \int d\mathbf{p}\,\mathrm{Re}\,\langle\mathbf{p}|\, G_0^3 - G_0^{3\dagger}\,|\mathbf{p}\rangle \times \langle\mathbf{p}|\, T_{34}^\dagger\,|\mathbf{p}\rangle\langle\mathbf{p}|\, T_{23}^\dagger\,|\mathbf{p}\rangle\langle\mathbf{p}|\, T_{12}^\dagger\,|\mathbf{p}\rangle \\ &= \int d\mathbf{p}\,\tfrac{1}{2}\,\mathrm{Re}\,\delta''[\epsilon - \epsilon(\mathbf{p})]\,\langle\mathbf{p}|\, T_{34}^\dagger\,|\mathbf{p}\rangle \times \langle\mathbf{p}|\, T_{23}^\dagger\,|\mathbf{p}\rangle\langle\mathbf{p}|\, T_{12}^\dagger\,|\mathbf{p}\rangle. \quad \text{(D34)}\end{aligned}$$

The forward T-matrix elements depend on the off-shell energy variable ϵ. Thus, we have

$$\begin{aligned}(b_4)_c &= -4^{\frac{3}{2}}\beta\int d\epsilon e^{-\beta\epsilon}(1/2\pi)\,\mathrm{Im}\,(\mathrm{Tr}\ln S)_c \\ &= -4^{\frac{3}{2}}\beta\int d\mathbf{p}\,\frac{1}{2}\left[\frac{\partial^2}{\partial\epsilon^2} e^{-\beta\epsilon}\langle\mathbf{p}|\, T_{34}^\dagger\,|\mathbf{p}\rangle \times \langle\mathbf{p}|\, T_{23}^\dagger\,|\mathbf{p}\rangle\langle\mathbf{p}|\, T_{12}^\dagger\,|\mathbf{p}\rangle\right]_{\epsilon=\epsilon(\mathbf{p})}. \quad \text{(D35)}\end{aligned}$$

Equation (D35) is, of course, the same as (D33).

* Alfred P. Sloan Foundation Fellow.

† Supported in part by the U.S. Atomic Energy Commission under contract AT(11-1)-GEN-10, P.A. 11.

‡ Research sponsored in part by the National Science Foundation, Grant No. GP-16147.

[1] This subject was originated by Beth and Uhlenbeck in Physica **3**, 729 (1936).

[2] Reference 3 contains a list of more recent papers which discuss the extension of Beth and Uhlenbeck's work to the third and higher virial coefficients. In addition, there is some recent work done by S. Y. Larson and P. L. Mascheroni, "The Quantum Mechanical 3rd Virial Coefficient and Three Body Phase Shifts," to be published. Bedeaux's work cited in Ref. 3 has now been published in Physica **45**, 469 (1969).

[3] R. Dashen, S. Ma, and H. J. Bernstein, Phys. Rev. **187**, 345 (1969). Referred to in the text as I.

[4] R. Dashen and S. Ma, J. Math. Phys. **11**, 1136 (1970). Referred to in the text as II.

[5] Here one obtains results essentially identical to those found ten years ago by T. D. Lee and C. N. Yang, Phys. Rev. **117**, 12 (1960).

[6] See, for example, K. Huang, *Statistical Mechanics* (Wiley, New York, 1963), Sec. 15.3.

[7] There are six terms like that written in (4.2), corresponding to the various possibilities for successive collisions among three particles. Assuming the particles to be identical, we concentrate on one term and drop a factor of $(3!)^{-1}$.

[8] A rigorous account of the properties of Feynman integrals applied to statistical mechanics can be found in J. Ginibre, J. Math. Phys. **6**, 238, 252 (1965). See also, E. Nelson, J. Math. Phys. **5**, 332 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Gauge Fields with Positive-Definite Energy Density

C. A. UZES
Department of Physics, University of Georgia, Athens, Georgia 30601

(Received 29 May 1970)

The relationship between the couplings of the non-Abelian gauge field and the internal holonomy group is investigated on the classical level, under the requirements of full local gauge invariance and positive definiteness of the gauge field's energy density. For the free gauge field it is found that each solution to the gauge field equation with self-coupling is associated with a simple compact internal holonomy group. When the gauge field is coupled to other multiplet fields, an internal symmetry group is present in addition to the holonomy group $\mathcal{H}$. In this case the internal holonomy group must also be simple and compact. In addition, homogeneity and isotropy of event space for closed systems of fields require that $\mathcal{H}$ be identical in group structure with either the symmetry group or a subgroup of the symmetry group and that all the field equations can be decoupled into sets of equations acting on multiplet fields transforming under irreducible representations of $\mathcal{H}$. Thus, in local gauge theory, multiplet fields can be classified with respect to simple and compact internal holonomy groups. Some comments are offered on the relationship between internal holonomy and symmetry groups.

I. INTRODUCTION

In local gauge theory the geometric properties of the gauge field and potentials allow a Lie group, the internal holonomy group $\mathcal{H}$, to be assigned to each analytic solution to the gauge field equation.[1] The use of the ordinary holonomy group in studies of solutions to the Einstein field equation[2-4] suggests that the internal holonomy group will be of interest in considering properties of solutions to the gauge field equation. To be sure, this group has proven a useful tool in investigating the gauge field's ability to carry internal charge,[5] the properties of gauge fields with spherical[6] and plane symmetry,[7] and the electromagnetic part of the free gauge field.[8] Compactness of $\mathcal{H}$ also insures the positive definiteness of the gauge field's energy density.[9] Combined with the physical criteria of positive definiteness of the gauge field's energy density, $\mathcal{H}$ has also proven useful in establishing conditions for nonexistence theorems on solutions to the free gauge field equation[10] and in establishing the precise nature of any internal charge carried by the gauge field.[11] Indeed, it is apparent that there must be an intimate connection between the possible internal holonomy groups and the physical properties of gauge fields. In order to determine more precisely this connection, the present paper is concerned with the relationship between the gauge field's couplings and internal holonomy groups, under the restrictions of full local gauge invariance and the positive definiteness of its energy density.

The investigation begins with a discussion of the concept of positive definiteness of the free (self-coupled) gauge field's energy density for classes of solutions to the gauge field equation and of the consequences of this restriction upon the Lie algebraic structure of the fields, i.e., the structure of $\mathcal{H}$. Keeping in mind that no internal symmetry group is present in the free field case, we consider the structure and decomposition of the free gauge field equation.

When the gauge field is coupled to other multiplet fields, it is necessary to introduce an internal metric in order to define a gauge invariant Lagrangian for the system of fields. This automatically defines an internal symmetry group. Of interest is how theories of the free gauge field fit into theories involving coupling between the gauge field and other multiplet fields defined by internal symmetry groups. A clarification of this point should be desirable as a basis for determining how symmetry breaking can come about in a locally gauge invariant theory, particularly if the asymptotic identification is employed to define the electromagnetic part of the gauge field.[8] The principle of homogeneity and isotropy of event space for closed systems serves here to clarify the relationship between the holonomy and symmetry groups. It also clarifies the connection between the classification of gauge fields given here via internal holonomy groups, and previous classifications of gauge fields coupled to other multiplet fields in terms of internal symmetry groups, based upon generalizations of the so-called Yang–Mills trick.[12,13]

In the following section some of the properties of the internal holonomy group employed below are briefly reviewed.[14] Full local gauge invariance is assumed throughout and the development is on the classical level. No quantized systems are considered. The event space is taken as Minkowskian, with x^κ, $\kappa = 0, 1, 2, 3$, standing for the event coordinates and ∂_κ for $\partial/\partial x^\kappa$.

II. THE INTERNAL GEOMETRY

In local gauge theory the particle multiplet fields are at each event defined with respect to a distinct n-dimensional internal space. The collection of bases

over event space is referred to as the internal base or gauge.[15] Under coordinate-dependent internal base transformations, covariant internal vectors, linear internal operators, and the $n \times n$ matrix fields of the gauge potential transform, respectively, according to the rules

$$\psi' = S^{-1}\psi, \quad P' = S^{-1}PS, \quad \Gamma'_\mu = S^{-1}\Gamma_\mu S - S^{-1}\partial_\mu S. \tag{1}$$

The transformation property of the gauge potential allows the introduction of the gauge covariant derivatives

$$\nabla_\mu\psi = \partial_\mu\psi - \Gamma_\mu\psi, \quad \nabla_\mu P = \partial_\mu P - [\Gamma_\mu, P], \tag{2}$$

which transform under (1) in the same manner as the quantities upon which they operate. The gauge field is defined in terms of the gauge potential by

$$\phi_{\mu\nu} = \partial_\mu\Gamma_\nu - \partial_\nu\Gamma_\mu - [\Gamma_\mu, \Gamma_\nu] \tag{3}$$

and transforms under (1) as a linear internal operator. By virtue of its definition, it also satisfies identically the "Bianchi" identity

$$\nabla_\mu\phi_{\kappa\lambda} + \nabla_\lambda\phi_{\mu\kappa} + \nabla_\kappa\phi_{\lambda\mu} = 0. \tag{4}$$

The internal holonomy group is defined in terms of the equivalence transport of internal vectors. The internal vector $\psi(x^\kappa + dx^\kappa)$ is defined to be equivalent to $\psi(x^\kappa)$ provided that the gauge-invariant condition

$$dx^\kappa\nabla_\kappa\psi = 0 \tag{5}$$

is satisfied at x^κ. In this case $\psi(x^\kappa + dx^\kappa)$ may be considered the result of the equivalence transport of $\psi(x^\kappa)$ along dx^κ to $x^\kappa + dx^\kappa$.

The element $H_\ell(x^\kappa)$ of $\mathcal{H}(x^\kappa)$ affects through the transformation

$$\psi' = H_\ell(x^\kappa)\psi \tag{6}$$

the equivalence transport of covariant internal vectors about the closed loop ℓ in event space originating at x^κ.

Consider the potentials to be analytic functions of the event coordinates. At each event the gauge field and its covariant derivatives span the Lie algebra $\mathfrak{L}$ of $\mathcal{H}$. One can always choose an internal base such that the linearly independent generators of $\mathcal{H}(x^\kappa)$, $L(x^\kappa)_m$, $m = 1, \cdots, p$, are numerically the same at all events.[1] In addition, the internal base can be chosen such that $\Gamma_\kappa \in \mathfrak{L}$.[16] This insures the existence of the expansions

$$\Gamma_\kappa = b_\kappa{}^m L_m, \quad \phi_{\kappa\lambda} = b_{\kappa\lambda}{}^m L_m,$$
$$\nabla_\mu \cdots \nabla_\nu\phi_{\kappa\lambda} = b_{\kappa\lambda,\nu\cdots\mu}{}^m L_m, \tag{7}$$

where $\partial_k L_m = 0$, and the $b_\kappa{}^m$, $b_{\kappa\lambda}{}^m$, etc., are real functions of the event coordinates. The following discussion is in terms of such a basis.

III. POSITIVE DEFINITENESS AND $\mathcal{H}$

The action principle applied to the gauge field Lagrangian[17]

$$L_F = -\tfrac{1}{4}\,\mathrm{Tr}\,\phi_{\kappa\lambda}\phi^{\kappa\lambda} \tag{8}$$

yields the free gauge field equation

$$\nabla_\nu\phi^{\mu\nu} = 0. \tag{9}$$

Note that (8) and (9) are automatically invariant under invertible transformations of the type (1), and thus that the notion of an internal symmetry group is not involved in discussions of properties of the free gauge field.[18] The criterion of positive definiteness is motivated by considering the initial value problem for (9) at $x^0 = 0$.

The constraint equation, the zero component of (9), sets restrictions on the initial values that the ϕ_{jk} and $\Gamma_j, j, k = 1, 2, 3$, may assume on the surface $x^0 = 0$. Consider a gauge in which $\Gamma_0 = 0$ at $x^0 = 0$. It is easily seen that such a gauge, consistent with (7), always exists. The constraint equation reduces to

$$\partial_j\phi^{0j} - [\Gamma_j, \phi^{0j}] = 0, \tag{10}$$

where

$$\phi^{0j} = \partial^0\Gamma^j.$$

Given the Γ_j, (10) is then a linear differential equation for ϕ^{0j}, and, in principle, solutions can easily be found at $x^0 = 0$.[19]

Using the identity

$$\nabla_\mu \cdots \nabla_\beta(\nabla_0\nabla_\gamma - \nabla_\gamma\nabla_0)\nabla_\delta \cdots \nabla_\alpha\phi_{\kappa\lambda}$$
$$= \nabla_\mu \cdots \nabla_\beta[\nabla_\delta \cdots \nabla_\alpha\phi_{\kappa\lambda}, \phi_{0\gamma}],$$

we can permute the time covariant derivatives in

$$\nabla_\mu \cdots \nabla_\nu\phi_{\kappa\lambda}$$

with the spatial ones and eliminate them, using either (4) or (9). Since in the gauge $\Gamma_0 = 0$ the time covariant derivative is just the partial time derivative, it is clear that an analytic solution to (10) determines a solution to (9), at least in the neighborhood of the surface $x^0 = 0$. Therefore, one should expect a large number of possible solutions to (9), corresponding to the large number of initial configurations of the gauge field.

The gauge invariant Yang–Mills energy–momentum tensor density is

$$T^{\kappa\lambda} = -\mathrm{Tr}\,(\phi^{\kappa\mu}\phi_\mu{}^\lambda + \tfrac{1}{4}g^{\kappa\lambda}\phi_{\beta\gamma}\phi^{\beta\gamma}).$$

It is a generalization of that used in electromagnetic theory, and is determined in the usual fashion from the free Lagrangian (8) by using Noether's theorem. In

the basis (7), the energy density takes the form

$$T^{00} = -\tfrac{1}{2} l_{mn}\left(\sum_j b_{0j}{}^m b_{0j}{}^n + \tfrac{1}{2}\sum_{jk} b_{jk}{}^m b_{jk}{}^n\right), \quad (11)$$

where

$$l_{mn} = \mathrm{Tr}\, L_m L_n$$

is a tensor in the group space of $\mathcal{K}$.

In a particular solution to the gauge field equation, not all Lie algebraic components of $\phi_{\kappa\lambda}$ need contribute to (11), since the gauge fields by themselves need not span $\mathfrak{L}$. However, it is clear from the above discussion that one can always construct a solution to the field equation such that the expansion coefficients $b_{\kappa\lambda}{}^m$ with respect to any desired generator L_m are nonvanishing. The $b_{\kappa\lambda}{}^m$ are real since $\phi_{\kappa\lambda}$ belongs to the real Lie algebra of $\mathcal{K}$. Since l_{mn} is symmetric, the energy density can be real for the whole class of solutions for a given $\mathcal{K}$ if and only if l_{mn} is also real. But then l_{mn} may be placed in diagonal form by a real algebra base transformation. It follows that the energy density will be positive definite for the class of solutions with a given $\mathcal{K}$ if and only if l_{mn} is a negative definite tensor in the group space of $\mathcal{K}$.

The structure constants of $\mathcal{K}$ are defined through the commutation relations

$$[L_m, L_n] = c_{mn}{}^p L_p. \quad (12)$$

The quantity

$$f_{mnp} = \mathrm{Tr}\,[L_m, L_n] L_p = c_{mn}{}^q l_{qp} \quad (13)$$

is totally antisymmetric, since the trace of any commutator vanishes. The negative definiteness of l_{mn} allows an "orthogonal" algebra basis to be constructed in the sense that

$$l_{mn} = -\delta_{mn},$$

or in which

$$c_{mn}{}^p = -f_{mnp}. \quad (14)$$

Thus, if the energy density of the gauge field is positive definite, there exists an algebra base in which the structure constants of $\mathcal{K}$ can be considered as totally antisymmetric in their group space indices.

$\mathcal{K}$ may be written as the direct product of its invariant subgroups. Construct for the invariant subgroup an "orthogonal" set of generators, and construct the remaining generators of $\mathcal{K}$ from the orthogonal complement of the subspace spanned by them. This can always be done. Let the indices m, n, and p refer to generators of the invariant subgroup and t and s to any of the remaining generators of $\mathcal{K}$. Then, by definition,

$$[L_t, L_m] = c_{tm}{}^n L_n, \quad [L_m, L_n] = c_{mn}{}^p L_p,$$

since $c_{tm}{}^s = c_{mn}{}^t = 0$. However, consistency with (14) requires

$$c_{tm}{}^s = c_{mn}{}^t = c_{ts}{}^m = c_{tm}{}^n = 0. \quad (15)$$

It follows that if the energy density of the gauge field is positive definite for the class of solutions with a given $\mathcal{K}$, then $\mathcal{K}$ may in its most general form be expressed as a direct product of an Abelian group $\mathfrak{C}$, its center, and simple groups $\mathcal{R}_r$. The direct product of simple groups, of course, constitutes a semisimple group, denoted here by $\mathcal{R}$.

$\mathcal{K}$ is a connected Lie group, each closed loop ℓ defining a transformation of $\mathcal{K}$ may be continuously shrunk to a point. All of its transformations may be expressed in the form

$$H_\ell(x^\kappa) = \exp\,[\eta_\ell(x^\kappa)^m L_m], \quad (16)$$

where the $\eta_\ell(x^\kappa)^m$ are the group parameters of $\mathcal{K}$ defining the equivalence transport of internal vectors about some closed loop ℓ in event space originating at x^κ. It follows that the adjoint and linear adjoint groups of $\mathcal{K}$, defined by all possible similarity transformations on the elements of $\mathcal{K}$ and $\mathfrak{L}$ of the form $H_\ell H_{\ell'} H_\ell^{-1}$ and $H_\ell L_m H_\ell^{-1}$, respectively, are isomorphic. But they are not necessarily isomorphic to $\mathcal{K}$ itself;

$$H_1 H_\ell H_1^{-1} = H_2 H_\ell H_2^{-1}$$

for all H_ℓ if the product $H_1^{-1}H_2$ commutes with all transformations of $\mathcal{K}$, i.e., if $\mathcal{K}$ has a center.[20] However, the adjoint and linear adjoint groups are isomorphic to $\mathcal{R}$.

The generators of the linear adjoint group are the structure constants of $\mathcal{K}$. From the antisymmetry of f_{mnp} and (13) it follows that

$$c_{mn}{}^p l_{pq} + c_{mq}{}^p l_{pn} = 0. \quad (17)$$

Thus l_{mn} is invariant under all transformations of the linear adjoint group. Let w_m and v_m be any two vectors in the Lie algebra space of $\mathcal{K}$, the space in which the linear adjoint group operates, and let w'_m and v'_m be the vectors resulting from them under any transformation of this group. It follows from (17) that

$$v'_m l^{mn} w'_n = v_m l^{mn} w_n.$$

Thus, there exists a definite scalar product in the group space of $\mathcal{K}$, and the linear adjoint group of $\mathcal{K}$ is compact.[21] But the latter is isomorphic to $\mathcal{R}$. Hence, positive definiteness implies that $\mathcal{K}$ is a direct product of an Abelian group, its center, and compact simple groups.

IV. POSITIVE DEFINITENESS AND THE FREE GAUGE FIELD

The Lie algebra of $\mathcal{K}$ is the sum of the Lie algebras of each of its invariant subgroups. Using the basis (7)

and taking into account (15), one can write

$$^{\tau}\phi_{\kappa\lambda} = \partial_\kappa {}^{\tau}\Gamma_\lambda - \partial_\lambda {}^{\tau}\Gamma_\kappa - [{}^{\tau}\Gamma_\kappa, {}^{\tau}\Gamma_\lambda],$$
$$^{\tau}\nabla_\mu\phi_{\kappa\lambda} = \partial_\mu {}^{\tau}\phi_{\kappa\lambda} - [{}^{\tau}\Gamma_\mu, {}^{\tau}\phi_{\kappa\lambda}],$$

and also

$$^{c}\phi_{\kappa\lambda} = \partial_\kappa {}^{c}\Gamma_\lambda - \partial_\lambda {}^{c}\Gamma_\kappa, \quad {}^{c}\nabla_\mu\phi_{\kappa\lambda} = \partial_\mu {}^{c}\phi_{\kappa\lambda},$$

where the left superscript τ refers to the parts of Γ_κ, $\phi_{\kappa\lambda}$, and $\nabla_\mu\phi_{\kappa\lambda}$ belonging to the Lie algebra of $\mathcal{R}_\tau$ and the left superscript c refers to those parts belonging to the Lie algebra of $\mathcal{C}$.

Due to the linear independence of the generators of $\mathcal{H}$, the free gauge field equation and Bianchi identity decouple into the τ sets of uncoupled equations

$$^{\tau}\nabla_\mu\phi^{\kappa\mu} = 0, \quad {}^{\tau}\nabla_\mu\phi_{\kappa\lambda} + {}^{\tau}\nabla_\lambda\phi_{\mu\kappa} + {}^{\tau}\nabla_\kappa\phi_{\lambda\mu} = 0,$$

while in addition the $^{c}\phi_{\kappa\lambda}$ satisfy Maxwell's equations.

For a theory in which there is free gauge field and nothing else, the $^{c}\phi_{\kappa\lambda}$ are of no interest. One is interested in localized solutions to the gauge field equation, blobs of free gauge field energy density which may interact with one another. The $^{c}\phi_{\kappa\lambda}$ cannot participate in the self-interaction of the gauge field. In this case it should be sufficient to consider only gauge fields with simple and compact $\mathcal{H}$.[22]

In the free field case,the nature of the internal space is not of interest, the field equations may be expressed solely in terms of the Lie algebraic components of the gauge fields and potentials, the $b_\kappa{}^m$ and $b_{\kappa\lambda}{}^m$. Of interest is whether in a situation in which the gauge field is coupled to other particle multiplet fields a similar decoupling of field equations can be made with respect to the irreducible representations of simple and compact internal holonomy groups in internal space.

V. POSITIVE DEFINITENESS AND COUPLED FIELDS

When the gauge field is coupled to other multiplet fields, the energy density of the gauge field itself must still be positive definite for any class of solutions with a given $\mathcal{H}$. It is clear that this condition is sufficient to insure that also in the coupled case $\mathcal{H}$ can at most be the direct product of an Abelian group $\mathcal{C}$, its center, and compact simple groups $\mathcal{R}_\tau$. For definiteness, the gauge field in the following is considered coupled to a Dirac multiplet field. The coupling is introduced by replacing in the free spinor Lagrangian all partial derivatives by their gauge invariant counterparts, i.e., through the generalized principle of minimal coupling.[12,13,23] It will be clear, however, that the results below will apply equally well to any gauge invariant coupling between the gauge field and arbitrary multiplet fields.

Let $\psi(x^\kappa)_a$, $a = 1, \cdots, n$, (spinor indices suppressed)[24] denote the Dirac multiplet field in question. In order to construct a free spinor Lagrangian invariant under internal base transformations, one needs an internal (hybrid) metric. Let $g_{\bar{a}b}$ be such a metric, transforming under internal base transformations according to the rule (in component notation)

$$g_{\bar{a}'b'} = S^{\bar{a}}_{\bar{a}'} g_{\bar{a}b} S^{b}_{b'},$$

where $S^{\bar{a}}_{\bar{a}}$, is the complex conjugate of $S^{a}_{a'}$ and where $g_{\bar{a}'b'}$ denotes the components of $g_{\bar{a}b}$ in the new internal base. The Lagrangian for the closed system of interacting fields can be expressed in the form

$$L_T = -\tfrac{1}{2}i[\bar{\psi}_{\bar{a}}\gamma^\mu g^{\bar{a}b}\nabla_\mu\psi_b - (\nabla_\mu\bar{\psi}_{\bar{a}})\gamma^\mu g^{\bar{a}b}\psi_b] - M\bar{\psi}_{\bar{a}}g^{\bar{a}b}\psi_b - L_F, \quad (18)$$

where[25]

$$\nabla_\mu\psi_b = (\partial_\mu\delta^c_b - \Gamma_{\mu b}{}^{c})\psi_c,$$

and where $\bar{\psi}_{\bar{a}}$ denotes the Dirac adjoint spinor $\psi_{\bar{a}}\gamma^0$. Note that no internal metric is involved in the expression for L_F, [(8)].

The representation of $\mathcal{H}$ in internal space is its self-representation, which may be reducible. The principle problem of this section is to determine the structure of the "interaction" or coupling in the Lagrangian (18) and resulting field equations with respect to the subspaces in the internal space transforming invariantly under the self-representations of the internal holonomy group and also with respect to the compact simple groups $\mathcal{R}_\tau$. It is resolved by invoking the principle that the event space must be gauge-invariantly homogeneous and isotropic with respect to an arbitrary displacement of the closed system of fields.

The condition of gauge-invariant homogeneity and isotropy of space is satisfied if, at each and every event when each internal multiplet field is equivalence-displaced along an arbitrary line increment dx^κ (line increments emanating from different events being parallel), the total Lagrangian remains invariant. The equivalence instead of the ordinary displacement is required from local gauge invariance, since at each event the multiplet fields are defined with respect to a distinct internal base. In making the displacement, it is necessary to drag along with the fields the internal base used. The internal metric is not displaced, however, since it is not a dynamical field.

The gauge field and potentials are included as part of the closed system of fields. It is necessary, therefore, to define the equivalence transport of the gauge potential. This is done by requiring that the equivalence displaced field $\phi'_{\mu\nu}$ be defined in terms of the

displaced potentials Γ'_μ according to (3). Since by definition

$$\phi'_{\mu\nu} = \phi_{\mu\nu} - dx^\kappa[\Gamma_\kappa, \phi_{\mu\nu}], \tag{19}$$

one can easily verify that this condition requires the definition

$$\Gamma'_\mu = \Gamma_\mu + \partial_\mu(dx^\kappa\Gamma_\kappa) + [dx^\kappa\Gamma_\kappa, \Gamma_\mu]. \tag{20}$$

Using the definitions (5), (19), and (20) for the equivalence transport of fields along dx^κ, one finds that the Lagrangian for the closed system of fields will remain invariant provided that the internal metric satisfies the condition

$$\partial_\mu g_{\bar{a}b} - \Gamma_{\mu\bar{a}}{}^{\bar{c}}\, g_{\bar{c}b} - \Gamma_{\mu b}{}^{c}\, g_{\bar{a}c} = 0,$$

i.e., that

$$\nabla_\mu g_{\bar{a}b} = 0.$$

Geometrically, this means that the same internal metric must be employed at each event. It follows from (6) that $g_{\bar{a}b}$ must be invariant under all transformations of $\mathcal{K}$, that is,

$$H_{\bar{a}}{}^{b}\, g_{bc} H_{\bar{d}}{}^{c} = g_{\bar{a}\bar{d}}. \tag{21}$$

Hence $\mathcal{K}$ must be identical to or a subgroup of the group of gauge transformations which leave the internal metric invariant; $\mathcal{K}$ is identical to or a subgroup of the internal symmetry group.[26]

Since $\mathcal{R}$ is compact, its self-representation in the internal space may be taken to be unitary. Then there exists an invariant bilinear form

$$\theta_a h^{ab} \zeta_b,$$

where θ_a and ζ_b are internal vectors, ζ_b is the complex conjugate of ζ_b, and where h^{ab} is a positive-definite Hermitian tensor in internal space, the holonomy metric. The representation of $\mathcal{R}$ in internal space is then completely reducible. The internal space may be written as the sum of orthogonal (with respect to h^{ab}) invariant subspaces, each transforming under an irreducible representation of $\mathcal{R}$ in internal space. Of importance is whether these subspaces are also orthogonal with respect to the internal metric g_{ab} appearing in (18). If so, only the components of the Dirac multiplet field lying in the same irreducible subspace are coupled by the non-Abelian gauge field in the Lagrangian (18).

The nonsingular linear internal operator A defined by

$$A_a{}^c = g_{ab} h^{bc}$$

satisfies

$$AR - RA = 0,$$

where R is any transformation of $\mathcal{R}$. The eigenvectors of A with common eigenvalue then form invariant subspaces under the unitary transformations of $\mathcal{R}$, and the different invariant subspaces are orthogonal with respect to the holonomy metric. From Schurr's lemma it follows that an internal base can be chosen such that A has diagonal form. Without loss of generality, the base can be chosen such that

$$h_{ab} = \delta_{ab}.$$

Let these irreducible invariant subspaces in internal space be labeled by I_α. Then g_{ab} is proportional to h_{ab} in each I_α. Thus the subspaces transforming under the irreducible representations of $\mathcal{R}$ in internal space are orthogonal with respect to both the holonomy metric and the metric defining the internal symmetry group g_{ab}.

The generators of the center of $\mathcal{K}$ commute among themselves and with all transformations of $\mathcal{R}$. Therefore, it is also possible to decompose the internal space into invariant orthogonal subspaces transforming irreducibly under $\mathcal{R}$, where in each of these subspaces, say $I_{\alpha'}$, the generators of $\mathcal{C}$ are multiples of the identity operator. Commuting linear operators have simultaneous eigenvectors.

The irreducible subspaces $I_{\alpha'}$ so defined must coincide with the I_α defined above by the diagonal form of A, i.e., A must commute with the transformations of $\mathcal{C}$; for in the opposite situation at least two irreducible subspaces would be required to have a nontrivial intersection, leading to a contradiction. The I_α and $I_{\alpha'}$ were chosen to be irreducible.

Since the transformations of $\mathcal{C}$ commute with A and leave invariant the metric g^{ab}, they also leave invariant h^{ab}, which is definite in each I_α. It follows that if $\mathcal{K}$ has a center, then the center of $\mathcal{K}$ is also a compact group. Then the irreducible invariant subspaces transforming under irreducible representations of $\mathcal{K}$, coinciding with those of $\mathcal{R}$, are orthogonal with respect to the internal metric. This result would hold, for example, even if g_{ab} were indefinite.

Suppose the procedure for decomposing the internal space into the sum of irreducible subspaces (under $\mathcal{K}$) has been carried out. Then in the basis (7) the L_m, the Γ_κ, and the $\phi_{\kappa\lambda}$ must have a block diagonal form with respect to the I_α. Furthermore, one has

$$\psi(\alpha)_{\bar{a}} g^{\bar{a}b} L_{mb}{}^{c}\, \theta(\beta)_c = 0,$$

where $\psi(\alpha)$ and $\theta(\beta)$ are, respectively, internal vectors belonging to the irreducible subspaces I_α and I_β. Therefore, the Lagrangian (18) may be simplified to

$$L_T = \sum_\alpha L(\alpha)_D + L(\alpha)_F,$$

$$L(\alpha)_F = -\tfrac{1}{4} \operatorname{Tr} \phi(\alpha)_{\kappa\lambda} \phi(\alpha)^{\kappa\lambda}, \tag{22}$$

where

$$L(\alpha)_D = -\tfrac{1}{2}i[\bar{\psi}(\alpha)_{\bar{a}}\gamma^\mu g^{\bar{a}b}\nabla_\mu\psi(\alpha)_b - (\nabla_\mu\bar{\psi}(\alpha)_b)\gamma^\mu g^{\bar{a}b}\psi(\alpha)_b] - M\bar{\psi}(\alpha)_{\bar{a}}g^{\bar{a}b}\psi(\alpha)_b \tag{23}$$

and where

$$\nabla_\mu\psi(\alpha)_b = (\partial_\mu\delta^c_b - \Gamma(\alpha)_{\mu b}{}^c)\psi(\alpha)_c.$$

The remaining problem is to clarify the relationship between the irreducible self-representations of $\mathcal{K}$ and those of the $\mathcal{R}_\tau$. The self-representations of any $\mathcal{R}_\tau$ may be broken down into irreducible representations in the internal space, and all these representations must be faithful. The field equation

$$J(\alpha)^\kappa = \nabla_\lambda\phi(\alpha)^{\kappa\lambda},$$

where

$$J(\alpha)^\kappa{}_b{}^c = i\bar{\psi}(\alpha)_{\bar{a}}\gamma^\mu\psi(\alpha)_b g^{\bar{a}c}$$

is the external source current, and the Bianchi identity (4) may be looked upon from two points of view, either in terms of their Lie algebraic components or in terms of their matrix elements in internal space. Consistency requires faithful irreducible representations of each $\mathcal{R}_\tau$.

Suppose the subspaces transforming under irreducible representations of a particular group, say $\mathcal{R}_1$, have been identified. The matrices representing the generators of this subgroup must be in block diagonal form with respect to these invariant subspaces. The generators of the other subgroups commute with those of $\mathcal{R}_1$, and their eigenvectors can be used to label the subspaces invariant under $\mathcal{R}_1$ (not necessarily uniquely). The subspaces so labeled must coincide for all the generators of each $\mathcal{R}_\tau$ with $\tau > 1$, since the subspaces are irreducible. Otherwise one would have a nonnull intersection of two irreducible subspaces, itself an invariant subspace, leading to a contradiction. This means that the matrices representing the generators of the groups $\mathcal{R}_\tau$ with $\tau > 1$ must be multiples of the identity operator in each of the invariant subspaces. However, since each representation must be faithful, this requires that, in the subspaces transforming under irreducible representations of $\mathcal{R}_1$, the matrices representing the generators of each $\mathcal{R}_\tau$ with $\tau > 1$ all vanish. Hence the irreducible self-representations of $\mathcal{R}$ coincide with those of an $\mathcal{R}_\tau$. The self-representation of $\mathcal{R}$ is constructed by taking the direct sum of irreducible self-representations of simple compact holonomy groups in the internal space. The subspaces I_α are those transforming under irreducible representations of these groups.

The internal holonomy group is further restricted by considering the nature of the possible interactions and conserved charges generated by the components of ${}^c\phi_{\kappa\lambda}$, belonging to the Lie algebra of the center of $\mathcal{K}$. Such components satisfy Maxwell's equations. Since in each I_α the generators of $\mathcal{C}$ are multiples of the identity operator, all components of the external multiplet field in I_α have equal weight in contributing to the external source current

$$j(\alpha)^\kappa{}_m = \mathrm{Tr}\,{}^cL(\alpha)J(\alpha)^\kappa, \quad {}^cL \in {}^c\mathfrak{L}.$$

Therefore, all the particles represented by the different components of $\psi(\alpha)$ can be expected to have long-range repulsive Coulomb-like interactions. Such a situation would be in conflict with experimental data, the interaction being the same as that which would be expected to arise due to a baryon number field and associated conservation law.[27] Thus the center of $\mathcal{K}$ must be trivial.

Due to the total antisymmetry of the structure constants of $\mathcal{K}$ and to the orthogonality of the irreducible subspaces I_α, the gauge field equation can be decomposed into the form

$${}^rJ(\alpha)^\kappa = \nabla_\lambda\,{}^r\phi(\alpha)^{\kappa\lambda},$$

while a similar decomposition holds for the Lagrangian (18) and the field equations for other multiplet fields. Thus one must conclude that, in a locally gauge-invariant theory, gauge fields of physical interest have simple compact $\mathcal{K}$ and that all the multiplet fields may be classified according to the irreducible representations of $\mathcal{K}$.

VI. DISCUSSION

Via the fundamental Lie algebraic and group properties of the gauge field present in any locally gauge-invariant theory, the features of solutions to the gauge field equation were investigated under the requirement that their resultant energy density must always come out positive. For a theory involving only the self-coupled gauge field, this was done independently of the notion of an internal symmetry group. The results indicate that free gauge fields of physical interest must have simple and compact $\mathcal{K}$.

In considering the case where the gauge field is both self-coupled and coupled to other multiplet fields, it was necessary to introduce the notion of an internal symmetry group. The precise nature of the coupling between the gauge field and the other multiplet fields, and of the symmetry group, was left unspecified. Gauge-invariant homogeneity and isotropy of event space for closed systems led to the condition that $\mathcal{K}$ must be either identified with or a subgroup of the symmetry group. The possibility of the latter being a subgroup of $\mathcal{K}$, allowed through the requirement

that $\mathcal{H}$ and the symmetry group be compatible,[1] is not allowed.

Positive definiteness taken alone requires in the case of coupled fields only that $\mathcal{H}$ be the direct product of an Abelian group $\mathcal{C}$, its center, and compact simple groups $\mathcal{R}$. However, the subspaces transforming under the irreducible self-representations of $\mathcal{H}$ and $\mathcal{R}$ must coincide. It was further found that the presence of a nontrivial center could be expected to lead to unphysical pseudo-electromagneticlike interactions between external multiplet fields. Hence, in locally gauge-invariant theories involving either coupled or free gauge fields, gauge fields of physical interest should have simple compact $\mathcal{H}$. Since the multiplet fields transforming under different irreducible representations of $\mathcal{H}$ are orthogonal with respect to the internal metric, all the multiplet fields can be classified according to the irreducible representations of this group.

In the classification of gauge fields allowed by generalizations of the Yang–Mills trick, global symmetries of the first kind were generalized through minimal coupling to local symmetries, following the lead of Yang and Mills. The result was that theories derived along these lines should have simple symmetry groups, with the gauge field transforming under the adjoint representation of the symmetry group. The internal holonomy group played no role. In the present investigation, the classification problem was approached from the point of view of how to generalize possible theories of the free self-coupled gauge field and imbed them into symmetry theories involving other multiplet fields. $\mathcal{H}$ was treated as fundamental. It is interesting to note that local gauge invariance alone (with recognition of the homogeneity and isotropy of event space) is sufficient to require that the gauge potentials are anti-Hermitian with respect to the internal metric and transform under the adjoint representation of the internal symmetry group. These assumptions do not have to be made at the outset, as is customary in generalizing the Yang–Mills trick.

The precise relationship between the internal holonomy and symmetry groups requires further investigation. The statement that $\mathcal{H}$ must be either identical to or a subgroup of the symmetry group refers only to their structure and not to the physical notions associated with them. This is particularly important when one recognizes that both groups may have a strong coordinate dependence.

Although the $\mathcal{H}(x^\kappa)$ at different events are isomorphic, they may tend asymptotically away from a localized region in event space to an Abelian group of lower dimensionality. This fact has played an important role in leading to the asymptotic identification for the electromagnetic part of the gauge field.[8] Therein, $\mathcal{H}$ tends away from the sources of the gauge field to an Abelian group, and the long-range component of the gauge field is identified with the electromagnetic field. Crucial is the coordinate dependence of the internal symmetry group. If one requires that the theory be invariant under the full symmetry group at all events, there clearly can be no electromagnetic breaking of isospin symmetry. In this case it would appear necessary to reject the asymptotic identification. On the other hand, in the context of the asymptotic identification, the charge measuring operators can be shown to restrict the allowed symmetry and holonomy group transformations as one moves away from the sources of the gauge field. Since the latter governs the dynamic symmetry of the interactions between sources, one can construct a model in which the symmetries of the interactions between sources decreases with their increased separation. Thus by not attempting to *a priori* identify the gauge field with a particle multiplet field, one can conjecture at a theory in which the gauge field presents a unified approach to electromagnetic and strong interactions within the context of local gauge theory and which at the same time accounts for electromagnetic and medium strength breaking of strong interactions.[28]

[1] H. G. Loos, J. Math. Phys. **8**, 2114 (1967).
[2] J. F. Schell, J. Math. Phys. **2**, 202 (1961).
[3] R. P. Kerr and J. N. Goldberg, J. Math. Phys. **2**, 332 (1961).
[4] J. N. Goldberg and R. P. Kerr, J. Math. Phys. **2**, 327 (1961).
[5] C. A. Uzes, Ann. Phys. (N.Y.) **50**, 534 (1968).
[6] H. G. Loos, Nucl. Phys. **72**, 677 (1967).
[7] C. A. Uzes, J. Math. Phys. **10**, 1885 (1969).
[8] H. G. Loos, Nuovo Cimento **53A**, 365 (1968).
[9] R. P. Treat, Nuovo Cimento **50A**, 871 (1967); Ph.D. thesis, University of California, Riverside, 1967.
[10] H. G. Loos, Nuovo Cimento **53A**, 201 (1968).
[11] C. A. Uzes, Nuovo Cimento **62A**, 157 (1969).
[12] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[13] S. L. Glashow and M. Gell-Mann, Ann. Phys. (N.Y.) **15**, 437 (1961).
[14] For a more complete discussion of these properties, see Ref. 1.
[15] The term gauge invariance as employed here refers to invariance of the theory under coordinate dependent changes in this set of internal bases, and should be distinguished from coordinate-dependent symmetry (object) transformations on multiplet fields.
[16] E. Lubkin, private communication via H. G. Loos.
[17] Tr denotes the trace of linear internal operators.
[18] The transformation properties of the gauge fields and potentials are the same for both coordinate-dependent gauge and symmetry transformations.
[19] An algebraic solution to (10) is given in Ref. 11.
[20] J. A. Schouten, *Ricci Calculus* (Springer-Verlag, Berlin, 1954), 2nd. ed., Chap. IV.
[21] L. S. Pontryagin, *Topological Groups* (Gordon and Breach, New York, 1966), 2nd. ed., Chap. II.
[22] It follows that the nonexistence theorem of Ref. 10 applies to all localized solutions to the free gauge field equation of physical interest provided that they satisfy the restrictions therein on time dependence, etc.
[23] Other than the defined transformation property (3), no restriction need be placed here at the outset on the manner in which the

gauge potential appears in the gauge covariant derivative (2), in distinction to investigations based upon generalizations of the Yang–Mills trick. The latter assume that either the potentials transform under the adjoint representation of an internal symmetry group as in Ref. 13 or some equivalent condition such as the invertibility of certain matrices, as in Ref. 12.

[24] The following convention is used below. The indices $m, n, \cdots, t$ refer to the group space of $\mathcal{K}$, while the indices a, b, c, and d refer to the internal space.

[25] $g^{a\bar{b}}$ denotes the inverse of $g_{a\bar{b}}$, and it is assumed that $g^{a\bar{b}}$ is Hermitian.

[26] Since invariance of the mass term in (18) under an equivalence transport of the fields also leads to (21), it is evident that (21) must hold for arbitrary gauge-invariant couplings between the gauge and other multiplet fields. One also has that gauge invariance alone is sufficient to insure the reality of the interaction energy density ($g^{a\bar{b}}\Gamma_{\mu b}{}^{c} = -\Gamma_{\mu\bar{b}}{}^{\bar{a}}g^{\bar{b}c}$) and also that the potentials can always be expanded in terms of the generators of the internal symmetry group.

[27] C. N. Yang and T. D. Lee, Phys. Rev. **98**, 1501 (1955).

[28] C. A. Uzes, "Asymmetry in the Asymptotic Identification" (unpublished).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Electromagnetic Diffraction by a Thin Conducting Annular Disk*

D. L. JAIN AND R. P. KANWAL
Department of Mathematics, Pennsylvania State University, University Park, Pennsylvania 16802

(Received 9 March 1970)

The problem of diffraction of time harmonic electromagnetic waves by a perfectly conducting thin annular disk is solved when the incident wave is a plane wave traveling in a direction perpendicular to the annulus. The electromagnetic problem is reduced to two scalar problems with the help of the Hertz vector representation formulas. Each of these scalar problems is formulated in terms of a Fredholm integral equation of the first kind, which is subsequently reduced to a system of four simultaneous Fredholm integral equations of the second kind. This system is then solved by the straightforward iteration scheme. Low-frequency approximations are presented for various physical quantities such as the components of the induced surface current density at the disk and the scattering cross section.

1. INTRODUCTION

The problem of electromagnetic diffraction by a thin circular conducting disk has been solved by many authors and by widely differing techniques. An account of these attempts is given in Refs. 1–6. However, there has been no attempt made so far to solve the corresponding problem of an annular disk. The reason is that it is a three-part boundary value problem, and until recently there were not many mathematical techniques available for solving such a problem. With the appearance of a paper by Gubenko and Mossakovskii[7] a great interest has been aroused to solve three-part boundary value problems. In the process, several interesting methods have been devised. An integral equation technique which has its origin in the researches of Copson,[8] Williams,[9] Thomas,[10] and the present authors[11,12] appears to be very suitable for the purpose. In this paper we extend that technique and present the solution for the problem of electromagnetic diffraction by an ideally conducting circular annulus.

We use a Hertz vector formulation as used by Bazer and Rubenfeld[6] in their study of the electromagnetic diffraction by a circular aperture in a plane screen. Thereby, the present problem reduces to two scalar problems which are linked by three arbitrary constants. These constants are subsequently evaluated with the help of the edge and continuity conditions. Each of these two scalar problems is reduced to the solution of a Fredholm integral equation of the first kind which is subsequently reduced to two Volterra integral equations and a system of four simultaneous Fredholm integral equations of the second kind. The former have a rather elementary kernel and therefore can be readily inverted, while the latter are solved by the standard iteration procedure. Although these integral equations are valid for all wavelengths and all the ratios of the inner and outer radii of the annulus, they are specially useful when the waves have long wavelengths compared with the outer radius and the inner radius is much smaller than the outer radius.

In Sec. 2 we present the mathematical formulation of the problem. Section 3 contains the mathematical technique for solving this problem, and we have included it for the sake of completeness as well as for ease of reference. We present the solution in Sec. 4. The edge conditions are studied in Sec. 5, wherein the three constants mentioned above are evaluated. In

gauge potential appears in the gauge covariant derivative (2), in distinction to investigations based upon generalizations of the Yang–Mills trick. The latter assume that either the potentials transform under the adjoint representation of an internal symmetry group as in Ref. 13 or some equivalent condition such as the invertibility of certain matrices, as in Ref. 12.

[24] The following convention is used below. The indices $m, n, \cdots, t$ refer to the group space of $\mathcal{K}$, while the indices a, b, c, and d refer to the internal space.

[25] $g^{a\bar{b}}$ denotes the inverse of $g_{a\bar{b}}$, and it is assumed that $g^{a\bar{b}}$ is Hermitian.

[26] Since invariance of the mass term in (18) under an equivalence transport of the fields also leads to (21), it is evident that (21) must hold for arbitrary gauge-invariant couplings between the gauge and other multiplet fields. One also has that gauge invariance alone is sufficient to insure the reality of the interaction energy density ($g^{a\bar{b}}\Gamma_{\mu b}{}^{c} = -\Gamma_{\mu\bar{b}}{}^{\bar{a}}g^{\bar{b}c}$) and also that the potentials can always be expanded in terms of the generators of the internal symmetry group.

[27] C. N. Yang and T. D. Lee, Phys. Rev. **98**, 1501 (1955).

[28] C. A. Uzes, "Asymmetry in the Asymptotic Identification" (unpublished).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Electromagnetic Diffraction by a Thin Conducting Annular Disk*

D. L. JAIN AND R. P. KANWAL
Department of Mathematics, Pennsylvania State University, University Park, Pennsylvania 16802

(Received 9 March 1970)

The problem of diffraction of time harmonic electromagnetic waves by a perfectly conducting thin annular disk is solved when the incident wave is a plane wave traveling in a direction perpendicular to the annulus. The electromagnetic problem is reduced to two scalar problems with the help of the Hertz vector representation formulas. Each of these scalar problems is formulated in terms of a Fredholm integral equation of the first kind, which is subsequently reduced to a system of four simultaneous Fredholm integral equations of the second kind. This system is then solved by the straightforward iteration scheme. Low-frequency approximations are presented for various physical quantities such as the components of the induced surface current density at the disk and the scattering cross section.

1. INTRODUCTION

The problem of electromagnetic diffraction by a thin circular conducting disk has been solved by many authors and by widely differing techniques. An account of these attempts is given in Refs. 1–6. However, there has been no attempt made so far to solve the corresponding problem of an annular disk. The reason is that it is a three-part boundary value problem, and until recently there were not many mathematical techniques available for solving such a problem. With the appearance of a paper by Gubenko and Mossakovskii[7] a great interest has been aroused to solve three-part boundary value problems. In the process, several interesting methods have been devised. An integral equation technique which has its origin in the researches of Copson,[8] Williams,[9] Thomas,[10] and the present authors[11,12] appears to be very suitable for the purpose. In this paper we extend that technique and present the solution for the problem of electromagnetic diffraction by an ideally conducting circular annulus.

We use a Hertz vector formulation as used by Bazer and Rubenfeld[6] in their study of the electromagnetic diffraction by a circular aperture in a plane screen. Thereby, the present problem reduces to two scalar problems which are linked by three arbitrary constants. These constants are subsequently evaluated with the help of the edge and continuity conditions. Each of these two scalar problems is reduced to the solution of a Fredholm integral equation of the first kind which is subsequently reduced to two Volterra integral equations and a system of four simultaneous Fredholm integral equations of the second kind. The former have a rather elementary kernel and therefore can be readily inverted, while the latter are solved by the standard iteration procedure. Although these integral equations are valid for all wavelengths and all the ratios of the inner and outer radii of the annulus, they are specially useful when the waves have long wavelengths compared with the outer radius and the inner radius is much smaller than the outer radius.

In Sec. 2 we present the mathematical formulation of the problem. Section 3 contains the mathematical technique for solving this problem, and we have included it for the sake of completeness as well as for ease of reference. We present the solution in Sec. 4. The edge conditions are studied in Sec. 5, wherein the three constants mentioned above are evaluated. In

Sec. 6 we present the formulas for the components of the induced surface current density at the disk and the scattering cross section.

2. MATHEMATICAL FORMULATION OF THE PROBLEM

Cylindrical polar coordinates (ρ, φ, z) are chosen with the origin at the center of the annular disk so that the disk occupies the region $z = 0$, $b \leq \rho \leq a$, for all φ. The medium has the permittivity ϵ and permeability μ, while the speed of propagation of electromagnetic waves is $c = (\epsilon\mu)^{-\frac{1}{2}}$. Maxwell equations in source-free regions (in rationalized MKS units) are

$$\begin{aligned} &\operatorname{curl} \mathbf{E} - i\omega\mu\mathbf{H} = 0, \quad \operatorname{div} \mathbf{E} = 0, \\ &\operatorname{curl} \mathbf{H} + i\omega\epsilon\mathbf{E} = 0, \quad \operatorname{div} \mathbf{H} = 0, \end{aligned} \tag{2.1}$$

where $\mathbf{E}$ and $\mathbf{H}$ are the electric field and magnetic field intensities, respectively. In the above equations and in the sequel, a time dependence $\exp(-i\omega t)$ is understood and omitted.

Let an incident plane wave traveling along the z axis, due to sources in the half-space $z < 0$, impinge on the circular annulus. If $\mathbf{E}^{(0)}$ and $\mathbf{H}^{(0)}$ are the time-independent parts of the electric field and magnetic field intensities of this incident wave, then

$$\begin{aligned} \mathbf{E}^{(0)} &= (E_\rho^{(0)}, E_\varphi^{(0)}, E_z^{(0)}) \\ &= (\cos\varphi, -\sin\varphi, 0)\exp(ikz), \\ \mathbf{H}^{(0)} &= (H_\rho^{(0)}, H_\varphi^{(0)}, H_z^{(0)}) \\ &= (\epsilon/\mu)^{\frac{1}{2}}(\sin\varphi, \cos\varphi, 0)\exp(ikz), \end{aligned} \tag{2.2}$$

and $k = \omega/c$. The time-independent parts of the electric and magnetic intensities of the total field are

$$\mathbf{E} = \mathbf{E}^{(0)} + \mathbf{E}^{(1)}, \quad \mathbf{H} = \mathbf{H}^{(0)} + \mathbf{H}^{(1)}, \tag{2.3}$$

where $\mathbf{E}^{(1)}$ and $\mathbf{H}^{(1)}$ correspond to the diffracted field. We can represent the diffracted field in terms of a magnetic Hertz vector $\mathbf{M}$:

$$\begin{aligned} \mathbf{M} &= (M_\rho, M_\varphi, M_z) \\ &= \{M_1(\rho, z)\sin\varphi, M_1(\rho, z)\cos\varphi, M_2(\rho, z)\sin\varphi\}, \end{aligned} \tag{2.4}$$

$$\mathbf{E}^{(1)} = i\omega\mu \operatorname{curl}\mathbf{M}, \quad \mathbf{H}^{(1)} = (\operatorname{grad}\operatorname{div}\mathbf{M}) + k^2\mathbf{M}, \tag{2.5}$$

where

$$(\nabla^2 + k^2)\mathbf{M} = 0$$

and

$$(\nabla^2 + k^2)M_1(\rho, z) = 0$$

or

$$\frac{\partial^2 M_1}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial M_1}{\partial\rho} + \frac{\partial^2 M_1}{\partial z^2} + k^2 M_1 = 0, \tag{2.6}$$

$$(\nabla^2 + k^2)[M_2(\rho, z)\sin\varphi] = 0,$$

or

$$\frac{\partial^2 M_2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial M_2}{\partial\rho} - \frac{M_2}{\rho^2} + \frac{\partial^2 M_2}{\partial z^2} + k^2 M_2 = 0. \tag{2.7}$$

Besides, from (2.4) and (2.5), we have

$$\begin{aligned} E_\rho^{(1)} &= i\omega\mu\left(\frac{M_2}{\rho} - \frac{\partial M_1}{\partial z}\right)\cos\varphi, \\ H_\rho^{(1)} &= \left(\frac{\partial^2 M_1}{\partial\rho^2} + \frac{\partial^2 M_2}{\partial\rho\partial z} + k^2 M_1\right)\sin\varphi, \\ E_\varphi^{(1)} &= i\omega\mu\left(\frac{\partial M_1}{\partial z} - \frac{\partial M_2}{\partial\rho}\right)\sin\varphi, \\ H_\varphi^{(1)} &= \left(\frac{1}{\rho}\frac{\partial M_1}{\partial\rho} + \frac{1}{\rho}\frac{\partial M_2}{\partial z} + k^2 M_1\right)\cos\varphi, \\ E_z^{(1)} &= i\omega\mu\left(\frac{\partial M_1}{\partial\rho}\right)\cos\varphi, \\ H_z^{(1)} &= \left(\frac{\partial^2 M_1}{\partial\rho\partial z} - \frac{\partial^2 M_2}{\partial\rho^2} - \frac{1}{\rho}\frac{\partial M_2}{\partial\rho} + \frac{M_2}{\rho^2}\right)\sin\varphi. \end{aligned} \tag{2.8}$$

The boundary conditions are:

(i) The tangential component of E vanishes on the annulus;

(ii) E and H are continuous across the region $z = 0$, $\rho > a$, $0 \leq \rho < b$;

(iii) M_1 and M_2 satisfy the radiation condition at infinity.

In terms of the scalar fields M_1 and M_2, these boundary conditions take a simple form. Indeed, for M_1 we have

$$\left[\frac{\partial M_1}{\partial z}\right]_{z=0+} = \left[\frac{\partial M_1}{\partial z}\right]_{z=0-} = A - \frac{i}{\omega\mu}, \quad b \leq \rho \leq a, \tag{2.9}$$

$$M_1 \text{ and } \frac{\partial M_1}{\partial z} \text{ are continuous across } z = 0, \quad \rho > a, \tag{2.10}$$

$$[M_1(\rho, z)]_{z=0+} - [M_1(\rho, z)]_{z=0-} = B, \quad 0 \leq \rho < b, \tag{2.11}$$

$$\left[\frac{\partial M_1}{\partial z}\right]_{z=0+} - \left[\frac{\partial M_1}{\partial z}\right]_{z=0-} = C, \quad 0 \leq \rho < b, \tag{2.12}$$

where A, B, and C are three unknown constants which shall be eventually evaluated with the help of the edge conditions and continuity considerations of the quantities $[M_1(\rho, z)]_{z=0\pm}$ and $[M_2(\rho, z)]_{z=0\pm}$, which are functions of ρ.

The corresponding boundary conditions for M_2 are

$$[M_2]_{z=0+} = [M_2]_{z=0-} = A\rho, \quad b \leq \rho \leq a, \quad (2.13)$$

M_2 and $\dfrac{\partial M_2}{\partial z}$ are continuous across $z = 0, \quad \rho > a,$ (2.14)

$$\left[\frac{\partial M_2}{\partial z}\right]_{z=0+} - \left[\frac{\partial M_2}{\partial z}\right]_{z=0-} = -k^2 B\rho, \quad 0 \leq \rho < b, \quad (2.15)$$

$$[M_2]_{z=0+} - [M_2]_{z=0-} = C\rho, \quad 0 \leq \rho < b. \quad (2.16)$$

The electromagnetic problem has thus been reduced to two scalar problems: one, to determine M_1 from (2.6) subject to the boundary conditions (2.9)–(2.12) and, second, to determine M_2 from (2.7) subject to the boundary conditions (2.13)–(2.16). These two scalar problems are linked by three arbitrary constants A, B, and C.

The next step is to give both these scalar problems an integral equation formulation. We start with M_1 and define the jump

$$I(\rho) = [M_1(\rho, z)]_{z=0+} - [M_1(\rho, z)]_{z=0-}. \quad (2.17)$$

From (2.10) and (2.11) it follows that

$$I(\rho) = 0, \quad \rho > a, \quad I(\rho) = B, \quad 0 \leq \rho < b. \quad (2.18)$$

It is traditionally known by applying the Green's function method[1,13] that the differential equation (2.6) subject to the boundary conditions (2.9)–(2.11) has the integral representation formula

$$M_1(\rho, z) = \frac{1}{4\pi}\int_0^a\int_0^{2\pi} tI(t)\left\{\frac{\partial}{\partial z_1}\left(\frac{e^{ikR}}{R}\right)\right\}_{z_1=0} d\varphi_1\, dt - \frac{C}{4\pi}\int_0^b\int_0^{2\pi} t\left\{\frac{e^{ikR}}{R}\right\}_{z_1=0} d\varphi_1\, dt, \quad (2.19)$$

where (t, φ_1, z_1) denotes the point on the annular disk and

$$R = [\rho^2 + t^2 - 2\rho t\cos(\varphi - \varphi_1) + (z - z_1)^2]^{\frac{1}{2}}.$$

Using the boundary condition (2.9) in (2.19), we obtain

$$\begin{aligned} A - \frac{i}{\omega\mu} &= -\frac{1}{4\pi}\int_0^a\int_0^{2\pi} tI(t)\left\{\frac{\partial^2}{\partial z^2}\left(\frac{e^{ikR}}{R}\right)\right\}_{\substack{z_1=0\\ z=0}} d\varphi_1\, dt \\ &= \frac{1}{2}\left(\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + k^2\right)\left[\int_0^a tI(t) \times \int_0^\infty \frac{pe^{-\gamma|z|}J_0(p\rho)J_0(pt)}{\gamma}\, dp\, dt\right]_{z=0}, \quad b \leq \rho \leq a, \end{aligned} \quad (2.20)$$

where

$$\begin{aligned} \gamma &= -i(k^2 - p^2)^{\frac{1}{2}}, \quad k \geq p, \\ &= (p^2 - k^2)^{\frac{1}{2}}, \quad p \geq k. \end{aligned} \quad (2.21)$$

But

$$\left(\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho}\right)J_0(p\rho) = -p^2 J_0(p\rho);$$

therefore, (2.20) becomes

$$\begin{aligned} 2\left(A - \frac{i}{\omega\mu}\right) &= -\left[\int_0^a tI(t)\int_0^\infty p\gamma e^{-\gamma|z|}J_0(p\rho)J_0(pt)\, dp\, dt\right]_{z=0} \\ &= -\frac{1}{\rho}\frac{\partial}{\partial\rho}\left[\rho\left(\int_0^b B\frac{d}{dt}\int_0^\infty \frac{t\gamma}{p}e^{-\gamma|z|}J_1(p\rho)J_1(pt)\, dp\, dt + \int_b^a I(t)\frac{d}{dt}\int_0^\infty t\frac{\gamma}{p}e^{-\gamma|z|}J_1(p\rho)J_1(pt)\, dp\, dt\right)\right]_{z=0}, \\ & \qquad b \leq \rho \leq a, \end{aligned} \quad (2.22)$$

where we have used the boundary condition (2.18).

It follows from the integral representation formula (2.19) for the function $M_1(\rho, z)$ that $[M_1(\rho, z)]_{z=\pm 0}$ are continuous functions of ρ for all values of ρ if and only if the function $I(\rho)$ defined by the relation (2.17) is also continuous for all values of ρ, and we obtain from (2.18)

$$I(a) = 0, \quad B = I(b). \quad (2.23)$$

Furthermore, it is interesting to observe that if the relation (2.23) did not hold, then we would have ended up getting a divergent integral (as $\rho \to b$) on the right side of (2.22).

From (2.22) and (2.23) it follows that

$$2\rho\left(A - \frac{i}{\omega\mu}\right) = \frac{\partial}{\partial\rho}\left(\rho\int_b^a t\phi_1(t)\int_0^\infty \frac{\gamma}{p}J_1(p\rho)J_1(pt)\, dp\, dt\right), \quad b \leq \rho \leq a, \quad (2.24)$$

where

$$\phi_1(\rho) = \frac{d}{d\rho}[I(\rho)] \quad \text{and} \quad I(a) = 0. \quad (2.25)$$

Now integrate (2.24) with respect to ρ, and divide both sides by ρ to get the Fredholm integral equation of the first kind,

$$f_1(\rho) = \left(A - \frac{i}{\omega\mu}\right)\rho = \int_b^a t\phi_1(t)K_{11}(t, \rho)\, dt, \quad b \leq \rho \leq a, \quad (2.26)$$

where the kernel $K_{11}(t, \rho)$ is

$$K_{11}(t, \rho) = \int_0^\infty \frac{\gamma}{p} J_1(p\rho) J_1(pt)\, dp. \tag{2.27}$$

Let us now attend to finding the corresponding integral representation formula for M_2 from the boundary value problem as embodied in the differential equation (2.7) and the boundary conditions (2.13)–(2.16). This is accomplished by defining the jump

$$\phi_2(\rho) = \left[\frac{\partial M_2(\rho, z)}{\partial z}\right]_{z=0+} - \left[\frac{\partial M_2(\rho, z)}{\partial z}\right]_{z=0-}. \tag{2.28}$$

Thus from (2.14) and (2.15) it follows that

$$\phi_2(\rho) = 0, \quad a < \rho < \infty,$$
$$\phi_2(\rho) = -Bk^2\rho = -k^2\rho I(b), \quad 0 \le \rho < b. \tag{2.29}$$

Again by applying the usual Green's function approach, we find that the boundary value problem for M_2 is governed by the formula

$$M_2(\rho, z) \sin\varphi = -\frac{1}{4\pi}\int_0^a\int_0^{2\pi} t\phi_2(t)\left\{\frac{e^{ikR}}{R}\right\}_{z_1=0} \sin\varphi_1\, d\varphi_1\, dt + \frac{C}{4\pi}\int_0^b\int_0^{2\pi} t^2\left\{\frac{\partial}{\partial z_1}\left(\frac{e^{ikR}}{R}\right)\right\}_{z_1=0} \sin\varphi_1\, d\varphi_1\, dt. \tag{2.30}$$

By setting $\varphi_1 - \varphi = \psi$, after simplification we obtain

$$M_2(\rho, z) = -\frac{1}{4\pi}\int_0^a\int_0^{2\pi} t\phi_2(t)\left(\frac{e^{ik(\rho^2+t^2-2\rho t\cos\psi+z^2)^{\frac{1}{2}}}}{(\rho^2 + t^2 - 2\rho t\cos\psi + z^2)^{\frac{1}{2}}}\right)\cos\psi\, d\psi\, dt + \frac{C}{4\pi}\int_0^b\int_0^{2\pi} t^2\left\{\frac{\partial}{\partial z_1}\left[\frac{e^{ik[\rho^2+t^2-2\rho t\cos\psi+(z-z_1)^2]^{\frac{1}{2}}}}{(\rho^2 + t^2 - 2\rho t\cos\psi + (z - z_1)^2]^{\frac{1}{2}}}\right]_{z_1=0}\right\}\cos\psi\, d\psi\, dt. \tag{2.31}$$

The boundary conditions (2.13) and (2.29) lead to the Fredholm integral equation of the first kind,

$$f_2(\rho) = \int_b^a t\phi_2(t)K_{21}(t, \rho)\, dt, \quad b \le \rho \le a, \tag{2.32}$$

where

$$K_{21}(t, \rho) = \int_0^\infty \frac{pJ_1(p\rho)J_1(pt)\, dp}{\gamma} \tag{2.33}$$

and

$$f_2(\rho) = -2A\rho + Bk^2\int_0^b t^2\int_0^\infty \frac{p}{\gamma} J_1(p\rho)J_1(pt)\, dp\, dt, \quad b \le \rho \le a. \tag{2.34}$$

It is interesting to note that Eqs. (2.26) and (2.32) do not involve the constant C.

In the next section we present a mathematical technique for solving the integral equations of the type (2.26) and (2.32).

3. MATHEMATICAL TECHNIQUE

Both the integral equations (2.26) and (2.32) are of the form

$$f(\rho) = \int_b^a K_1(t, \rho)g(t)\, dt, \quad b \le \rho \le a, \tag{3.1}$$

where the functions f and K are given and $g(t)$ is to be found. We follow Gubenko and Mossakovskii[7] and set

$$f(\rho) = \sum_{r=-\infty}^{\infty} a_r\rho^r = f_1(\rho) + f_2(\rho), \quad b \le \rho \le a, \tag{3.2}$$

where

$$f_1(\rho) = \sum_{r=0}^{\infty} a_r\rho^r, \quad 0 \le \rho \le a,$$
$$f_2(\rho) = \sum_{r=-\infty}^{-1} a_r\rho^r, \quad b \le \rho < \infty. \tag{3.3}$$

Furthermore, we define two more functions $g_1(\rho)$ and $g_2(\rho)$ such that

$$g_1(\rho) + g_2(\rho) = 0, \quad 0 \le \rho < b,$$
$$= g(\rho), \quad b \le \rho \le a,$$
$$= 0, \quad a < \rho < \infty. \tag{3.4}$$

With the help of the relations (3.2)–(3.4), the integral equation (3.1) becomes equivalent to the pair of integral equations

$$f_1(\rho) = \int_0^\infty g_1(t)K_1(t, \rho)\, dt, \quad 0 < \rho \le a, \tag{3.5}$$

$$f_2(\rho) = \int_0^\infty g_2(t)K_1(t, \rho)\, dt, \quad b \le \rho < \infty. \tag{3.6}$$

The present method is based on perturbing the kernel $K_1(t, \rho)$ on a known suitable kernel $K_0(t, \rho)$:

$$K_1(t, \rho) = K_0(t, \rho) + G(t, \rho), \tag{3.7}$$

where the difference kernel $G(t, \rho)$ is in some sense smaller than the dominant part $K_0(t, \rho)$. Putting (3.7)

in (3.5) and (3.6), we have

$$\int_0^\infty K_0(t,\rho)g_1(t)\,dt = f_1(\rho) - \int_0^\infty G(t,\rho)g_1(t)\,dt, \quad 0 < \rho \leq a, \tag{3.8}$$

$$\int_0^\infty K_0(t,\rho)g_2(t)\,dt = f_2(\rho) - \int_0^\infty G(t,\rho)g_2(t)\,dt, \quad b \leq \rho < \infty. \tag{3.9}$$

The choice of $K_0(t,\rho)$ is such that for all $g(t)$ it satisfies the relation

$$\begin{aligned}\int_0^\infty K_0(t,\rho)g(t)\,dt &= h_{11}(\rho)\int_0^\rho K_2(w,\rho)[h_{12}(w)]^2 \\ &\quad\times\int_w^\infty K_2(w,t)g(t)h_{13}(t)\,dt\,dw, \quad 0<\rho<\infty, \\ &= h_{21}(\rho)\int_\rho^\infty K_2(\rho,w)[h_{22}(w)]^2 \\ &\quad\times\int_0^w K_2(t,w)g(t)h_{23}(t)\,dt\,dw, \quad 0<\rho<\infty,\end{aligned} \tag{3.10}$$

where the functions h_{ij}, $i = 1, 2, j = 1, 2, 3$, and K_2 are known functions. Besides, the kernel $K_2(t,\rho)$ is such that the Volterra equations

$$\int_0^\rho K_2(t,\rho)g(t)\,dt = f(\rho), \quad 0<\rho<\infty, \tag{3.11}$$

$$\int_\rho^\infty K_2(\rho,t)g(t)\,dt = f(\rho), \quad 0<\rho<\infty, \tag{3.12}$$

possess explicit unique solutions for g in terms of f for all arbitrary differentiable functions f. This method also necessitates the introduction of two more functions $L_1(v,w)$ and $L_2(v,w)$ such that

$$\begin{aligned}G(t,\rho) &= h_{11}(\rho)h_{13}(t)\int_0^\rho\int_0^t K_2(w,\rho)K_2(v,t) \\ &\quad\times h_{12}(w)h_{12}(v)L_1(v,w)\,dv\,dw \\ &= h_{21}(\rho)h_{23}(t)\int_\rho^\infty\int_t^\infty K_2(\rho,w)K_2(t,v) \\ &\quad\times h_{22}(w)h_{22}(v)L_2(v,w)\,dv\,dw.\end{aligned} \tag{3.13}$$

The result of substituting (3.10) and (3.13) in the relations (3.8) and (3.9) is the integral equations

$$\begin{aligned}h_{11}(\rho)\int_0^\rho K_2(w,\rho)[h_{12}(w)]^2\int_w^\infty K_2(w,t)g_1(t)h_{13}(t)\,dt\,dw \\ = f_1(\rho) - h_{11}(\rho)\int_0^\rho K_2(w,\rho)h_{12}(w)\int_0^\infty L_1(v,w)h_{12}(v) \\ \times\int_v^\infty K_2(v,t)g_1(t)h_{13}(t)\,dt\,dv\,dw, \quad 0<\rho\leq a,\end{aligned} \tag{3.14}$$

and

$$\begin{aligned}h_{21}(\rho)\int_\rho^\infty K_2(w,\rho)[h_{22}(w)]^2\int_0^w K_2(t,w)g_2(t)h_{23}(t)\,dt\,dw \\ = f_2(\rho) - h_{21}(\rho)\int_\rho^\infty K_2(\rho,w)h_{22}(w)\int_0^\infty L_2(v,w)h_{22}(v) \\ \times\int_0^v K_2(t,v)g_2(t)h_{23}(t)\,dt\,dv\,dw, \quad b\leq\rho<\infty,\end{aligned} \tag{3.15}$$

where we have assumed that various orders of integration can be interchanged.

Let us now define functions $S_1(\rho)$, $S_2(\rho)$, $T_1(\rho)$, $T_2(\rho)$, $C_1(\rho)$, and $C_2(\rho)$ such that

$$\begin{aligned}h_{12}(\rho)\int_\rho^\infty K_2(\rho,t)g_1(t)h_{13}(t)\,dt &= S_1(\rho), \quad 0<\rho\leq a, \\ &= -T_1(\rho), \quad a<\rho<\infty,\end{aligned} \tag{3.16}$$

$$\begin{aligned}h_{22}(\rho)\int_0^\rho K_2(t,\rho)g_2(t)h_{23}(t)\,dt &= -T_2(\rho), \quad 0<\rho<b, \\ &= S_2(\rho), \quad b\leq\rho<\infty,\end{aligned} \tag{3.17}$$

$$h_{11}(\rho)\int_0^\rho K_2(w,\rho)C_1(w)h_{12}(w)\,dw = f_1(\rho), \quad 0<\rho\leq a, \tag{3.18}$$

$$h_{21}(\rho)\int_\rho^\infty K_2(\rho,w)C_2(w)h_{22}(w)\,dw = f_2(\rho), \quad b\leq\rho<\infty. \tag{3.19}$$

The integral equations (3.16)–(3.19) are of the form (3.11) and (3.12), which can be inverted in view of our earlier assumptions. Thus it follows from Eqs. (3.18) and (3.19) that the functions $C_1(\rho)$ and $C_2(\rho)$ are known in terms of $f_1(\rho)$ and $f_2(\rho)$. Moreover, from (3.14), (3.16), and (3.18) we derive the equation

$$\begin{aligned}S_1(\rho) + \int_0^a L_1(v,\rho)S_1(v)\,dv \\ = C_1(\rho) + \int_a^\infty L_1(v,\rho)T_1(v)\,dv, \quad 0<\rho\leq a.\end{aligned} \tag{3.20}$$

Similarly, the result of combining (3.15), (3.17), and (3.19) is

$$\begin{aligned}S_2(\rho) + \int_b^\infty L_2(v,\rho)S_2(v)\,dv \\ = C_2(\rho) + \int_0^b L_2(v,\rho)T_2(v)\,dv, \quad b\leq\rho<\infty.\end{aligned} \tag{3.21}$$

Since we have four unknown functions S_1, S_2, T_1, and T_2, we need two more equations to supplement (3.20) and (3.21). They are readily obtained as follows: From the relations (3.4) and (3.16) we get the integral equation

$$T_1(\rho) = h_{12}(\rho)\int_\rho^\infty K_2(\rho,t)g_2(t)h_{13}(t)\,dt, \quad a<\rho<\infty, \tag{3.22}$$

while the relations (3.4) and (3.17) yield the equation

$$T_2(\rho) = h_{22}(\rho)\int_0^\rho K_2(t, \rho)g_1(t)h_{23}(t)\,dt, \quad 0 < \rho < b. \tag{3.23}$$

Inverting (3.17), we find $g_2(t)$ in terms of T_2 and S_2, and the result of substituting this value in (3.22) is an integral equation containing the unknown functions S_2, T_1, and T_2. Similarly, inverting (3.16) yields $g_1(t)$, which when substituted in (3.23) results in an integral equation containing the unknown functions S_1, T_1, and T_2. Thereby, we have succeeded in deducing a system of four simultaneous integral equations involving four unknown functions S_1, S_2, T_1, and T_2. These equations are the Fredholm integral equations of the second kind and, as such, can be solved by the Neumann iteration scheme. Once we have solved this determinate system, we can recover the values of the functions $g_1(\rho)$ and $g_2(\rho)$ from the integral equations (3.16) and (3.17). Finally, the relation (3.4) gives us the value of the unknown function $g(\rho)$ of the original integral equation.

4. SOLUTION OF THE PROBLEM

Comparing (2.26) with (3.1), we have

$$g(\rho) = \rho\phi_1(\rho), \quad K_{11}(t, \rho) = \int_0^\infty \frac{\gamma J_1(p\rho)J_1(pt)\,dp}{p},$$
$$f_1(\rho) = (A - i/\omega\mu)\rho, \tag{4.1}$$
$$f_{11}(\rho) = (A - i/\omega\mu)\rho, \quad 0 \le \rho \le a,$$
$$f_{12}(\rho) = 0, \qquad b \le \rho < \infty,$$
$$g_1(\rho) = \rho\phi_{11}(\rho), \quad g_2(\rho) = \rho\phi_{12}(\rho), \tag{4.2}$$

where

$$\begin{aligned}\phi_{11}(\rho) + \phi_{12}(\rho) &= 0, && 0 \le \rho < b,\\ &= \phi_1(\rho), && b \le \rho \le a,\\ &= 0, && a < \rho < \infty,\end{aligned} \tag{4.3}$$

and where the subscript 1 has been introduced to signify that we are dealing with the scalar problem for M_1.

We split K_{11} as in (3.7):

$$K_{11}(t, \rho) = K_0(t, \rho) + G_1(t, \rho),$$

where

$$K_0(t, \rho) = \int_0^\infty J_1(p\rho)J_1(pt)\,dp, \tag{4.4}$$

$$G_1(t, \rho) = \int_0^\infty \left(\frac{\gamma}{p} - 1\right)J_1(p\rho)J_1(pt)\,dp. \tag{4.5}$$

When we use the formulas

$$J_n(p\rho) = \left(\frac{2p}{\pi}\right)^{\frac{1}{2}}\frac{1}{\rho^n}\int_0^\rho \frac{J_{n-\frac{1}{2}}(pw)w^{n+\frac{1}{2}}\,dw}{(\rho^2 - w^2)^{\frac{1}{2}}}, \tag{4.6}$$

$$= \left(\frac{2p}{\pi}\right)^{\frac{1}{2}}(\rho)^n\int_\rho^\infty \frac{J_{n+\frac{1}{2}}(pw)w^{-n+\frac{1}{2}}\,dw}{(w^2 - \rho^2)^{\frac{1}{2}}}, \tag{4.7}$$

and

$$\int_0^\infty pJ_{n+\frac{1}{2}}(pw)J_{n+\frac{1}{2}}(pv)\,dp = \frac{\delta(w - v)}{(wv)^{\frac{1}{2}}}, \tag{4.8}$$

with δ being the Dirac delta function, we find from (4.4) that, for all $g(t)$,

$$\begin{aligned}&\int_0^\infty K_0(t, \rho)g(t)\,dt\\ &= \frac{2}{(\pi\rho)}\int_0^\rho \frac{w^2}{(\rho^2 - w^2)^{\frac{1}{2}}}\int_w^\infty \frac{t^{-1}g(t)\,dt\,dw}{(t^2 - w^2)^{\frac{1}{2}}}, \quad 0 < \rho < \infty,\\ &= \frac{2}{\pi}\rho\int_0^\infty \frac{w^{-2}}{(w^2 - \rho^2)^{\frac{1}{2}}}\int_0^w \frac{tg(t)\,dt\,dw}{(w^2 - t^2)^{\frac{1}{2}}}, \qquad 0 < \rho < \infty.\end{aligned} \tag{4.9}$$

Similarly, from (4.5) we have

$$\begin{aligned}G_1(t, \rho) &= \frac{2}{(\pi\rho t)}\int_0^\rho\int_0^t \frac{(wv)^{\frac{3}{2}}}{(\rho^2 - w^2)^{\frac{1}{2}}(t^2 - v^2)^{\frac{1}{2}}}\\ &\quad\times\left(\int_0^\infty p\left(\frac{\gamma}{p} - 1\right)J_{\frac{1}{2}}(pv)J_{\frac{1}{2}}(pw)\,dp\right)dv\,dw\\ &= \frac{(2pt)}{\pi}\int_\rho^\infty\int_t^\infty \frac{(wv)^{-\frac{1}{2}}}{(w^2 - \rho^2)^{\frac{1}{2}}(v^2 - t^2)^{\frac{1}{2}}}\\ &\quad\times\left(\int_0^\infty p\left(\frac{\gamma}{p} - 1\right)J_{\frac{3}{2}}(pv)J_{\frac{3}{2}}(pw)\,dp\right)dv\,dw.\end{aligned} \tag{4.10}$$

Comparing (4.9) and (4.10) with (3.10) and (3.13) gives

$$\begin{aligned}&h_{111}(\rho) = 2/\pi\rho, \quad h_{112}(\rho) = \rho, \quad h_{113}(\rho) = 1/\rho,\\ &h_{121}(\rho) = 2\rho/\pi, \quad h_{122}(\rho) = 1/\rho, \quad h_{123}(\rho) = \rho,\\ &K_2(t, \rho) = (\rho^2 - t^2)^{-\frac{1}{2}},\end{aligned} \tag{4.11}$$

$$L_{11}(v, w) = (vw)^{\frac{1}{2}}\int_0^\infty (\gamma - p)J_{\frac{1}{2}}(pv)J_{\frac{1}{2}}(pw)\,dp, \tag{4.12}$$

$$L_{12}(v, w) = (vw)^{\frac{1}{2}}\int_0^\infty (\gamma - p)J_{\frac{3}{2}}(pv)J_{\frac{3}{2}}(pw)\,dp. \tag{4.13}$$

Since the kernel $K_2 = (\rho^2 - t^2)^{-\frac{1}{2}}$ is a rather simple kernel, the Volterra integral equations (3.11) and (3.12) can be readily inverted. Thus the splitting of the kernel $K_{11}(t, \rho)$ satisfies all the requirements of the previous section. We can, therefore, follow that analysis, and the integral equation (2.26) leads to a set of the following four simultaneous Fredholm

integral equations of the second kind:

$$T_{11}(\rho) = l_{12}(\rho) + \frac{1}{\rho(\pi)^{\frac{1}{2}}\Gamma(\frac{5}{2})} \times \int_0^b \frac{u^2 T_{12}\, {}_2F_1(\frac{1}{2}, 1; \frac{5}{2}; u^2/\rho^2)\, du}{(\rho^2 - u^2)}, \quad a < \rho < \infty, \tag{4.14}$$

$$T_{12}(\rho) = l_{11}(\rho) + \frac{\rho^2}{(\pi)^{\frac{1}{2}}\Gamma(\frac{5}{2})} \times \int_a^\infty \frac{u^{-1} T_{11}(u)\, {}_2F_1(\frac{1}{2}, 1; \frac{5}{2}; \rho^2/u^2)\, du}{(u^2 - \rho^2)}, \quad 0 < k < b, \tag{4.15}$$

$$S_{11}(\rho) + \int_0^a S_{11}(v) L_{11}(v, \rho)\, dv = C_{11}(\rho) + \int_a^\infty T_{11}(v) L_{11}(v, \rho)\, dv, \quad 0 < \rho \leq a, \tag{4.16}$$

$$S_{12}(\rho) + \int_b^\infty S_{12}(v) L_{12}(v, \rho)\, dv = C_{12}(\rho) + \int_0^b T_{12}(v) L_{12}(v, \rho)\, dv, \quad b \leq \rho < \infty, \tag{4.17}$$

where

$$\rho \int_\rho^\infty \frac{\phi_{11}(t)\, dt}{(t^2 - \rho^2)^{\frac{1}{2}}} = S_{11}(\rho), \quad 0 < \rho \leq a, \qquad = -T_{11}(\rho), \quad a < \rho < \infty, \tag{4.18}$$

$$\frac{1}{\rho} \int_0^\rho \frac{t^2 \phi_{12}(t)\, dt}{(\rho^2 - t^2)^{\frac{1}{2}}} = -T_{12}(\rho), \quad 0 < \rho < b, \qquad = S_{12}(\rho), \quad b \leq \rho < \infty, \tag{4.19}$$

$$l_{11}(\rho) = -\frac{2}{\pi\rho} \int_0^\rho \frac{t^2}{(\rho^2 - t^2)^{\frac{1}{2}}} \frac{d}{dt} \times \int_t^a \frac{S_{11}(u)\, du\, dt}{(u^2 - t^2)^{\frac{1}{2}}}, \quad 0 < \rho < b, \tag{4.20}$$

$$l_{12}(\rho) = \frac{2\rho}{\pi} \int_\rho^\infty \frac{t^{-2}}{(t^2 - \rho^2)^{\frac{1}{2}}} \frac{d}{dt} \times \int_b^t \frac{u^2 S_{12}(u)\, du\, dt}{(t^2 - u^2)^{\frac{1}{2}}}, \quad a < \rho < \infty, \tag{4.21}$$

$$C_{11}(\rho) = \frac{1}{\rho} \frac{d}{d\rho} \int_0^\rho \frac{t^2 f_{11}(t)\, dt}{(\rho^2 - t^2)^{\frac{1}{2}}}, \quad 0 < \rho \leq a, \tag{4.22}$$

$$C_{12}(\rho) = -\rho \frac{d}{d\rho} \int_\rho^\infty \frac{f_{12}(t)\, dt}{(t^2 - \rho^2)^{\frac{1}{2}}}, \quad b \leq \rho < \infty, \tag{4.23}$$

and ${}_2F_1$ is a hypergeometric function

$${}_2F_1(\tfrac{1}{2}, 1, \tfrac{5}{2}; x^2/y^2) = (3y/4x^3) \times \{2xy - (y^2 - x^2) \log [(y + x)/(y - x)]\}, \quad x < y.$$

Since $f_{11}(\rho) = (A - i/\omega\mu)\rho$ and $f_{12}(\rho) = 0$, Eqs. (4.22) and (4.23) yield

$$C_{11}(\rho) = 2(A - i/\omega\mu)\rho, \quad C_{12}(\rho) = 0. \tag{4.24}$$

Now, if we can solve the four integral equations (4.14)–(4.17) for S_{11}, S_{12}, T_{11}, and T_{12}, then we can invert the integral equations (4.18) and (4.19) and obtain the values of ϕ_{11} and ϕ_{12}, which in turn give the value of the required function ϕ_1 from the relation (4.3):

$$\phi_1(\rho) = -\frac{2}{\pi} \frac{d}{d\rho} \left(\int_\rho^a \frac{S_{11}(u)\, du}{(u^2 - \rho^2)^{\frac{1}{2}}} - \int_a^\infty \frac{T_{11}(u)\, du}{(u^2 - \rho^2)^{\frac{1}{2}}} \right) + \frac{2}{\pi\rho^2} \frac{d}{d\rho} \left(-\int_\rho^b \frac{u^2 T_{12}(u)\, du}{(\rho^2 - u^2)^{\frac{1}{2}}} + \int_b^\rho \frac{u^2 S_{12}(u)\, du}{(\rho^2 - u^2)^{\frac{1}{2}}} \right), \quad b \leq \rho \leq a. \tag{4.25}$$

The system of integral equations (4.14)–(4.17) is valid for all wavelengths and the ratios b/a. They are specially useful for applying the iteration scheme when the waves have long wavelengths compared with the outer radius a of the annulus and when the inner radius b is much smaller than a. Thus we have the perturbation parameters

$$\alpha = ak, \quad \beta = bk, \quad \lambda = b/a = \beta/\alpha.$$

We shall, therefore, assume in the sequel that $\alpha \ll 1$, $\alpha = O(\lambda)$, and $\beta = \alpha\lambda = O(\alpha^2)$. Fortunately, the kernels of the Eqs. (4.14)–(4.17) tend to zero when λ, α, and β tend to zero. Indeed, by using Noble's contour integration technique,[14] we have

$$L_{11}(v, \rho) = -i(v\rho)^{\frac{1}{2}} \int_0^k (k^2 - \rho^2)^{\frac{1}{2}} H_{\frac{1}{2}}^{(1)}(p\rho) J_{\frac{1}{2}}(pv)\, dp, \quad \rho \geq v, \qquad = -i(v\rho)^{\frac{1}{2}} \int_0^k (k^2 - \rho^2)^{\frac{1}{2}} J_{\frac{1}{2}}(p\rho) H_{\frac{1}{2}}^{(1)}(pv)\, dp, \quad v \geq \rho, \tag{4.26}$$

where $H_r^{(1)}$ is a Hankel function of the first kind. Since the Bessel and Hankel functions of order $\frac{1}{2}$ can be expressed in terms of elementary functions, the formula (4.26) yields the following approximate expansion:

$$L_{11}(v, \rho) = \frac{-k^2 v}{2} - \frac{2i\rho v k^3}{3\pi} + \frac{(v^3 + 3\rho^2 v)k^4}{48} + \frac{2i(\rho^3 v + \rho v^3)k^5}{45\pi} + O(k^6), \quad \rho \geq v, \qquad = \frac{-k^2 \rho}{2} - \frac{2i\rho v k^3}{3\pi} + \frac{(\rho^3 + 3\rho v^2)k^4}{48} + \frac{2i(\rho v^3 + \rho^3 v)k^5}{45\pi} + O(k^6), \quad v \geq \rho. \tag{4.27}$$

Similarly,

$$L_{12}(\rho, v) = -k^2v^2/6\rho + O(k^3), \quad \rho \geq v,$$
$$= -k^2\rho^2/6v + O(k^3), \quad v \geq \rho. \quad (4.28)$$

With this much data in our hands, we proceed to solve the system of integral equations (4.14)–(4.17) and start with (4.16). Setting

$$S_{11}(\rho) = X_{11}(\rho) + W_{11}(\rho), \quad (4.29)$$

we split Eq. (4.16) into two simpler integral equations

$$X_{11}(a\rho) = 2a\left(A - \frac{i}{\omega\mu}\right)\rho - a\int_0^1 X_{11}(av)L_{11}(av, a\rho)\,dv, \quad 0 < \rho \leq 1, \quad (4.30)$$

and

$$W_{11}(a\rho) = a\int_1^\infty L_{11}(av, a\rho)T_{11}(av)\,dv - a\int_0^1 L_{11}(av, a\rho)W_{11}(av)\,dv, \quad 0 < \rho \leq 1. \quad (4.31)$$

Equation (4.30) contains only one unknown function and can be solved by the straightforward iteration scheme. In fact, when we substitute the value of $aL_{11}(av, a\rho)$ from (4.27) in (4.30), an approximate value of $X_{11}(a\rho)$ is obtained as

$$X_{11}(a\rho) = 2a(A - i/\omega\mu) \times [c_{11}(\alpha)\rho + c_{13}(\alpha)\rho^3 + c_{15}(\alpha)\rho^5 + O(\alpha^6)], \quad 0 < \rho \leq 1, \quad (4.32)$$

$$c_{11}(\alpha) = 1 + \frac{\alpha^2}{4} + \frac{2i\alpha^3}{9\pi} + \frac{7\alpha^4}{192} + \frac{41i\alpha^5}{450\pi},$$
$$c_{13}(\alpha) = -\frac{\alpha^2}{12} - \frac{\alpha^4}{32} - \frac{i\alpha^5}{30\pi}, \quad (4.33)$$
$$c_{15}(\alpha) = \frac{\alpha^4}{320}.$$

The unknown functions occurring in the relation (4.14)–(4.21) have to be evaluated in the order

$$X_{11}, l_{11}, T_{12}, S_{12}, l_{12}, T_{11}, W_{11}, S_{11}. \quad (4.34)$$

Having found X_{11}, we proceed to evaluate the other functions in the above sequence and end up with the following approximate values of the unknown functions S_{11}, S_{12}, T_{11}, and T_{12} as

$$S_{11}(a\rho) = X_{11}(a\rho) + O((A - i/\omega\mu)\alpha^2\lambda^5), \quad 0 < \rho \leq 1, \quad (4.35)$$

$$S_{12}(b\rho) = -(4a/45\pi)(A - i/\omega\mu)\alpha^2\lambda^4[\rho^{-1} + O(\alpha^2)], \quad 1 \leq \rho < \infty, \quad (4.36)$$

$$T_{11}(a\rho) = 32a(A - i/\omega\mu)(\lambda^5/45\pi^2)[\rho^{-3} + O(\alpha^2)], \quad 1 < \rho < \infty, \quad (4.37)$$

$$T_{12}(b\rho) = (8a/3\pi)(A - i/\omega\mu)\lambda^2\rho^2 \times \{[c_{11}(\alpha) - c_{13}(\alpha) - \tfrac{1}{3}c_{15}(\alpha)] + \tfrac{2}{5}\lambda^2\rho^2 \times [c_{11}(\alpha) + 3c_{13}(\alpha)] + \tfrac{9}{35}\lambda^4\rho^4 + O(\alpha^5)\}, \quad 0 < \rho < 1. \quad (4.38)$$

Let us now solve the boundary value problem for M_2 as embodied in the integral equation (2.32). The kernel K_{21} is split as

$$K_{21}(t, \rho) = K_0(t, \rho) + G_2(t, \rho), \quad (4.39)$$

where $K_0(t, \rho)$ is given by (4.4) and

$$G_2(t, \rho) = \int_0^\infty \left(\frac{p}{\gamma} - 1\right)J_1(p\rho)J_1(pt)\,dp. \quad (4.40)$$

It can be easily verified that this splitting of the kernel K_{21} of the integral equation (2.32) satisfies all the requirements of the previous section as for (2.26), and, comparing the corresponding expressions in Sec. 3, we have in this case

$$g(\rho) = \rho\phi_2(\rho), \quad h_{211}(\rho) = 2/\pi\rho, \quad h_{212}(\rho) = \rho,$$
$$h_{213}(\rho) = 1/\rho, \quad h_{221}(\rho) = 2\rho/\pi, \quad h_{222}(\rho) = 1/\rho,$$
$$h_{223}(\rho) = \rho, \quad K_2(t, \rho) = (\rho^2 - t^2)^{-\frac{1}{2}},$$
$$g_1(\rho) = \rho\phi_{21}(\rho), \quad g_2(\rho) = \rho\phi_{22}(\rho), \quad (4.41)$$

$$L_{21}(v, w) = (vw)^{\frac{1}{2}}\int_0^\infty p\left(\frac{p}{\gamma} - 1\right)J_{\frac{1}{2}}(pv)J_{\frac{1}{2}}(pw)\,dp, \quad (4.42)$$

$$L_{22}(v, w) = (vw)^{\frac{1}{2}}\int_0^\infty p\left(\frac{p}{\gamma} - 1\right)J_{\frac{3}{2}}(pv)J_{\frac{3}{2}}(pw)\,dp, \quad (4.43)$$

$$\phi_{21}(\rho) + \phi_{22}(\rho) = 0, \quad 0 \leq \rho < b,$$
$$= \phi_2(\rho), \quad b \leq \rho \leq a,$$
$$= 0, \quad a < \rho < \infty. \quad (4.44)$$

The next step requires that we write $f_2(\rho) = f_{21}(\rho) + f_{22}(\rho)$ as in the relation (3.2). To accomplish that, we have to expand the integral occurring in (2.34) in powers of ρ. Now

$$\int_0^b t^2\int_0^\infty \frac{p}{\gamma}J_1(p\rho)J_1(pt)\,dp\,dt = \int_0^b t^2\int_0^\infty J_1(p\rho)J_1(pt)\,dp\,dt + \int_0^b t^2\int_0^\infty \left(\frac{p}{\gamma} - 1\right)J_1(p\rho)J_1(pt)\,dp\,dt. \quad (4.45)$$

Using the relations (4.6) and (4.8), we have

$$\begin{aligned}
&\int_0^b t^2 \int_0^\infty J_1(p\rho) J_1(pt)\, dp\, dt \\
&= \frac{2}{\pi\rho} \int_0^b t \int_0^t \frac{w^2\, dw\, dt}{(\rho^2 - w^2)^{\frac{1}{2}}(t^2 - w^2)^{\frac{1}{2}}} \\
&= \frac{2}{\pi\rho} \int_0^b \frac{w^2}{(\rho^2 - w^2)^{\frac{1}{2}}} \int_w^b \frac{t\, dt\, dw}{(t^2 - w^2)^{\frac{1}{2}}} \\
&= \frac{2}{\pi\rho^2} \int_0^b w^2 (b^2 - w^2)^{\frac{1}{2}} \left(1 - \frac{w^2}{\rho^2}\right)^{-\frac{1}{2}} dw \\
&= \frac{2}{\pi\rho^2} \int_0^b w^2 (b^2 - w^2)^{\frac{1}{2}} \left(1 + \sum_{n=1}^{\infty} A_n \frac{w^{2n}}{\rho^{2n}}\right) dw \\
&= \frac{b^4}{8\rho^2}\left(1 + 2\sum_{n=1}^{\infty} \frac{b^{2n}}{\rho^{2n}} \frac{(2n+1)A_n^2}{(n+1)(n+2)}\right), \quad b \le \rho \le a,
\end{aligned} \tag{4.46}$$

where

$$A_n = \frac{1 \times 3 \times 5 \times \cdots \times (2n-1)}{2 \times 4 \times 6 \times \cdots \times (2n)}. \tag{4.47}$$

Similarly,

$$\begin{aligned}
&\int_0^b t^2 \int_0^\infty \left(\frac{p}{\gamma} - 1\right) J_1(p\rho) J_1(pt)\, dp\, dt \\
&= \int_0^b t^2 \int_0^\infty \left(\frac{p}{\gamma} - 1\right) \frac{2p}{\pi pt} \\
&\quad \times \int_0^\rho \int_0^t \frac{J_{\frac{1}{2}}(pw) J_{\frac{1}{2}}(pv)(wv)^{\frac{3}{2}}}{(\rho^2 - w^2)^{\frac{1}{2}}(t^2 - v^2)^{\frac{1}{2}}}\, dv\, dw\, dp\, dt \\
&= \frac{2}{\pi\rho} \int_0^b t \int_0^\rho \int_0^t \frac{wv L_{21}(v, w)\, dv\, dw\, dt}{(\rho^2 - w^2)^{\frac{1}{2}}(t^2 - v^2)^{\frac{1}{2}}} \\
&= \frac{2}{\pi\rho} \int_0^\rho \frac{w}{(\rho^2 - w^2)^{\frac{1}{2}}} \int_0^b v L_{21}(v, w) \int_v^b \frac{t\, dt\, dv\, dw}{(t^2 - v^2)^{\frac{1}{2}}} \\
&= \frac{2}{\pi\rho} \int_0^\rho \frac{w}{(\rho^2 - w^2)^{\frac{1}{2}}} \int_0^b v L_{21}(v, w)(b^2 - v^2)^{\frac{1}{2}}\, dv\, dw \\
&= \frac{2}{\pi\rho} \left\{ \int_0^b v(b^2 - v^2)^{\frac{1}{2}} \int_v^\rho \frac{w L_{21}(v, w)\, dw\, dv}{(\rho^2 - w^2)^{\frac{1}{2}}} \right. \\
&\quad + \int_0^b \frac{w}{(\rho^2 - w^2)^{\frac{1}{2}}} \\
&\quad \left. \times \int_w^b v L_{21}(v, w)(b^2 - v^2)^{\frac{1}{2}}\, dv\, dw \right\}, \quad b \le \rho \le a.
\end{aligned} \tag{4.48}$$

The kernel $L_{21}(v, w)$ can be expanded in powers of k by applying Noble's contour integration technique[14] and we have

$$\begin{aligned}
L_{21}(v, w) &= \begin{cases} i(vw)^{\frac{1}{2}} \int_0^k \dfrac{p^2}{(k^2 - p^2)^{\frac{1}{2}}} \\ \quad \times H_{\frac{1}{2}}^{(1)}(pw) J_{\frac{1}{2}}(pv)\, dp, & w \ge v \\ i(vw)^{\frac{1}{2}} \int_0^k \dfrac{p^2}{(k^2 - p^2)^{\frac{1}{2}}} \\ \quad \times J_{\frac{1}{2}}(pw) H_{\frac{1}{2}}^{(1)}(pv)\, dp, & v \ge w \end{cases} \\
&= \begin{cases} \dfrac{k^2 v}{2} + \dfrac{4ik^3 vw}{3\pi} - \dfrac{k^4}{16}(3w^2 v + v^3) \\ \quad - \dfrac{8ik^5}{45\pi}(wv^3 + w^3 v) + O(k^6), & w \ge v \\ \dfrac{k^2 w}{2} + \dfrac{4ik^3 vw}{3\pi} - \dfrac{k^4}{16}(3wv^2 + w^3) \\ \quad - \dfrac{8ik^5}{45\pi}(wv^3 + w^3 v) + O(k^6), & v \ge w \end{cases}
\end{aligned} \tag{4.49}$$

Substituting it in (4.48), we derive

$$\begin{aligned}
&\int_0^b t^2 \int_0^\infty \left(\frac{p}{\gamma} - 1\right) J_1(p\rho) J_1(pt)\, dp\, dt \\
&= \frac{k^2}{\pi\rho}\left(\int_0^b v(b^2 - v^2)^{\frac{1}{2}} \int_v^\rho \frac{vw\, dw\, dv}{(\rho^2 - w^2)^{\frac{1}{2}}} \right. \\
&\quad \left. + \int_0^b \frac{w}{(\rho^2 - w^2)^{\frac{1}{2}}} \int_w^b wv(b^2 - v^2)^{\frac{1}{2}}\, dv\, dw \right) \\
&\quad + \frac{8ik^3}{3\pi^2\rho}\left(\int_0^\rho \frac{w^2\, dw}{(\rho^2 - w^2)^{\frac{1}{2}}} \right) \\
&\quad \times \left(\int_0^b v^2(b^2 - v^2)^{\frac{1}{2}}\, dv \right) + O(k^4) \\
&= \frac{k^2}{\pi\rho}\left(\int_0^b v^2(b^2 - v^2)^{\frac{1}{2}}(\rho^2 - v^2)^{\frac{1}{2}}\, dv + \tfrac{1}{3} \int_0^b \frac{w^2}{(\rho^2 - w^2)^{\frac{1}{2}}} \right. \\
&\quad \left. \times (b^2 - w^2)^{\frac{3}{2}}\, dw \right) + \tfrac{1}{24} ik^3 b^4 \rho + O(k^4) \\
&= (k^2/\pi\rho)\{\rho[\tfrac{1}{16} b^4 \pi + O(b^6/\rho^2)] + (1/3\rho)[O(b^6)]\} \\
&\quad + \tfrac{1}{24} ik^3 b^4 \rho + O(k^4), \quad b \le \rho \le a.
\end{aligned} \tag{4.50}$$

Substituting the values of integrals from Eqs. (4.46) and (4.50) in (4.45), we readily obtain from (2.34) the values of the functions f_{21} and f_{22} in terms of the dimensionless parameters α and λ in the form

$$f_{21}(a\rho) = -2aA\rho + B[\alpha^4\lambda^4(\tfrac{1}{16} + \tfrac{1}{24} i\alpha\rho) + O(\alpha^{10})], \quad 0 \le \rho \le 1, \tag{4.51}$$

$$\begin{aligned}
&f_{22}(b\rho) \\
&= B\left[\frac{\alpha^2\lambda^2}{8\rho^2}\left(1 + 2\sum_{1}^{\infty} \frac{(2n+1)A_n^2}{\rho^{2n}(n+1)(n+2)}\right) + O(\alpha^4\lambda^4)\right], \\
&\qquad 1 \le \rho < \infty.
\end{aligned} \tag{4.52}$$

We are now ready to apply the technique of the last section for reducing the Fredholm integral equation of the first kind [(2.32)] to a set of four simultaneous integral equations of the second kind for four unknown functions S_{21}, S_{22}, T_{21}, and T_{22}. These equations are

$$T_{21}(\rho) = l_{22}(\rho) + \frac{1}{\rho(\pi)^{\frac{1}{2}}\Gamma(\frac{5}{2})} \times \int_0^b \frac{u^2 T_{22}(u)\, {}_2F_1(\frac{1}{2}, 1; \frac{5}{2}; u^2/\rho^2)\, du}{(\rho^2 - u^2)}, \quad a < \rho < \infty, \tag{4.53}$$

$$T_{22}(\rho) = l_{21}(\rho) + \frac{\rho^2}{(\pi)^{\frac{1}{2}}\Gamma(\frac{5}{2})} \times \int_a^\infty \frac{T_{21}(u)\, {}_2F_1(\frac{1}{2}, 1; \frac{5}{2}; \rho^2/u^2)\, du}{u(u^2 - \rho^2)}, \quad 0 < \rho < b, \tag{4.54}$$

$$S_{21}(\rho) + \int_0^a L_{21}(v, \rho) S_{21}(v)\, dv = C_{21}(\rho) + \int_a^\infty L_{21}(v, \rho) T_{21}(v)\, dv, \quad 0 < \rho \le a, \tag{4.55}$$

$$S_{22}(\rho) + \int_b^\infty L_{22}(v, \rho) S_{22}(v)\, dv = C_{22}(\rho) + \int_0^b L_{22}(v, \rho) T_{22}(v)\, dv, \quad b \le \rho < \infty, \tag{4.56}$$

where

$$l_{21}(\rho) = -\frac{2}{\pi\rho} \int_0^\rho \frac{t^2}{(\rho^2 - t^2)^{\frac{1}{2}}} \frac{d}{dt} \times \int_t^a \frac{S_{21}(u)\, du\, dt}{(u^2 - t^2)^{\frac{1}{2}}}, \quad 0 < \rho < b, \tag{4.57}$$

$$l_{22}(\rho) = \frac{2\rho}{\pi} \int_\rho^\infty \frac{1}{t^2(t^2 - \rho^2)^{\frac{1}{2}}} \frac{d}{dt} \times \int_b^t \frac{u^2 S_{22}(u)\, du\, dt}{(t^2 - u^2)^{\frac{1}{2}}}, \quad a < \rho < \infty, \tag{4.58}$$

$$\rho \int_\rho^\infty \frac{\phi_{21}(t)\, dt}{(t^2 - \rho^2)^{\frac{1}{2}}} = \begin{cases} S_{21}(\rho), & 0 < \rho \le a \\ -T_{21}(\rho), & a < \rho < \infty \end{cases}, \tag{4.59}$$

$$\rho^{-1} \int_0^\rho \frac{t^2 \phi_{22}(t)\, dt}{(\rho^2 - t^2)^{\frac{1}{2}}} = \begin{cases} -T_{22}(\rho), & 0 < \rho < b \\ S_{22}(\rho), & b \le \rho < \infty \end{cases}, \tag{4.60}$$

$$C_{21}(\rho) = \frac{1}{\rho} \frac{d}{d\rho} \int_0^\rho \frac{t^2 f_{21}(t)\, dt}{(\rho^2 - t^2)^{\frac{1}{2}}}, \quad 0 < \rho \le a, \tag{4.61}$$

$$C_{22}(\rho) = -\rho \frac{d}{d\rho} \int_\rho^\infty \frac{f_{22}(t)\, dt}{(t^2 - \rho^2)^{\frac{1}{2}}}, \quad b \le \rho < \infty, \tag{4.62}$$

and L_{21} and L_{22} are defined by the relations (4.42) and (4.43). We have already expanded L_{21} in powers of k in Eq. (4.49). Similarly,

$$L_{22}(v, \rho) = (k^2\rho^2/6v) + O(k^3), \quad v \ge \rho, \qquad = (k^2 v^2/6\rho) + O(k^3), \quad \rho \ge v. \tag{4.63}$$

From the relations (4.44), (4.59), and (4.60), it follows that

$$\phi_2(\rho) = -\frac{2}{\pi}\frac{d}{d\rho}\left(\int_\rho^a \frac{S_{21}(u)\, du}{(u^2 - \rho^2)^{\frac{1}{2}}} - \int_a^\infty \frac{T_{21}(u)\, du}{(u^2 - \rho^2)^{\frac{1}{2}}}\right) + \frac{2}{\pi\rho^2}\frac{d}{d\rho}\left(-\int_0^b \frac{u^2 T_{22}(u)\, du}{(\rho^2 - u^2)^{\frac{1}{2}}} + \int_b^\rho \frac{u^2 S_{22}(u)\, du}{(\rho^2 - u^2)^{\frac{1}{2}}}\right), \quad b \le \rho \le a. \tag{4.64}$$

Putting the values of f_{21} and f_{22} from (4.51) and (4.52) into (4.61) and (4.62), we have

$$C_{21}(a\rho) = -4aA\rho + B[\alpha^4\lambda^4(\tfrac{1}{3\cdot 2}\pi + \tfrac{1}{1\cdot 2} i\alpha\rho) + O(\alpha^{10})], \quad 0 < \rho \le 1, \tag{4.65}$$

$$C_{22}(b\rho) = B\left[\frac{\alpha^2\lambda^2}{4}\left(\frac{1}{\rho^2} + 2\sum_1^\infty \frac{A_n}{\rho^{2n+2}(n+2)}\right) + O(\alpha^4\lambda^4)\right], \quad 1 \le \rho < \infty. \tag{4.66}$$

The steps of solving the system of Eqs. (4.53)–(4.56) are the same as for the system (4.14)–(4.17). First, we set

$$S_{21}(\rho) = X_{21}(\rho) + W_{21}(\rho), \tag{4.67}$$

which splits the Eq. (4.55) into two equations:

$$X_{21}(a\rho) = C_{21}(a\rho) - a\int_0^1 L_{21}(av, a\rho) X_{21}(av)\, dv, \quad 0 < \rho \le 1, \tag{4.68}$$

$$W_{21}(a\rho) = a\int_1^\infty L_{21}(av, a\rho) T_{21}(av)\, dv - a\int_0^1 L_{21}(av, a\rho) W_{21}(av)\, dv, \quad 0 < \rho \le 1. \tag{4.69}$$

Equation (4.68) is easily solved to yield

$$X_{21}(a\rho) = -4Aa[c_{21}(\alpha)\rho + c_{23}(\alpha)\rho^3 + c_{25}(\alpha)\rho^5 + O(\alpha^6)] + B\alpha^4\lambda^4[\tfrac{1}{3\cdot 2}\pi + \tfrac{1}{1\cdot 2} i\alpha\rho + O(\alpha^2)], \tag{4.70}$$

where

$$c_{21}(\alpha) = 1 - \frac{\alpha^2}{4} - \frac{4i\alpha^3}{9\pi} + \frac{19\alpha^4}{192} + \frac{53i\alpha^5}{225\pi}, \qquad c_{23}(\alpha) = \frac{\alpha^2}{12} + \frac{\alpha^4}{96} + \frac{i\alpha^5}{45\pi}, \quad c_{25}(\alpha) = \frac{-\alpha^4}{960}. \tag{4.71}$$

The next step is to proceed in the order as given by the sequence (4.34) with first subscript changed to 2.

Thereby, the required solutions for S_{21}, S_{22}, T_{21}, and T_{22} are

$$S_{21}(a\rho) = X_{21}(a\rho) + O(\alpha^2\lambda^5 A) + O(\alpha^4\lambda^5 B), \quad 0 < \rho \leq 1, \quad (4.72)$$

$$S_{22}(b\rho) = -\frac{8A(a\alpha^2\lambda^4)}{45\pi}\left(\frac{1}{\rho} + O(\alpha^2)\right) + B\left[\frac{\alpha^2\lambda^2}{4} \times \left(\frac{1}{\rho^2} + 2\sum_1^\infty \frac{A_n}{\rho^{2n+2}(n+2)}\right) + O(\alpha^4\lambda^4)\right]$$
$$= -[8A(a\alpha^2\lambda^4)/45\pi][\rho^{-1} + O(\alpha^2)] + B\{\alpha^2\lambda^2[-(1-\rho^{-2})^{\frac{1}{2}} - \tfrac{2}{3}\rho^2(1-\rho^{-2})^{\frac{3}{2}} + \tfrac{2}{3}\rho^2] + O(\alpha^4\lambda^4)\}, \quad 1 \leq \rho < \infty, \quad (4.73)$$

$$T_{21}(a\rho) = -(64Aa\lambda^5/45\pi^2)[\rho^{-3} + O(\alpha^2)] + (8B\alpha^2\lambda^5/45\pi)[\rho^{-3} + O(\lambda^2)], \quad 1 < \rho < \infty, \quad (4.74)$$

$$T_{22}(b\rho) = -(16Aa\lambda^2\rho^2/3\pi)\{[c_{21}(\alpha) - c_{23}(\alpha) - \tfrac{1}{3}c_{25}(\alpha)] + \tfrac{2}{5}\lambda^2\rho^2[c_{21}(\alpha) + 3c_{23}(\alpha)] + \tfrac{9}{35}\lambda^4\rho^4 + O(\alpha^5)\} + O(\alpha^8 B), \quad 0 < \rho < 1. \quad (4.75)$$

5. EVALUATION OF THE CONSTANTS A, B, AND C

We found in Sec. 2 that the continuity of $[M, (\rho, z)]_{z=\pm 0}$ at the inner edge $\rho = b$ helped us to get the relation $B = I(b)$. We shall now evaluate A with the help of the outer edge condition and then evaluate B from the solution. To accomplish that, let us denote the total induced surface current density at the disk by $\mathbf{j}(\rho, \varphi, 0) = (j_\rho, j_\varphi, 0)$. Then the outer edge condition demands that

$$j_\rho(\rho, \varphi, 0) = O((a^2 - \rho^2)^{\frac{1}{2}}) \quad \text{as} \quad \rho \to a, \quad (5.1)$$

$$j_\varphi(\rho, \varphi, 0) = O((a^2 - \rho^2)^{-\frac{1}{2}}) \quad \text{as} \quad \rho \to a. \quad (5.2)$$

Now at the disk

$$j_\rho(\rho, \varphi, 0) = [H_\varphi^{(1)}]_{z=0-} - [H_\varphi^{(1)}]_{z=0+} = -\cos\varphi\{\rho^{-1}[\phi_1(\rho) + \phi_2(\rho)] + k^2 I(\rho)\}, \quad b \leq \rho \leq a, \quad (5.3)$$

$$j_\varphi(\rho, \varphi, 0) = [H_\rho^{(1)}]_{z=0+} - [H_\rho^{(1)}]_{z=0-} = \sin\varphi[\phi_1'(\rho) + \phi_2'(\rho) + k^2 I(\rho)], \quad b \leq \rho \leq a. \quad (5.4)$$

The behavior of the function $\phi_1(\rho)$ at $\rho = a$ is easily determined if we appeal to the relations (4.3), (4.18), (4.19), and (4.25). We get (see Appendix)

$$\phi_1(\rho) = \frac{2}{\pi}\left(\frac{S_{11}(a) + T_{11}(a)}{(a^2 - \rho^2)^{\frac{1}{2}}} + O((a^2 - \rho^2)^{\frac{1}{2}})\right) \quad \text{as} \quad \rho \to a. \quad (5.5)$$

Similarly,

$$\phi_2(\rho) = \frac{2}{\pi}\left(\frac{S_{21}(a) + T_{21}(a)}{(a^2 - \rho^2)^{\frac{1}{2}}} + O((a^2 - \rho^2)^{\frac{1}{2}})\right) \quad \text{as} \quad \rho \to a. \quad (5.6)$$

It therefore follows from (5.3)–(5.6) that the edge conditions (5.1) and (5.2) are satisfied if

$$S_{11}(a) + T_{11}(a) + S_{21}(a) + T_{21}(a) = 0, \quad (5.7)$$

since we have already made use of relation $I(a) = 0$ to insure the uniqueness of the solution $I(\rho)$ of the differential equation (2.25). Substituting the values of the functions S_{11}, T_{11}, S_{21}, and T_{21} from the last section in (5.7), we obtain

$$2a(A - i/\omega\mu)[c_{11}(\alpha) + c_{13}(\alpha) + c_{15}(\alpha) + O(\alpha^6)] + 32a(A - i/\omega\mu)(\lambda^5/45\pi^2)[1 + O(\alpha^2)] = 4Aa[c_{21}(\alpha) + c_{23}(\alpha) + c_{25}(\alpha) + O(\alpha^6)] - B\alpha^4\lambda^4[\tfrac{1}{3}\tfrac{1}{2}\pi + \tfrac{1}{1}\tfrac{1}{2}i\alpha + O(\alpha^2)] + (64Aa\lambda^5/45\pi^2) \times [1 + O(\alpha^2)] - (8B\alpha^2\lambda^5/45\pi)[1 + O(\lambda^2)], \quad (5.8)$$

where $B = I(b)$.

Now we explain the procedure to be followed for determining the unknown constants A and B. Substituting the values of S_{11}, S_{12}, T_{11}, and T_{12} obtained in the last section into (4.25), we determine the value of ϕ_1 in terms of the unknown constant A. Substituting this value of ϕ_1 in (2.25), we solve it to find the solution $I(\rho)$ in terms of the unknown constant A. From this expression of $I(\rho)$, we obtain $I(b)$ in terms of A. Thus we have determined the value of B in terms of A, which when substituted in (5.8) gives the required value of the unknown constant A, and thereby $I(\rho)$ is completely determined. Putting $\rho = b$ in this value of $I(\rho)$, we also obtain the value of the second constant $B = I(b)$.

Following the above procedure, we obtain

$$A = \frac{-i}{\omega\mu}\left(1 + \frac{2\alpha^2}{3} + \frac{4i\alpha^3}{3\pi} + \frac{2\alpha^4}{15} + \frac{136i\alpha^5}{135\pi} + O(\alpha^6)\right), \quad (5.9)$$

$$B = \frac{8ia}{\pi\omega\mu}\left[1 + \frac{5\alpha^2}{9} - \frac{\lambda^2}{3} + O(\alpha^3)\right]. \quad (5.10)$$

Having determined both the constants A and B, we can write down the approximate expansions of X_{i1}, S_{ij}, and Tij, $i = 1, 2, j = 1, 2$, from (4.32), (4.35)–(4.38), (4.70), and (4.72)–(4.75):

$$X_{11}(a\rho) = -(4ai/\omega\mu) \times [d_{11}(\alpha)\rho + d_{13}(\alpha)\rho^3 + d_{15}(\alpha)\rho^5 + O(\alpha^6)], \quad 0 < \rho \leq 1, \quad (5.11)$$

where

$$d_{11}(\alpha) = 1 + \frac{7\alpha^2}{12} + \frac{8i\alpha^2}{9\pi} + \frac{179\alpha^4}{960} + \frac{188i\alpha^5}{225\pi},$$
$$d_{13}(\alpha) = -\frac{\alpha^2}{12} - \frac{17\alpha^4}{288} - \frac{4i\alpha^5}{45\pi}, \quad (5.12)$$
$$d_{15}(\alpha) = \frac{\alpha^4}{320},$$

$$S_{11}(a\rho) = X_{11}(a\rho) + O(\alpha^7), \quad 0 < \rho \leq 1, \quad (5.13)$$

$$S_{12}(b\rho) = (8ai\alpha^2\lambda^4/45\pi\omega\mu)[\rho^{-1} + O(\alpha^2)], \quad 1 \leq \rho < \infty, \quad (5.14)$$

$$T_{11}(a\rho) = -(64ia\lambda^5/45\pi^2\mu\omega)[\rho^{-3} + O(\alpha^2)], \quad 1 < \rho < \infty, \quad (5.15)$$

$$T_{12}(b\rho) = -(16ia/3\pi\omega\mu) \times \{\lambda^2\rho^2[d_{11}(\alpha) - d_{13}(\alpha) - \tfrac{1}{3}d_{15}(\alpha)] + \tfrac{2}{5}\lambda^4\rho^4[d_{11}(\alpha) + 3d_{13}(\alpha)] + \tfrac{9}{35}\lambda^6\rho^6 + O(\alpha^7)\}, \quad 0 < \rho < 1, \quad (5.16)$$

$$X_{21}(a\rho) = (4ia/\omega\mu) \times [d_{21}(\alpha)\rho + d_{23}(\alpha)\rho^3 + d_{25}(\alpha)\rho^5 + O(\alpha^6)], \quad 0 < \rho \leq 1, \quad (5.17)$$

with

$$d_{21}(\alpha) = 1 + \frac{5\alpha^2}{12} + \frac{8i\alpha^3}{9\pi} + \frac{21\alpha^4}{320} + \frac{138i\alpha^5}{225\pi},$$
$$d_{23}(\alpha) = \frac{\alpha^2}{12} + \frac{19\alpha^4}{288} + \frac{2i\alpha^5}{15\pi}, \quad d_{25}(\alpha) = -\frac{\alpha^4}{960}, \quad (5.18)$$

$$S_{21}(a\rho) = X_{21}(a\rho) + O(\alpha^7), \quad 0 < \rho \leq 1, \quad (5.19)$$

$$S_{22}(b\rho) = (8ia\alpha^2\lambda^4/45\pi\omega\mu)[\rho^{-1} + O(\alpha^2)] + (8ia/\pi\omega\mu)\{\alpha^2\lambda^2[1 + \tfrac{5}{9}\alpha^2 - \tfrac{1}{3}\lambda^2 + O(\alpha^3)] \times [-(1-\rho^{-2})^{\frac{1}{2}} - \tfrac{2}{5}\rho^2(1-\rho^{-2})^{\frac{3}{2}} + \tfrac{2}{5}\rho^2] + O(\alpha^4\lambda^4)\}, \quad 1 \leq \rho < \infty, \quad (5.20)$$

$$T_{21}(a\rho) = (64ia\lambda^5/45\pi^2\omega\mu)[\rho^{-3} + O(\alpha^2)], \quad 1 < \rho < \infty, \quad (5.21)$$

$$T_{22}(b\rho) = (16ia/3\pi\omega\mu) \times \{\lambda^2\rho^2[d_{21}(\alpha) - d_{23}(\alpha) - \tfrac{1}{3}d_{25}(\alpha)] + \tfrac{2}{5}\lambda^4\rho^4[d_{11}(\alpha) + 3d_{13}(\alpha)] + \tfrac{9}{35}\lambda^6\rho^6 + O(\alpha^7)\}, \quad 0 < \rho < 1. \quad (5.22)$$

Substituting these values in (4.25) and (4.64), we obtain

$$\phi_1(\rho) = \frac{8i}{\pi\omega\mu}\left(\left(\frac{-\rho/a}{(1-\rho^2/a^2)^{\frac{1}{2}}}\right)\right) \times \left[\left(1 + \frac{\alpha^2}{2} + \frac{8i\alpha^3}{9\pi} + \frac{47\alpha^4}{360} + \frac{168i\alpha^5}{225\pi}\right) + \left(1 - \frac{\rho^2}{a^2}\right)\left(\frac{\alpha^2}{6} + \frac{19\alpha^4}{180} + \frac{8i\alpha^5}{45\pi}\right) + \left(1 - \frac{\rho^2}{a^2}\right)^2\frac{\alpha^4}{120} + O(\alpha^6)\right] + \frac{8\lambda^5}{45\pi^2}\frac{a^3}{\rho^3} \times \left[\frac{3a}{\rho}\sin^{-1}\frac{\rho}{a} - \left(1 - \frac{\rho^2}{a^2}\right)^{\frac{1}{2}} - \frac{2}{(1-\rho^2/a^2)^{\frac{1}{2}}}\right] + \frac{4}{3\pi}\left\{e_{12}\frac{\lambda}{2}\left(\frac{3\rho}{b}\sin^{-1}\frac{b}{\rho} - \left(1 - \frac{b^2}{\rho^2}\right)^{\frac{1}{2}} - \frac{2}{(1-b^2/\rho^2)^{\frac{1}{2}}}\right) + e_{14}\frac{\lambda^3}{8}\frac{\rho^2}{b^2}\left[\frac{15\rho}{b}\sin^{-1}\frac{b}{\rho} + 2\left(1 - \frac{b^2}{\rho^2}\right)^{\frac{3}{2}} - 9\left(1 - \frac{b^2}{\rho^2}\right)^{\frac{1}{2}} - \frac{8}{(1-b^2/\rho^2)^{\frac{1}{2}}}\right] + e_{16}\frac{\lambda^5\rho^4}{48b^4} \times \left[105\frac{\rho}{b}\sin^{-1}\frac{b}{\rho} - 8\left(1 - \frac{b^2}{\rho^2}\right)^{\frac{5}{2}} + 38\left(1 - \frac{b^2}{\rho^2}\right)^{\frac{3}{2}} - 87\left(1 - \frac{b^2}{\rho^2}\right)^{\frac{1}{2}} - \frac{48}{(1-b^2/\rho^2)^{\frac{1}{2}}}\right]\right\} + \frac{2\alpha^2\lambda^4 ab}{45\pi\rho^2(1-b^2/\rho^2)^{\frac{1}{2}}} + O(\alpha^6)\Bigg), \quad (5.23)$$

where

$$e_{12} = d_{11} - d_{13} - \tfrac{1}{3}d_{15} = 1 + \tfrac{2}{3}\alpha^2 + (8i\alpha^3/9\pi) + \tfrac{11}{45}\alpha^4 + 208i\alpha^5/225\pi,$$
$$e_{14} = \tfrac{2}{5}(d_{11} + 3d_{13}) = \tfrac{2}{5}(1 + \tfrac{1}{3}\alpha^2 + 8i\alpha^3/9\pi),$$
$$e_{16} = \tfrac{9}{35} \quad (5.24)$$

and

$$\phi_2(\rho) = \frac{8i}{\pi\omega\mu}\left(\left(\frac{\rho/a}{(1-\rho^2/a^2)^{\frac{1}{2}}}\right)\right) \times \left[\left(1 + \frac{\alpha^2}{2} + \frac{8i\alpha^3}{9\pi} + \frac{47\alpha^4}{360} + \frac{168i\alpha^5}{225\pi}\right) + \left(1 - \frac{\rho^2}{a^2}\right)\left(-\frac{\alpha^2}{6} - \frac{23\alpha^4}{180} - \frac{4i\alpha^5}{15\pi}\right) + \left(1 - \frac{\rho^2}{a^2}\right)^2\left(\frac{-\alpha^4}{360}\right) + O(\alpha^6)\right] + \frac{8\lambda^5}{45\pi^2}\frac{a^3}{\rho^3} \times \left[-\frac{3a}{\rho}\sin^{-1}\frac{\rho}{a} + \left(1 - \frac{\rho^2}{a^2}\right)^{\frac{1}{2}} + \frac{2}{(1-\rho^2/a^2)^{\frac{1}{2}}}\right] - \frac{4}{3\pi}\left\{e_{22}\frac{\lambda}{2}\left[\frac{3\rho}{b}\sin^{-1}\frac{b}{\rho} - \left(1 - \frac{b^2}{\rho^2}\right)^{\frac{1}{2}} - \frac{2}{(1-b^2/\rho^2)^{\frac{1}{2}}}\right] + e_{24}\frac{\lambda^3}{8}\frac{\rho^2}{b^2}\left[15\frac{\rho}{b}\sin^{-1}\frac{b}{\rho} + 2\left(1 - \frac{b^2}{\rho^2}\right)^{\frac{3}{2}} - 9\left(1 - \frac{b^2}{\rho^2}\right)^{\frac{1}{2}} - \frac{8}{(1-b^2/\rho^2)^{\frac{1}{2}}}\right] + e_{26}\frac{\lambda^5}{48}\frac{\rho^4}{b^4}$$

$$\times\left[105\frac{\rho}{b}\sin^{-1}\frac{b}{\rho}-8\left(1-\frac{b^2}{\rho^2}\right)^{\frac{5}{2}}+38\left(1-\frac{b^2}{\rho^2}\right)^{\frac{3}{2}}\right.$$
$$\left.-87\left(1-\frac{b^2}{\rho^2}\right)^{\frac{1}{2}}-\frac{48}{(1-b^2/\rho^2)^{\frac{1}{2}}}\right]\Bigg\}$$
$$-\frac{2\alpha^2\lambda}{3\pi}\left[1+\frac{5\alpha^2}{9}-\frac{\lambda^2}{3}\right]\left[\frac{3\rho}{b}\sin^{-1}\frac{b}{\rho}-\left(1-\frac{b^2}{\rho^2}\right)^{\frac{1}{2}}\right.$$
$$\left.-\frac{2}{(1-b^2/\rho^2)^{\frac{1}{2}}}\right]+\frac{2\alpha^2\lambda^4 ab}{45\pi\rho^2(1-b^2/\rho^2)^{\frac{1}{2}}}+O(\alpha^6)\Bigg), \tag{5.25}$$

where

$$e_{22}=1+\tfrac{1}{3}\alpha^2+(8i\alpha^3/9\pi)+108i\alpha^5/225\pi,$$
$$e_{24}=\tfrac{2}{5}(1+\tfrac{2}{3}\alpha^2+8i\alpha^3/9\pi),\quad e_{26}=\tfrac{9}{35}. \tag{5.26}$$

Finally, substituting the value of $\phi_1(\rho)$ from (5.23) into the system (2.25), we obtain $I(\rho)$, given by

$$I(\rho)=\frac{8ia}{\pi\omega\mu}\Bigg(\left[\left(1-\frac{\rho^2}{a^2}\right)^{\frac{1}{2}}\right]$$
$$\times\left[\left(1+\frac{\alpha^2}{2}+\frac{8i\alpha^3}{9\pi}+\frac{47\alpha^4}{360}+\frac{168i\alpha^5}{225\pi}\right)\right.$$
$$+\frac{1}{3}\left(1-\frac{\rho^2}{a^2}\right)\left(\frac{\alpha^2}{6}+\frac{19}{180}\alpha^4+\frac{8i\alpha^5}{45\pi}\right)$$
$$\left.+\left(1-\frac{\rho^2}{a^2}\right)^2\frac{\alpha^4}{600}+O(\alpha^6)\right]$$
$$-\frac{8\lambda^5}{45\pi^2}\frac{a^3}{\rho^3}\left[\sin^{-1}\frac{\rho}{a}-\frac{\rho}{a}\left(1-\frac{\rho^2}{a^2}\right)^{\frac{1}{2}}\right]$$
$$+\frac{4}{3\pi}\left\{e_{12}\frac{\lambda^2}{2}\left[\frac{3\rho^2}{2b^2}\sin^{-1}\frac{b}{\rho}\right.\right.$$
$$\left.-\frac{3}{2}\left(\frac{\rho^2}{b^2}-1\right)^{\frac{1}{2}}-\sin^{-1}\frac{b}{\rho}\right]$$
$$+e_{14}\frac{\lambda^4}{8}\left[\frac{15}{4}\frac{\rho^4}{b^4}\sin^{-1}\frac{b}{\rho}-\frac{15}{4}\left(\frac{\rho^2}{b^2}-1\right)^{\frac{3}{2}}\right.$$
$$\left.\left.\left.-\frac{25}{4}\left(\frac{\rho^2}{b^2}-1\right)^{\frac{1}{2}}-2\sin^{-1}\frac{b}{\rho}\right]\right\}+O(\alpha^6)\Bigg). \tag{5.27}$$

Thus we have completely determined the values of the unknown functions I, ϕ_1, and ϕ_2 which, when substituted in the Eqs. (5.3) and (5.4), give rise to the values of j_ρ and j_φ at the disk, as explained in the next section.

Note that the edge conditions (5.7) and (5.8) do not involve the constant C at all. However, to evaluate C, we observe from the integral representation formula (2.31) for the function $M_2(\rho, z)$ that $[M_2(\rho, z)]_{z=\pm 0}$ are continuous functions of ρ for all values of ρ if and only if C vanishes. It also follows from the continuity consideration of these functions at $\rho = b$ and the relations (2.13) and (2.16) that

$$C=0. \tag{5.28}$$

6. INDUCED SURFACE CURRENT DENSITY AND SCATTERING CROSS SECTION

The components of the total induced surface current density at the annulus are given by the formulas (5.3) and (5.4). Substituting the values of $\phi_1(\rho)$, $\phi_2(\rho)$, and $I(\rho)$ from Eqs. (5.23), (5.25), and (5.27) in the formulas (5.3) and (5.4), we readily obtain the values of j_ρ and j_φ at the disk. It follows from these values that

$$j_\rho(\rho, \varphi, 0)=O((\rho^2-b^2)^{-\frac{1}{2}})\quad\text{as}\quad\rho\to b,$$
$$j_\varphi(\rho, \varphi, 0)=O((\rho^2-b^2)^{-\frac{3}{2}})\quad\text{as}\quad\rho\to b.$$

We are not giving the values of j_ρ and j_φ explicitly, for the economy of space, and these can be determined as explained above.

Finally, to determine the scattering cross section, we first find the far field amplitudes of M_1 and M_2. In terms of spherical polar coordinates (r, θ, φ),

$$\rho=r\sin\theta,\quad z=r\cos\theta,$$

the far field amplitude for M_i, $i = 1, 2$, is defined as

$$M_i(\rho, z)=A_i(\theta)(e^{ikr}/r)+O(r^{-2}),\quad r\to\infty. \tag{6.1}$$

Comparing it with the integral representation formulas for M_1 and M_2 as given by (2.19) and (2.30), we obtain, after a slight simplification,

$$A_1(\theta)=-\frac{i\cot\theta}{2}\int_0^a I(t)\frac{d}{dt}\left[tJ_1(kt\sin\theta)\right]dt$$
$$=\frac{i\cot\theta}{2}\int_0^a t\frac{d}{dt}\left[I(t)\right]J_1(kt\sin\theta)\,dt$$
$$=\frac{i\cot\theta}{2}\int_b^a t\phi_1(t)J_1(kt\sin\theta)\,dt$$
$$=\frac{i\cot\theta}{2}\int_0^\infty t[\phi_{11}(t)+\phi_{12}(t)]J_1(kt\sin\theta)dt, \tag{6.2}$$

where we have used the relations (2.18), (2.25), (4.3), and (5.28). Now via the relations (4.6), (4.7), (4.18), and (4.19), there results from (6.2) after some manipulations

$$A_1(\theta)=ia\cos\theta\left(\frac{\alpha}{2\pi\sin\theta}\right)^{\frac{1}{2}}$$
$$\times\left(\int_0^1 v^{\frac{1}{2}}S_{11}(av)J_{\frac{1}{2}}(\alpha\sin\theta v)\,dv\right.$$
$$-\int_1^\infty v^{\frac{1}{2}}T_{11}(av)J_{\frac{1}{2}}(\alpha\sin\theta v)\,dv-\lambda^{\frac{3}{2}}$$
$$\times\int_0^1 v^{\frac{1}{2}}T_{12}(bv)J_{\frac{3}{2}}(\beta v\sin\theta)\,dv$$
$$\left.+\lambda^{\frac{3}{2}}\int_1^\infty v^{\frac{1}{2}}S_{12}(bv)J_{\frac{3}{2}}(\beta v\sin\theta)\,dv\right). \tag{6.3}$$

Similarly,

$$A_2(\theta) = -\tfrac{1}{2}i\Big(Bk^2\int_0^b t^2 J_1(kt\sin\theta)\,dt - \int_b^a t\phi_2(t)J_1(kt\sin\theta)\,dt\Big)$$
$$= -\tfrac{1}{16}iBa\alpha^3\lambda^4\sin\theta[1+O(\alpha^4)] + ia\Big(\frac{\alpha\sin\theta}{2\pi}\Big)^{\frac{1}{2}} \times \Big(\int_0^1 v^{\frac{1}{2}}S_{21}(av)J_{\frac{1}{2}}(\alpha\sin\theta v)\,dv - \int_1^\infty v^{\frac{1}{2}}T_{21}(av)J_{\frac{1}{2}}(\alpha v\sin\theta)\,dv - \lambda^{\frac{3}{2}} \times \int_0^1 v^{\frac{1}{2}}T_{22}(av)J_{\frac{3}{2}}(\beta v\sin\theta)\,dv + \lambda^{\frac{3}{2}}\int_1^\infty v^{\frac{1}{2}}S_{22}(av)J_{\frac{3}{2}}(\beta v\sin\theta)\,dv\Big). \quad (6.4)$$

Substituting the values of S_{11}, S_{12}, T_{11}, and T_{12} from the Eqs. (5.13)–(5.16) in (6.3), we obtain, after some simplification,

$$A_1(\theta) = \frac{4a^2\alpha\cos\theta}{3\pi\omega\mu}\Big[1+\frac{8\alpha^2}{15}+\frac{8i\alpha^3}{9\pi} + \frac{16\alpha^4}{105}+\frac{176i\alpha^5}{225\pi}-\frac{\alpha^2}{10}\Big(1+\frac{11\alpha^2}{21}+\frac{8i\alpha^3}{9\pi}\Big)\sin^2\theta + \frac{\alpha^4}{280}\sin^4\theta - \frac{16\lambda^5}{15\pi^2}+O(\alpha^6)\Big]. \quad (6.5)$$

Similarly,

$$A_2(\theta) = -\frac{4\alpha a^2\sin\theta}{3\pi\omega\mu}\Big[1+\frac{11\alpha^2}{30}+\frac{8i\alpha^3}{9\pi} + \frac{17\alpha^4}{280}+\frac{408i\alpha^5}{675\pi}+\cos^2\theta\Big(\frac{\alpha^2}{10}+\frac{17\alpha^4}{420}+\frac{4i\alpha^5}{45\pi}\Big) + \frac{\alpha^4}{280}\cos^4\theta - \frac{16\lambda^5}{15\pi^2}+O(\alpha^6)\Big]. \quad (6.6)$$

Having found $A_1(\theta)$ and $A_2(\theta)$, we use the formula

$$\sigma = \frac{\omega^2}{a^2}(\mu k)^2 \times \int_0^\pi [|A_1(\theta)|^2 + \cos^2\theta\,|A_1(\theta)+A_2'(\theta)|^2]\sin\theta\,d\theta, \quad (6.7)$$

where

$$A_2'(\theta) = -\tan\theta A_2(\theta),$$

for calculating the scattering cross section σ. After substituting the values from (6.5) and (6.6) and simplifying, we obtain

$$\sigma = \frac{128\alpha^4}{27\pi^2}\Big(1+\frac{22\alpha^2}{25}+\frac{7312\alpha^4}{18375}-\frac{32\lambda^5}{15\pi^2}+O(\alpha^6)\Big). \quad (6.8)$$

When $\lambda\to 0$, (6.8) and the values of j_ρ and j_φ agree with the known results for the whole circular disk.[1]

APPENDIX

We give here the proof of result (5.5).

One readily deduces from the Eqs. (4.3), (4.18), and (4.19) that

$$\frac{2}{\pi}\frac{d}{d\rho}\Big(\int_\rho^\infty \frac{T_{11}(u)\,du}{(u^2-\rho^2)^{\frac{1}{2}}}\Big) + \frac{2}{\pi\rho^2}\frac{d}{d\rho} \times \Big(-\int_0^b \frac{u^2T_{12}(u)\,du}{(\rho^2-u^2)^{\frac{1}{2}}} + \int_b^\rho \frac{u^2S_{12}(u)\,du}{(\rho^2-u^2)^{\frac{1}{2}}}\Big) = 0, \quad a<\rho<\infty. \quad (A1)$$

If the function $T_{11}(\rho)$ is extended over the range $b<\rho\le a$, the Eq. (A1) will also hold for the range $b<\rho\le a$. But then the Eq. (4.25) yields

$$\phi_1(\rho) = -\frac{2}{\pi}\frac{d}{d\rho}\Big(\int_\rho^a \frac{[S_{11}(u)+T_{11}(u)]}{(u^2-\rho^2)^{\frac{1}{2}}}\,du\Big), \quad b<\rho\le a. \quad (A2)$$

This equation gives, after some manipulations, the required result

$$\phi_1(\rho) = \frac{2}{\pi}\Big(\frac{S_{11}(a)+T_{11}(a)}{(a^2-\rho^2)^{\frac{1}{2}}}+O((a^2-\rho^2)^{\frac{1}{2}})\Big) \quad \text{as} \quad \rho\to a. \quad (A3)$$

Similarly, the result (5.6) can be established. We have verified that the solution $\phi_1(\rho)$ satisfies the formulas (A2) as well as (A3).

* This work was sponsored by a National Science Foundation Grant.

[1] C. J. Bouwkamp, Philips Res. Rept. **5**, 401 (1950).
[2] K. A. Lur'e, Zh. Tekh. Fiz. **29**, 1421 (1959) [Sov. Phys. Tech. Phys. **4**, 1313 (1959)].
[3] W. E. Williams, Cambridge Phil. Soc. **58**, 625 (1962).
[4] S. Inawashiro, J. Phys. Soc. (Japan) **18**, 273 (1963).
[5] J. Boersma, Composito Math. **16**, 205 (1965).
[6] J. Bazer and L. Rubenfeld, J. Soc. Ind. Appl. Math. **13**, 558 (1965).
[7] V. S. Gubenko and V. I. Mossakovskii, Prikl. Mat. Mekh. **24**, 477 (1960).
[8] E. T. Copson, Proc. Edinburgh Math. Soc. **8**, 14 (1947).
[9] W. E. Williams, Proc. Edinburgh Math. Soc., II **13**, 235 (1963).
[10] D. P. Thomas, Intern. J. Eng. Sci. **3**, 405 (1965).
[11] D. L. Jain and R. P. Kanwal, Intern. J. Eng. Sci. **8**, 687 (1970).
[12] D. L. Jain and R. P. Kanwal, J. Eng. Math. **4**, 219 (1970).
[13] I. Stakgold, *Boundary Value Problems of Mathematical Physics* (Macmillan, New York, 1968), Vol. II.
[14] B. Noble, *Electromagnetic Waves*, R. E. Langer, Ed. (U. of Wisconsin Press, Madison, Wisconsin, 1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Relation between the Singularities of the S Matrix and the L^2 Class of Solutions in Potential Theory

C. A. NELSON, A. K. RAJAGOPAL, AND C. S. SHASTRY
Department of Physics and Astronomy, Louisiana State University
Baton Rouge, Louisiana 70803

(Received 15 October 1970)

Ma has shown that, in the case of an exponential potential, in the s-wave S matrix there exist poles that do not contribute to the completeness, even though they appear in the same part of the k plane as the bound state poles. These poles were called "redundant" poles. Subsequently, other examples of redundant poles have appeared in the literature. Recently their importance with regard to the concept of "shadow" states has been stressed by Sudarshan. If we construct the Green's function for the Schrödinger equation with the boundary conditions of regularity at the origin and the outgoing spherical waves at large distances, the singularities of it in the k plane completely determine the L^2 class of eigensolutions. From this Green's function, the T matrix (and hence the S matrix) is constructed explicitly. This S matrix is found to be the same as the one defined through the usual Jost solution, and it is shown that the singularities of the S matrix, besides corresponding to those appearing in the Green's function, also contain the "redundant" singularities. Thus it is shown that the wavefunctions associated with the "redundant" singularities do not belong to the L^2 class. By a careful derivation, we resolve the Ma paradox concerning the Heisenberg identity. In the particular cases of the exponential potential and the Eckart potential, the redundant solutions correspond to the vanishing of a Wronskian pointing to the breakdown of the linear independence of the starting wavefunctions in defining the S matrix through the Jost functions. It is stressed that when singularities appear in the S matrix other than those corresponding to the bound and scattering states of the given problem, the "redundant" singularities can only be understood by a "dynamical" equation such as the Schrödinger equation.

I. INTRODUCTION

Almost a quarter of a century ago, Ma[1] and, soon after, Jost[2] discovered that, for exponential potential for s waves, the singularities of the S matrix in the physical plane are of two kinds. One set of poles corresponds to genuine bound states, and another set does not correspond to anything physical, in the sense that they do not appear in the completeness statement concerning the entire set of solutions for this Hamiltonian. This was exhibited in terms of a "Heisenberg identity,[3]" which is derived from the completeness statement concerning the L^2 class of solutions of the problem, when one applied it to the asymptotic region of the configuration space where the scattering solutions in this region are written in terms of the phase shift in the usual way. The Heisenberg identity (HI) thus derived relates the spatial Fourier transform of the S matrix, defined by $\exp 2i\delta(k)$ [$\delta(k)$ is the phase shift] for real k, to the sum over the square of the modulus of the asymptotic bound state solutions. This integral can also be computed by a method of contour integration when $S(k)$ is assumed to be the suitably analytically continued function in the upper half k-complex plane which coincides with $\exp 2i\delta(k)$ for real k. This evaluation then led to the paradox that certain new terms now appear besides the known bound state contributions. Since the complete solutions of this problem can be written down, Ma verified explicitly that the completeness of the set of solutions does not contain terms corresponding to these "redundant" states. We resolve this paradox in this paper by showing that HI is, in actual fact, a tautology when the terms left out in the asymptotic expansions are examined with care. Subsequently, other examples of redundant poles appeared in the literature.[4,5] Peierls[6] later showed that if the potential is a cutoff exponential, these redundant poles do not appear at all. Moreover, by arguments of perturbative character, by considering a superposition of exponential potentials, he associated the "left-hand cut" for the Yukawa potential with the "redundant" states, in that this does not appear in the completeness of solutions for this case. Quite recently, Biswas, Pradhan, and Sudarshan[7] have investigated these poles to elucidate the concept of "shadow" states.

In this paper, we wish to put forward an explanation of these "redundant" singularities of the S-matrix in terms of the associated wavefunctions belonging to a non-L^2 class. Moreover, by a careful consideration of the derivation of HI, we explicitly show that, in actuality, a correct derivation yields a tautology and not HI. This resolves Ma's paradox. In *nonrelativistic* potential theory, a *complete* identification of the states for given boundary conditions can be made by constructing the Green's function for the system.[8] *From the Green's function the completeness of the entire set of solutions is derived;* also,

following the usual operator relationship *between the T matrix* and the *full Green's function, one may derive the S matrix* for real values of the momenta and energy. *This is done purely to bring out the appearance of an extra term in the S matrix. Also this gives us a check on the usual identification of S from the Jost solutions.* All these are given in the next section. In the third section, two examples discussed in the literature are re-examined in the light of our approach, and it is shown that the "redundant poles" correspond to the breakdown of linear independence of two basic solutions employed in constructing the usual S matrix. Thus the redundant poles correspond here to vanishing of the relevant Wronskian. A summary of the results is given at the end.

II. GREEN'S FUNCTION, COMPLETENESS RELATION, AND S MATRIX FOR SHORT-RANGED POTENTIAL

The solutions and definitions employed here are those of Newton.[4] For simplicity, we will throughout discuss the s-state Schrödinger equation only, and a discussion of the higher angular momentum states does not in any way alter the conclusions drawn from this simple case. Since most of the details of constructing the Green's function can be found in either Ref. 4 or Ref. 8, we here give only the results. The radial Schrödinger equation to be solved is

$$-\psi'' + \mathfrak{U}\psi = k^2\psi, \tag{1}$$

where $\mathfrak{U} = 2\mu V$ and other symbols have their usual significance.[9] We will assume that $r = 0$ is "regular" in the sense of the theory of ordinary differential equations of second order, so that it will be assumed that $\lim r^2\mathfrak{U}(r) = 0$ as $r \to 0$. Only *regular* solutions are acceptable for the discussion of the physical problems. If $f_\pm(k, r)$ exist such that

$$\lim_{r\to\infty} e^{\mp ikr} f_\pm(k, r) = 1 \tag{2}$$

and are solutions of (1), then, if these *two solutions are linearly independent*, we can construct a general *regular solution* of (1) as

$$\varphi(k, r) = (2ik)^{-1}[f_-(k, 0)f_+(k, r) - f_+(k, 0)f_-(k, r)]. \tag{3}$$

$f_\pm(k, 0)$ is $f_\pm(k, r)$ evaluated for $r = 0$ and is called the Jost function. Note that $\phi(k, r) \to 0$ as $r \to 0$. We now construct the Green's function associated with (1), and it satisfies the equation

$$\left(-\frac{\partial^2}{\partial r^2} + \mathfrak{U}(r) - k^2\right)\mathfrak{G}^{(+)}(k; r\ r') = -\delta(r - r'), \tag{4}$$

with the boundary conditions that it be regular at $r = 0$ and contains only *outgoing waves* for $r \to \infty$. Then, in terms of (2) and (3), we have[4]

$$\mathfrak{G}^{(+)}(k; r, r') = -[f_+(k, 0)]^{-1}[f_+(k, r)\varphi(k, r')\eta_+(r - r') + f_+(k, r')\varphi(k, r)\eta_+(r' - r)], \tag{5}$$

where $\eta_+(r - r') = 1$ for $r > r'$ and is zero otherwise.

We may also state here that the functions $f_\pm(k, r)$ are such that their Wronskian

$$W(f_+, f_-) = -2ik, \tag{6}$$

and $f_\pm(k, 0)$ are the Wronskians

$$W(f_\pm(k, r), \varphi(k, r)) = f_\pm(k, 0), \tag{7}$$

$W(\psi_1, \psi_2) = \psi_1\psi_2' - \psi_1'\psi_2$. The linear independence of the various solutions are implied by nonzero values of the corresponding Wronskians. The completeness of the entire set of solutions that obey the requirements of regularity at the origin and outgoing wave at infinity is at once arrived at from (5) in the usual way. Defining the normalized bound state solutions and the "outgoing" scattering solutions by

$$\psi^{(n)}(r) = \varphi^{(n)}(r)/N_n \equiv \varphi(iK_n, r)/N_n \tag{8a}$$

and

$$\psi^{(+)}(k, r) = k\varphi(k, r)/f_+(k, 0), \tag{8b}$$

where N_n is the normalization constant, we then find that the completeness statement is

$$\sum_n \psi^{(n)}(r)\psi^{(n)*}(r') + \frac{2}{\pi}\int_0^\infty dk\psi^{(+)}(k, r)\psi^{(+)*}(k, r') = \delta(r - r'). \tag{9}$$

This also determines the complete set of the L^2 class of solutions for this problem.[8] The iK_n here are the poles in $\mathfrak{G}^{(+)}(k, r, r')$ or the zeros of $f_+(k, 0)$ in the upper half k-complex plane. It may be noted that, for a *short-ranged* potential, $f_+(k, r)$ and $f_+(k, 0)$ are analytic in the upper half-plane in k but $\phi(k, r)$ is *analytic in* the entire k plane.[2] From (5) then, *only the zeros of* $f_+(k, 0)$ *contribute to the completeness of the* L^2 *class of solutions of Eq. (1).* It may be remarked that, for $k = iK_n$, $\phi(k, r)$ is such that it vanishes for $r \to \infty$.

From (5) we now proceed to construct the T matrix by employing the relation

$$T = \mathfrak{U} + \mathfrak{U}\mathfrak{G}^{(+)}\mathfrak{U}. \tag{10}$$

We take the diagonal matrix element of (10) corresponding to the energy k^2, so that we directly obtain T on the mass shell. Thus, for s waves, we need to compute

$$\langle k|\, T\, |k\rangle = B_0 + M,$$

where B_0 is the term corresponding to $\mathcal{U}$, the usual Born term, and M stands for the matrix element of the second term of (10). Here $\langle r \mid k\rangle = j_0(kr)$. Thus

$$B_0 = \int_0^\infty dr\, r j_0^2(kr)\mathcal{U}(r)$$

and

$$M = \int_0^\infty r\, dr \int_0^\infty r'\, dr' \times j_0(kr)\mathcal{U}(r)\mathcal{G}^{(+)}(k; r, r')\mathcal{U}(r')j_0(kr'). \quad (11)$$

We now employ (5) in evaluating M. We use the Schrödinger equation obeyed by $f_+(k, r)$ in the presence of $\mathcal{U}$ and that obeyed by $j_0(kr)$ in the absence of $\mathcal{U}$, to simplify the above expression for M. After some algebra, we obtain

$$M = -\frac{f_-(k,0)}{2ikf_+(k,0)}[f_+(k,0)-1]^2 - B_0 + \frac{f_-(k,0)}{2ik}[f_+(k,0)-1] - \frac{1}{2ik}[f_-(k,0)-1]. \quad (12)$$

Thus we get

$$\langle k|\, T\,|k\rangle = (2ik)^{-1}[1 - f_-(k,0)/f_+(k,0)] \quad (13)$$

after clearing the algebraic expression. It is interesting to note that the *usual* Born term is cancelled by a corresponding term in M and the remainder adds up to (13). Using the relationship

$$S(k) = 1 - 2ik\,\langle k|\, T(k)\,|k\rangle, \quad (14)$$

we arrive at the familiar S matrix, which could be identified even at the starting point when $\phi(k, r)$ is defined by (3). The point of this calculation was to show explicitly that the expression (9), which displays only the appearance of $f_+(k, 0)$ through $\mathcal{G}^{(+)}$, has hidden in it the usual $S(k)$. The algebra outlined above shows another feature that the Born term in (9) does not appear explicitly if T is evaluated by making use of $\mathcal{G}^{(+)}$.

This demonstration serves to indicate the appearance of the singularities of $f_-(k, 0)$ in $S(k)$ but not in $\mathcal{G}^{(+)}(k; rr')$. The corresponding eigenfunctions thus do not appear in the completeness statement, and hence by exclusion they do not belong to the L^2 class of solutions of (1). In the next section we will show two examples where the corresponding $f_-(k, 0)$ have poles, which are then explicitly identified to be zeros of the Wronskian of $f_+(k, r)$ and $f_-(k, r)$ when appropriately redefined.

We will now show that Ma's puzzle is really not a puzzle at all by re-examining with care the derivation of the Heisenberg identity. Let us write

$$f_\pm(k, r) = e^{\pm ikr} + [f_\pm(k, r) - e^{\pm ikr}] \equiv e^{\pm ikr} + \tilde{f}_\pm(k, r). \quad (15)$$

$\tilde{f}_\pm(k, r)$ is designed to vanish for $r \to \infty$. From (3), we then have

$$\varphi(k, r) = \varphi^0(k, r) + \tilde{\varphi}^0(k, r), \quad (16)$$

where

$$\varphi^0(k, r) = (2ik)^{-1}[f_-(k, 0)e^{ikr} - f_+(k, 0)e^{-ikr}] \quad (16'a)$$

and

$$\tilde{\varphi}(k, r) = (2ik)^{-1}[f_-(k, 0)\tilde{f}_+(k, r) - f_+(k, 0)\tilde{f}_-(k, r)]. \quad (16'b)$$

Let us substitute (16) in (8b) and subsequently in the completeness relationship (9). Then an exact restatement of the completeness relation (9) is

$$\sum_n \psi^{(n)}(r)\psi^{(n)*}(r') + \frac{2}{\pi}\int_0^\infty dk\,\psi_0^{(+)}(k, r)\psi_0^{(+)*}(k, r') + A = \delta(r - r'), \quad (17)$$

where

$$\psi^{(+)}(k, r) = \psi_0^{(+)}(k, r) + \tilde{\psi}^{(+)}(k, r) = \frac{k\varphi^{(0)}(k, r)}{f_+(k, 0)} + \frac{k\tilde{\varphi}^{(0)}(k, r)}{f_+(k, 0)} \quad (18)$$

and so

$$A = \frac{2}{\pi}\int_0^\infty dk[\psi_0^{(+)}(k, r')\tilde{\psi}^{(+)*}(k, r') + \tilde{\psi}^{(+)}(k, r)\psi_0^{(+)*}(k, r') + \tilde{\psi}^{(+)}(k, r)\tilde{\psi}^{(+)*}(k, r')]. \quad (19)$$

By construction, when (17) is applied to r, r' in the asymptotic region, we obtain

$$\frac{1}{2\pi}\sum_n |c_n|^2 e^{-|K_n|r} + \frac{2}{\pi}\int_0^\infty dk \sin[kr + \delta(k)]\sin[kr' + \delta(k)] + A' = \delta(r - r'). \quad (20)$$

A' now contains A plus the left-out terms of the second term in (17). If it is *assumed* that if $A' \to 0$, then one arrives at the Heisenberg identity upon using the definitions $S(k) = e^{2i\delta(k)} = 1/S(-k)$:

$$\int_{-\infty}^\infty dk S(k)e^{ik(r+r')} = \sum_n |c_n|^2 e^{-|K_n|(r+r')}. \quad (21)$$

By a contour integration method, the integral on the left side can also be calculated, in which case we would obtain an extra term corresponding to the

redundant poles. Ma suspected that, even though $\tilde{\psi}^{(+)}$ is zero, in the asymptotic region, an actual evaluation of A (or equivalently A') may *yield* the missing contribution due to the redundant poles, thus giving the result of the contour integration method. This was Ma's puzzle. However, if one explicitly *calculates A by using only* the definitions and properties of $f_{\pm}(k, r)$, $f_{\pm}(k, 0)$, and $S(k)$ *but not evaluating anywhere* the Fourier transform of $S(k)$, one merely obtains, after straightforward algebra, the result that

$$A = -\sum_n \psi^{(n)}(r)\psi^{(n)*}(r') + \frac{2}{\pi}\int_{-\infty}^{\infty} dk\, S(k)e^{ik(r+r')}. \tag{22}$$

Note that we have not used any asymptotic expansions. When (22) is substituted in (17), we merely obtain the well-known identity

$$\frac{2}{\pi}\int_0^{\infty} dk \sin[kr + \delta(k)] \sin[kr' + \delta(k)] \equiv \delta(r - r') - \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, S(k)e^{ik(r+r')}, \tag{23}$$

which thus leads us to a tautology!

To make this a little more explicit, we reconstruct the above calculation for the simple case of Eckart potential, where all the terms can be explicitly written out. The pitfall then is an actual evaluation of certain integrals that make up A by *contour methods*, in which case one merely verifies Cauchy's theorem on contour integration for $S(k)$. Thus we have resolved the original paradox posed by Ma. This tautology also reaffirms the fact that, for a given set of parameters, the non-L^2 solutions are distinct from the L^2 set and by such methods as asymptotic expansions of the L^2 set one cannot glean at the structure of the non-L^2 solutions.

III. THE EXPONENTIAL AND THE ECKART POTENTIALS

A. The Exponential Potential[1]

Here

$$V(r) = -v_0 e^{-r/a}. \tag{24}$$

The complete solutions for s waves exist, and

$$f_+(k, r) = e^{iak\ln(a^2v_0)}\Gamma(1 - 2iak)J_{-2iak}(2av_0^{\frac{1}{2}}e^{-r/2a}),$$
$$f_-(k, r) = f_+(-k, r). \tag{25}$$

Then

$$S(k) = \frac{f_-(k, 0)}{f_+(k, 0)} = e^{-2iak\ln(a^2v_0)}\frac{\Gamma(1 + 2iak)J_{2iak}(2av_0^{\frac{1}{2}})}{\Gamma(1 - 2iak)J_{-2iak}(2av_0^{\frac{1}{2}})}. \tag{26}$$

The zeros of $f_+(k, 0)$ corresponding to

$$J_{-2iak}(2av_0^{\frac{1}{2}}) = 0 \tag{27}$$

define the usual bound states. The poles of $f_-(k, 0)$ occur when $\Gamma(1 + 2iak) = \infty$, or when $1 + 2iak =$ zero or negative integer, or, equivalently,

$$2iak = -n, \quad n \text{ integer } (\neq 0). \tag{28}$$

The Wronskian of $J_\nu(z)$ and $J_{-\nu}(z)$, which are actually the two solutions $f_{\pm}(k, r)$ without the factors in (25), is given by

$$W[J_\nu(z), J_{-\nu}(z)] = -2\sin(\nu\pi)/\pi z. \tag{29}$$

This therefore vanishes when ν is an integer. Thus, the starting solutions (15), where J_{-2iak} and J_{2iak} are assumed to be linearly independent, are not so at values of k given by (28). In fact, $\phi(k, r)$, defined by (3) and (25), for such values of k can be computed, and, after some algebra, one obtains

$$\varphi(k, r) \xrightarrow[2iak=-n]{} -a[J_n(x)Y_n(xe^{-r/2a}) - J_n(xe^{-r/2a})Y_n(x)] \tag{30}$$

with $x = 2a(v_0)^{\frac{1}{2}}$. One thus obtains J_n and Y_n as the two linearly independent solutions.[10] $\phi(k, r)$ does not have the required behavior as $r \to \infty$ since it now involves $Y_n(0)$, not as for bound states.

B. Eckart Potential

Here

$$V(r) = -2\beta\lambda^2 e^{-\lambda r}/(1 + \beta e^{-\lambda r})^2, \quad 0 < \lambda, \quad -1 < \beta < 0. \tag{31}$$

Then,

$$f_+(k, r) = e^{ikr}\{[2k - i\lambda g(r)]/(2k + i\lambda)\}, \tag{32}$$

with

$$g(r) = (\beta e^{-\lambda r} - 1)/(\beta e^{-\lambda r} + 1),$$

and other quantities follow as before:

$$S(k) = (2k + i\lambda)[2k + i\lambda g(0)]/[2k - i\lambda g(0)](2k - i\lambda). \tag{33}$$

The bound states correspond to zeros of $f_+(k, 0)$, and we locate them at

$$k = i\tfrac{1}{2}\lambda g(0) \tag{34}$$

and the corresponding $\phi(k, r) \to 0$ as $r \to \infty$. For the present choice of β, this pole does not occur in the upper half-plane. A pole of $f_-(k, 0)$ is found at

$$k = i\lambda/2. \tag{35}$$

Defining $\bar{f}_+(k, r) = e^{ikr}[2k - i2g(r)]$ to be the solution instead of (32), we find that the Wronskian of the

two basic solutions is

$$W(\bar{f}_+(k,r),\bar{f}_-(k,r)) = 2ik(4k^2+\lambda^2), \quad (36)$$

and this vanishes for $k = i\lambda/2$. Computing $\phi(k,r)$ for $k = i\lambda/2$, we see that

$$\varphi(k,r) \xrightarrow[k=i\lambda/2]{} \frac{1}{2\lambda}\{\lambda r h(r) + (e^{\lambda r/2} - e^{-\lambda r/2}) \times [1 + g(r)g(0)]\}, \quad (37)$$

where

$$h(r) = 4\beta e^{-\lambda r/2}/(\beta+1)(\beta e^{-\lambda r}+1).$$

Again, $\phi(k,r)$ diverges as $r \to \infty$.

It may also be noted that the location of these poles are independent of the strength of the potential in contrast to the bound state poles.

We now demonstrate explicitly that the redundant poles do not appear in any computation of the sort employed by Ma. From (32) we note that we may rewrite it in the form

$$f_+(k,r) = e^{ikr} - e^{ikr}i\lambda[1+g(r)]/(2k+i\lambda) \equiv e^{ikr} + \bar{f}_+(k,r), \quad (32')$$

and similarly for $f_-(k,r)$. Hence one can explicitly compute A, given by (18). All these integrals can be done by the methods of *contour integration* or *otherwise*, leading in actuality to the statement that

$$A = (2/\pi)\int_{-\infty}^{\infty} dk S(k) e^{ik(r+r')} = 2\pi i \operatorname{Res}[(2/\pi)S(k)e^{ik(r+r')}]_{k=i\lambda/2}.$$

This is obviously true. If we do not evaluate the integrals by any technique, but rewrite it in terms of $S(k)$, we merely obtain (23). This then verifies our resolution of the Ma paradox for the Eckart potential.

IV. CONCLUSIONS

From the analysis presented here, it is now possible to make a definite statement in potential theory about the nature of the solutions corresponding to the singularities that occur in the S matrix, viz., that the poles of $f_-(k,0)$ correspond to non-L^2 class of solutions of the Schrödinger equation and the zeros of $f_+(k,0)$ to the L^2 class of solutions. This is in conformity with one's intuitive feeling that, had there been L^2 class of functions corresponding to the redundant poles, it should have been manifest in the completeness relation. From the two examples given above, for the exponential type potential, poles appear in $f_-(k,0)$ which correspond to vanishing of the Wronskian of the solutions, $f_\pm(k,r)$, and hence their linear independence required for the definition of $\phi(k,r)$; hence the definition of $S(k)$ breaks down. Since $f_\pm(k,r)$ were chosen to obey (2), the new linear independent set will necessarily correspond to those which do not obey (2), and hence, also, for such a case, the non-L^2 type of solutions appear. It is known[6] that for any cutoff potential of the exponential type, poles do not occur at all for $f_-(k,0)$, and, in fact, $f_\pm(k,0)$ do not contain any singularities. Following the argument of Peierls,[6] the cut replaces the redundant poles in the case of the Yukawa potential, and the corresponding wavefunction seems to diverge at infinity, just as for the redundant poles. We have thus resolved the original Ma puzzle and given a meaning to the redundant poles by characterizing the corresponding wavefunction.

ACKNOWLEDGMENTS

We wish to thank Professor E. C. G. Sudarshan for stressing that a complete resolution of the Ma paradox necessitates an understanding of the missing terms of the asymptotic expansion. One of us (C. A. N.) wishes to thank Professor E. C. G. Sudarshan for introducing him to "shadow" states and for the warm hospitality at the Center for Particle Theory, the University of Texas at Austin, during the summer of 1970.

[1] S. T. Ma, Phys. Rev. **69**, 668 (1946) and **71**, 195 (1947). Ma employs a different normalization in defining $\phi(k,r)$ which makes it vanish at the redundant poles. This, therefore, is not useful in bringing out the explicit linear dependence of the solutions at such poles. See also K. Kodeira, Am. J. Math. **71**, 921 (1949); **72**, 502 (1950).

[2] R. Jost, Helv. Phys. Acta **20**, 256 (1947); see also A. Martin, Nuovo Cimento **14**, 403 (1959).

[3] W. Heisenberg, Z. Naturforsch. **11/12**, 607 (1946).

[4] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

[5] A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decay in Nonrelativistic Quantum Mechanics* (Israel Program for Scientific Translations, Jerusalem, 1969), p. 63. See also A. Bhattacharjee and E. C. G. Sudarshan, Nuovo Cimento **25**, 864 (1962).

[6] R. E. Peierls, Proc. Roy. Soc. (London) **A253**, 16 (1959).

[7] S. N. Biswas, T. Pradhan, and E. C. G. Sudarshan, Center for Particle Theory, Austin, Texas Preprint CPT-70 AEC-20, 1970. This paper inspired the present investigation.

[8] B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956).

[9] A potential is called short range if it decreases exponentially at infinity, i.e., if there exists an $a > 0$ such that

$$\int_0^\infty r\,dr\,e^{ar}|\mathfrak{V}(r)| < \infty.$$

[10] This is also noted in Ref. 5.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Erratum: Properties of Overcomplete and Nonorthogonal Basis Vectors

[J. Math. Phys. **10,** 1774 (1969)]

D. J. ROWE
Department of Physics, University of Toronto, Ontario, Canada

(Received 6 October 1970)

In this paper a simple alternative orthogonalization procedure to the well-known Schmidt procedure was presented with the principal objective of showing how equations-of-motion calculations could be performed for excitation operators expanded in a space of nonorthogonal, and even over-complete, basis operators. It has since come to my attention that this orthogonalization procedure is identical to the "Method of Canonical Orthonormalization" presented by Löwdin[1] many years ago and used by him in subsequent papers.[2] I wish, therefore, to draw the attention of the reader to these papers and to apologize to Löwdin for my failure to refer to his work.

[1] P.-O. Löwdin, Advan. Phys. **5,** 1 (1956).
[2] P.-O. Löwdin, Rev. Mod. Phys. **39,** 259 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Erratum: Maxwell's Equations Having a Gradient as Source

[J. Math. Phys. **11,** 2075 (1970)]

M. MURASKIN
Department of Physics, University of North Dakota, Grand Forks, North Dakota 58201

(Received 6 October 1970)

A gradient source also appears in the work of J. S. Dowker and Y. P. Dowker [Proc. Roy. Soc. (London) **A294,** 175 (1966)] and of J. Frenkel [*Wave Mechanics—Advanced Theory* (Oxford U.P., Oxford, 1934), pp. 266ff]. Other pertinent references are J. M. Whittaker, Proc. Cambridge Phil. Soc. **24,** 501 (1928), G. Rumer, *Z.* Physik **65,** 244 (1930), and M. Sachs and S. Schwebel, J. Math. Phys. **3,** 843 (1962). This last reference has some expressions similar to ours, except that $f = 0$ is found there. I am grateful to Dr. J. S. Dowker for drawing my attention to these references.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 4 APRIL 1971

Erratum: Direct Canonical Transformations

[J. Math. Phys. **11,** 2776 (1970)]

DAVID P. STERN
Laboratory for Space Physics, Goddard Space Flight Center, Greenbelt, Maryland 20771

(Received 8 October 1970)

In this article the claim was made that a perturbation method derived by Lacina[1] was in error by neglecting the lower limit of an integral. I am indebted to Dr. Lacina for pointing out and proving that the contribution of this limit vanishes. The proof is by induction, as follows (numbers refer to equations in the author's paper).

To show that one possible solution in Eq. (42) is obtained by setting $\mu_i^{(k)}$ equal to zero, suppose that this has already been proved for $m \leq k - 1$; then, for those orders, $\zeta_i^{(m)}$ vanishes at $\bar{y}_1 = C$. Now $\mathbf{f}^{(k)}$ depends on these $\zeta_i^{(m)}$ and on their derivatives, and any term in $\mathbf{f}^{(k)}$ contains at least one such component in undifferentiated form, associated with $\partial/\partial y_i$. Therefore, with the given choice of $\mu_i^{(m)}$ of lower orders, $\mathbf{f}^{(k)}$ ($\bar{y}_1 = C$) vanishes. Choosing $\psi^{(k)} = 0$ in (45) then shows that $\mu_i^{(k)}$ also vanishes. To show this inductive process can be started, it is only necessary to observe that $\mathbf{f}^{(1)} = 0$, not only for $\bar{y}_1 = C$ but also in general.

[1] J. Lacina, Ann. Phys. (N.Y.) **51,** 381 (1969).