S-Matrix Singularity Structure in the Physical Region. II. Unitarity Integrals*

M. J. W. BLOXHAM, D. I. OLIVE, AND J. C. POLKINGHORNE
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England

(Received 1 August 1967)

The general techniques developed in an earlier paper [J. Math. Phys. **10**, 494 (1969)] are applied to evaluate the singularities and discontinuities of unitarity integrals. The results are conveniently expressed in terms of what we call mechanism (or M) diagrams.

1. INTRODUCTION

A unitarity integral is regarded as an integral over real loop momenta of an integrand which is a product of mass-shell δ functions and mass-shell amplitudes possessing the Landau–Cutkosky singularity structure. This paper applies the results of a previous paper,[1] hereafter called I, to the singularities of unitarity integrals occurring for physical momenta. The main problem in doing this is the framing of a suitable language or notation in which to deal with combinatorial complications. Since the anticipated singularities lie on arcs of Landau curves which can be related to Landau diagrams, and since the unitarity integrals themselves can be represented by unitarity diagrams,[2] it is natural to develop a diagrammatic technique. In fact, it is already known[3] (and here confirmed in greater detail) that, roughly speaking, the singularities of unitarity integrals correspond to diagrams obtained by replacing the "bubbles" of the unitarity diagram by Landau diagrams corresponding to possible singularities of the bubbles. The diagram so obtained is called a mechanism (or M) diagram, and plays a central role in providing simple rules for further important properties of the singularities, namely, (i) in distinguishing which arcs of a Landau curve may be singular, (ii) in determining whether (and if so in what sense) the evaluations of the unitarity integral on the two sides of the singularity $I_>$ and $I_<$ are analytically related, (iii) in determining (where relevant) the discontinuity of the unitarity integral, paying special attention to the integration region and over-all numerical factors, and (iv) in providing general rules for the quantities $I_> - I_{<\eta-i\epsilon}$, where $I_{<\eta-i\epsilon}$ is a continuation of $I_<$ to the same side as $I_>$.

The work applies to any unitarity integral, however many particles are involved, and to any singularity, providing that the corresponding Landau diagram has no more than one line joining any two bubbles (a restriction we intend to remove in a future paper). We exclude points where the singular curve has a node or intersects other singular curves. In Sec. 2 we recall how Landau singularities and unitarity integrals are conveniently related to Landau and unitarity diagrams, and in Sec. 3 we use the idea of a mechanism diagram to find what singularities unitarity integrals may have. The M diagram is relevant for each of the three classes of singularity distinguished in I—generative, explicit, and regenerative—and it also provides a rule for the natural distortion prescribing the analytic relation, if any, between $I_<$ and $I_>$. In Sec. 4 we find that the main difficulty in using the results of I to evaluate the discontinuity due to a specific generative mechanism is that, when we have identical particles, a given singularity may be generated by many different sets of singularity surfaces, and we must have a technique for counting the different pinch points that arise. This is provided by a simple group-theoretic treatment of permutational symmetries in the relevant diagrams. In Sec. 5 we derive rules for evaluating $I_> - (I_<)_{\eta-i\epsilon}$ that are valid, no matter how the singularity originates. This is our main result, and it is employed in the subsequent paper, in which the unitarity equations are used to determine the singularities of the amplitude itself.

2. LANDAU DIAGRAMS

Consider a diagram representing a sequence of intermediate scattering processes that may occur for a given number of initial and final external particles. By energy-momentum conservation at each vertex, each internal line carries a four-momentum q, which is a linear combination of the external momenta p and loop momenta k, just as in a Feynman diagram.

* The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research under Grant AF EOAR 65-36 through the European Office of Aerospace Research (OAR), United States Air Force.

[1] M. J. W. Bloxham, D. I. Olive, and J. C. Polkinghorne, J. Math. Phys. **10**, 494 (1969).

[2] See D. I. Olive, Phys. Rev. **135**, B745 (1964).

[3] S. Mandelstam, Phys. Rev. **112**, 1349 (1958); R. E. Cutkosky, J. Math. Phys. **1**, 429 (1960); J. C. Polkinghorne, Nuovo Cimento **23**, 360 (1962); **25**, 901 (1962); H. P. Stapp, Phys. Rev. **125**, 2139 (1962).

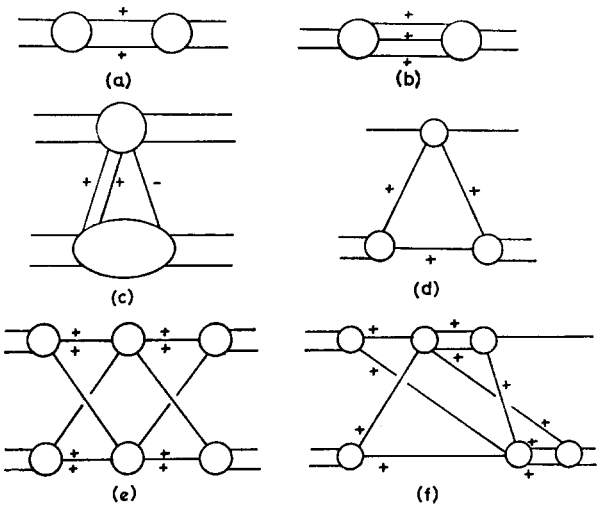


FIG. 1. Examples of Landau diagram; (a), (b), and (f) are non-simple. With n_L and n_F as explained in the text we have (a) $n_F = n_L = 2$, (b) $n_F = n_L = 6$, (c) $n_F = 6, n_L = 2$, (d) $n_F = n_L = 1$, (e) $n_F = n_L = 2$, and (f) $n_F = 8$ $n_L = 4$.

Associated with one such diagram will be the set of Landau equations:

$$\begin{aligned} q^2 - m^2 &= 0 \quad \text{for each internal line,} \\ \sum_k \alpha q &= 0 \quad \text{for each loop } i \end{aligned} \tag{2.1}$$

(where the summation runs over lines carrying k and is measured in the sense of the loop), which implicitly define a Landau curve $L(p) = 0$. If the equations have a solution for physical p (so that the Landau curve enters the physical region for the amplitude) and for real q, then each α is real and (making some choice of over-all factor) has a definite sign, as does each q_0. Corresponding to a particular solution of this sort, we draw a diagram orientated so that positive energy flows in a definite sense, say from right to left, and we label the lines with the signs of the α's. This diagram we call the Landau diagram, and examples are given in Fig. 1. Note that to preserve the sense of energy flow we may be obliged to stretch out a vertex, as in the so-called "pseudothreshold" of 1(c).

Our Landau diagram thus corresponds to a particular solution of the Landau equations rather than to the curve itself. If a diagram has a symmetry that allows internal lines to be permuted without altering its structure (while holding external lines fixed and paying no regard to line labels or sense), this tells us that any new association of q's to internal lines which is effected by the permutation is also a solution to the equations at the same p. These symmetry operations form a group which we call G_F, of order n_F. It is also convenient to define a group G_L, of order n_L, which is the subgroup of G_F which respects both the sense of the internal lines and their labels $(+, -)$. Illustrations appear in Fig. 1.

In what follows we shall assume, for simplicity, that only one Landau diagram corresponds to the Landau curve under consideration, and that the various solutions at a particular p are all related by symmetry operations which are elements of G_F. This restriction is not essential, as we note in Sec. 5.

As explained in I, we can express the Landau curve as

$$L(p) \equiv \sum \alpha(q^2 - m^2) = 0 \tag{2.2}$$

and define a normal variable η by

$$d\eta = \frac{\partial L}{\partial p}\, dp = \left(\sum_p \alpha q\right) dp, \tag{2.3}$$

where $\sum_p$ runs over lines carrying p and a summation over the p's is understood. This variable depends, of course, on the choice of over-all sign for the α's. (If the α's have the same sign, we shall conventionally choose them to be all positive.)

Unitarity Integrals

Unitarity integrals are conveniently represented by diagrams, called *I diagrams*, of which examples are given in Fig. 2. The contribution (also called *I*) that a unitarity integral makes to the right-hand side of the unitarity equation is given by the rules[2]:

(i) A^+ or A^- for each + or − bubble, respectively,
(ii) $-2\pi i\delta^+(q^2 - m^2)$ for each internal line,
(iii) $\int i(2\pi)^{-4}\, d^4k$ for each loop,
(iv) $(n_I)^{-1}$ where n_I is the symmetry number of the diagram,
(v) $-(-1)^{\text{number of minus bubbles}}$.

(2.4)

The symmetry number n_I is defined as the number of permutations of internal lines that leave the structure

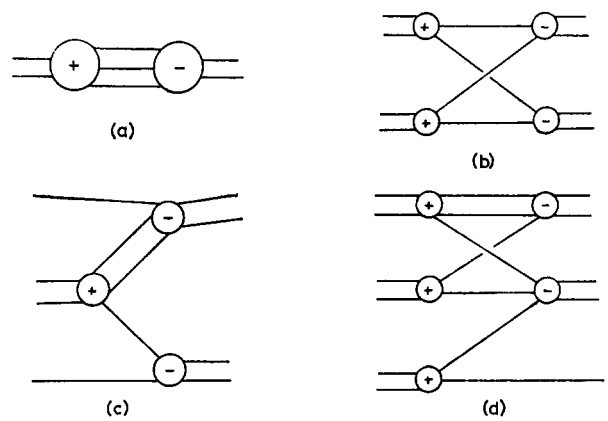


FIG. 2. Examples of *I* diagrams.

of the diagram unchanged, and is made up of factors n for each n-particle intermediate state joining two bubbles.

The integrand thus has δ-functions constraints

$$D(p, k) \equiv q^2 - m^2 = 0 \tag{2.5}$$

and real singularities which are the real singularities of the bubbles. These we assume to be the positive-α parts of Landau curves.[4] They can be written

$$S(p, k) \equiv \sum \hat{\alpha}(q^2 - m^2) = 0, \tag{2.6}$$

where the $\hat{\alpha}$'s are all positive and the q's are the momenta of the lines of the corresponding Landau diagram for the bubble. $\hat{\alpha}$ and q are solutions to the Landau equations for that diagram and so are expressed in terms of the external and loop momenta p and k that appear in I itself. We further assume, as for Feynman integrals, the + and − amplitudes are $\eta + i\epsilon$ and $\eta - i\epsilon$ limits, respectively, on to η real. Hence the singularities S of the integrand obey as $S + i\epsilon$ or $S - i\epsilon$ prescription accordingly as they occur in a + or − bubble, that is, the sign of the prescription is that of the bubble label.

For "simple" singularities (those corresponding to Landau diagrams with only single lines joining any two bubbles) we assume that the discontinuity in η of the + amplitude is given in terms of the corresponding Landau diagram by the Cutkosky rules:

(i) A^+ for each bubble,
(ii) $-2\pi i\delta^+(q^2 - m^2)$ for each line,
(iii) $\int i(2\pi)^{-4}\, d^4k$ for each loop,
(iv) $(n_L)^{-1}$, where n_L is defined above. (2.7)

This agrees with the result obtained in perturbation theory (see I, Sec. 4), if we note that any inequality of n_L and n_F will be compensated by the appearance of δ^+ rather than δ in the rules above. This reduces the number of pinch points counted, according to arguments similar to ones that will be presented in Sec. 4.

In the case when more than one singularity coming from a particular bubble participates in the pinch, it is the multiple discontinuity at their intersection which is relevant. As we see later, this vanishes unless the singularities can be put together in one hinged diagram [such as Figs. 2(c) or 2(d)], in which case the multiple discontinuity is given by the rules above.[5]

3. SINGULARITIES OF UNITARITY INTEGRALS

Unitarity integrals satisfy the appropriate conditions for Theorems 1 and 2 of I. We distinguish four possibilities:

(a) *Generative singularity*, in which the S's coming from a particular bubble can be represented by one diagram for that bubble (so that if more than one S comes from a particular bubble, they are to correspond to the hinged diagram just discussed). In the unitarity diagram I we replace each bubble by the Landau diagram corresponding to its particular participating singularity, if any, and label the new bubbles with the sign of the parent bubble, obtaining what we shall call the *M diagram*. According to I, Sec. 2, the resulting singularity has equation $L(p) \equiv \sum \sigma S + \sum \alpha D = 0$. By (2.2) and (2.6) this is

$$L(p) \equiv \sum \alpha(q^2 - m^2) = 0, \tag{3.1}$$

where the summation extends over the lines of the M diagram, the q being its internal momenta. Further,

$$\alpha = \begin{cases} \alpha & \text{for phase space lines appearing in } M \text{ (zero if they do not participate),} \\ \hat{\alpha}\sigma & \text{for lines that appear in } M \text{ through having been inserted in the bubbles of } I. \end{cases} \tag{3.2}$$

We consider the lines of M as being labeled with the signs of the α's, if nonzero, and with the label o if $\alpha = 0$ (which can only apply to phase-space lines). Then from (2.2) we recognize (3.1) as the equation of the Landau singularity corresponding to the Landau diagram obtained by contracting out the o lines. In Fig. 3 we give examples of the procedure $I \to M \to L$.

(b) *Generative singularities not of type* (a). In this case two or more S's coming from a particular bubble cannot be represented by a single diagram. The resultant singularity cannot therefore be represented by a Landau diagram, and would not appear to be a Landau singularity. Although this sort of pinch is

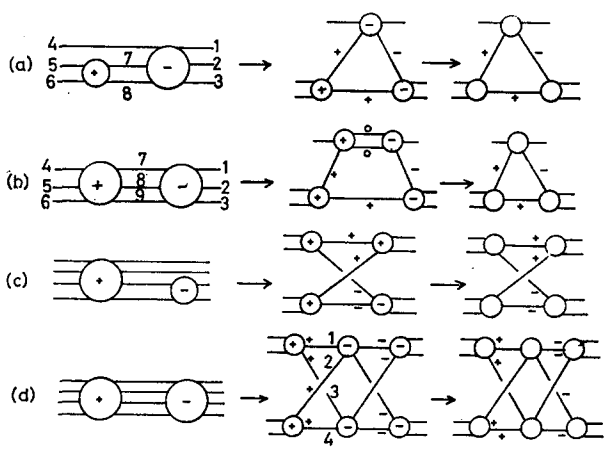


FIG. 3. Examples of the procedure $I \to M \to L$ for generative mechanisms.

[4] This and following assumptions are proved in succeeding papers. Though the proof employs results derived in this paper, it is inductive, and the argument not therefore circular.

[5] This is associated with what Pham (Ref. 7) calls a fiber product.

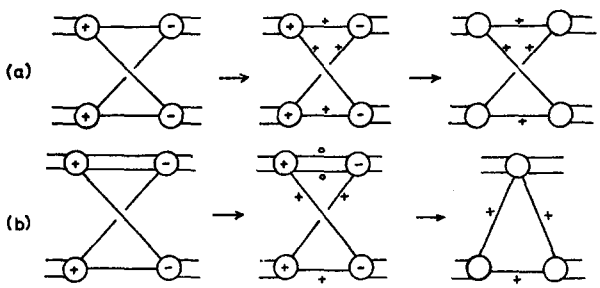


FIG. 4. Examples of the procedure $I \to M \to L$ for explicit mechanisms.

genuine, we shall see that such singularities are spurious as they have zero discontinuity.[6,7]

(c) *Explicit singularity.* This is just case (a) with no S's and hence no insertions, and can lead only to Landau singularities (see Fig. 4).

(d) *Regenerative singularity.* In this case we have a singularity of a bubble which does not depend on the integration variables. Diagrammatically, the effect is that the M diagram is hinged in a characteristic way, as in Fig. 5. We sum up possibilities (a), (c), and (d) in

Rule 1: Physical unitarity integrals may be singular on the Landau curves corresponding to diagrams obtained by inserting some (or no) physical subdiagrams into the bubbles and contracting some (or no) phase space lines.

Natural Distortions

According to the first theorem of I, to be singular a particular mechanism must satisfy

$$\text{sign } \sigma_1\epsilon_1 = \sin \sigma_2\epsilon_2 = \cdots = \text{sign } \epsilon_{\text{nat}}, \quad (3.3)$$

that is, the quantities $\sigma\epsilon$ take the same sign for each singularity, and ϵ_{nat} is an increment of that sign. If the singularity is explicit, we understand $\epsilon_{\text{nat}} = 0$, as this equation suggests. Furthermore, according to Theorem 2 of paper I, the integrals $I_>$ and $I_<$ (evaluated in $\eta > 0$ and $\eta < 0$) are not analytically related if $\epsilon_{\text{nat}} = 0$. For brevity we shall say the mechanism is *singular* + if $\epsilon_{\text{nat}} \geq 0$, and *singular* − if $\epsilon_{\text{nat}} \leq 0$. With this definition explicit mechanisms are both. As the sign of the ϵ_i associated with each S_i in our unitarity integral is simply the label of the bubble, (3.3) implies that, for singularity, sign σ = sign ϵ_{nat} times the label of the bubble. By (3.2), as $\hat{\alpha} > 0$, this is also the line label of any line inserted into the bubble. So we have

Rule 2: The mechanism corresponding to a particular M diagram is

singular + if every line joining + bubbles has label + and every line joining − bubbles has label −,

singular − if every line joining + bubbles has label − and every line joining − bubbles has label +,

and nonsingular otherwise.

Note that although the line labeling of an M diagram has an arbitrary over-all sign, the rule itself is intrinsic because the notion of singular ± also depends on this sign through the definition of η. The mechanism diagrams in Figs. 3 and 5 are singular +.

Effects of Contracting Out o Lines

According to Theorem 1 of I, there is the further necessary condition for singularity that, for nonparticipating D's, $D = 0$. In terms of the M diagram, this means that each o line must satisfy both momentum conservation and $q^2 - m^2 = 0$. First we note that if one o line joins two bubbles, then all other lines joining these two bubbles must be o lines, if L is to be sensible.

Suppose only one o line joins two bubbles. Its momentum is determined by momentum conservation in terms of the other lines of M, and hence, on the Landau curve, in terms of the external momenta p. When $q^2 = m^2$, which is necessary for singularity, the Landau equations which involve this line in addition are also satisfied, but with the corresponding α zero. Thus the Landau curve corresponding to the diagram got by contracting a single o line is nonsingular except at the effective intersection with the higher-order curve corresponding to the participation of the line.

Now consider a case where many o lines join two given bubbles. If k are the loop momenta they form, then the participating S's and D's are independent of k and we have the infinite degeneracy situation considered in Sec. 3 of paper I. Reality and the mass shell conditions for the o lines imply

$$(\sum q)^2 \geq (\sum m)^2,$$

where $\sum$ runs over the o lines. Consequently, singularity occurs only on the part of the Landau curve

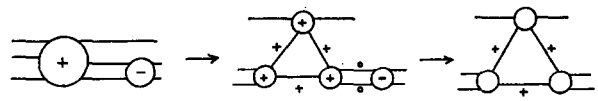


FIG. 5. An example of the procedure $I \to M \to L$ for a regenerative mechanism.

[6] I. T. Drummond, Nuovo Cimento **29**, 720 (1963).
[7] F. Pham, Ann. Inst. H. Poincaré **6**, 89 (1967).

where this is true. The transition between singularity and nonsingularity takes place when the equality holds, in which case the Landau equations for the higher-order curve in which these o lines participate, with zero α's, are also satisfied. Thus we have the antihierarchial effect,[8] mentioned in I, that the singularity of unitarity integrals may be affected by effective intersections with higher-order Landau curves.

These two phenomena are quite awkward in that we can no longer classify singularities solely by the α signs on L. Fortunately, they will each be automatically accounted for by our subsequent discontinuity formula, and need not be explicitly considered.

From our analysis we discover that unitarity integrals are in general singular on different parts of Landau curves from Feynman integrals, so that we cannot say they have global $i\epsilon$ prescriptions, both because the sense in which they are limits on to $\eta = 0$ depends upon the signs of the α's and because they may be singular in regions containing places where the normal variable η may cease to be defined.

4. DISCONTINUITIES OF UNITARITY INTEGRALS

According to the ideas of paper I, we can isolate contributions to a given discontinuity that are due to different mechanisms by choosing suitable R regions to surround the pinch points. The first main discontinuity result of this paper is that the contribution to $\mathrm{dif}_\eta^> I$ arising from a given generative mechanism is given by

$$M_> - (M_<)_{\eta - i\epsilon_{\mathrm{nat}}}, \tag{4.1}$$

where M is an integral obtained from the M diagram corresponding to the generative mechanism by the diagrammatic rules below. $M_>$ is M evaluated in $\eta > 0$ and $(M_<)_{\eta - i\epsilon_{\mathrm{nat}}}$ is the continuation of $M_<$ evaluated in $\eta < 0$ into the region $\eta > 0$ following an $\eta - i\epsilon_{\mathrm{nat}}$ detour at the singularity $\eta = 0$. The rules are

(i) A^+ or A^- for each + or − bubble;
(ii) $-2\pi i\delta^+(q^2 - m^2)$ for each line;
(iii) $\int i(2\pi)^{-4}\, d^4k$ for each loop;
(iv) $(n_M)^{-1}$ where n_M is the symmetry number of the M diagram;
(v) $-(-1)^{\text{number of minus bubbles}}$;
(vi) the integration is restricted to regions in the space of loop momenta of M that are not contracted out. These regions surround the pinch points corresponding to the line labeling of the M diagram.

(4.2)

[8] P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, Nuovo Cimento 43, 444 (1966).

The symmetry number n_M is the number of permutations of the internal lines of M that preserve M, holding external lines fixed and taking into account the labeling of its bubbles and of its lines. The resemblances between these rules and the unitarity and Cutkosky rules (2.4) and (2.7) are of course significant and fortunate for the ease of future analysis. We now derive these rules.

First Steps in the Proof

In paper I (Theorem 3) we showed that the contribution to the discontinuity of the integral arising from a particular generative mechanism characterized by a pinch of singularity surfaces S and constraint surfaces D is given by

$$C_> - (C_<)_{\eta - i\epsilon_{\mathrm{nat}}}, \tag{4.3}$$

where

$$C \equiv \int_{R(k)>0} \prod_i \mathrm{dif}_{S_i}^{\gamma_i} f\, dk, \tag{4.4}$$

where $R > 0$ is a region surrounding the pinch point $\bar{k}$, γ_i are $>$ or $<$ chosen arbitrarily, and f is understood to contain the constraints. In our case f is a product of A^+'s, A^-'s, and mass-shell δ functions. In $S > 0$, $\mathrm{dif}_S^> f$ for an S occurring in a + bubble is got by replacing that A^+ by its Cutkosky discontinuity (2.7). If S occurs in a − bubble, then as − is found by Hermitian conjugation of a + amplitude, so is its dif: $\mathrm{dif}\, A^- = (\mathrm{dif}\, A^+)^\dagger$. Under Hermitian conjugation of the rules (2.7), each A^+ becomes A^-. According to the rules, there is also an imaginary factor $(i)^{l-m}$, where l and m are the numbers of loops and lines of L, which under conjugation becomes $(-i)^{l-m}$, thereby introducing a factor $(-1)^{m-l}$. Now for any connected diagram, $m - l = v - 1$, where v is the number of vertices. Therefore in $S > 0$ $\mathrm{dif}_S^> A^-$ is given directly in terms of the L diagram corresponding to S by the rules

(i) A^- for each bubble,
(ii) $-2\pi i\delta^+(q^2 - m^2)$ for each line,
(iii) $\int i(2\pi)^{-4}\, d^4k^4$ for each loop,
(iv) $(n_L)^{-1}$ as in (2.7),
(v) $-(-1)^{\text{number of minus bubbles}}$.

(4.5)

Since in Eq. (2.6) we adopted the convention $\hat{\alpha} > 0$, it is in the region $S > 0$ that the Cutkosky formula is valid with real contours of integration. Further, the Cutkosky integrals vanish in $S < 0$ since, according to Pham,[7] there is a real vanishing cycle in $S > 0$. Accordingly, we choose to set each γ_i in (4.4) equal to $>$. Inserting (2.7) and (4.5) into (4.4), we see that we do indeed get an A^+ or A^- factor for each (+) or (−)

bubble of the M diagram. Further, each $(-)$ bubble brings an extra factor (-1), so we have understood aspects (i), (ii), (iii), and (v) of (4.2). We go on to discuss (iv) in the next subsection. As for (vi), the restriction on the uncontracted loop momenta appearing on the original integral follows from the R region ideas of I, whilst the fact that the loop momenta which correspond to the use of the Cutkosky formula for dif's of the integrand are similarly restricted is due to the existence of the vanishing cycle already noted. The argument is unaffected by infinite degeneracies associated with multiple o lines; compare the discussion of this point in I.

The Symmetry Factor

The factor $(n_I \prod n_C)^{-1}$ (where n_C are symmetry numbers associated with inserted Cutkosky formulas) for one contribution is not, in general, the n_M^{-1} given by our rules. This is because the M diagram may represent more than one mechanism. For example, in Fig. 3(a) for the equal-mass case L is generated in I not only by the pole at $s_{238} = m^2$, but also by the pole at $s_{237} = m^2$ and the contributions to the discontinuity from the two mechanisms are the same. The numerical factor is therefore twice the $(n_I n_C)^{-1}$ of each contribution, or $2 \times (2 \times 1)^{-1}$, which is n_M^{-1}, n_M the symmetry number of the M diagram, as predicted by the rules. Similarly, in Fig. 3(b) the three generative mechanism in I involving poles in (s_{569} and s_{239}), (s_{568} and s_{238}), and (s_{567} and s_{237}) yield a contribution which when expressed as an integral over just one R region carries a numerical factor $3 \times (6 \times 1)^{-1}$, $= n_M^{-1}$.

For identical particles, symmetries imply that pinch points can occur together in sets—for if there is one mechanism making I singular at a given $\bar{p}$, then there may be many, as in the above examples. Furthermore, as a symmetry in the M diagram reflects a symmetry in the Landau equations, a given set of S's may pinch at several places at once. We have chosen to represent all these symmetry-related pinches by one M diagram; consequently, to find the numerical factor associated with its contribution we shall have to count how many such pinch points there are. The first step is to count the number of mechanisms, that is, the number of distinct insertions of Landau singularities into the bubbles of I that lead to a given M.

Let us number the internal lines of the unitarity diagram I in order that we may distinguish them, and let G_I be the group of permutations of these lines which preserve the structure of I when the external lines are held fixed. The order of G_I is n_I, the number appearing in the rules (2.4). Let us similarly label the corresponding lines of the M diagram and permute these labels according to elements of G_I to get a new diagram. If the permutation is such that the resultant diagram can be rearranged into its original form while holding external lines fixed, then the singularities of the unitarity integrand indicated by these two M diagrams are the same. Examples are (78) in Fig. 3(b) and (13)(24) in Fig. 3(d). Such permutations form a group which we call G_S, a subgroup of G_I. Permutations for which this rearrangement is impossible [e.g., (78) in Fig. 3(a)] lead to different singularities of the unitarity integrand which nevertheless generate the same singularity Γ unless we break our assumption that Γ corresponds to a unique Landau diagram.

The cosets of G_S in G_I are in one-to-one correspondence with the distinct sets of similar singularities of the I integrand giving Γ. By Lagrange's theorem, there are n_I/n_S of these sets. In order usefully to re-express this number n_S we now number all internal lines of the M diagram and define

$\hat{G}_M$ = group of permutations of lines of M leaving it invariant, while holding external lines fixed and respecting bubble labels,

$G_+(G_-)$ = group of permutations of lines of M joining $+$ $(-)$ bubbles that leave M invariant, holding all other lines fixed.

By invariant we mean that the final diagram can be rearranged into the original diagram without permuting any lines.

Then $G_+ \otimes G_-$ is an invariant subgroup of $\hat{G}_M$ and the quotient group $\hat{G}_M/G_+ \otimes G_-$ is isomorphic to the G_S defined above.

According to Bose (or Fermi) statistics, the unitarity integrand is invariant under permutations of phase-space momenta given by S_I, and so the $n_I/n_S = n_I n_+ n_- / \hat{n}_M$ symmetry-related sets of S's contribute equally to the discontinuity across Γ. The discontinuity contributed by one set has a numerical factor $(n_I)^{-1}$ coming from I [see (2.4)] and factors $(n_+)^{-1}$ and $(n_-)^{-1}$ coming from the inserted Cutkosky formulas (2.7) and (4.5). Hence, adding contributions, the over-all numerical factor associated with the M diagram contribution is

$$\frac{1}{n_I}\frac{1}{n_+}\frac{1}{n_-}\frac{n_I n_+ n_-}{\hat{n}_M} = \frac{1}{\hat{n}_M},$$

provided we integrate the M integral over the R regions associated with all the pinch points involving one particular set of S's.

In (4.2) we integrated over what may be a smaller

number of R regions, namely, those associated with pinch points corresponding to a specific line labeling of the phase space lines in M, and we had a numerical factor $(n_M)^{-1}$ rather than the above. In order to get this form, which is useful in later work (paper III of this series, the following paper), we define

G_M = group of permutations of internal lines of M which leave M invariant while holding external lines fixed and taking into account both bubble and line labels (+, −, or o).

Then G_M is a subgroup of $\hat{G}_M$, though not an invariant one, and its left cosets in $\hat{G}_M$ correspond to elements of G_S giving distinct rearrangements of labels (+, −, and o) on the phase-space lines. Thus for each pinch point corresponding to a specific line labeling there is a family of $\hat{n}_M/n_M$ distinct pinch points involving the same S's but different line labelings, as obtained by these cosets. Each member of such a family contributes a numerically equal amount to the total discontinuity, and if we integrate over the R regions associated with just one member of each family, the numerical factor is

$$\frac{\hat{n}_M}{n_M}\frac{1}{\hat{n}_M} = \frac{1}{n_M},$$

as quoted in (4.2).

We have implicitly assumed that all pinch points discussed so far can be treated independently. By a later result dif's for singularities in different generating sets annihilate each other at their intersection, while pinch points for the same S's but corresponding to different line labelings must be distinct, in the cases we are considering.

The pinch points we are left with, namely, those corresponding to a specific line labeling, are permuted among themselves by elements of G_M, but we cannot in general tell whether or not they are distinct, and so do not separate their contributions further.

5. EVALUATION OF $I_> - I_{<\eta-i\epsilon}$

Regenerative and Explicit Mechanisms

In I we saw that the integral formula we have just used to get our first result (4.2) generalizes to the situation in which many generative mechanisms and a regenerative or explicit mechanism operate at the same time in the integral to give (Theorem 4 of I)

$$I_> - I_{<\eta-i\epsilon} = \left(\sum_+ C\right)_> - \left(\sum_+ C\right)_{<\eta-i\epsilon},$$

where

$$\left.\begin{aligned} C &= \int_{R>0} \prod \mathrm{dif}_{S_i}^{\gtrless} f\, dk && \text{for generative mechanisms,}\\ C &= \int \mathrm{dif}_S^{\gtrless} f\, dk && \text{for regenerative mechanisms,}\\ C &= \int_{R>0} f\, dk && \text{for explicit mechanisms,}\end{aligned}\right\} \tag{5.1}$$

and $\sum_+$ is the sum over all singular + mechanisms. We now translate this into M-diagram terms to obtain our final result.

In the case of an explicit mechanism, as in Fig. 4, the M diagram is I itself. M is both singular + and singular −. The numerical factor is n_I^{-1}, and the only symmetries of such an I are found by permuting lines joining the same two bubbles, which must carry the same line labels (in the case where they are not contracted out, this is implied by $\sum \alpha q = 0$). In this case therefore $n_M = n_I$ and the numerical factor is still given by n_M^{-1}.

A regenerative mechanism corresponds to L itself being inserted in a bubble of I to form an M diagram. L is recovered from M by contracting all the phase-space lines, which therefore have label o. The inserted lines will all carry labels + if we take it that we know L is singular only on the positive-α Landau curves. The contribution to the discontinuity is then given by the M-diagram rules, for in this case $n_I n_L = n_M$, while the R-region specification is made automatic because of the existence of a vanishing cycle for the L discontinuity.

Finally, in order to apply Theorem 4 of I, we must confirm that when generative and a regenerative mechanism act together,

$$\mathrm{dif}_L \prod_i \mathrm{dif}_{S_i} f = 0.$$

Since $\prod_i \mathrm{dif}_{S_i} f$ is given by the Cutkosky rules, the nonvanishing of the expression above would be equivalent to saying that the M diagram for the generative process has L as a regenerative singularity. This would mean that the diagrams M and L could be hinged, and indeed that L could be hinged with itself, since M goes into L by contracting internal lines. The two vertices of L which are put together in the hinging process must have external lines exhausting all external lines of L, if the hinging is to be possible. This cannot be so unless L has only two vertices with external lines; the result thus follows for all other graphs, and, in particular, for all simple graphs.

To sum up,

$$I_{>} - I_{<\eta-i\epsilon} = \left(\sum_{+} M\right)_{>} - \left(\sum_{+} M\right)_{<\eta-i\epsilon}, \quad (5.2)$$

where the sum runs over all singular + M diagrams, which are the diagrams got by inserting some (or no) + lines into + bubbles and some (or no) − lines into − bubbles of I, which, by contracting some phase space lines (those labeled o), give L, and whose remaining lines are labeled with the sign they have in L.

The Spurious Singularities

We now consider case (b) of Sec. 3, when one of the bubbles A of I contributes two singularities, say S_1 and S_2, to a given generative mechanism. By (5.1) the resulting contribution to the discontinuity has a factor $\mathrm{dif}_{S_1} \mathrm{dif}_{S_2} A$ in the integrand which is integrated over a neighborhood of the intersection of S_1 and S_2. This is nonzero only if $\mathrm{dif}_{S_2} A$, which is a Cutkosky integral, is singular on S_1. Since S_1 is a Landau singularity and since our previous results apply equally well to Cutkosky integrals, we consider the possible M diagrams by which this could be so.

Consider the possibility of a generative mechanism. Since $\mathrm{dif}_{S_2} A$ is evaluated close to S_2, which is the positive-α arc of a Landau curve, the internal momenta very nearly satisfy the Landau equations. The same applies to the corresponding lines of the M diagram. Hence the internal momenta of the M diagram must satisfy two different sets of Landau equations, corresponding to S_1 and S_2. Then a third set of Landau equations is also satisfied, namely, those of M itself, and so if the double discontinuity is nonzero because of a generative mechanism, we are on another Landau curve, a situation we have excluded from consideration.

Similar arguments apply to explicit mechanisms, while if S_1 is a regenerative singularity of $\mathrm{dif}_{S_2} A$, then the M diagram is a hinged diagram, and we have case (a) of Sec. 3, which has already been considered.

6. COMMENTS

Notice that our formulas (5.2) remain true when there is the form of nonsingularity, discussed in Sec. 3, associated with the presence of o lines in M. This occurred because the nonparticipating δ-function constraints could not be satisfied simultaneously with the Landau equations. By virtue of the R region, the momenta of the M integral corresponding to + and − lines are constrained to lie close to the pinch values given by the Landau equations; by the same token it is impossible to find real momenta which also satisfy the δ-function constraints corresponding to the o lines. Hence the integrand has zero support and the M functions $M_{>}$ and $M_{<}$ both vanish.

This means that an important simplification has occurred. Although the arcs of possible singularity are not solely classified by the α assignments of the lines of L, in that transitions between singularity and nonsingularity occur at effective intersections with higher-order curves, nevertheless the form of the discontinuity formula which is applicable can be classified according to the α signs of L.

Finally, we recall our assumption that our Landau corresponds to just one Landau diagram L. If, as may be the case, it corresponds to several such diagrams, the modification to our procedure is obvious, and we add separately the M diagram contributions corresponding to each L.

S-Matrix Singularity Structure in the Physical Region. III. General Discussion of Simple Landau Singularities*

M. J. W. BLOXHAM, D. I. OLIVE, AND J. C. POLKINGHORNE
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England

(Received 1 August 1967)

Drawing on the understanding of unitarity integrals acquired in previous papers [J. Math. Phys. **10**, 494, 545 (1969)], we show that unitarity demands that connected amplitudes be singular on the positive-α arcs of all "simple" Landau curves in their physical regions. Further, we show that the amplitudes are nonsingular on mixed-α arcs, while on the positive-α arcs their discontinuities are given by the Cutkosky rule. This confirms arguments from perturbation theory and demonstrates how a weak analyticity assumption can generate in an exact way a singularity scheme relating to causality.

1. INTRODUCTION

In this paper we apply the results of the preceding papers[1,2] of the series to determine the physical-region singularities of connected S-matrix amplitudes, showing them to be the positive-α arcs of Landau curves with associated discontinuities given by the Cutkosky formula. We restrict our attention to singularities corresponding to simple Landau curves, that is, curves whose Landau diagrams have no more than one line joining any two bubbles. The discussion of the nonsimple case requires refinements of technique which we postpone to a later paper.

The procedure is a generalization of one already carried out successfully in the study of particular examples.[3] It consists of comparing the physical unitarity equation holding on one side of a Landau curve with an analytic continuation of the unitarity equation holding on the other side. The two are, in general, different because of the mechanisms which make unitarity integrals singular, which were discussed in II. From their simultaneous validity in Sec. 2 we are able to deduce an equation of the form

$$\sum_R (\mathrm{dif}\, A)(1 - B^-) = \sum_{G,E}^{+} M_{>} - \sum_{G,E}^{+} M_{<\eta - i\epsilon}. \quad (1.1)$$

The terms on the right involve known quantities arising from the various generative and explicit mechanisms operating in the unitarity integrals, while the terms on the left arise from regenerative singularity mechanisms and involve unknown quantities, the discontinuities of several amplitudes A across singularities related to the Landau curve under consideration. The number of terms to be included in the original unitarity equation depends upon the values of external momenta, and this is also true of (1.1), but the argument is phrased in a way that will take this fact into account.

The first step in solving (1.1) is to simplify the right-hand side by means of a cancellation theorem obtained in Sec. 3 which states that a certain sum of M terms adds to zero. The second step (Sec. 4) is to invert the operators on the left-hand side of (1.1) and obtain a unique solution for dif A, the required discontinuities. Section 5 deals with alternative continuation procedures and the relaxation of certain assumptions made earlier. Section 6 sums up our conclusions.

We wish to emphasize two points about our argument, one methodological, the other more fundamental. To evaluate the right-hand side of (1.1) we need to know discontinuities of amplitudes across singularities corresponding to subdiagrams of the diagram under discussion. These subdiagrams involve fewer loops and their properties have been determined at an earlier stage of the inductive argument. To invert the equations we consider successive intervals between thresholds in the total energy, and since the energy at which the discontinuity of a subdiagram is evaluated never exceeds the total energy, we can indeed follow a well-defined induction procedure.

* The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research under Grant AF EOAR 65-36 through the European Office of Aerospace Research (OAR), United States Air Force.

[1] M. J. W. Bloxham, D. I. Olive, and J. C. Polkinghorne, J. Math. Phys. **10**, 494 (1969). We shall call this paper I.

[2] M. J. W. Bloxham, D. I. Olive, and J. C. Polkinghorne, J. Math. Phys. **10**, 545 (1969). We shall refer to this paper as II.

[3] (a) D. I. Olive, Nuovo Cimento **29**, 326 (1963); P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, J. Math. Phys. **7**, 1593 (1966); (b) D. I. Olive, Phys. Rev. **135**, B745 (1964); R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, London, 1966); (c) P. V. Landshoff and D. I. Olive, J. Math. Phys. **7**, 1464 (1966); (d) M. J. W. Bloxham, Nuovo Cimento **44**, 794 (1966); (e) P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, J. Math. Phys. **7**, 1600 (1966); J. K. Storrow, Nuovo Cimento **48**, 593 (1967). (a) and (b) deal with normal thresholds and poles, respectively, which arise only through explicit and regenerative mechanisms. (c) and (e) deal with triangle singularities arising only through generative and regenerative mechanisms, while (d) deals with a singularity arising through all three mechanisms. Our analysis in this paper generalizes all these arguments except for those of (a) and (e), which deal with non-simple graphs.

Finally, we emphasize that our discussion uses only unitarity, connectedness structure, and a relatively weak analyticity assumption. The latter is simply that the physical amplitudes defined in any neighborhood of the physical region are analytically related, and that the $\eta + i\epsilon$ prescription provides the correct analytic relation for A^+ across the positive-α arcs of Landau curves (which can actually be proved in specific examples). No use is made in this work of crossing or Hermitian analyticity.

2. THE ANALYTIC CONTINUATION OF UNITARITY RELATIONS

Paper II indicated that unitarity integrals are singular on Landau curves. We therefore consider one such Landau curve Γ which corresponds to a Landau diagram L which is simple, as defined in the introduction, and not hinged.[4] We consider points of Γ which lie on no other Landau curves and which are not themselves nodes. Our technique is to compare the physical-unitarity relations holding one side of Γ with the analytic continuation of physical unitarity from the other side of the curve. In making the comparison we shall only continue the infinitesimal distance necessary to cross the curve.

Specifically, we shall consider the $\eta - i\epsilon$ continuation of unitarity written in the $SS^\dagger$ form, that is, the form in which $-$ (i.e., "minus") bubbles appear to the right of $+$ bubbles. For the moment we assume that $\eta - i\epsilon$ is the correct path of continuation for the $-$ bubbles. Later we show that arguments which were based on $\eta - i\epsilon$ or $\eta + i\epsilon$ continuation for either the $SS^\dagger$ or $S^\dagger S$ form would lead to the same singularity structure, and that in fact no assumption need be made as to the correct analytic continuation for the amplitudes, except on positive-α arcs. The unitarity relations are of the form

$$A(+) - A(-) = \sum_\alpha I^\alpha_>, \quad \eta > 0, \tag{2.1a}$$

$$A(+) - A(-) = \sum_\alpha I^\alpha_<, \quad \eta < 0, \tag{2.1b}$$

where, on the right-hand side, we sum over all integrals in the unitarity relation for given external momenta. If there are new terms in unitarity which appear as we cross Γ (corresponding to explicit singularities), the corresponding $I^\alpha_<$ will be zero, but this is taken account of in the formalism of II. If (2.1b) is now continued across Γ in an $\eta - i\epsilon$ sense, $A(-)$ will continue into itself but $A(+)$ will be taken on to the wrong sheet of Γ, where we denote its value by $A(i)$. Subtracting the continued equation from (2.1a) gives

$$A(+) - A(i) = \sum_\alpha (I^\alpha_> - I^\alpha_{<\eta - i\epsilon}). \tag{2.2}$$

The left-hand side is what we wish to evaluate. If it is zero, Γ must in fact be nonsingular; if it is nonzero, Γ is singular and (2.2) evaluates the discontinuity across it. According to (5.2) of II, (2.2) can be rewritten as

$$A(+) - A(i) = \sum_\alpha \overset{+}{\sum} (M^\alpha_> - M^\alpha_{<\eta - i\epsilon}), \tag{2.3}$$

where the sum $\overset{+}{\sum}$ extends over all the singular $+$ mechanism diagrams associated with the singularity of the integral I^α on Γ. Note that, as explained in II, this sum may include terms with o lines which represent integrals that in fact vanish over a range of external momenta. An example of this is provided by single-particle o lines: an M diagram containing such a line contributes nothing except at an effective intersection with the Landau curve whose diagram includes this line. Similarly, the formalism automatically takes antihierarchy into account, the switch from singularity to nonsingularity of an integral being matched by the vanishing of the contribution from one or more M diagram terms. Thus, no explicit account need be taken of this phenomenon.[5] It will sometimes be convenient for the manipulations not to remove such diagrams from (2.3), even though their contributions vanish.

The Iterative Procedure

We now examine more carefully what information is used in evaluating the right-hand side of (2.3). In obtaining the explicit terms nothing is assumed. In obtaining the generative terms we have assumed Cutkosky discontinuity formulas for the singularities participating in the generative mechanism. Such a singularity corresponds to a subset of the internal lines of the diagram L whose Landau curve we are discussing. This defines a partial ordering of singularities which allows the Cutkosky assumption to be made inductively. To evaluate the regenerative contributions, on the other hand, we would need the Cutkosky formulas for a set of singularities with the same internal lines as L and with different numbers of external lines attached to the bubble of L in ways

[4] For any simple Landau diagram there may be other diagrams leading to the same Landau curve which are found by replacing single-particle lines by pseudothreshold sets $(2m - m)^2$, etc., if energy–momentum conservation permits. These further diagrams, however, correspond to pinch points of the unitarity integrals which are distinct from those of the simple Landau diagram under discussion. Their contributions can, therefore, be considered separately, though we defer this until paper IV of this series. However, we note that, since pseudothresholds necessarily correspond to mixed α's, these further diagrams are expected to cancel in any unitarity equation and so not affect the amplitude.

[5] A particular example of the effect has recently been discussed by D. Branson, Nuovo Cimento **54A**, 217 (1968).

which are consistent with the given external momenta. These singularities must be considered simultaneously with L, and it obviously is not permissable to assume a Cutkosky formula for them without proof. We shall simply assume that each singularity has the same $i\epsilon$ prescription in the appropriate amplitude A and write its discontinuity as an unknown quantity dif A. Returning to (2.3), this means that if L has positive α's, we must replace the Cutkosky form in each regenerative M diagram by a connected bubble labeled dif A where A is the relevant amplitude. If the M diagram is CB^-, where C represents the Cutkosky form and B^- the remainder of the diagram, then we replace it with (dif A)B^-. If L has mixed α's, we expect we shall find it to be nonsingular, and there are nonsingular regenerative M diagrams in (2.3) according to the definitions of these M diagrams given in II. However, as we cannot assume L to be nonsingular from the outset, we must retain the set of regenerative terms (dif A)B^- exactly as in the positive-α case. Finally, we note that regenerative and generative mechanisms in the same I^α do yield a sum of M integrals for each mechanism separately, as $\mathrm{dif}_L \mathrm{dif}_S A = 0$ with S a singularity participating in a generative singularity. The argument is given in II. The equation we have to discuss in the inductive argument is therefore

$$\mathrm{dif}\, A - \sum_\alpha \sum_R (\mathrm{dif}\, A)B^- = \sum_\alpha \sum_{E,G}^{+} (M^\alpha_> - M^\alpha_{<\eta-i\epsilon}), \tag{2.4}$$

where the sum may run over many A's and B^-'s.

Partitions and Explosions

We can regard the M diagram as being partitioned, the internal lines cut by the partition being the phase space lines of the original unitarity integral.[6] According to II, the mechanism diagrams which are singular + for L have the following properties:

(i) Each line cut by the partition bears positive energy from right to left and is labeled with a +, −, or o;

(ii) internal lines and bubbles to the left of the partition carry + labels, while those to the right of the partition carry − labels (this is for the $SS^\dagger$ form of unitarity);

(iii) contraction of o lines and omission of bubble labels gives the diagram L.

These properties are valid regardless of whether the mechanism concerned is generative, explicit, or regenerative. Indeed, for a given L we may define M diagrams as labeled diagrams with associated partitions satisfying the properties above. Any such diagram must be an M diagram in the original sense, because, by contracting all lines not cut by the partition, we obtain the unitarity integral which would possess the singularity mechanism. Further, it is easy to recognize whether the mechanism is generative, regenerative, or explicit.

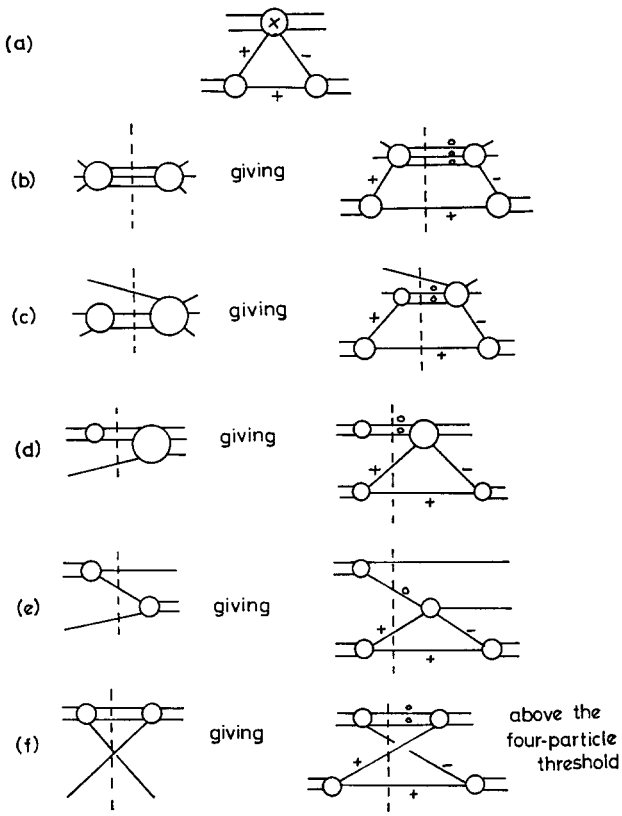


FIG. 1. A diagram (a) together with possible explosions (b)–(f) of the vertex X. The dotted line is the partition.

We can reverse this procedure and enumerate all possible M diagrams by introducing the concept of explosions at vertices of the L diagram. An *explosion* of a vertex is given by replacing it with a + bubble or − bubble or any other term occurring in the unitarity equation for the process represented by the bubble at the vertex (with o labels on all the new lines introduced). Having done this at each vertex, we retain those diagrams which satisfy the conditions above and which are compatible with energy–momentum conservation for the values of external momenta considered. Figure 1 illustrates the explosion procedure.

Disposable Vertices

A particularly important type of vertex in the original Landau diagram is that at which all the incoming

[6] We shall allow a partition which cuts only external lines, classing this as a regenerative M diagram. It corresponds to minus the dif A term on the left of (2.4).

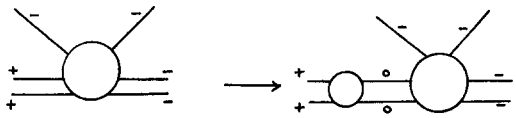


FIG. 2. A vertex which has a unique exploded version (below the three-particle subenergy threshold).

internal lines are − lines and all the outgoing internal lines are + lines. We call such vertex a *disposable vertex* because it will be possible to form mechanism diagrams in which the partition passes either to the right or to the left of the vertex. Moreover, in such a case there will also be diagrams in which the partition passes through lines introduced by exploding the vertex, and the set of all possible explosions will exactly correspond to the set of all terms in the unitarity equation appropriate to the process represented by the original vertex. For this reason disposable vertices play a vital role in the cancellation effects discussed in the succeeding section.

Explosions can also be made at vertices which are not disposable; an example is given in Fig. 2. In these cases, however, the set of possible explosions is limited, and it is not possible to have M diagrams with partitions running to either side of the vertex, though, as Fig. 2 illustrates, it may be possible to displace the partition to one of the sides. We call these vertices *explodable vertices*. Their existence can lead to possible partitions which are not obvious on inspection of the original Landau diagram. An example is given in Fig. 3.

3. THE CANCELLATION THEOREM

The right-hand side of Eq. (2.4) can now be read as a sum over all possible M diagrams that correspond to singular + generative and explicit mechanisms, without reference to the particular unitarity integrals I^{α} in which they arise. To simplify this sum we prove the following cancellation theorem:

$$\sum_{G,E,R}^{+} M = 0. \tag{3.1}$$

That is, the sum of all the M integrals corresponding to singular + M diagrams for L vanishes when the

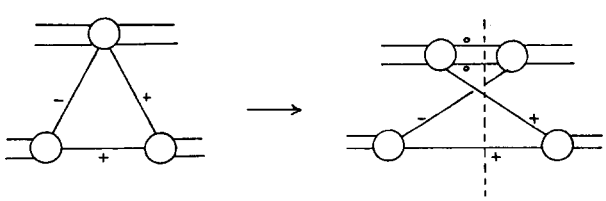


FIG. 3. A Landau diagram with an associated partitioned mechanism diagram made possible by the existence of an explodable vertex.

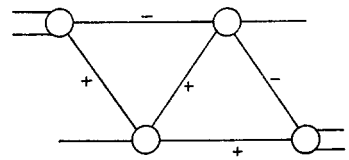


FIG. 4. A Landau diagram without a disposable vertex.

integration regions are suitably standardized in the way discussed below. The result holds on either side of Γ, that is, for $M_>$ or $M_<$. The basic idea is that all unitarity terms appear at a disposable vertex, and their sum vanishes by the unitarity equation valid for that vertex. The first step is to show that in the physical region the Landau diagram always contains at least one disposable vertex. At first sight this might not seem to be so, for it is easy to write down Landau diagrams which do not have this property; Fig. 4 is a simple example. However, the Landau equations cannot be satisfied by real internal momenta in the physical region with the α assignments of Fig. 4. This is because the assignment of positive energy to the internal momentum vectors directed as shown prevents the condition

$$\sum \alpha q_0 = 0 \tag{3.2}$$

being satisfied for the upper triangular loop. This generalizes in the following way. In the Landau diagram suppose we choose a − line and follow to the left through a sequence of − lines as far as is possible. The sequence stops at a vertex at which the only outgoing lines are either + or external lines. If there is no disposable vertex, then there must be at least one + line which is incoming at this vertex. We then follow back to the right through this and succeeding + lines as far as possible. The chain ends at a vertex at which the incoming lines are all − lines or external lines. If there is no disposable vertex there must be at least one outgoing − line at this vertex. This we follow to the left, and so on. Eventually, because of the finite number of lines in the diagram, the sequence must close to form a loop. By construction the momenta of this loop do not satisfy the $\sum \alpha q_0 = 0$ condition as each term is strictly negative. We conclude that for possible arcs of Landau curves in the physical region there must be a disposable vertex.

Standardizing the R Region

In order to obtain the cancellation (3.1) it is necessary that all the integrals corresponding to each M diagram should be taken over the same region of integration. We first show that this is possible.

Different M diagrams will give different status to a loop momentum k which occurs in the Landau diagram. For some it will be a loop momentum in the

corresponding unitarity integral, in which case, as explained in I and II, its region of integration is confined to a small real arbitrary region R surrounding the pinch point (or pinch points). In other mechanism diagrams k may correspond to an integration in a Cutkosky formula inserted into the unitarity integrand. Its range of integration in that case is not confined by R region considerations but, as was noted in II, the fact that there is a real vanishing cycle on positive arcs means that in this case also the integration is in fact confined to a real neighborhood of the value (or values) of k corresponding to the solution of the Landau equations for L. The loop momenta which must be contracted out to form L from M have an unrestricted region of integration since they correspond to an infinite degeneracy situation as discussed in I.

It is therefore possible to choose a standardized R region for the momenta appearing in L which is the same for all M diagrams, the remaining momenta being unrestricted. In fact, this standard R region can be taken to be the same as an R region for the Feynman integral associated with L. Of course, the unitarity integrals have more singularities than the Feynman integral, but these are irrelevant—provided we are near a point on a single Landau curve, since in that case no other singularities can intersect the pinch point.

The Cancellation

Let us imagine that the integrations over the loops to be contracted out are carried out first. These are unrestricted and hence are genuine unitarity integrations, dependent only upon the momenta carried by the remaining lines of the M diagrams. The cancellation actually occurs before the remaining integrations are performed, so let us suppose the momenta borne by the + and − lines take some fixed value lying in the R region discussed above. Initially we consider the case when L has associated symmetry number (n_L) unity so that there is no ambiguity as to which lines of the M diagram correspond to which lines of the L diagram.

The explosions permissable at each L vertex depend solely on the momenta which we have now fixed, and so are independent of each other. We select a disposable vertex V and group the terms of (3.1) into sets corresponding to diagrams which have the same explosions at all other vertices. Our aim is to show that unitarity at V will cause the sum of contributions from each such set to vanish. If there is more than one disposable vertex, the choice of which to use is arbitrary. According to the rules of M-diagram contributions given in II, when the partition passes to the left of V the term to be inserted at V is $-A^-$, where A is the amplitude for that vertex, and when it passes to the right of V, the appropriate term is A^+; when the partition runs through lines introduced by exploding V, the term is minus the appropriate contribution to the right-hand side of the unitarity equation for the bubble at V. Thus we get precisely all the terms in the unitarity equation valid for the momenta incident on vertex V, because all such terms give genuine singular $+M$ diagrams. All the factors corresponding to the other lines and vertices are common factors within a given set and hence we can indeed use unitarity at V to get the set to cancel to zero. Examples are given in Figs. 5 and 6. Hence the whole of (3.1) does cancel to zero, at least if the numerical factors are correct. These we now proceed to discuss.

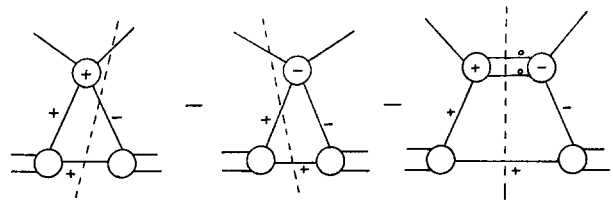
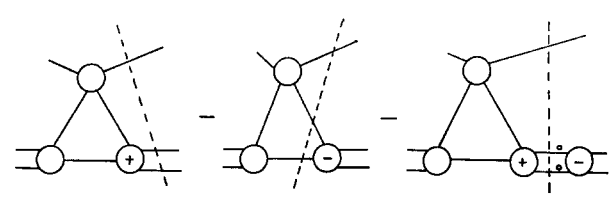


FIG. 5. A simple example of the cancellation mechanism on a mixed-α arc. All contributions are generative.

The Case When L Has Symmetries

In this case there will be more than one way of identifying M lines with L lines.

Let X be the diagram obtained from the M diagram by contracting out o lines and labeling the fused vertices to indicate the unitarity term applicable there. Thus X differs from the diagram L only in having its bubbles labeled. Following II, let G_M be the group of permutations of internal lines of M that leave it invariant, while respecting line and bubble labels, and let G_0 be the subgroup permuting only o lines. Then G_0 is an invariant subgroup of G_M and the quotient group G_M/G_0 is isomorphic to G_X, the group of permutations of internal lines of X that leave X invariant while respecting line and bubble labels. Now G_X is itself a subgroup (though not an invariant



(a) (b) (c)

FIG. 6. A simple example of the cancellation mechanism on a positive-α arc. (a) Regenerative: the left-hand side of Eq. (2.3); (b) generative; (c) regenerative.

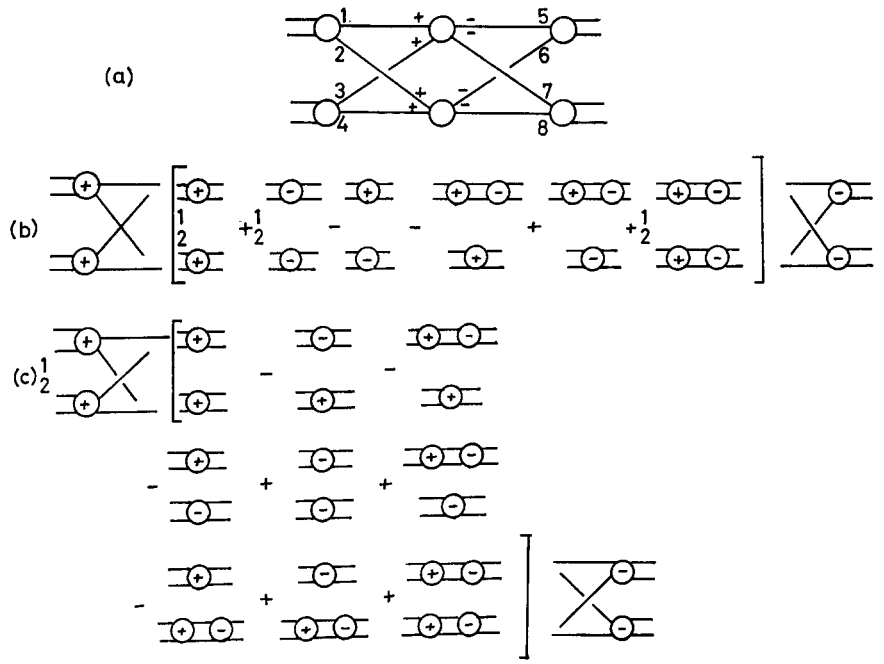


FIG. 7. An L diagram with symmetry factor two and M diagram contributions.

one) of G_L, the group of permutations of internal lines of L that leaves L invariant, respecting line labels, and the cosets of G_X in G_L correspond to orientations of the diagram X which differ if the internal lines are identified, but not otherwise.

In the cancellation argument we want all these different orientations so as to ensure that, once the lines are identified, each vertex indeed appears with all possible explosions. Therefore we rewrite a particular M diagram as the sum over the n_L/n_X orientations of itself, compensating for this by a numerical factor n_X/n_L. That is, we write

$$\frac{1}{n_M}\int_R M(q) = \frac{1}{n_M}\frac{n_X}{n_L}\int_R \sum_i M(p_i q), \qquad (3.3)$$

where $\sum_i$ indicates the sum over the n_L/n_X cosets of G_X in G_L and P_i is an element of the ith such coset. (It is irrelevant whether or not $P_i q = q$.) We have used the fact that the integration region R is invariant (by choice) under G_L. Since $n_X = n_M/n_0$, the over-all numerical factor associated with each oriented term is not n_M^{-1} but $(n_0 n_L)^{-1}$, which is just right, as n_L^{-1} is common to all the terms and n_0^{-1} is the numerical factor appropriate to the unitarity integrals associated with the exploded vertices. We can now pick out the values of the integrands for chosen $q \in R$ and cancel as before, recognizing a specific vertex by the lines attached to it.

As an illustration, we consider the L given in Fig. 7(a). This has $n_L = 2$ because of the symmetry (12)(34)(56)(78), which effectively interchanges the two disposable vertices. In Fig. 7(b) we have written down all the singular $+M$ diagram terms appearing below the 5-particle threshold, writing the numerical factor associated with the possibility of the above permutation explicitly, i.e., $\frac{1}{2}$ if the disposable vertices have identical labels and 1 otherwise. Using (3.3), we rewrite these terms as in Fig. 7(c). Now we see that both the rows and the columns in the array cancel, corresponding to whether we choose to cancel on the top or bottom disposable vertex.

4. INVERSION AND UNIQUENESS

We recall the equation derived so far:

$$\sum_R (\mathrm{dif}\, A)(1 - B^-) = \sum_{E,G}^{+} (M_> - M_{<\eta - i\epsilon}), \qquad (4.1)$$

$$\sum_{E,G,R} M = 0 \quad \text{in} \quad \eta > 0 \quad \text{or} \quad \eta < 0. \qquad (4.2)$$

Using (4.2) with $\eta > 0$ and $\eta < 0$, (4.1) becomes

$$\sum_R \mathrm{dif}\, A(1 - B^-) = -\sum_R^{+} (M_> - M_{<\eta - i\epsilon}).$$

Now for positive α's, the contribution from a regenerative mechanism diagram has the form CB, with C a Cutkosky integral. $C = 0$ in $\eta < 0$ because of the existence of the vanishing cycle. So

$$\sum_R \mathrm{dif}\, A(1 - B^-) = -\sum_R^{+} M_> . \qquad (4.3)$$

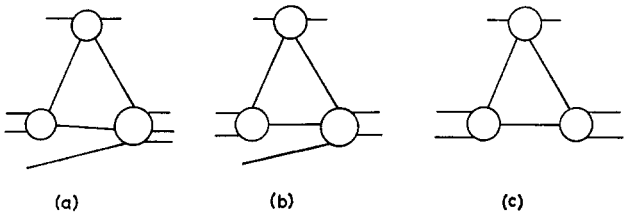


FIG. 8. Landau diagrams discussed in this section.

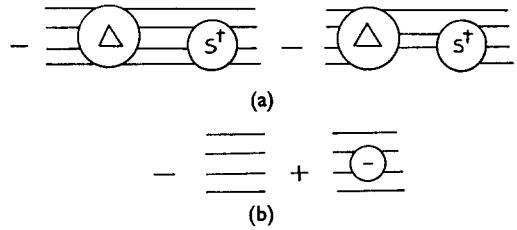


FIG. 10. Collections of terms appearing in Fig. 9.

This equation also holds for mixed α's, as there are no singular + regenerative M diagrams, as we noted earlier, and the right-hand side is actually zero. Notice how the terms $M_{<\eta-i\epsilon}$ have disappeared. We have not had the problem of examining the details of the analytic continuation because the terms appeared in combination whose sum vanished in $\eta < 0$.

We now note that (4.3) is consistent with what we wish to prove. The terms on the left-hand side were defined in Sec. 2 exactly by the property that if we replace each dif A by the Cutkosky formula $C_>$ on the positive-α arc, then

$$\sum_R \text{dif } A(1 - B^-) \quad \text{becomes} \quad -\sum_R^+ M_> .$$

The Cutkosky formula is thus a manifest solution to (4.3) in the positive-α case.[7] On the mixed-α arcs our anticipated solution dif $A = 0$ also satisfied (4.3), as both sides vanish. It remains only to be shown that these solutions are unique.

Let us define new bubbles, labeled Δ, which represent dif A^+ where L has mixed α, and dif $A^+ - C$ near positive-α arcs. Then our result (4.3) can be summed up as

$$\sum_R \Delta(1 - B^-) = 0, \tag{4.4}$$

where each term is obtained from a positive-α singular + regenerative M diagram by replacing the Cutkosky formula part with a Δ bubble and several Δ bubbles with different numbers of external lines may appear

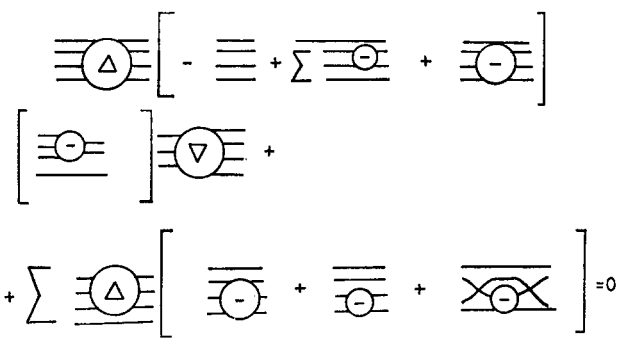


FIG. 9. Equation (4.4) for Fig. 8(a).

in the sum. We wish to conclude from (4.4) that each $\Delta = 0$.

Let us write out in full Eq. (4.4) for the specific example of the analysis of the singularity associated with Fig. 8(a). This is done in Fig. 9. The Δ bubbles in the equation correspond to Figs. 8(a), 8(b), 8(c), respectively. As this example illustrates, either all left-hand lines are connected to a Δ bubble or only some of them are. This is in contrast with right-hand lines of Δ, which may be greater in number than the incoming lines if there is sufficient energy for the relevant intermediate state. Δ's, which have the same number of outgoing lines as the original unitarity equation, appear attached to operators $(1 - B^-)$ which sum to give matrix elements of $-S^\dagger$. For example, the first two terms in Fig. 9 may be rewritten as in Fig. 10(a). Δ's which appear with fewer left-hand lines do not have this property. For example, in the third term of Fig. 9 we lack the factors shown in Fig. 10(b), because these would not give a connected contribution. However, the fact that at least one outgoing momentum line is not attached to the Δ bubble means that it is evaluated in an energy range which will already have been dealt with at an earlier stage in the argument. Such Δ bubbles are thus already known to be zero. Consequently, the surviving terms in (4.4) take the form

$$\sum_i \Delta_i S^\dagger(0_1) S^\dagger(0_2) \cdots S^\dagger(0_n) = 0, \tag{4.5}$$

where the right-hand lines of each Δ bubble have been formed into sets $0_1 \cdots 0_n$ according to the vertex of the Landau diagram at which they are incident [for Figure 8(a), 0_1 is the top right-hand line, 0_2 the remaining three]. $S^\dagger(0_i)$ operates on the subspace of the lines 0_i and the sum includes all phase-space integrations that correspond to possible intermediate states, as in Fig. 10(a). The Δ_i's in (4.5) have the same left-hand and internal lines, but have different numbers of lines in the sets 0_i, corresponding to the different intermediate states. Equation (4.5) was derived from the unitarity equation for a particular

[7] A special case of this argument has been given for the pole by H. P. Stapp in *High-Energy Physics and Elementary Particles* (IAEA, Vienna, 1965); J. Math. Phys. **9**, 1548 (1968).

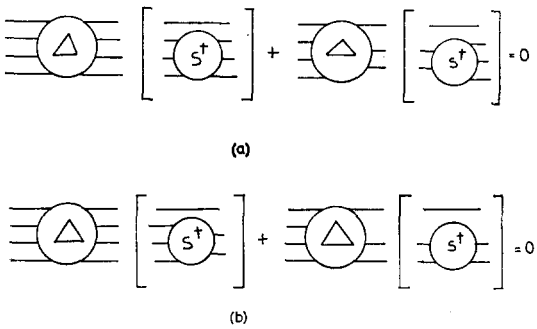


FIG. 11. Equations (4.5) for the singularities of Figs. 8(a) and 8(b).

amplitude, and the Δ bubble for this amplitude will be one of the Δ_i that appear. If we start from unitarity equations for amplitudes corresponding to each of the Δ_i, we obtain a closed set of equations like (4.5) which can readily be solved by postmultiplication by $S(0_i)$'s and the use of unitarity to give the unique solution $\Delta_i = 0$, each i.

For example, the surviving terms of Fig. 9 may be written as in Fig. 11(a). A similar analysis applied to the singularity associated with Fig. 8(b) yields Fig. 11(b). Multiplication of the expression in Fig. 11(a) on the right by the expression in Fig. 12(a) and of the expression in Fig. 11(b) by the expression in Fig. 12(b) and adding yields, from the application of unitarity, the equation in Fig. 12(c).

As the sequence of argument is well defined and the inversion problem now explicitly solved, our result holds for all the simple Landau singularities in the physical region of any amplitude.

5. OTHER CONTINUATIONS

So far we have only considered the $\eta - i\epsilon$ continuation of unitarity written in the $SS^\dagger$ form. We could also consider the $\eta + i\epsilon$ continuation of this form,

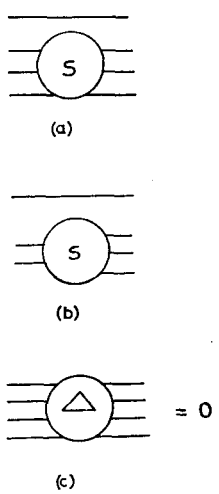


FIG. 12. Quantities relevant to the manipulation of Fig. 11.

or either $\eta \pm i\epsilon$ continuation of the $S^\dagger S$ form in which + bubbles lie to the right of − bubbles. We now discuss all these possibilities, retaining for the moment the supposition that $\eta + i\epsilon$ is the correct path of continuation for + bubbles. We do not wish to recapitulate the complexities of the preceding discussion, so we write the equations in a symbolic way which adequately displays their structure. The four possible equations are

$$\eta - i\epsilon, \quad SS^\dagger: \quad \sum \text{dif}\, A^+(1 - B^-) = \sum_{GE}^{+} \text{dif}\, I, \tag{5.1a}$$

$$\eta + i\epsilon, \quad SS^\dagger: \quad \sum -(1 + B^+)\, \text{dif}\, A^- = \sum_{GE}^{-} \text{dif}\, I, \tag{5.1b}$$

$$\eta - i\epsilon, \quad S^\dagger S: \quad \sum (1 - B^+)\, \text{dif}\, A^+ = \sum_{GE}^{+} \text{dif}\, I', \tag{5.1c}$$

$$\eta + i\epsilon, \quad S^\dagger S: \quad \sum -\text{dif}\, A^-(1 + B^+) = \sum_{GE}^{-} \text{dif}\, I', \tag{5.1d}$$

in which I represents unitarity integrals written in the $SS^\dagger$ form and I', integrals in the $S^\dagger S$ form, and by dif I we really mean $I_> - I_{<\eta \pm i\epsilon}$. The integrals I and I' are related by the transformation $(+) \leftrightarrow -(-)$ applied to each bubble, as are $-B^-$ and B^+. There is a one-to-one correspondence between integrals I having $+(-)$ singularity mechanisms and integrals I' having $-(+)$ mechanisms. Thus, if (5.1a) leads to dif A^+ as a Cutkosky integral with + bubbles, then (5.1d) gives dif A^- as the same integral with + bubbles replaced by − bubbles and an over-all factor $(-1)^{n+1}$, where n is the number of bubbles. This agrees with the rules stated on the basis of Hermitian conjugation in II. The relation between (5.1a) and (5.1c) may be stated by saying that the arguments which, when read from right to left, diagrammatically apply to (5.1a), apply to (5.1c) when read from left to right. A similar relation exists between (5.1b) and (5.1d). Thus all procedures lead to the same picture of the singularity structure.

Finally, we consider the effect of changing the assumption that $\eta + i\epsilon$ is the correct prescription for the + amplitude. Supposing that $\eta - i\epsilon$ is in fact the correct continuation interchanges the left-hand sides of (5.1a) and (5.1b), and of (5.1c) and (5.1d). On mixed-α arcs, where all these right-hand sides are zero, this makes no difference. On positive-α arcs, however, the interchange would have serious results, for the inversion procedure would no longer give for dif A^+ a Cutkosky formula with only + bubbles. At the next stage of the induction argument this altered

Cutkosky form would spoil the cancellation mechanism even for the mixed-α arcs. We conclude therefore that $\eta + i\epsilon$ is necessary on positive-α arcs. This assumption for each Landau singularity is not quite as extensive as it seems, for if the scattering amplitude is to be the real boundary value of an analytic function almost everywhere in the physical region, the continuation prescriptions across different Landau curves are linked in the sense that they must agree at effective intersections between the curves. Thus an $\eta + i\epsilon$ prescription for one curve implies a similar prescription for all curves with which it has an effective intersection, all curves with which these other curves have effective intersections, and so on. This implication is limited by the noncontinuity of η (and hence of the prescription) across cusps or nodes of the curves. The positive arcs are believed to be free of these singular points, so that this hierarchial constraint works to maximum effect. On the other hand, it is highly satisfactory that either an $\eta + i\epsilon$ or an $\eta - i\epsilon$ prescription can be adopted on the mixed-α arcs, for these are known to be capable of having such cusps and nodes.

6. CONCLUSION

In this paper and its successor (which will deal with Landau diagrams having multiple lines) we will have given a complete description of the singularity structure in the physical region which is required by unitarity and the weak assumption of analyticity described in the introduction. This structure may be described as the minimum singularity structure. It is obtained by taking the singularities represented by explicit terms and adding to them all the singularities generated through the iterative mechanism. It is possible to conceive of further mechanisms, such as the sequences of complex singularities approaching the physical region discussed by Martin,[8] which would give further singularities consistent with unitarity. We have excluded these in the hope that the criterion of simplicity, or mimimal singularity, is the one chosen by nature.

In S-matrix theory it seems necessary at present to define the "minimum singularity structure" through an iterative mechanism such as that described above. We have been able to obtain precise results in the physical region, because crossing and Hermitian analyticity do not come into play and because the analytic continuations required are only infinitesimal and consequently make less demand on technical resources. Outside the physical region a similar but more involved iteration procedure is expected to operate.[9,10] We believe that it is an important aspect of our discussion that it shows how such iteratively defined singularity structure can be made to yield precise results.

Finally, we note that results similar to ours have been linked by some authors to causal properties of wavepackets.[11] The fact that we can deduce it from unitarity and a weak analyticity assumption is therefore some indication of how analyticity leads to causality. The reverse direction of the argument is still, of course, completely ununderstood.

[8] A. Martin, "Inability of Field Theory to Exploit the Full Unitarity Condition" (CERN Preprint Th. 727).

[9] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, London, 1966), Sec. 4.10.

[10] For an example outside the physical region, see J. C. Polkinghorne, J. Math. Phys. **7**, 2230 (1966).

[11] B. N. Valuev, Sov. Phys. JETP **20**, 433 (1965); S. Coleman and R. E. Norton, Nuovo Cimento **38**, 438 (1965); A. Peres, Ann. Phys. (N.Y.) **37**, 179 (1966); F. Pham, Ann. Inst. H. Poincaré, **6**, 89 (1967); D. Iagolnitzer, "S-Matrix and Classical Description of Interactions" (Saclay preprint, November, 1966); C. Chandler and H. P. Stapp, "S-Matrix Causality Conditions and Physical-Region Analyticity Properties," reported in *Proceedings of XIIIth International Conference on High-Energy Physics* (Univ. of California Press, Los Angeles, 1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Lower Bounds to the Ground-State Energy of Systems Containing Identical Particles

F. CALOGERO
Istituto di Fisica dell'Università di Roma, Istituto Nazionale di Fisica Nucleare Sezione di Roma, Rome, Italy

AND

C. MARCHIORO
Istituto di Fisica dell'Università di Roma, Rome, Italy

(Received 4 April 1968)

Lower bounds to the ground-state energy of a system composed of N identical particles interacting by two-body forces (and possibly also with an external potential) are derived. They are expressed in terms of the ground-state energies of (simpler) systems composed of a smaller number of particles having masses and/or interactions different from the original ones. This reduction process may be continued all the way down to systems involving one and two particles only. The generalization to systems containing some identical and some distinguishable particles is also discussed.

1. INTRODUCTION

The framework of this paper is the nonrelativistic quantum theory of many identical particles interacting among themselves by two-body forces, and possibly also with an external potential. The scope of this paper is to provide *lower* bounds to the ground-state energy of the system composed of N such particles, in terms of the ground state energies of simpler systems. The mathematical tool employed to obtain these results is the Rayleigh–Ritz minimum principle, which is, however, used in a somewhat unconventional way, so that it yields a *lower* rather than an *upper* bound to the quantity of interest.

Throughout this paper, by ground-state energy we mean the energy of the lowest-lying state of the system; this state may be a (normalizable) bound state or it may just mark the beginning of a continuum, the first case being generally the most interesting one.

No restriction on the two-body potential or on the external potential is detailed, the only requirement for the validity of the results being the applicability of the Rayleigh–Ritz principle. In particular the results hold both for potentials which diverge at large distances (corresponding to closed problems) and for potentials which vanish asymptotically.

The approach and results of this paper are analogous to, but more general than, those of Post.[1] The main difference between the present approach and that of Post (which is restricted to the case with only interparticle forces present and no external potential) is that he performs an exact separation of the center-of-mass motion which we do not make. This separation restricts the applicability of the approach and therefore eliminates the possibility to obtain some of the results we get. But in those cases where both approaches apply, the previous elimination of the center-of-mass motion is an advantage which is reflected in Post's bounds being more stringent than those derived here. It would of course be easy to modify our approach in such cases so as to reproduce Post's results, but because this would merely correspond to a repetition of his work we confine ourselves to pointing out which of our results should be discarded as being merely a less stringent version of the results previously obtained by Post.

In the following section the results are derived and discussed. Some examples, and in particular the limit of large N, are treated in Sec. 3; they demonstrate, at least in some cases, the stringency of the bounds obtained. Section 4 contains some final comments, including a discussion of the generalization to systems containing some identical and some distinguishable particles.

We use throughout a unit system such that $\hbar = 1$.

2. RESULTS

Let us consider the system composed of N identical particles of mass m interacting with an external potential $W(r_i)$ and among themselves via the two-body potential $V(r_i; r_j)$. Here and in the following, r_i labels the position of the ith particle in 3-space and all other internal degrees of freedom which this particle may possess (spin, isotopic spin, etc.); generally the

[1] H. R. Post, Proc. Phys. Soc. (London) **A69**, 936 (1956). See also: H. R. Post, Proc. Phys. Soc. (London) **A79**, 819 (1962); R. L. Hall and H. R. Post, Proc. Phys. Soc. (London) **A90**, 381 (1967); R. L. Hall, Proc. Phys. Soc. (London) **A91**, 787 (1967); O. Ya. Savchenko, Yad. Fiz. **6**, 645 (1967) [Sov. J. Nucl. Phys. **6**, 468 (1968)]. Other papers which consider similar problems but employ different techniques are: N. W. Bazley and D. W. Fox, Phys. Rev. **124**, 483 (1961); R. Sugar and R. Blankenbecler, Phys. Rev. **136**, B492 (1964); F. J. Dyson and A. Lenard, J. Math. Phys. **8**, 423 (1967); F. Calogero and Yu. A. Simonov, Nuovo Cimento **56B**, 71 (1968). Additional papers may be traced from the references in these.

potentials are multiplicative functions of the interparticle distance times operators acting on the (discrete) variables which label these additional degrees of freedom.[2]

We indicate by E_N the ground-state energy of this system. This is the quantity we want to minorize.

Let us then consider a comparison system composed of n particles of masses $m_i = m/h_i$, interacting between themselves via the two-body potentials $V_{ij}(r_i; r_j) \equiv g_{ij}V(r_i; r_j)$; the ith particle also interacts with the external potential $f_iW(r_i)$. We indicate with $E_n(h; g; f)$ the ground-state energy of this system, the arguments h, g, and f indicating respectively the sets $\{h_i\}$, $\{g_{ij}\}$, $\{f_i\}$.

If all the scale constants h_i coincide, $h_i = h$, the identity $E_n(h; g; f) = hE_n(1; g/h; f/h)$ holds, the arguments g/h and f/h indicating of course the sets $\{g_{ij}/h\}$ and $\{f_i/h\}$. This identity, whose proof follows by inspection from the respective Hamiltonians or, equivalently, from a dimensional count, is often used in the following to express the results in the most convenient way.

The results of this paper are based on the following lemma.

Lemma: If the number n of particles of the comparison system does not exceed the number N of original particles

$$n \leq N, \tag{2.1a}$$

and if the *scale constants* h_i and the *coupling constants* g_{ij} and f_i are adjusted so that *the sum total of, respectively, the kinetic and the potential energies are the same in the two systems*, namely,

$$\sum_{i=1}^{n} h_i = \sum_{i=1}^{N} 1 \equiv N, \tag{2.1b}$$

$$\sum_{i,j=1}^{n}{}' g_{ij} = \sum_{i,j=1}^{N}{}' 1 \equiv N(N-1), \tag{2.1c}$$

$$\sum_{i=1}^{n} f_i = \sum_{i=1}^{N} 1 \equiv N, \tag{2.1d}$$

then the ground-state energy $E_n(h; g; f)$ of the comparison system provides a lower bound to the ground-state energy E_N of the original system:

$$E_N \geq E_n(h; g; f). \tag{2.2}$$

Here and in the following the prime on a double sum indicates that the diagonal terms should be omitted.

[2] The 2-body potentials are assumed to depend (and operate) upon the degrees of freedom of each of the two particles, but not upon the state of the two-particle subsystem. In particular, exchange forces are excluded (although some of the results below remain valid even when exchange forces are present).

Proof: The Rayleigh–Ritz principle, applied to the comparison system, implies that

$$\left\langle \Psi, \left[\sum_{i=1}^{n} h_i T_i + \sum_{i,j=1}^{n}{}' g_{ij} V(r_i; r_j) + \sum_{i=1}^{n} f_i W(r_i) \right] \Psi \right\rangle \geq E_n(h; g; f), \tag{2.3a}$$

where $h_i T_i \equiv -(2m/h_i)^{-1}\nabla_i^2$ is the kinetic energy of the ith particle and Ψ is any normalized trial function. If we now choose Ψ to be a function of the N coordinates r_i, which is completely symmetric or antisymmetric under the exchange of any two coordinates r_i, r_j, we may, in Eq. (2.3a), freely substitute any other label in place of the index i of T_i and of the indices i and j in the arguments of V and W [note the role played at this point by the restriction of Eq. (2.1a); it is required that *all* the n coordinates which appear in the Hamiltonian of Eq. (2.3a) be treated symmetrically or antisymmetrically]. Taking advantage of this possibility and using Eqs. (2.1) we may therefore rewrite Eq. (2.3a) in the form

$$\left\langle \Psi, \left[\sum_{i=1}^{N} T_i + \sum_{i,j=1}^{N}{}' V(r_i; r_j) + \sum_{i=1}^{N} W(r_i) \right] \Psi \right\rangle \geq E_n(h; g; f). \tag{2.3b}$$

If we now choose Ψ to be the (unknown, but certainly symmetric or antisymmetric) ground-state wavefunction of the original N-particle system, the lhs of this inequality becomes the lhs of Eq. (2.2). Q.E.D.

Before proceeding to derive more specific results from this lemma, there is one important remark to be made, which implies, in some cases, a strengthening of the lemma.

It refers to the case in which also the comparison system contains some particles which are equal, namely which have identical masses and interactions. *The lemma may then be strengthened by requiring that, in evaluating the ground-state energy $E_n(h; g; f)$ of the comparison system, these particles be treated as identical particles, obeying the same statistic as the original particles.* The validity of this remark is easily justified through a re-examination of the proof of the lemma. The fact that it may modify the lemma only in the sense of making its conclusion more stringent is implied by the observation that any restriction, such as a (partial) symmetry requirement, upon the ground state of the comparison system can only raise (or leave unchanged) its energy.

The arbitrariness in the choice of the constants n, h_i, g_{ij}, and f_i, which is implicit in the formulation of the lemma, provides great flexibility. This is exploited below to derive a few significant examples of the kind

of results which are implied by the lemma, and by the remark following it. These examples, far from exhausting the potentiality of the lemma, are merely meant to illustrate it.

The first choice we consider is

$$n = N - 1, \quad h_i = f_i = h = N/(N-1),$$
$$g_{ij} = g = N/(N-2), \quad 1 \leq i,j \leq n. \tag{2.4}$$

It is consistent with Eqs. (2.1) and corresponds to the elimination of the Nth particle from the system, its kinetic and potential energy being evenly distributed among the other particles. It yields the following theorem.

Theorem 1: The ground-state energy of the system composed of N identical particles interacting with an external potential $W(r_i)$ and through the interparticle two-body potential $V(r_i; r_j)$ is not less than $N/(N-1)$ times the ground state energy of the system composed of $N-1$ such particles interacting with the same external potential but through a two-body potential which is $(N-1)/(N-2)$ times stronger than that of the N-body case:

$$E_N \geq E_{N-1}\left(\frac{N}{N-1}; \frac{N}{N-2}; \frac{N}{N-1}\right) = \frac{N}{N-1} E_{N-1}\left(1; \frac{N-1}{N-2}; 1\right). \tag{2.5}$$

Repeated application of this theorem yields the following corollary.

Corollary 1.1: The ground-state energy E_N of the system composed of N identical particles of mass m interacting via the two-body interparticle potential $V(r_i; r_j)$ and with the external potential $W(r_i)$ is not less than $N/2$ times the ground-state energy of the system composed of two such particles interacting with the same external potential and among themselves through a two-body interparticle potential, which is $N-1$ times stronger than the original interparticle potential:

$$E_N \geq E_2[\tfrac{1}{2}N; \tfrac{1}{2}N(N-1); \tfrac{1}{2}N] = \tfrac{1}{2}NE_2(1; N-1; 1). \tag{2.6}$$

Thus a lower bound to the ground-state energy of an N-particle system is obtained from the solution of a much simpler problem. Of course, this corollary follows directly from the lemma if one sets $n = 2$, $h_i = f_i = \frac{1}{2}N$, $g_{12} = g_{21} = \frac{1}{2}N(N-1)$, $i = 1, 2$.

In the case without external potential, namely if $W = 0$, this result may of course be written with zero in place of the argument f, but it provides a less stringent bound than Post's result,[1] which in our notation reads

$$E_N \geq E_2(N-1; \tfrac{1}{2}N(N-1), 0) = (N-1)E_2(1; \tfrac{1}{2}N; 0). \tag{2.7}$$

As mentioned in the introduction, the difference is due to the preliminary elimination, performed by Post, of the center-of-mass motion; this can, of course, be performed only if no external potential is present. Note, however, that even in this case such an elimination prevents the establishment of more general results, such as the lemma and theorem given above and also some of the theorems and corollaries given below. The fact that the result of Post is more stringent follows from the obvious inequality

$$E_2(N-1; g; 0) \geq E_2(\tfrac{1}{2}N; g; 0), \tag{2.8}$$

which is a consequence of the positivity of the kinetic energy and of the fact that for all values of N larger than two, $N-1$ is larger than $\frac{1}{2}N$; for $N = 2$ both results, Eqs. (2.6) and (2.7), hold with the equality sign and reduce to trivial identities.

Another choice of the constants, which is also consistent with Eqs. (2.1), is

$$n = N, \quad h_i = h, \quad 0 \leq h \leq (N-1)/N,$$
$$h_N = N - h(N-1), \quad g_{ij} = N/(N-2),$$
$$g_{iN} = g_{Ni} = 0, \quad f_i = 0, \quad f_N = N, \tag{2.9}$$
$$1 \leq i, \quad j \leq N-1.$$

This choice corresponds to decoupling the Nth particle from all the others and transferring to it all the interaction with the external potential. It yields the following theorem.

Theorem 2: The ground-state energy E_N of the system composed of N identical particles of mass m interacting via the two-body interparticle potential $V(r_i; r_j)$ and with the external potential $W(r_i)$ is not less than the sum of (a) h times the ground-state energy of the system composed of $N-1$ such particles, interacting via the two-body interparticle potential $N[h(N-2)]^{-1}V(r_i; r_j)$ but with no external potential, and (b) $[N - h(N-1)]$ times the ground state energy of one such particle interacting only with the external potential $N[N-h(N-1)]^{-1}W(r)$, h being an arbitrary nonnegative number less than $(N-1)/N$:

$$E_N \geq E_{N-1}\left(h; \frac{N}{N-2}; 0\right) + E_1(N - h(N-1); 0; N)$$
$$= hE_{N-1}\left(1; \frac{N}{h(N-2)}; 0\right) + [N - h(N-1)]E_1\left(1; 0; \frac{N}{N-h(N-1)}\right). \tag{2.10}$$

Combining this result with that of Post, Eq. (2.7), we obtain the following corollary, which we write directly as an inequality.

Corollary 2.1:

$$E_N \geq (N-2)hE_2\left(1; \frac{N(N-1)}{2h(N-2)}; 0\right) + [N - h(N-1)]E_1\left(1; 0; \frac{N}{N-h(N-1)}\right). \tag{2.11}$$

Here $E_2(1, g, 0)$ is the ground-state energy of the system composed of two identical particles having the same mass and obeying the same statistic as the original particles but interacting through a 2-body interparticle potential which is g times the original interparticle potential. Note that this corollary reduces the evaluation of a lower bound for the ground-state energy of a system composed of N identical particles to that of the ground state of one- and two-body systems only. The arbitrariness in the choice of h [except for the inequality in Eq. (2.9)] should also be emphasized.

Another interesting choice of the constants is the following one:

$$\begin{gathered} n = N, \quad h_1 = h_2 = h_i = 1, \\ g_{12} = g_{21} = g, \quad g_{1i} = g_{2i} = g_{i1} = g_{i2} = 0, \\ g_{ij} = g' = [(N-2)(N-3)]^{-1}[N(N-1) - 2g], \\ f_1 = f_2 = f, \quad f_i = f' = (N-2)^{-1}(N-2f), \\ 3 \leq i, j \leq N. \end{gathered} \tag{2.12}$$

This is also consistent with Eqs. (2.1); it corresponds to particles 1 and 2 interacting only between themselves, through the potential $gV(r_1; r_2)$, and with the external potential $fW(r_{1,2})$, and the remaining $N-2$ particles interacting between themselves through the potential $g'V(r_i; r_j)$ and with the external potential $f'W(r_i)$, with g' and f' adjusted to satisfy Eqs. (2.1). It yields, therefore, the following theorem.

Theorem 3: The ground-state energy of the system composed of N identical particles interacting via the two-body interparticle potential $V(r_i; r_j)$ and with the external potential $W(r_i)$ is not less than the sum of (a) the ground-state energy of the system composed of $N-2$ such particles interacting via a two-body interparticle potential which is $[(N-2)(N-3)]^{-1} \times [N(N-1) - 2g]$ times the original one, and with an external potential which is $(N-2)^{-1}(N-2f)$ times the original one, and (b) the ground-state energy of the system composed of two such particles interacting through a two-body interparticle potential which is g times the original one and an external potential which is f times the original one, g and f being arbitrary constants:

$$E_N \geq E_{N-2}(1; [(N-2)(N-3)]^{-1}[N(N-1) - 2g]; [N-2]^{-1}[N-2f]) + E_2(1; g; f). \tag{2.13}$$

The special choice

$$g = \tfrac{1}{2}[N(N-1) - (N-2)(N-3)] = 2N - 3,$$

with $f = 1$ and $f = 0$, yields two corollaries which we write directly in the form of inequalities.

Corollary 3.1:

$$E_N \geq E_{N-2} + E_2(1; 2N-3; 1). \tag{2.14}$$

Corollary 3.2:

$$E_N \geq E_{N-2}(1; 1; N/(N-2)) + E_2(1; 2N-3; 0). \tag{2.15}$$

Note that the first one yields a lower bound to the ground-state energy of the system composed of N identical particles in terms of the ground-state energy of the system composed of $N-2$ such particles *with the same forces*, and it might, therefore, be directly useful for phenomenological analysis.

For potentials which vanish at large distances, some additional interesting results may be easily obtained. To this end let us introduce the minimum value G necessary and sufficient for the two-body problem with the same particles and with the interparticle potential $GV(r_1; r_2)$ to possess one (zero-energy) bound state. Of course, G is less than unity if $V(r_1; r_2)$ itself is sufficiently attractive to sustain a two-body bound state. Similarly, let us introduce the minimum value F necessary and sufficient for the single-particle external potential $FW(r)$ to sustain a (zero-energy) bound state. We assume, of course, that at least one of the two potentials V, W is attractive at least in some region, otherwise the ground-state energy of any N-body system is simply zero; and in the following discussion we consider for simplicity only positive values of G and F, although this restriction is not really necessary.

Having introduced these definitions, we remark that they imply that

$$E_2(1; G; 0) = E_2(1; 0; F) = 0. \tag{2.16}$$

The combination of this equation with the previous inequalities yields some interesting results whose explicit derivation is left to the reader. We mention explicitly only the following theorem, which follows from Post's result, Eq. (2.7).

Theorem 4: If the potential $V(r_i; r_j)$ through which certain identical particles interact pairwise is too weakly attractive to sustain a two-body bound state, the minimum number of such particles which is necessary to build up sufficient attraction to yield a negative-energy ground-state is the smallest integer not less than $2G$, where G is such that the potential $GV(r_1; r_2)$ is just sufficient to sustain a (zero-energy) two-body bound state.

This theorem may, for instance, be used to obtain a lower bound for the number of ^{4}He atoms which constitutes the smallest "droplet" of liquid helium at zero temperature.[3]

In the theorems and corollaries given above we have concentrated upon the results which obtain by "eliminating" from the N-body system one particle (Theorem 1), or "factoring out" one particle (Theorem 2) or a two-particle compound (Theorem 3). This procedure was dictated by the aim to eventually reduce the N-body problem to subproblems involving at most two bodies, such problems being of course substantially easier. But, of course, the scope of the lemma is much broader, as it is exemplified by the following theorem, which is written directly as an inequality.

Theorem 5:

$$E_N \geq E_M(1; g; f) + E_{N-M}(1; [(N-M)(N-M-1)]^{-1} \times [N(N-1) - M(M-1)g]; [N-M]^{-1} \times [N - Mf]), \quad N \geq M+2, \quad M \geq 2. \quad (2.17)$$

This theorem follows from the lemma (with $n = N$ and $h_i = 1$) and corresponds to a choice of the constants g_{ij} and f_i such that, in the comparison system, the first M particles interact between themselves via the two-body potential gV and with the external potential fW, while the $N - M$ remaining particles interact between themselves via the potential $g'V$ and with the external potential $f'W$, with g' and f' adjusted so that the sum total, respectively, of the internal and external interaction energy is unchanged, i.e., Eqs. (2.1c) and (2.1d) are fulfilled. Again the values of g and f are arbitrary. Note that Theorem 3 is merely a special case of this theorem, corresponding to $M = 2$.

[3] A preliminary computation yields for this lower bound the value 11 ± 1, the error reflecting the uncertainty on the (phenomenological) potential. One of us (F. C.) wishes to acknowledge a suggestive conversation with Professor G. Morpurgo about the possibility to perform a computation of this kind.

In the following section we give some examples of applications of these results. Here we discuss in a qualitative manner what one might expect as regards the stringency of these bounds. These comments will also bring out the difference between the fermion and the boson case, which has been hidden up to now.

The essence of the lemma is the fact that a uniform distribution of kinetic and potential energy among all the particles of the system—and the requirement of total symmetry or antisymmetry of the wavefunction—which are characteristic of the system composed of N identical particles, raise the energy of the corresponding state, relative to that of a comparison system in which the kinetic and the potential energies are distributed less uniformly among all the particles, and to which no symmetry requirement applies (or it applies only partially). Of course, this statement is true only if the sum totals of the kinetic and potential energies, respectively, in the two systems are the same, the precise meaning of this equality being defined by the formulation of the lemma.

Once this basic point is understood, it is easy to anticipate which cases are more likely to be characterized by stringent bounds. They are just those in which the symmetry requirement is less important. Generally, for a system composed of N identical particles with no internal degrees of freedom and obeying Bose statistics, the symmetry requirement is irrelevant for the determination of the ground-state energy, because in this case for "dynamical" reasons, the ground-state wavefunction would be completely symmetrical even if the particles were distinguishable. Just the opposite situation prevails in the case of Fermi statistics, provided the number of particles is sufficiently large for the Pauli principle to play a dynamical role. (This is not the case for systems composed of up to four nucleons, since there are just four available states for each nucleon, corresponding to the internal degrees of freedom of spin and isospin, and the forces depend very weakly on these degrees of freedom; for instance, the ground-state wavefunction of the alpha particle is spatially symmetrical, just as it would be if the four nucleons were distinguishable particles.)

In conclusion, one may hope that results, such as those of Corollaries 1.1 and 2.1 and Theorem 4 above, which "reduce" the N-body system all the way down to two-body systems, yield stringent bounds also for values of N much larger than 2, only if the ground-state energy of the original system is not strongly affected by symmetry requirements; generally, this is the case if the ground-state wavefunction of the original system is spatially symmetrical, because

generally this is directly implied by the dynamics, even without an extra symmetry requirement. As regards results such as those of Theorems 1, 2, and 3, and of Corollaries 3.1 and 3.2, this consideration, although always valid, is less important. These conclusions are perhaps best illustrated by the discussion of an explicit example, say N identical fermions without internal degrees of freedom and interacting by 2-body attractive forces. For large values of N the ground-state energy of such a system is much raised by the Pauli principle, which forces the wavefunction to be completely antisymmetrical and, therefore, to have many zeros (resulting in a possible decrease of the average potential energy) and a lot of curvature (resulting in an increase of the average kinetic energy). Thus a lower bound such as that of, say, Corollary 2.1, is certainly very poor, because it neglects almost completely the effect of the Pauli principle. On the other hand, a result such as, say, Corollary 3.1, is for large N much more stringent, because the lower bound still retains almost fully the effect of the Pauli principle; namely, it retains that effect for $N-2$ particles. But in any case, this result cannot be very stringent, because the effect of the Pauli principle between 2 and $N-2$ particles is not taken into account. On the other hand, if the particles were bosons, the symmetry requirement would have no effect at all on the ground-state wavefunction, and therefore the hope that all bounds be reasonably good would not be unrealistic. That this hope is fulfilled, at least in some cases, is indeed confirmed by the examples treated in the following section.

3. EXAMPLES

In this section we consider some examples to give an idea of the stringency of the bounds considered in this paper. We limit ourselves to very simple cases, which can be managed without any analytical or numerical effort.

We begin with the limit of large N, restricting the consideration to the boson case which is the only one in which one may hope to obtain reasonably stringent bounds.

We note first of all that, if the two-body interparticle potential $V(r_i; r_j)$ has a finite negative minimum $-|V_{\min}|$, the reduction of the N-body problem to one- and two-body problems through any one of the theorems and corollaries of the previous section leads eventually to the trivial result:

$$E_N \geq -\tfrac{1}{2}N^2 |V_{\min}| [1 + O(N)^{-1}], \quad N \to \infty. \quad (3.1)$$

If, on the other hand, the potential does not have a finite minimum, the theorems and corollaries may yield more interesting results. For instance, in the case of an interparticle potential

$$V(r_i; r_j) = -e^2/|\mathbf{r}_i - \mathbf{r}_j|,$$

corresponding to an attractive Coulomb-like interaction between all pairs (Newtonian forces), and no external potential, the result of Post, Eq. (2.7), implies the inequality

$$E_N \geq -(16)^{-1}N^2(N-1)me^4. \quad (3.2)$$

Here and below m is the mass of the particles. This lower bound may be compared with the upper bound

$$E_N \leq -(6\pi)^{-1}N^2(N-1)me^4. \quad (3.3)$$

Both these results have been given by Post[1]; the second one is obtained by a straightforward application of the Rayleigh–Ritz variational principle with a Gaussian trial function depending only upon the sum of the squares of all interparticle distances. An upper bound which is certainly more stringent is[4]

$$E_N \leq -[4/(9\pi)][N(N-1)(N-2)^{-1}(3N-4)^{-1} \times \Gamma(\tfrac{3}{2}N - \tfrac{3}{2})/\Gamma(\tfrac{3}{2}N - 3)]^2 me^4. \quad (3.4)$$

For large N, these results yield

$$E_N = -c^{-1}N^3e^4m, \quad N \to \infty, \quad (3.5)$$

with

$$16 \leq c \leq 6\pi \quad (3.6)$$

leading to a determination of E_N accurate to $\pm 8\%$. Note that the upper bounds, Eqs. (3.3) and (3.4), coincide in this limit.

The combination of Post's lower bound, Eq. (3.2), with the upper bound of Eq. (3.4) yields, for small values of N, a determination of the ground-state energy of the system under consideration which is even more accurate. In fact for $N = 3$ and $N = 4$ one finds, respectively,

$$-\tfrac{9}{8}me^4 \leq E_3 \leq -\left(\frac{16}{5\pi}\right)^2 me^4, \quad (3.7)$$

$$-3me^4 \leq E_4 \leq -\left(\frac{105}{64}\right)^2 me^4, \quad (3.8)$$

which are accurate to $\pm 4\%$ and $\pm 5.5\%$, respectively. Other examples for small values of N may be found in the paper by Hall and Post.[1]

In the fermion case, and always considering particles which interact via the two-body interparticle potential $V(r_i; r_j) = -e^2/|\mathbf{r}_i - \mathbf{r}_j|$, one obtains in place of Eq.

[4] F. Calogero and Yu. A. Simonov, Phys. Rev. **169**, 789 (1968). The upper bound of this paper is the most stringent bound that may be obtained by inserting in the Rayleigh–Ritz principle a trial function which depends only on the sum of the squares of all interparticle distances.

(3.7) the result

$$-\frac{9}{32}me^4 \leq E_3 \leq -\left(\frac{1252704}{1276275\pi}\right)^2 me^4. \quad (3.9)$$

Here the lower bound is again obtained from the Post result, Eq. (2.7), taking into account the requirement of antisymmetry on the two-body problem, which implies that the lowest P-wave state has to be used in place of the S-wave ground state; this accounts for the difference (a factor of 4) between the lhs of Eqs. (3.7) and (3.9). The upper bound is obtained instead by applying the technique of Ref. 4 to the 3-fermion problem (with vanishing total angular momentum). These bounds imply a determination of the ground-state energy accurate to $\pm 50\%$. It should again be emphasized that these results are reported here only to give an idea of the stringency of the bounds in the various cases. Note, incidentally, that one is considering fermions having no internal degrees of freedom (and therefore no spin).

4. FINAL REMARKS

As we have already emphasized, the lemma of Sec. 2 constitutes a rather general result. However, it is not the most general one that may be formulated, as it is restricted to systems composed of identical particles only. In fact, the extension to systems which contain some identical and some distinguishable particles is already implicit in it, since the presence of the distinguishable particles may be considered, from the point of view of the identical particles, as merely contributing to the "external" potential $W(r_i)$. Thus in conclusion it may be quite generally stated that the essence of all the results of this paper is *the possibility, in order to establish a lower bound to the ground-state energy of a system containing some identical particles, to distribute at one's convenience among them the kinetic and potential energies, with the only constraint of keeping their sum totals unchanged.* The extraordinary flexibility afforded by this possibility is best illustrated by the consideration of an example.

Consider a system composed of four identical particles of mass m_a, five identical particles of mass m_b, and 13 identical particles of mass m_c. All these particles interact among themselves via two-body interparticle potentials; specifically, $V^{aa}(r_i; r_j)$ is the potential between two particles of mass m_a, $V^{ab}(r_i; r_j)$ is the interparticle potential between one particle of mass m_a and one of mass m_b, etc. This is clearly a complicated system. But the previous remark implies that the ground-state energy of this system is not less than the ground-state energy of a comparison system, composed of four particles of mass m_a, four particles of mass $m'_b = (\frac{4}{5})m_b$, and four particles of mass $m'_c = (\frac{4}{13})m_c$, interacting as follows: The first two particles of mass m_a interact only between themselves, via the potential $6V^{aa}$; the first two particles of mass m'_b interact only between themselves, via the potential $10V^{bb}$; the first two particles of mass m'_c interact only between themselves, via the potential $78V^{cc}$; the third particle of mass m_a and the third particle of mass m'_b interact only between themselves, via the potential $20V^{ab}$; the fourth particle of mass m_a and the third particle of mass m'_c interact only between themselves, via the potential $52V^{ac}$; and the fourth particle of mass m'_b interacts only with the fourth particle of mass m'_c, via the potential $65V^{bc}$. Clearly this system is a very simple one, being in fact only the union of six completely disconnected 2-body systems. Thus a lower bound to the ground-state energy of the original very complicated system may be obtained solving only two-body problems.

Obviously this reduction to problems involving only two bodies can be applied to any system composed of N_1 identical particles of mass m_1, N_2 identical particles of mass m_2, etc., up to N_K identical particles of mass m_K, all interacting between themselves, provided $N_I \geq K + 1$, $I = 1, 2, \cdots, K$; the reduced problem may consist of K 2-body problems with appropriate masses and potentials $\frac{1}{2}I(I-1)V^{II}$ and of $\frac{1}{2}K(K-1)$ 2-body problems with appropriate masses and potentials IJV^{IJ} (here of course V^{II} indicates the potential between identical particles of type I, and V^{IJ} the interparticle potential between a particle of type I and one of type J). It is also clear that the reduction we have indicated is merely one possibility among several. Which one of these is likely to produce a more stringent bound must be discussed in each case, keeping in mind the considerations which have been offered at the end of Sec. 2.

It should finally be emphasized that the remarks referring to the (partial) symmetry constraints which may be imposed upon the comparison system (in particular the remark after the lemma in Sec. 2) also apply in this more general context.

Note Added in Proof: In the case without external potential another interesting choice for the constants in the lemma is the following one:

$$n = N, \quad h_i = h = N/(N-1), \quad h_N = 0,$$
$$g_{iN} = g_{Ni} = g = \tfrac{1}{2}N, \quad g_{ij} = 0,$$
$$1 \leq i, \quad j \leq N-1.$$

One is thereby attributing an infinite mass to the Nth particle, so that it does not move, and is letting the remaining particles interact only with this one, so

that they become essentially independent from one another. In the boson case, this choice leads again to the result (2.6) (with $f = 0$), but in the fermion case the remarks after the lemma leads to the inequality

$$E_N \geq \sum_{n=1}^{N-1} E_2^{(n)}(\tfrac{1}{2}N/(N-1); \tfrac{1}{2}N)$$
$$= \tfrac{1}{2}N(N-1)^{-1}\sum_{n=1}^{N-1} E_2^{(n)}(1; N-1),$$

where $E_2^{(n)}(h; g)$ is the nth energy level of the problem with two distinguishable particles having masses m/h and interacting through the interparticle potential gV [or, equivalently, of the problem with one particle of mass $m/(2h)$ in the external fixed potential gV]. Of course in the sum, each level must enter as many times as required by its multiplicity. This result, however, for $N > 3$, is superseded by that of Hall,[5] who by a previous separation of the center-of-mass motion obtains

$$E_N \geq \sum_{n=1}^{N-1} E_2^{(n)}(\tfrac{3}{4}; \tfrac{1}{2}N) = \frac{3}{4}\sum_{n=1}^{N-1} E_2^{(n)}(1; \tfrac{2}{3}N).$$

[5] R. L. Hall, Proc. Phys. Soc. (London) **A91**, 16 (1967); other references are given in this paper.

The advantage of this approach is that it retains more fully the effect of the Pauli principle; it is, therefore, particularly convenient at large N. For instance, for $V(r_i, r_j) = -e^2/|\mathbf{r}_i - \mathbf{r}_j|$, the last formula yields the lower bound[6]

$$E_N \geq -\tfrac{1}{6}me^4N^2(V+1); \quad V \text{ integral},$$
$$\tfrac{1}{3}V(V+\tfrac{1}{2})(V+1) \leq N-1 < \tfrac{1}{3}(V+1)(V+\tfrac{3}{2})(V+2).$$

ACKNOWLEDGMENTS

One of us (F. C.) wishes to acknowledge a CERN Traveling Fellowship which made possible, in February 1968, a two-week visit to the Institute of Theoretical and Experimental Physics (ITEP) in Moscow, during which this paper was partly written. It is a pleasure to thank all the staff of the Theoretical Group at ITEP for the warm hospitality, and Dr. Yu. A. Simonov for several illuminating discussions on the argument of this paper.

[6] Lévy-Leblond obtains a similar result by an approach analogous to that described in this paper; the result given here is more stringent, for $N > 3$, because it incorporates Hall's improvement. Asymptotically ($N \to \infty$) the improvement is by a factor $2/3$. J.-M. Lévy-Leblond, J. Math. Phys. **10**, 806 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Group-Theoretic Approach to the New Conserved Quantities in General Relativity

MOSHE CARMELI

General Physics Research Laboratory, Aerospace Research Laboratories, Wright-Patterson AFB, Ohio

(Received 1 July 1968)

The Newman–Penrose formalism for obtaining the recent conserved quantities in general relativity is discussed and a group-theoretic interpretation is given to it. This is done by relating each triad of the orthonormal vectors on the sphere to an orthogonal matrix g. As a result, the spin-weighted quantities η become functions on the group of three-dimensional rotation, $\eta = \eta(g)$, where $g \in O_3$. An explicit form for the matrix g is given and a prescription for rewriting $\eta(g)$ as functions of the spherical coordinates is also given. We show that a quantity of spin weight s can be expanded as a series in the matrix elements T^j_{sm} of the irreducible representation of O_3, where s is fixed. Infinite- and finite-dimensional representations of the group SU_3 are then realized in the spaces of η's and T^j_{sm}. It is shown that the infinite-dimensional representation is not irreducible; its decomposition into irreducible parts leads to the expansion of η in the T^j_{sm}, the latter providing invariant subspaces in which irreducible representations act.

1. INTRODUCTION

Recent research on conservation laws for the electromagnetic and gravitational fields by Newman and Penrose[1] features a new differential operator ð and a class of functions ${}_sY_{jm}(\phi, \theta)$, called *spin-s spherical harmonics*, all defined on a sphere. The new operator and functions appear in the study of the Bondi–Metzner–Sachs group. The ${}_sY_{jm}$ form a complete orthonormal set for each value of s in the sense that certain field functions η, called quantities of spin-weight s, can be expanded in series in the ${}_sY_{jm}$. Quantities η of spin-weight s are those field functions

[1] E. T. Newman and R. Penrose, J. Math. Phys. 7, 863 (1966).

that they become essentially independent from one another. In the boson case, this choice leads again to the result (2.6) (with $f = 0$), but in the fermion case the remarks after the lemma leads to the inequality

$$E_N \geq \sum_{n=1}^{N-1} E_2^{(n)}(\tfrac{1}{2}N/(N-1); \tfrac{1}{2}N)$$
$$= \tfrac{1}{2}N(N-1)^{-1}\sum_{n=1}^{N-1} E_2^{(n)}(1; N-1),$$

where $E_2^{(n)}(h; g)$ is the nth energy level of the problem with two distinguishable particles having masses m/h and interacting through the interparticle potential gV [or, equivalently, of the problem with one particle of mass $m/(2h)$ in the external fixed potential gV]. Of course in the sum, each level must enter as many times as required by its multiplicity. This result, however, for $N > 3$, is superseded by that of Hall,[5] who by a previous separation of the center-of-mass motion obtains

$$E_N \geq \sum_{n=1}^{N-1} E_2^{(n)}(\tfrac{3}{4}; \tfrac{1}{2}N) = \frac{3}{4}\sum_{n=1}^{N-1} E_2^{(n)}(1; \tfrac{2}{3}N).$$

[5] R. L. Hall, Proc. Phys. Soc. (London) **A91**, 16 (1967); other references are given in this paper.

The advantage of this approach is that it retains more fully the effect of the Pauli principle; it is, therefore, particularly convenient at large N. For instance, for $V(r_i, r_j) = -e^2/|\mathbf{r}_i - \mathbf{r}_j|$, the last formula yields the lower bound[6]

$$E_N \geq -\tfrac{1}{6}me^4N^2(V+1); \quad V \text{ integral},$$
$$\tfrac{1}{3}V(V+\tfrac{1}{2})(V+1) \leq N-1 < \tfrac{1}{3}(V+1)(V+\tfrac{3}{2})(V+2).$$

ACKNOWLEDGMENTS

One of us (F. C.) wishes to acknowledge a CERN Traveling Fellowship which made possible, in February 1968, a two-week visit to the Institute of Theoretical and Experimental Physics (ITEP) in Moscow, during which this paper was partly written. It is a pleasure to thank all the staff of the Theoretical Group at ITEP for the warm hospitality, and Dr. Yu. A. Simonov for several illuminating discussions on the argument of this paper.

[6] Lévy-Leblond obtains a similar result by an approach analogous to that described in this paper; the result given here is more stringent, for $N > 3$, because it incorporates Hall's improvement. Asymptotically ($N \to \infty$) the improvement is by a factor $2/3$. J.-M. Lévy-Leblond, J. Math. Phys. **10**, 806 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Group-Theoretic Approach to the New Conserved Quantities in General Relativity

MOSHE CARMELI

General Physics Research Laboratory, Aerospace Research Laboratories, Wright-Patterson AFB, Ohio

(Received 1 July 1968)

The Newman–Penrose formalism for obtaining the recent conserved quantities in general relativity is discussed and a group-theoretic interpretation is given to it. This is done by relating each triad of the orthonormal vectors on the sphere to an orthogonal matrix g. As a result, the spin-weighted quantities η become functions on the group of three-dimensional rotation, $\eta = \eta(g)$, where $g \in O_3$. An explicit form for the matrix g is given and a prescription for rewriting $\eta(g)$ as functions of the spherical coordinates is also given. We show that a quantity of spin weight s can be expanded as a series in the matrix elements T^j_{sm} of the irreducible representation of O_3, where s is fixed. Infinite- and finite-dimensional representations of the group SU_3 are then realized in the spaces of η's and T^j_{sm}. It is shown that the infinite-dimensional representation is not irreducible; its decomposition into irreducible parts leads to the expansion of η in the T^j_{sm}, the latter providing invariant subspaces in which irreducible representations act.

1. INTRODUCTION

Recent research on conservation laws for the electromagnetic and gravitational fields by Newman and Penrose[1] features a new differential operator ð and a class of functions ${}_sY_{jm}(\phi, \theta)$, called *spin-s spherical harmonics*, all defined on a sphere. The new operator and functions appear in the study of the Bondi–Metzner–Sachs group. The ${}_sY_{jm}$ form a complete orthonormal set for each value of s in the sense that certain field functions η, called quantities of spin-weight s, can be expanded in series in the ${}_sY_{jm}$. Quantities η of spin-weight s are those field functions

[1] E. T. Newman and R. Penrose, J. Math. Phys. 7, 863 (1966).

obtained by contraction of tensor (of any rank) fields with a certain complex triad.

Consequently, the operator ð and the functions ${}_sY_{jm}$ were studied by Goldberg, Macfarlane, Newman, Rohrlich, and Sudarshan.[2] They showed that ð is related to the angular-momentum operator **K** and the ${}_sY_{jm}$ are related[3] to the matrix elements of the irreducible representations T^j_{sm} of the three-dimensional rotation group O_3. The relevance of ð and ${}_sY_{jm}$ to the rotation group was stimulated by the similarity of certain relations appearing in the Newman–Penrose formalism to those appearing in the usual theory of angular momentum. It was shown that, by choosing the Euler angles appearing in T^j_{mn} and **K** in a certain way, an identification with ${}_mY_{jn}$ and ð is obtained. If T^j_{mn} and **K** were to be written in terms of some three variables other than Euler's angles,[4] one might still find a way to make such an identification possible.

The question arises as to whether the Newman–Penrose formalism has a closer connection to O_3. This question becomes increasingly important because of the unclear physical meaning of the new conserved quantities. Giving a group-theoretic interpretation to the formalism could lead to a better understanding of these conserved quantities which might be of importance to astrophysics and probably to the rest of physics. In this paper we answer the above question positively, thus giving the Newman–Penrose formalism a group-theoretic basis.

In general there exists a close connection between the usual theory of spherical harmonics and that of group representations which was first pointed out by both Cartan and Weyl, though it is not our purpose to go into this question here.[5]

It should be mentioned that Gel'fand and Shapiro[6,7] have presented a method for expansion of certain combinations of components of vector and tensor fields. The expansion was made in terms of *generalized spherical functions*, the latter are the representation matrix elements T^j_{mn}, written as functions of Euler's angles. When the field quantities are evaluated on the surface of a sphere, one has again to make certain identification of its variables, similar to what has been shown in Ref. 2.

Finally, we point out that *generalized surface harmonics* were also introduced and discussed by Moses.[8] We only mention that Moses' generalized surface harmonics, denoted by $Y^{mn}_j(\phi, \theta)$, are related to the spin-s spherical harmonics ${}_sY_{jm}(\phi, \theta)$ of Newman and Penrose by

$$Y^{sm}_j(\phi, \theta) = {}_{-s}Y_{jm}(\phi, \theta)e^{-is\phi}.$$

In Sec. 2, we define quantities η as functions over the group O_3. These quantities are defined in the same way Newman and Penrose define their quantities of spin-weight s. However, whereas previously the additional rotational degree of freedom was fixed, we here do not make this convention and our η's depend on three angles.

In Sec. 3, we relate these three angles to the three variables of the elements of O_3. The expansion of the functions η in terms of the matrix elements of the irreducible representations of O_3 is given in Sec. 4. The last section is devoted to the connection between infinite- and finite-dimensional representation of O_3 in the spaces of the η's and the T^j_{mn}, respectively.

2. QUANTITIES OF SPIN WEIGHT s

In the three-dimensional Euclidean space with spherical coordinates r, θ, ϕ, one introduces a triad[9] of unit vectors ξ_1, ξ_2, ξ_3 on each point of a sphere of radius r. The two vectors ξ_1 and ξ_2 are taken to be in the tangent plane to the sphere at the spherical angles ϕ and θ, whereas ξ_3 is taken to be normal to the sphere there. The vectors ξ_1 and ξ_2 are defined up to a rotation with an angle, which we denote by ϕ_2, in the tangent plane about an axis in the direction of ξ_3. The rotation of ξ_1 and ξ_2 about ξ_3 is given a definite mathematical expression in the sequel (see Sec. 3).

By introducing the above rotation, we have added a new variable ϕ_2 upon which the two vectors ξ_1 and ξ_2 depend. Accordingly, these two vectors depend on the spherical angles ϕ, θ as well as the new angle ϕ_2:

$$\xi_1 = \xi_1(\phi, \theta, \phi_2), \quad \xi_2 = \xi_2(\phi, \theta, \phi_2). \tag{2.1}$$

The vector ξ_3, on the other hand, depends only on ϕ and θ:

$$\xi_3 = \xi_3(\phi, \theta). \tag{2.2}$$

Of particular interest to us is the behavior of ξ_1 and ξ_2 under the rotation about ξ_3. Such a rotation

[2] J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. **8**, 2155 (1967).

[3] In Ref. 2, the relationship of the functions spin-s spherical harmonics to the four-dimensional rotation group and to the Lorentz group was also indicated.

[4] Representations of the three-dimensional rotation group in terms of direction and angle of rotation was given by H. E. Moses, Ann. Phys. (N.Y.) **37**, 224 (1966); M. Carmeli, J. Math. Phys. **9**, 1987 (1968).

[5] See, for example, R. Godement, Trans. Am. Math. Soc. **73**, 496 (1952).

[6] I. M. Gel'fand and Z. Ya. Shapiro, Usp. Mat. Nauk **7**, 3 (1952); English translation in Am. Math. Soc. Transl. (2) **2**, 207 (1956).

[7] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon Press, Inc., New York, 1963).

[8] H. E. Moses, Ann. Phys. (N.Y.) **41**, 166 (1967).

[9] In Ref. 2, these three vectors were denoted by **a**, **b**, and **c**.

can be represented by the orthogonal matrix

$$\Omega \equiv \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.3}$$

Denoting the column of the three vectors ξ_1, ξ_2, ξ_3 by ξ,

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}, \tag{2.4}$$

then under the rotation Ω we have

$$\xi' = \Omega\xi. \tag{2.5}$$

It is convenient to introduce the complex vector ξ_+ and its complex conjugate ξ_- defined by

$$\xi_\pm = 2^{-\frac{1}{2}}(\xi_1 \mp i\xi_2). \tag{2.6}$$

Under the rotation (2.3) these complex vectors transform as

$$\xi'_\pm = e^{\pm i\psi}\xi_\pm. \tag{2.7}$$

A quantity η is now said to be of spin weight s if under the rotation Ω it transforms as

$$\eta \to \eta' = e^{is\psi}\eta, \tag{2.8}$$

where s is an integer.

Examples of quantities of spin weights 1, 0, and -1 are obtained by scalar multiplication of a vector field $\mathbf{v}$ with $\xi_\mp$, ξ_3:

$$\eta_{\pm 1}(\phi, \theta, \phi_2) = \mathbf{v}(\phi, \theta) \cdot \xi_\pm(\phi, \theta, \phi_2),$$
$$\eta_0(\phi, \theta) = \mathbf{v}(\phi, \theta) \cdot \xi_3(\phi, \theta). \tag{2.9}$$

Further examples of quantities of spin weights of any order can be obtained in a similar way from tensor fields.

In the following, we restrict ourselves to quantities η obtained by contraction of tensor fields with the triad $\xi_\pm$, ξ_3. In other words, the η functions are components of tensor fields along the complex triad. These components generally depend on the three angles ϕ, θ, ϕ_2. By relating these angles to the three parameters appearing in the three-dimensional rotation group (such as Euler's angles or direction and angle of rotation[4]), the quantities η can be considered as functions of g:

$$\eta = \eta(g), \tag{2.10}$$

where g is an element of O_3.

Using the well-known relationship between the rotation group and the special unitary group of order two, SU_2, we consider η as functions on the group SU_2 also:

$$\eta = \eta(u), \tag{2.11}$$

where $u \in SU_2$.

In the next section, we give the explicit dependence of g on the angles ϕ, θ, and ϕ_2. We also find out what should be substituted for g in order that η be written as a linear combination of components of the original tensor field.

3. RELATION OF η FUNCTIONS TO THE ROTATION GROUP

We now relate the quantities η to the elements g of the rotation group O_3.

First we relate the triad field ξ, given by Eq. (2.4), to $g \in O_3$ in such a way that for each triad ξ there corresponds a rotation $g \in O_3$. This correspondence can be achieved, using a method similar to that outlined by Gel'fand and Shapiro,[6] by assigning to each triad ξ a rotation $g \in O_3$ which transforms a certain given triad on the sphere at $\phi = \theta = 0$ to ξ. To find the rotation g, we proceed as follows.

We introduce a new triad of unit vectors

$$\mathbf{e} = \begin{pmatrix} \mathbf{e}_\theta \\ \mathbf{e}_\phi \\ \mathbf{e}_r \end{pmatrix}, \tag{3.1}$$

whose vectors are directed along the coordinates θ, ϕ, and r, respectively, and whose origin coincides with that of the triad ξ. The triad of vectors $\mathbf{e}$ is a function of the spherical coordinates, $\mathbf{e} = \mathbf{e}(\phi, \theta)$. The particular triad at $\phi = \theta = 0$ is denoted by us by

$$\mathbf{e}^0 = \begin{pmatrix} \mathbf{e}^0_\theta \\ \mathbf{e}^0_\phi \\ \mathbf{e}^0_r \end{pmatrix}. \tag{3.2}$$

Accordingly, we have

$$\mathbf{e}^0 = [\mathbf{e}(\phi, \theta)]_{\phi=\theta=0}. \tag{3.3}$$

It is easily seen that the three vectors $\mathbf{e}^0_\theta$, $\mathbf{e}^0_\phi$, and $\mathbf{e}^0_r$ are pointing in the same directions as the Cartesian coordinates x, y, and z of the fixed system. The transformation g is then defined as that one which transforms the triad $\mathbf{e}^0$ into the triad ξ:

$$\xi = g\mathbf{e}^0. \tag{3.4}$$

Now the vectors $\mathbf{e}_\theta$, $\mathbf{e}_\phi$, and $\mathbf{e}_r$ can be decomposed along the Cartesian coordinates, hence along $\mathbf{e}^0_\theta$, $\mathbf{e}^0_\phi$, and $\mathbf{e}^0_r$. One easily finds that the matrix of rotation R which transforms the triad $\mathbf{e}^0$ into the triad $\mathbf{e}$,

$$\mathbf{e} = R\mathbf{e}^0, \tag{3.5}$$

is given by

$$R = \begin{pmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \\ \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{pmatrix}. \tag{3.6}$$

The above matrix R can be written as a product of three orthogonal matrices

$$R = DBA', \tag{3.7}$$

where

$$D = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.8}$$

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \tag{3.9}$$

and where

$$A' = \begin{pmatrix} \sin\phi & -\cos\phi & 0 \\ \cos\phi & \sin\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.10}$$

By changing the variable ϕ, the matrix A' may be written as

$$A = \begin{pmatrix} \cos\phi_1 & -\sin\phi_1 & 0 \\ \sin\phi_1 & \cos\phi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.11}$$

where $\phi_1 = \pi/2 - \phi$.

It remains to find the transformation from the triad $\mathbf{e}$ to the triad $\boldsymbol{\xi}$. Since the angle of rotation ϕ_2 of the two vectors $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ was left undetermined, we define it by

$$\boldsymbol{\xi} = CD^{-1}\mathbf{e}, \tag{3.12}$$

where the matrix C is given by

$$C = \begin{pmatrix} \cos\phi_2 & -\sin\phi_2 & 0 \\ \sin\phi_2 & \cos\phi_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.13}$$

For later calculations we have to know the relation between the triad $\boldsymbol{\xi}$ when $\phi_2 = 0$ and the triad $\mathbf{e}$. From Eq. (3.12) we obtain

$$\begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \\ \boldsymbol{\xi}_3 \end{pmatrix}_{|\phi_2=0} = \begin{pmatrix} -\mathbf{e}_\phi \\ \mathbf{e}_\theta \\ \mathbf{e}_r \end{pmatrix}. \tag{3.14}$$

Using Eqs. (3.12), (3.5), (3.7), and (3.11), we find the explicit form of the rotation $g \in O_3$ which transforms the fixed triad $\mathbf{e}^0$ into the triad $\boldsymbol{\xi}$:

$$g(\phi_1, \theta, \phi_2) = C(\phi_2)B(\theta)A(\phi_1). \tag{3.15}$$

The transformation $g(\phi_1, \theta, \phi_2)$ represents three rotations with Euler's angles ϕ_1, θ, and ϕ_2 around the z, x, and z axis, respectively, where $\phi_1 = \pi/2 - \phi$. Hence, for each value of the variables ϕ, θ, and ϕ_2 of the triad $\boldsymbol{\xi}$ there corresponds a rotation $g(\pi/2 - \phi, \theta, \phi_2) \in O_3$, and any function of these variables can be considered as a function of $g \in O_3$.

In particular, the vectors $\boldsymbol{\xi}_\pm$, $\boldsymbol{\xi}_3$ can be considered as functions over the group O_3:

$$\boldsymbol{\xi}_\pm = \boldsymbol{\xi}_\pm(g), \quad \boldsymbol{\xi}_3 = \boldsymbol{\xi}_3(g). \tag{3.16}$$

A direct calculation shows that

$$\begin{aligned} \boldsymbol{\xi}_\pm(g) &= -2^{-\frac{1}{2}}[\mathbf{e}_\phi(\phi_1, \theta) \pm i\mathbf{e}_\theta(\phi_1, \theta)]e^{\mp i\phi_2}, \\ \boldsymbol{\xi}_3(g) &= \mathbf{e}_r(\phi_1, \theta), \end{aligned} \tag{3.17}$$

where $\phi_1 = \pi/2 - \phi$.

The above considerations show that all functions η obtained by contraction of tensor fields with the complex vectors (3.16), such as those given by Eqs. (2.9), are functions of $g \in O_3$:

$$\eta = [\eta(g)]_{\phi_1=\pi/2-\phi}. \tag{3.18}$$

For example, the functions η_s, $s = 1, 0, -1$, given by Eqs. (2.9), are given by

$$\begin{aligned} \eta_{\pm 1}(g) &= -2^{-\frac{1}{2}}[v_\phi(\phi_1, \theta) \pm iv_\theta(\phi_1, \theta)]e^{\mp i\phi_2}, \\ \eta_0(g) &= v_r(\phi_1, \theta). \end{aligned} \tag{3.19}$$

When we put $\phi_2 = 0$, the functions η become functions of the spherical coordinates ϕ and θ only:

$$\eta(\phi, \theta) = [\eta(g)]_{\phi_1=\pi/2-\phi, \phi_2=0}. \tag{3.20}$$

Using Eqs. (3.19), we obtain the result for η_s, $s = 1, 0, -1$, for example:

$$\begin{aligned} \eta_{\pm 1}(\phi, \theta) &= 2^{-\frac{1}{2}}(v_\phi \pm iv_\theta), \\ \eta_0(\phi, \theta) &= v_r. \end{aligned} \tag{3.21}$$

In the next section, we relate these functions to the matrix elements of the irreducible representations of O_3 and SU_2.

4. EXPANSION OF QUANTITIES OF SPIN WEIGHT s

Let $T^j_{mn}(u)$ be the matrix elements of the irreducible representations of the special unitary group of order two, SU_2, in their canonical basis (called the generalized spherical functions by Gel'fand and Shapiro). Here, $m, n = -j, -j+1, \cdots, j$, where j, the weight of the representation, is a nonnegative integer or half-integer, $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$. The three variables ϕ_1, θ, ϕ_2 are employed here also so that a function $f(u)$ over the group SU_2 means a function of the Euler angles, $f(u) = f(\phi_1, \theta, \phi_2)$. Although the indices j, m, n can be half-integers, in general they are restricted in our case to integers only. This is so since we have limited ourselves to quantities η having only integral weights (compare Sec. 2). The use of SU_2, whose relationship to O_3 is well known, makes it easy to extend our results to quantities with half-integral weights.

We mention without proof[10] that every continuous

[10] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon Press, Inc., New York, 1964).

function $f(u)$ in the group SU_2 is the limit of a uniformly convergent sequence of finite linear combinations of the functions $T^j_{mn}(u)$.[11] This means that the functions T^j_{mn} form a complete orthogonal system for the aggregate of all functions $f(u)$ whose modulus square is integrable with the measure du.[12] Accordingly, we have

$$f(u) = \sum_{j=0}^{\infty} \sum_{m,n=-j}^{j} \alpha^j_{mn} T^j_{mn}(u), \tag{4.1}$$

where α^j_{mn} are constants given by

$$\alpha^j_{mn} = (2j+1)A^j_{mn},$$
$$A^j_{mn} = \int f(u)\, \overline{T^j_{mn}(u)}\, du. \tag{4.2}$$

We now examine the behavior of different quantities under the rotation Ω, Eq. (2.3). The unitary matrix $\gamma \in SU_2$ which corresponds to the rotation $\Omega \in O_3$ is given by[10]

$$\gamma = \mp \begin{pmatrix} e^{-i\psi/2} & 0 \\ 0 & e^{i\psi/2} \end{pmatrix}. \tag{4.3}$$

Furthermore, the matrix T^j, when j is an integer, satisfies[10]

$$T^j_{mn}(\gamma u) = e^{im\psi} T^j_{mn}(u),$$
$$T^j_{mn}(u\gamma) = e^{in\psi} T^j_{mn}(u), \tag{4.4}$$

where γ is given by Eq. (4.3) and $u \in SU_2$.

Accordingly, under the rotation Ω, the function $f(u)$ transforms into

$$f(\gamma u) = \sum_j \sum_{m,n} a^j_{mn} T^j_{mn}(\gamma u)$$
$$= \sum_j \sum_{m,n} \alpha^j_{mn} e^{im\psi} T^j_{mn}(u). \tag{4.5}$$

In particular, for a quantity of spin weight s, we have

$$\eta(u) = \sum_j \sum_{m,n} \beta^j_{mn} T^j_{mn}(u), \tag{4.6}$$

$$\eta(\gamma u) = \sum_j \sum_{m,n} \beta^j_{mn} e^{im\psi} T^j_{mn}(u). \tag{4.7}$$

On the other hand, η satisfies the relation (2.8) which, in our present notation, reads

$$\eta(\gamma u) = e^{is\psi} \eta(u). \tag{4.8}$$

Hence, using Eq. (4.6), we have

$$\eta(\gamma u) = e^{is\psi} \sum_j \sum_{m,n} \beta^j_{mn} T^j_{mn}(u). \tag{4.9}$$

Comparing the two expressions for $\eta(\gamma u)$ given by Eqs. (4.7) and (4.9), we obtain

$$\sum_j \sum_{m,n} \beta^j_{mn} (e^{im\psi} - e^{is\psi}) T^j_{mn}(u) = 0. \tag{4.10}$$

Using the orthogonality relation that the matrices T^j satisfy,

$$\int T^j_{mn}(u)\, \overline{T^{j'}_{m'n'}(u)}\, du = (2j+1)^{-1} \delta_{jj'} \delta_{mn'} \delta_{nn'}, \tag{4.11}$$

we obtain

$$\beta^j_{mn} (e^{im\psi} - e^{is\psi}) = 0. \tag{4.12}$$

Thus, we have

$$\beta^j_{mn} = \delta_{ms} \beta^j_{sn}. \tag{4.13}$$

Accordingly, using Eqs. (4.6) and (4.13), we obtain for any quantity of spin weight s:

$$\eta(u) = \sum_{j=|s|}^{\infty} \sum_{n=-j}^{j} \beta^j_{sn} T^j_{sn}(u). \tag{4.14}$$

It remains to write this formula in terms of the original spherical angles. Since, similarly to $g \in O_3$, u is expressed in terms of ϕ_1, θ, and ϕ_2, we obtain the desired expressions by substituting $\phi_1 = \pi/2 - \phi$ and $\phi_2 = 0$. Accordingly,[13]

$$\eta(\phi, \theta) = \sum_{j=|s|}^{\infty} \sum_{n=-j}^{j} \beta^j_{sn} T^j_{sn}(\pi/2 - \phi, \theta, 0). \tag{4.15}$$

For example, a vector field $\mathbf{v}$ can be expanded in the following way:

$$v_{\pm}(\phi, \theta) = \sum_{j,n} \beta^j_{\pm 1,n} T^j_{\pm 1,n}(\pi/2 - \phi, \theta, 0),$$
$$v_0(\phi, \theta) = \sum_{j,n} \beta^j_{0,n} T^j_{0,n}(\pi/2 - \phi, \theta, 0), \tag{4.16}$$

where we have used the notation

$$v_{\pm} = -2^{-\frac{1}{2}}(v_\phi \pm i v_\theta),$$
$$v_0 = v_r. \tag{4.17}$$

A tensor field $\mathbf{W}$ of rank two provides nine quantities W_s of weights $s = 2, 1, 0, -1, -2$. These are obtained by contraction of the tensor field $\mathbf{W}$ with $\boldsymbol{\xi}_{\mp}$, $\boldsymbol{\xi}_3$ and inserting $\phi_2 = 0$ [using Eq. (3.14)].[14]

[11] This theorem is valid for compact groups in general. See, for example, L. S. Pontrjagin, *Topological Groups* (Princeton University Press, Princeton, N.J., 1946); M. A. Naimark, *Normed Rings* (P. Noordhoff Ltd., Groningen, The Netherlands, 1959).

[12] The aggregate of all measurable functions $f(u)$ satisfying the condition $\int |f(u)|^2\, du < \infty$ provides a complete Hilbert space denoted by $L^2(SU_2)$. In this space, the addition of functions and multiplication by a number are defined in the usual way, the scalar product being defined by $(f_1, f_2) = \int f_1(u)\overline{f_2(u)}\, du$.

[13] Comparing the expansion given by Eq. (4.15) with the similar one in terms of the spin-s spherical harmonics ${}_sY_{jm}(\phi, \theta)$ of Newman and Penrose (Ref. 1), we conclude that

$${}_sY_{jm}(\phi, \theta) \sim T^j_{sm}(\pi/2 - \phi, \theta, 0).$$

The relation comparable to this one in Ref. 2 was shown to be somewhat different, i.e., given by ${}_sY_{jm}(\phi, \theta) \sim T^j_{s,m}(\phi, \theta, 0)$ [see Eq. (3.11) of Ref. 2]. The reason for this difference is the way Euler's angles are chosen. Our rotation with the angle θ is taken about the x axis, whereas that of Ref. 2 was taken about the y axis.

[14] The field quantities that were expanded by Newman and Penrose are those obtained from contraction of the vector field $\mathbf{v} = \mathbf{E} + i\mathbf{B}$ for the electromagnetic case. For the gravitational field, in both the linear and the full theory of relativity, the quantity analogous to the Maxwell tensor $F_{\mu\nu}$ is the Weyl tensor $C_{\alpha\beta\gamma\delta}$. Just as $F_{\mu\nu}$ corresponds to the two vectors $\mathbf{E}$ and $\mathbf{B}$, the $C_{\alpha\beta\gamma\delta}$ corresponds to two traceless symmetric three-dimensional tensors U_{ij} and V_{ij}. The quantities to be expanded are then those obtained from the tensor $W_{ij} = U_{ij} + iV_{ij}$. Since W_{ij} is symmetric and traceless, one obtains only five quantities instead of the nine quantities which are usually obtained. See, for example, E. T. Newman and R. Penrose, "Some New Gravitationally Conserved Quantities," in *Research on Solutions of the Gravitational Field Equations*, Aerospace Research Laboratories Technical Report No. ARL 67-0053, 1967, p. 115.

TABLE I. Various spin-weight functions W_s obtained from the tensor field **W**. We also give their expansion modes in T^j_{sn}, the latter being evaluated at $\phi_1 = \pi/2 - \phi$ and $\phi_2 = 0$.

Weight s	W_s	Tensor components	Expanded in T^j_{sn}
± 2	$\mathbf{W}\cdot\xi_\pm\xi_\pm$	$W_{\phi\phi} - W_{\theta\theta} \pm i(W_{\theta\phi} + W_{\phi\theta})$	$T^j_{\pm 2,n}(\pi/2-\phi,\theta,0)$
± 1	$\mathbf{W}\cdot\xi_\pm\xi_3$	$-(W_{\phi r} \pm iW_{\theta r})$	$T^j_{\pm 1,n}(\pi/2-\phi,\theta,0)$
	$\mathbf{W}\cdot\xi_3\xi_\pm$	$-(W_{r\phi} \pm iW_{r\theta})$	
0	$\mathbf{W}\cdot\xi_+\xi_-$	$W_{\phi\phi} + W_{\theta\theta} + i(W_{\theta\phi} - W_{\phi\theta})$	$T^j_{0,n}(\pi/2-\phi,\theta,0)$
	$\mathbf{W}\cdot\xi_3\xi_3$	W_{rr}	
	$\mathbf{W}\cdot\xi_-\xi_+$	$W_{\phi\phi} + W_{\theta\theta} - i(W_{\theta\phi} - W_{\phi\theta})$	

We give in Table I the various spin-weight functions W_s obtained from the tensor field **W**. We also give their expansion modes in T^j_{sn}; the latter are evaluated at $\phi_1 = \pi/2 - \phi$ and $\phi_2 = 0$.

The above results were also obtained by Gel'fand and Shapiro[6] by using a very tedious method.

The considerations of this section show how closely related the functions η are to the matrix elements T^j_{mn}. This relationship has even a group-theoretic meaning when T^j_{mn} and η are considered as functions on the group O_3 or the group SU_2. This meaning is pointed out in the next section.

5. CONCLUDING REMARKS

We conclude our discussion by giving a group-theoretic meaning to the expansion of the functions η in terms of the $T^j_{mn}(u)$. It is shown that the transformations which connect the η's realize an infinite-dimensional representation of SU_2 in the space of all functions η. This infinite-dimensional representation is not irreducible; its decomposition into irreducible parts leads to the expansion of $\eta(u)$ in the $T^j_{mn}(u)$, the latter providing invariant subspaces in which irreducible representations of SU_2 act.

We notice that the set of all measurable functions $\eta(u)$ of weight s [i.e., satisfying Eq. (4.8)] which satisfy the condition

$$\int |\eta(u)|^2\, du < \infty \tag{5.1}$$

form a Hilbert space.[10] It is denoted by $L_2^{2s}(SU_2)$.[15]

We now assign for each $u_1 \in SU_2$ an operator V_{u_1} defined in the Hilbert space $L_2^{2s}(SU_2)$ by

$$V_{u_1}\eta(u) = \eta(uu_1), \tag{5.2}$$

where $\eta(u) \in L_2^{2s}(SU_2)$. The correspondence $u_1 \to V_{u_1}$ then realizes an infinite-dimensional representation of SU_2 in the space $L_2^{2s}(SU_2)$. It is, in fact, a unitary representation.[16]

The decomposition of this infinite-dimensional representation into irreducible parts can easily be done if we notice that the matrix elements $T^j_{sn}(u) \in L_2^{2s}(SU_2)$ since they satisfy Eq. (4.8). Therefore,

$$V_{u_1}T^j_{sm}(u) = T^j_{sm}(uu_1). \tag{5.3}$$

Since T^j_{sm} are matrix elements of the irreducible representations of SU_2, we have

$$T^j_{sm}(uu_1) = \sum_{n=-j}^{j} T^j_{sn}(u)T^j_{nm}(u_1). \tag{5.4}$$

Using Eq. (5.3), we obtain

$$V_{u_1}T^j_{sm}(u) = \sum_{n=-j}^{j} T^j_{sn}(u)T^j_{nm}(u_1). \tag{5.5}$$

Accordingly, the transformation V_{u_1} realizes a representation of SU_2 in the space R^s of $(2j+1)$-functions of the sth row of the matrix T^j. Also, the matrix elements of V_{u_1} are $T^j_{nm}(u_1)$. The representation $u_1 \to V_{u_1}$ in the space of functions $T^j_{sn}(u)$, $n = -j, -j+1, \cdots, j$, is irreducible, and the $T^j_{sn}(u)$ form a canonical basis in this space.[17]

The above considerations show that the infinite-dimensional representation of SU_2 in the Hilbert space $L_2^{2s}(SU_2)$ is decomposed into irreducible parts defined in the subspaces R^s of the matrix elements $T^j_{sn}(u)$, where j and s are fixed. In fact these are the only irreducible components of the infinite-dimensional representation. This explains the meaning of the expansions of η in terms of T^j_{sn}.

[15] The scalar product in this Hilbert space is defined by $(\eta, \eta') = \int \eta(u)\overline{\eta'(u)}\, du$, for any $\eta, \eta' \in L_2^{2s}(SU_2)$. It can be shown that the space $L_2^{2s}(SU_2)$ is a closed subspace of the Hilbert space $L^2(SU_2)$ and, therefore, is complete.

[16] This representation is related to the principal series of representations of the unimodular group of order two. See, e.g., Ref. 10.

[17] The operators $H_\pm$, H_3 of this representation are given by

$$H_\pm = e^{\mp i\phi_1}\left(\pm \operatorname{cotan}\theta\frac{\partial}{\partial\phi_1} + i\frac{\partial}{\partial\theta} \mp \operatorname{cosec}\theta\frac{\partial}{\partial\phi_2}\right),$$
$$H_3 = i\frac{\partial}{\partial\phi_1};$$

they satisfy the following relations with respect to the canonical basis $T^j_{m,-j}, T^j_{m,-j+1}, \cdots, T^j_{m,j}$:

$$H_\pm T^j_{mn} = [(j \pm n + 1)(j \mp n)]^{\frac{1}{2}}T^j_{m,n\pm 1},$$
$$H_3 T^j_{mn} = nT^j_{mn},$$

where m, n take the values $-j, -j+1, \cdots, j$. By changing the variables ϕ_1 into ϕ_2 and vice versa and using the relation

$$T^j_{nm}(\phi_2, \theta, \phi_1) = T^j_{mn}(\phi_1, \theta, \phi_2),$$

we obtain

$$K_\pm T^j_{mn} = [(j \pm m + 1)(j \mp m)]^{\frac{1}{2}}T^j_{m\pm 1,n},$$
$$K_3 T^j_{mn} = mT^j_{mn},$$

where $K_\pm$, K_3 are given by $H_\pm$, H_3 with ϕ_1 and ϕ_2 exchanged:

$$K_\pm = e^{\mp i\phi_2}\left(\pm \operatorname{coton}\theta\frac{\partial}{\partial\phi_2} + i\frac{\partial}{\partial\theta} \mp \operatorname{cosec}\theta\frac{\partial}{\partial\theta_1}\right),$$
$$K_3 = i\frac{\partial}{\partial\phi_2}.$$

It then follows that K_+ is most related to $\eth$. See Ref. 2.

Electron in a Given Time-Dependent Electromagnetic Field*

ANTON Z. CAPRI
Palmer Physical Laboratory, Princeton University and
Theoretical Physics Institute, University of Alberta†

(Received 19 June 1968)

The theory of a quantized Dirac field interacting with a classical electromagnetic field is considered. The resulting q-number problem is reduced to a closely related c-number problem. The theory is then shown to be without divergences. The interpolating field is shown to exist and is local. Also, the S matrix is shown to be unitary.

I. INTRODUCTION

The problem of an electron in an external electromagnetic field has been considered by various people, particularly Salam and Matthews[1] and Schwinger.[2] It was shown by their use of the Fredholm theory that the one-particle Green's functions exist for an electron interacting with an external electromagnetic field.

It is also implicit in the work of Salam and Matthews, and explicit in that of Schwinger, that all the Green's functions of the theory can be expressed in terms of the two-point function. However, it is not obvious how the full apparatus of the theory is to be recovered from the Green's functions. The difficulty is that the theory has, in general, two distinct vacuum states: the in vacuum Ψ^0_{in} and the out vacuum Ψ^0_{out}. The two-point Green's function as defined by Schwinger is

$$\frac{(\Psi^0_{\text{out}}, (\psi_{(x)}\psi^*_{(y)})_+ \Psi^0_{\text{in}})}{(\Psi^0_{\text{out}}, \psi^0_{\text{in}})},$$

and it is not obvious how to use the standard reconstruction theorem of Wightman[3] to recover the Hilbert space and field operators. That theorem assumes a unique vacuum state.

We have avoided the difficulties mentioned above by working directly with the fields. These fields are operator-valued distributions[4] and therefore have a meaning only when smeared with suitably smooth test functions. They satisfy a system of linear partial differential equations. The problem of existence of a solution of these equations is reduced here to the same problem for a closely related system for complex-valued functions.

II. THE FREE DIRAC FIELD

The free Dirac operators $\psi_{\text{in}}(x)$ are given in the usual Fock-space representation as presented, for example, in Ref. 5. The Hilbert space $H^{(\text{in})}$ on which they act is given by

$$H^{(\text{in})} = \bigoplus_{n=0}^{\infty} H_n^{(\text{in})} \tag{1}$$

and

$$H_n^{(\text{in})} = \overset{n}{\wedge} H_1^{(\text{in})}, \tag{2}$$

where $\overset{n}{\wedge}$ means the n-fold completely antisymmetric tensor product. The no-particle (vacuum) $H_0^{(\text{in})}$ consists of elements which are complex numbers with the inner product

$$(\Phi_0, \Psi_0) = \Phi_0^* \Psi_0 \tag{3}$$

and

$$H_1^{(\text{in})} = H^{(+)} \oplus H^{(-)}. \tag{4}$$

A state in $H^{(\epsilon)}$, $\epsilon = \pm$, is described by a pair of complex-valued functions of the momentum p labeled by an index α taking on the values $\pm\frac{1}{2}$. (Strictly speaking, the states are equivalence classes of functions. Two functions belong to the same equivalence class if they differ only off the mass shell.) Under the proper orthochronous Lorentz group, these states transform according to the representations $D^{(\frac{1}{2},0)}$ or $D^{(0,\frac{1}{2})}$ of $SL(2, c)$ depending upon whether ϵ is $+$ or $-$, respectively. The inner products in H^{ϵ} are given by

$$(\Phi_\alpha^{(+)}(p), \Psi_\alpha^{(+)}(p)) = \sum_{\alpha\beta} \int_{V_+} \Phi_\alpha^{(+)*}(p) \left(\frac{\tilde{p}}{m}\right)_{\alpha\beta} \Psi_\beta^{(+)}(p)\, d\Omega_m(p), \tag{5a}$$

$$(\Phi_\alpha^{(-)}(p), \Psi_\alpha^{(-)}(p)) = \sum_{\alpha\beta} \int_{V_+} \Phi_\alpha^{(-)*}(p) \left(\frac{\underset{\sim}{p}}{m}\right)_{\alpha\beta} \Psi_\beta^{(-)}(p)\, d\Omega_m(p). \tag{5b}$$

Here,

$$\tilde{p} = p_0 1 - \mathbf{p}\cdot\boldsymbol{\sigma}, \tag{6a}$$

$$\underset{\sim}{p} = p_0 1 + \mathbf{p}\cdot\boldsymbol{\sigma}. \tag{6b}$$

Convenient orthonormal basis functions for H_1^{ϵ} are

$$b^{(+)}_{\alpha,k}(p) = \sum_{\beta=\frac{1}{2}}^{\frac{1}{2}} (\underset{\sim}{p}/m)_{\alpha,\beta} H_{k_\beta}(\mathbf{p})(|\mathbf{p}|^2 + m^2)^{\frac{1}{4}} \quad \text{for } (+), \tag{7a}$$

$$b^{(-)}_{\alpha,k}(p) = \sum_{\beta} (\tilde{p}/m)_{\alpha\beta} H_{k_\beta}(\mathbf{p})(|\mathbf{p}|^2 + m^2)^{\frac{1}{4}} \quad \text{for } (-). \tag{7b}$$

* This work is part of the author's Princeton Ph.D thesis, 1967.
† Present address.
[1] A. Salam and P. T. Matthews, Phys. Rev. **90**, 690 (1953).
[2] J. Schwinger, Phys. Rev. **93**, 615 (1954).
[3] A. S. Wightman, Phys. Rev. **101**, 860 (1956).
[4] A. S. Wightman and L. Gårding, Arkiv Fysik **28**, 129 (1964).
[5] S. S. Schweber, *An Introduction To Relativistic Quantum Field Theory* (Harper and Row Publishers, Inc., New York, 1962).

The $H_{k_\beta}(\mathbf{p})$ may be taken as pairs of ordinary Hermite functions of the three-vector $\mathbf{p}$ so that each k_β corresponds to a triple $(l_\beta, m_\beta, n_\beta)$ of integers.

The basis vectors in $H_n^{(\mathrm{in})\epsilon}$ are conveniently taken as determinants of the b_ϵ's:

$$b^\epsilon_{(\alpha_1 k_1, \cdots, \alpha_n k_n)}(p, \cdots, p_n) = \frac{1}{(n!)^{\frac{1}{2}}} \det\,(b^\epsilon_{\alpha_1 k_1}(p_1), \cdots, b^\epsilon_{\alpha_n k_n}(p_n)). \quad (8)$$

We have occasion to use such a basis set later on.

The field operators $\psi_{\mathrm{in}}(f)$, $\bar\psi_{\mathrm{in}}(f)$ are defined by

$$\begin{aligned}(\psi_{\mathrm{in}}(f)\Phi)^{(n,r)}_{\alpha_1,\cdots,\alpha_n,\beta_1,\cdots,\beta_r}(p_1, \cdots, p_n, q_1, \cdots, q_r) &= \int_{V_+} d\Omega_m(p)\Big\{\sum_\alpha (n+1)^{\frac{1}{2}} \tilde f(p) u_\alpha(p) \\ &\times \Phi^{(n+1,r)}_{\alpha\,\alpha_1,\cdots,\alpha_n,\beta_1,\cdots,\beta_r}(p\, p_1, \cdots, p_n, q_1, \cdots, q_r)\Big\} \\ &+ r^{-\frac{1}{2}} \sum_{j=1}^{r} (-1)^{n+j+1} \tilde f(-q_j) v_{\beta_j}(q_j) \\ &\times \Phi^{(n,r-1)}_{\alpha_1,\cdots,\alpha_n,\beta_1,\cdots,\hat\beta_j,\cdots,\beta_n} \\ &\times (p_1, \cdots, p_n, q_1, \cdots, \hat q_j, \cdots, q_r), \end{aligned} \quad (9a)$$

where $\hat{}$ over a letter means "omit this" and the u_α, v_β are the positive- and negative-frequency solutions of the Dirac equation. The $\tilde f(p)$ are the Fourier transforms of four-component test functions. Throughout this paper we use test functions from the space $\mathcal{S}$ of infinitely differentiable functions that are rapidly decreasing at infinity. We could, in fact, use a larger space consisting of functions which, together with their derivatives, are continuous and L^2. For simplicity, we have avoided this added complication.

With the notation $\bar u = u^* \gamma^0$ we have

$$\begin{aligned}(\bar\psi_{\mathrm{in}}(f)\Phi)^{(n,r)}_{\alpha_1,\cdots,\alpha_n,\beta_1,\cdots,\beta_r}(p_1, \cdots, p_n, q_1, \cdots, q_r) &= \int_{V_+} d\Omega_m(q)\Big\{\sum_\beta (r+1)^{\frac{1}{2}} (-1)^n \bar v_\beta(q) \tilde f(q) \\ &\times \Phi^{(n;r+1)}_{\alpha_1,\cdots,\alpha_n,\beta\,\beta_1,\cdots,\beta_r}(p_1, \cdots, p_n, q\, q_1, \cdots, q_r)\Big\} \\ &+ n^{-1} \sum_{j=1}^{n} (-1)^{j+1} \bar u_{\alpha_j}(p_j) \tilde f(-p_j) \\ &\times \Phi^{(n-1,r)}_{\alpha_1,\cdots,\hat\alpha_j,\cdots,\alpha_n,\beta_1,\cdots,\beta_r} \\ &\times (p_1, \cdots, \hat p_j, \cdots, p_n, q_1, \cdots, q_r). \end{aligned} \quad (9b)$$

Also,

$$\psi_{\mathrm{in}}(f)^* = \bar\psi_{\mathrm{in}}(\bar f); \quad (10)$$

and

$$[\psi_{\mathrm{in}}(f), \psi_{\mathrm{in}}(g)^*]_+ = -iS(f, g), \quad (11)$$

where

$$S(x - y) = -(i\partial\!\!\!/ + m)\Delta(x - y) \quad (12)$$

and

$$S(x - y) = S_a(x - y) - S_r(x - y), \quad (13)$$

where S_a and S_r are the advanced and retarded Green's functions, respectively, for the free Dirac equation.

The fields $\psi_{\mathrm{in}}(x)$ and $\bar\psi_{\mathrm{in}}(x)$ satisfy the equations

$$(-i\gamma \cdot \partial + m)\psi(x) = 0, \quad (14a)$$

$$i\partial_\mu \bar\psi(x)\gamma^\mu + m\bar\psi(x) = 0. \quad (14b)$$

As domain of the field operators we may take all Ψ such that $\Psi^{(n,r)} = 0$ for sufficiently large n, r and such that $\Psi^{(n,r)} \in \mathcal{S}$ when restricted to the direct product of the relevant mass hyperboloids.

III. THE INTERPOLATING FIELD

The most general gauge-invariant interaction of a Dirac particle with an external electromagnetic field is described by the field equations

$$(-i\gamma \cdot \partial + m)\psi = (eA\!\!\!/ + \mu\sigma^{\alpha\beta}F_{\alpha\beta})\psi. \quad (15)$$

Here, $eA\!\!\!/ + \mu\sigma^{\alpha\beta}F_{\alpha\beta} = B$,

$$\sigma^{\alpha\beta} = i^{-1}[\gamma^\alpha, \gamma^\beta]. \quad (16)$$

Using the retarded or advanced free Green's functions, this equation leads to the Källén–Yang–Feldman[6,7] integral equations:

$$\psi(x) = \psi_{\mathrm{in}}(x) + \int S_r(x - y)B(y)\psi(y)\, dy, \quad (17a)$$

$$\psi(x) = \psi_{\mathrm{out}}(x) + \int S_a(x - y)B(y)\psi(y)\, dy. \quad (17b)$$

It is at this point that our procedure differs from the usual one (perturbation expansion). In fact, we convert this q-number problem to a c-number problem.

Since the solutions $\psi(x)$ of these equations (if they exist) define operator-valued distributions, it is natural to consider the smeared equations. We continue to use the notation

$$\psi(f) = \sum_\alpha \int f_\alpha(x)\psi_\alpha(x)\, d^4x, \quad (18)$$

where α labels the spinor components of ψ.

Thus (suppressing the spinor indices),

$$\begin{aligned}\psi(f) &= \psi_{\mathrm{in}}(f) + \int f(x)S_r(x - y)B(y)\psi(y)\, dy\, dx \\ &= \psi_{\mathrm{in}}(f) + \psi[(f * S_r)B],\end{aligned}$$

where the $*$ means convolution. So that, finally,

$$\psi(T_r f) = \psi_{\mathrm{in}}(f), \quad (19a)$$

where

$$T_r f = f - (f * S_r)B \quad (20a)$$

is a mapping of the test-function space $\mathcal{S}$ into itself if

[6] G. Källén, Arkiv Fysik **2**, 371 (1950).
[7] C. N. Yang and D. Feldman, Phys. Rev. **79**, 972 (1950).

B is properly restricted. In fact, if we choose $A_\mu \in \mathcal{S}$ then $B \in \mathcal{S}$.

Now, S_r is a tempered distribution and, therefore, $f * S_r \in O_M$, the space of infinitely differentiable functions of slow growth. Hence, with $B \in \mathcal{S}$ we see that $(f * S_r)B \in \mathcal{S}$. Furthermore, this mapping T_r is continuous from $\mathcal{S}$ into $\mathcal{S}$.

In a similar manner we obtain

$$\psi(T_a f) = \psi_{\text{out}}(f), \tag{19b}$$

with

$$T_a f = f - (f * S_a)B \tag{20b}$$

a continuous mapping from $\mathcal{S}$ into $\mathcal{S}$.

We use the notation

$$Df = i\partial_\mu f\gamma^\mu + mf, \tag{21}$$

$$D_B f = Df + fB. \tag{22}$$

It then follows that

$$T_r(Df) = T_a(Df) = D_B f. \tag{23}$$

IV. THE MAPPINGS T_r AND T_a

We now show that the use of the mappings T_r and T_a reduces the question of the existence of solutions for the fields to a c-number problem. To find the interpolating fields requires finding the inverse mappings T_r^{-1} and T_a^{-1}. If these inverse mappings exist and define a continuous mapping of $\mathcal{S}$ onto $\mathcal{S}$, we are finished. The continuity of the inverse mappings follows from a theorem in Gel'fand and Schilow[8] that states:

If a continuous linear operator A maps a complete, countably normed space $\mathbf{X}$ in a one-to-one invertible fashion onto the space $\mathbf{Y}$, then the inverse operator A^{-1} is also continuous.

Thus, we need only show that the inverse mappings exist from $\mathcal{S}$ onto $\mathcal{S}$. This means that we must show that, for all $h \in \mathcal{S}$, we can find an $f \in \mathcal{S}$ such that

$$T_r f = h \tag{24a}$$

or

$$T_a f = h. \tag{24b}$$

We restrict the discussion to T_r since the discussion for T_a is identical.

Define the auxiliary function

$$g = f * S_r. \tag{25}$$

Then, $g \in O_M$ and due to the retardedness property of S_r, g vanishes as $x^0 \to -\infty$. Furthermore,

$$Dg = f. \tag{26}$$

With zero initial conditions, the unique solution of this equation in O_M is

$$g = f * S_r.$$

Replacing f by Dg in $T_r f$, we get

$$Dg + gB = h. \tag{27}$$

So g must be a solution of the inhomogeneous, interacting field equation with zero initial data. In the case of T_a^{-1} we have zero final data.

The above system of differential equations is strictly hyperbolic and the Cauchy problem is well posed. It is well known[9] that in this case there is a unique solution for $g \in O_M$.

Therefore, T_r^{-1} and T_a^{-1} exist and define continuous one-to-one mappings of $\mathcal{S}$ onto $\mathcal{S}$.

It also follows from the strict hyperbolicity of the operator D_B that for zero initial data

$$\text{supp } g \subset V_+ \text{ supp } h \tag{28a}$$

and for zero final data

$$\text{supp } g \subset V_- \text{ supp } h. \tag{28b}$$

Here, $V_\pm$ supp h is the forward (backward) light cone subtended by the support of h.

From this we can conclude that

$$\text{supp } T_r^{-1} f \subset V_+ \text{ supp } f, \tag{29a}$$

$$\text{supp } T_a^{-1} f \subset V_- \text{ supp } f. \tag{29b}$$

V. RECIPROCITY RELATIONS

In this section we prove and list several identities relating the Green's functions for the Dirac equation and the mappings T_r, T_a.

We start with the following well-known identity for the Dirac Green's functions:

$$(\gamma^0 S_r^* \gamma^0)(-x) = \bar{S}_r(x) = S_a(x), \tag{30a}$$

$$(\gamma^0 S_a^* \gamma^0)(-x) = \bar{S}_a(x) = S_r(x). \tag{30b}$$

Also, since

$$\gamma^0 \gamma^{\mu\dagger} \gamma^0 = \gamma^\mu$$

and

$$\gamma^0 \sigma^{\mu\nu\dagger} \gamma^0 = \sigma^{\mu\nu},$$

we get

$$\overline{B(y)} = B(y). \tag{31}$$

It then follows by writing out the following expression that

$$S_r(T_r f - f, g) = S_r(f, T_a g - g). \tag{32}$$

Whence it follows that

$$S_r(T_r f, g) = S_r(f, T_a g), \tag{33a}$$

[8] I. M. Gel'fand and G. E. Schilow, *Verallgemeinerte Funktionen II* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1962).

[9] *Partial Differential Equations*, Bers, John, and Schechter, Eds. (Interscience Publishers, Inc., New York, 1964).

and replacing f by $T_r^{-1}f$ and g by $T_a^{-1}g$, we get

$$S_r(f, T_a^{-1}g) = S_r(T_r^{-1}f, g). \tag{33b}$$

In a similar manner, it is possible to prove a lot more identities, of which we now make a list since we need many of them later:

$$S_r(T_r f, g) = S_r(f, T_a g), \tag{33a}$$

$$S_a(T_a f, g) = S_a(f, T_r g), \tag{34a}$$

$$S_r(T_a f, g) = S_r(f, T_r g), \tag{35a}$$

$$S_a(T_r f, g) = S_a(f, T_a g), \tag{36a}$$

$$S_a(T_r f, g) - S_r(f, T_r g) = S_a(f, g) - S_r(f, g), \tag{37a}$$

$$S_a(f, T_a g) - S_r(T_a f, g) = S_a(f, g) - S_r(f, g). \tag{38a}$$

The corresponding identities in terms of the inverse mappings are

$$S_r(f, T_a^{-1}g) = S_r(T_r^{-1}f, g), \tag{33b}$$

$$S_a(f, T_r^{-1}g) = S_a(T_a^{-1}f, g), \tag{34b}$$

$$S_r(f, T_r^{-1}g) = S_r(T_a^{-1}f, g), \tag{35b}$$

$$S_a(f, T_a^{-1}g) = S_a(T_r^{-1}f, g), \tag{36b}$$

$$S_a(f, T_r^{-1}g) - S_r(T_r^{-1}f, g) = S_a(T_r^{-1}f, T_r^{-1}g) - S_r(T_r^{-1}f, T_r^{-1}g), \tag{37b}$$

$$S_a(T_a^{-1}f, g) - S_r(f, T_a^{-1}g) = S_a(T_a^{-1}f, T_a^{-1}g - S_r(T_a^{-1}f, T_a^{-1}g). \tag{38b}$$

VI. THE INTERACTING GREEN'S FUNCTION

We now obtain the smeared Green's functions for the interacting Dirac equation

$$\overline{D_B} S_r^B(x, y) = \delta(x - y), \tag{39a}$$

$$\overline{D_B} S_a^B(x, y) = \delta(x - y), \tag{40a}$$

where the differentials act on x and

$$\overline{D_B} = (-i\gamma \cdot \partial + m + B)$$

and

$$D_B S_r^B(x, y) = \delta(x - y), \tag{39b}$$

$$D_B S_a^B(x, y) = \delta(x - y), \tag{40b}$$

where the differentials act on y. We now define the two distributions

$$S_r^B(f, g) = S_r(T_r^{-1}f, g), \tag{41a}$$

$$S_a^B(f, g) = S_a(T_a^{-1}f, g), \tag{41b}$$

and we show that these are the required retarded and advanced Green's functions, respectively.

That the above defined distributions satisfy the correct equations (39) and (40) follows from the identities (33b) and (34b) and the fact that

$$T_r^{-1} D_B f = T_a^{-1} D_B f = Df, \tag{42}$$

so that

$$\begin{aligned} \overline{D_B} S_r^B(f, g) &= S_r(T_r^{-1} D_B f, g) \\ &= S_r(Df, g) \\ &= \delta(f, g). \end{aligned}$$

The proof for the other expressions follows in a similar manner.

That these distributions have the correct support properties follows from (29a) and (29b). Thus, if

$$V_+ \operatorname{supp} f \cap \operatorname{supp} g = \phi \tag{43a}$$

or

$$\operatorname{supp} f \cap V_- \operatorname{supp} g = \phi, \tag{43b}$$

then

$$S_r^B(f, g) = 0, \tag{44}$$

since S_r has support in the forward light cone. This proves that S_r^B has the same support as S_r, namely, the forward light cone.

Similarly, one shows that S_a^B has the same support, the backward light cone, as S_a. Therefore, S_r^B and S_a^B are indeed the retarded and advanced Green's functions, respectively, for the operator D_B.

VII. LOCALITY—COMMUTATION RELATIONS

We constructed ψ_{in} so that

$$[\psi_{\text{in}}(f), \psi_{\text{in}}(g)^*]_+ = iS_r(f, g) - iS_a(f, g).$$

Also, we showed that

$$\psi(f) = \psi_{\text{in}}(T_r^{-1}f),$$

so that

$$\begin{aligned} [\psi(f), \psi(g)^*]_+ &= iS_r(T_r^{-1}f, T_r^{-1}g) - iS_a(T_r^{-1}f, T_r^{-1}g) \\ &= iS_r(T_r^{-1}f, g) - iS_a(f, T_r^{-1}g) \\ &= iS_r(T_r^{-1}f, g) - iS_a(T_a^{-1}f, g) \\ &= iS_r^B(f, g) - iS_a^B(f, g). \end{aligned}$$

This proves that the interpolating field is local. We have used the identities (37b) and (34b) and Eqs. (41a) and (41b).

We now show that the in-field and the out-field satisfy the same commutation relations, that is,

$$[\psi_{\text{out}}(f), \psi_{\text{out}}(g)^*]_+ = [\psi_{\text{in}}(f), \psi_{\text{in}}(g)^*]_+. \tag{45}$$

We recall that

$$\psi(f) = \psi_{\text{in}}(T_r^{-1}f),$$
$$\psi_{\text{out}}(f) = (T_a f).$$

Therefore,

$$\psi_{\text{out}}(f) = \psi_{\text{in}}(T_r^{-1}T_a f) \tag{46}$$

and

$$[\psi_{\text{out}}(f), \psi_{\text{out}}(g)^*]_+ = [\psi_{\text{in}}(T_r^{-1}T_a f), \psi_{\text{in}}(T_r^{-1}T_a f^*)]_+ = iS_r(T_r^{-1}T_a f, T_r^{-1}T_a g) - iS_a(T_r^{-1}T_a f, T_r^{-1}T_a g).$$

But from (37b) and (38b) we conclude that

$$S_r(T_r^{-1}f, T_r^{-1}g) - S_a(T_r^{-1}f, T_r^{-1}g) = S_r(T_a^{-1}f, T_a^{-1}g) - S_a(T_a^{-1}f, T_a^{-1}g) \quad (47)$$

and, hence, that

$$S_r(T_r^{-1}T_a f, T_r^{-1}T_a g) - S_a(T_r^{-1}T_a f, T_r^{-1}T_a g) = S_r(f, g) - S_a(f, g), \quad (48)$$

which proves our assertion that the out-field and the in-field satisfy the same commutation relations.

VIII. UNITARITY OF THE S MATRIX—THE OUT-VACUUM

Since the out-field and the in-field satisfy the same commutation relations, we need only show that, corresponding to the out-field, there exists a vacuum state in the Hilbert space of in-states to conclude that the two fields are unitarily equivalent.[10] The unitary operator connecting them is the S matrix. Recall that the in-field was explicitly constructed to have a vacuum state (corresponding to no particles being present).

We now show that the out-field has a vacuum. As we saw before,

$$\psi_{\text{out}}(f) = \psi_{\text{in}}(T_r^{-1}T_a f).$$

We first show that

$$T_r^{-1}T_a f = f + f_0, \quad (49)$$

where $\text{supp} f_0 \subset \text{supp } B$.

But

$$T_r^{-1}T_a f = Dg, \quad (50)$$

where g is the solution of

$$D_B g = T_a f \quad (51)$$

corresponding to zero initial conditions.

A particular integral of (51) is

$$g_{\text{particular}} = f * S_a. \quad (52)$$

So

$$g = f * S_a + g_0, \quad (53)$$

where

$$D_B g_0 = 0 \quad (54)$$

with initial conditions $f * (S_r - S_a)$.

Hence,

$$f_0 = Dg_0 = -g_0 B \quad (55)$$

and

$$\text{supp} f_0 \subset \text{supp } B,$$

as claimed.

If f^+ is a positive-frequency test function [that is, the Fourier transform $\tilde{f}^+(p_0, \mathbf{p})$ of f^+ vanishes for $p_0 \leq 0$], then by construction

$$\psi_{\text{in}}(f^+)\Phi_{\text{in}}^{(0,0)} = 0. \quad (56)$$

We now show that the equation

$$\psi_{\text{out}}(f^+)\Phi_{\text{out}}^0 = 0 \quad (57)$$

has a solution

$$\Phi_{\text{out}}^0 \in H^{(\text{in})}.$$

Using a complete orthonormal set of basis vectors, as described previously, we look for a solution of the form

$$\Phi_{\text{out}}^0 = \sum_{n,r} a_{n,r}^{(\alpha_1 k_1, \cdots, \alpha_n k_n, \beta_1 j_1, \cdots, \beta_r j_r)} \times b_{(\alpha_1 k_1, \cdots, \alpha_n k_n, \beta_1 j_1, \cdots, \beta_r j_r)}^{(n,r)}(p_1, \cdots, p_n, q_1, \cdots, q_r). \quad (58)$$

We suppress the indices and simply write

$$\Phi_{\text{out}}^0 = \sum_{n,r} a_{n,r} b^{(n,r)}. \quad (59)$$

We now set

$$\psi_{\text{out}}(f^+)\Phi_{\text{out}}^0 = 0.$$

Then we get

$$\psi_{\text{in}}(f^+ + f_0^+)\Phi_{\text{out}}^0 = 0, \quad (60)$$

and component by component this reads

$$a_{n+1,r}\int d\Omega_m(p) \times \left\{\sum_\alpha (n+1)^{\frac{1}{2}}[\tilde{f}^+(p) + \tilde{f}_0^+(p)]u_\alpha(p) b^{(n+1,r)}\right\} = -a_{n,r-1} r^{-\frac{1}{2}} \sum_{j=1}^{r} (-1)^{n+j+1} \tilde{f}_0^+(-q_j) v_{\beta j}(q_j) b^{(n,r-1)}. \quad (61)$$

If we take the inner product of this expression with itself, we get

$$|a_{n+1,r}|^2 I_{n+1,r} = \frac{|a_{n,r-1}|^2}{(n+1)r} K_{n,r-1}, \quad (62)$$

where

$$I_{n+1,r} = \left\| \int d\Omega_m(p) \left\{\sum_\alpha [\tilde{f}^+(p) + \tilde{f}_0^+(p)]u_\alpha(p)\right\} \times b_{(\alpha k \alpha_1 k_1, \cdots, \alpha_n k_n, \beta_1 l_1, \cdots, \beta_r l_r)}^{(n+1,r)} (p p_1, \cdots, p_n, q_1, \cdots, q_r) \right\|^2 \quad (63)$$

[10] L. Gårding and A. S. Wightman, Proc. Natl. Acad. Sci. U.S. **40**, 617 (1954).

and

$$K_{n,r-1} = \Big\| \sum_{j=1}^{r} (-1)^{n+j+1} \tilde{f}_0^+(-q_j) v_{\beta_j}(q_j) \times b^{(n,r-1)}_{(\alpha_1 k_1, \cdots, \alpha_n k_n, \beta_1 l_1, \cdots, \hat{\beta}_j \hat{l}_j, \cdots, \beta_r l_r)} \times (p_1, \cdots, p_n, q_1, \cdots, \hat{q}_j, \cdots, q_r) \Big\|^2. \tag{64}$$

Using (62), we can express $|a_{n+r,r}|^2$ in terms of $|a_{n,0}|^2$ and $|a_{n,r+n}|^2$ in terms of $|a_{0,r}|^2$. But if we consider the expression

$$(\psi_{\text{in}}(\tilde{f}^+ + \tilde{f}_0^+)\Phi_{\text{in}})^{(n-1,0)} = \int d\Omega_m(p) \Big\{ \sum_\alpha (n)^{\frac{1}{2}} [\tilde{f}^+(p) + \tilde{f}_0^+(p)] u_\alpha(p) \Big\} \times \Phi^{(n,0)}_{\alpha\alpha_1, \cdots, \alpha_{n-1}}(p\, p_1, \cdots, p_{n-1}) \tag{65}$$

and use (60), we get

$$a_{n,0} = 0, \quad n \geq 1, \tag{66a}$$

and from the adjoint of (65) we get

$$a^*_{0,r} = 0, \quad r \geq 1. \tag{66b}$$

Therefore, the only nonvanishing coefficients are of the form $a_{n,n}$ and we call them a_n. The conditions (66) state that particles are created and destroyed only in pairs.

We must now estimate $I_{n+1,n+1}$ and $K_{n,n}$. Integrating out the variables that can be done trivially, we get

$$I_{n+1,n+1} = \Big\| \int d\Omega_m(p) \Big\{ \sum_\alpha [\tilde{f}^+(p) + \tilde{f}_0^+(p)] u_\alpha(p) \Big\} \times [(n+1)!]^{-\frac{1}{2}} \det [b^{(+)}_{\alpha_1 k_1, \ldots}(p_1), \cdots, b^{(+)}_{\alpha_{n+1} k_{n+1}}(p_{n+1})] \Big\|^2 \tag{67}$$

$$\geq \frac{1}{n+1} \sum_{i=1}^{n+1} \int |\tilde{f}^+(p) + \tilde{f}_0^+(p)] u_\alpha(p) b_{\alpha_i k_i}(p)|^2 \, d\Omega_m(p) \tag{68}$$

$$\geq \frac{n+1}{n+1} |\theta_2|^2 = |\theta_2|^2 > 0. \tag{69}$$

The first inequality comes from the fact that, in squaring the determinant, the cross terms obtained are either zero or positive due to the orthogonality of the basis functions.

In the expression for $K_{n,n}$, we expand $\tilde{f}_0^+(-q_j)v_{\beta_j}(q_j)$ in terms of the basis vectors for $H_1^{(-)}$:

$$\tilde{f}_0^+(-q_j) v_{\beta_j}(q_j) = \sum_{r=1}^{\infty} C_r b^{(-)}_{\beta_j k_r}(q_j). \tag{70}$$

Then,

$$K_{n,n} = (n+1) \, \| \sum C_r [(n+1)!]^{-\frac{1}{2}} \times \det [b^{(-)}_{\beta_1 k_1}(q_1), \cdots, b^{(-)}_{\beta_{n+1} k_{n+1}}(q_{n+1})] \| \tag{71}$$

$$= (n+1) \sum |C_r|^2 \tag{72}$$

$$= (n+1) \, |\theta_1|^2 < \infty; \tag{73}$$

since

$$f_0^+ \in \mathcal{S},$$

combining the estimates (69) and (73), we get that

$$|a_{n+1}|^2 \leq \frac{|a_n|^2}{n+1} \left| \frac{\theta_1}{\theta_2} \right|^2.$$

Let

$$\left| \frac{\theta_1}{\theta_2} \right| = \theta. \tag{74}$$

Then,

$$|a_{n+1}|^2 \leq \frac{|a_0|^2}{(n+1)!} \theta^{2(n+1)}. \tag{75}$$

Hence,

$$\sum |a_n|^2 \leq |a_0|^2 \sum \frac{(\theta^2)^n}{n!} = |a_0|^2 \, e^{\theta^2} < \infty,$$

$$\therefore \quad \Phi^0_{\text{out}} \in H^{\text{in}}$$

and the out-field has a vacuum state. It is clear that the out-field is irreducible since it satisfies the free Dirac equation and therefore the S matrix is unitary.

IX. CONCLUSION

We have considered the theory of a Dirac field interacting with a classical electromagnetic field. We have shown that the interpolating field exists and is local and furthermore that there is a *unitary* S matrix. The theory is completely free of any divergences and, hence, no renormalization is needed. This is due to the fact that the field equations are *linear* and it is not necessary to give a meaning to products of field operators. If one wishes to define local observables such as the current, then a renormalization is needed. This renormalization amounts to extracting the "correct singular part" from the product of distributions. No statement was made about the definition of such quantities.

The essential point (that seems to have been missed up till now) is that, in a simple theory of this sort, the question of the existence of the fields can be clearly separated from the question of how one should define local observables.

ACKNOWLEDGMENTS

It is my sincere pleasure to acknowledge the help and encouragement of Professor A. S. Wightman during the research described in this paper. I am also grateful to Princeton University and the Theoretical Physics Institute of the University of Alberta for financial support.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 MARCH 1969

Invariant Imbedding and Case Eigenfunctions*

S. PAHOR† AND P. F. ZWEIFEL‡
Department of Nuclear Engineering, The University of Michigan, Ann Arbor, Michigan

(Received 20 June 1968)

A new approach to the solution of transport problems, based on the ideas introduced into transport theory by Ambarzumian, Chandrasekhar, and Case, is discussed. To simplify the discussion, the restriction to plane geometry and one-speed isotropic scattering is made. However, the method can be applied in any geometry with any scattering model, so long as a complete set of infinite-medium eigenfunctions is known. First, the solution for the surface distributions is sought. (In a number of applications this is all that is required.) By using the infinite-medium eigenfunctions, a system of singular integral equations together with the uniqueness conditions is derived for the surface distributions in a simple and straightforward way. This system is the basis of the theory. It can be reduced to a system of Fredholm integral equations or to a system of nonlinear integral equations, suitable for numerical computations. Once the surface distributions are known, the complete solution can be found by quadrature by using the full-range completeness and orthogonality properties of the infinite-medium eigenfunctions. The method is compared with the standard methods of invariant imbedding, singular eigenfunctions, and a new procedure recently developed by Case.

I. INTRODUCTION

In the past 50 years or so, a number of methods have been devised for solving the neutron (or radiation) transport equations. Excluding strictly approximation procedures such as spherical-harmonics expansions, discrete-ordinate methods, etc.,[1-3] the most important schemes are the Wiener–Hopf method, which is described in detail in Ref. 2, the invariant-imbedding technique, first introduced to transport theory by Ambarzumian[4] and developed extensively by Chandrasekhar[1] and others,[5] and the Case eigenfunction-expansion method.[3,6]

Historically, the first exact method was the Wiener–Hopf method. Because it was basically simpler, the invariant-imbedding method became more popular after its introduction. Eventually, the eigenfunction-expansion approach became more widely used than either of those methods for a number of reasons which are discussed below. (The Wiener–Hopf method is in fact identical with Case's method in the sense that any problem which can be solved by one method can be solved also by the other. Because Case's method is simpler and more familiar, we will not discuss the Wiener–Hopf method further.)

We first note that the traditional derivations of the equations of invariant imbedding are based upon intuitive physical arguments which, by virtue of the known existence of unique solutions of the transport equations,[7] are, in fact, spurious. However, this approach has some real advantages for numerical computation. On the other hand, it does not give complete knowledge of the neutron distribution in a given medium, but only the reflected and transmitted intensities. (Admittedly, in a number of applications these are all that are required.)

A really more serious disadvantage of the invariant-imbedding equations is that they are, in general, not uniquely soluble. To guarantee a unique solution, additional conditions must be imposed.[1] These conditions cannot be obtained from the original invariant-imbedding arguments, and so must be introduced in a somewhat arbitrary manner.

The Case method,[6] on the other hand, has the virtue of simplicity and familiarity, since it is based on an eigenfunction-expansion technique which is already well known to physicists from applications in "classical" boundary-value problems. Furthermore, no intuitive arguments and no extraneous conditions are necessary in order to derive the equations and to guarantee unique solutions. However, by straightforward application of this method, more information is frequently obtained than is really required (as, for example, the neutron distribution everywhere rather than at a surface) and reducing the results to numerics is highly nontrivial.[8]

The major purpose of the present paper is to rederive the nonlinear integral equations of invariant

* Work supported by the National Science Foundation.

† On leave from the University of Ljubljana, Yugoslavia.

‡ Present address: Dept. of Physics, Virginia Polytechnic Institute, Blacksburg, Virginia 24060.

[1] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950).

[2] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).

[3] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1967).

[4] V. A. Ambarzumian, *Theoretical Astrophysics* (Pergamon Press, Inc., New York, 1958).

[5] R. E. Bellman, H. H. Kagiwada, R. E. Kalaba, and M. C. Prestrud, "Invariant Imbedding and Time-Dependent Transport Processes," The Rand Corporation, R-423-ARPA, 1964.

[6] K. M. Case, Ann. Phys. (N.Y.) 9, 1 (1960).

[7] K. M. Case and P. F. Zweifel, J. Math. Phys. 4, 1367 (1963).

[8] M. R. Mendelson, thesis, The University of Michigan, 1964.

imbedding and the uniqueness conditions in a fashion which does not suffer from the deficiencies noted above. This is accomplished by using the Case infinite-medium eigenfunctions. The nonlinear integral equations follow from a system of singular integral equations, which are themselves derived in a simple and straightforward way from Case's eigenfunctions. It is interesting to compare our derivation with those of Sobolev,[9] Busbridge,[10] and Mullikin.[11]

We deal primarily with slab problems—in the limit, of course, half-space results are obtained. The familiar restriction to plane geometry and one-speed isotropic scattering is made. However, the method can be applied in any geometry with any scattering model (e.g., multivelocity anisotropic scattering) so long as a complete set of infinite-medium eigenfunctions is known.

The results we obtain are not new. However, we do feel that our approach yields a coherent, mathematically satisfying, and simple derivation of singular integral equations and equivalent invariant-imbedding nonlinear integral equations, together with the conditions which guarantee unique solution.

In Sec. II, we give a brief review of Case's eigenfunctions and their properties. Then, in Sec. III, the system of singular integral equations and the nonlinear integral equations—together with the conditions guaranteeing uniqueness—are derived for the slab. In Sec. IV, some remarks are made for the half-space problems.

II. THE CASE EIGENFUNCTIONS

We begin with the Case eigenfunctions of the one-speed one-dimensional transport equation with isotropic scattering

$$\left(\mu\frac{\partial}{\partial x}+1\right)\psi(x,\mu)=\frac{c}{2}\int_{-1}^{1}\psi(x,\mu')\,d\mu'. \tag{1}$$

These eigenfunctions may be written in the form[3,6]

$$\psi_\nu(x,\mu)=\phi(\nu,\mu)e^{-x/\nu}, \tag{2a}$$

with

$$\phi(\nu,\mu)=\frac{c\nu}{2}P\frac{1}{\nu-\mu}+\lambda(\nu)\delta(\nu-\mu),\quad \nu\in(-1,1), \tag{2b}$$

$$\phi(\pm\nu_0,\mu)=\frac{c\nu_0}{2}\frac{1}{\nu_0\mp\mu}, \tag{2c}$$

$$\int_{-1}^{1}\phi(\nu,\mu)\,d\mu=1,\quad \nu\in(-1,1),\quad \nu=\pm\nu_0. \tag{2d}$$

Here the discrete eigenvalue ν_0 is a root of the dispersion function

$$\Lambda(z)=1-\frac{cz}{2}\int_{-1}^{1}\frac{d\mu}{z-\mu}. \tag{3}$$

The quantity $\lambda(\nu)$ which appears in Eq. (2b) is related to the boundary values of the dispersion function $\Lambda(z)$ on the branch cut $(-1, 1)$. In fact,

$$\lambda(\nu)=\tfrac{1}{2}[\Lambda^+(\nu)+\Lambda^-(\nu)], \tag{4}$$

where

$$\Lambda^\pm(\nu)=\lim_{0<\epsilon\to 0}\Lambda(\nu\pm i\epsilon),\quad \nu\in(-1,1). \tag{5}$$

We note that c, the mean number of neutrons emitted per collision, will always be assumed to be such that the slab is "subcritical." For $c<1$, this is certainly true for all slab thickness.

The eigenfunctions are orthogonal in the sense that

$$\int_{-1}^{1}\mu\phi(\nu,\mu)\phi(\nu',\mu)\,d\mu=0,\quad \nu\neq\nu'. \tag{6}$$

In fact, the normalization integrals are also known:

$$\int_{-1}^{1}\mu\phi^2(\pm\nu_0,\mu)\,d\mu=\pm N(\nu_0), \tag{7a}$$

$$\int_{-1}^{1}\mu\phi(\nu,\mu)\phi(\nu',\mu)\,d\mu=N(\nu)\delta(\nu-\nu'),\quad \nu\in(-1,1), \tag{7b}$$

where

$$N(\nu_0)=\frac{c}{2}\nu_0^2\left.\frac{\partial\Lambda(z)}{\partial z}\right|_{z=\nu_0}, \tag{8a}$$

$$N(\nu)=\nu\Lambda^+(\nu)\Lambda^-(\nu). \tag{8b}$$

(All of the above results, which are well known, are restated merely for convenience.)

We now consider the solution of the so-called albedo problem for a slab. This is the problem of determining the distribution of neutrons everywhere in a source-free slab due to an incident beam. We seek the solution, denoted as $\psi(0,\mu_0\to x,\mu;\tau)$, to the homogeneous transport equation (1) subject to the boundary conditions

$$\psi(0,\mu_0\to 0,\mu;\tau)=\delta(\mu_0-\mu),\quad \mu_0>0,\quad \mu>0,$$
$$\psi(0,\mu_0\to\tau,-\mu;\tau)=0,\quad \mu>0, \tag{9}$$

where τ is the thickness of a slab.

We expand the solution $\psi(0,\mu_0\to x,\mu;\tau)$ in terms of the eigenfunctions. That is,

$$\psi(0,\mu_0\to x,\mu;\tau)=A(\nu_0)\phi(\nu_0,\mu)e^{-x/\nu_0}+A(-\nu_0)\phi(-\nu_0,\mu)e^{x/\nu_0}+\int_{-1}^{1}A(\nu)\phi(\nu,\mu)e^{-x/\nu}\,d\nu. \tag{10}$$

The Case procedure is to determine the expansion coefficient as discussed earlier.

[9] V. V. Sobolev, *A Treatise on Radiative Transfer* (D. Van Nostrand Inc., Princeton, N.J., 1963).
[10] I. W. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge University Press, London, 1960).
[11] T. W. Mullikin, Astrophys. J. **136**, 627 (1962); **139**, 379, 1267 (1964).

However, by using the set of eigenfunctions in another way, we are led at once to a system of singular integral equations for the reflected and transmitted intensities.

III. DERIVATION OF INVARIANT-IMBEDDING EQUATIONS

A. Albedo Problem for a Slab

Let us first consider the solution of the albedo problem defined in the previous section, for it will be shown that the solution of any problem can be expressed in terms of the albedo solution (Sec. III.B). We consider a slab whose left-hand surface is at $x = 0$, and whose right-hand surface is at $x = \tau$.

We are primarily interested in the reflected and the transmitted distributions $\psi(0, \mu_0 \to 0, -\mu; \tau)$ and $\psi(0, \mu_0 \to \tau, \mu; \tau)$, $\mu > 0$. From the reciprocity theorem for one-speed theory[3,12] it follows that these distributions satisfy the relations

$$\mu\psi(0, \mu_0 \to 0, -\mu; \tau) = \mu_0\psi(0, \mu \to 0, -\mu_0; \tau), \quad (11)$$

$$\mu\psi(0, \mu_0 \to \tau, \mu; \tau) = \mu_0\psi(0, \mu \to \tau, \mu_0; \tau), \quad \mu_0 > 0, \quad \mu > 0. \quad (12)$$

In view of these relations, it is convenient to introduce so-called Ambarzumian–Chandrasekhar's S and T functions, defined as[1]

$$(1/2\mu)S(\tau; \mu_0, \mu) = \psi(0, \mu_0 \to 0, -\mu; \tau) \quad (13)$$

and

$$(1/2\mu)T(\tau; \mu_0, \mu) + \delta(\mu_0 - \mu)e^{-\tau/\mu_0} = \psi(0, \mu_0 \to \tau, \mu; \tau). \quad (14)$$

Both functions are symmetric:

$$S(\tau; \mu_0, \mu) = S(\tau; \mu, \mu_0), \quad (15)$$

$$T(\tau; \mu_0, \mu) = T(\tau; \mu, \mu_0). \quad (16)$$

The reflected and transmitted distributions of an albedo problem $\psi(0, -\mu; \tau)$ and $\psi(\tau, \mu; \tau)$, $\mu > 0$, for a given incident distribution $\psi(0, \mu; \tau)$, $\mu > 0$, can be then expressed as

$$\psi(0, -\mu; \tau) = \frac{1}{2\mu}\int_0^1 S(\tau; \mu', \mu)\psi(0, \mu'; \tau)\, d\mu', \quad (17)$$

$$\psi(\tau, \mu; \tau) = \psi(0, \mu; \tau)e^{-\tau/\mu} + \frac{1}{2\mu}\int_0^1 T(\tau; \mu', \mu)\psi(0, \mu'; \tau)\, d\mu', \quad \mu > 0. \quad (18)$$

We now derive a system of singular integral equations for S and T by using the intuitive invariant-imbedding arguments.[13,14]

Let us take any exponentially decreasing infinite-medium eigenfunction

$$\phi(\nu, \mu)e^{-x/\nu}, \quad \nu \in (0, 1), \quad \nu = \nu_0. \quad (19)$$

The function $\phi(\nu, \mu)e^{-x/\nu}$ describes a distribution of neutrons for the infinite medium. At $x = 0$, the angular density $\phi(\nu, -\mu)$, $\mu > 0$, can be thought of as resulting from the reflection of the "incident" distribution $\phi(\nu, \mu)$, $\mu > 0$, on the slab of thickness τ, and from the transmission of the "incident" distribution $\phi(\nu, -\mu)e^{-\tau/\nu}$, $\mu > 0$, at $x = \tau$, through the same slab. Therefore, in view of Eqs. (17) and (18), we have

$$[1 - e^{-\tau(1/\nu+1/\mu)}]\phi(\nu, -\mu) = \frac{1}{2\mu}\int_0^1 S(\tau; \mu', \mu)\phi(\nu, \mu')\, d\mu' + \frac{e^{-\tau/\nu}}{2\mu}\int_0^1 T(\tau; \mu', \mu)\phi(\nu, -\mu')\, d\mu', \quad \mu > 0, \quad \nu \in (0, 1), \quad \nu = \nu_0. \quad (20)$$

Similarly, by taking any exponentially increasing eigenfunction

$$\phi(-\nu, \mu)e^{x/\nu}, \quad \nu \in (0, 1), \quad \nu = \nu_0, \quad (21)$$

and reasoning as before, we get

$$(e^{-\tau/\nu} - e^{-\tau/\mu})\phi(\nu, \mu) = \frac{e^{-\tau/\nu}}{2\mu}\int_0^1 S(\tau; \mu', \mu)\phi(-\nu, \mu')\, d\mu' + \frac{1}{2\mu}\int_0^1 T(\tau; \mu', \mu)\phi(\nu, \mu')\, d\mu', \quad \mu > 0, \quad \nu \in (0, 1), \quad \nu = \nu_0. \quad (22)$$

For $\nu \in (0, 1)$, the above equations constitute a system of singular integral equations for S and T, while for $\nu = \nu_0$ we obtain two conditions which must be satisfied by S and T.

Because Eqs. (20) and (22) are the basis for our further discussion, we now rederive them rigorously, without appealing to the above intuitive invariant-imbedding arguments. Actually, the rigorous derivation is even simpler than the intuitive one given above.

To see this, let us define an albedo problem by the following boundary conditions:

$$\psi(0, \mu; \tau) = \phi(\nu, \mu), \quad \psi(\tau, -\mu; \tau) = \phi(\nu, -\mu)e^{-\tau/\nu}, \quad \mu > 0, \quad \nu \in (0, 1), \quad \nu = \nu_0, \quad 0 \le x \le \tau. \quad (23)$$

[12] K. M. Case, Rev. Mod. Phys. **29**, 651 (1957).

[13] S. Pahor and I. Kuščer, Astrophys. J. **143**, 888 (1966).

[14] S. Pahor, Nucl. Sci. Eng. **29**, 248 (1967).

It can be easily verified by inspection that the *unique* solution of this particular problem is simply

$$\psi(x, \mu; \tau) = \phi(\nu, \mu)e^{-x/\nu} \tag{24}$$

(because it is a solution of the transport equation and obeys the boundary conditions).

By applying Eqs. (17) and (18) to this solution, we get Eqs. (20) and (21). This represents a rigorous derivation of Eqs. (20) and (21). The same system of singular integral equations, including anisotropic scattering, was already derived by Sobolev[9] and Mullikin.[11] However, our derivation of these equations is much simpler than that of Sobolev and Mullikin; furthermore, it is evident how the described technique could be applied to any geometry and scattering model, once the complete set of infinite medium eigenfunctions is known. (Even if the set is not complete, we obtain in this way some information on the surface distribution. However, the resulting equations are not uniquely soluble.)

It is interesting to compare the present approach with the approach recently developed by Case,[15] where the infinite-medium Green's function is used as a starting point. In both cases, first the integral equations for the surface distributions are derived. However, the corresponding equations are different, though equivalent, and the kernels of Eqs. (20) and (21), yielded by the present method, are somewhat simpler.

The functions $S(\tau; \mu_0, \mu)$ and $T(\tau; \mu_0, \mu)$ can be expressed in terms of Ambarzumian–Chandrasekhar's $X(\mu)$ and $Y(\mu)$ functions of a single variable, with τ as a parameter, which are more suitable for numerical computations than $S(\tau; \mu_0, \mu)$ and $T(\tau; \mu_0, \mu)$.

Let us integrate Eqs. (20) and (22) over μ from 0 to 1. Defining new $X(\mu)$ and $Y(\mu)$ functions as

$$X(\mu) = 1 + \frac{1}{2}\int_0^1 S(\tau; \mu', \mu)\frac{d\mu'}{\mu'}, \tag{25}$$

$$Y(\mu) = e^{-\tau/\mu} + \frac{1}{2}\int_0^1 T(\tau; \mu', \mu)\frac{d\mu'}{\mu'}, \tag{26}$$

and using the normalization condition (2d), we get a system of equations for $X(\mu)$ and $Y(\mu)$:

$$1 = \int_0^1 X(\mu')\phi(\nu, \mu')\,d\mu' + e^{-\tau/\nu}\int_0^1 Y(\mu')\phi(\nu, -\mu')\,d\mu', \tag{27}$$

$$1 = \int_0^1 X(\mu')\phi(\nu, -\mu')\,d\mu' + e^{\tau/\nu}\int_0^1 Y(\mu')\phi(\nu, \mu')\,d\mu', \quad \nu \in (0, 1), \quad \nu = \nu_0. \tag{28}$$

By introducing new functions $Z(\mu)$ and $W(\mu)$ as

$$Z(\mu) = X(\mu) + Y(\mu), \tag{29}$$

$$W(\mu) = X(\mu) - Y(\mu), \tag{30}$$

we obtain for these functions two uncoupled equations

$$1 + e^{-\tau/\nu} = \int_0^1 Z(\mu')\phi(\nu, \mu')\,d\mu' + e^{-\tau/\nu}\int_0^1 Z(\mu')\phi(\nu, -\mu')\,d\mu', \tag{31}$$

$$1 - e^{-\tau/\nu} = \int_0^1 W(\mu')\phi(\nu, \mu')\,d\mu' - e^{-\tau/\nu}\int_0^1 W(\mu')\phi(\nu, -\mu')\,d\mu', \quad \nu \in (0, 1), \quad \nu = \nu_0. \tag{32}$$

Singular integral equations, such as Eq. (31) or Eq. (32), with the condition for $\nu = \nu_0$ included, are equivalent to certain Fredholm integral equations.[16] These Fredholm integral equations were studied in detail by Leonard and Mullikin[17] and they derived conditions which guarantee the existence and uniqueness of the solution. In our case of isotropic scattering, these conditions are satisfied for all subcritical c and certainly for $c < 1$. Therefore, also Eqs. (31) and (32), Eqs. (27) and (28), and Eqs. (20) and (22) are uniquely soluble.

What remains to be done is to express $S(\tau; \mu_0, \mu)$ and $T(\tau; \mu_0, \mu)$ in terms of $X(\mu)$ and $Y(\mu)$. In deriving these relations, we obtain for $X(\mu)$ and $Y(\mu)$ a system of nonlinear integral equations which are convenient for numerical computations.

We introduce two new functions $R(\tau; \mu_0, \mu)$ and $U(\tau; \mu_0, \mu)$ as

$$\left(\frac{1}{\mu_0} + \frac{1}{\mu}\right)S(\tau; \mu_0, \mu) = cR(\tau; \mu_0, \mu), \tag{33}$$

$$\left(\frac{1}{\mu_0} - \frac{1}{\mu}\right)T(\tau; \mu_0, \mu) = cU(\tau; \mu_0, \mu), \tag{34}$$

and we substitute them for $S(\tau; \mu_0, \mu)$ and $T(\tau; \mu_0, \mu)$ in Eqs. (20) and (22). By using the explicit form of eigenfunctions (2b) and (2c) for $\nu \in (0, 1)$ and $\nu = \nu_0$, we get, after a partial-fraction analysis and taking into

[15] K. M. Case, *Proceedings of the Symposium on Transport Theory, April, 1967* (American Mathematical Society, Providence, R.I.) (to be published).

[16] N. I. Muskelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953).

[17] A. Leonard and T. W. Mullikin, J. Math. Phys. **5**, 399 (1964).

account Eqs. (25) and (26), the following four equations:

$$X(\mu) - e^{-\tau/\nu}Y(\mu) = \lambda(\nu)R(\tau;\nu,\mu) + \frac{c\nu}{2}P\int_0^1 \frac{R(\tau;\mu',\mu)}{\nu-\mu'}\,d\mu' - \frac{c\nu}{2}\int_0^1 \frac{U(\tau;\mu',\mu)}{\nu+\mu'}\,d\mu' e^{-\tau/\nu}, \quad (35)$$

$$e^{-\tau/\nu}X(\mu) - Y(\mu) = -\lambda(\nu)U(\tau;\nu,\mu) - \frac{c\nu}{2}P\int_0^1 \frac{U(\tau;\mu',\mu)}{\nu-\mu'}\,d\mu' - \frac{c\nu}{2}e^{-\tau/\nu}\int_0^1 \frac{R(\tau;\mu',\mu)}{\nu+\mu'}\,d\mu', \quad (36)$$

$$X(\mu) - e^{-\tau/\nu_0}Y(\mu) = \frac{c\nu_0}{2}\int_0^1 \frac{R(\tau;\mu',\mu)}{\nu_0-\mu'}\,d\mu' - \frac{c\nu_0}{2}\int_0^1 \frac{U(\tau;\mu',\mu)}{\nu_0+\mu'}\,d\mu' e^{-\tau/\nu_0}, \quad (37)$$

$$e^{-\tau/\nu_0}X(\mu) - Y(\mu) = -\frac{c\nu_0}{2}\int_0^1 \frac{U(\tau;\mu',\mu)}{\nu-\mu'}\,d\mu' - \frac{c\nu_0}{2}e^{-\tau/\nu_0}\int_0^1 \frac{R(\tau;\mu',\mu)}{\nu+\mu'}\,d\mu'. \quad (38)$$

Now, if we first multiply Eq. (27) by $X(\mu)$ and subtract Eq. (28) multiplied by $Y(\mu)$, then multiply Eq. (27) by $Y(\mu)$ and subtract Eq. (28) multiplied by $X(\mu)$, we get equations identical to Eqs. (35) to (38), except that

$$R(\tau;\mu_0,\mu) \to X(\mu)X(\mu_0) - Y(\mu)Y(\mu_0), \quad (39)$$

$$U(\tau;\mu_0,\mu) \to Y(\mu)X(\mu_0) - Y(\mu_0)X(\mu). \quad (40)$$

Thus, the above bilinear expressions are solutions of Eqs. (35)–(38). These solutions are also unique, because Eqs. (35)–(38) uniquely determine $R(\tau;\mu_0,\mu)$ and $U(\tau;\mu_0,\mu)$.

By expressing $S(\tau;\mu_0,\mu)$ and $T(\tau;\mu_0,\mu)$ in Eqs. (25) and (26) in terms of $X(\mu)$ and $Y(\mu)$ [via Eqs. (39), (40), (33), and (34)], we get a system of nonlinear integral equations for $X(\mu)$ and $Y(\mu)$,

$$X(\mu) = 1 + \frac{c\mu}{2}\int_0^1 \frac{X(\mu)X(\mu') - Y(\mu)Y(\mu')}{\mu+\mu'}\,d\mu', \quad (41)$$

$$Y(\mu) = e^{-\tau/\mu} + \frac{c\mu}{2}\int_0^1 \frac{Y(\mu)X(\mu') - X(\mu)Y(\mu')}{\mu-\mu'}\,d\mu', \quad (42)$$

with the conditions, which must be satisfied by $X(\mu)$ and $Y(\mu)$,

$$1 = \frac{c\nu_0}{2}\int_0^1 \frac{X(\mu')}{\nu_0-\mu'}\,d\mu' + \frac{c\nu_0}{2}e^{-\tau/\nu_0}\int_0^1 \frac{Y(\mu')}{\nu_0+\mu'}\,d\mu', \quad (43)$$

$$1 = \frac{c\nu_0}{2}\int_0^1 \frac{X(\mu')}{\nu_0+\mu'}\,d\mu' + \frac{c\nu_0}{2}e^{\tau/\nu_0}\int_0^1 \frac{Y(\mu')}{\nu_0-\mu'}\,d\mu', \quad (44)$$

following from Eqs. (27) and (28) for $\nu = \nu_0$.

Let us now also show that the system of nonlinear integral equations (41) and (42), together with the conditions (43) and (44), uniquely determine $X(\mu)$ and $Y(\mu)$.

First, we note that $X(\mu)$ and $Y(\mu)$ can be analytically continued outside the interval (0, 1), by using Eqs. (41) and (42). It can be easily verified[10] that if $X(\mu)$ and $Y(\mu)$ satisfy Eqs. (41) and (42), but not necessarily (43) and (44), they also satisfy the integral equations

$$\Lambda(z)X(z) = 1 - \frac{cz}{2}\int_0^1 \frac{X(\mu)}{z-\mu}\,d\mu - \frac{zc}{2}e^{-\tau/z}\int_0^1 \frac{Y(\mu)}{z+\mu}\,d\mu, \quad (45)$$

$$\Lambda(z)Y(z) = e^{-\tau/z}\left[1 - \frac{cz}{2}\int_0^1 \frac{X(\mu)}{z+\mu}\,d\mu\right] - \frac{zc}{2}\int_0^1 \frac{Y(\mu)}{z-\mu}\,d\mu, \quad z \notin (-1,1). \quad (46)$$

By applying the Plemelj formula[16] to the above equations for $z \in (0,1)$, we get the singular integral equations (27) and (28). Since these singular integral equations, together with the conditions (43) and (44), uniquely determine $X(\mu)$ and $Y(\mu)$, the same is true for the nonlinear integral equations (41) and (42) combined with the conditions (43) and (44).

We can now easily prove that $X(z)$ and $Y(z)$ are analytic functions in the whole complex plane, except at $z = 0$, where they have an essential singularity.

Since $\Lambda(z) = \Lambda(-z)$, we see at once from Eqs. (45) and (46) that $X(z)$ and $Y(z)$ satisfy the relation[10]

$$Y(z) = e^{-\tau/z}X(-z), \quad (47)$$

which is valid in the whole complex plane. In view of Eqs. (41), (42), (45), (46), and the conditions (43) and (44), the $X(z)$ and $Y(z)$ could be singular only for $z = -\nu_0$ and $z \in (-1, 0)$. However, since $X(z)$ and $Y(z)$ are analytic for $z = \nu_0$ and $z \in (0, 1)$, $z = 0$ excluded, the same is true also for $z = -\nu_0$ and $z \in (-1, 0)$, because of Eq. (47), while it follows from Eqs. (41) and (42) that, for $z = 0$, the functions $X(z)$ and $Y(z)$ have an essential singularity.

B. Green's Function for a Slab

We now show how other slab problems can be solved with the help of the solution for the albedo problem. Evidently, what we need is the solution of the Green's function problem, defined by the non-homogeneous transport equation

$$\left(\mu\frac{\partial}{\partial x}+1\right)G(x_0,\mu_0\to x,\mu;\tau)$$
$$=\frac{c}{2}\int_{-1}^{1}G(x_0,\mu_0\to x,\mu';\tau)\,d\mu'$$
$$+\delta(\mu_0-\mu)\delta(x_0-x),\quad 0<x_0<\tau, \tag{48}$$

with boundary conditions

$$G(x_0,\mu_0\to 0,\mu;\tau)=0,\quad \mu>0,$$
$$G(x_0,\mu_0\to \tau,-\mu;\tau)=0,\quad \mu>0. \tag{49}$$

In order to determine the emergent distributions $G(x_0,\mu_0\to 0,-\mu;\tau)$ and $G(x_0,\mu_0\to\tau,\mu;\tau)$, $\mu>0$, we need the infinite-medium Green's function $G(x_0,\mu_0\to x,\mu;\infty)$ which satisfies Eq. (48). This function can be solved in terms of Case's eigenfunctions[3,6] and is, therefore, considered as known.

Let us seek the solution of our problem in the form

$$G(x_0,\mu_0\to x,\mu;\tau)$$
$$=G(x_0,\mu_0\to x,\mu;\infty)+\psi(x,\mu;\tau). \tag{50}$$

By substituting Eq. (50) into Eqs. (48) and (49), it follows that the unknown function $\psi(x,\mu;\tau)$ must satisfy the homogeneous transport equation (1) and the boundary conditions

$$\psi(0,\mu;\tau)=-G(x_0,\mu_0\to 0,\mu;\infty),\quad \mu>0,$$
$$\psi(\tau,-\mu,\tau)=-G(x_0,\mu_0\to\tau,-\mu;\infty),\quad \mu>0. \tag{51}$$

In this way, the Green's function problem for a slab is reduced to determining the infinite-medium Green's function and to solving two albedo problems discussed in the previous section.

C. Distribution Inside a Slab

Once the surface distributions for a slab problem are known, the inside distribution can be determined by using the full-range completeness and orthogonality relations of Case's eigenfunctions. In view of the results of the previous section, it is sufficient to show how the albedo problem, defined by Eqs. (1) and (9), can be solved completely in terms of the function $S(\tau;\mu_0,\mu)$ or $T(\tau;\mu_0,\mu)$ and Case's eigenfunctions.

We start with the eigenfunction expansion given by Eq. (10). By using the full-range orthogonality relations of Case's eigenfunctions (7a), (7b), and (6), we can determine the expansion coefficients with the help of the function $S(\tau;\mu_0,\mu)$, for instance, by setting $x=0$ in Eq. (10). In this way we get

$$N(\nu)A(\nu)=\mu_0\phi(\nu,\mu_0)-\frac{1}{2}\int_0^1 S(\tau;\mu_0,\mu)\phi(\nu,-\mu)\,d\mu,$$
$$\nu\in(-1,1),\quad \nu=\pm\nu_0. \tag{52}$$

On the other hand, by using the function $T(\tau;\mu_0,\mu)$ and setting $x=\tau$ in Eq. (10), we obtain

$$N(\nu)A(\nu)=\mu_0\phi(\nu,\mu_0)e^{\tau(1/\nu-1/\mu_0)}$$
$$+\tfrac{1}{2}e^{\tau/\nu}\int_0^1 T(\tau;\mu_0,\mu)\phi(\nu,\mu)\,d\mu,$$
$$\nu\in(-1,1),\quad \nu=\pm\nu_0. \tag{53}$$

By using Eqs. (20) and (22), satisfied by $S(\tau;\mu_0,\mu)$ and $T(\tau;\mu_0,\mu)$, it can be easily verified that the rhs of Eqs. (52) and (53) are indeed identical.

IV. HALF-SPACE PROBLEMS

We now briefly discuss half-space problems and show how they can be solved exactly in closed form.

The equations for the half-space problems can be formally obtained from the slab equations of the previous section by limiting τ to infinity and writing

$$\lim S(\tau;\mu_0,\mu)=S(\mu_0,\mu), \tag{54}$$
$$\lim X(\mu)=H(\mu), \tag{55}$$
$$\lim T(\tau;\mu_0,\mu)=0, \tag{56}$$
$$\lim Y(\mu)=0. \tag{57}$$

(We assume, of course, that $c<1$.) The resulting half-space equations are much simpler than the equations of the previous section. In fact, it will be shown that a closed-form solution for $H(\mu)$ can be obtained. Once $H(\mu)$ is known, all other half-space problems can be solved exactly in terms of $H(\mu)$ and Case's eigenfunctions.

To show that, let us consider the explicit form of the singular integral equation for the function $H(\mu)$:

$$\lambda(\nu)H(\nu)=1-\frac{c\nu}{2}P\int_0^1\frac{H(\mu)}{\nu-\mu}\,d\mu, \tag{58}$$

together with the condition

$$0=1-\frac{c\nu_0}{2}\int_0^1\frac{H(\mu)}{\nu_0-\mu}\,d\mu, \tag{59}$$

resulting from Eq. (27) of the previous section by letting τ approach infinity.

We assume that a solution of Eq. (58) exists and that it satisfies a Hölder condition[16] for $\mu\in(0,1)$ and

the condition (59). Guided by the form of the singular integral equation (58), we define an analytic function $F(z)$ in the complex plane cut from -1 to 1 as

$$\Lambda(z)F(z) = 1 + \frac{cz}{2}\int_0^1 \frac{H(\mu)}{\mu - z}\, d\mu. \tag{60}$$

Since $\Lambda(z)$ is analytic in the whole cut plane, with $\Lambda(z) \neq 0$, except for $z = \pm\nu_0$, the same is true also for $F(z)$, in view of our assumption on $H(\mu)$ for $\mu \in (0, 1)$.[16] For $z = \pm\nu_0$, $\Lambda(z)$ has simple zeros, so $F(z)$ may have simple poles there. However, it follows from Eq. (59) that $F(z)$ is analytic also for $z = \nu_0$.

By applying the Plemelj formula[16] to Eq. (60), and taking into account Eqs. (58) and (4), it follows that

$$F^+(x) = F^-(x) = H(x), \quad x \in (0, 1), \tag{61}$$

and

$$F^+(x)\Lambda^+(x) = F^-(x)\Lambda^-(x), \quad x \in (-1, 0). \tag{62}$$

We see from Eq. (61) that $F(z)$ is the analytic continuation of $H(\mu)$, $\mu \in (0, 1)$. Therefore,

$$H(z) = F(z) \tag{63}$$

and $H(z)$ is analytic in whole complex plane, cut from -1 to 0, except for $z = -\nu_0$, where it may have a simple pole.

Let us now consider the product $H(z)H(-z)\Lambda(z)$. This is an even function of z, analytic in the whole complex plane cut from -1 to 1, since $H(z)$ has at most a simple pole for $z = -\nu_0$. Moreover, this product is, in view of Eq. (60), also continuous across the cut $(-1, 1)$, with $H^2(0)\Lambda(0) = 1$, as follows from Eqs. (58) and (3). Hence $H(z)H(-z)\Lambda(z)$ is analytic in the whole complex plane and we have

$$H(z)H(-z)\Lambda(z) = 1. \tag{64}$$

Two important results follow at once from the above relation. First, we see that $H(z)$ has indeed a simple pole for $z = -\nu_0$. Second, by combining Eqs. (60) and (64), we get the nonlinear integral equation for the function $H(z)$:

$$H(z)\left[1 + \frac{cz}{2}\int_0^1 \frac{H(\mu)}{z + \mu}\, d\mu\right] = 1, \quad z \notin (-1, 0). \tag{65}$$

Now, we turn our attention to Eq. (62). We see that $H(z)$ is also the solution of the homogeneous Hilbert problem.[16] By requiring that the solution is analytic in the whole complex plane, cut from -1 to 0, with a simple pole at $z = -\nu_0$, we can solve this problem uniquely in a closed form. We obtain[18]

$$H(z) = \frac{1 + z}{1 + z/\nu_0} \exp\left[\frac{z}{2\pi i}\int_0^1 \ln\frac{\Lambda^+(x)}{\Lambda^-(x)}\frac{dx}{(z + x)x}\right]. \tag{66}$$

[18] S. Pahor, Nucl. Sci. Eng. 26, 192 (1966).

In deriving the above solution we have also justified the assumptions, made in the beginning of this discussion, that a solution of Eq. (58) exists and satisfies a Hölder condition for $\mu \in (0, 1)$.

It is obvious now, from the results of the previous section, how to express the emergent distribution for the albedo problem in terms of the function $H(\mu)$ and how to determine the emergent distribution for the half-space Green's function problem.

However, there is the so-called Milne problem, characteristic for the half-space, which should be mentioned. It turns out that for the half-space the homogeneous transport equation (1) has solutions even for a zero incident distribution, if we drop the condition that solutions are bounded at infinity. We may say that in this case we have sources at infinity.

The Milne problems [whose solution is defined as $\psi(x, \mu; \nu)$] are conveniently defined by the homogeneous transport equation (1) and the boundary conditions

$$\begin{aligned} &\psi(0, \mu; \nu) = 0, \quad \mu > 0, \\ &\psi(x, \mu; \nu) \to \phi(-\nu, \mu)e^{x/\nu}, \\ &\qquad x \to \infty, \quad \nu \in (0, 1), \quad \nu = \nu_0. \end{aligned} \tag{67}$$

We want to determine the emergent distribution $\psi(0, -\mu; \nu)$, $\mu > 0$. To do that, let us define the following "albedo" problem:

$$\begin{aligned} &\psi(0, \mu) = \phi(-\nu, \mu), \quad \mu > 0, \\ &\psi(x, \mu) \to \phi(-\nu, \mu)e^{x/\nu}, \\ &\qquad x \to \infty, \quad \nu \in (0, 1), \quad \nu = \nu_0. \end{aligned} \tag{68}$$

Obviously, the solution of this problem is

$$\psi(x, \mu) = \phi(-\nu, \mu)e^{x/\nu}, \quad \mu \in (-1, 1). \tag{69}$$

Let us decompose the solution $\psi(x, \mu)$ into two parts:

$$\psi(x, \mu) = \psi_1(x, \mu) + \psi_2(x, \mu), \tag{70}$$

where

$$\begin{aligned} &\psi_1(0, \mu) = 0, \quad \mu > 0, \\ &\psi_1(x, \mu) \to \phi(-\nu, \mu)e^{x/\nu}, \quad x \to \infty, \end{aligned} \tag{71}$$

and

$$\begin{aligned} &\psi_2(0, \mu) = \phi(-\nu, \mu), \quad \mu > 0, \\ &\psi_2(x, \mu) \to 0, \quad x \to \infty, \end{aligned} \tag{72}$$

with $\psi_1(x, \mu)$ and $\psi_2(x, \mu)$ satisfying the transport equation (1). Evidently, $\psi_1(x, \mu)$ is just the solution of our Milne problem, while $\psi_2(x, \mu)$ is the solution of a "proper" albedo problem with $\psi_2(\infty, \mu) = 0$.

Therefore, we may apply Eq. (17), with $\tau = \infty$, to $\psi_2(0, \mu)$. In this way we get

$$\psi_2(0, -\mu) = \phi(\nu, \mu) - \psi(0, -\mu; \nu) = \frac{1}{2\mu}\int_0^1 S(\mu', \mu)\psi_2(0, \mu')\, d\mu', \quad \mu > 0, \quad (73)$$

or, in view of Eq. (72),

$$\psi(0, -\mu; \nu) = \phi(\nu, \mu) - \frac{1}{2\mu}\int_0^1 S(\mu', \mu)\phi(-\nu, \mu')\, d\mu', \quad \mu > 0, \quad \nu \in (0, 1), \quad \nu = \nu_0. \quad (74)$$

Now, we express the function $S(\mu', \mu)$ in Eq. (74) in terms in the function $H(\mu)$ by using Eqs. (33) and (39). By taking into account Eqs. (2c), (2d), and (65), it follows that $\psi(0, -\mu; \nu)$ can be expressed in terms of the function $H(z)$ as

$$\psi(0, -\mu; \nu) = \frac{c}{2} P \frac{\nu}{\nu - \mu} \frac{H(\mu)}{H(\nu)} + \lambda(\nu)\delta(\nu - \mu), \quad \mu > 0, \quad \nu \in (0, 1), \quad \nu = \nu_0. \quad (75)$$

Of course, the only physically meaningful solution is that for $\nu = \nu_0$. However, the other solutions are useful for constructing the half-space solutions inside the medium.

Once the surface distribution for any particular half-space problem is known, the complete solution can be obtained by using the full-range orthogonality relations (6), (7a), and (7b).

For instance, let us construct the complete solution of the albedo problem. This solution, denoted as $\psi(0, \mu_0 \to x, \mu)$, satisfies the transport equation (1) and the boundary conditions

$$\psi(0, \mu_0 \to 0, \mu) = \delta(\mu_0 - \mu), \quad \mu > 0, \qquad \psi(0, \mu_0 \to x, \mu) \to 0, \quad x \to \infty. \quad (76)$$

The emergent distribution $\psi(0, \mu_0 \to 0, -\mu)$, $\mu > 0$, can be expressed in terms of the function $S(\mu_0, \mu)$, in view of Eq. (13), as

$$\psi(0, \mu_0 \to 0, -\mu) = (1/2\mu)S(\mu_0, \mu). \quad (77)$$

Because of the condition (76) at infinity, we expand $\psi(0, \mu_0 \to x, \mu)$ only in terms of the exponentially decreasing eigenfunctions

$$\psi(0, \mu_0 \to x, \mu) = A(\nu_0)\phi(\nu_0, \mu)e^{-x/\nu_0} + \int_0^1 A(\nu)\phi(\nu, \mu)e^{-x/\nu}\, d\nu. \quad (78)$$

By setting $x = 0$ and expressing $\psi(0, \mu_0 \to 0, -\mu)$ using Eq. (77), we determine the expansion coefficients as has been explained. Taking into account Eq. (74), we finally get

$$\frac{1}{\mu_0}\psi(0, \mu_0 \to x, \mu) = \frac{\psi(0, -\mu_0; \nu_0)}{N(\nu_0)}\phi(\nu_0, \mu)e^{-x/\nu_0} + \int_0^1 \frac{\psi(0, -\mu_0; \nu)}{N(\nu)}\phi(\nu, \mu)e^{-x/\nu}\, d\nu, \quad (79)$$

and this represents the complete solution of the half-space albedo problem.

V. CONCLUSION

The method presented in this paper is based on the ideas introduced into transport theory by Ambarzumian, Chandrasekhar, and Case. First, the solution for the surface distributions is sought. (In a number of applications this is all that is required.) By using the infinite-medium eigenfunctions, a system of singular integral equations together with the uniqueness conditions is derived for the surface distributions in a simple and straightforward way. This system is the basis of the whole theory.

One could stop there and determine the surface distributions by solving numerically the system of singular integral equations combined with the uniqueness conditions. Or, this system can be reduced to certain uncoupled Fredholm integral equations which can be then used for numerical computations. Finally, the surface distributions can be also computed by using the nonlinear integral equations. It is evident that the question of how to compute the surface distributions is the most important one, since once these are known, the complete solution can be found by quadrature.

As far as we know, the system of singular integral equations (27) and (28) ($\nu = \nu_0$ included) has not been used to compute $X(\mu)$ and $Y(\mu)$. For numerical computations, this system can be rearranged so that the principle-value integrals disappear. Then it could be solved approximately, for instance, by reducing it to a system of linear algebraic equations.

The other possibility, to solve numerically the above mentioned Fredholm integral equations, was considered by Leonard and Mullikin.[17] They showed that these Fredholm integral equations converge rapidly under iteration for all c and τ. Unfortunately, the kernels of these equations are not simple functions and to compute them requires quite a lot of work.

So, it seems that the simplest way to obtain numerical values for the surface distributions is the straightforward iteration of the nonlinear integral equations

(41) and (42). This was done successfully by Chandrasekhar and others.[1,19] Since the system (41) and (42) is not uniquely soluble, the conditions (43) and (44) should be used as a check. At the same time, this would give an estimate of the accuracy of the iterations.

[19] S. Chandrasekhar, D. Elbert, and A. Franklin, Astrophys. J. **115**, 244 (1952).

ACKNOWLEDGMENTS

We wish to thank Professor G. C. Summerfield for stimulating discussions. One of us (S. P.) wishes to express his gratitude to the members of the Department of Nuclear Engineering, The University of Michigan, for hospitality shown him and to the Institute of Science and Technology for financial support.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Realization of Chiral Symmetry in a Curved Isospin Space

KURT MEETZ
Physikalisches Institut der Universität Bonn, Bonn, Germany

(Received 14 September 1968)

The nonlinear realizations of the chiral group $SU(2) \otimes SU(2)$ are studied from a geometric point of view. The three-dimensional nonlinear realization, associated with the pion field, is considered as a group of coordinate transformations in a three-dimensional isospin space of constant curvature, leaving invariant the line element. Spinor realizations in general coordinates are constructed by combined coordinate–spin-space transformations in analogy to Pauli's method for spinors in general relativity. The description of vector mesons and possible chiral-invariant Lagrangians, which yield the various nonlinear models in specific frames of general coordinates, are discussed.

1. INTRODUCTION

Chiral-invariant Lagrangians are currently used as a practical tool to study the implications of current algebra.[1,2] The Lagrangians are to be constructed as functionals of fields, which have definite transformation properties under the chiral group $SU(2) \otimes SU(2)$. Because there does not exist a three-dimensional linear representation of the group, it has been suggested[3,4] that the pion field transforms according to the three-dimensional nonlinear realization. This implies that chiral symmetry is a pure interaction symmetry not shared by the asymptotic fields.

A systematic development of the nonlinear realizations can depart from different points of view. While the transformation laws are nonlinear, the transformations are still implemented by unitary operators in quantum theory. Weinberg has studied the most general form for the commutators of generators and fields.[5] On the other hand, for a better understanding of the mathematical nature of nonlinear realizations, it seems worthwhile to keep the analogy to linear representations as close as possible.

A nonlinear realization is a representation of the group in a curved instead of Euclidean space. We show in Sec. 2 that the chiral group is the invariance group of the metric in a three-dimensional space with constant curvature $K = 1/f^2$. This "fundamental" nonlinear realization is associated with the pion field. While the field components are the coordinates in the curved space, space–time derivatives are tangents and transform as contravariant vectors under coordinate transformations. The Riemannian geometry of the curved space replaces the Euclidean geometry of linear-representation spaces. Following Pauli's treatment of spinors in general relativity,[6] we study in Sec. 3 spinor realizations of the chiral group in general coordinates by combined coordinate–spin space transformations. The realizations associated with vector mesons are discussed in Sec. 4.

The various nonlinear models treated in the literature[7–12] result from a specific choice of general pion coordinates. This is in complete agreement with Weinberg,[5] but we think that our more geometric

[1] S. Weinberg, Phys. Rev. Letters **18**, 188 (1967).
[2] B. W. Lee and H. T. Nieh, Phys. Rev. **166**, 1507 (1968).
[3] J. Schwinger, Phys. Letters **24B**, 473 (1967).
[4] J. Wess and B. Zumino, Phys. Rev. **163**, 1727 (1967).
[5] S. Weinberg, Phys. Rev. **166**, 1568 (1968).
[6] W. Pauli, Ann. Phys. (Leipzig) **18**, 337 (1933).
[7] G. Kramer, H. Rollnik, and B. Stech, Z. Physik **154**, 564 (1959).
[8] M. Gell-Mann and M. Lévy, Nuovo Cimento **16**, 705 (1960).
[9] F. Gürsey, Nuovo Cimento **16**, 230 (1960).
[10] F. Gürsey, Ann. Phys. (N.Y.) **12**, 91 (1961).
[11] P. Chang and F. Gürsey, Phys. Rev. **164**, 1752 (1967).
[12] L. S. Brown, Phys. Rev. **163**, 1802 (1967).

(41) and (42). This was done successfully by Chandrasekhar and others.[1,19] Since the system (41) and (42) is not uniquely soluble, the conditions (43) and (44) should be used as a check. At the same time, this would give an estimate of the accuracy of the iterations.

[19] S. Chandrasekhar, D. Elbert, and A. Franklin, Astrophys. J. **115**, 244 (1952).

ACKNOWLEDGMENTS

We wish to thank Professor G. C. Summerfield for stimulating discussions. One of us (S. P.) wishes to express his gratitude to the members of the Department of Nuclear Engineering, The University of Michigan, for hospitality shown him and to the Institute of Science and Technology for financial support.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Realization of Chiral Symmetry in a Curved Isospin Space

KURT MEETZ
Physikalisches Institut der Universität Bonn, Bonn, Germany

(Received 14 September 1968)

The nonlinear realizations of the chiral group $SU(2) \otimes SU(2)$ are studied from a geometric point of view. The three-dimensional nonlinear realization, associated with the pion field, is considered as a group of coordinate transformations in a three-dimensional isospin space of constant curvature, leaving invariant the line element. Spinor realizations in general coordinates are constructed by combined coordinate–spin-space transformations in analogy to Pauli's method for spinors in general relativity. The description of vector mesons and possible chiral-invariant Lagrangians, which yield the various nonlinear models in specific frames of general coordinates, are discussed.

1. INTRODUCTION

Chiral-invariant Lagrangians are currently used as a practical tool to study the implications of current algebra.[1,2] The Lagrangians are to be constructed as functionals of fields, which have definite transformation properties under the chiral group $SU(2) \otimes SU(2)$. Because there does not exist a three-dimensional linear representation of the group, it has been suggested[3,4] that the pion field transforms according to the three-dimensional nonlinear realization. This implies that chiral symmetry is a pure interaction symmetry not shared by the asymptotic fields.

A systematic development of the nonlinear realizations can depart from different points of view. While the transformation laws are nonlinear, the transformations are still implemented by unitary operators in quantum theory. Weinberg has studied the most general form for the commutators of generators and fields.[5] On the other hand, for a better understanding of the mathematical nature of nonlinear realizations, it seems worthwhile to keep the analogy to linear representations as close as possible.

A nonlinear realization is a representation of the group in a curved instead of Euclidean space. We show in Sec. 2 that the chiral group is the invariance group of the metric in a three-dimensional space with constant curvature $K = 1/f^2$. This "fundamental" nonlinear realization is associated with the pion field. While the field components are the coordinates in the curved space, space–time derivatives are tangents and transform as contravariant vectors under coordinate transformations. The Riemannian geometry of the curved space replaces the Euclidean geometry of linear-representation spaces. Following Pauli's treatment of spinors in general relativity,[6] we study in Sec. 3 spinor realizations of the chiral group in general coordinates by combined coordinate–spin space transformations. The realizations associated with vector mesons are discussed in Sec. 4.

The various nonlinear models treated in the literature[7–12] result from a specific choice of general pion coordinates. This is in complete agreement with Weinberg,[5] but we think that our more geometric

[1] S. Weinberg, Phys. Rev. Letters **18**, 188 (1967).
[2] B. W. Lee and H. T. Nieh, Phys. Rev. **166**, 1507 (1968).
[3] J. Schwinger, Phys. Letters **24B**, 473 (1967).
[4] J. Wess and B. Zumino, Phys. Rev. **163**, 1727 (1967).
[5] S. Weinberg, Phys. Rev. **166**, 1568 (1968).
[6] W. Pauli, Ann. Phys. (Leipzig) **18**, 337 (1933).
[7] G. Kramer, H. Rollnik, and B. Stech, Z. Physik **154**, 564 (1959).
[8] M. Gell-Mann and M. Lévy, Nuovo Cimento **16**, 705 (1960).
[9] F. Gürsey, Nuovo Cimento **16**, 230 (1960).
[10] F. Gürsey, Ann. Phys. (N.Y.) **12**, 91 (1961).
[11] P. Chang and F. Gürsey, Phys. Rev. **164**, 1752 (1967).
[12] L. S. Brown, Phys. Rev. **163**, 1802 (1967).

point of view might help towards a better understanding of nonlinear realizations and open the way for the study of other symmetry groups.

2. PIONS

We start with the σ model,[8,13] i.e., a real four-component field $p^\alpha(x)$ ($\alpha = 0, 1, 2, 3$) in the $(\frac{1}{2}, \frac{1}{2})$ representation of $SU(2) \otimes SU(2)$. This means that at every space–time point x the components $p^\alpha(x)$ span a four-dimensional Euclidean space E_4 and have the transformation law

$$p'^\alpha = R^\alpha_\beta p^\beta, \tag{2.1}$$

where R is an orthogonal 4×4 matrix such that

$$RR^{\mathrm{T}} = 1; \quad |R| = 1. \tag{2.2}$$

Because

$$p^\alpha p_\alpha = f^2 = (p^0)^2 + (\mathbf{p})^2 \tag{2.3}$$

is invariant under (2.1), the three-dimensional surface of the upper or lower half-sphere of radius f,

$$p^0 = \pm(f^2 - \boldsymbol{\pi}^2)^{\frac{1}{2}}, \qquad \mathbf{p} = \boldsymbol{\pi}, \tag{2.4}$$

is mapped onto itself by (2.1). Equation (2.4) gives a special parameterization of a three-dimensional curved space V_3 with constant curvature $K = 1/f^2$. We take the parameters π^i ($i = 1, 2, 3$), i.e., the coordinates in V_3, as the three components of the pion field. This is a generalization of the usual setup, where the field components are coordinates in an Euclidean space. The group $SU(2) \otimes SU(2)$ is represented in V_3 by the coordinate transformations

$$\pi'^i = R^i_l \pi^l \pm R^i_0 (f^2 - \boldsymbol{\pi}^2)^{\frac{1}{2}}. \tag{2.5}$$

The representation of the isospin subgroup ($R^i_0 = 0$) is still linear. The inner geometry of V_3 is concerned with the properties which are independent of the embedding into the space E_4. Its fundamental quantity is the metric tensor

$$g_{ik}(\pi) = \frac{\partial p^\alpha}{\partial \pi^i} \frac{\partial p_\alpha}{\partial \pi^k} = \delta_{ik} + \frac{\pi_i \pi_k}{f^2 - \boldsymbol{\pi}^2}, \tag{2.6}$$

if we use the parameterization (2.4). g_{ik} is a covariant tensor with the general transformation law

$$g'_{ik}(\pi') = g_{rs}(\pi) \frac{\partial \pi^r}{\partial \pi'^i} \frac{\partial \pi^s}{\partial \pi'^k} \tag{2.7}$$

under coordinate transformations

$$\pi'^i = \pi'^i(\pi). \tag{2.8}$$

The contravariant tensor g^{ik} is determined by

$$g^{ir} g_{rk} = g^i_k = \delta^i_k. \tag{2.9}$$

[13] J. Schwinger, Ann. Phys. (N.Y.) 2, 407 (1957).

It can easily be checked that the transformations (2.5) leave the metric invariant

$$g'_{ik}(\pi') = g_{ik}(\pi'), \tag{2.10}$$

i.e., $SU(2) \otimes SU(2)$ is the invariance group of the quadratic form

$$g_{ik}(\pi)\, d\pi^i\, d\pi^k = ds^2 \tag{2.11}$$

in V_3.

Each component of the space–time derivative $\partial_\mu \pi^i$ ($\mu = 0, 1, 2, 3$) is a contravariant vector with respect to coordinate transformations in V_3 with the transformation law

$$\partial_\mu \pi'^i = \frac{\partial \pi'^i}{\partial \pi^k} \partial_\mu \pi^k. \tag{2.12}$$

The corresponding contravariant vector is defined by

$$\partial_\mu \pi_i = g_{ik} \partial_\mu \pi^k. \tag{2.13}$$

Hence, the only choice for an $[SU(2) \otimes SU(2)]$-invariant Lagrangian of second order in $\partial_\mu \pi^i$ is

$$L_{\pi\pi} = \tfrac{1}{2} g_{ik}(\pi) \partial_\mu \pi^i \partial^\mu \pi^k = \tfrac{1}{2} \partial_\mu \pi_i \partial^\mu \pi^i. \tag{2.14}$$

The nonlinear models treated in the literature[7–12] correspond to different systems of coordinates in V_3. Weinberg[5] calls this a redefinition of the pion field. With (2.6) we obtain

$$L_{\pi\pi} = \tfrac{1}{2} \partial_\mu \boldsymbol{\pi} \partial^\mu \boldsymbol{\pi} + \frac{1}{2} \frac{(\boldsymbol{\pi} \cdot \partial_\mu \boldsymbol{\pi})(\boldsymbol{\pi} \cdot \partial^\mu \boldsymbol{\pi})}{f^2 - \boldsymbol{\pi}^2}, \tag{2.15}$$

where $\boldsymbol{\pi}$ is the isovector π^i. A transformation to stereographic coordinates $\bar{\pi}^i$,

$$\pi^i = \frac{\bar{\pi}^i}{1 + \bar{\boldsymbol{\pi}}^2/4f^2} \tag{2.16}$$

leads to

$$\bar{g}_{ik}(\bar{\pi}) = \frac{\delta_{ik}}{1 + \bar{\boldsymbol{\pi}}^2/4f^2} \tag{2.17}$$

and

$$L_{\pi\pi} = \frac{1}{2} \frac{\partial_\mu \bar{\boldsymbol{\pi}} \cdot \partial^\mu \bar{\boldsymbol{\pi}}}{1 + \bar{\boldsymbol{\pi}}^2/4f^2}, \tag{2.18}$$

which is another nonlinear model. The Lagrangians (2.15) and (2.18) are form-invariant under $SU(2) \otimes SU(2)$ because of (2.10).

The covariant field equations of (2.14) read

$$\pi^i + \begin{Bmatrix} i \\ mn \end{Bmatrix} \partial_\mu \pi^m \partial^\mu \pi^n = 0, \tag{2.19}$$

where $\{^i_{mn}\}$ is a Christoffel symbol of the second kind. The field equation (2.19) is a direct generalization of the flat isospin-space equation

$$\Box \pi^i = 0. \tag{2.20}$$

While $\partial_\mu \pi^i$ is a contravariant vector, a second space-time differentiation must be performed in an invariant fashion. The invariant Bianchi derivative of a vector a^i is defined by

$$D_\mu a^i = a^i_{\|l} \partial_\mu \pi^l = \partial_\mu a^i + \begin{Bmatrix} i \\ kl \end{Bmatrix} a^k \partial_\mu \pi^l, \quad (2.21)$$

where

$$a^i_{\|l} = \frac{\partial a^i}{\partial \pi^l} + \begin{Bmatrix} i \\ kl \end{Bmatrix} a^k \quad (2.22)$$

is the usual covariant differentiation. If a^i depends explicitly on x,

$$a^i = a^i(\boldsymbol{\pi}(x), x), \quad (2.23)$$

we write

$$D_\mu a^i = \partial'_\mu a^i + a^i_{\|l} \partial_\mu \pi^l, \quad (2.24)$$

where ∂'_μ acts only on the explicitly appearing variables x. Then $D_\mu a^i$ transforms as a contravariant vector. Hence, we may write instead of (2.19):

$$D_\mu \partial^\mu \pi^i = D_\mu D^\mu \pi^i = 0, \quad (2.25)$$

if we define

$$D^\mu f = \partial^\mu f \quad (2.26)$$

for a scalar f.

3. NUCLEONS

Our next task is to induce a representation of $SU(2) \otimes SU(2)$ in the two-dimensional spin space of the $I = \frac{1}{2}$ representation of the isotopic $SU(2)$. To do so, we generalize a method developed by Pauli[6] for spinors in general relativity.

The two-dimensional spin space is related to the three-dimensional Euclidean space E_3 of the $I = 1$ representation by the Clifford algebra

$$E_3: \quad [\tau_i, \tau_k]_+ = 2\delta_{ik}. \quad (3.1)$$

Instead of (3.1), we consider the Clifford algebra of V_3,

$$V_3: \quad [\Gamma^i(\pi), \Gamma^k(\pi)]_+ = 2g^{ik}(\pi), \quad (3.2)$$

where the $\Gamma^i(\pi)$ are 2×2 spin matrices depending on the pion field $\boldsymbol{\pi}$. We define the transformation law of the Γ^i under coordinate transformations in V_3 by

$$\Gamma'^i(\pi') = \frac{\partial \pi'^i}{\partial \pi^k} \Gamma^k(\pi). \quad (3.3)$$

Because the $SU(2) \otimes SU(2)$ transformations (2.5) leave the metric tensor unchanged, there must exist a nonsingular matrix S for every transformation (2.5) so that

$$\Gamma'^i(\pi') = S^{-1} \Gamma^i(\pi') S. \quad (3.4)$$

If we define the combined coordinate–spin space transformation laws of $SU(2) \otimes SU(2)$ spinors and the Γ^i matrices by

$$\psi'(\pi') = S\psi(\pi), \quad \psi'^+(\pi') = \psi^+(\pi) S^{-1},$$
$$\Gamma'^i(\pi') = S\Gamma'^i(\pi') S^{-1} = \Gamma^i(\pi'), \quad (3.5)$$

bilinear quantities constructed with the Γ^i transform as ordinary tensors, e.g.,

$$\psi'^+(\pi')\Gamma^i(\pi')\psi'(\pi') = \psi^+(\pi)\Gamma^k(\pi)\psi(\pi) \frac{\partial \pi'^i}{\partial \pi^k}. \quad (3.6)$$

A representation of the Clifford algebra (3.2) can be obtained in the form

$$\Gamma_i(\pi) = h_{im}(\pi)\tau_m, \quad (3.7)$$

with

$$h_{im}(\pi) h_{km}(\pi) = g_{ik}(\pi). \quad (3.8)$$

It must be emphasized that only the first index of h_{im} is covariant under coordinate transformations. The conditions (3.8) do determine the coefficients h_{im} only up to an orthogonal transformation acting on the second index. We take the solution corresponding to the expansion

$$g_{ik} = \delta_{ik} + \varphi_{ik},$$
$$h_{im} = (g_{im})^{\frac{1}{2}} = (\delta_{im} + \tfrac{1}{2}\varphi_m + \cdots). \quad (3.9)$$

In case of (2.6) we have

$$h_{im} = \delta_{im} + \frac{f - (f^2 - \boldsymbol{\pi}^2)^{\frac{1}{2}}}{\boldsymbol{\pi}^2 (f^2 - \boldsymbol{\pi}^2)^{\frac{1}{2}}} \pi_i \pi_m. \quad (3.10)$$

We now turn to the representation of infinitesimal $SU(2) \otimes SU(2)$ transformations in the spin space. The isotopic subgroup need not be discussed. It has the usual representation. For infinitesimal chiral transformations we may write, according to (2.5),

$$\pi'^i = \pi^i + \delta\Lambda^i(\pi). \quad (3.11)$$

If we put

$$S = 1 + T, \quad (3.12)$$

we obtain, from (3.3) and (3.4),

$$-[T, \Gamma^i(\pi)] = \Gamma^k \frac{\partial \delta\Lambda^i}{\partial \pi^k} - \frac{\partial \Gamma^i}{\partial \pi^k} \delta\Lambda^k. \quad (3.13)$$

By differentiation of (3.2) we find that the derivatives of Γ^i have the structure

$$\frac{\partial \Gamma^i}{\partial \pi^l} = \tfrac{1}{2}\Gamma_r \frac{\partial g^{ri}}{\partial \pi^l} + [\Delta_l, \Gamma^i], \quad (3.14)$$

where the spin-space operator Δ_l is determined up to the product of the identity and a gradient function, if we add the integrability conditions

$$\frac{\partial^2 \Gamma^i}{\partial \pi^l \partial \pi^m} = \frac{\partial^2 \Gamma^i}{\partial \pi^m \partial \pi^l}. \quad (3.15)$$

We find

$$\Delta_l = -\frac{1}{8}\left[\Gamma_j, \frac{\partial \Gamma^j}{\partial \pi^l}\right]. \tag{3.16}$$

The solution of (3.13) is

$$T = \Delta_l \delta\Lambda^l + \tfrac{1}{8}[\Gamma_l, \Gamma^m]\frac{\partial \delta\Lambda^l}{\partial \pi^m}, \tag{3.17}$$

a generalization of the well-known spinor-transformation law for linear transformations. If we take

$$\delta\Lambda^k = \epsilon^k (f^2 - \boldsymbol{\pi}^2)^{\frac{1}{2}} \tag{3.18}$$

and use (2.6), we arrive at Weinberg's transformation law[5] for our particular choice of coordinates in V_3:

$$T = \epsilon^k i\epsilon_{krn}\frac{\tau_r}{2}\pi_n \frac{1}{f + (f^2 - \boldsymbol{\pi}^2)^{\frac{1}{2}}}. \tag{3.19}$$

In order to introduce a covariant derivation of spinors, we first consider the usual covariant derivative of Γ^i:

$$\Gamma^i_{\|l} = \frac{\partial \Gamma^i}{\partial \pi^l} + \begin{Bmatrix} i \\ lk \end{Bmatrix}\Gamma^k. \tag{3.20}$$

Because

$$g^{ik}_{\|l} = 0, \tag{3.21}$$

we conclude by comparison with (3.14) that there must exist an isospin matrix Ω_l so that

$$\Gamma^i_{\|l} = [\Omega_l, \Gamma^i]. \tag{3.22}$$

It is given by

$$\Omega_l = \Delta_l + \tfrac{1}{8}[\Gamma^m, \Gamma^n]\frac{\partial g_{ln}}{\partial \pi^m} = -\tfrac{1}{8}[\Gamma_j, \Gamma^j_{\|l}]. \tag{3.23}$$

With Pauli,[6] we define a covariant spinor differentiation (denoted by "$.l$") by the following conditions:

$$\begin{aligned} &\text{(a)}\ \ \Gamma^i_{.l} = \Gamma^i_{\|l} + [\Gamma^i, \Omega_l] = 0, \\ &\text{(b)}\ \ (\psi^+\Gamma^i\psi)_{.l} = \psi^+_{.l}\Gamma^i\psi + \psi^+\Gamma^i\psi_{.l}, \\ &\text{(c)}\ \ (\psi^+\Gamma^i\psi)_{.l} = (\psi^+\Gamma^i\psi)_{\|l}. \end{aligned} \tag{3.24}$$

These conditions lead to

$$\begin{aligned} \psi_{.l} &= \frac{\partial \psi}{\partial \pi^l} - \Omega_l\psi, \\ \psi^+_{.l} &= \frac{\partial \psi^+}{\partial \pi^l} + \psi^+\Omega_l. \end{aligned} \tag{3.25}$$

It is then clear that an invariant space–time differentiation of a spinor has to be defined by

$$\begin{aligned} D'_\mu\psi &= \partial_\mu\psi - \partial_\mu\pi^l\Omega_l\psi, \\ D'_\mu\psi^+ &= \partial_\mu\psi^+ + \psi^+\Omega_l\partial_\mu\pi^l, \end{aligned} \tag{3.26}$$

where

$$\partial_\mu\psi = \partial_\mu\pi^l\frac{\partial\psi}{\partial\pi^l} + \partial'_\mu\psi, \tag{3.27}$$

if ψ depends explicitly on x.

The Lagrangian

$$L_{\psi\psi} = -\bar\psi i\gamma_\mu D'^\mu\psi + m\bar\psi\psi, \tag{3.28}$$

where ψ is a four-component space–time spinor and a two-component isospinor, leads to the invariant field equation

$$(-i\gamma_\mu D'^\mu + m)\psi = 0. \tag{3.29}$$

If we use the coordinate system (2.6), we find

$$\Omega_l = i\epsilon_{lsv}\pi_s\tau_v\frac{1}{2f}\frac{1}{f + (f^2 - \boldsymbol{\pi}^2)^{\frac{1}{2}}}, \tag{3.30}$$

$$D'_\mu\psi = \partial_\mu\psi - i\boldsymbol{\tau}\cdot(\partial_\mu\boldsymbol{\pi}\times\boldsymbol{\pi})\frac{1}{2f}\frac{1}{f + (f^2 - \boldsymbol{\pi}^2)^{\frac{1}{2}}}\psi, \tag{3.31}$$

which coincides with Weinberg's "covariant derivative"[5] and corresponds to Gürsey's second nonlinear model.[11] Equations (2.25) and (3.29) are free-field equations, covariant under $SU(2)\otimes SU(2)$ in an isospin space of constant curvature $K = 1/f^2$. The π–π interaction and the π–N interaction embedded in (2.25) and (3.29), respectively, is induced and completely determined by the curvature. In the limit $f\to\infty$, the space becomes flat, the interactions vanish, and the symmetry group $SU(2)\otimes SU(2)$ is contracted to isotopic $SU(2)$.

The most simple form for interacting pion and nucleon fields is the sum of (2.14) and (3.28):

$$L = -\bar\psi(-i\gamma\cdot D' + m)\psi + \tfrac{1}{2}\partial_\mu\pi_i\partial^\mu\pi^i. \tag{3.32}$$

While the field equations for the nucleon field remain unchanged,

$$(-i\gamma\cdot D' + m)\psi = 0, \tag{3.33}$$

there appears an additional term in the pion field equation,

$$\Box\pi^i + \begin{Bmatrix} i \\ mn \end{Bmatrix}\partial_\mu\pi^m\partial^\mu\pi^n = \tfrac{1}{8}R^i{}_{lmn}\bar\psi i\gamma_\mu[\Gamma^m, \Gamma^n]\psi\partial^\mu\pi^l, \tag{3.34}$$

where

$$R^i{}_{lmn} = g^{ir}(\pi)R_{rlmn}(\pi) \tag{3.35}$$

is the Riemannian curvature tensor depending only on the pion field.

All interaction terms of (3.32) are due to the curvature of the pion-isospin space and vanish in the flat-space limit $f\to\infty$. But we may add further terms that do not have this property. A chiral-invariant "trilinear" interaction term must be of the form

$$L_I = g\bar\psi\gamma_\mu\gamma_5\Gamma^i(\pi)\psi\partial^\mu\pi_i. \tag{3.36}$$

It reduces to the gradient coupling term

$$L_I = g\bar\psi\gamma_\mu\gamma_5\boldsymbol{\tau}\psi\cdot\partial^\mu\boldsymbol{\pi} \tag{3.37}$$

in the flat-space limit.

4. ρ MESONS

The ρ meson is usually introduced instead of the direct vector interaction of pion pairs in the covariant derivative of the ψ field

$$D'_\mu\psi = \partial_\mu\psi - ig_0\boldsymbol{\rho}_\mu \cdot (\boldsymbol{\tau}/2)\psi. \tag{4.1}$$

To guarantee the covariance of (4.1) under spin-space transformations

$$D'_\mu S\psi = SD'_\mu\psi, \tag{4.2}$$

$\boldsymbol{\rho}_\mu$ has to obey a particular transformation law that can be deduced from the transformation law of the interaction term in (3.26):

$$D'_\mu\psi = \partial_\mu\psi - \partial_\mu\pi^l\Omega_l(\pi)\psi. \tag{4.3}$$

The spin-space operator Ω_l is defined by (3.22). Because the matrices $\Gamma^i(\pi)$ have the same functional dependance in all coordinate systems related by chiral transformations [see Eq. (3.5)]:

$$\Gamma'^i(\pi') = \Gamma^i(\pi'), \tag{4.4}$$

Ω_l in the primed system is given by

$$\Gamma^i(\pi')_{\|l} = [\Omega_l(\pi'), \Gamma^i(\pi')]. \tag{4.5}$$

Hence,

$$D'_\mu\psi' = \partial_\mu\psi' - \partial_\mu\pi'^l\Omega_l(\pi')\psi'. \tag{4.6}$$

The interaction term in (4.3) may be written as

$$\partial_\mu\pi^l\Omega_l(\pi) = \tfrac{1}{2}(D_\mu h^j_m(\pi))h_{jn}(\pi)\tfrac{1}{4}[\tau_m, \tau_n]. \tag{4.7}$$

Because of

$$h^j_m h_{jn} = \delta_{mn}, \tag{4.8}$$

the quantity

$$(D_\mu h^j_m(\pi))h_{jn}(\pi) \to g_0\rho_{\mu,mn}(\pi) \tag{4.9}$$

is a skew symmetric tensor that may be replaced by the ρ-meson field in the spirit of (4.1). g_0 is some coupling constant. We have already remarked earlier that the indices m, n are, in contrast to j, not covariant with respect to general coordinate transformations in V_3. Let us now discuss the transformation properties of the ρ field. According to (4.6) we have in a primed system

$$(D_\mu h^j_m(\pi'))h_{jn}(\pi') \to g_0\rho_{\mu,mn}(\pi'). \tag{4.10}$$

Referring to (2.10) and (3.8), we see that

$$h_{ir}(\pi')h_{kr}(\pi') = g_{ik}(\pi') = h'_{ir}(\pi')h'_{kr}(\pi') = g'_{ik}(\pi'). \tag{4.11}$$

Hence, there must exist an orthogonal matrix S such that

$$h_{im}(\pi') = S_{mr}h'_{ir}(\pi'). \tag{4.12}$$

For infinitesimal transformations it is found to be

$$S = 1 + T, \quad T = T_l t_l, \quad (t_l)_{mn} = i\epsilon_{mln}. \tag{4.13}$$

This should be compared with the spin-space transformation matrix in the spinor case (3.12) that can be written as

$$T = T_l(\tau_l/2). \tag{4.14}$$

The transformation law for the ρ-meson field now follows from (4.10) and (4.12). Using an obvious matrix notation we may write

$$g_0\rho_\mu(\pi') = Sg_0\rho_\mu(\pi)S^{-1} + (\partial_\mu S)S^{-1}. \tag{4.15}$$

Hence, the ρ field does not transform as a tensor under spin-space transformations nor does the curl of the ρ field. But the quantity

$$F_{\mu\nu} = \partial_\nu\rho_\mu - \partial_\mu\rho_\nu + g_0(\rho_\mu\rho_\nu - \rho_\nu\rho_\mu) \tag{4.16}$$

does transform as a tensor

$$F_{\mu\nu}(\pi') = SF_{\mu\nu}(\pi)S^{-1}. \tag{4.17}$$

This leads to the invariant kinetic Lagrangian

$$L_{\rho\rho} = \tfrac{1}{4}(F_{\mu\nu}F^{\mu\nu}), \tag{4.18}$$

used by Schwinger[3] and Weinberg.[5]

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Solution of the Bethe-Salpeter Equation with a Square-Well Potential

R. F. KEAM
Department of Physics, University of Auckland, Auckland, New Zealand

(Received 20 August 1968)

The ladder approximation Bethe–Salpeter equation for (i) a bound spin-0 boson–boson system of zero total mass and (ii) a bound spin-$\frac{1}{2}$ fermion–antifermion system of zero total mass is solved for a four-dimensional square-well potential.

1. INTRODUCTION

The only known exact solutions of the ladder-approximation Bethe–Salpeter equation[1,2] for a bound spin-$\frac{1}{2}$ fermion–antifermion system, expressible in terms of rational or commonly occurring higher transcendental functions, are for zero total mass and with the mass of the exchange particle also zero.[3–5] Attention is therefore becoming directed to numerical approximation methods and interesting results have already been obtained.[6,7] So far, field-theoretical potentials have been used, and these have a singularity[2,8,9] serious enough to modify considerably the form of the wavefunctions at the origin of the Euclidean relative-coordinate space in which the problem is described after the Wick rotation[10] has been performed. One consequence is the familiar "Goldstein problem" and it appears that in order to overcome it one must in some fashion cut off the high momenta.[11,12] It has seemed worthwhile, therefore, to investigate a model nonsingular potential since, as Mandelstam has shown,[9] this gives rise to a discrete set of coupling constants and problems at the origin of relative-coordinate space are avoided. We choose the four-dimensional square-well potential as the simplest to deal with analytically. As well, however, this choice introduces a high momentum cutoff in a very real way; we show in Sec. 3 that in place of the single-particle exchange k^{-2} asymptotic momentum dependence, the square well gives $k^{-2}J_2(kA)$ so that the k^{-2} dependence is modulated by an oscillating factor of magnitude $k^{-\frac{1}{2}}$. We perform the analysis for zero total mass since, in this case, exact analytical solutions are found for all Dirac-space sectors for all types of coupling and for all angular momenta, in both the internal and the external regions, and the numerical work is reduced to matching the values and gradients of the internal and external forms for each variable at the well boundary. We arrive at relationships between well width and well depth in order that the system described has zero total mass. It is hoped that the solutions form the starting point for a perturbation treatment of bound systems with small but nonvanishing total mass E.

The notation and conventions in the following sections are those of a previous paper[13] hereafter referred to as I. In Sec. 2, the parity and charge-parity properties of the zero-mass solutions in I are derived. In Sec. 3 we consider a boson–boson system with a square well and compare the results with those for a potential due to particle exchange. Sections 4 and 5 detail properties of the square well for a fermion–antifermion system and the analytical solutions of the corresponding Bethe–Salpeter equation. Finally in Sec. 6 we consider certain numerical solutions.

2. PARITY AND CHARGE CONJUGATION PROPERTIES

We wish to identify the operators $\bar{\mathcal{P}}$, $\bar{\mathcal{C}}$ which, acting on the wavefunction $\chi(x_1, x_2)$, are equivalent, respectively, to the Hilbert-space parity operator $\bar{P}$ and charge conjugation operator $\bar{C}$ acting on the state vector $|B\rangle$; i.e.,

$$\begin{aligned}\bar{\mathcal{P}}\chi &= \langle 0|\, T\{\psi(x_1)\bar{\psi}(x_2)\}\bar{P}\,|B\rangle,\\ \bar{\mathcal{C}}\chi &= \langle 0|\, T\{\psi(x_1)\bar{\psi}(\psi_2)\}\bar{C}\,|B\rangle.\end{aligned} \tag{1}$$

Using the well-known commutation properties of $\bar{C}$, $\bar{P}$ with ψ, $\bar{\psi}$, we find, indeed,[14] that

$$\bar{\mathcal{P}}\chi = \hat{\gamma}_4\chi(x_1', x_2'), \tag{2}$$

$$\bar{\mathcal{C}}\chi = C\chi_{\mathrm{T}}(x_2, x_1)C^{-1}, \tag{3}$$

where $x_1' = (-\mathbf{x}_1, (x_1)_4)$, $x_2' = (-\mathbf{x}_2, (x_2)_4)$, a subscript T stands for "transpose," and C is a 4×4

[1] H. A. Bethe and E. E. Salpeter, Phys. Rev. **84**, 1232 (1951).
[2] M. Gell-Mann and F. E. Low, Phys. Rev. **84**, 350 (1951).
[3] J. S. Goldstein, Phys. Rev. **91**, 1516 (1953).
[4] A. Bastai, L. Bertocchi, G. Furlan, and M. Tonin, Nuovo Cimento **30**, 1532 (1963).
[5] W. Kummer, Nuovo Cimento **31**, 219 (1964).
[6] P. Narayanaswamy and A. Pagnamenta, Nuovo Cimento **53A**, 635 (1968).
[7] C. H. Llewellyn Smith [submitted to Ann. Phys. (N.Y.)].
[8] A. Bastai, L. Bertocchi, S. Fubini, G. Furlan, and M. Tonin, Nuovo Cimento **30**, 1512 (1963).
[9] S. Mandelstam, Proc. Roy. Soc. (London) **A233**, 248 (1955).
[10] G. C. Wick, Phys. Rev. **96**, 1124 (1954).
[11] Reference 6, p. 652.
[12] An alternative possibility has been considered by the present author (submitted to J. Math. Phys.).
[13] R. F. Keam, J. Math. Phys. **9**, 1462 (1968).
[14] Confer J. Harte, Nuovo Cimento **45A**, 179 (1966).

matrix satisfying

$$C^\dagger = C^{-1}, \quad C_{\mathrm{T}} = -C, \qquad C^{-1}\gamma_\mu C = -\gamma_{\mu\mathrm{T}}. \tag{4}$$

In terms of the wavefunction in the relative coordinate $f(x)$ [cf. Eq. (I.6)] these become

$$\bar{\mathcal{P}} f(\mathbf{x}, x_4) = \hat{\gamma}_4 f(-\mathbf{x}, x_4), \tag{5}$$

$$\bar{\mathcal{C}} f(\mathbf{x}, x_4) = C f_{\mathrm{T}}(-\mathbf{x}, -x_4) C^{-1}. \tag{6}$$

From (6) we can deduce that

$$\bar{\mathcal{C}} = (-1)^{\frac{1}{4}(\mathcal{R}+3)}\mathcal{C}, \tag{7}$$

where the operators $\mathcal{R}$, $\mathcal{C}$ are as defined in I. $\bar{\mathcal{P}}$ commutes with $\mathcal{B}_E$, $\mathcal{U}$, $\mathcal{R}$, $\mathcal{C}$; but

$$\bar{\mathcal{P}} J_i^\pm = J_i^\mp \bar{\mathcal{P}} \tag{8}$$

and there are similar relations with $J_i^\pm$ replaced by $L_i^\pm$ or $S_i^\pm$.

In Eq. (I.54), namely

$$f_{j^+j^-Jm} = \sum_{m^+m^-} (j^+j^-m^+m^- \mid Jm) f_{j^+j^-m^+m^-}, \tag{9}$$

we wrote down the form of a general solution of the ladder approximation Wick-rotated Bethe–Salpeter equation for a massless bound system of a spin-$\frac{1}{2}$ fermion and spin-$\frac{1}{2}$ antifermion of equal mass, where the suffices in the left member are the quantum numbers corresponding respectively to the operators $(J^+)^2$, $(J^-)^2$, J^2, J_z.

In the S–V sector (where, say, $j^+ = j^- = j$) the explicit form of f^{SV}_{jjJm} is found by substituting into the right member of (9) from Eq. (I.30).

In the T–A sector, Case A gives the disjoint solutions

$$\begin{aligned} f^{TA-}_{jjJm} = \sum_{m^+m^-} (jjm^+m^- \mid Jm) \Big[& 2^{-\frac{1}{2}}(2j+1)^{-\frac{1}{2}} w(R) \\ & \times \Big\{ \sum_{m_s^+} (j1m_l^+m_s^+ \mid jm^+) Z_{jm_l^+m^-} \Sigma^+_{m_s^+} \\ & + \sum_{m_s^-} (j1m_l^-m_s^- \mid jm^-) Z_{jm^+m_l^-} \Sigma^-_{m_s^-} \Big\} \\ & + j^{-\frac{1}{2}} a_1(R) \sum_{m_s^+m_s^-} (j + \tfrac{1}{2}\,\tfrac{1}{2} m_l^+ m_s^+ \mid jm^+) \\ & \times (j + \tfrac{1}{2}\,\tfrac{1}{2} m_l^- m_s^- \mid jm^-) \\ & \times Z_{j+\frac{1}{2}m_l^+m_l^-} i\gamma_5 \Lambda_{m_s^+m_s^-} + (j+1)^{-\frac{1}{2}} a_2(R) \\ & \times \sum_{m_s^+m_s^-} (j - \tfrac{1}{2}\,\tfrac{1}{2} m_l^+ m_s^+ \mid jm^+) \\ & \times (j - \tfrac{1}{2}\,\tfrac{1}{2} m_l^- m_s^- \mid jm^-) \\ & \times Z_{j-\frac{1}{2}m_l^+m_l^-} i\gamma_5 \Lambda_{m_s^+m_s^-} \Big], \end{aligned} \tag{10}$$

TABLE I. Parity and charge parity of solutions of the Bethe–Salpeter equation.

Solution	Parity	Charge parity	
f^{SV}	$(-1)^J$	$(-1)^{2j}$	
f^{TA+}	$(-1)^J$	$(-1)^{2j+1}$	$(j \neq 0)$
f^{TA-}	$(-1)^{J+1}$	$(-1)^{2j+1}$	
g^{TA+}	$(-1)^J$	$(-1)^{2j+1}$	$(J \neq 0)$
g^{TA-}	$(-1)^{J+1}$	$(-1)^{2j+1}$	$(J \neq 0)$
f^P	$(-1)^{J+1}$	$(-1)^{2j}$	

$$\begin{aligned} f^{TA+}_{jjJm} = \sum_{m^+m^-} (jjm^+m^- \mid Jm) & 2^{-\frac{1}{2}}(2j+1)^{-\frac{1}{2}} \bar{w}(R) \\ & \times \Big\{ \sum_{m_s^+} (j1m_l^+m_s^+ \mid jm^+) Z_{jm_l^+m^-} \Sigma^+_{m_s^+} \\ & - \sum_{m_s^-} (j1m_l^-m_s^- \mid jm^-) Z_{jm^+m_l^-} \Sigma^-_{m_s^-} \Big\}. \end{aligned} \tag{11}$$

Cases B and C similarly give solutions $f^{TA}_{j\,j+1\,Jm}$ and $f^{TA}_{j+1\,j\,Jm}$. To obtain parity eigenfunctions we must take the linear combinations

$$g^{TA\pm}_{j\,j+1\,Jm} = f^{TA}_{j\,j+1\,Jm} \pm f^{TA}_{j+1\,j\,Jm}. \tag{12}$$

In the P sector there is just the solution

$$f^P_{jjJm} = p(R) \sum_{m^+m^-} (jjm^+m^- \mid Jm) Z_{jm^+m^-} \tfrac{1}{2}\gamma_5. \tag{13}$$

The parity and charge parity of these solutions, all of which have total angular momentum J, are given in Table I. (The suffices j^+j^-Jm are omitted.)

3. BOSON–BOSON SYSTEM WITH SQUARE WELL

Written as a differential equation in relative coordinates, the ladder-approximation Wick-rotated Bethe–Salpeter equation for a system of two spinless bosons of equal mass m bound to zero total mass is

$$(\Box - m^2)^2 f(x) = -\lambda \mathcal{U} f(x), \tag{14}$$

where $\mathcal{U}$ is the potential function and λ a coupling constant. Writing the potential due to exchange of spinless quanta of mass μ explicitly, we have $\lambda = g^2/(4\pi)^2$ and[15]

$$\begin{aligned} -\tfrac{1}{4}\mathcal{U} &= (2\pi)^{-4} \int_{-\infty}^{\infty} e^{ik\cdot x} (k^2 + \mu^2)^{-1}\, d^4k \\ &= (2\pi)^{-2} \int_0^\infty R^{-1} K^2 (K^2 + \mu^2)^{-1} J_1(KR)\, dK \\ &= (\mu/R) K_1(\mu R) \\ &\equiv U(\mu, R). \end{aligned} \tag{15}$$

[15] The notation used throughout for the various higher transcendental functions is that of *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun, Eds. (Dover Publications, Inc., New York, 1965).

Here R is the radial coordinate in configuration space, and K the corresponding radial coordinate in the Euclidean relative-momentum space.

Separating (14) in four-dimensional spherical-polar coordinates (R, ψ, θ, ϕ), we have solutions of the form[16]

$$f(x) = g_n(R)\Omega_{nlm}(\psi, \theta, \phi), \tag{16}$$

where

$$\left\{\frac{d^2}{dR^2} + \frac{3}{R}\frac{d}{dR} - \frac{n(n+2)}{R^2} - m^2\right\}^2 g_n(R) = -\lambda \mathcal{V} g_n(R), \tag{17}$$

and n is a nonnegative integer.

We now define the potential to be a square well of depth W and radius $A = a/m$ when

$$\begin{aligned} \lambda\mathcal{V} &= -W, \quad 0 \le R < A, \\ \lambda\mathcal{V} &= 0, \qquad\quad R > A. \end{aligned} \tag{18}$$

Let us take $\rho = mR$ as the independent variable and write

$$d_\alpha^\pm = \frac{d}{d\rho} \pm \frac{2(j+\alpha)}{\rho} = \frac{1}{m} D_\alpha^\pm, \tag{19}$$

$$\Delta = d_{\frac{1}{2}}^+ d_0^- = d_{-\frac{1}{2}}^- d_1^+, \tag{20}$$

where $j = \frac{1}{2}n$. If we then put $\kappa = Wm^{-4}$, Eq. (17) for the square well becomes

$$(\Delta - 1)^2 g = \kappa g, \quad 0 \le \rho < a, \tag{21a}$$

$$(\Delta - 1)^2 g = 0, \qquad \rho > a, \tag{21b}$$

where we have dropped the suffix n from g_n. Solutions of both (21a) and (21b) may be found by comparison with the equation investigated in the Appendix A. In the external region $(\rho > a)$, only solutions g_e vanishing as $\rho \to \infty$ are acceptable and we have

$$g_e = C_1 K_{2j}(\rho) + C_2 \rho^{-1} K_{2j+1}(\rho) \tag{22}$$

for some constants C_1, C_2. In the internal region $(\rho < a)$, we must select solutions g_i of (21a) which remain finite as $\rho \to 0$. For $\kappa \ge 0$, one such solution is

$$g_{i1} = \rho^{-1} I_{2j+1}(\eta\rho), \tag{23}$$

where $\eta = (1 + \kappa^{\frac{1}{2}})^{\frac{1}{2}}$. Another solution g_{i2} has a form depending on κ:

$$\begin{aligned} g_{i2} &= I_{2j}(\rho), && \kappa = 0, \\ g_{i2} &= \rho^{-1} I_{2j+1}(|\beta|\,\rho), && 0 < \kappa < 1, \\ g_{i2} &= \rho^{2j}, && \kappa = 1, \\ g_{i2} &= \rho^{-1} J_{2j+1}(\beta\rho), && \kappa > 1, \end{aligned} \tag{24}$$

where $\beta = (\kappa^{\frac{1}{2}} - 1)^{\frac{1}{2}}$.

[16] For an explicit form of the function Ω_{nlm}, see, e.g., C. Schwartz, Phys. Rev. **137**, B717 (1965), Eq. (13). ($\Omega_{nlm} \equiv |nlm\rangle$.)

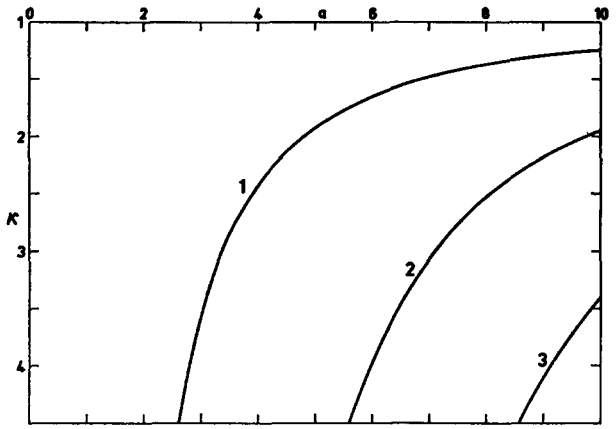


FIG. 1. Curves of zero total mass for two spin-0 bosons each of mass m interacting through a square-well potential of radius a/m and depth κm^4.

For $\kappa < 0$,

$$\begin{aligned} g_{i1} &= \mathrm{Re}(\rho^{-1} J_{2j+1}(\beta\rho)), \\ g_{i2} &= \mathrm{Im}(\rho^{-1} J_{2j+1}(\beta\rho)). \end{aligned} \tag{25}$$

Then for all cases

$$g_i = C_3 g_{i1} + C_4 g_{i1} \tag{26}$$

for some constants C_3, C_4.

Since (17) is a fourth-order differential equation, we must match the values at $\rho = a$ of g_e, g_i and of their first, second, and third derivatives. Three of these relations determine the ratios of C_1, C_2, C_3, and C_4, and the fourth determines a discrete set of depths that the well must have for a given width in order that the bound state has zero total mass.

Testing the cases $\kappa \ge 0$ for $j = 0$ with a computer reveals that solutions exist only for $\kappa > 1$, and the zero-total-mass curves for small values of κ, a are shown in Fig. 1. We have not attempted to test the case $\kappa < 0$.[17]

It would be interesting if we could match in some fashion the results for the square well with those for the theoretical potential of Eq. (15). Certainly there should be qualitative agreement. In relative-momentum space it is a matter of comparing:

$$\text{(i) Field theoretical potential } \frac{\lambda}{K^2 + \mu^2}, \tag{27}$$

$$\text{(ii) Square-well potential } \frac{\kappa m^2 a^2}{4K^2} J_2\left(\frac{aK}{m}\right). \tag{28}$$

We assume that the least value of κ for a given a corresponds to the physical case and restrict our attention to the curve marked "1" in Fig. 1. Its general shape suggests that the quantity $h = a^2(\kappa - 1)$

[17] The remarks at the end of Sec. 6 are applicable here and indicate that solutions are likely to be found for $\kappa < 0$.

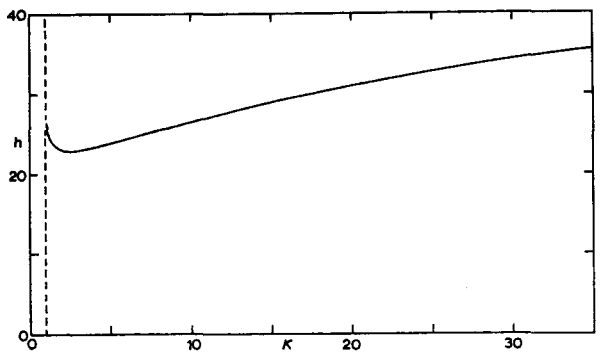


FIG. 2. The quantity $h = a^2(\kappa - 1)$ for curve "1" of Fig. 1, plotted as a function of κ.

may be slowly varying, and in Fig. 2 we plot h as a function of κ. Presumably, well depth corresponds to coupling strength, and if we choose $\kappa = \lambda/2m^2$, since h does not vanish as $\kappa \to 1$, we have for $\lambda \to 2m^2$ that $a \to \infty$. This result is physically reasonable since $\lambda = 2m^2$ corresponds to the exchange particle mass μ vanishing[10] and the field-theoretical potential, therefore, having infinite range. We can put the correspondence between the ranges of the two potentials on a more definite basis by choosing $a = \xi(m/\mu)$ for some constant ξ. Narayanaswamy and Pagnamenta[18] have given computed values of λ for varying m/μ at zero total mass, and it appears that in order to obtain the same computed value of λ at $m/\mu = 1$ we must take $\xi^2 \sim 30$; but if we wish to obtain the same computed value of λ at $m/\mu = 2$ we must take $\xi^2 \sim 20$. Thus, the agreement is more qualitative than quantitative. From the definitions of h and a we have

$$\frac{\lambda}{m^2} = 2 + \frac{1}{\xi^2} \cdot \frac{\mu^2}{m^2} h, \tag{29}$$

which is to be compared with Eq. (18) of Ref. 6.

4. FERMION–ANTIFERMION SQUARE-WELL POTENTIAL

The ladder-approximation Wick-rotated Bethe–Salpeter equation for a spin-$\frac{1}{2}$ fermion–antifermion system of total mass E is [Eq. (I.8)][18a]:

$$(\gamma \cdot \partial - \mu_a E\gamma_4 + m_a)f(x) \times (\gamma \cdot \overleftarrow{\partial} + \mu_b E\gamma_4 + m_b) = -\lambda \mathfrak{V} f(x), \tag{30}$$

where m_a, m_b are, respectively, the fermion and antifermion masses.

Absorbing λ into $\mathfrak{V}$ we put for the general four-dimensional radial potential function [cf. Eq. (I.12)]:

$$\lambda \mathfrak{V} = \sum_i V_i(R)\hat{\Gamma}_i. \tag{31}$$

We now define the potential to be a square well of depth V, radius a/m, and interaction type i when

$$\begin{aligned} V_j(R) &= -V\delta_{ij}, \quad 0 \leq R < a/m, \\ V_j(R) &= 0, \qquad\qquad R > a/m. \end{aligned} \tag{32}$$

We consider only the equal-mass case $m_a = m_b = m$ and, putting $\kappa = V/m^2$, we define in the interior region

$$\begin{aligned} \alpha_{ij} &= -(m^2 + \lambda \bar{V}_j)/m^2 \\ &= c_{ij}\kappa - 1, \end{aligned} \tag{33}$$

where the c_{ij} are the coefficients of Table I in I.

In nonrelativistic quantum theory the sign that a potential term should have in order that the force field be attractive is unambiguous. This is not the case, however, for a relativistic theory such as we are considering here, specifically because E occurs quadratically in the left member of (30), whereas λ appears linearly in the right member. There is as yet no guarantee of the sign of $dE/d\lambda$ and, indeed, for two members E_1, E_2 of a discrete energy spectrum, no guarantee for a given λ that $dE_1/d\lambda$ and $dE_2/d\lambda$ have even the same sign. A direct test, therefore, as to whether the potential is attractive for a particular sign of the potential term is simply to see if a bound state does exist for that sign. Consequently in the following work we examine for both signs of V (or κ) whether zero-mass solutions exist.

Granted this point we note, however, that field theoretical interaction Lagrangians corresponding to each interaction type i do give a quite definite sign for $\mathfrak{V}$. In fact for an exchange field A^i of type i (where $i = S$, V, A, or P) whose quanta have mass μ, the right member of (30) becomes $4\epsilon\lambda U(\mu, R)\hat{\Gamma}_i f(x)$, where interaction type i, interaction Lagrangian $\mathfrak{L}$, and ϵ are as shown in Table II and $\lambda = g^2/(4\pi)^2$ as before.

The sign of ϵ depends simply on whether i appears explicitly in front of the expression for the Hermitian $\mathfrak{L}$ as listed, since in ladder approximation the potential comes from the lowest-order contributing term in the S-matrix expansion and this contains the square of $\mathfrak{L}$.

TABLE II. Corresponding forms for i, $\mathfrak{L}$, and ϵ.

i	$\mathfrak{L}$	ϵ
S	$g\bar{\psi}A^S\psi$	1
V	$ig\bar{\psi}\gamma_\mu A^V_\mu\psi$	−1
A	$g\bar{\psi}i\gamma_5\gamma_\mu A^A_\mu\psi$	1
P	$ig\bar{\psi}\gamma_5 A^P\psi$	−1

[18] See Ref. 6, p. 640, Table I and p. 647, Fig. 1.

[18a] Through a typographical error, the left member of this equation was given incorrectly in Eq. (I.8).

The tensor interaction has been excluded from the above comparisons, since the interaction Lagrangian $\mathfrak{L}_T = g\bar{\psi}\sigma_{\mu\nu}A^T_{\mu\nu}\psi$, which would be the direct analog of those appearing in Table II, is intrinsically different. The exchange-field contribution to $\mathfrak{L}_T$ comes only from the antisymmetric part of $A^T_{\mu\nu}$, and this corresponds to the reducible $\mathfrak{D}^{01} + \mathfrak{D}^{10}$ representation of the proper Lorentz group, i.e., a spin-1 field. But basically a spin-1 field is a vector representation, so that, in order to describe it by second-rank tensor components $A^T_{\mu\nu}$, the latter necessarily involve *derivatives* of the basic vector field components. In other words $\mathfrak{L}_T$ describes a derivative coupling which implies a singularity of the potential of even higher order at the origin. It is exemplified by the electromagnetic field where $A^T_{\mu\nu} = F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and where $F_{\mu\nu}$, A_μ have their usual meanings.

5. FORMAL ANALYTICAL SOLUTIONS

When E vanishes, Eq. (30) splits into Dirac-space S–V, T–A, and P sectors. Angular variables may be separated and the equation reduces to sets of coupled radial equations which in I are written down explicitly for the various possible cases. In what follows, each set is solved analytically in the interior and exterior regions for the potential (32).

A. S–V Sector

Using (19) and (20), Eqs. (I.43) become

$$\begin{aligned}
&(\Delta - \alpha_{i1})s + 2(d^+_{\frac{1}{2}}v_1 + d^-_{-\frac{1}{2}}v_2) = 0,\\
&d^-_0\{d^+_{\frac{1}{2}}v_1 + 2(j+1)(d^-_{-\frac{1}{2}}v_2 + s)\} = (2j+1)\alpha_{i2}v_1,\\
&d^+_1\{-d^-_{-\frac{1}{2}}v_2 + 2j(d^+_{\frac{1}{2}}v_1 + s)\} = (2j+1)\alpha_{i1}v_2,\\
&\qquad 0 \le \rho < a, \quad j \neq 0, \qquad (34)
\end{aligned}$$

in the interior region and are of the same form in the exterior region but with α_{i1}, α_{i2} replaced by -1.

If v_1, v_2 are eliminated we find

$$\{\Delta^2 - (\alpha_{i1} + \alpha_{i2} + 4)\Delta + \alpha_{i1}\alpha_{i2}\}s = 0, \qquad (35)$$

the solutions of which follow by comparison with the equation investigated in the Appendix A.

Given a solution for s, solutions for v_1, v_2 are found from the equations

$$\begin{aligned}
&(2j+1)(d^-_0 d^+_{\frac{1}{2}} + \alpha_{i2})v_1 = -(j+1)d^-_0(\Delta - \alpha_{i1} - 2)s,\\
&(2j+1)(d^+_1 d^-_{-\frac{1}{2}})v_2 + \alpha_{i2} = -jd^+_1(\Delta - \alpha_{i1} - 2)s, \qquad (36)
\end{aligned}$$

which readily follow from (34). Mandelstam's analysis[9] shows that a solution will be physically acceptable only if for some positive ϵ it is $O(R^{\epsilon-1})$ as $R \to 0$, and a detailed check shows that s, v_1, v_2 must in fact all be bounded for the square-well potential as $R \to 0$.

The various possible real forms of coupled physically acceptable formal solutions are displayed in Table III.

When the roots of the quadratic equation corresponding to (35) are real and unequal, to any positive root η^2 corresponds a solution (a), to any negative root $-\beta^2$ corresponds a solution (b), and to a zero root (when $\alpha_{i2} \neq 0$) corresponds a solution (c). The two roots give two distinct solutions, and a third is given by (d) or (e) according as α_{i2} is positive or negative. When $\alpha_{i2} = 0$, instead of (c) and (d) or (e), we have (f) and (g).

When the roots of the quadratic equation are real and equal, if they are positive we have: (a), (h), and (d) or (e); and, if negative: (b), (i), and (d) or (e).

When the roots are complex, we have from the consideration of this case in the Appendix and from the reality of the operators in (36) that two real

TABLE III. Solutions of the coupled equations (34).

s	v_1	v_2		
$\frac{1}{\rho}I_{2j+1}(\sigma_1)$	$\frac{1}{2}\left(\frac{j+1}{2j+1}\right)(\alpha_{i1} - \eta^2)\frac{1}{\sigma_1}I_{2j+2}(\sigma_1)$	$\frac{1}{2}\left(\frac{j}{2j+1}\right)(\alpha_{i1} - \eta^2)\frac{1}{\sigma_1}I_{2j}(\sigma_1)$	$\sigma_1 = \eta\rho$	(a)
$\frac{1}{\rho}J_{2j+1}(\sigma)$	$\frac{1}{2}\left(\frac{j+1}{2j+1}\right)(\alpha_{i1} + \beta^2)\frac{1}{\sigma}J_{2j+2}(\sigma)$	$-\frac{1}{2}\left(\frac{j}{2j+1}\right)(\alpha_{i1} + \beta^2)\frac{1}{\sigma}J_{2j}(\sigma)$	$\sigma = \beta\rho$	(b)
ρ^{2j}	0	$(4j/\alpha_{i2})\rho^{2j-1}$		(c)
0	$\rho^{-1}J_{2j+2}(\alpha^{\frac{1}{2}}_{i2}\rho)$	$\rho^{-1}J_{2j}(\alpha^{\frac{1}{2}}_{i2}\rho)$	$\alpha_{i2} > 0$	(d)
0	$\rho^{-1}I_{2j+2}(\lvert\alpha_{i2}\rvert^{\frac{1}{2}}\rho)$	$-\rho^{-1}I_{2j}(\lvert\alpha_{i2}\rvert^{\frac{1}{2}}\rho)$	$\alpha_{i2} < 0$	(e)
$(2j+1)\rho^{2j}$	$\{(\alpha_{i1} - 4j)/[8(j+1)]\}\rho^{2j+1}$	$\frac{1}{2}j(\alpha_{i1} + 2)\rho^{2j+1}$		(f)
0	0	ρ^{2j-1}		(g)
$\eta I_{2j}(\sigma_1)$	$\left(\frac{j+1}{2j+1}\right)[\frac{1}{2}(\alpha_{i1} - \eta^2)I_{2j+1}(\sigma_1) - \alpha_{i1}\frac{1}{\sigma_1}I_{2j+2}(\sigma_1)]$	$\left(\frac{j}{2j+1}\right)[\frac{1}{2}(\alpha_{i1} - \eta^2)I_{2j-1}(\sigma_1) - \eta^2\frac{1}{\sigma_1}I_{2j}(\sigma_1)]$		(h)
$\beta J_{2j}(\sigma)$	$\left(\frac{j+1}{2j+1}\right)[\frac{1}{2}(\alpha_{i1} + \beta^2)J_{2j+1}(\sigma) - \alpha_{i1}\frac{1}{\sigma}J_{2j+2}(\sigma)]$	$-\left(\frac{j}{2j+1}\right)[\frac{1}{2}(\alpha_{i1} + \beta^2)J_{2j-1}(\sigma) + \eta^2\frac{1}{\sigma}J_{2j}(\sigma)]$		(i)

solutions are obtained from the real and imaginary parts of (b) and the third solution is (d) or (e).

The nature of the roots for the various ranges and values of κ for each type of interaction is detailed in Table IV. The general physically acceptable solution in any particular case will then be some linear combination of the corresponding three formal solutions.

Solutions for the exterior region may be found by putting $\kappa = 0$ in (34) *et seq.* and selecting forms which vanish as $\rho \to \infty$. The three acceptable forms are displayed in Eq. (37) where for brevity we have omitted the argument ρ of each modified Bessel function:

$$\begin{array}{lccc} & \text{(a)} & \text{(b)} & \text{(c)} \\ s = (2j+1)K_{2j} - \rho^{-1}K_{2j+1} & & \rho^{-1}K_{2j+1} & \rho^{-1}K_{2j+1}, \\ v_1 = (j+1)K_{2j+1} & & \rho^{-1}K_{2j+2} & 0, \\ v_2 = jK_{2j-1} & & 0 & \rho^{-1}K_{2j}. \end{array} \tag{37}$$

When $j = 0$, in place of (34) we have

$$\begin{gathered} d_{\frac{3}{2}}^{+}(d_0^{-}s + 2v_1) = \alpha_{i1}s, \\ d_0^{-}(d_{\frac{3}{2}}^{+}v_1 + 2s) = \alpha_{i2}v_1, \\ 0 \leq \rho < a, \end{gathered} \tag{38}$$

and the corresponding exterior region equations obtained by putting $\kappa = 0$. We may again derive Eq. (35). Formal solutions corresponding to Table III (a), (b), (c), (f), (h), and (i) may be derived and are in fact identical to the forms found for s and v_1 in that table when we put $j = 0$ there. Exterior-region formal solutions corresponding to (37) are:

$$\begin{array}{lcc} & \text{(a)} & \text{(b)} \\ s = \rho^{-1}K_1 & & K_0, \\ v = \rho^{-1}K_2 & & \rho^{-1}K_2 + K_1. \end{array} \tag{39}$$

TABLE IV. The character of the roots of the quadratic equation corresponding to (35).

κ	Roots		
	S interaction		
> 1	+	+	
$= 1$	+	0	$(\alpha_{i1} = \alpha_{i2} = 0)$
$< 1, > 0$	+	+	
$= 1$	+	+	(equal)
< 0	complex		
	V interaction		
$> \frac{1}{4}$	+	−	
$= \frac{1}{4}$	+	0	$(\alpha_{i1} = 0)$
$< \frac{1}{4}, > 0$	+	+	
$= 0$	+	+	(equal)
$< 0, > -\frac{4}{9}$	complex		
$= -\frac{4}{9}$	+	+	(equal)
$< -\frac{4}{9}, > -\frac{1}{2}$	+	+	
$= -\frac{1}{2}$	+	0	$(\alpha_{i2} = 0)$
$< -\frac{1}{2}$	+	−	
	T interaction		
$> \frac{1}{6}$	+	−	
$= \frac{1}{6}$	+	0	$(\alpha_{i1} = 0)$
$< \frac{1}{6}, > 0$	+	+	
$= 0$	+	+	(equal)
$< 0, > -\frac{4}{3}$	complex		
$= -\frac{4}{3}$	−	−	(equal)
$< -\frac{4}{3}$	−	−	
	A interaction		
$> \frac{1}{2}$	+	+	
$= \frac{1}{2}$	+	0	$(\alpha_{i2} = 0)$
$< \frac{1}{2}, > \frac{1}{4}$	+	−	
$= \frac{1}{4}$	+	0	$(\alpha_{i1} = 0)$
$< \frac{1}{4}, > 0$	+	+	
$= 0$	+	+	(equal)
$< 0, > -12$	complex		
$= -12$	−	−	(equal)
< -12	−	−	
	P interaction		
> 1	+	−	
$= 1$	+	0	$(\alpha_{i1} = 0)$
$< 1, > 0$	+	+	
$= 0$	+	+	(equal)
$< 0, > -1$	+	+	
$= -1$	+	0	$(\alpha_{i2} = 0)$
< -1	+	−	

B. *T–A* Sector

From Eq. (I.46), for Case A, $j \neq 0$, we find in the interior region that w satisfies

$$\{\Delta^2 - (\alpha_{i3} + \alpha_{i4} + 4)\Delta + \alpha_{i3}\alpha_{i4}\}w = 0, \tag{40}$$

which is very similar to Eq. (35). One can follow the argument for the *S–V* sector from Eq. (35) to Eq. (37); but with the replacements $s \to w$, $v_1 \to a_1$, $v_2 \to a_2$, $\alpha_{i1} \to \alpha_{i3}$, $\alpha_{i2} \to \alpha_{i4}$, an explicit factor $j \rightleftarrows$ an explicit factor $(j+1)$, an explicit factor $(2j+1)$ is unaltered, and exponents and Bessel-function orders are unaltered. There is a single exception—the analog of the form for v_1 in Table III (f) is

$$a_1 = -\frac{\alpha_{i3} + 4(j+1)}{8(j+1)}\rho^{2j+1}.$$

The nature of the roots for the various ranges and values of κ for each type of interaction is shown in Table V.

We defer consideration of $\bar{w}$ and of the $j = 0$ case [Eqs. (I.47) and (I.52)] and treat these with the *P*-sector solutions below.

From Eq. (I.46) for Case B we find that in the interior region a satisfies

$$\{\Delta_1^2 - (\alpha_{i3} + \alpha_{i4} + 4)\Delta_1 + \alpha_{i3}\alpha_{i4}\}a = 0, \tag{41}$$

where $\Delta_1 = d_2^{+}d_{\frac{1}{2}}^{-} = d_0^{-}d_{\frac{3}{2}}^{+}$. Δ_1 is also Δ with j

TABLE V. The character of the roots of the quadratic equation corresponding to (40).

κ	Roots			κ	Roots		
	S interaction				T or A interaction		
>1	+	+		>4	−	−	
$=1$	+	0	$(\alpha_{i3}=\alpha_{i4}=0)$	$=4$	−	−	(equal)
$<1, >0$	+	+		$<4, >0$	complex		
$=0$	+	+	(equal)	$=0$	+	+	(equal)
<0	complex			$<0, >-\frac{1}{2}$	+	+	
				$=-\frac{1}{2}$	+	0	($\alpha_{i3}=0$ for T $\alpha_{i4}=0$ for A)
				$<-\frac{1}{2}$	+	−	
	V interaction				P interaction		
$>\frac{1}{2}$	+	−		>1	+	−	
$=\frac{1}{2}$	+	0	$(\alpha_{i4}=0)$	$=1$	+	0	$(\alpha_{i3}=0)$
$<\frac{1}{2}, >0$	+	+		$<1, >0$	+	+	
$=0$	+	+	(equal)	$=0$	+	+	(equal)
$<0, >-4$	complex			$<0, >-1$	+	+	
$=-4$	−	−	(equal)	$=-1$	+	0	$(\alpha_{i4}=0)$
<-4	−	−		<-1	+	−	

replaced by $j+\frac{1}{2}$. The analysis proceeds as for Case A, and the forms of solution are displayed in Table VI. Each form is the direct analog of the solution appearing in Table III which is indexed by the same letter [(a), (b), etc.]. The exterior-region solutions are

$$\begin{array}{lccc} & \text{(a)} & \text{(b)} & \text{(c)} \\ a = & K_{2j+1} & \rho^{-1}K_{2j+2} & \rho^{-1}K_{2j+2}, \\ t_1 = & K_{2j+2} & 2\rho^{-1}K_{2j+3} & 0, \\ t_2 = & K_{2j} & 0 & 2\rho^{-1}K_{2j+1}. \end{array} \tag{42}$$

It is not necessary to consider the T–A-sector Case C of I separately, because of the symmetry shown in that paper to exist between Cases B and C.

C. P Sector

From Eq. (I.53) p satisfies

$$(\Delta+\alpha_{i5})p=0, \quad 0\leq\rho<a, \tag{43}$$

for which the only solutions bounded as $\rho\to 0$ are

$$\alpha_{i5}>0, \quad p=\sigma^{-1}J_{2j+1}(\sigma), \quad \sigma=(\alpha_{i5})^{\frac{1}{2}}\rho, \tag{44a}$$

$$\alpha_{i5}=0, \quad p=\rho^{2j}, \tag{44b}$$

$$\alpha_{i5}<0, \quad p=\sigma^{-1}I_{2j+1}(\sigma), \quad \sigma=(-\alpha_{i5})^{\frac{1}{2}}\rho. \tag{44c}$$

The exterior-region solution is

$$p=\rho^{-1}K_{2j+1}(\rho). \tag{45}$$

From Eqs. (I.47) and (I.52), the T–A-sector Case A equations for $\bar{w}$ and a_1, respectively, are

$$(\Delta+\alpha_{i3})\bar{w}=0, \quad j\neq 0, \tag{46}$$

$$(\Delta_1+\alpha_{i4})a_1=0, \quad j=0 \tag{47}$$

in the interior region. The solution of each follows immediately from (44) upon replacing α_{i5} with α_{i3}, α_{i4} respectively, and also j by $\frac{1}{2}$ in the latter case. In the exterior region, $\bar{w}$ satisfies (45) and

$$a_1=\rho^{-1}K_2(\rho). \tag{48}$$

6. NUMERICAL RESULTS

We consider first the cases discussed in Sec. 5C of this paper.

The boundary conditions relating internal and external solutions are that the radial functions and their derivatives are continuous at the well radius $\rho=a$. Thus for α_{i5} positive, from (44) and (45) we

TABLE VI. Solutions of the T–A Sector, Case B equations.

a	t_1	t_2	
$\rho^{-1}I_{2j+2}(\sigma_1)$	$\frac{1}{2}(\alpha_{i4}-\eta^2)\sigma^{-1}I_{2j+2}(\sigma_1)$	$\frac{1}{2}(\alpha_{i4}-\eta^2)\sigma_1^{-1}I_{2j+1}(\sigma_1)$	(a)
$\rho^{-1}J_{2j+2}(\sigma)$	$\frac{1}{2}(\alpha_{i4}+\beta^2)\sigma^{-1}J_{2j+2}(\sigma)$	$-\frac{1}{2}(\alpha_{i4}+\beta^2)\sigma^{-1}J_{2j+1}(\sigma)$	(b)
ρ^{2j+1}	0	$8(j+1)\rho^{2j}/\alpha_{i3}$	(c)
0	$\rho^{-1}J_{2j+2}(\alpha_{i3}^{\frac{1}{2}}\rho)$	$-\rho^{-1}J_{2j+1}(\alpha_{i3}^{\frac{1}{2}}\rho)$	(d)
0	$\rho^{-1}I_{2j+2}(\lvert\alpha_{i3}\rvert^{\frac{1}{2}}\rho)$	$\rho^{-1}I_{2j+1}(\lvert\alpha_{i3}\rvert^{\frac{1}{2}}\rho)$	(e)
$(2j+3)\rho^{2j+1}$	$-\rho^{2j+2}$	$\frac{1}{2}(2j+3)(\alpha_{i4}+2)\rho^{2j+2}$	(f)
0	0	ρ^{2j}	(g)
$\eta I_{2j+1}(\sigma_1)$	$\frac{1}{2}(\alpha_{i4}-\eta^2)I_{2j+2}(\sigma_1)-\alpha_{i4}\sigma_1^{-1}I_{2j+2}(\sigma_1)$	$\frac{1}{2}(\alpha_{i4}-\eta^2)I_{2j}(\sigma_1)-\eta^2\sigma^{-1}I_{2j+1}(\sigma_1)$	(h)
$\beta J_{2j+1}(\sigma)$	$\frac{1}{2}(\alpha_{i4}+\beta^2)J_{2j+2}(\sigma)-\alpha_{i4}\sigma^{-1}J_{2j+2}(\sigma)$	$-\frac{1}{2}(\alpha_{i4}+\beta^2)J_{2j}(\sigma)-\beta^2\sigma^{-1}J_{2j+1}(\sigma)$	(i)

TABLE VII. Numerical values of $z_{j,s}$.

	s	$j=0$	$\frac{1}{2}$	1	$\frac{3}{2}$
$a=1$	1	2.65	3.92	5.18	6.40
	2	5.64	7.07	8.44	9.78
	3	8.73	10.21	11.64	13.03
	4	11.85	13.35	14.81	16.23
	5	14.97	16.49	17.97	19.42
$a=2$	1	2.88	4.07	5.27	6.47
	2	5.79	7.16	8.51	9.82
	3	8.83	10.28	11.68	13.06
	4	11.93	13.40	14.85	16.26
	5	15.04	16.54	18.00	19.44

have at $\rho = a$, putting $\beta = (\alpha_{i5})^{\frac{1}{2}}$,

$$a\frac{K_{2j}(a)}{K_{2j+1}(a)} = -\beta a\frac{J_{2j}(\beta a)}{J_{2j+1}(\beta a)}. \tag{49}$$

Since the left member is positive for all values of a, β must be chosen so that $J_{2j}(\beta a)/J_{2j+1}(\beta a)$ is negative. If $j_{\nu,s}$ is the sth positive zero of $J_\nu(z)$[19]:

$$j_{\nu,s} < j_{\nu+1,s} < j_{\nu,s+1}.$$

From this it follows that $zJ_{2j}(z)/J_{2j+1}(z)$ decreases from zero to minus infinity as z goes from $j_{2j,s}$ to $j_{2j+1,s}$. Thus given a, there is a solution of (49) for some value of βa lying between $j_{2j,s}$ and $j_{2j+1,s}$ for each $s = 1, 2, \cdots$. Furthermore, if $z_{j,s}$ is this value,

$$\beta = z_{j,s}/a \tag{50}$$

and since $z_{j,s}$ varies only over the small range $j_{2j,s} < z_{j,s} < j_{2j+1,s}$, we have approximately

$$\beta \propto a^{-1}$$

for each of the (discrete infinity of) values of β corresponding to a given a. This approximation is least good for small j and s; $z_{0,0}$ varies by about 60% over the complete range of a, but for example, $z_{3,3}$ has a 10% variation and $z_{0,20}$ a 2% variation. Numerical values for some $z_{j,s}$ are shown in Table VII for two values of a.

When α_{i5} is negative we have I instead of J functions in the right member of (49). But (49) then has no solution, since its left and right members are, respectively, positive and negative. A somewhat similar result follows when $\alpha_{i5} = 0$. Solutions therefore exist only for $\alpha_{i5} > 0$. For S, T, and P interactions ($i = 1, 3, 5$) this implies that the well depth κ must exceed a certain value κ_m (1, $\frac{1}{6}$, 1, respectively) in order that the system can bind to zero mass. For $\kappa > \kappa_m$ there are infinite series of widths a possible, and from (50)

$$a^2 = \frac{z_{j,s}^2}{c_{i5}}\frac{1}{\kappa-\kappa_m} \dot{\propto} \frac{1}{\kappa-\kappa_m}. \tag{51}$$

The similarity to results for the boson–boson system of Sec. 3 is clear with $z_{j,s}^2$ being the analog of h. In contrast to h, $z_{j,s}$ is, however, strictly bounded. For V and A interactions, only negative values of κ less than $-\frac{1}{4}(= -\kappa_m'$, say) give zero-mass states and in place of (51) we have

$$a^2 = \frac{z_{j,s}^2}{(-c_{i5})}\frac{1}{\kappa'-\kappa_m'} \dot{\propto} \frac{1}{\kappa'-\kappa_m'}, \tag{52}$$

where $\kappa' = -\kappa$. Comparing the signs of κ with the signs of ϵ in Sec. 4, we see that the exchange potentials there correspond with square-well potentials that produce bound states (of zero mass) for S or V interactions, but not A or P interactions. Let us consider this more carefully. In place of (43), the exchange potential gives rise to the radial-differential equation

$$(\Delta - 1)g(\rho) = -4\epsilon c_{i5}\lambda\cdot\frac{\mu}{m\rho}\cdot K_1\left(\frac{\mu}{m}\cdot\rho\right)g(\rho). \tag{53}$$

For $\lambda \geq \lambda_0 = (2j+1)^2/(4\epsilon c_{i5})$, neither solution satisfies the Mandelstam criterion at the origin. For $\lambda < \lambda_0$, one solution does not satisfy the Mandelstam criterion and the other does. However, at least for $\mu = 0$, the latter solution increases at infinity and is therefore unacceptable. Hence we have the "Goldstein problem."

If we try to match the field-theoretical potential with the square well, it seems natural to associate the dividing value λ_0 with κ_m or κ_m'. Consequently we assume

$$\frac{\kappa}{\lambda} = \frac{1}{c_{i5}}\left[\frac{(2j+1)^2}{4\epsilon c_{i5}}\right]^{-1},$$

or

$$\kappa = 4\epsilon\lambda/(2j+1)^2. \tag{54}$$

If we again take $a = \xi(m/\mu)$ we obtain as the analog of (29):

$$\lambda = \frac{\epsilon(2j+1)^2}{4c_{i5}}\left[1 + \frac{1}{\xi^2}\cdot\frac{\mu^2}{m^2}\cdot z_{j,s}^2\right]. \tag{55}$$

Since λ must be positive, this result can obtain, as mentioned above, only for S or V interactions.

We see that, even allowing a cutoff procedure by matching with a square well, one cannot obtain P-sector zero-mass states with A or P interactions and, in particular, no 0^- state is produced this way.

[19] See Ref. 15, p. 370, inequality (9.5.2).

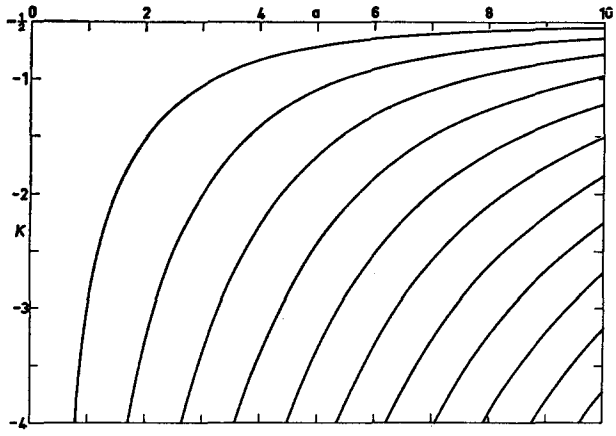


FIG. 3. Curves of zero total mass for a spin-½ fermion–anti-fermion 0^+ system with a vector-interaction square-well potential of radius a/m and depth $\epsilon\kappa m^2$.

For $\bar{w}$, only S, T, P interactions can produce bound states and, of these, only S has the same well-depth sign as for the exchange potential.

From (47), for a_1 with $j = 0$, there is no bound state with a T interaction and of the others only S and P interactions agree with the exchange potential signs.

As a further example let us consider the numerical solutions for $j = 0$ in the S–V sector for each interaction type. According to Table I, the bound system has $J = 0$, positive parity, and positive charge parity. Using a computer for all regions of κ where the corresponding quadratic has real roots (see Table IV for these regions), we find sets of solutions for zero total mass in the following seven cases:

(i) V interaction $\kappa > \frac{1}{4}$
(ii)* V interaction $\kappa < -\frac{1}{2}$
(iii) T interaction $\kappa > \frac{1}{6}$
(iv) T interaction $\kappa < -\frac{4}{3}$
(v) A interaction $\kappa \lesssim -12$
(vi) P interaction $\kappa > 1$
(vii)* P interaction $\kappa < -1$.

The asterisks mark cases where the well has the same sign as the corresponding model field-theoretical potential. Graphs of well-width vs well-depth for cases (ii) and (vii) are given in Figs. 3 and 4 respectively. The curves are clearly again of the approximate form of (51).

Defining $h_V = -a^2(\kappa + \frac{1}{2})$, $h_P = -a^2(\kappa + 1)$ for (ii) and (vii) we demonstrate the comparatively small variation of these quantities in Fig. 5 for the curve of smallest a for given κ. Also, using analytical forms for the functions involved near the origin, we have been able to obtain the result that asymptotically as $a \to 0$, $h_V \to \frac{12}{7}$, $h_R \to 8$.

It is clear that for V, T, and P interactions the potential is attractive for *both* signs of κ.

No attempt has been made to carry out numerical calculations for regions where the corresponding quadratic has complex solutions, but we can perhaps gain some idea of what to expect by the following considerations. We notice that no zero-mass solutions are found for cases where the corresponding quadratic has real roots which are either both positive or one positive and one zero. On the other hand, zero-mass solutions are always found where there are real roots at least one of which is negative and where the corresponding range of κ is infinite. Positive roots correspond to monotonic functions and negative roots to oscillating functions in the interior of the well. For complex arguments the functions are of the type $\mathrm{Re}J_n(z)$, $\mathrm{Im}J_n(z)$, and asymptotically

$$\mathrm{Re}J_n(z) \sim (2\pi r)^{-\frac{1}{2}} e^{r \sin(|\alpha|)} \cos\theta,$$
$$\mathrm{Im}J_n(z) \sim (2\pi r)^{-\frac{1}{2}} e^{r \sin(|\alpha|)} (\alpha/|\alpha|) \sin\theta,$$

where

$$z = re^{i\alpha},$$
$$\theta = r\cos\alpha + \tfrac{1}{2}|\alpha| - \tfrac{1}{2}(n + \tfrac{1}{2})\pi.$$

Thus for $\alpha \neq \frac{1}{2}\pi$, both $\mathrm{Re}J_n(z)$ and $\mathrm{Im}J_n(z)$ are oscillating. From this it seems likely that massless bound states will be found for the S interaction when $\kappa < 0$, but not for the complex regions for the V or T interactions. With the A interaction the region $0 > \kappa > -12$ may give massless states, since it is contiguous to the infinite region $\kappa \lesssim -12$ which does have such states. Neither for the S nor the A interaction, however, must we necessarily expect that solutions will be found right up to the region boundary at $\kappa \to 0$.

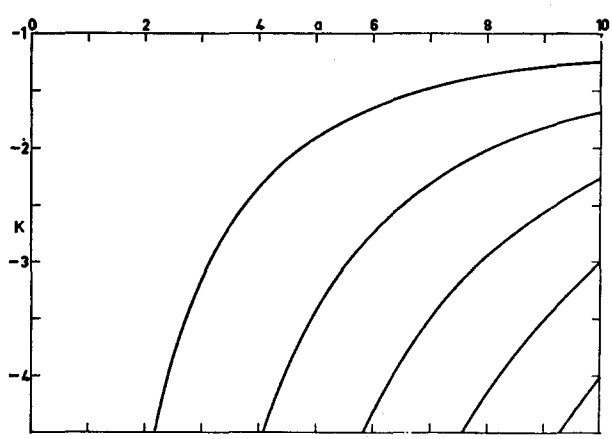


FIG. 4. Curves of zero total mass for a spin-½ fermion–anti-fermion 0^+ system with a pseudoscalar-interaction square-well potential of radius a/m and depth $\epsilon\kappa m^2$.

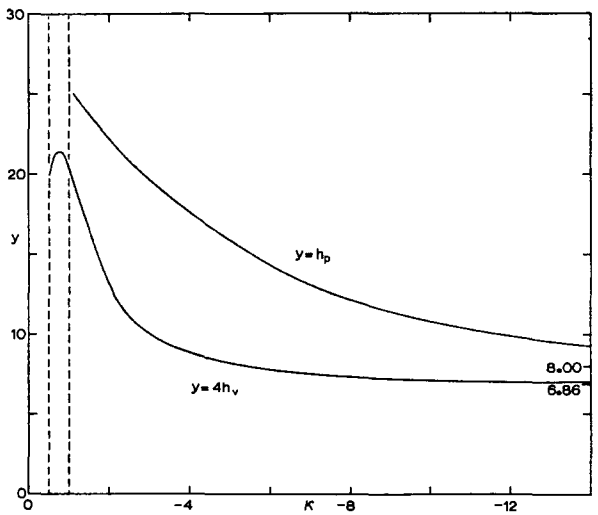


FIG. 5. The quantities $4h_V$, h_P for the curves of least a for given κ from Figs. 3 and 4 respectively, plotted as a function of κ. The asymptotic values $h_V \to \frac{1}{7}^2(4h_V \to 6.86)$, $h_P \to 8$ are indicated.

ACKNOWLEDGMENTS

I should like to thank Dr. P. C. M. Yock for many helpful discussions and Professor R. H. Dalitz for his interest in this work.

APPENDIX

Let the quadratic equation

$$F(X) \equiv X^2 + 2bX + c = 0 \tag{A1}$$

have real coefficients b, c and suppose its roots are $-\beta^2$, $-\gamma^2$ so that we may write

$$X^2 + 2bX + c \equiv (X + \beta^2)(X + \gamma^2). \tag{A2}$$

Consider the general fourth-order differential equation in Δ with real coefficients

$$F(\Delta)f(\rho) = 0. \tag{A3}$$

The character of its solutions depends on that of the roots of (A1) and we consider the various cases separately.

Case A: $c \neq 0$, $b^2 - c \neq 0$. β^2, γ^2 are nonzero and unequal and the solutions of (A3) are those of

$$(\Delta + \beta^2)f = 0, \tag{A4}$$

together with those of

$$(\Delta + \gamma^2)f = 0. \tag{A5}$$

A complete set is therefore given by

$$\begin{aligned} &\rho^{-1}J_{2j+1}(\beta\rho), \quad \rho^{-1}Y_{2j+1}(\beta\rho), \\ &\rho^{-1}J_{2j+1}(\gamma\rho), \quad \rho^{-1}Y_{2j+1}(\gamma\rho), \end{aligned} \tag{A6}$$

where β, γ are convenient square roots of the possibly complex quantities β^2, γ^2.

If $b^2 - c > 0$ then β^2, γ^2 are real, and if they are positive then taking the positive square roots gives the set (A6) of solutions in real form. If (say) β^2 is negative, the solutions of (A3) obtained from (A4) are real if we take them to be

$$\rho^{-1}I_{2j+1}(|\beta|\,\rho), \quad \rho^{-1}K_{2j+1}(|\beta|\,\rho). \tag{A7}$$

If $b^2 - c < 0$, $\gamma = \bar{\beta}$ and since $J_\nu(\bar{z}) = \overline{J_\nu(z)}$, $Y_\nu(\bar{z}) = \overline{Y_\nu(z)}$, a *real* complete set of solutions is

$$\begin{aligned} &\mathrm{Re}(\rho^{-1}J_{2j+1}(\beta\rho)), \quad \mathrm{Im}(\rho^{-1}J_{2j+1}(\beta\rho)), \\ &\mathrm{Re}(\rho^{-1}Y_{2j+1}(\beta\rho)), \quad \mathrm{Im}(\rho^{-1}Y_{2j+1}(\beta\rho)), \end{aligned} \tag{A8}$$

and we suppose $-\frac{1}{2}\pi < \arg(\beta) < \frac{1}{2}\pi$.

Case B: $c \neq 0$, $b^2 - c = 0$. Equation (A3) becomes

$$(\Delta + \beta^2)^2 f = 0. \tag{A9}$$

If f_1 satisfies (A4), it satisfies (A9); and it is easily shown that $[\partial/\partial(\beta^2)]f_1$ also satisfies (A9). A complete real set for β positive is, therefore,

$$\rho^{-1}J_{2j+1}(\beta\rho), \quad \rho^{-1}Y_{2j+1}(\beta\rho), \quad J_{2j}(\beta\rho), \quad Y_{2j}(\beta\rho) \tag{A10}$$

and, for β^2 negative,

$$\begin{aligned} &\rho^{-1}I_{2j+1}(|\beta|\,\rho), \quad \rho^{-1}K_{2j+1}(|\beta|\,\rho), \\ &I_{2j}(|\beta|\,\rho), \quad K_{2j}(|\beta|\,\rho). \end{aligned} \tag{A11}$$

Case C: $c = 0$, $b \neq 0$. Just one root of (A1) vanishes and the corresponding solutions are

$$\rho^{2j}, \quad \rho^{-2(j+1)}. \tag{A12}$$

Case D: $c = 0$, $b = 0$. The complete set of solutions is

$$\rho^{2(j+1)}, \quad \rho^{2j}, \quad \rho^{-2j}, \quad \rho^{-2(j+1)}. \tag{A13}$$

Method for the Evaluation of Certain Time-Dependent Thermal Averages

DAVID W. HOWGATE
Solid State Physics Branch, Physical Sciences Laboratory, Redstone Arsenal, Alabama

(Received 7 July 1968)

A simple method is presented for the evaluation of time-dependent thermal averages relating to the Debye–Waller form.

Recently, Terwiel[1] has presented a method of evaluating averages of the Debye–Waller form

$$\langle e^{\theta}\rangle \equiv \mathrm{Tr}\, e^{-\beta H}e^{\theta}/\mathrm{Tr}\, e^{-\beta H}, \quad \beta = 1/k_BT,$$
$$H = \sum_i \omega_i(a_i^\dagger a_i + \tfrac{1}{2}), \quad \theta = \sum_i (C_i a_i + D_i a_i^\dagger),$$
$$[a_i, a_j^\dagger] = \delta_{ij}, \quad [a_i, a_j] = 0, \quad \text{etc.}, \tag{1}$$

which, apart from the cyclic property of the trace, employs only the relations

$$e^{-\beta H}a_i e^{\beta H} = e^{\beta\omega}a_i, \tag{2a}$$
$$e^{\lambda\theta}a_i e^{-\lambda\theta} = a_i - \lambda D_i. \tag{2b}$$

One calculates $\langle e^{\theta}\rangle$ by using the expression

$$F(\lambda) \equiv \langle e^{\lambda\theta}\rangle \tag{3}$$

and by utilizing the properties of the differential form

$$dF/d\lambda = \sum_i f_i. \tag{4}$$

Terwiel's approach to the calculation of thermal averages—illustrated in relations (2) through (4)—is not limited solely to the evaluation of expressions in the Debye–Waller form involving only a thermal factor. For example, his methods can easily be extended to include an evaluation of $K(t)$, such that

$$K(t) = \int_{-\infty}^{\infty} d\omega e^{-i\omega t}K(\omega),$$

were $K(t)$ is the transform of the radiative spectral-absorption function $K(\omega)$ corresponding to the absorption of external radiation by a defect or impurity electron trapped in a crystalline lattice. In fact, it is possible to calculate a variety of thermally averaged multiple-time functions by using an extension of his procedure.

To see how this can be done, first consider a general time-dependent thermal average $\langle A(q, p; \{t\})\rangle$, subject to the conditions[2]

$$\langle A(q, p; \{t\})\rangle = \prod_i \langle A_i(q, p; \{t\})\rangle, \tag{5a}$$

$$A_i(q, p; \{t\}) = e^{\theta_i(q,t_q)}e^{\theta_i(q+1,t_{q+1})}\cdots e^{\theta_i(p,t_p)}, \quad p \geq q \text{ are integers}, \tag{5b}$$

$$\theta_i(q, t) = e^{iHt}\theta_i(q)e^{-iHt}, \quad \theta_i(q) = c_i(q)a_i + d_i(q)a_i^\dagger, \tag{5c}$$

$$[\theta_i(q), \theta_j(q')] = 0. \tag{5d}$$

We are particularly interested in the sequence $\langle A(1, p; \{t\})\rangle$. In the case of the example[3] just mentioned, $K(t)$ would become $\langle A(1, 2; \{t\})\rangle$, with

$$c_i(1) = -[v_i(g) - v_i(e)]/\omega_i,$$
$$d_i(1) = [v_i(g) - v_i(e)]/\omega_i,$$
$$c_i(2) = -[v_i(e) - v_i(g)]/\omega_i,$$
$$d_i(2) = [v_i(e) - v_i(g)]/\omega_i,$$

where $v_i(g)$ and $v_i(e)$ are the distortions in the equilibrium position of the ith lattice mode with the trapped electron in the ground (g) and excited (e) states, respectively. In place of Terwiel's expressions (2b), (3), and (4), we use

$$e^{\theta_i(q)}a_i e^{-\theta_i(q)} = a_i - d_i(q), \quad e^{\theta_i(q)}a_i^\dagger e^{-\theta_i(q)} = a_i^\dagger + c_i(q), \tag{6}$$

and

$$(d/dt_q)\langle A_i(1, p; \{t\})\rangle = f_i^q(1, p; \{t\}). \tag{7}$$

With the aid of (1), (5c), and (6), we first write the commutator expressions

$$[a_i(t), e^{\pm\theta_i(q,t')}] = \pm d_i(q)e^{-i\omega_i(t-t')}e^{\pm\theta_i(q,t')}, \tag{8a}$$
$$[a_i^\dagger(t), e^{\pm\theta_i(q,t')}] = \mp c_i(q)e^{i\omega_i(t-t')}e^{\pm\theta_i(q,t')}, \tag{8b}$$

from which we obtain

$$[a_i(t), A_i(q, p; \{t\})] = \sum_{x=q}^{p} d_i(x)e^{-i\omega_i(t-t_x)}A_i(q, p; \{t\}), \tag{9a}$$

$$[a_i^\dagger(t), A_i(q, p; \{t\})] = -\sum_{x=q}^{p} c_i(x)e^{i\omega_i(t-t_x)}A_i(q, p; \{t\}), \tag{9b}$$

[1] R. H. Terwiel, J. Math. Phys. **8**, 926 (1967).

[2] Defining $\langle\cdots\rangle_i = \mathrm{Tr}\, e^{\beta H_i}\cdots/\mathrm{Tr}\, e^{-\beta H_i}$, $H_i = \omega_i(a_i^\dagger a_i + \frac{1}{2})$, one can substitute $\langle A_i\rangle_i$ for $\langle A_i\rangle$ in (5a). Relation (5a) is applicable to any operator function capable of being factored over the index i.

[3] See, e.g., M. Lax, J. Chem. Phys. **20**, 1752 (1952); J. J. Markham, Rev. Mod. Phys. **31**, 956 (1959).

upon applying (5b). Using (2a), (9a), and the cyclic property of the trace, we then obtain the expression

$$\langle A_i(1, q; \{t\})a_i(t)A_i(q+1, p; \{t\})\rangle$$
$$= \Big[N_i \sum_{x=1}^{q} d_i(x)e^{-i\omega_i(t-t_x)} + (N_i+1)\sum_{x=q+1}^{p} d_i(x)e^{-i\omega_i(t-t_x)}\Big]\langle A_i(1, p; \{t\})\rangle,$$
$$N_i = (e^{\beta\omega_i}-1)^{-1}. \quad (10)$$

An equivalent expression is constructed for $a_i^\dagger(t)$ from (10) by interchanging N_i and N_i+1, replacing $d_i(x)$ by $c_i(x)$, and affixing a positive sign to the time-dependent exponential. Applying the identity

$$e^{\theta_i(q,t_q)} \equiv e^{iHt_q}e^{\theta_i(q)}e^{-iHt_q}$$

to $A_i(1, p; \{t\})$, we differentiate $A_i(1, p; \{t\})$ with respect to t_q to obtain

$$f_i^q(1, p; \{t\}) = i\omega_i\langle A_i(1, q; \{t\})[d_i(q)a_i^\dagger(t_q) - c_i(q)a_i(t_q) - c_i(q)d_i(q)]A_i(q+1, p; \{t\})\rangle \quad (11)$$

with the aid of (8a) and (8b) and (7). Inserting (10) along with the equivalent expression for $a_i^\dagger(t)$ into (11) we arrive at a first-order differential equation:

$$\frac{d}{dt_q}\langle A_i(1, p; \{t\})\rangle$$
$$= i\omega_i\Big\{\sum_{x=1}^{q-1}[d_i(q)c_i(x)(N_i+1)e^{i\omega_i(t_q-t_x)} - c_i(q)d_i(x)N_ie^{-i\omega_i(t_q-t_x)}] + \sum_{x=q+1}^{p}[d_i(q)c_i(x)N_ie^{i\omega_i(t_q-t_x)} - c_i(q)d_i(x)(N_i+1)e^{-i\omega_i(t_q-t_x)}]\Big\}\langle A_i(1, p; \{t\})\rangle,$$

whose solution using (5d) may be expressed as

$$\langle A_i(1, p; \{t\})\rangle = \exp\Big\{\frac{1}{2}\sum_{q=1}^{p}\sum_{q'=1}^{p} c_i(q)d_i(q')[N_ig_i^-(t_q, t_{q'}) + (N_i+1)g_i^+(t_q, t_{q'})]\Big\} \quad (12)$$

in terms of an ordered function

$$g_i^\pm(t_q, t_{q'}) = e^{\pm i\omega_i(t_q-t_{q'})}, \quad q > q',$$
$$= 1, \quad q = q',$$
$$= e^{\pm i\omega_i(t_{q'}-t_q)}, \quad q < q'.$$

If we arbitrarily assign, for example,

$$\sum_i\sum_{q=1}^{p}\theta_i(q) = \lambda\theta \equiv \lambda\sum_i(C_ia_i + D_ia_i^\dagger),$$

then employing (5a) and (12) we get

$$\langle A(1, p; \{0\})\rangle = F(\lambda) = \exp\Big[\tfrac{1}{2}\lambda^2\sum_i C_iD_i\coth\tfrac{1}{2}\beta\omega_i\Big],$$

in agreement with Terwiel. An alternate form for $\langle A_i(1, p; \{t\})\rangle$ may be obtained from (12) by defining a constant

$$M_i(q) = \langle e^{\theta_i(q,t_q)}\rangle = \exp[\tfrac{1}{2}c_i(q)d_i(q)\coth\tfrac{1}{2}\beta\omega_i] \quad (13a)$$

and a time-dependent function

$$G_i(q, q'; t_q - t_{q'}) = \langle e^{\theta_i(q,t_q)}e^{\theta_i(q',t_{q'})}\rangle/M_i(q)M_i(q')$$
$$= \exp\{c_i(q)d_i(q')[N_ie^{i\omega_i(t_q-t_{q'})} + (N_i+1)e^{-i\omega_i(t_q-t_{q'})}]\}. \quad (13b)$$

In terms of (13a) and (13b), expression (12) becomes

$$\langle A_i(1, p; \{t\})\rangle = \prod_{q=1}^{p}\Big[M_i(q)\prod_{x=q+1}^{p}G_i(q, x; t_q - t_x)\Big]. \quad (14)$$

If expressions for averages involving products of A, a, and $a^\dagger$ are desired (such as one encounters in electron-lattice theory upon introducing a canonical transformation to displace the phonon oscillators[4]), we can define

$$c_i(q) = \lambda_q\bar{c}_i(q), \quad d_i(q) = \lambda_q\bar{d}_i(q), \quad (15)$$

such that, after differentiating with respect to λ_q and, using (11), we can construct the relations

$$O(a_i; q, \lambda_q, t_q)e^{\theta_i(q,t_q)} \equiv -\frac{1}{2}\Big[\frac{1}{i\omega_ic_i(q)}\frac{d}{dt_q} - \frac{1}{\bar{c}_i(q)}\frac{d}{d\lambda_q} + d_i(q)\Big]e^{\theta_i(q,t_q)} = e^{\theta_i(q,t_q)}a_i(t_q), \quad (16a)$$

$$O(a_i^\dagger; q, \lambda_q, t_q)e^{\theta_i(q,t_q)} \equiv \frac{1}{2}\Big[\frac{1}{i\omega_id_i(q)}\frac{d}{dt_q} + \frac{1}{\bar{d}_i(q)}\frac{d}{d\lambda_q} + c_i(q)\Big]e^{\theta_i(q,t_q)} = e^{\theta_i(q,t_q)}a_i^\dagger(t_q). \quad (16b)$$

For example, consider the average

$$\langle e^{\theta_i(1,t_1)}a_i(t_1)e^{\theta_i(2,t_2)}a_i^\dagger(t_2)\rangle.$$

Applying (16a) and (16b), we conclude that

$$\langle e^{\theta_i(1,t_1)}a_i(t_1)e^{\theta_i(2,t_2)}a_i^\dagger(t_2)\rangle = O(a_i; 1, \lambda_1, t_1)O(a_i^\dagger; 2, \lambda_2, t_2)\langle A_i(1, 2; \{t\})\rangle,$$

and upon inserting (14) we obtain

$$\langle e^{\theta_i(1,t_1)}a_i(t_1)e^{\theta_i(2,t_2)}a_i^\dagger(t_2)\rangle = O(a_i; 1, \lambda_1, t_1)O(a_i^\dagger; 2, \lambda_2, t_2)M_i(1)M_i(2) \times G_i(1, 2; t_1 - t_2),$$

[4] E. O. Kane, Phys. Rev. 119, 40 (1960).

which can be evaluated by using (13a) and (13b). A large variety of thermal averages can be evaluated in a similar fashion by using combinations of (16a), (16b), and (8a), (8b) in conjunction with (5a), (14), (13a), and (13b).

ACKNOWLEDGMENT

The author wishes to thank Dr. C. A. Coulter, Clark University, Worcester, Massachusetts, for suggestions and comments offered in the preparation of the manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Unitary Representations of the Poincaré Group Contained in a Class of Representations of the Conformal Group

WILLIAM H. KLINK
Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa

(Received 12 February 1968)

A unitary irreducible class of representations of the conformal group is constructed and reduced with respect to the Poincaré group to see which unitary irreducible representations of the Poincaré group it contains. In particular, it is shown that this class of representations of the conformal group does not contain the continuous-spin representations of the Poincaré group. It is concluded that the representations of the conformal group cannot be used to eliminate the continuous-spin representations.

I. INTRODUCTION

In the past several years there has been a renewed interest in the conformal group, not only because of its relationship to Maxwell's equations[1] and the corresponding quantum-mechanical mass-zero wave equations,[2] but also because of its possible use with non-mass-zero particles.[3] The fact that the Poincaré group is a subgroup of the conformal group has raised the question of how representations of the Poincaré group are contained in representations of the conformal group; in general, one has tried to answer this question by using Lie algebra techniques.[4] In this work, purely group-theoretical techniques are used, relying heavily on Mackey's induced representation theory.[5] It is well known that all of the unitary irreducible representations of the Poincaré group can be written as induced representations[6]; it is not as well known that large classes of unitary irreducible representations of semisimple Lie groups (including, in particular, the conformal group) can also be written as induced representations.

The relationship of representations of the Poincaré group to those of the conformal group also leads to the following question: By an appropriate choice of representations of the conformal group, is it possible to eliminate the continuous-spin mass-zero representations of the Poincaré group? In his classic analysis of the Poincaré group Wigner[6] showed that for the mass-zero representations, two types of spin representations arise: those corresponding to a discrete spin which are associated with such particles as the photon and neutrino, and those corresponding to a continuous spin which seem to have no counterpart in nature, because, perhaps, they would have peculiar physical properties.[7]

Now, whenever a representation of a group is decomposed into the representations of a subgroup, not all of the representations of the subgroup appear. The question to be raised here is whether or not the continuous-spin representations are contained in representations of the conformal group, and it is, in fact, shown in Sec. IV that not only the continuous-spin, but also the discrete-spin representations do not appear in a class of irreducible representations of the conformal group.

In order to show this result it is first necessary, in

[1] T. Fulton, F. Rohrlich, and L. Witten, Rev. Mod. Phys. **34**, 442 (1962), and references cited therein.

[2] L. Gross, J. Math. Phys. **5**, 687 (1964); J. Wess, Nuovo Cimento **18**, 1086 (1960).

[3] H. A. Kastrup, Phys. Rev. **150**, 1183 (1966); references to Kastrup's earlier work on the conformal group are given in this reference.

[4] Y. Murai, Progr. Theoret. Phys. (Kyoto) **9**, 147 (1953) and **11**, 441 (1954); A. Kihlberg, V. F. Muller, and F. Halbwachs, Commun. Math. Phys. **3**, 194 (1966); M. L. Graev, Dokl. Akad. Nauk SSSR **98**, 517 (1954); A. Esteve and P. G. Sona, Nuovo Cimento **32**, 473 (1964); L. H. Thomas, Ann. Math. **42**, 113 (1941); J. Fischer and R. Raczka, Commun. Math. Phys. **3**, 233 (1966); **4**, 8 (1967).

[5] G. W. Mackey, *The Theory of Group Representations*, (Department of Mathematics, The University of Chicago, Chicago, Ill., 1955); Am. J. Math. **73**, 576 (1951).

[6] E. P. Wigner, Ann. Math. **40**, 149 (1939); P. Moussa and R. Stora, *Lectures in Theoretical Physics* (University of Colorado Press, Boulder, Colorado, 1964). Vol. VIIa.

[7] E. P. Wigner, *Theoretical Physics* (IAEA, Vienna, 1963), p. 70.

which can be evaluated by using (13a) and (13b). A large variety of thermal averages can be evaluated in a similar fashion by using combinations of (16a), (16b), and (8a), (8b) in conjunction with (5a), (14), (13a), and (13b).

ACKNOWLEDGMENT

The author wishes to thank Dr. C. A. Coulter, Clark University, Worcester, Massachusetts, for suggestions and comments offered in the preparation of the manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Unitary Representations of the Poincaré Group Contained in a Class of Representations of the Conformal Group

WILLIAM H. KLINK
Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa

(Received 12 February 1968)

A unitary irreducible class of representations of the conformal group is constructed and reduced with respect to the Poincaré group to see which unitary irreducible representations of the Poincaré group it contains. In particular, it is shown that this class of representations of the conformal group does not contain the continuous-spin representations of the Poincaré group. It is concluded that the representations of the conformal group cannot be used to eliminate the continuous-spin representations.

I. INTRODUCTION

In the past several years there has been a renewed interest in the conformal group, not only because of its relationship to Maxwell's equations[1] and the corresponding quantum-mechanical mass-zero wave equations,[2] but also because of its possible use with non-mass-zero particles.[3] The fact that the Poincaré group is a subgroup of the conformal group has raised the question of how representations of the Poincaré group are contained in representations of the conformal group; in general, one has tried to answer this question by using Lie algebra techniques.[4] In this work, purely group-theoretical techniques are used, relying heavily on Mackey's induced representation theory.[5] It is well known that all of the unitary irreducible representations of the Poincaré group can be written as induced representations[6]; it is not as well known that large classes of unitary irreducible representations of semisimple Lie groups (including, in particular, the conformal group) can also be written as induced representations.

The relationship of representations of the Poincaré group to those of the conformal group also leads to the following question: By an appropriate choice of representations of the conformal group, is it possible to eliminate the continuous-spin mass-zero representations of the Poincaré group? In his classic analysis of the Poincaré group Wigner[6] showed that for the mass-zero representations, two types of spin representations arise: those corresponding to a discrete spin which are associated with such particles as the photon and neutrino, and those corresponding to a continuous spin which seem to have no counterpart in nature, because, perhaps, they would have peculiar physical properties.[7]

Now, whenever a representation of a group is decomposed into the representations of a subgroup, not all of the representations of the subgroup appear. The question to be raised here is whether or not the continuous-spin representations are contained in representations of the conformal group, and it is, in fact, shown in Sec. IV that not only the continuous-spin, but also the discrete-spin representations do not appear in a class of irreducible representations of the conformal group.

In order to show this result it is first necessary, in

[1] T. Fulton, F. Rohrlich, and L. Witten, Rev. Mod. Phys. **34**, 442 (1962), and references cited therein.

[2] L. Gross, J. Math. Phys. **5**, 687 (1964); J. Wess, Nuovo Cimento **18**, 1086 (1960).

[3] H. A. Kastrup, Phys. Rev. **150**, 1183 (1966); references to Kastrup's earlier work on the conformal group are given in this reference.

[4] Y. Murai, Progr. Theoret. Phys. (Kyoto) **9**, 147 (1953) and **11**, 441 (1954); A. Kihlberg, V. F. Muller, and F. Halbwachs, Commun. Math. Phys. **3**, 194 (1966); M. L. Graev, Dokl. Akad. Nauk SSSR **98**, 517 (1954); A. Esteve and P. G. Sona, Nuovo Cimento **32**, 473 (1964); L. H. Thomas, Ann. Math. **42**, 113 (1941); J. Fischer and R. Raczka, Commun. Math. Phys. **3**, 233 (1966); **4**, 8 (1967).

[5] G. W. Mackey, *The Theory of Group Representations*, (Department of Mathematics, The University of Chicago, Chicago, Ill., 1955); Am. J. Math. **73**, 576 (1951).

[6] E. P. Wigner, Ann. Math. **40**, 149 (1939); P. Moussa and R. Stora, *Lectures in Theoretical Physics* (University of Colorado Press, Boulder, Colorado, 1964). Vol. VIIa.

[7] E. P. Wigner, *Theoretical Physics* (IAEA, Vienna, 1963), p. 70.

Sec. II, to analyze the conformal group, while in Sec. III those elements of Mackey's theory which are needed are presented. Finally, in Sec. IV one class of unitary irreducible representations of the conformal group, the principal nondegenerate series, is analyzed.

II. FINITE-DIMENSIONAL (NONUNITARY) REPRESENTATIONS OF THE CONFORMAL AND POINCARÉ GROUPS

The conformal group $SO(4, 2)$ is usually defined[8] as the set of 6×6 matrices leaving invariant the form

$$\begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & -1 & \\ & & & & & -1 \end{pmatrix}$$

or, what is equivalent,

$$\begin{pmatrix} -1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & -1 \end{pmatrix}$$

(four plus ones, two minus ones). The conformal group is a 15-parameter semisimple Lie group which contains the 10-parameter Poincaré group.

To show that the Poincaré group, defined as

$$X'^{\mu} = L^{\mu}_{\nu} X^{\nu} + T^{\mu}, \quad \mu, \nu = 0, 1, 2, 3, \tag{1}$$

where L^{μ}_{ν} is a Lorentz transformation and T^{μ} a translation, is indeed a subgroup of the conformal group, Murai[8] introduces a 6-dimensional vector η^A, $A = 0, 1, 2, 3, 5, 6$, such that, under a linear transformation of η^A, the form $(\eta^1)^2 + (\eta^2)^2 + (\eta^3)^2 - (\eta^0)^2 + (\eta^5)^2 - (\eta^6)^2$ is left invariant. Murai sets

$$X^{\mu} = \eta^{\mu}\left(\frac{\eta^6 + \eta^5}{\sqrt{2}}\right)^{-1}, \quad \tfrac{1}{2} g_{\mu\nu} X^{\mu} X^{\nu} = \frac{\eta^6 - \eta^5}{\eta^6 + \eta^5}, \tag{2}$$

where

$$g_{\mu\nu} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}, \quad X^2 = g_{\mu\nu} X^{\mu} X^{\nu},$$

and shows that this leads to the following set of transformations [Eqs. (3) and (4) of Ref. 8]:

$$\begin{aligned} \eta'^{\mu} &= L^{\mu}_{\nu}\eta^{\nu} + 2^{-\frac{1}{2}} T^{\mu}(\eta^5 + \eta^6), \\ \eta'^{5} &= \eta^5 - 2^{-\frac{1}{2}} L^{\mu}_{\lambda} T_{\mu} \eta^{\lambda} - \tfrac{1}{4} T^{\mu} T_{\mu} (\eta^5 + \eta^6), \\ \eta'^{6} &= \eta^6 + 2^{-\frac{1}{2}} L^{\mu}_{\lambda} T_{\mu} \eta^{\lambda} + \tfrac{1}{4} T^{\mu} T_{\mu} (\eta^5 + \eta^6). \end{aligned} \tag{3}$$

These transformations are all linear and can be replaced by a matrix. Thus, a 6-dimensional nonunitary matrix representation of the Poincaré group is

$$P = \begin{pmatrix} L^{\mu}_{\nu} & 2^{-\frac{1}{2}} T^{\mu} & 2^{-\frac{1}{2}} T^{\mu} \\ -[2^{-\frac{1}{2}} T_{\mu} L^{\mu}_{\nu}]^{\mathrm{T}} & 1 - \frac{1}{4} T^2 & -\frac{1}{4} T^2 \\ [2^{-\frac{1}{2}} T_{\mu} L^{\mu}_{\nu}]^{\mathrm{T}} & \frac{1}{4} T^2 & 1 + \frac{1}{4} T^2 \end{pmatrix} = \begin{pmatrix} I_4 & 2^{-\frac{1}{2}} T^{\mu} & 2^{-\frac{1}{2}} T^{\mu} \\ -[2^{-\frac{1}{2}} T^{\mu}]^{\mathrm{T}} & 1 - \frac{1}{4} T^2 & -\frac{1}{4} T^2 \\ [2^{-\frac{1}{2}} T^{\mu}]^{\mathrm{T}} & \frac{1}{4} T^2 & 1 + \frac{1}{4} T^2 \end{pmatrix} \begin{pmatrix} L^{\mu}_{\nu} & 0 & 0 \\ 0 & & \\ 0 & & I_2 \end{pmatrix}, \tag{4}$$

where T means transpose and I_j is the j-dimensional identity matrix. It is readily checked that Eq. (4) is a representation of the Poincaré group and that, further, it leaves invariant the form

$$\begin{pmatrix} -1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & -1 \end{pmatrix}$$

It is shown in Sec. III that it is necessary to know the Iwasawa decomposition[9] for the conformal group in order to generate the requisite unitary irreducible representations. The Iwasawa decomposition is written as ANK_m, where A is a noncompact Abelian subgroup, N is a nilpotent subgroup, and K_m is the maximal compact subgroup. A can be written as[10]

$$\begin{pmatrix} \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix} & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & \begin{pmatrix} \cosh\rho & \sinh\rho \\ \sinh\rho & \cosh\rho \end{pmatrix} \end{pmatrix}, \tag{5}$$

where α and ρ are real parameters. The maximal compact subgroup of the conformal group is $SO(4) \otimes SO(2)$ [where $SO(n)$ is the orthogonal group in n dimensions with determinant one]; it is sufficient to know that four of the six parameters of N are the

[8] Y. Murai, Progr. Theoret. Phys. (Kyoto) **11**, 443 (1954).

[9] R. Hermann, *Lie Groups for Physicists* (W. A. Benjamin, Inc., New York, 1966); A. Kihlberg, V. F. Muller, and F. Halbwachs, Commun. Math. Phys. **3**, 194 (1966).

[10] W. H. Klink, J. Math. Phys. **9**, 1669 (1968).

translations of the Poincaré group

$$T = \begin{pmatrix} I_4 & 2^{-\frac{1}{2}}T^\mu & 2^{-\frac{1}{2}}T^\mu \\ -[2^{-\frac{1}{2}}T^\mu]^{\mathrm{T}} & 1-\frac{1}{4}T^2 & -\frac{1}{4}T^2 \\ [2^{-\frac{1}{2}}T^\mu]^{\mathrm{T}} & \frac{1}{4}T^2 & 1+\frac{1}{4}T^2 \end{pmatrix}. \quad (4)$$

In Sec. IV it proves convenient to have another decomposition for the conformal group besides the Iwasawa decomposition. This decomposition corresponds to the one usually given for the Lie algebra[3] of the conformal group, in which there are, besides the six-parameter Lorentz transformations and the four-parameter translations of Eq. (4), another four-parameter translation called C and a scalar transformation S. These transformations have the matrix representation

$$S = \begin{pmatrix} I_4 & 0 & 0 \\ 0 & \cosh\rho & \sinh\rho \\ 0 & \sinh\rho & \cosh\rho \end{pmatrix},$$

$$C = \begin{pmatrix} I_4 & -2^{-\frac{1}{2}}C^\mu & 2^{-\frac{1}{2}}C^\mu \\ [2^{-\frac{1}{2}}C^\mu]^{\mathrm{T}} & 1-\frac{1}{4}C^2 & \frac{1}{4}C^2 \\ [2^{-\frac{1}{2}}C^\mu]^{\mathrm{T}} & -\frac{1}{4}C^2 & 1+\frac{1}{4}C^2 \end{pmatrix}. \quad (6)$$

The decompositions of the conformal group given thus far have been matrix decompositions. One final way of writing the conformal group, which is essential in the following analysis, is as a transformation group, written symbolically as

$$X'^\mu = f^\mu(X^\nu, g); \quad (7)$$

f^μ is a nonlinear function carrying the point X^ν into X'^μ under the action of a group element g.

To get the form of f^μ, we invert Eq. (2) to find η^μ, η^5, η^6 as a function of X^μ, $g_{\mu\nu}X^\mu X^\nu = X^2$. Thus,

$$\eta^\mu = (4/\sqrt{2})[X^\mu\eta^6/(2+X^2)],$$
$$\eta^5 = [(2-X^2)/(2+X^2)]\eta^6. \quad (8)$$

Now, the elements of the conformal group given in Eqs. (4) and (6) carry the 6-dimensional vector $\eta^A = (\eta^\mu, \eta^5, \eta^6)$ linearly into the vector $\eta'^A = (\eta'^\mu, \eta'^5, \eta'^6)$. Denoting any of these elements by the matrix O^A_B, we have

$$\begin{pmatrix} \eta'^\mu \\ \eta'^5 \\ \eta'^6 \end{pmatrix} = \begin{pmatrix} O^\mu_\nu & O^\mu_5 & O^\mu_6 \\ O^5_\nu & O^5_5 & O^5_6 \\ O^6_\nu & O^6_5 & O^6_6 \end{pmatrix} \begin{pmatrix} \eta^\mu \\ \eta^5 \\ \eta^6 \end{pmatrix}. \quad (9)$$

Replacing the η and η' variables by X and X' variables of Eq. (8) gives

$$\begin{pmatrix} (4/2^{\frac{1}{2}})[X'^\mu/(2+X'^2)]\eta'^6 \\ [(2-X'^2)/(2+X'^2)]\eta'^6 \\ \eta'^6 \end{pmatrix} = \begin{pmatrix} O^\mu_\nu & O^\mu_5 & O^\mu_6 \\ O^5_\nu & O^5_5 & O^5_6 \\ O^6_\nu & O^6_5 & O^6_6 \end{pmatrix} \begin{pmatrix} (4/2^{\frac{1}{2}})[X^\mu/(2+X^2)]\eta^6 \\ [(2-X^2)/(2+X^2)]\eta^6 \\ \eta^6 \end{pmatrix}. \quad (10)$$

Finally, after some tedious algebra, the transformation from X^μ to X'^μ can be written as

$$X'^\mu = \frac{O^\mu_\nu X^\nu + \frac{1}{2}(2)^{\frac{1}{2}}(-O^\mu_5 + O^\mu_6)(\frac{1}{2}X^2) + \frac{1}{2}(2)^{\frac{1}{2}}(O^\mu_5 + O^\mu_6)}{2^{-\frac{1}{2}}(O^5_\nu + O^6_\nu)X^\nu + \frac{1}{2}(-O^5_5 - O^6_5 + O^5_6 + O^6_6)(\frac{1}{2}X^2) + \frac{1}{2}(O^5_5 + O^6_5 + O^5_6 + O^6_6)}. \quad (11)$$

Letting $K_\pm = (\pm O^5_5 \pm O^6_5 + O^5_6 + O^6_6)$ gives

$$X'^\mu = \frac{2K_+^{-1}O^\mu_\nu X^\nu + 2^{\frac{1}{2}}K_+^{-1}(-O^\mu_5 + O^\mu_6)\frac{1}{2}X^2 + 2^{\frac{1}{2}}K_+^{-1}(O^\mu_5 + O^\mu_6)}{2^{\frac{1}{2}}K_+^{-1}(O^5_\nu + O^6_\nu)X^\nu + K_-K_+^{-1}\frac{1}{2}X^2 + 1}, \quad (12)$$

and this is chosen as the canonical form of the transformation f^μ.

As a check on the form of f^μ, when O^A_B is restricted to elements of the Poincaré group, Eq. (12) must reduce to Eq. (1). To see that this indeed happens, set $O^\mu_\nu = L^\mu_\nu$, $O^\mu_5 = O^\mu_6 = 2^{-\frac{1}{2}}T^\mu$, $-O^5_\nu = O^6_\nu = 2^{-\frac{1}{2}}T_\nu$, and $K_\pm = \pm(1-\frac{1}{4}T^2) \pm \frac{1}{4}T^2 - \frac{1}{4}T^2 + \frac{1}{4}T^2$ so that $K_+ = 2$, $K_- = 0$. Then substituting into Eq. (12) gives

$$X'^\mu = \frac{2(2)^{-1}L^\mu_\nu X^\nu + \frac{1}{2}(2)^{\frac{1}{2}}(-2^{-\frac{1}{2}}T^\mu + 2^{-\frac{1}{2}}T^\mu)(\frac{1}{2}X^2) + \frac{1}{2}(2)^{\frac{1}{2}}(2^{-\frac{1}{2}}T^\mu + 2^{-\frac{1}{2}}T^\mu)}{\frac{1}{2}(2)^{\frac{1}{2}}(-2^{-\frac{1}{2}}T_\nu + 2^{-\frac{1}{2}}T_\nu)X^\nu + 1}$$
$$= L^\mu_\nu X^\nu + T^\mu, \quad (13)$$

in agreement with Eq. (1).

To get the form of f^μ corresponding to the scalar element in Eq. (6), set $O^\mu_\nu = \delta^\mu_\nu$, $O^\mu_5 = O^\mu_6 = 0$, $O^5_\nu = O^6_\nu = 0$, $K_+ = 2(\cosh\rho + \sinh\rho) = 2e^\rho$, and $K_- = 0$. Then,

$$X'^\mu = X^\mu/e\rho = e^{-\rho}X^\mu, \quad (14)$$

which is the usual form of the dilation transformation as given, for example, in Murai.[8]

Finally, there is the element C of Eq. (6); for this element $O^\mu_\nu = \delta^\mu_\nu$, $-O^\mu_5 = O^\mu_6 = 2^{-\frac{1}{2}}C^\mu$, $O^5_\nu = O^6_\nu = 2^{-\frac{1}{2}}C_\nu$, $K_+ = 2$, and $K_- = C^\mu C_\mu = C^2$, so that

$$X'^\mu = \frac{X^\mu + \frac{1}{2}(2)^{\frac{1}{2}}(2(2)^{-\frac{1}{2}}C^\mu)(\frac{1}{2}X^2)}{C_\nu X^\nu + \frac{1}{2}C^2(\frac{1}{2}X^2) + 1} = \frac{X^\mu + C^\mu(\frac{1}{2}X^2)}{C_\nu X^\nu + \frac{1}{2}C^2(\frac{1}{2}X^2) + 1}. \tag{15}$$

The set of transformations (15) are called special conformal transformations,[11] and it is not hard to check that combining the transformations (15) with those of the Lorentz transformations gives a nonlinear-transformation group representation of the Poincaré group, as is already easily ascertained by looking at the commutation relations of the Lie algebra corresponding to the group elements C and L^μ_ν.[3]

To get all the elements in the canonical form (12), it is merely necessary to combine (13) and (14) to get

$$X'^\mu = e^{-\rho}L^\mu_\nu X^\nu + T^\mu, \tag{16}$$

which, with (15), gives

$$X'^\mu = \frac{(e^{-\rho}L^\mu_\nu + T^\mu C_\nu)X^\nu + (e^{-\rho}L^\mu_\nu C^\nu + \frac{1}{2}T^\mu C^2)(\frac{1}{2}X^2) + T^\mu}{C_\nu X^\nu + \frac{1}{2}C^2(\frac{1}{2}X^2) + 1} \tag{17}$$

as the form of the conformal group which will be used throughout this paper.

III. THE DECOMPOSITION OF INDUCED REPRESENTATIONS

The classes of unitary irreducible representations being considered in this paper are all induced representations. It has been shown[6] that all unitary irreducible representations of the Poincaré group can be written as induced representations, while large classes of unitary irreducible representations of the conformal group can be written as induced representations.[12]

Therefore, this section deals with the notion of induced representations and how it is possible to decompose a reducible induced representation into a direct integral of irreducible induced representations. All of the ideas sketched here can be found in Mackey's lecture notes.[13]

Let G be a given noncompact group (in this paper, G is the conformal group) and H_1 a subgroup of G (H_1 is given in Sec. IV and varies with the classes of representations of G being considered.) Let $\mathcal{K}_1$ be an irreducible representation of H_1 acting on the vector space $\mathfrak{V}(\mathcal{K}_1)$. Consider functions f which map elements g of G[14] into $\mathfrak{V}(\mathcal{K}_1)$:

$$g \xrightarrow{f} \mathfrak{V}(\mathcal{K}_1), \quad f(g) \in \mathfrak{V}(\mathcal{K}_1), \tag{18}$$

such that f satisfies the condition $f(h_1 g) = \mathcal{K}_1(h_1)f(g)$. This set of functions forms a new vector space $\hat{\mathfrak{V}}(\mathcal{K}_1)$:

$$\mathfrak{V}(\mathcal{K}_1) = \{f \,|\, f(g) \in \mathfrak{V}(\mathcal{K}_1), \quad f(h_1 g) = \mathcal{K}_1(h_1)f(g) \text{ for all } h_1 \in H_1, \quad g \in G\}. \tag{19}$$

The representation $U^{\mathcal{K}_1}(g')f(g) = \rho(g', g)f(gg')$ on the vector space $\hat{\mathfrak{V}}(\mathcal{K}_1)$ is called the induced representation of G. The function $\rho(g', g)$ is chosen in such a way as to make $U^{\mathcal{K}_1}(g')$ a unitary representation. However, in order to make $U^{\mathcal{K}_1}(g')$ a unitary representation on $\hat{\mathfrak{V}}(\mathcal{K}_1)$, it is first necessary to make $\hat{\mathfrak{V}}(\mathcal{K}_1)$ into a Hilbert space.

$\hat{\mathfrak{V}}(\mathcal{K}_1)$ is readily made into a Hilbert space for the class of representations being considered in this paper, namely, the principal nondegenerate series of representations. For this class of representations, an inner product is defined as

$$(f(g), f'(g)) = \int_{G/H_1} f^*(g)f'(g)\, d\mu, \tag{20}$$

where * means complex conjugation and $d\mu$ is the measure associated with G/H_1, which is inherited from the Haar measure of G.[15]

After H_1 has been chosen so as to induce a unitary irreducible representation $U^{\mathcal{K}_1}(G)$ of G, we are interested in seeing how $U^{\mathcal{K}_1}(G)$ decomposes into unitary irreducible representations of a subgroup H_2 of G (in this paper H_2 is the Poincaré group).

Mackey has shown[13] how to decompose $U^{\mathcal{K}_1}(H_2)$ into direct integrals over double cosets

$$G = \bigcup_D H_1 g_D H_2,$$

where $\{g_D\}$ are elements of G not in H_1 or H_2 so that

$$U^{\mathcal{K}_1}(H_2) \cong \int d\mu(g_D) U^{\mathcal{J}_D}(H_2); \tag{21}$$

[11] H. A. Kastrup, Phys. Rev. **150**, 1183 (1966); J. Rosen, "On Conformal Groups and Transformations of Trajectories under the Space–time Conformal Group," Brown University preprint, NYO-2262TA-151, 1966.

[12] E. M. Stein, *High Energy Physics and Elementary Particles* (IAEA, Vienna, 1965); see also M. L. Graev, Ref. 4.

[13] G. W. Mackey, Ref. 5, pp. 135ff.

[14] Lower-case letters denote individual elements of a group while capital letters denote sets of elements. Script letters denote either representations of the groups or the spaces on which these representations operate.

[15] See Ref. 13, p. 119, for the proper definition of the measure $d\mu$ and the function $\rho(g', g)$.

here $d\mu(g_D)$ is the double coset measure, $\cong$ means "is equivalent to," and $\mathfrak{d}_D$ is the representation $\mathfrak{K}_1(g_D J_D g_D^{-1})$ of the group $J_D = g_D^{-1} H_1 g_D \cap H_2$. In general, $U^{\mathfrak{d}_D}(H_2)$ is not an irreducible representation of H_2, so that it is necessary to further decompose $U^{\mathfrak{d}_D}(H_2)$ into irreducible induced representations of H_2.

The calculation of J_D is fairly difficult with the definition of J_D given above. For that reason, we actually compute J_D in the following way. Combine $H_1 g_D$ with J_D so that

$$H_1 g_D J_D = H_1 g_D g_D^{-1} H_1 g_D \cap H_1 g_D H_2,$$

which gives $H_1 g_D J_D = H_1 g_D \cap H_1 g_D H_2$. Now, in general, $H_1 g_D h_2$, $h_2 \in H_2$ sends g_D into a new double coset $H_1 g_{D'}$. If $D' \neq D$, then $H_1 g_D \cap H_1 g_{D'}$ is empty, so that it is necessary to find those h_2 such that $H_1 g_D h_2 = H_1 g_D$. The set of all h_2 leaving the double coset g_D invariant is precisely J_D[16]; that is,

$$J_D = \{h_2 \mid h_2 \in H_2,\ H_1 g_D h_2 = H_1 g_D\}. \tag{22}$$

It is sufficient to actually calculate $g_D h_2 = h_1 g_D$. Thus, the procedure in Sec. IV is to find H_1, decompose G into right cosets g_c with respect to H_1, such that

$$G = \bigcup_c H_1 g_c, \tag{23}$$

and then compute $g_c h_2 = h_1 g_{c'}$. The double cosets are a subset of the right cosets and, further, those double cosets which are left invariant under the action of h_2 generate the subgroup J_D.

Finally, it is necessary to decompose $U^{\mathfrak{d}_D}(H_2)$ of (21) into irreducible representations of H_2. Since we are concerned with whether the continuous-spin representations of the Poincaré group are actually contained in $U^{\mathfrak{d}_D}(H_2)$, it is sufficient to use the Frobenius reciprocity theorem,[17] which says that if $\mathfrak{d}_D$ is an irreducible representation of J_D, then $U^{M,s}(H_2)$ is contained as many times in $U^{\mathfrak{d}_D}(H_2)$ as $U^{M,s}(J_D)$ contains $\mathfrak{d}_D$. Here M and s are the unitary irreducible representation labels of the Poincaré group.

IV. ANALYSIS OF THE PRINCIPAL NONDEGENERATE SERIES OF REPRESENTATIONS OF THE CONFORMAL GROUP

For the principal nondegenerate series of representations of the conformal group, $H_1 = ANC(A)$,[18] where

$$C(A) = \{k \mid k \in K_m,\ ak = ka \text{ for all } a \in A\}. \tag{24}$$

The form of A is given in Eq. (5); in general, α is not equal to ρ, so that those elements of K_m which commute with A can be written as

$$C(A) = \begin{pmatrix} I_2 & 0 & 0 \\ 0 & SO(2) & 0 \\ 0 & 0 & I_2 \end{pmatrix} = \begin{pmatrix} I_2 & 0 & 0 \\ 0 & \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} & 0 \\ 0 & 0 & I_2 \end{pmatrix}. \tag{25}$$

Four of the six elements of N are the translations of the Poincaré group, Eq. (4); the other two elements are the nilpotent elements of the Lorentz group. To write these elements explicitly, it is convenient to consider the covering group $SL(2, c)$ of the Lorentz group.[19] Then the α parameter in A [Eq. (5)] and the θ parameter of $C(A)$ can be written together as $\left(\begin{smallmatrix} \alpha_1 & 0 \\ 0 & \alpha_1^{-1} \end{smallmatrix}\right)$, α_1 complex, with $\cosh \alpha = \frac{1}{2}(|\alpha_1|^2 + |\alpha_1|^{-2})$ and $\arg \alpha_1 = \frac{1}{2}\theta$. Further, the two remaining nilpotent elements of N can be written in $SL(2, c)$ as $\left(\begin{smallmatrix} 1 & \beta_1/\alpha_1 \\ 0 & 1 \end{smallmatrix}\right)$, β_1 complex. Combining these four elements gives

$$\bar{L}_1 \equiv \begin{pmatrix} \alpha_1 & \beta_1 \\ 0 & \alpha_1^{-1} \end{pmatrix}. \tag{26}$$

Call L_1 those Lorentz transformations corresponding to the elements $\bar{L}_1$; also label all elements of H_1 with the subscript 1. Then it is possible to write H_1 as a transformation group, so that

$$H_1 : X'^{\mu} = e^{-\rho_1} L_{1\nu}^{\mu} X^{\nu} + T_1^{\mu}. \tag{27}$$

H_1 is a 9-parameter group, corresponding to the four elements T_1^{μ}, the four elements L_1 as determined by $\bar{L}_1$ of Eq. (26), and the scalar $e^{-\rho_1}$.

Once H_1 is chosen, a convenient choice of right-coset labels is

$$\{g_c\} : X'^{\mu} = \frac{L_{c\nu}^{\mu} X^{\nu} + C^{\mu}(\frac{1}{2}X^2)}{C_{\nu} X^{\nu} + \frac{1}{2}C^2(\frac{1}{2}X^2) + 1}, \tag{28}$$

where L_c is the Lorentz transformation corresponding to the covering group element

$$\bar{L}_c = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}, \quad c \text{ complex}. \tag{29}$$

The decomposition of $SL(2, c)$ into $\bar{L}_1$ and $\bar{L}_c$ corresponds to the decomposition given in Ref. 19 and is, in fact, used to induce a class of representations of the group $SL(2, c)$.

It can now be shown that in combining H_1, Eq. (27), with $\{g_c\}$, Eq. (28) uniquely covers every element

[16] This result is carried out in the appendix using homogeneous spaces. See I. M. Gel'fand and M. A. Naimark, *Unitäre Darstellungen der Klassischen Gruppen* (Akademie-Verlag, Berlin, 1957).

[17] See Ref. 13, p. 129.

[18] See Ref. 12 and references cited therein.

[19] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon Press, Inc., New York, 1964), p. 120.

of the conformal group, such that

$$g = h_1 g_c : X''^{\mu} = f^{\mu}(X'^{\alpha}, g) = e^{-\rho_1} L^{\mu}_{1\nu} X'^{\nu} + T^{\mu}_1,$$
$$X'^{\nu} = \frac{L^{\nu}_{c\alpha} X^{\alpha} + C^{\nu}(\frac{1}{2}X^2)}{C_{\beta}X^{\beta} + \frac{1}{2}C^2(\frac{1}{2}X^2) + 1}, \tag{30}$$
$$X''^{\mu} = \frac{e^{-\rho_1} L^{\mu}_{1\nu}[L^{\nu}_{c\alpha} X^{\alpha} + C^{\nu}(\frac{1}{2}X^2)]}{C_{\alpha}X^{\alpha} + \frac{1}{2}C^2(\frac{1}{2}X^2) + 1} + T^{\mu}_1,$$

or

$$X''^{\mu} = \frac{(e^{-\rho_1} L^{\mu}_{1\nu} L^{\nu}_{c\alpha} + T^{\mu}_1 C_{\alpha}) X^{\alpha} + [e^{-\rho_1} L^{\mu}_{1\nu} L^{\nu}_{c\alpha} C^{\alpha} + T^{\mu}_1(\frac{1}{2}C^2)]\frac{1}{2}X^2 + T^{\mu}_1}{C_{\alpha}X^{\alpha} + \frac{1}{2}C^2(\frac{1}{2}X^2) + 1}. \tag{31}$$

Comparing this with the canonical form of the conformal group (17), we see that

$$T^{\mu}_1 = T^{\mu},$$
$$e^{-\rho_1} L^{\mu}_{1\nu} L^{\nu}_{c\alpha} + T^{\mu}_1 C_{\alpha} = e^{-\rho} L^{\mu}_{\alpha} + T^{\mu} C_{\alpha}, \tag{32}$$

so that

$$e^{-\rho} L^{\mu}_{\alpha} = e^{-\rho_1} L^{\mu}_{1\nu} L^{\nu}_{c\alpha} + T^{\mu}_1 C_{\alpha} - T^{\mu} C_{\alpha}$$
$$= e^{-\rho_1} L^{\mu}_{1\nu} L^{\nu}_{c\alpha}, \tag{33}$$
$$\rho_1 = \rho,$$
$$L^{\mu}_{1\nu} L^{\nu}_{c\alpha} = L^{\mu}_{\alpha}.$$

But this last expression corresponds precisely to the decomposition of $SL(2, c) = \bar{L}_1 \bar{L}_c$ of Eqs. (26) and (29). Thus, Eqs. (32) and (33) uniquely specify any group element of the conformal group.

Before calculating the double cosets g_D, it is necessary to know how $h_2 \in H_2$ sends a right-coset element into another coset element; thus, we must calculate $g_c h_2 = h_1 g_{c'}$. But Eqs. (32) and (33) show how to take any element of the conformal group and see what right coset it is in. Thus, it is only necessary to calculate $g_c h_2 = g$.

Now, Eq. (1) is the transformation group for H_2, the Poincaré group. g_c is given by Eq. (28), so that we have

$$g = g_c h_2 : X''^{\mu} = \frac{L^{\mu}_{c\nu} X'^{\nu} + C^{\mu}(\frac{1}{2}X'^2)}{C_{\nu}X'^{\nu} + \frac{1}{2}C^2(\frac{1}{2}X'^2) + 1},$$
$$X'^{\nu} = L^{\nu}_{2\alpha} X^{\alpha} + T^{\nu}_2, \tag{34}$$

where the $L^{\nu}_{2\alpha}$, T_2 are the 10 parameters of H_2. Then,

$$X''^{\mu} = \frac{L^{\mu}_{c\nu} L^{\nu}_{2\alpha} X^{\alpha} + L^{\mu}_{c\nu} T^{\nu}_2 + \frac{1}{2}C^{\mu}(L^{\alpha}_{2\beta} X^{\beta} + T^{\alpha}_2)^2}{C_{\nu}(L^{\nu}_{2\alpha} X^{\alpha} + T^{\nu}) + \frac{1}{2}C^2[\frac{1}{2}(L^{\alpha}_{2\beta} X^{\beta} + T^{\alpha}_2)^2] + 1}$$
$$= \frac{(L^{\mu}_{c\beta}\delta^{\beta}_{\nu} + C^{\beta} T_{2_{\nu}}) L^{\nu}_{2\alpha} X^{\alpha} + C^{\mu}(\frac{1}{2}X^2) + [L^{\mu}_{c\nu} T^{\nu}_2 + C^{\mu}(\frac{1}{2}T^2_2)]}{[C_{\nu} + (\frac{1}{2}C^2) T_{2_{\nu}}] L^{\nu}_{2\alpha} X^{\alpha} + \frac{1}{2}C^2(\frac{1}{2}X^2) + D(C^{\mu}, T^{\mu}_2)}, \tag{35}$$

where $D(C^{\mu}, T^{\mu}_2) = C_{\mu} T^{\mu}_2 + \frac{1}{2}C^2(\frac{1}{2}T^2_2) + 1$.

To get Eq. (35) in canonical form, it is necessary to divide the numerator and denominator by $D(C^{\mu}, T^{\nu}_2)$. Thus,

$$X''^{\mu} = \frac{\frac{1}{D}(L^{\mu}_{c\beta}\delta^{\beta}_{\nu} + C^{\beta} T_{2_{\nu}}) L^{\nu}_{2\alpha} X^{\alpha} + \frac{1}{D} C^{\mu} \frac{X^2}{2} + \frac{1}{D}\left[L^{\mu}_{c\nu} T^{\nu}_2 + C^{\mu} \frac{T^2_2}{2}\right]}{\frac{1}{D}\left(C_{\nu} + \frac{C^2}{2} T_{2_{\nu}}\right) L^{\nu}_{2\alpha} X^{\alpha} + \frac{1}{D}\frac{C^2}{2}\frac{X^2}{2} + 1}. \tag{36}$$

Comparing Eq. (36) with Eq. (17), we see that

$$\frac{1}{D}\left(C_{\nu} + \frac{C^2}{2} T_{2_{\nu}}\right) L^{\nu}_{2\alpha}$$

corresponds to C_{α},

$$\frac{1}{D}\left[L^{\mu}_{c\nu} T^{\nu}_2 + C^{\mu} \frac{T^2_2}{2}\right]$$

corresponds to T^{μ}, and $e^{-\rho} L^{\mu}_{\nu} + T^{\mu} C_{\nu}$ corresponds to

$$\frac{1}{D}[L^{\mu}_{c\beta}\delta^{\beta}_{\alpha} + C^{\beta} T_{2_{\alpha}}] L^{\alpha}_{2\nu}.$$

Therefore, using Eqs. (32) and (33), we get that the equation $g_c h_2 = h_1 g_{c'}$ can be written as

$$C_{\alpha} \xrightarrow{h_2} C'_{\alpha} = \frac{1}{D(C^{\mu}, T^{\mu}_2)}\left[C_{\nu} + \frac{C^2}{2} T_{2_{\nu}}\right] L^{\nu}_{2\alpha} \tag{37}$$

for the transformation of C_{α} into C'_{α}. The transformation of L_c into $L_{c'}$ under the action of H_2 is also easily obtained, but is quite complicated and, furthermore, is not needed in the rest of the analysis. It is sufficient to note that, if $T_{2_{\nu}}$ is zero, then

$$L^{\mu}_{c_{\nu}} \xrightarrow{L_2} L^{\mu}_{c_{\nu'}}; \quad L^{\mu}_{c_{\alpha}} L^{\alpha}_{2_{\nu}} = L^{\mu}_{1_{\beta}} L^{\beta}_{c_{\nu'}}. \tag{38}$$

But Eq. (38) can be written in the covering group representation as

$$L_c \xrightarrow{L_2} L_{c'}; \quad L_c L_2 = L_1 L_{c'}. \tag{39}$$

In matrix notation, this becomes

$$\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}\begin{pmatrix} \alpha_2 & \beta_2 \\ \gamma_2 & \delta_2 \end{pmatrix} = \begin{pmatrix} \alpha_1 & \beta_1 \\ 0 & \alpha_1^{-1} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ c' & 1 \end{pmatrix}, \tag{40}$$

so that

$$\begin{pmatrix} \alpha_2 & \beta_2 \\ c\alpha_2 + \gamma_2 & c\beta_2 + \delta_2 \end{pmatrix} = \begin{pmatrix} \alpha_1 + \beta_1 c' & \beta_1 \\ c'\alpha_1^{-1} & \alpha_1^{-1} \end{pmatrix}. \tag{41}$$

Therefore, under the action of L_2, $c\alpha_2 + \gamma_2 = c'\alpha_1^{-1}$, $c\beta_2 + \delta_2 = \alpha_1^{-1}$ so that

$$c \xrightarrow{L_2} c' = \frac{c\alpha_2 + \gamma_2}{c\beta_2 + \delta_2}. \tag{42}$$

Now, before choosing the double cosets, let us see if any of the right cosets are left invariant under the action of T_r^μ. With $L^\nu_{2\alpha} = \delta^\nu_\alpha$, Eq. (37) shows that, in order for C_α to equal C_α, it is necessary that

$$1/[D(C^\mu, T_2^\mu)] = 1 \quad \text{and} \quad (\tfrac{1}{2}C^2)T_{2_\nu} = 0.$$

But calculating $C'_\alpha C^{\alpha'} = C'^2$ in Eq. (37) gives

$$C'^2 = C^2/[D(C^\mu, T_2^\mu)], \tag{43}$$

so that setting C^2 equal to zero means C'^2 also is zero. Thus, if $C^2 = 0$, there is no transformation h_2 which carries C^μ to C'^μ such that C'^μ has a length different from zero. Therefore, the equation $(\frac{1}{2}C^2)T_{2_\nu} = 0$ can be satisfied only if $T_{2_\nu} = 0$, so that $T_{2_\nu} = 0$ is the only translation vector which leaves C^μ invariant and, thus, there are no elements of T_{2_ν} in the subgroup J_D [Eq. (22)].

The same result does not hold for L_2, however. If T_{2_ν} is set equal to zero, $D(C^\mu, T_2^\mu) = 1$ and Eq. (37) shows that $C'_\alpha = C_\nu L^\nu_{2\alpha}$. If C_ν is chosen as

$$C_\nu = (1, 0, 0, 0), \tag{44}$$

then the rotation subgroup $SO(3)$ of L_2 leaves C_ν invariant.

The question, then, is whether the rotation subgroup of L_2 also leaves L_c invariant. Now, the covering group of $SO(3)$ is $SU(2)$,[19] so that we must check whether $SU(2) = \left(\begin{smallmatrix} \alpha_2 & \beta_2 \\ -\beta_2^* & \alpha_2^* \end{smallmatrix}\right)$, $|\alpha_2|^2 + |\beta_2|^2 = 1$ leaves c of Eq. (42) invariant:

$$c \xrightarrow{SU(2)} c' = \frac{c\alpha_2 - \beta_2^*}{c\beta_2 + \alpha_2^*}. \tag{45}$$

Clearly, only the identity element $\alpha_2 = 1$, $\beta_2 = 0$ leaves c invariant. However, if $c = 0$, Eq. (45) becomes

$$c = 0 \xrightarrow{SU(2)} c' = -\beta_2^*/\alpha_2^*, \tag{46}$$

so that, if $\beta_2 = 0$, $c = 0$ is left invariant. But $\beta_2 = 0$ corresponds to the subgroup $SO(2)$ of $SO(3)$, so we have just found that the subgroup $SO(2)$ of L_2 leaves the cosets $C_\nu = (1, 0, 0, 0)$ and $c = 0$ invariant.

We thus choose as our double coset g_D precisely this element[20]

$$g_D: L_c = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad C_\nu = (1, 0, 0, 0). \tag{47}$$

It only remains to be seen if $g_D H_2$ carries g_D into every right coset $g_{c'}$. Equation (42) shows that if $c = 0$, $c' = \gamma_2/\delta_2$, so that c' can take on any complex value, since γ_2 is arbitrary. Further, the pure Lorentz transformations carry $C_\nu = (1, 0, 0, 0)$ into an arbitrary C'_α such that $C'^2 = 1$. To get an arbitrary length for C'_α, note that Eq. (43) shows that, even if $C^2 = 1$, $C'^2 = 1/[D(C^\mu, T_2^\mu)]$ and, since T_2^μ is arbitrary, it is possible to make C'^2 any length desired.

Thus, it has been shown that, with g_D given in Eq. (47), the subgroup J_D is $SO(2)$. We now use this fact to decompose $U^{\mathcal{H}_1}(H_2)$. In order to decompose $U^{\mathcal{H}_1}(H_2)$, it is necessary to give the representation $\mathcal{H}_1$ of $H_1 = ANC(A)$. Since A and $C(A)$ form an Abelian subgroup and N is nilpotent, $\mathcal{H}_1$ is

$$h_1 \to \mathcal{H}_1 = |\alpha_1|^{i\sigma_1}\left(\frac{\alpha_1}{|\alpha_1|}\right)^m |\rho_1|^{i\sigma_2}, \tag{48}$$

where α_1 is defined in Eq. (26) and ρ_1 corresponds to the scalar element of Eq. (27). Thus, the principal nondegenerate series of representations of the conformal group require three labels: two continuous ones, σ_1 and σ_2; and one, m, an integer or half-integer. From Eq. (48) and (21), it follows that

$$\mathcal{J}_D = (\alpha_1/|\alpha_1|)^m = e^{im\theta} \tag{49}$$

which is an irreducible representation of $J_D = SO(2)$.

Thus, since there is only one double coset, the Mackey subgroup theorem, Eq. (21), says that

$$U^{\mathcal{H}_1}(H_2) \equiv U^{\sigma_1 m \sigma_2}(H_2) \cong U^{\mathcal{J}_D}(H_2) \equiv U^m(H_2). \tag{50}$$

We wish to see whether any of the mass-zero representations of the Poincaré group occur in the reducible representation $U^m(H_2)$, induced by the subgroup $J_D = SO(2)$. Now Rideau[21] has carried out the decomposition of the regular representation of the Poincaré group and finds that the mass-zero representations do not occur in this decomposition. But the regular representation can be written as an induced representation, induced by the identity

[20] There are other double cosets, but they are not relevant in our analysis. These double cosets have been tabulated by E. Thieleker (private communication).

[21] G. Rideau, Commun. Math. Phys. 3, 218 (1966).

subgroup of the Poincaré group. Further, by using Mackey's notion of "inducing by stages," [22] we can write the regular representation of the Poincaré group equivalently as being induced by the regular representation of $SO(2)$. Now by the Frobenius reciprocity theorem,[17] each representation of $SO(2)$, labeled by $m = 0, \pm 1, \pm 2, \cdots$, occurs once in the regular representation of $SO(2)$. Therefore, since the mass-zero representations do not occur in the regular representation of the Poincaré group, they do not occur in any of the representations $U^m(H_2)$ induced by the subgroup $SO(2)$.

V. CONCLUSION

At first sight, it seems rather puzzling that the conformal group, the group leaving the sourceless Maxwell equations invariant, should not single out the mass-zero representations of the Poincaré group. Certainly in the case of the principal nondegenerate series the mass-zero representations do not even appear.

What seems to be happening is that different classes of representations of the conformal group single out either mass-zero *or* non-mass-zero representations of the Poincaré group. Thus, considering those five elements of the conformal group not appearing in the Poincaré group, namely the dilations [Eq. (14)] and the four vector accelerations [Eq. (15)], it is clear that these five elements treat the mass-zero and non-mass-zero representations very differently. Since mass-zero systems travel at the speed of light, the acceleration operators obviously transform them differently from non-mass-zero systems. Also, the dilation operator can change the mass scale for a non-mass-zero system, whereas it cannot for a mass-zero system.

Further, it is to be noted that the inducing subgroup of the mass-zero representation, consisting of the translations T_μ [Eq. (1)], $SO(2)$ [Eq. (25)], and the nilpotent element β_1 of Eq. (26)—all are contained in N and $C(A)$ [see Eq. (27)]. This suggests that the representations of the subgroup inducing the mass-zero representations of the Poincaré group also generate a class of representations of the conformal group. This class of representations contains only mass-zero representations of the Poincaré group and no non-mass-zero representations.[23]

The statements of the last two paragraphs have not been proven; nevertheless, it seems safe to conclude that it is impossible to split the continuous-spin representations off from the other "physical" Poincaré group representations (including the finite-spin mass-zero representations) by choosing appropriate representations of the conformal group.

ACKNOWLEDGMENTS

The author wishes to thank Dr. E. Thieleker for helpful discussions and the referee for pointing out the incorrect conclusions originally drawn from Eq. (50).

APPENDIX A

The results of Eq. (22) can be shown most clearly by using the notion of homogeneous spaces.[9] Consider a group G, and a subgroup H_1, and call $\mathcal{M}$ the homogeneous space G/H_1. Then a point m can be identified with a right coset H_1g. The stabilizer point m_0 of $\mathcal{M}$ corresponds to the identity coset H_1.

Now consider the action of elements g of G on points m of $\mathcal{M}$: $mg = m' \in \mathcal{M}$. In particular, $m_0H_1 = m_0$. To show that J_D consists of those elements of H_2 which leave H_1g_D invariant, define $m_D = m_0g_D$; then,

$$\begin{aligned} m_D g_D^{-1} H_1 g_D &= m_0 H_1 g_D \\ &= m_0 g_D \\ &= m_D, \end{aligned} \tag{51}$$

so that $g_D^{-1}H_1g_D$ leaves m_D invariant. Therefore, the intersection of $g_D^{-1}H_1g_D$ with H_2 (which defines J_D) acting on m_D gives

$$m_D[g_D^{-1}H_1g_D \cap H_2] = m_D \cap m_DH_2, \tag{52}$$

and in order for this not to be empty, m_Dh_2 must equal m_D for some $h_2 \in H_2$. The set of these $h_2 \in H_2$ then gives J_D.

[22] See Ref. 13, p. 121.

[23] *Note Added in Proof:* A class of representations of the conformal group containing only mass-zero representations of the Poincaré group is given by G. Mack and I. Todorov, "*Irreducibility of the Ladder Representations of U(2, 2) when restricted to its Poincaré Subgroup*" IAEA Preprint IC-68-86, Trieste, Italy, 1968.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Powers of the D Functions

JAGANNATHAN GOMATAM AND F. ROHRLICH
Department of Physics, Syracuse University, Syracuse, New York

(Received 18 October 1968)

While powers of distributions in general do not exist, $[D_+]^n$ and $[D_-]^n$ are the exceptions among all D functions; they do exist and are given explicitly. A "modified" power $[D_\Gamma]_n$ based on the representation of D_Γ as analytic functionals can be defined. It exists for *all* homogeneous D_Γ, i.e., for $D_\pm$, D, and D_1. The relation of $[D_\pm]_n$ to $[D_\pm]^n$ is given, and explicit expressions are found in x space and in p space. The Källén–Lehmann–Umezawa–Kamefuchi representation of these distributions is derived. The extent to which these considerations are applicable to the $\Delta_\Gamma(x, m)$, $m > 0$, is discussed.

1. INTRODUCTION

It is well known that the ordinary product of two distributions at the same argument is, in general, not defined. The convolution product exists, but only under rather limiting conditions. The great physical importance of these distributions, usually called "Green's functions" in field theory, requires one to pay special attention to the few exceptional cases where the ordinary product involving these functions *is* defined. In particular, the distributions $[\Delta_\pm]^n(x, m)$ are known to exist[1] for all positive integers n. In the present paper we consider, among other distributions, the $m = 0$ case of the above, i.e., $[D_\pm]^n(x)$.

However, the fact that these distributions can be expressed as limits of analytic functionals[2] permits one to define the power of such a distribution as the limit of the power of that analytic functional. We then have to distinguish between the power so defined (modified power) and the usual definition of power (direct power); this is done in Sec. 2. The modified power exists for a larger class of distributions than the direct power. It is so defined that it is equal to the direct power in those cases where both exist.

While we are directing our attention here to modified powers of the distributions $D_\Gamma(x)$, i.e., to x space, it is important to recall that modified powers of distributions in p space have been defined earlier.[2] The similarity in form of the distributions $D_\Gamma(x)$ and $\tilde{D}_\Gamma(p)$, as is evident from the appendix of the work by Rohrlich and Strocchi,[3] permits one to relate the two. Such relations exist both in x space and in p space (Sec. 3). Finally, the spectral representations of the various powers are given in Sec. 4. The last section generalizes these considerations to the $\Delta_\Gamma(x, m)$ with $m > 0$.

2. DIRECT AND MODIFIED POWERS

The usual definition of $D_\Gamma(x)$ in terms of its Fourier-integral representation is given in Ref. 4, whose notation we follow. An alternative but completely equivalent definition is given in Ref. 3, namely,

$$\eta_\Gamma D_\Gamma(x) = (4\pi^2 i x^2)^{-1}\big|_{C_\Gamma}. \tag{2.1}$$

Here C_Γ is a contour in the complex x^0 plane, given in Ref. 3, and $\eta_\Gamma = 1$ for $C_\Gamma = C_\pm$, C, C_A, C_R, C_{1A}, C_{1R}, $\eta_P = 2$, and $\eta_1 = i/2$. Equation (2.1) expresses symbolically the fact that $D_\Gamma(x)$ is obtained as the limit of an analytic functional as the complex contour C_Γ approaches the real axis. For details the reader is referred to Ref. 2.

Concentrating first on $D_\pm(x)$, we see that one can define a "modified power" of $D_\pm(x)$ by

$$[D_\pm]_n(x) \equiv [\mp(4\pi^2 i x^2)^{-1}]^n\big|_{\mp C_\pm}. \tag{2.2}$$

The choice of phase factor is a matter of convenience. This power is clearly to be distinguished from

$$[D_\pm]^n(x) \equiv [\mp(4\pi^2 i x^2)^{-1}\big|_{\mp C_\pm}]^n, \tag{2.3}$$

which we call "direct power." There is, of course, no reason why (2.2) and (2.3), provided they exist, should be equal.

Before discussing the existence of these expressions we recall that $\tilde{D}_\Gamma$, the p-space representation of the D functions, is given by[4]

$$\tilde{D}_\Gamma(p) = (2\pi)^{-2} p^{-2}\big|_{\tilde{C}_\Gamma}. \tag{2.4}$$

The modified power of $\tilde{D}_\Gamma$ was studied in Ref. 2. More specifically, the distribution $\tilde{D}^n_\Gamma(p)$ was defined there by

$$\tilde{D}^n_\Gamma(p) \equiv (2\pi)^{-2}(p^{-2})^{n+1}\big|_{\tilde{C}_\Gamma}, \tag{2.5}$$

and this is related to the modified power by

$$[\tilde{D}_\Gamma]_n(p) = [1/(4\pi^2)^{n-1}]\tilde{D}^{n-1}_\Gamma(p). \tag{2.6}$$

[1] Walter E. Thirring, *Principles of Quantum Electrodynamics* (Academic Press Inc., New York, 1958), Appendix II.
[2] V. Georgé and F. Rohrlich, J. Math. Phys. **8**, 1748 (1967).
[3] F. Rohrlich and F. Strocchi, Phys. Rev. **139**, B476 (1965).
[4] J. M. Jauch and F. Rohrlich, *Theory of Photons and Electrons* (Addison-Wesley Publ. Co., Reading, Mass., 1959; second, corrected, printing), Appendix I.

There is a one-to-one relation between the contours $\tilde{C}_\Gamma$ and C_Γ (if both refer to the same complex plane):

$$\mp C_{+/-} = \tilde{C}_{A/R}, \quad C = \tilde{C}, \quad \mp C_{A/R} = \tilde{C}_{-/+},$$
$$-C_1 = \tilde{C}_P, \quad \pm C_{1A/1R} = \tilde{C}_{1R/1A}, \quad C_P = \tilde{C}_1. \quad (2.7)$$

Now the existence of (2.5) has been discussed[2] and it was found that $\tilde{D}^n_\Gamma$ exists only for those contours $\tilde{C}_\Gamma$ which are not pinched in the limit in which the poles coalesce. This means that (2.5) does *not* exist for $\tilde{D}^n_{1R}$, $\tilde{D}^n_{1A}$, $\tilde{D}^n_1$, $\tilde{D}^n_\pm$, and $n > 0$.

This information also answers the questions concerning the existence of the modified powers of $D_\Gamma(x)$. Using (2.7), we see that the modified powers of all the homogeneous $D_\Gamma(x)$ do exist and those for the inhomogeneous $D_\Gamma(x)$ ($[D_{1A}]_n$, $[D_{1R}]_n$, $[D_P]_n$, $[D_A]_n$, $[D_R]_n$) do not exist. "Homogeneous" and "inhomogeneous" refer here to the differential equations which the D_Γ satisfy.

The existing modified powers thus (for $n \geq 1$) are

$$[D_\pm]_n(x) = \left(\mp\frac{1}{4\pi^2 i}\right)^n (2\pi)^2 \tilde{D}^{n-1}_{A/R}(x)$$
$$= \left(\mp\frac{1}{4\pi^2 i}\right)^n \left[R\left(\frac{1}{x^2}\right)^n \mp \frac{i\pi}{(n-1)!}\epsilon(x^0)\delta^{(n-1)}(x^2)\right], \quad (2.8)$$

$$[D]_n(x) \equiv \left(\frac{1}{4\pi^2 i}\frac{1}{x^2}\right)^n\Bigg|_C = \left(\frac{1}{4\pi^2 i}\right)^n (2\pi)^2 \tilde{D}^{n-1}(x)$$
$$= \left(\frac{1}{4\pi^2 i}\right)^n \frac{2\pi i}{(n-1)!}\epsilon(x^0)\delta^{(n-1)}(x^2), \quad (2.9)$$

$$[D_1]_n(x) \equiv \left(\frac{1}{2\pi^2}\frac{1}{x^2}\right)^n\Bigg|_{-C_1} = \left(\frac{1}{2\pi^2}\right)^n (2\pi)^2 \tilde{D}^{n-1}_P(x)$$
$$= \left(\frac{1}{2\pi^2}\right)^n R\left(\frac{1}{x^2}\right)^n. \quad (2.10)$$

The explicit expressions are taken from Ref. 2. One easily verifies the following interrelations between the modified powers for $n \geq 1$:

$$[D_+]_{2n} + [D_-]_{2n} = 2(-\tfrac{1}{4})^n [D_1]_{2n}, \quad (2.11)$$
$$[D_+]_{2n-1} + [D_-]_{2n-1} = [D]_{2n-1}, \quad (2.12)$$
$$[D_+]_{2n} - [D_-]_{2n} = -[D]_{2n}, \quad (2.13)$$
$$[D_+]_{2n-1} - [D_-]_{2n-1} = i(-\tfrac{1}{4})^{n-1}[D_1]_{2n-1}. \quad (2.14)$$

We now turn to the question of existence of direct powers. The direct powers of $D_\pm(x)$ were considered by Klaiber.[5] His result (in our notation) is

$$[D_\pm]^n(x) = \left(\mp\frac{1}{4\pi^2 i}\right)^n \lim_{\epsilon\to+0} \frac{1}{[\mathbf{x}^2 - (x^0 \mp i\epsilon)^2]^n}. \quad (2.15)$$

Comparison with (2.2) shows that the direct and modified powers are equal:

$$[D_\pm]^n(x) = [D_\pm]_n(x). \quad (2.16)$$

It follows that the modified powers of D and D_1 [Eqs. (2.9) and (2.10)] can be expressed in terms of the direct powers of D_+ and D_- via (2.11) to (2.14).

The direct powers of D and D_1, however, do not exist. This follows from the fact that the cross terms $[D_+]^k[D_-]^l$, which occur in $[D]^n = [D_+ + D_-]^n$ and $[iD_1]^n = [D_+ - D_-]^n$, are undefined for all $k, l \geq 1$.

It is obvious that the existence of the direct powers of D_+ and D_- permits the existence of analytic functionals of D_+ and D_-. Furthermore, "modified analytic functionals" of all homogeneous D functions exist as series of modified powers.

3. FOURIER TRANSFORM

In (2.6) we had the modified power $[\tilde{D}_\Gamma]_n$ of $\tilde{D}_\Gamma$. This distribution must be carefully distinguished from the Fourier transform of the modified power of D_Γ, i.e., $[D_\Gamma]^\sim_n$. We have[6]

$$[D_\pm]^\sim_n(p) = \frac{1}{(2\pi)^2}\int e^{-ipx}[D_\pm]_n(x)\, d^4x$$
$$= \left(\mp\frac{1}{4\pi^2 i}\right)^n \int e^{-ipx}\tilde{D}^{n-1}_{A/R}(x)\, d^4x$$
$$= \left(\mp\frac{1}{4\pi^2 i}\right)^n (2\pi)^2 D^{n-1}_{A/R}(-p)$$
$$= \mp\left(\mp\frac{1}{4\pi^2 i}\right)^{n-1}\frac{2}{i}\,\theta(\pm p^0)D^{n-1}_p(-p), \quad n > 1. \quad (3.1)$$

From Ref. 2 we learn that

$$D^n_P(x) = \frac{\theta(-x^2)(|x^2|/4)^{n-1}}{16\pi(n-1)!\,n!}, \quad n \geq 1. \quad (3.2)$$

Thus,

$$[D_\pm]^\sim_n(p) = \mp\left(\mp\frac{1}{4\pi^2 i}\right)^{n-1}\frac{\theta(\pm p^0)\theta(-p^2)(|p^2|/4)^{n-2}}{8\pi i(n-2)!\,(n-1)!}$$
$$= \mp\left(\mp\frac{1}{16\pi^2 i}\right)^{n-1}\frac{\theta(\pm p^0)\theta(-p^2)\,|p^2|^{n-2}}{2\pi i(n-2)!\,(n-1)!}, \quad n > 1. \quad (3.3)$$

The distributions $[D_+]^\sim_m$ and $[D_-]^\sim_n$ thus have no common support for any $m, n \geq 1$. Similarly, one finds

$$[D]^\sim_n(p) = (4\pi^2 i)^{-n}(2\pi)^2 D^{n-1}(-p)$$
$$= -\left(\frac{1}{16\pi^2 i}\right)^{n-1}\frac{\epsilon(p^0)\theta(-p^2)\,|p^2|^{n-2}}{2\pi i(n-2)!\,(n-1)!}, \quad n > 1, \quad (3.4)$$
$$= [D_-]^\sim_n(p) - (-)^n[D_+]^\sim_n(p), \quad n \geq 1, \quad (3.5)$$

[5] B. Klaiber, Nuovo Cimento **36**, 165 (1965).

[6] The first of Eqs. (4.11), Ref. 2, is used here. A misprint distorted this equation which should read $\Delta^n_{R,A}(x) = \pm\theta(\pm x^0)\Delta^n(x) = 2\theta(\pm x^0)\Delta^n_P(x)$.

in agreement with (2.12) and (2.13). Finally,

$$[D_1]_n^{\sim}(p) = \left(\frac{1}{2\pi^2}\right)^{n-1} 2D_P^{n-1}(-p)$$
$$= \left(\frac{1}{8\pi^2}\right)^{n-1} \frac{\theta(-p^2)\,|p^2|^{n-2}}{2\pi(n-2)!\,(n-1)!} \tag{3.6}$$
$$= \tfrac{1}{2}(2i)^n([D_-]_n^{\sim}(p) + (-)^n[D_+]_n^{\sim}(p)), \quad n \geq 1. \tag{3.7}$$

The direct powers of $D_\pm$ are related to the modified powers in p space the same way as in x space, (2.16):

$$[D_\pm]^{\tilde{n}}(p) = [D_\pm]_n^{\sim}(p). \tag{3.8}$$

While no other D_Γ functions have direct powers, as mentioned before, there are $\tilde{D}_\Gamma$ functions which do. Writing (2.8) with argument p and setting $n = 1$, we obtain, by taking the direct power,

$$[\tilde{D}_{A/P}]^n(p) = (\mp i)^n[D_\pm]^n(p)$$
$$= (\mp i)^n(\mp 4\pi^2 i)^{-n}(2\pi)^2\tilde{D}_{A/P}^{n-1}(p)$$
$$= [\tilde{D}_{A/R}]_n(p). \tag{3.9}$$

In the last equality, we used (2.6).

4. THE KLUK REPRESENTATION OF $[D_\Gamma]_n$

The result (3.3) leads to the Källén–Lehmann–Umezawa–Kamefuchi (KLUK) representation of $[D_\pm]^n(x)$. The Fourier transform of (3.3) gives

$$[D_\pm]^n(x) = \frac{1}{(2\pi)^2}\int e^{ipx}[D_\pm]^{\tilde{n}}(p)\,d^4p$$
$$= \left(\mp\frac{1}{16\pi^2 i}\right)^{n-1}\frac{1}{(n-2)!\,(n-1)!}\frac{\pm i}{(2\pi)^3}$$
$$\times\int e^{ipx}\theta(\pm p^0)\delta(p^2+\kappa^2)\kappa^{2n-4}\,d^4p\,d\kappa^2$$
$$= \left(\mp\frac{1}{16\pi^2 i}\right)^{n-1}\frac{1}{(n-2)!\,(n-1)!}$$
$$\times\int_0^\infty \Delta_\pm(x,\kappa^2)\kappa^{2n-4}\,d\kappa^2, \tag{4.1}$$

which agrees with Ref. 5. Similarly,

$$[D]_n(x) = \left(\frac{1}{16\pi^2 i}\right)^{n-1}\frac{1}{(n-2)!\,(n-1)!}$$
$$\times\int_0^\infty \Delta(x,\kappa^2)\kappa^{2n-4}\,d\kappa^2, \tag{4.2}$$

$$[D_1]_n(x) = \left(\frac{1}{8\pi^2}\right)^{n-1}\frac{1}{(n-2)!\,(n-1)!}$$
$$\times\int_0^\infty \Delta_1(x,\kappa^2)\kappa^{2n-4}\,d\kappa^2. \tag{4.3}$$

From (4.1) and (3.9) it follows that for the direct power of $\tilde{D}_{A/R}$ we get

$$[\tilde{D}_{A/R}]^n(p) = \mp i\left(\frac{1}{16\pi^2}\right)^{n-1}\frac{1}{(n-2)!\,(n-1)!}$$
$$\times\int_0^\infty \Delta_\pm(p,\kappa^2)\kappa^{2n-4}\,d\kappa^2. \tag{4.4}$$

5. THE POWERS OF $\Delta_\Gamma(x, m)$

Since the $\Delta_\Gamma(x, m)$, $m > 0$, cannot be written as analytic functionals similar to (2.1), modified powers cannot be defined in the same way. The direct powers of $\Delta_\pm(x, m)$ are known to exist. Their KLUK representation follows from Thirring's result[1]:

$$\Delta_\pm(x,a)\Delta_\pm(x,b) = \pm i\int_0^\infty d\rho_2(a,b,\kappa^2)\Delta_\pm(x,\kappa),$$
$$d\rho_2(a,b,\kappa^2) = \frac{d\kappa^2}{16\pi^2}\frac{[(\kappa^2-a^2-b^2)^2-(2ab)^2]^{\frac{1}{2}}}{\kappa^2}$$
$$\times\,\theta[\kappa^2-(a+b)^2]. \tag{5.1}$$

One has

$$[\Delta_\pm]^n(x,m) = (\pm i)^{n-1}\int_0^\infty d\rho_n(m,\kappa^2)\Delta_\pm(x,\kappa), \tag{5.2}$$

where $d\rho_n(m,\kappa^2)$ is determined by iteration from $d\rho_2(m,\kappa^2) = d\rho_2(m,m,\kappa^2)$ and

$$d\rho_n(m\kappa'^2) = \int d\rho_2(m,\kappa,\kappa'^2)\,d\rho_{n-1}(m,\kappa^2). \tag{5.3}$$

No direct powers of other $\Delta_\Gamma(x, m)$ exist as in the $m = 0$ case. In p space,

$$\tilde{\Delta}_\Gamma(p,m) = \frac{1}{(2\pi)^2}\frac{1}{p^2+m^2}\Big|_{\mathcal{C}_\Gamma}, \tag{5.3'}$$

so that modified powers *can* be defined. As in (2.6), we have

$$[\tilde{\Delta}_\Gamma]_n(p,m) = \left(\frac{1}{(2\pi)^2}\frac{1}{p^2+m^2}\right)^n\Big|_{\mathcal{C}_\Gamma}$$
$$= \left(\frac{1}{4\pi^2}\right)^{n-1}\tilde{\Delta}_\Gamma^{n-1}(p,m). \tag{5.4}$$

The latter are the distributions given in Ref. 2.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Curvature Collineations: A Fundamental Symmetry Property of the Space-Times of General Relativity Defined by the Vanishing Lie Derivative of the Riemann Curvature Tensor

GERALD H. KATZIN,* JACK LEVINE,* AND WILLIAM R. DAVIS
Departments of Physics and Mathematics, North Carolina State University, Raleigh, North Carolina

(Received 18 March 1968)

A Riemannian space V_n is said to admit a particular symmetry which we call a "curvature collineation" (CC) if there exists a vector ξ^i for which $\pounds_\xi R^i_{jkm} = 0$, where R^i_{jkm} is the Riemann curvature tensor and $\pounds_\xi$ denotes the Lie derivative. The investigation of this symmetry property of space-time is strongly motivated by the all-important role of the Riemannian curvature tensor in the theory of general relativity. For space-times with zero Ricci tensor, it follows that the more familiar symmetries such as projective and conformal collineations (including affine collineations, motions, conformal and homothetic motions) are subcases of CC. In a V_4 with vanishing scalar curvature R, a covariant conservation law generator is obtained as a consequence of the existence of a CC. This generator is shown to be directly related to a generator obtained by means of a direct construction by Sachs for null electromagnetic radiation fields. For pure null-gravitational space-times (implying vanishing Ricci tensor) which admit CC, a similar covariant conservation law generator is shown to exist. In addition it is found that such space-times admit the more general generator (recently mentioned by Komar for the case of Killing vectors) of the form $(\sqrt{-g}\, T^{ijkm}\xi_i\xi_j\xi_k)_{;m} = 0$, involving the Bel-Robinson tensor T^{ijkm}. Also it is found that the identity of Komar, $[\sqrt{-g}(\xi^{i;j} - \xi^{j;i})]_{;i;j} = 0$, which serves as a covariant generator of field conservation laws in the theory of general relativity appears in a natural manner as an essentially trivial necessary condition for the existence of a CC in space-time. In addition it is shown that for a particular class of CC, $\pounds_\xi K$ is proportional to K, where K is the Riemannian curvature defined at a point in terms of two vectors, one of which is the CC vector. It is also shown that a space-time which admits certain types of CC also admits cubic first integrals for mass particles with geodesic trajectories. Finally, a class of null electromagnetic space-times is analyzed in detail to obtain the explicit CC vectors which they admit.

1. INTRODUCTION

In accord with the early observations of Nöther,[1] the existence of certain geometric symmetry properties described by continuous groups of motions or collineations lead to conservation laws in the form of first integrals (i.e., constants of the motion) of a dynamical system. Indeed, the fundamental importance of groups of motions (Killing vectors) in space-time and their relation to the conservation laws of energy, linear momentum, and angular momentum for particles and fields is well known.[2] The relation between the existence of affine or projective collineations in the Riemannian space-times of general relativity and the existence of concomitant particle and field conservation laws was developed in several recent papers.[3]

In this paper we are concerned with a symmetry property of space-time which we call a "curvature collineation" (CC). A Riemannian space-time V_n is said to admit a CC if there exists an infinitesimal transformation $\bar{x}^i = x^i + \epsilon\xi^i(x)$ for which $\pounds R^i_{jkm} = 0$, where ϵ is a positive infinitesimal, $\pounds$ denotes the Lie derivative with respect to the vector ξ^i, and R^i_{jkm} is the Riemannian curvature tensor. Clearly, the investigation of this symmetry property is strongly motivated by the all-important role of the Riemann curvature tensor in the general theory of relativity.

Our preliminary investigations have led to several results of physical and geometrical interest. We have shown that the existence of a certain type of CC leads directly to the existence of a cubic first integral of a mass particle with geodesic trajectory. In addition we have found that if a V_4 with nonvanishing Ricci tensor R_{ij} and with vanishing scalar curvature R admits a CC then a field conservation law results. This conservation law is directly related to a conservation law obtained by Sachs[4] for null electromagnetic radiation fields. Also, it is shown that our result can be extended to pure null gravitation radiation fields which were also treated by Sachs.

Furthermore, it turns out that the identity of Komar,[5] which serves as a covariant generator of field conservation laws in the theory of general relativity, appears in a natural manner as an essentially trivial necessary condition for the existence of a CC in a V_n.

* Supported by National Science Foundation Grant No. GP 6876.

[1] E. Nöther, Nachr. Akad. Wiss. Göttingen. II Math. Physik Kl., Vol. 235 (1918).

[2] For a discussion of the relation between groups of motions admitted by space-time and conservation laws of particle mechanics see, for example, W. R. Davis and G. H. Katzin, Am. J. Phys. 30, 750 (1962).

[3] W. R. Davis and M. K. Moss, Nuovo Cimento 38, 1558 (1965); G. H. Katzin and J. Levine, J. Math. Phys. 9, 8 (1968).

[4] R. K. Sachs, Z. Phys. 157, 462 (1960).

[5] A. Komar, Phys. Rev. 113, 934 (1959).

For those space-times with $R_{ij} = 0$ it is found that, when they are admitted, projective collineations (including affine collineations and motions as subcases) and conformal collineations (including conformal and homothetic motions as subcases) are subcases of CC. Thus, it appears that the CC is the fundamental symmetry property of space-time to consider in the study of conservation laws pertaining to gravitational radiation.

Indeed, quite aside from the question of new conservation laws, it follows that studies of CC could provide an important invariant classification of Riemannian space-times which would include and extend far beyond the familiar classification that has been given by Petrov[6] on the basis of groups of motions.

A general picturesque geometric interpretation of a CC is difficult to obtain. However it will be shown that the deformation of the Riemannian curvature K defined by $\pounds K$ takes a particularly simple form if the Lie derivative is formed with respect to a particular type of CC vector. In addition, we give a diagram (Fig. 1) which shows the relationship between CC and other more familiar symmetry properties of space-time.

In this paper we give several theorems concerning CC in Einstein spaces. Finally, a particular class of null electromagnetic space-times is analyzed in detail to obtain the explicit CC vectors which they admit. In a subsequent paper, we shall treat CC in conformally flat spaces, including a treatment of the group structure of the CC.

Throughout this paper we need to refer to the equations describing motions, affine collineations, projective collineations, homothetic motions, conformal motions, and conformal collineations. We, therefore, now give a summary of these well-known space-time symmetries. For most of these symmetries we find it convenient to express the necessary and sufficient conditions in several alternative forms. It is to be understood that the symbol $\pounds$ denotes Lie differentiation with respect to a vector ξ^i (sometimes written $\pounds_\xi$ to stress the vector dependence).

Motion[7] (M): A V_n is said to admit a M provided there exists a (Killing) vector ξ^i such that[8]

$$h_{ij} \equiv \pounds g_{ij} = \xi_{i;j} + \xi_{j;i} = 0. \tag{1.1}$$

Affine Collineation[9] (AC): A V_n is said to admit an AC provided there exists a vector ξ^i such that

$$\pounds \left\{ {k \atop ij} \right\} \equiv \xi^k_{;ji} + \xi^m R^k_{jmi}$$

$$\equiv \tfrac{1}{2} g^{kl}(h_{li;j} + h_{lj;i} - h_{ij;l}) = 0, \tag{1.2}$$

where $\{{k \atop ij}\}$ is the Christoffel symbol of the second kind and where the Riemannian curvature tensor[10]

$$R^l_{ijk} \equiv \left\{ {l \atop ik} \right\}_{,j} - \left\{ {l \atop ij} \right\}_{,k} + \left\{ {m \atop ik} \right\} \left\{ {l \atop mj} \right\} - \left\{ {m \atop ij} \right\} \left\{ {l \atop mk} \right\}.$$

Alternatively, the necessary and sufficient condition (1.2) for an AC may be expressed in the form

$$h_{ij;k} = 0. \tag{1.3}$$

Obviously every M is an AC. We use the terminology proper AC (Prop AC) to denote those AC which are not M.

Projective Collineations[11] (PC): A V_n is said to admit a PC provided there exists a vector ξ^i such that

$$\pounds \Pi^i_{jk} = 0, \tag{1.4}$$

where the projective connection

$$\Pi^i_{jk} \equiv \left\{ {i \atop jk} \right\} - \frac{1}{n+1} \left(\delta^i_j \left\{ {h \atop hk} \right\} + \delta^i_k \left\{ {h \atop hj} \right\} \right).$$

Alternatively, we may express (1.4) in the form

$$\pounds \left\{ {i \atop jk} \right\} = \delta^i_j \phi_{;k} + \delta^i_k \phi_{;j}, \tag{1.5}$$

where

$$\phi_{;j} = (n+1)^{-1} \xi^m_{;mj}. \tag{1.6}$$

It follows from (1.5) that for a PC we have

$$h_{ij;k} = 2g_{ij}\phi_{;k} + g_{ik}\phi_{;j} + g_{jk}\phi_{;i}. \tag{1.7}$$

In addition we find that for every PC we have

$$\pounds W^i_{jkl} = 0, \tag{1.8}$$

---

[6] See, for example, A. Z. Petrov, "Invariant Classification of Gravitational Fields," in *Recent Developments in General Relativity* (The Macmillan Co., New York, 1962), pp. 371–378.

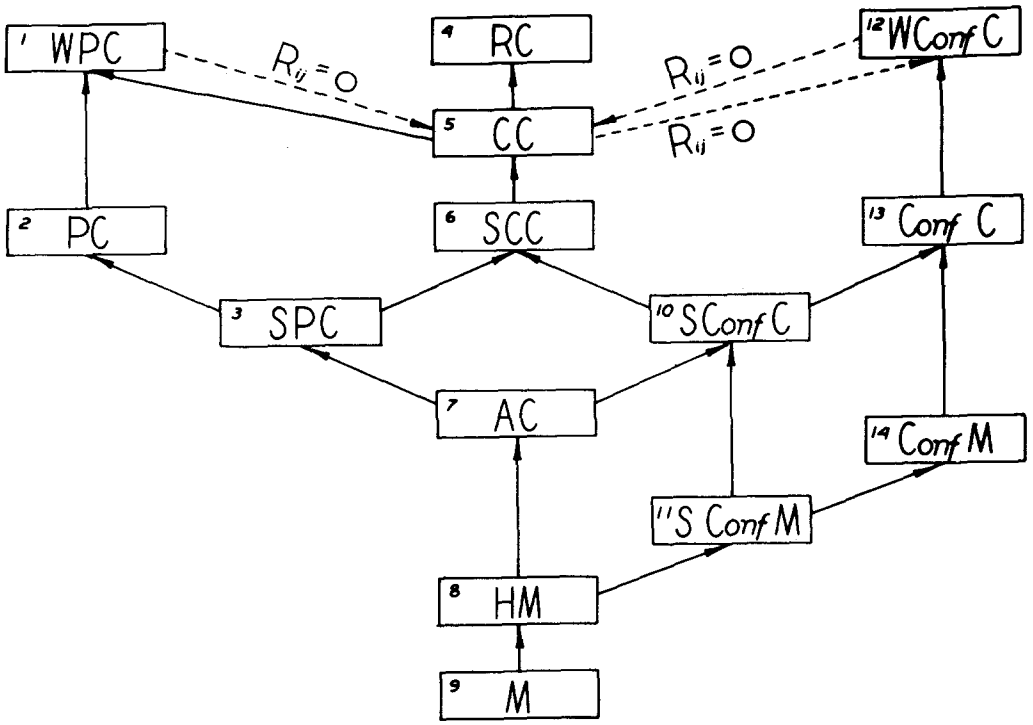
[7] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, N.J., 1926).

[8] Indices take on values $1, 2, \cdots, n$ and the Einstein summation convention is used. Covariant differentiation is indicated by a semicolon (;) and partial differentiation by a comma (,).

[9] K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland Publ. Co., Amsterdam, 1957).

[10] The definition of R^i_{jkl} used in this paper is given by Eisenhart (see Ref. 7) and differs in sign from the definition used by Yano (see Ref. 9). Throughout this paper we have adjusted equations obtained from Ref. 9 to be consistent with Eisenhart's definition of the curvature tensor.

[11] See Ref. 9 and J. A. Schouten [*Ricci-Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed.] for a thorough discussion of PC. Also, G. H. Katzin and Jack Levine [Tensor, New Series **19**, 162 (1968)] give an alternative derivation of the necessary and sufficient conditions for PC in terms of the Lie derivative of the parameter-independent form of the path equation.



1. WPC - Weyl Projective Collineation - $\mathcal{L}W^{i}_{jkh}=0$.

2. PC - Projective Collineation - $\mathcal{L}\{^{i}_{jk}\}=\delta^{i}_{j}\phi_{;k}+\delta^{i}_{k}\phi_{;j}$.

3. SPC - Special Projective Collineation - $\mathcal{L}\{^{i}_{jk}\}=\delta^{i}_{j}\phi_{;k}+\delta^{i}_{k}\phi_{;j}$, $\phi_{;jk}=0$.

4. RC - Ricci Collineation - $\mathcal{L}R_{ij}=0$.

5. CC - Curvature Collineation - $\mathcal{L}R^{i}_{jkh}=0$.

6. SCC - Special Curvature Collineation - $(\mathcal{L}\{^{i}_{jk}\})_{;h}=0$.

7. AC - Affine Collineation - $\mathcal{L}\{^{i}_{jk}\}=0$.

8. HM - Homothetic Motion - $\mathcal{L}g_{ij}=2\sigma g_{ij}$, $\sigma=\text{const}$.

9. M - Motion - $\mathcal{L}g_{ij}=0$.

10. SConf C - Special Conformal Collineation - $\mathcal{L}\{^{i}_{jk}\}=\delta^{i}_{j}\sigma_{;k}+\delta^{i}_{k}\sigma_{;j}-g_{jk}g^{ih}\sigma_{;h}$, $\sigma_{;jk}=0$.

11. S Conf M - Special Conformal Motion - $\mathcal{L}g_{ij}=2\sigma g_{ij}$, $\sigma_{;jk}=0$.

12. W Conf C - Weyl Conformal Collineation - $\mathcal{L}C^{i}_{jkh}=0$.

13. Conf C - Conformal Collineation - $\mathcal{L}\{^{i}_{jk}\}=\delta^{i}_{j}\sigma_{;k}+\delta^{i}_{k}\sigma_{;j}-g_{jk}g^{ih}\sigma_{;h}$.

14. Conf M - Conformal Motion - $\mathcal{L}g_{ij}=2\sigma g_{ij}$.

FIG. 1. Relation between space–time symmetries.

where the Weyl projective curvature tensor is given by

$$W^h_{ijk} = R^h_{ijk} - (n-1)^{-1}(\delta^h_k R_{ij} - \delta^h_j R_{ik}). \quad (1.9)$$

By inspection, every AC is a PC (i.e., a PC with $\phi_{;k} = 0$). We use the terminology proper PC (Prop PC) to denote those PC which are not AC.

Conformal Motion[9] (Conf M): A V_n is said to admit a Conf M provided there exists a vector ξ^i such that

$$\pounds(\mathfrak{g}^{-1/n} g_{ij}) = 0, \quad \mathfrak{g} \equiv |\mathrm{Det}\,(g_{ij})|. \quad (1.10)$$

Equivalently we have

$$h_{ij} = 2\sigma g_{ij}, \quad (1.11)$$

where σ is a scalar expressible in the form

$$\sigma = n^{-1}\xi^k_{;k}. \quad (1.12)$$

It follows that every Conf M must satisfy

$$\pounds\begin{Bmatrix} i \\ jk \end{Bmatrix} = \delta^i_j\sigma_{;k} + \delta^i_k\sigma_{;j} - g_{jk}g^{im}\sigma_{;m}. \quad (1.13)$$

It can also be shown that every Conf M satisfies $\pounds K^i_{jk} = 0$, where the conformal connection K^i_{jk} is formed with the relative tensor $(\mathfrak{g}^{-1/n} g_{ij})$ in the same manner that the Christoffel symbol $\{^i_{jk}\}$ is constructed with the metric tensor g_{ij}. Alternatively, K^i_{jk} may be expressed in the form

$$K^i_{jk} \equiv \begin{Bmatrix} i \\ jk \end{Bmatrix} - \frac{1}{n}\left(\delta^i_j\begin{Bmatrix} m \\ mk \end{Bmatrix} + \delta^i_k\begin{Bmatrix} m \\ mj \end{Bmatrix} - g_{jk}g^{im}\begin{Bmatrix} h \\ hm \end{Bmatrix}\right).$$

We use the notation Proper Conf M (Prop Conf M) for those Conf M with $\sigma \neq$ const.

Homothetic Motions[9] (HM): A V_n is said to admit an HM if there exists a vector ξ^i such that (1.11) holds with σ a nonzero constant.

Conformal Collineations[12] (Conf C): A V_n is said to admit a Conf C if there exists a vector ξ^i for which (1.13) holds. It follows that every Conf M is a Conf C, but not necessarily conversely. It can be shown that the necessary and sufficient condition (1.13) for a Conf C may be expressed in the equivalent form

$$h_{ij;k} = 2\sigma_{;k}g_{ij}, \quad (1.14)$$

and that every Conf C (see Ref. 9, p. 160) must satisfy

$$\pounds C^h_{ijk} = 0, \quad (1.15)$$

where the conformal curvature tensor C^h_{ijk} is defined by[7]

$$C^h_{ijk} \equiv R^h_{ijk} + (n-2)^{-1}(\delta^h_j R_{ik} - \delta^h_k R_{ij} + g_{ik}R^h_j - g_{ij}R^h_k) + R[(n-1)(n-2)]^{-1}(\delta^h_k g_{ij} - \delta^h_j g_{ik}). \quad (1.16)$$

[Consistent with the notation of Eisenhart (see Ref. 10) we define the Ricci tensor by $R_{ij} \equiv R^h_{ijh}$, and the scalar curvature by $R \equiv R^i_i$.]

As a means of succinctly summarizing the relations between the well-known symmetries discussed above and several additional symmetries which are considered in this paper, we have constructed a block diagram (Fig. 1). The diagram should be read in the following sense: When it exists, (i.e., is admitted by the V_n) the symmetry described in any given block is automatically a subcase of the symmetries described in those adjacent blocks indicated by the arrows leading from the given block. Thus for example, if an HM is admitted by the V_n, then the transformation which defines the HM also satisfies the requirements for being a Conf M and AC and so on through the diagram. Note that the dashed arrows should only be considered when the Ricci tensor vanishes. The several blocks containing symmetries which were not discussed earlier will be explained at appropriate places in the text.

We remark in passing that there are several interesting possibilities for expanding this diagram in the sense of defining additional new symmetries. Here, however, we have limited the diagram to include only those symmetries which we discuss because of their relation to CC.

2. NECESSARY AND SUFFICIENT CONDITIONS FOR CURVATURE COLLINEATIONS

The infinitesimal transformation

$$\bar{x}^i = x^i + \xi^i(x)\delta t, \quad (2.1)$$

where δt is a positive infinitesimal, defines a curvature collineation (CC) provided the Riemannian space V_n (of general signature) admits a vector field $\xi^i(x)$ such that

$$\pounds_\xi R^k_{jhi} = 0. \quad (2.2)$$

In general, the solution to (2.2) consists of a set of r vectors $\xi^i_{(\alpha)}$ $\alpha = 1, \cdots, r$ which define an r-parameter invariance group.[9] However, in this paper we shall not investigate the group properties of CC *per se*.

Next we present several useful forms for the Lie derivative of the curvature tensor. Formally, we have (see Ref. 9):

$$\pounds R^k_{jhi} \equiv R^k_{jhi,m}\xi^m + R^k_{mhi}\xi^m_{,j} + R^k_{jmi}\xi^m_{,h} + R^k_{jhm}\xi^m_{,i} - R^m_{jhi}\xi^k_{,m}. \quad (2.3)$$

[12] Y. Tashiro, Math. J. Okayamo Univ. **10**, 75 (1960).

By use of the definition of covariant differentiation, we obtain

$$\pounds R^k_{jhi} = R^k_{jhi;m}\xi^m + R^k_{mhi}\xi^m_{;j} + R^k_{jmi}\xi^m_{;h} + R^k_{jhm}\xi^m_{;i} - R^m_{jhi}\xi^k_{;m}. \quad (2.4)$$

If we employ the Bianchi and Ricci identities (see Ref. 7) and use (1.2), we find (2.4) can be expressed in the forms

$$\pounds R^k_{jhi} = \left(\pounds\left\{{k \atop ij}\right\}\right)_{;h} - \left(\pounds\left\{{k \atop hj}\right\}\right)_{;i} \quad (2.5)$$

and

$$\pounds R^k_{jhi} = \tfrac{1}{2}g^{km}[(h_{im;j} + h_{mj;i} - h_{ij;m})_{;h} - (h_{hm;j} + h_{mj;h} - h_{hj;m})_{;i}]. \quad (2.6)$$

By substitution of $\pounds R^k_{jhi}$ as given by (2.6) into (2.2) we obtain (after multiplying by g_{kl} to lower the index k):

Theorem 2.1: A necessary and sufficient condition for a Riemannian space V_n to admit a CC is that there exist a transformation of the form (2.1) such that the vector ξ^i satisfies

$$(h_{im;j} + h_{mj;i} - h_{ij;m})_{;h} - (h_{hm;j} + h_{mj;h} - h_{hj;m})_{;i} = 0, \quad (2.7)$$

where

$$h_{ij} \equiv \xi_{i;j} + \xi_{j;i}.$$

We may express (2.7) in an equivalent but simpler form by returning to (2.2) and substituting (2.5) into (2.2) and then using the first expression for $\pounds\{{i \atop jk}\}$ given by (1.2) along with the Ricci identity to obtain

$$(\xi_{i;mj} + \xi_{m;ji} - \xi_{i;jm})_{;h} - (\xi_{h;mj} + \xi_{m;jh} - \xi_{h;jm})_{;i} = 0. \quad (2.8)$$

Although (2.8) is a simpler equation than (2.7), we find (2.7) to be more useful for most of our considerations.

From (2.2) we observe by contracting on the indices k and i that every CC vector ξ^i satisfies

$$\pounds R_{jh} = 0. \quad (2.9)$$

In general, if a V_n admits a vector ξ^i such that (2.9) holds we say that the V_n admits a "Ricci collineation" (RC). Thus,

Theorem 2.2: In a V_n every CC is an RC.

In (2.7) if we interchange the indices j and m and add the resulting equation to (2.7) we obtain

Theorem 2.3: A necessary condition for a transformation of the form (2.1) to define a CC is that

$$h_{jm;ih} - h_{jm;hi} = 0. \quad (2.10)$$

It is of interest to note that (2.10) could also be obtained by starting with

$$g_{ia}R^a_{jkm} + g_{ja}R^a_{ikm} \equiv 0. \quad (2.11)$$

By taking the Lie derivative of (2.11) it follows that, if (2.2) holds, we obtain

$$h_{ia}R^a_{jkm} + h_{ja}R^a_{ikm} = 0 \quad (2.12)$$

which by means of the Ricci identity reduces to (2.10).

The necessary condition (2.10) for a CC leads directly to an identity that has been of special interest in the formulation of the field conservation laws of general relativity. In particular, if the condition (2.10) is multiplied by $\mathfrak{g}^{\frac{1}{2}}g^{jh}g^{mi}$, where $\mathfrak{g} \equiv |\mathrm{Det}\, g_{ij}|$, one obtains

$$[\mathfrak{g}^{\frac{1}{2}}(\xi^{i;j} - \xi^{j;i})]_{;ji} \equiv \{[\mathfrak{g}^{\frac{1}{2}}(\xi^{i;j} - \xi^{j;i})]_{;j}\}_{,i} \equiv 0. \quad (2.13)$$

This is the covariant identity first discussed in the literature by Komar[5] and given further interpretation by Komar[13] and others[14] in terms of its role as a conservation law generator when the space–time admits symmetry properties. Since this tensor expression is obviously a vanishing identity[15] for all ξ^i, it follows that this necessary condition for a CC places no restriction on ξ^i. Nonetheless, this result emphasizes, as expected, that CC are necessarily symmetry properties of space–time that are embraced by the group of general curvilinear coordinate transformations in space–time.[16] In this paper, no attempt is made to explicitly formulate or interpret the field conservation laws that would follow from (2.13) for

[13] A. Komar, Phys. Rev. **127**, 1411 (1962).

[14] W. R. Davis and M. K. Moss, Nuovo Cimento **27**, 1492 (1963); **38**, 1531, 1558 (1965).

[15] For a derivation of this conservation law generator that is connected with the transformation properties of the Lagrangian underlying the variational formulation of general relativity, see W. R. Davis and M. K. Moss, J. Math. Phys. **7**, 975 (1966). Of course, it is to be recognized that this evident identity can be obtained in a number of ways not obviously involving its direct construction with the help of the Ricci identities.

[16] In so far as M can be regarded as "rigid" displacements in space–time, it is clear that CC can be regarded as nonrigid deformations that leave the components of the Riemannian curvature tensor invariant in detailed functional form.

space-times that admit CC. However, it is of some interest to note that, for pure fields of gravitational radiation, (2.13) would not, in general, lead to trivial results for "proper" CC as do their counterparts corresponding to M (Killing vectors).[17] We discuss CC in V_n with $R_{ij} = 0$ in Sec. 4 and CC for null metrics in Sec. 7.

3. RELATIONS BETWEEN CC AND OTHER SYMMETRIES

From the definition (1.1) of a M in a V_n it is immediate that we may state

Theorem 3.1: In a V_n every M is a CC.

Similarly, from the definition (1.2) of an AC it follows that we may state

Theorem 3.2: In a V_n every AC is a CC.

Also it follows immediately from the definition of HM that h_{ij} from (1.11) satisfies (1.3) and, hence, as a consequence of Theorem 3.2 we state

Theorem 3.3: In a V_n every HM is a CC.

From Yano (Ref. 9, p. 167) it is known that, if a transformation is both a Conf M and a PC, then it is an HM. Hence, we have the following as a consequence of Theorem 3.3.

Theorem 3.4: In a V_n if a transformation is both a Conf M and a PC then it is a CC.

Next, let us consider under what conditions a PC is a CC. We therefore require that $\pounds\{^{i}_{jk}\}$ be given by (1.5) and substitute for $\pounds\{^{i}_{jk}\}$ in (2.5). If we then demand that $\pounds R^{k}_{ijh} = 0$, we obtain

$$\delta^{k}_{i}\phi_{;jh} - \delta^{k}_{h}\phi_{;ji} = 0. \tag{3.1}$$

We set $k = i$ and sum in (3.1) to obtain $\phi_{;jh} = 0$. We call a projective collineation with $\phi_{;jh} = 0$ a special projective collineation (SPC). It follows immediately by a covariant differentiation of (1.5) that an SPC satisfies

$$(\pounds\{^{k}_{ij}\})_{;h} = 0. \tag{3.2}$$

[17] Komar (Ref. 13) first pointed out that $(\xi^{i;j} - \xi^{j;i})_{;j}$ vanishes for motion ξ^i if $R_{ij} = 0$ everywhere.

In general, if a V_n admits a vector ξ^i such that (3.2) holds, we say that the V_n admits a special curvature collineation (SCC). Thus, every SPC is a SCC. We summarize the above by stating

Theorem 3.5: The necessary and sufficient condition for a PC to be a CC is for

$$\phi_{;jh} = 0, \tag{3.3}$$

where $\phi_{;jh} = (n+1)^{-1}\xi^{i}_{;ijh}$; that is, the PC must be an SPC.

Corollary 3.1: If a V_n admits a SPC then it admits a parallel field of vectors $\phi_j \equiv (n+1)^{-1}\xi^{i}_{;ij}$, where ξ^i defines the SPC.

We next turn our attention to the conditions for a Conf C to be a CC. We thus assume the V_n admits a Conf C, that is, (1.14) holds. We now use (1.13) to evaluate $\pounds\{^{k}_{ij}\}$ in (2.5) and require that $\pounds R^{k}_{ijh} = 0$. We immediately obtain

$$\delta^{k}_{i}\sigma_{;jh} - \delta^{k}_{h}\sigma_{;ji} - g_{ij}g^{km}\sigma_{;mj} + g_{hj}g^{km}\sigma_{;mi} = 0. \tag{3.4}$$

We set $k = i$ and sum in (3.4) to obtain

$$(n-2)\sigma_{;jh} + g_{hj}g^{im}\sigma_{;mi} = 0. \tag{3.5}$$

In (3.5) we multiply by g^{jh} and sum to obtain

$$g^{jh}\sigma_{;jh} = 0. \tag{3.6}$$

It follows from (3.5) and (3.6) that $\sigma_{;ij} = 0$. We call a conformal collineation with $\sigma_{;ij} = 0$ a special conformal collineation (S Conf C).[18] It follows immediately by covariant differentiation of (1.13) that an S Conf C satisfies (3.2). Thus every S Conf C is a SCC. We now summarize the above by stating

Theorem 3.6: The necessary and sufficient condition for a Conf C to be a CC is for

$$\sigma_{;jh} = 0, \tag{3.7}$$

where $\sigma_{;jh} = n^{-1}\xi^{i}_{;ijh}$; that is, the Conf C must be a S Conf C.

Corollary 3.2: If a V_n admits a S Conf C, then it admits a parallel field of vectors $\sigma_j \equiv n^{-1}\xi^{i}_{;ij}$, where ξ^i defines the S Conf C.

We define special conformal motion (S Conf M) as a Conf M with $\sigma_{;ij} = 0$. Hence, we have

[18] It follows from (1.11) and (1.13) that a special conformal motion (S Conf M) i.e., a conformal motion with $\sigma_{;ij} = 0$, is an S Conf C.

Theorem 3.7: Every S Conf M is a S Conf C.

By consideration of the deformation $\pounds_\xi K$ of the Riemannian curvature K under an S Conf M, we are able to gain a little more insight into this particular symmetry. We recall the Riemannian curvature associated with an elementary two-space is the invariant[19]

$$K = \epsilon(\eta)\epsilon(p)R^a_{ijk}g_{am}p^i\eta^j p^k\eta^m, \tag{3.8}$$

where η^i and p^i are orthogonal unit vectors, and $\epsilon(\eta)$ and $\epsilon(p)$ their indicators. We thus assume the V_n admits an S Conf M with respect to the vector ξ^i and then define the unit vector η^i by

$$\eta^i \equiv \mu^{-1}\xi^i, \quad \mu^2 \equiv \xi^j\xi_j, \tag{3.9}$$

where μ is a scalar. In addition, we take the unit vector p^i to be the tangent to a curve such that $p^i \equiv dx^i/ds$, where ds is the differential path parameter.

The deformation in K with respect to ξ^i takes the form

$$\begin{aligned}\pounds K = \epsilon(\eta)\epsilon(p)[&(\pounds R^a_{ijk})g_{am}p^i\eta^j p^k\eta^m \\ &+ R^a_{ijk}(\pounds g_{am})p^i\eta^j p^k\eta^m + 2R^a_{ijk}g_{am}\pounds(p^i)\eta^j p^k\eta^m \\ &+ 2R^a_{ijk}g_{am}p^i\pounds(\eta^i)p^k\eta^m].\end{aligned} \tag{3.10}$$

From Yano (Ref. 9, p. 89), we have

$$\pounds p^i \equiv \pounds\left(\frac{dx^i}{ds}\right) = -p^i\frac{(\pounds ds)}{ds}. \tag{3.11}$$

Next we formulate $\pounds\eta^j$ and find by use of (3.9) that

$$\pounds\eta^j = \eta^j_{;l}\xi^l - \eta^l\xi^j_{;l} = -\mu_{;l}\eta^l\eta^j. \tag{3.12}$$

Our assumption that ξ^i defines an S conf M assures us that $\pounds_\xi R^a_{ijk} = 0$ and $\pounds_\xi g_{ij} = 2\sigma g_{ij}$. Hence, with this assumption and use of (3.11) and (3.12) we obtain

$$\pounds K = 2[\sigma - (\pounds ds)/ds - \mu_{;l}\eta^l]K. \tag{3.13}$$

From the fundamental form of the V_n,

$$ds^2 = g_{ij}\,dx^i\,dx^j, \tag{3.14}$$

we obtain

$$2ds(\pounds ds) = (\pounds g_{ij})\,dx^i\,dx^j, \tag{3.15}$$

which, for a Conf M (see 1.11), reduces to

$$2ds(\pounds ds) = 2\sigma g_{ij}\,dx^i\,dx^j = 2\sigma ds^2. \tag{3.16}$$

Thus in (3.13) we use $(\pounds ds)/ds = \sigma$ to obtain

$$\pounds K = -2\mu_{;l}\eta^l K = -2[(\pounds\mu)/\mu]K. \tag{3.17}$$

From the definition of μ^2 [See Eq. (3.9)], we have

$$2\mu\mu_{;l} = \xi^j_{;l}\xi_j + \xi^j\xi_{j;l} = 2\xi_{j;l}\xi^j. \tag{3.18}$$

Therefore, from (3.18) we obtain (using the definition of η^i):

$$\begin{aligned}2\mu_{;l}\eta^l = 2\xi_{j;l}\eta^l\eta^j &= (\xi_{j;l} + \xi_{l;j})\eta^l\eta^j \\ &= (\pounds g_{jl})\eta^l\eta^j = 2\sigma.\end{aligned} \tag{3.19}$$

Thus (3.17) becomes

$$\pounds K = -2\sigma K = -\frac{\pounds(ds^2)K}{ds^2} \tag{3.20}$$

and we find that the deformation in Riemannian curvature takes a particularly simple form.

Assuming $K > 0$ in (3.20) we find that, if distances are stretched by the deformation, then the Riemannian curvature decreases.

4. CC IN SPECIAL RIEMANNIAN SPACES

A. V_n with Vanishing Ricci Tensor

Let us now consider those V_n for which

$$R_{ij} = 0. \tag{4.1}$$

We denote such spaces by V^0_n. From the definition of the Weyl projective curvature tensor W^h_{ijk}, [see Eq. (1.9)] we find, for a V^0_n, that

$$W^h_{ijk} = R^h_{ijk}. \tag{4.2}$$

As mentioned in Sec. 1, it can be shown (Ref. 9, p. 134) if a V_n admits a PC then the vector ξ^i defining the PC must satisfy $\pounds_\xi W^h_{ijk} = 0$. Hence, we may state

Theorem 4.1: In a V^0_n (i.e., a V_n with $R_{ij} = 0$) every PC is a CC.

It follows immediately from Theorem 4.1, Theorem 3.5, and Corollary 3.1 that we may state

Corollary 4.1: If a V^0_n admits a PC, then the PC must be an SPC.

Corollary 4.2: If a V^0_n admits a PC, then it also admits a field of parallel vectors.

Next we consider the conformal curvature tensor C^h_{ijk} [see Eq. (1.16)] and observe that in a V^0_n:

$$C^h_{ijk} = R^h_{ijk}. \tag{4.3}$$

As mentioned in Sec. 1, if a V_n admits a Conf C, then the vector ξ^i defining the Conf C must satisfy $\pounds_\xi C^h_{ijk} = 0$. Hence, we may state

Theorem 4.2: In a V^0_n every Conf C is a CC.

It follows immediately from Theorem 4.2, Theorem 3.6, and Corollary 3.2 that we may state

[19] J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, 1949), p. 95; also Ref. 7, p. 79.

Corollary 4.3: If a V_n^0 admits a Conf C, then the Conf C must be a S Conf C.

Corollary 4.4: If a V_n^0 admits a Conf C, then it also admits a field of parallel vectors.

B. Einstein Spaces

Next consider a V_n which is an Einstein space. It then follows (assuming $R \neq 0$) that[20]

$$R_{ij} = (R/n)g_{ij} = (n-1)\kappa g_{ij}, \tag{4.4}$$

where κ is the scalar curvature which is constant for $n > 2$. Let us consider first the case $n > 2$ and assume the space admits a Ricci collineation (RC). It follows from (2.9) and (4.4) that $\pounds g_{ij} = 0$. Hence we may state

Theorem 4.3: Every RC in an Einstein space ($R \neq 0$, $n > 2$) is an M.

Since every space of constant curvature is an Einstein space we have from Theorem 4.3:

Corollary 4.5: Every RC in a space of constant curvature ($R \neq 0$, $n > 2$) is an M.

Since every harmonic space is an Einstein space[21] we obtain from Theorem 4.3:

Corollary 4.6: Every RC in a harmonic space ($R \neq 0$, $n > 2$) is an M.

From Theorem 2.2 we know that in a V_n every CC is a RC. Hence, we have

Corollary 4.7: Every CC in an Einstein space ($R \neq 0$, $n > 2$) is an M.

Corollary 4.8: Every CC in a space of constant curvature ($R \neq 0$, $n > 2$) is an M.

Corollary 4.9: Every CC in a harmonic space ($R \neq 0$, $n > 2$) is an M.

Next we consider the case $n = 2$. Since, for every V_2, Eq. (4.4) is applicable (i.e., every V_2 is an Einstein space, and where R is a scalar) we have from (4.4) and (2.9), assuming the space admits an RC,

$$(\pounds R)g_{ij} + R\pounds g_{ij} = 0. \tag{4.5}$$

Therefore (still assuming $R \neq 0$) we find

$$\pounds g_{ij} = -(\pounds R/R)g_{ij}. \tag{4.6}$$

Hence we have

Theorem 4.4: Every RC in a V_2 is a Conf M (or M as a subcase when $\pounds R = 0$).

It follows immediately by means of Theorem 2.2 that we may state another result.

Corollary 4.10: Every CC in a V_2 is a Conf M (or M as subcase when $\pounds R = 0$).

5. SCC AND FIRST INTEGRALS OF THE GEODESIC EQUATIONS

We now wish to examine more closely the class of CC which we have called SCC. It will be recalled [see Eq. (3.2)] that these symmetries were characterized by the existence of a vector ξ^i for which

$$\left(\pounds_\xi \left\{ {k \atop ij} \right\}\right)_{;h} = 0. \tag{5.1}$$

If the $\pounds\{{k \atop ij}\}$ is expressed in terms of the h_{ij} by use of (1.2), it then follows that (5.1) implies

$$(h_{im;j} + h_{mj;i} - h_{ij;m})_{;k} = 0. \tag{5.2}$$

By interchanging the indices i and m in (5.2) and adding the resulting equation to (5.2), we obtain

$$h_{im;jk} = 0. \tag{5.3}$$

It is easily seen that (5.3) also implies (5.1); hence, we may state

Theorem 5.1: The necessary and sufficient condition for a V_n to admit an SCC (i.e., a CC such that $(\pounds_\xi\{{i \atop jk}\})_{;l} = 0$) is that there exist a vector ξ^i for which $h_{ij;kl} = 0$, where $h_{ij} \equiv \xi_{i;j} + \xi_{j;i}$.

If we multiply (5.3) by g^{im} and sum, we obtain

$$\lambda_{j;k} \equiv \xi^m{}_{;mjk} = 0. \tag{5.4}$$

Hence, we may state

Corollary 5.1: If a V_n admits an SCC then it admits a parallel field of vectors $\lambda_j \equiv \xi^m{}_{;mj}$, where ξ^m is defined by (5.1).

It is well known (see Ref. 7) that the necessary and sufficient condition for the geodesic equation

$$\frac{dp^i}{ds} + \left\{ {i \atop jk} \right\} p^j p^k = 0, \quad p^i \equiv \frac{dx^i}{ds}, \tag{5.5}$$

[20] J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954) 2nd ed., p. 148.

[21] T. J. Willmore, *An Introduction to Differential Geometry* (Clarendon Press, Oxford, 1959), p. 238.

to admit an mth order first integral of the form

$$A_{i_1 i_2 \cdots i_m} p^{i_1} p^{i_2} \cdots p^{i_m} = \text{const}, \tag{5.6}$$

where $A_{i_1 \cdots i_m}$ is assumed symmetric on all indices, is[22]

$$P\{A_{i_1 \cdots i_m ; i_{m+1}}\} = 0. \tag{5.7}$$

Physically, such integrals are of interest in general relativity in that they are constants of the motion for a free mass particle.

For the case $m = 1$ we find from (5.6) and (5.7) that the geodesics have linear first integrals

$$A_{i_1} p^{i_1} = \text{const} \tag{5.8}$$

provided

$$A_{i_1 ; i_2} + A_{i_2 ; i_1} = 0. \tag{5.9}$$

We recognize (5.9) to be Killing's equations or the necessary and sufficient conditions for the V_n to admit a motion [See Eq. (1.1) and Ref. 2].

Similarly for the case $m = 2$ we find from (5.6) and (5.7) that the geodesics have quadratic constants of the motion of the form

$$A_{i_1 i_2} p^{i_1} p^{i_2} = \text{const} \tag{5.10}$$

provided

$$A_{i_1 i_2 ; i_3} + A_{i_3 i_1 ; i_2} + A_{i_2 i_3 ; i_1} = 0. \tag{5.11}$$

It was recently shown (Ref. 3) that PC and AC are basicly related to quadratic first integrals in much the same manner that M are related to linear first integrals.[23]

Taking $m = 3$ in (5.6) and (5.7), we find the geodesics have cubic first integrals of the form

$$A_{i_1 i_2 i_3} p^{i_1} p^{i_2} p^{i_3} = \text{const} \tag{5.12}$$

provided

$$A_{i_1 i_2 i_3 ; i_4} + A_{i_4 i_1 i_2 ; i_3} + A_{i_3 i_4 i_1 ; i_2} + A_{i_2 i_3 i_4 ; i_1} = 0. \tag{5.13}$$

We next define

$$A_{ijk} \equiv h_{ij;k} + h_{ki;j} + h_{jk;i} \tag{5.14}$$

and, by means of (5.3), observe that

$$A_{ijk;m} = 0. \tag{5.15}$$

Hence, we may state

Theorem 5.2: If a V_n admits an SCC with respect to the vector ξ^i, then the geodesics in the V_n admit a cubic first integral of the form $A_{ijk} p^i p^j p^k = \text{const}$, where A_{ijk} is defined by (5.14) and satisfies (5.15).

From Theorem 5.1 and Eqs. (5.4), (5.8), and (5.9), we may write

Theorem 5.3: If a V_n admits an SCC with respect to the vector ξ^i, then the V_n also admits an M (Killing vector) with respect to the vector λ^i, where $\lambda_j \equiv \xi^m{}_{;mj}$ and where $\lambda_{j;k} = 0$. In addition, the geodesics in the V_n admit a linear first integral $\lambda_j p^j = \text{const}$.

In Sec. 3, we found that every SPC is an SCC. Thus, by means (1.7) and (5.14) (assuming $\phi_{;jk} = 0$ as required for the PC to be an SPC),we construct the coefficient A_{ijk} of the cubic first integral which must exist and find

$$A_{ijk} = 4(g_{ij}\phi_{;k} + g_{jk}\phi_{;i} + g_{ki}\phi_{;j}). \tag{5.16}$$

We notice that (5.16) satisfies (5.15) and the resulting cubic first integral reduces to

$$g_{ij}\phi_{;k} p^i p^j p^k = \text{const}. \tag{5.17}$$

By inspection of (5.17) it is observed that this integral may be considered as degenerate[24] in that it is expressible as the product of the fundamental quadratic first integral $g_{ij} p^i p^j = 1$ and the linear first integral $\phi_{;k} p^k = \text{const}$ (which exists because $\phi_{;jk} = 0$).

Also in Sec. 3 we found that every S Conf C is an SCC. Thus, by using (1.14) and (5.14), we construct (using $\sigma_{;ik} = 0$ as required for an S Conf C) the coefficient A_{ijk} of a cubic first integral and obtain

$$A_{ijk} = 2(\sigma_{;i} g_{jk} + \sigma_{;j} g_{ki} + \sigma_{;k} g_{ij}) \tag{5.18}$$

which we observe satisfies (5.15). Thus the resulting cubic first integral reduces to

$$\sigma_{;i} g_{jk} p^i p^j p^k = \text{const}. \tag{5.19}$$

As with the cubic first integral resulting from the SPC, we also find the cubic first integral resulting from the V_n admitting an S Conf C is degenerate in that it too is a product of the fundamental quadratic first integral $g_{jk} p^j p^k = 1$ and the linear first integral $\sigma_{;i} p^i = \text{const}$ (which exists because $\sigma_{;jk} = 0$).

Although we have found that SPC and S Conf C (which we recall are subcases of SCC) lead to degenerate cubic first integrals, we have no reason to suspect that, in general, all cubic first integrals concomitant with the existence of SCC in a V_n are degenerate.

[22] The symbol $P\{\ \}$ indicates the sum of the terms obtained by cyclic permutation of all free indices within the braces.

[23] It has been shown (R. Morgan, M.S. Thesis, North Carolina State University at Raleigh, 1967) that the maximum number of linearly independent quadratic first integrals admitted by a Minkowski space–time may be directly related to the existence of PC and AC by means of the "related integral theorem" [G. Katzin and J. Levine (Ref. 3)]. Further discussion of first integrals in spaces of constant curvature may be found in papers by G. Katzin and J. Levine, Tensor, New Series, **16**, 97 (1965); **18**, 32 (1967); G. Katzin, J. Levine, and J. Halsey; **19**, 42 (1968).

[24] See W. R. Davis and M. K. Moss, Ref. 3.

6. A CONSERVATION LAW GENERATOR CONCOMITANT WITH THE EXISTENCE OF CC IN V_n WITH $R = 0$ AND $R_{ij} \neq 0$

We assume that a V_n with $R = 0$ and $R_{ij} \neq 0$ admits a Ricci collineation (RC), i.e., we assume there exists a vector ξ^i which satisfies (2.9). It therefore follows that

$$\pounds R \equiv \pounds(g^{ik}R_{ik}) = R_{ik}\pounds g^{ik} = 0. \tag{6.1}$$

Since $\pounds g^{ik} \equiv -g^{il}g^{mk}\pounds g_{lm}$, we can express (6.1) in the form

$$R^{lm}\pounds g_{lm} = 0. \tag{6.2}$$

Now, using $\pounds g_{lm} = \xi_{l;m} + \xi_{m;l}$ in (6.2), we obtain

$$R_l^m \xi^l{}_{;m} = 0. \tag{6.3}$$

From the twice-contracted Bianchi identity (see Ref. 7, p. 82) we find (using $R = 0$):

$$R^m_{l;m} = 0. \tag{6.4}$$

Combining (6.4) with (6.3) gives

$$(R_l^m \xi^l)_{;m} = 0. \tag{6.5}$$

In a space-time with $R = 0$, the Einstein field equations take the form

$$R_l^m = \kappa T_l^m, \tag{6.6}$$

where κ is a constant and T_l^m is an energy–momentum tensor with trace $T \equiv T_m^m = 0$. Substituting (6.6) in (6.5) gives

Theorem 6.1: If a space-time V_4, with $R = 0$ and $R_{ij} \neq 0$, admits an RC, then there exists a covariant conservation law generator of the form

$$(\mathfrak{g}^{\frac{1}{2}}T_l^m\xi^l)_{;m} \equiv (\mathfrak{g}^{\frac{1}{2}}\overset{\cdots}{T}_l\,\xi^l)_{,m} = 0, \quad \mathfrak{g} \equiv |\mathrm{Det}\, g_{ij}|, \tag{6.7}$$

where ξ^l is defined by $\pounds_\xi R_{ij} = 0$.

Since we have already found (see Theorem 2.2) that every CC is also an RC it follows that we may state

Corollary 6.1: If a space-time V_4 with $R = 0$ and $R_{ij} \neq 0$ admits a CC, then there exists a covariant conservation law generator of the form (6.7) where ξ^l is defined by $\pounds_\xi R^i_{jkl} = 0$.

We wish to point out that covariant conservation law generators of the form of (6.7) were first discussed by Trautmann[25] who showed they held for ξ^i representing Conf M in space-times with $R = 0$ and for M in space-times with $R \neq 0$.

7. CONSERVATION LAWS FOR NULL FIELDS

In this section, we consider the special case of space–times constituting null-electromagnetic and pure null-gravitational fields that admit CC. It will be shown that conservation laws of the type first directly constructed by Sachs[4] for these null V_4's actually follow from a symmetry argument when CC are admitted.

First the null-electromagnetic case will be considered; the matter tensor is given by[26] $T^{ij} = R^{ij} = \mu k^i k^j$, where μ is a scalar and $g_{ij}k^ik^j = 0$. It follows (see Ref. 4) that there is no loss of generality in taking $k^i{}_{;j}k^j = 0$. Assuming this type of space–time admits a CC and using Corollary 6.1 [cf. Eq. (6.3)] one can write

$$(\mathfrak{g}^{\frac{1}{2}}T^{ij}\xi_j)_{;i} = (\mathfrak{g}^{\frac{1}{2}}\mu k^i k^j \xi_j)_{;i} = (\mathfrak{g}^{\frac{1}{2}}\mu k^i)_{;i}k^j\xi_j = 0. \tag{7.1}$$

Thus, for[27] $k^j\xi_j \neq 0$ one finds

$$(\mathfrak{g}^{\frac{1}{2}}\mu k^i)_{;i} = (\mathfrak{g}^{\frac{1}{2}}\mu k^i)_{,i} = 0, \tag{7.2}$$

which gives the above mentioned conservation law in the form first found by Sachs. Here we have succeeded in relating this conservation law to a symmetry property that may be admitted by null metrics. Before discussing a nontrivial example of a null-electromagnetic metric which admits CC, it will be shown that these considerations can be extended to the case of a pure null field of gravitational radiation.

In accord with the close analogy that exists between null-electromagnetic and null-gravitational fields, it follows that this particular type of V_4^0 is defined by the Bel–Robinson tensor taking the algebraic form[28]

$$T^{ijkm} = 2R^{irjs}R^k{}_r{}^m{}_s = \nu k^i k^j k^k k^m, \tag{7.3}$$

where ν is a scalar and $g_{ij}k^ik^j = 0$. For vacuum spaces V_4^0 ($R_{ij} = 0$) it can be shown that $R_{ijkm} = -R^{**}_{ijkm} \equiv -(\frac{1}{2}\mathfrak{g}^{\frac{1}{2}})\epsilon_{ijrs}(\frac{1}{2}\mathfrak{g}^{\frac{1}{2}})\epsilon_{kmuv}R^{rsuv}$; and in addition, when (7.3) holds, we find $R_{ijrs}R^{rskm} = 0$, $R^*_{ijrs}R^{rskm} = 0$ which, along with the full Bianchi identities, imply

$$T^{ijkm}{}_{;m} = 0. \tag{7.4}$$

[25] A. Trautmann, "Conservation Laws of General Relativity" in *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962); also, see Bull. Acad. Polon. Sci. Cl. III **4**, 675, 679 (1965); **5**, 721 (1957).

[26] See, for example, J. Ehlers and R. Sachs, Z. Physik **115**, 1094 (1959); A. Peres, Phys. Rev. **118**, 1105 (1960).

[27] Here we note that $k^j\xi_j$ cannot, in general, vanish for all the independent curvature collineation vectors (corresponding to distinct sets of values of the independent parameters involved in the solutions ξ^i constituting the CC) which, of course, even include motions as special cases. This matter will be considered further on the basis of the specific example of the next section.

[28] See Ref. 4 and also I. Robinson, Report to the Royaumont Conference 1959 (unpublished).

With the help of these relations it will now be shown that

$$(g^{\frac{1}{2}}T^{ijkm}\xi_k)_{;m} = 0, \tag{7.5}$$

provided the given null metric admits CC with respect to the vector ξ^k. In addition, it will be shown that (7.5) is essentially related to a conservation law of the form $(g^{\frac{1}{2}}vk^i)_{;i} = (g^{\frac{1}{2}}vk^i)_{,i} = 0$ which holds for the case of a pure null-gravitational field. Writing out (7.5) gives

$$(g^{\frac{1}{2}}T^{ijkm}\xi_k)_{;m} = g^{\frac{1}{2}}T^{ijkm}\xi_{k;m} + (g^{\frac{1}{2}}T^{ijkm})_{;m}\xi_k. \tag{7.6}$$

Clearly, by use of (7.4), the last term in (7.6) vanishes. Thus (7.5) holds for CC provided it can be shown that

$$T^{ijkm}\xi_{k;m} \equiv 2R^{irjs}R^{(k\ m)}_{r\ s}\xi_{k;m} = 0. \tag{7.7}$$

Returning to Sec. 2, it will be recalled that a necessary condition for a CC was given by [cf. Eq. (2.12)]:

$$h_{km}R^{m}{}_{rjs} + h_{rm}R^{m}{}_{kjs} = 0,$$

where $h_{km} \equiv \xi_{k;m} + \xi_{m;k}$. If this equation is multiplied by g^{kj} and summed on k and j one finds (using $R_{ij} = 0$):

$$\xi_{k;m}R^{(m\ k)}_{i\ r} = 0, \tag{7.8}$$

for all ξ^k constituting CC of the given space–time.[29] Thus (7.8) shows that (7.7) and thereby (7.5) hold when CC are admitted by pure null-gravitational metrics. Using (7.8) it immediately follows that one can also write

$$(g^{\frac{1}{2}}T^{ijkm}\xi_i\xi_j\xi_k)_{;m} = 0$$

in addition to Eq. (7.5). Hence the following theorem can be stated:

Theorem 7.1: If a null V_4^0 [i.e., a V_4^0 with $T^{ijkm} = 2R^{irjs}R^{k\ m}_{r\ s} = vk^ik^jk^kk^m$, where $k^ik_i = 0$ and v is a nonzero scalar] admits a CC, then the null V_4^0 admits a "conservation law" of the form

$$(g^{\frac{1}{2}}T^{ijkm}\xi_k)_{;m} = 0,$$

or a proper conservation law generator of the form

$$(g^{\frac{1}{2}}T^{ijkm}\xi_i\xi_j\xi_k)_{,m} = (g^{\frac{1}{2}}T^{ijkm}\xi_i\xi_j\xi_k)_{;m} = 0.$$

When CC are admitted by a null V_4^0, this theorem, in turn, implies Sach's conservation law

$$(g^{\frac{1}{2}}vk^m)_{;m} = (g^{\frac{1}{2}}vk^m)_{,m} = 0,$$

for[30] $k^j\xi_j \neq 0$ with $k_{i;j}k^j = 0$. This result immediately follows by expressing (7.5) in the form

$$(g^{\frac{1}{2}}vk^ik^jk^lk^m\xi_l)_{;m} = (g^{\frac{1}{2}}vk^m)_{;m}k^ik^jk^l\xi_l = 0.$$

[29] It is of interest to note that (7.8) is equivalent to the demand $g_{ij}\pounds R^i{}_{rs}{}^j{}_s = 0$ for space-times in which (7.3) holds.

[30] In this connection see Ref. 27 and Ref. 4.

Also, in this connection, it should be mentioned that Komar[31] has recently pointed out that for a V_4^0, $P^m \equiv g^{\frac{1}{2}}T^{ijkm}\xi_i\xi_j\xi_k$ with $P^m{}_{,m} = 0$ gives a constant of the motion which generates a "proper canonical mapping" closely related to that of the linearized theory providing Killing vectors ξ^i are admitted. In terms of Theorem 7.1, it is now seen that this observation due to Komar can be extended to include the more general case[32] of CC for null V_4^0.

8. EXAMPLE OF A V_4 WHICH ADMITS A CC

We now wish to demonstrate the existence of a type of space–time which admits a proper CC, that is, a CC which is not degenerate in the sense of being at the same time a lesser symmetry such as an M or PC, etc. The fundamental form of the space–time we wish to examine is

$$ds^2 = -(dx^1)^2 - (dx^2)^2 - C(u)(dx^3)^2 + (dx^4)^2, \tag{8.1}$$

where $u \equiv x^1 - x^4$. Physically, (8.1) can be interpreted as the line element of a null-electromagnetic (plane-wave) radiation space–time of the type considered, for example, by Sach's (refer to the discussion of Sec. 7[33]). The line element (8.1) is a special case of a more general V_4 which is known to admit a five-parameter group of motions.[34]

A brief outline of the procedure for finding the CC vector ξ^i admitted by (8.1) is next presented. Starting with (2.3), the equations to be solved for the ξ^i can be expressed in the form

$$\pounds R^i_{jkl} = R^i_{jkl,m}\xi^m - R^m_{jkl}\xi^i{}_{,m} + R^i_{mkl}\xi^m{}_{,j} + R^i_{jml}\xi^m{}_{,k} + R^i_{jkm}\xi^m{}_{,l} = 0. \tag{8.2}$$

From the algebraic symmetries on the indices we find that, in a V_4, Eq. (8.2) formally represents 96 equations. Evaluation of these equations by use of the metric tensor defined by (8.1) leads to the following set of equations (redundant and trivial equations have

[31] A. Komar, Phys. Rev. **164**, 1595 (1967).

[32] Recall (Sec. 4) we found that in V_n^0 ($R_{ij} = 0$) that Conf C and PC are also CC. From Fig. 1 we see that this implies that in a null V_4^0 that the extension of Komar's result to include CC actually means that it now is applicable for all the symmetries in the diagram (excluding RC since $R_{ij} = 0$).

[33] The physical properties of this particular V_4 [Eq. (8.1)] were investigated by G. C. Duncan, M.S. Thesis, North Carolina State University at Raleigh, North Carolina, 1966.

[34] Petrov (Ref. 6) has shown that the V_4 with metric

$$ds^2 = 2\,d\bar{x}^1\,d\bar{x}^4 - \bar{\alpha}(\bar{x}^4)(d\bar{x}^2)^2 - 2\bar{\beta}(\bar{x}^4)\,d\bar{x}^2\,d\bar{x}^3 - \gamma(\bar{x}^4)(d\bar{x}^3)^2 \quad \text{(a)}$$

admits a five-parameter group of motions. By making the coordinate transformation $\bar{x}^1 = (1/\sqrt{2})(x^1 + x^4)$, $\bar{x}^2 = x^2$, $\bar{x}^3 = x^3$, $\bar{x}^4 = -(1/\sqrt{2})(x^1 - x^4)$, we obtain (a) in the form

$$ds^2 = -(dx^1)^2 - \alpha(u)(dx^2)^2 - 2\beta(u)\,dx^2\,dx^3 - C(u)(dx^3)^2 + (dx^4)^2. \tag{b}$$

With $\alpha = 1$ and $\beta = 0$, we find (b) reduces to (8.1).

been omitted):

$$\pounds R^3_{112} = 0 \Rightarrow \xi^3{}_{,2} = 0, \tag{8.3}$$

$$\pounds R^2_{113} = 0 \Rightarrow \xi^2{}_{,3} = 0, \tag{8.4}$$

$$\pounds R^1_{113} = 0 \Rightarrow \xi^1{}_{,3} + C\xi^3{}_{,1} = 0, \tag{8.5}$$

$$\pounds R^4_{113} = 0 \Rightarrow \xi^4{}_{,3} + C\xi^3{}_{,1} = 0, \tag{8.6}$$

$$\pounds R^3_{114} = 0 \Rightarrow \xi^3{}_{,1} + \xi^3{}_{,4} = 0, \tag{8.7}$$

$$\pounds R^3_{123} = 0 \Rightarrow \xi^1{}_{,2} - \xi^4{}_{,2} = 0, \tag{8.8}$$

$$\pounds R^2_{313} = 0 \Rightarrow \xi^2{}_{,1} + \xi^2{}_{,4} = 0, \tag{8.9}$$

$$\pounds R^1_{413} = 0 \Rightarrow \xi^1{}_{,3} - C\xi^3{}_{,4} = 0, \tag{8.10}$$

$$\pounds R^4_{434} = 0 \Rightarrow \xi^4{}_{,3} - C\xi^3{}_{,4} = 0, \tag{8.11}$$

$$\pounds R^3_{334} = 0 \Rightarrow \xi^1{}_{,3} - \xi^4{}_{,3} + C(\xi^3{}_{,1} + \xi^3{}_{,4}) = 0, \tag{8.12}$$

$$\pounds R^3_{113} = 0 \Rightarrow 2(\xi^1 - \xi^4)_{,1} + [\ln (C^{-1}R_{1313})]_{,u}(\xi^1 - \xi^4) = 0, \tag{8.13}$$

$$\pounds R^3_{434} = 0 \Rightarrow -2(\xi^1 - \xi^4)_{,4} + [\ln (-C^{-1}R_{1313})]_{,u}(\xi^1 - \xi^4) = 0, \tag{8.14}$$

$$\pounds R^3_{134} = 0 \Rightarrow (\xi^1 - \xi^4)_{,1} - (\xi^1 - \xi^4)_{,4} + [\ln (-C^{-1}R_{1313})]_{,u}(\xi^1 - \xi^4) = 0, \tag{8.15}$$

$$\pounds R^1_{313} = 0 \Rightarrow \xi^1{}_{,4} + \xi^4{}_{,1} - 2\xi^3{}_{,3} - [\ln R_{1313}]_{,u}(\xi^1 - \xi^4) = 0, \tag{8.16}$$

$$\pounds R^4_{313} = 0 \Rightarrow \xi^1{}_{,1} - 2\xi^4{}_{,1} - \xi^4{}_{,4} + 2\xi^3{}_{,3} + [\ln R_{1313}]_{,u}(\xi^1 - \xi^4) = 0, \tag{8.17}$$

$$\pounds R^1_{334} = 0 \Rightarrow \xi^1{}_{,1} + 2\xi^1{}_{,4} - \xi^4{}_{,4} - 2\xi^3{}_{,3} - [\ln R_{1313}]_{,u}(\xi^1 - \xi^4) = 0, \tag{8.18}$$

and

$$R_{1313} = -\tfrac{1}{2}C^{\frac{1}{2}}(C^{-\frac{1}{2}}C_{,u})_{,u}.$$

By inspection, we find the following relations exist between equations of the set (8.3)–(8.18):

$$2[\text{Eq. (8.15)}] - \text{Eq. (8.13)} - \text{Eq. (8.14)} = 0, \tag{8.19}$$

$$\text{Eq. (8.5)} - \text{Eq. (8.6)} - \text{Eq. (8.8)} + \text{Eq. (8.9)} = 0, \tag{8.20}$$

$$\text{Eq. (8.17)} - \text{Eq. (8.18)} + 2[\text{Eq. (8.16)}] = 0. \tag{8.21}$$

$$\text{Eq. (8.13)} - \text{Eq. (8.14)} - 2[\text{Eq. (8.16)}] - 2[\text{Eq. (8.17)}] = 0, \tag{8.22}$$

$$\text{Eq. (8.5)} - \text{Eq. (8.10)} - C^{-1}\,\text{Eq. (8.9)} = 0. \tag{8.23}$$

By means of (8.19)–(8.23) we may eliminate Eqs. (8.1), (8.11), (8.15), (8.17), and (8.18) from the set to be solved.

From (8.3) and (8.7) we find $\xi^3 = \xi^3(u, x^3)$. Equations (8.4) and (8.9) show that $\xi^2 = \xi^2(u, x^2)$. Next we define $\mu \equiv \xi^1 - \xi^4$ and by use of (8.8), (8.7), (8.12), (8.13), and (8.14) we obtain $\mu = \mu(u)$.

By assuming $\mu \neq 0$ (i.e., $A \neq 0$) we obtain from (8.13) that $\mu = AT^{-\frac{1}{2}}$ where A is an arbitrary constant and $T \equiv C^{-1}R_{1313}$. Also from the form of μ we observe that (8.5) and (8.6) are equivalent. Using $\xi^4 = \xi^1 - \mu$ we express (8.16) in the form

$$\xi^1{}_{,4} + \xi^1{}_{,1} - 2\xi^3{}_{,3} = M(u), \tag{8.24}$$

where $M(u) = (\ln R_{1313})_{,u}\mu + \mu_{,u}$. Integrating (8.5) with respect to x^3 gives $\xi^1 = \phi(u, x^3) + \psi(x^1, x^2, x^4)$, where ϕ and ψ are arbitrary functions. With this expression for ξ^1 substituted into (8.24) we are able to integrate (8.24) and find $\xi^3 = D(u)x^3 + E(u)$, where D and E are arbitrary. Substitution of this value of ξ^3 back into (8.5) allows (8.5) to be integrated giving $\xi^1 = -C(u)[\frac{1}{2}D_{,u}(x^3)^2 + E_{,u}x^3] + F(x^1, x^2, x^4)$, where F is arbitrary. Then substitution of ξ^1 and ξ^3 into (8.24) shows that $F = \frac{1}{2}(x^1 + x^4)[M(u) + 2D(u)] + P(u, x^2)$, where P is arbitrary. We now summarize the components of the CC vector for the case $\mu \neq 0$:

$$\xi^1 = -C(u)[\tfrac{1}{2}D(u)_{,u}(x^3)^2 + E(u)_{,u}x^3] + \tfrac{1}{2}(x^1 + x^4)[2D(u) + M(u)] + P(u, x^2), \tag{8.25a}$$

$$\xi^2 = \xi^2(u, x^2), \tag{8.25b}$$

$$\xi^3 = D(u)x^3 + E(u), \tag{8.25c}$$

$$\xi^4 = \xi^1 - A[R_{1313}/C(u)]^{-\frac{1}{2}}, \tag{8.25d}$$

where A is an arbitrary constant; $D(u)$, $E(u)$, and $P(u, x^2)$ are arbitrary functions; and

$$M(u) \equiv A\left\{\left(\frac{R_{1313}}{C(u)}\right)^{-\frac{1}{2}}(\ln R_{1313})_{,u} + \left[\left(\frac{R_{1313}}{C(u)}\right)^{-\frac{1}{2}}\right]_{,u}\right\}.$$

For the case in which $\mu = 0$, i.e., $A = 0$, which implies $\xi^1 = \xi^4$, we find the solution (8.25) is still valid.[35]

As a check on the solution we consider the necessary conditions (2.12) which for the metric (8.1) can be expressed as

$$\begin{aligned} h_{31} &= 0, \\ h_{32} &= 0, \\ h_{34} &= 0, \\ h_{11} - C^{-1}h_{33} + h_{41} &= 0, \\ h_{12} + h_{42} &= 0, \\ h_{44} + C^{-1}h_{33} + h_{41} &= 0. \end{aligned} \tag{8.26}$$

By use of (8.25) and the definition $h_{ij} = \xi_{i;j} + \xi_{j;i}$, we verify that Eqs. (8.26) are satisfied.

[35] In accord with the comments of Sec. 7 (see Ref. 27), it is easily seen for the space–time (8.1) that $k_i\xi^i = 0$, where $k_i = (1, 0, 0, -1)$ only in the very special case where the parameter $A = 0$ in (8.25).

We established that the vector ξ^1 is, in general, not a motion vector or affine collineation vector by showing that, in general, $h_{ij} \neq 0$ and $h_{ij;k} \neq 0$.

A simple calculation shows that, in general, $\xi^i{}_{;ijk} \neq 0$. Thus by Theorems 3.5 and 3.6 we conclude that in general ξ^i does not define a Conf M (including HM) or PC.

It follows that for special choices of the arbitrary functions that ξ can be forced to be a motion vector as expected (see Ref. 34).

Note Added in Proof: We have recently shown ("Curvature Collineations in Conformally Flat Spaces. I," submitted to Tensor, New Series) that the Einstein static cosmological space–time admits both proper CC and proper SCC. In addition, it was found that a class of field conservation laws could be constructed as a consequence of this space–time simultaneously, admitting both a proper SCC and a geodesic congruence with vanishing expansion, rotation, and shear. This is in accord with the symmetry approach to the formulation of conservation laws where it can be shown that there are fundamental connections between field conservation laws associated with the kinematic properties of curve congruences and the symmetry properties admitted by the given space–time, as already suggested by the results of Sec. 7. We plan to publish the results of further investigations in this area in the near future.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Renormalization of a Finite Matrix Hamiltonian*

ASHER PERES†
Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland College Park, Maryland

(Received 5 October 1968)

We investigate the eigenvalues of a finite matrix Hamiltonian $H = H_0 + g_0V$, where H_0 is diagonal with eigenvalues $1, 2, \cdots, N$, and where all the elements of V are equal to 1. We are interested in the case $N \to \infty$. The radius of convergence of the perturbation series is $(\ln N)^{-1}$, but nevertheless the exact eigenvalues of H tend to well-defined limits when $N \to \infty$. It is shown that if we define $g = (g_0^{-1} + \ln N)^{-1}$ and if we let $g_0 \to 0$ as $N \to \infty$ in such a way that g is constant, then it is possible to obtain a perturbation series with the "renormalized" coupling constant g, provided that suitable counter terms are introduced. We also investigate a different model (where $V_{mn} = mn$) and show that no such renormalization is possible there.

I. INTRODUCTION AND SUMMARY

The purpose of this paper is not to produce another new theory, but only to clarify the concept of renormalization by means of an elementary algebraic model.

The model consists in finding the eigenvalues[1] of a sequence of finite matrices H which we write as

$$H = H_0 + g_0V. \tag{1}$$

Here, H_0 represents a truncated harmonic oscillator[2]

$$H_0 = \begin{pmatrix} 1 & & & \\ & 2 & & \\ & & \ddots & \\ & & & N \end{pmatrix} \tag{2}$$

and V is also an $N \times N$ matrix which has the property that V^2 diverges (i.e., is not defined) when $N \to \infty$.

For instance, in Sec. II we take $V_{mn} = 1$ (for all m and n) so that $V^2 = NV$ has no limit for $N \to \infty$.[3]

We first investigate the "energy levels" in perturbation theory: in first order, they are all shifted by g_0, but the second-order shifts behave as $-g_0^2 \ln N$, i.e., diverge for $N \to \infty$. On the other hand, an exact treatment of the problem shows that all the energy levels, except possibly one, tend to the positive integers for $N \to \infty$, if g_0 is kept fixed and finite.

However, it is shown in Sec. III that if we let $g_0 \to 0$ in such a way that

$$g_0^{-1} + \ln N = g^{-1}, \tag{3}$$

remains constant, then all the energy levels (and corresponding eigenvectors) tend to fixed nontrivial values for $N \to \infty$. Moreover, for small g, these

* Supported in part by NASA Grant No. NSG-436.

† Permanent address: Department of Physics, Technion–Israel Institute of Technology, Haifa, Israel.

[1] In a recent paper, C. M. Bender and T. T. Wu [Phys. Rev. Letters 21, 406 (1968)] have also investigated the eigenvalues of a model Hamiltonian.

[2] H. A. Buchdahl, Am. J. Phys. 35, 210 (1967).

[3] This property is reminiscent of the divergent behavior in quantum field theory of $\mathcal{H}(x)\mathcal{H}(y)$ when $x \to y$.

We established that the vector ξ^1 is, in general, not a motion vector or affine collineation vector by showing that, in general, $h_{ij} \neq 0$ and $h_{ij;k} \neq 0$.

A simple calculation shows that, in general, $\xi^i{}_{;ijk} \neq 0$. Thus by Theorems 3.5 and 3.6 we conclude that in general ξ^i does not define a Conf M (including HM) or PC.

It follows that for special choices of the arbitrary functions that ξ can be forced to be a motion vector as expected (see Ref. 34).

Note Added in Proof: We have recently shown ("Curvature Collineations in Conformally Flat Spaces. I," submitted to Tensor, New Series) that the Einstein static cosmological space–time admits both proper CC and proper SCC. In addition, it was found that a class of field conservation laws could be constructed as a consequence of this space–time simultaneously, admitting both a proper SCC and a geodesic congruence with vanishing expansion, rotation, and shear. This is in accord with the symmetry approach to the formulation of conservation laws where it can be shown that there are fundamental connections between field conservation laws associated with the kinematic properties of curve congruences and the symmetry properties admitted by the given space–time, as already suggested by the results of Sec. 7. We plan to publish the results of further investigations in this area in the near future.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Renormalization of a Finite Matrix Hamiltonian*

ASHER PERES†
Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland College Park, Maryland

(Received 5 October 1968)

We investigate the eigenvalues of a finite matrix Hamiltonian $H = H_0 + g_0V$, where H_0 is diagonal with eigenvalues $1, 2, \cdots, N$, and where all the elements of V are equal to 1. We are interested in the case $N \to \infty$. The radius of convergence of the perturbation series is $(\ln N)^{-1}$, but nevertheless the exact eigenvalues of H tend to well-defined limits when $N \to \infty$. It is shown that if we define $g = (g_0^{-1} + \ln N)^{-1}$ and if we let $g_0 \to 0$ as $N \to \infty$ in such a way that g is constant, then it is possible to obtain a perturbation series with the "renormalized" coupling constant g, provided that suitable counter terms are introduced. We also investigate a different model (where $V_{mn} = mn$) and show that no such renormalization is possible there.

I. INTRODUCTION AND SUMMARY

The purpose of this paper is not to produce another new theory, but only to clarify the concept of renormalization by means of an elementary algebraic model.

The model consists in finding the eigenvalues[1] of a sequence of finite matrices H which we write as

$$H = H_0 + g_0V. \tag{1}$$

Here, H_0 represents a truncated harmonic oscillator[2]

$$H_0 = \begin{pmatrix} 1 & & & \\ & 2 & & \\ & & \ddots & \\ & & & N \end{pmatrix} \tag{2}$$

and V is also an $N \times N$ matrix which has the property that V^2 diverges (i.e., is not defined) when $N \to \infty$.

For instance, in Sec. II we take $V_{mn} = 1$ (for all m and n) so that $V^2 = NV$ has no limit for $N \to \infty$.[3]

We first investigate the "energy levels" in perturbation theory: in first order, they are all shifted by g_0, but the second-order shifts behave as $-g_0^2 \ln N$, i.e., diverge for $N \to \infty$. On the other hand, an exact treatment of the problem shows that all the energy levels, except possibly one, tend to the positive integers for $N \to \infty$, if g_0 is kept fixed and finite.

However, it is shown in Sec. III that if we let $g_0 \to 0$ in such a way that

$$g_0^{-1} + \ln N = g^{-1}, \tag{3}$$

remains constant, then all the energy levels (and corresponding eigenvectors) tend to fixed nontrivial values for $N \to \infty$. Moreover, for small g, these

* Supported in part by NASA Grant No. NSG-436.

† Permanent address: Department of Physics, Technion–Israel Institute of Technology, Haifa, Israel.

[1] In a recent paper, C. M. Bender and T. T. Wu [Phys. Rev. Letters 21, 406 (1968)] have also investigated the eigenvalues of a model Hamiltonian.

[2] H. A. Buchdahl, Am. J. Phys. 35, 210 (1967).

[3] This property is reminiscent of the divergent behavior in quantum field theory of $\mathcal{H}(x)\mathcal{H}(y)$ when $x \to y$.

values are *the same as we would compute from first-order perturbation theory* with $H = H_0 + gV$. We can therefore call g the "renormalized coupling constant." Higher orders can also be computed by introducing infinite counter-terms, just as in quantum field theory.[4]

In Sec. IV we consider a different "interaction," namely $V_{mn} = mn$. Again, the second-order perturbation diverges for $N \to \infty$ and fixed g_0. But in this case there is no possibility of introducing a renormalized coupling constant such that $H = H_0 + gV$ would simulate, in first order, the results of the exact theory. This Hamiltonian is not renormalizable.

II. THE MODEL

Consider $H = H_0 + g_0V$, where H_0 is given by Eq. (2) and V is a $N \times N$ matrix, all the elements of which are equal to 1. We have $V^2 = NV$, so that the eigenvalues of V are N and 0. The normalized eigenvectors of V are conveniently taken as

$$N^{-\frac{1}{2}}(\alpha, \alpha^2, \cdots, \alpha^N), \tag{4}$$

where

$$\alpha = e^{2\pi ik/N}, \quad k = 0, 1, \cdots, N-1. \tag{5}$$

The eigenvalue N of V is obtained when we set $k = 0$. All the other values of k yield the eigenvalue 0.

The scalar product of any eigenvector of H_0 with any of the eigenvectors (4) is equal to $N^{-\frac{1}{2}}$. When $N \to \infty$, these two *complete* sets of eigenvectors become, so to speak, orthogonal to each other. As we shall see in the sequel, this property is closely akin to the well-known Van Hove orthogonality in quantum field theory.[5]

We now consider g_0 as a small parameter and expand the eigenvalues of H in a perturbation series. We obtain

$$E_k = k + g_0 - g_0^2 \sum{}' (n-k)^{-1} + \cdots. \tag{6}$$

Obviously, the second-order perturbation diverges logarithmically for $N \to \infty$. (The *difference* between energy levels remains finite in second order, but not in third order, as can be seen by a straightforward calculation.[6])

On the other hand, it is not difficult to obtain the exact solution of $H\Psi = E\Psi$. We set

$$\Psi = \sum a_n u_n, \tag{7}$$

where $H_0u_n = nu_n$ and $Vu_n = \sum u_m$. We obtain, by virtue of the orthogonality of the u_n,

$$na_n + g_0 \sum a_m = Ea_n \tag{8}$$

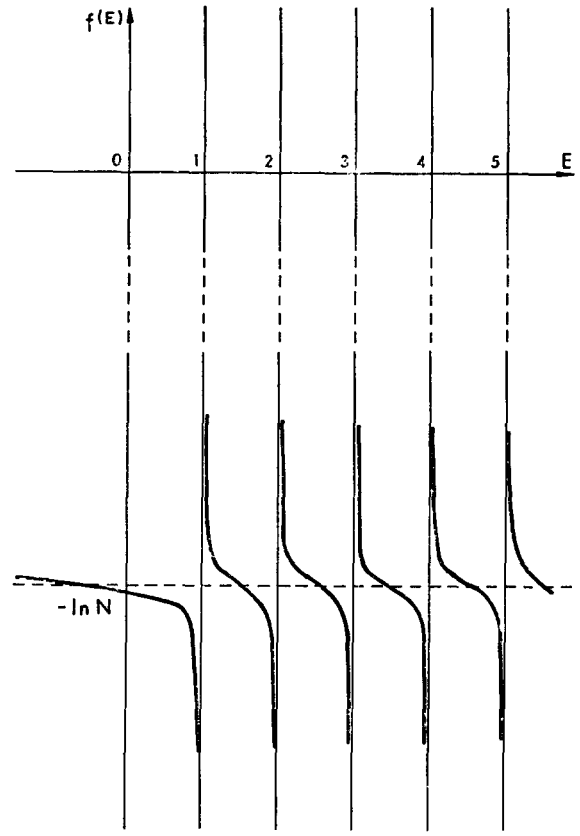


FIG. 1. Qualitative shape of the function $f(E) = \sum(E-n)^{-1}$ for large N.

or

$$a_n = g_0(E-n)^{-1} \sum a_m. \tag{9}$$

Summing over n, this gives the algebraic equation

$$\sum (E-n)^{-1} = g_0^{-1}. \tag{10}$$

The qualitative shape of the function

$$f(E) \equiv \sum (E-n)^{-1} \tag{11}$$

is given in Fig. 1 for large N. (The eigenvalues of H are obtained by drawing a horizontal line at height $1/g_0$.) Obviously, the graph has vertical asymptotes at $E = 1, 2, \cdots, N$. Moreover, for small $|E|$ (i.e., $|E| \ll N$), the curve can deviate appreciably from these asymptotes only around $f(E) = -\ln N$. This is easily seen from Eq. (11): if E is not too close to an integer, then $f(E) + \ln N$ must be finite.

If E is close to some finite integer k, then (10) gives

$$(E-k)^{-1} = g_0^{-1} + \ln N + \text{finite terms} \tag{12}$$

or

$$E = k + g_0/[1 + g_0(\ln N + \cdots)]. \tag{13}$$

This result can now be expanded in powers of g_0, whence it is readily seen that the radius of convergence of the perturbation series is $(\ln N)^{-1}$ and tends to zero

[4] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959), p. 376.

[5] L. Van Hove, Physica **18**, 145 (1952).

[6] See, e.g., L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon Press, Ltd., London, 1965), p. 132.

for $N \to \infty$. In other words, the eigenvalues of H become nonanalytic at $g_0 = 0$ when $N \to \infty$.

This is seen most strikingly if we consider the ground state of H, for fixed negative g_0 and $N \to \infty$. [All the other states are simply shifted by $(\ln N)^{-1}$, as seen from Eq. (13).[7]] Let us write (10) as

$$-(g_0)^{-1} = (1 - E)^{-1} + (2 - E)^{-1} + \cdots + (N - E)^{-1} \tag{14}$$

$$= (1 + |E|)^{-1} + (2 + |E|)^{-1} + \cdots + (N + |E|)^{-1} \tag{15}$$

$$\simeq -\ln |E| + \ln (N + |E|). \tag{16}$$

It follows that

$$E \simeq -N/(e^{-1/g_0} - 1) \simeq -Ne^{1/g_0} \tag{17}$$

which is conspicuously nonanalytic at g_0.[7,8]

A further insight into this situation is obtained by considering the eigenvectors of H. From (9), we have

$$a_n/a_m = (E - m)/(E - n) \tag{18}$$

so that, whenever E is very close to some integer, the corresponding Ψ is very close to the unperturbed Ψ even though perturbation theory is not valid.[7]

However, for the ground state which we considered above, we have

$$a_n = a_1(|E| + 1)/(|E| + n), \tag{19}$$

whence

$$\sum |a_n|^2 = 1 = |a_1|^2 (|E| + 1)^2 \sum (|E| + n)^{-2}. \tag{20}$$

A rough estimate of $\sum (|E| + n)^2$ is

$$\int_1^\infty (|E| + n)^{-2}\, dn = (|E| + 1)^{-1},$$

so that

$$|a_1|^2 \simeq (|E| + 1)^{-1} \simeq N^{-1}e^{-1/g_0}. \tag{21}$$

We see that a_1 is very small and, from (19), the other a_n are even smaller: the ground state of the perturbed Hamiltonian is "almost orthogonal" to all the eigenstates of the free Hamiltonian.[5]

III. RENORMALIZATION

We have seen that, when g_0 is held fixed and $N \to \infty$, all the finite eigenvalues of H tend to those of H_0. To obtain a nontrivial theory, we must, therefore, let g_0 tend to zero as $N \to \infty$ in such a way that

$$g_0^{-1} + \ln N \equiv g^{-1} \tag{22}$$

remains finite. (This is easily seen from Fig. 1.)

First, we note from Eq. (13) that, if g is small, all the eigenvalues of H can be written as $k + g$, i.e., they are correctly given by first-order perturbation theory, *as if* we had $H = H_0 + gV$.

More precisely, consider

$$H = H_0 + g(1 - g \ln N)^{-1} V \tag{23}$$

which is exactly equivalent to (1). Formally, we can write

$$H = H_0 + gV + g^2 \ln NV + g^3 (\ln N)^2 V + \cdots. \tag{24}$$

To first order in g, we have $E = k + g$, which is correct as we have seen. In the next order, we have two contributions: gV contributes

$$-g^2 \sum{}' (n - k)^{-1}, \tag{25}$$

as in Eq. (6), and $g^2 \ln NV$ gives a shift

$$+g^2 \ln N \tag{26}$$

for all energy levels. The total shift is

$$g^2[\ln N - \sum{}' (n - k)^{-1}], \tag{27}$$

which is *finite*.

The important point is that the cancellation of the divergence in (25) is due to the next term of the *divergent* series (24).[9] What we have actually done in (24) is to find, in a single stroke, *all the counter terms*[4] needed to cancel the ultraviolet divergences in all orders. Note that all these counter-terms are proportional to V (no other matrix is needed): we only need a renormalization of the coupling constant.[10]

We can thus obtain the energy levels as a power series in g. When $|g|$ becomes large, $1/g_0$ tends to $-\ln N$ and all the eigenvalues tend to well-defined limits. It would be interesting to see whether the power series in g has an infinite radius of convergence or is only asymptotic (a finite radius of convergence seems unlikely). These questions, however, are beyond the scope of this paper.

Anyway, we have seen that for small enough g, the exact theory (with g_0) can be *simulated* by a few

[7] A closer look at Fig. 1 shows that the other states are actually shifted by $-1 + (\ln N)^{-1}$ if $g_0 < 0$, and by $(\ln N)^{-1}$ if $g_0 > 0$. This is also not analytic at $g_0 = 0$.

[8] This is a rather delicate point, because (17) was derived by taking $N \to \infty$ for fixed negative g_0 and holds only if $-g_0 \gg (\ln N)^{-1}$. The nonanalyticity of the perturbation series in quantum field theory was first suggested by F. J. Dyson, Phys. Rev. **85**, 631 (1952). For other nonanalytic models, see Ref. 1, and A. Peres, J. Math. Phys. **4**, 332 (1963).

[9] On the other hand, the expansion of (23) into a convergent series (of negative powers of $g \ln N$) would be completely useless.

[10] As usual, the counter terms are not uniquely defined by the form of Eq. (1) and we can add to them arbitrary *finite* multiples of V. This, however, merely amounts to altering the value of g by a finite amount, i.e., to a redefinition of g. To obtain an unambiguous definition of g, we may set, for example, $E_1 = 1 + g$ *exactly*, and this fixes all the other E_k (as seen in Fig. 1).

perturbation terms of the "phenomenological" theory (with g). To complete the discussion, we must still show that this result holds not only for the eigenvalues of H, but also for the corresponding eigenvectors.

From Eq. (9), we have

$$1 = \sum |a_n|^2 = |g_0 \sum a_m|^2 \sum (E - n)^{-2}. \quad (28)$$

But $\sum (E - n)^{-2}$ converges and, if E is close to some integer k, is dominated by $(E - k)^{-2}$. It follows that $|g_0 \sum a_m| \simeq |E - k| \simeq |g|$. Substitution in (9) then gives (with proper account of signs):

$$a_n \simeq g/(k - n), \quad n \neq k, \quad (29)$$

in agreement with first-order perturbation theory.

The extension of this result to higher orders is left as an exercise for the skeptical reader.

IV. A NONRENORMALIZABLE MODEL

We now consider a different "interaction," namely $V_{mn} = mn$. This interaction is more singular than the previous one (the matrix elements grow faster) and more serious difficulties can be expected.

Equation (18) now becomes

$$na_n + g_0 n \sum m a_m = E a_n \quad (30)$$

or

$$na_n = g_0 n^2 (E - n)^{-1} \sum m a_m. \quad (31)$$

Summing over n, this gives

$$g_0^{-1} = \sum n^2/(E - n) \quad (32)$$

$$= E^2 f(E) - \tfrac{1}{2} N(N + 1) - NE, \quad (33)$$

where $f(E)$ is given by (11). We can also write (33) as

$$f(E) = [g_0^{-1} + \tfrac{1}{2} N(N + 1) + NE]/E^2, \quad (34)$$

so that the eigenvalues of H can easily be obtained as the points where the graph of Fig. 1 intersects the graph of $[g_0^{-1} + \frac{1}{2}N(N + 1) + NE]/E^2$. It is obvious that, if g_0 is held fixed and $N \to \infty$, all the eigenvalues of H simply coincide with the positive integers.

If we are interested in the nontrivial case where at least one eigenvalue of H is not an integer, we must make $g_0 \to 0$ in some definite way as in Sec. II. Let W be such an eigenvalue. The behavior of g_0 for large N is given by

$$g_0^{-1} = W^2 f(W) - \tfrac{1}{2} N(N + 1) - NW \quad (35)$$

and (34) can be rewritten as

$$f(E) = [(E - W)N + W^2 f(W)]/E^2. \quad (36)$$

Let us now investigate the other eigenvalues of H. The function on the rhs of (36) has, at the point $E = W$, a slope

$$NW^{-2} - 2f(W)W^{-1} \quad (37)$$

which tends to infinity for large N, because $f(W) \simeq -\ln N$ if W differs appreciably from an integer. It then follows from Fig. 1 that all the other eigenvalues of H are infinitely close to the positive integers when $N \to \infty$: *if one energy level has a finite shift, all the energy levels below it are not shifted at all, and all the energy levels above it are shifted by* -1.

Obviously, if we attempt to construct a "phenomenological" Hamiltonian giving the above result in first-order perturbation theory, it will bear no resemblance to the true Hamiltonian (for which perturbation theory gives energy shifts $g_0 k^2$, in first order). The theory is not renormalizable.[11]

V. OUTLOOK

In this paper, we have considered two models of finite matrix Hamiltonians in the limit when the order of the matrices tends to infinity.

We have found that the eigenvalues of the Hamiltonian (23) tend to well-defined (and nontrivial) limits which can be computed by perturbation theory, even though the series (24) diverges.

On the other hand, we have found no possibility of treating the model of Sec. IV by perturbation theory. (More precisely, we have shown that a phenomenological theory giving the same eigenvalues would bear no resemblance to the exact theory.)

This paper sheds, therefore, no light on the mystery of higher-order terms in nonrenormalizable interactions.

ACKNOWLEDGMENTS

It is a pleasure for the author to express his gratitude to Professor J. Weber for the warm hospitality shown to him at the University of Maryland. A friendly discussion with Professor O. W. Greenberg was helpful to clarify some points.

[11] Formally, this is due to the difference between (35) and the corresponding equation $g_0^{-1} = f(W)$ which holds for the previous model. As $N \to \infty$, the infinite part of $f(W)$, namely $-\ln N$, is independent of W, so that the same renormalization of g_0 can be used for all energy levels. This is no longer possible for the second model, which is therefore nonrenormalizable.

Massive Vector Meson Interacting with the Gravitational Field. I. General Formalism

G. E. TAUBER
Department of Physics, Tel-Aviv University, Tel-Aviv, Israel

(Received 10 October 1968)

Starting from the covariant field equations for a vector meson, the energy–momentum tensor entering in Einstein's field equations is derived. It is shown that its most general algebraic form involves two vector fields and two scalars. Specifying the formalism to the special cases for which the fields are either parallel or perpendicular to each other, it is found that the vector field cannot be described in terms of a perfect fluid involving only density and pressure, but includes an additional term involving the stresses. The conservation laws are given, which, in addition to the ones of relativistic hydrodynamics, also include the ones describing the streaming of the vector field.

1. INTRODUCTION

The description of a vector-meson field within the framework of general relativity is of interest in astrophysics whenever large gravitational fields and high densities are involved.[1] Some time ago, Zel'dovich[2] showed that a classical vector field interacting with stationary point charges can produce the most rigid equation of state possible, which is compatible with relativity, of the form

$$p/c^2 = \rho, \tag{1.1}$$

where p denotes the pressure and ρ the density.

It would, therefore, also be of interest to study a vector-meson field of finite mass interacting with the gravitational field and see what conclusions can be drawn about the effective equation of state. The problem is similar to the geometrization of the Maxwell equations which impose certain restrictions on the gravitational field.[3] In this case, too, the form of the energy–momentum tensor (2.10)—which contains a part identical in form to that of the Maxwell field—results in certain limitations on the gravitational field and the equation of state. In particular, it is found (cf. Sec. 3) that a massive vector meson cannot be described by a perfect fluid in terms of a pressure and density only, but involves at least an additional term describing stresses [cf. Eq. (4.17)].

In this paper, we develop the general formalism for a vector field in the presence of a gravitational field. Starting with the field equations for the vector meson (Sec. 2), the energy–momentum tensor is derived as usual from an invariant Lagrangian. By identifying this tensor with the one for a fluid and using the identities applying to the Maxwell tensor, we find (Sec. 3) the most general (algebraic) form involving two vector fields, the velocity field u^μ and the vector field A^μ, as well as two scalars related to the density and the pressure. In Sec. 4, two particular cases are studied, one for which the two fields are parallel and one for which they are orthogonal. It is here shown that the description as a fluid involves an additional term. Finally, in Sec. 5, the conservation laws and field equations are given. It is found that, in addition to the ones of ordinary relativistic hydrodynamics, another set appears describing the streaming of the vector field.

2. FIELD EQUATIONS

In ordinary field theory, the field equations for a vector meson of mass m are given by

$$\Box A_\nu = m^2 A_\nu, \tag{2.1}$$

together with the subsidiary conditions

$$\partial_\nu A^\nu = 0, \tag{2.1'}$$

where

$$\partial_\nu = \partial/\partial x^\nu, \quad \Box = \partial_\nu \partial^\nu,$$

and in our units $c = \hbar = 1$.

These two sets of equations can be combined into one,

$$[(\Box - m^2)\epsilon_{\mu\nu} - \partial_\mu \partial_\nu] A^\nu = 0, \tag{2.2}$$

and it can be seen easily that the subsidiary conditions (2.1') follow from (2.2) by a further differentiation with respect to x^μ.

The situation in general relativity is slightly more complicated owing to the fact that the covariant derivatives do not commute. In fact, we have

$$D_\mu D_\nu A^\mu - D_\nu D_\mu A^\mu = R_{\nu\lambda} A^\lambda, \tag{2.3}$$

[1] Cf. V. A. Ambartsurmayan and G. S. Saakyan, Astron. Zh. 37, 193 (1960) [Sov. Astron.—AJ 4, 187 (1960)]; A. G. W. Cameron, Astrophys. J. 130, 884 (1959); E. E. Salpeter, Ann. Phys. (N.Y.) 11, 393 (1960).

[2] Ya. B. Zel'dovich, Zh. Eksp. Teor. Fiz. 41, 1609 (1961) [Sov. Phys. JETP 14, 1143 (1962)].

[3] G. Y. Rainich, Trans. Am. Math. Soc. 27, 106 (1925).

where $R_{\nu\lambda}$ is the Ricci tensor and D_μ denotes the covariant derivative. A simple calculation shows that (2.2) has to be replaced by[4]

$$[(D_\alpha D^\alpha - m^2)g_{\mu\nu} - D_\nu D_\mu]A^\nu = 0. \tag{2.4}$$

The covariant subsidiary condition

$$D_\mu A^\mu = 0 \tag{2.5}$$

then follows by operating with D_μ on (2.4). Since the covariant derivatives of $g^{\mu\nu}$ are zero, we can write

$$D^\mu D_\alpha D^\alpha g_{\mu\nu} A^\nu = g^{\lambda\nu} g^{\alpha\beta} D_\nu D_\alpha D_\beta A_\lambda .$$

Applying now the commutation relation (2.3) to the tensor $D_\beta A_\lambda$, we obtain

$$D_\nu D_\alpha (D_\beta A_\lambda) = D_\alpha D_\nu (D_\beta A_\lambda) + R^{\sigma}_{\cdot\beta\nu\alpha} D_\sigma A_\lambda + R^{\sigma}_{\cdot\lambda\nu\alpha} D_\beta A_\sigma$$

and, hence,

$$D^\mu D_\alpha D^\alpha A_\mu = D^\alpha D_\mu D_\alpha A^\mu$$

as the last two terms cancel on account of the antisymmetry of the Riemann–Christoffel tensor

$$g^{\lambda\nu} g^{\alpha\beta} R^{\sigma}_{\cdot\beta\nu\alpha} = -R^{\sigma\lambda}, \quad g^{\lambda\nu} g^{\alpha\beta} R^{\sigma}_{\cdot\lambda\nu\alpha} = R^{\sigma\beta}.$$

Thus, so long as $m \neq 0$, the first and the third term in (2.4), after operating with D_μ, cancel and (2.5) follows.

Imposing now the subsidiary condition and making use of the commutation relations (2.3), the field equations (2.4) become

$$(D_\alpha D^\alpha g_{\mu\nu} - R_{\mu\nu})A^\nu = m^2 A_\mu , \tag{2.4'}$$

which differ from (2.1) (apart from the replacement of partial derivatives by covariant derivatives) in the explicit appearance of the Ricci tensor.[5] Alternately, the field equations (2.4) can be put in a more concise form by introducing the antisymmetric tensor

$$F_{\mu\nu} = D_\nu A_\mu - D_\mu A_\nu . \tag{2.6}$$

Raising and lowering appropriate indices then yields the familiar form

$$D_\nu F^{\mu\nu} = m^2 A^\mu . \tag{2.7}$$

These equations are seen to differ in form from the Maxwell equations (for empty space) only by the appearance of the vector potential A^μ on the right-hand side, although, of course, the meaning of the various terms is now different.

For a consistent description of the vector meson in a gravitational field we also have to consider Einstein's field equations

$$G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = \kappa T_{\mu\nu}, \tag{2.8}$$

where $T_{\mu\nu}$ is the energy–momentum tensor of the meson field. The form of this tensor can be obtained by variation of the Lagrangian density $\mathfrak{L}$ with respect of $g^{\alpha\beta}$. Since variation of $\mathfrak{L}$ with respect to A_α must also give the field equations (2.7), it follows that (apart from a divergence) the Lagrangian density is

$$\mathfrak{L} = \sqrt{-g}[\tfrac{1}{4} F_{\alpha\beta} F^{\alpha\beta} + \tfrac{1}{2} m^2 A_\alpha A^\alpha]. \tag{2.9}$$

Carrying out the variation with respect to $g^{\mu\nu}$ then yields the required form of the energy–momentum tensor:

$$T_{\mu\nu} = m[F_{\mu\alpha} F_{\nu\beta} g^{\alpha\beta} - \tfrac{1}{4} g_{\mu\nu} F_{\alpha\beta} F^{\alpha\beta}] + m^2[A_\mu A_\nu - \tfrac{1}{2} g_{\mu\nu} A_\alpha A^\alpha] \tag{2.10}$$

(which, again, differs in form, from the Maxwell energy–momentum tensor in the appearance of terms containing A explicitly). Substituting (2.10) into the Einstein equations (2.8), together with the field equations (2.7), then gives a codetermined description of the vector meson in a gravitational field.

3. THE ENERGY-MOMENTUM TENSOR

For a perfect fluid, the energy–momentum tensor can be written in terms of two scalars, the pressure p and density ρ, and the velocity field $u^\mu = dx^\mu/ds$ as

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - g_{\mu\nu} p. \tag{3.1}$$

The field equations of Einstein are not sufficient to determine these quantities uniquely and it is customary to impose a relation, an "equation of state," between the pressure p and the density ρ. In the presence of an external field, the particular form of the energy–momentum tensor and the equations satisfied by that field provide the additional information needed to solve the field equations (at least, in principle). In our case, too, it would be interesting to know whether (2.10) can be cast into the form (3.1) and, in particular, what, if any, limitations are imposed on ρ and p.

Instead of (3.1) we start with a more general form of the energy–momentum tensor allowing for possible viscous terms. Thus, we consider

$$T_{\mu\nu} = \sigma u_\mu u_\nu + S_{\mu\nu} \tag{3.2}$$

with

$$S_{\mu\nu} u^\nu = 0, \quad u^\nu u_\nu = 1. \tag{3.2'}$$

[Clearly, (3.1) is a special case of (3.2) with $\sigma = \rho$, $S_{\mu\nu} = (u_\mu u_\nu - g_{\mu\nu})p$.]

We have already remarked that part of the energy–momentum tensor (2.10) is of the form found in electromagnetic theory, i.e.,

$$E_{\mu\nu} = F_{\mu\alpha} F_{\nu\beta} g^{\alpha\beta} - \tfrac{1}{4} g_{\mu\nu} F_{\alpha\beta} F^{\alpha\beta},$$

[4] Note the order of the covariant derivatives. They can be put in the usual order by using the commutation relation (2.3).

[5] Of course, if we assume that the gravitational field is known and Einstein's equtions $R_{\mu\nu} = 0$ for empty space hold, then there is no change.

and is known to satisfy the identity[6]

$$E^{\alpha}_{\mu}E^{\nu}_{\alpha} = \tfrac{1}{4}g^{\nu}_{\mu}E_{\alpha\beta}E^{\alpha\beta}. \tag{3.3}$$

We substitute for $E_{\mu\nu}$, where

$$E_{\mu\nu} = T_{\mu\nu} - V_{\mu\nu}, \tag{3.3'}$$

where $T_{\mu\nu}$ is given by (3.2), and $V_{\mu\nu}$ is the part of the energy–momentum tensor containing A_{μ} explicitly and is defined by

$$V_{\mu\nu} = B_{\mu}B_{\nu} - \tfrac{1}{2}g_{\mu\nu}B^2, \quad B_{\mu} = mA_{\mu}, \quad B^2 = B_{\mu}B^{\mu}. \tag{3.3''}$$

We find that

$$u_{\mu}u^{\nu}(\sigma^2 + \sigma B^2) + S^{\alpha}_{\mu}S^{\nu}_{\alpha} + B^2S^{\nu}_{\mu} - \tfrac{1}{2}\sigma UU^{\nu}_{\mu} - X^{\nu}_{\mu}$$
$$= \tfrac{1}{4}g^{\nu}_{\mu}[\sigma^2 + S^2 + B^2(\sigma + S) - (\tfrac{1}{2}\sigma U^2 + X)]. \tag{3.4}$$

Here, X_{μ} is the vector

$$X_{\mu} = S^{\nu}_{\mu}B_{\nu}, \quad X_{\mu}u^{\mu} = 0,$$

while $U_{\mu\nu}$ and $X_{\mu\nu}$ are the tensors

$$U_{\mu\nu} = u_{\mu}B_{\nu} + u_{\nu}B_{\mu},$$
$$X_{\mu\nu} = B_{\mu}X_{\nu} + B_{\nu}X_{\mu}. \tag{3.5}$$

U, X, and S are the respective traces, and $S^2 = S_{\alpha\beta}S^{\alpha\beta}$.

These equations are clearly symmetric upon interchange of the indices, and the trace of each side being equal to each other. A further simplification can be achieved by multiplying and contracting in turn with u^{μ} and A^{μ}:

$$u_{\mu}b - \tfrac{1}{2}\sigma UB_{\mu} - \tfrac{1}{2}UX_{\mu} = 0 \tag{3.6a}$$

and

$$\tfrac{1}{2}U\sigma^2u_{\mu} + dB_{\mu} + Y_{\mu} = 0, \tag{3.6b}$$

where, in addition to the quantities (3.5), we have defined also the vector

$$Y_{\mu} = S^{\nu}_{\mu}X_{\nu} = S^{\nu}_{\mu}S^{\alpha}_{\nu}B_{\alpha}, \quad Y_{\mu}u^{\mu} = 0$$

and introduced the abbreviations

$$b = \tfrac{3}{4}(\sigma^2 + \sigma B^2) - \tfrac{1}{8}\sigma U^2 - \tfrac{1}{4}(B^2S + S^2) + \tfrac{1}{4}X,$$
$$d = b - (\sigma^2 + \sigma B^2) - \tfrac{1}{2}X. \tag{3.7}$$

We can generate additional relations between the various vector fields by operating on (3.6) with S^{μ}_{ν}. For example, from (3.6a) and (3.6b) we obtain

$$\sigma X_{\nu} + Y_{\nu} = 0 \tag{3.8a}$$

and

$$dX_{\nu} + Z_{\nu} = 0, \tag{3.8b}$$

where we have used (3.2'), and defined

$$Z_{\mu} = S^{\nu}_{\mu}Y_{\nu} = S^{\nu}_{\mu}S^{\alpha}_{\nu}X_{\alpha} = S^{\nu}_{\mu}S^{\alpha}_{\nu}S^{\beta}_{\alpha}B_{\beta},$$
$$Z_{\mu}u^{\mu} = 0.$$

Further multiplication of (3.8a) with S^{ν}_{λ} yields

$$\sigma Y_{\lambda} + Z_{\lambda} = 0. \tag{3.8c}$$

(In principle, we could continue thus to generate new equations, but, as we shall see later, the system is closed and no new information can be derived from further operation.) It follows from the above results that both vector fields Y_{μ} and Z_{μ} can be expressed in terms of X_{μ},

$$Z_{\mu} = -\sigma Y_{\mu} = \sigma^2X_{\mu}, \tag{3.9}$$

and that the system of equations is consistent, provided that

$$\sigma^2 + d = 0. \tag{3.10}$$

Contraction of (3.6a) with u^{μ} and A^{μ}, respectively, provides two additional limitations, i.e.,

$$b = \tfrac{1}{4}\sigma U^2 \tag{3.10'}$$

and

$$X = b - \sigma B^2 = \sigma(\tfrac{1}{4}U^2 - B^2), \tag{3.10''}$$

provided that $U \neq 0$.[7]

It can readily be seen that these are the only independent relations between the various quantities and that further contraction gives nothing new.

Making use of these results and (3.7), it follows that the Eqs. (3.6) reduce to one,

$$\tfrac{1}{2}\sigma Uu_{\mu} - \sigma B_{\mu} - X_{\mu} = 0, \tag{3.11}$$

which expresses X_{μ} (and therefore also Y_{μ} and Z_{μ}) in terms of the *two* vector fields u_{μ} and B_{μ}. Furthermore, from (3.7) and (3.10) we find

$$S^2 = S_{\alpha\beta}S^{\alpha\beta} = 2\sigma^2 + (\sigma + B)^2 - \sigma U^2, \tag{3.12a}$$

while the trace S is given by

$$S = -(\sigma + B^2), \tag{3.12b}$$

which follows from (3.2) and the fact that E^{ν}_{μ} is a traceless tensor.

We still have to satisfy (3.4) and the question arises, whether there exists an *algebraic* form of S which satisfied that condition. Since we have two vector fields u^{μ} and B_{μ} in addition to the fundamental tensor $g_{\mu\nu}$, it is suggestive to assume for S^{ν}_{μ} the form

$$S^{\nu}_{\mu} = \alpha g^{\nu}_{\mu} + \beta U^{\nu}_{\mu} + \gamma B_{\mu}B^{\nu} + \delta u_{\mu}u^{\nu}, \tag{3.13}$$

where the functions α, β, γ, and δ are yet to be determined. From the properties of S^{ν}_{μ} it follows that we must impose the following conditions:

(i) That derived from Eq. (3.2'):

$$S^{\nu}_{\mu}u_{\nu} = 0;$$

(ii) That derived from Eqs. (3.5) and (3.11):

$$S^{\nu}_{\mu}B_{\nu} = X_{\mu} + \sigma(\tfrac{1}{2}Uu_{\mu} - B_{\mu});$$

[6] See, e.g., C. Misner and J. A. Wheeler, Ann. Phys. (N.Y.) 2, 525 (1957).

[7] The case $U = 0$ is an important subcase which will be considered later (Section 4).

(iii) Eq. (3.12b):

$$S = -(\sigma + B^2).$$

These result in the following system of equations:

$$\begin{aligned} \alpha + \tfrac{1}{2}\beta U + \delta &= 0, \\ \beta + \tfrac{1}{2}\gamma U &= 0, \\ \beta B^2 + \tfrac{1}{2}\delta U &= \tfrac{1}{2}\sigma U, \\ \alpha + \tfrac{1}{2}\beta U + \gamma B^2 &= -\sigma, \\ 4\alpha + \beta U + \gamma B^2 + \delta &= -\sigma - B^2. \end{aligned} \tag{3.14}$$

Although these are five equations for four functions, they are found to be consistent and have as their solution

$$\begin{aligned} \alpha &= -\tfrac{1}{2}B^2, \\ \beta &= -\tfrac{1}{2}\gamma U = -\tfrac{1}{2}\Delta U(\sigma + \alpha), \\ \gamma &= \Delta(\sigma + \alpha), \\ \delta &= \Delta(\tfrac{1}{4}\sigma U^2 + \alpha B^2), \end{aligned} \tag{3.14'}$$

with

$$\Delta^{-1} = \tfrac{1}{4}(U^2 - 4B^2),$$

where the three functions B^2, U, and σ are still arbitrary. It can be shown by a straightforward, although tedious, calculation that there are no further conditions imposed on S_μ^ν by its properties. For example, the conditions

$$\begin{aligned} S_\mu^\nu S_\nu^\alpha u_\alpha &= 0, \\ S_\mu^\nu S_\nu^\alpha B_\alpha &= -\sigma^2(\tfrac{1}{2}u_\mu U - B_\mu) \end{aligned}$$

are identically satisfied by (3.13) and (3.14).

Finally, inserting (3.13) into (3.14) gives one additional relation between the three functions

$$(\sigma + \tfrac{1}{2}B^2)^2 - \tfrac{1}{2}\sigma U^2 = 0, \tag{3.15}$$

from which one can be eliminated. If we eliminate U and use α instead of B^2, it can be shown that (3.13) now takes the form

$$\begin{aligned} S_\mu^\nu = (\sigma + \alpha)^{-1}[&\alpha(\sigma + \alpha)g_\mu^\nu + \sigma(\sigma - 3\alpha)u_\mu u^\nu \\ &+ 2\sigma B_\mu B^\nu - (2\sigma)^{\frac{1}{2}}(\sigma - \alpha)U_\mu^\nu] \end{aligned} \tag{3.16}$$

involving two vector fields u^μ and B^μ as well as two scalars σ and α. This is the most general algebraic form of S_μ^ν and we now consider two special cases.

4. PARALLEL AND ORTHOGONAL FIELDS

If the two vectors u^μ and B^μ are not independent, they may be either parallel or perpendicular to each other. In the first case we can set

$$B^\mu = \tfrac{1}{2}Uu^\mu, \quad B^2 = \tfrac{1}{4}U^2 \tag{4.1}$$

and, consequently,

$$\begin{aligned} X_\mu &= S_\mu^\nu A_\nu = 0, \\ U_\mu^\nu &= Uu_\mu u^\nu, \quad X_\mu^\nu = 0. \end{aligned} \tag{4.2}$$

With this assumption, (3.4) becomes

$$\begin{aligned} u_\mu u^\nu(\sigma^2 - \tfrac{1}{4}\sigma U^2) + S_\mu^\alpha S_\alpha^\nu + \tfrac{1}{4}U^2 S_\mu^\nu \\ = \tfrac{1}{4}g_\mu^\nu[(\sigma^2 + S^2) + \tfrac{1}{4}U^2(S - \sigma)]. \end{aligned} \tag{4.3}$$

Contracting this equation with u^μ (or B^μ) results in

$$3(\sigma^2 - \tfrac{1}{4}\sigma U^2) = S^2 + \tfrac{1}{4}SU^2. \tag{4.4}$$

Since we now only have one arbitrary vector field u^μ, Eq. (4.13) takes the simple form

$$S_\mu^\nu = \tilde{\alpha}g_\mu^\nu + \tilde{\delta}u_\mu u^\nu, \tag{4.5}$$

where the two functions $\tilde{\alpha}$ and $\tilde{\delta}$ are to be determined from the conditions

$$S_\mu^\nu u_\nu = (\tilde{\alpha} + \tilde{\delta})u_\mu = 0 \tag{4.6a}$$

and

$$S = 4\tilde{\alpha} + \tilde{\delta} = -(\sigma + \tfrac{1}{4}U^2). \tag{4.6b}$$

This, together with (4.4), enables us to express all functions in terms of σ:

$$\begin{aligned} \tilde{\alpha} &= -\tilde{\delta} = -\sigma, \\ U^2 &= 8\sigma. \end{aligned} \tag{4.7}$$

A simple calculation then shows that (4.3) is identically satisfied and no further conditions are imposed on S_μ^ν. Therefore, the energy–momentum tensor $T^{\mu\nu}$ (3.2) is now given by

$$T^{\mu\nu} - \sigma u^\mu u^\nu + S^{\mu\nu} = 2\sigma u^\mu u^\nu - \sigma g^{\mu\nu} \tag{4.8}$$

which corresponds to a *perfect* fluid (3.1) with

$$\rho = p = \sigma. \tag{4.8'}$$

However, if we introduce (4.1) also into (3.3″), we find

$$V_{\mu\nu} = \tfrac{1}{4}U^2 u_\mu u_\nu - \tfrac{1}{8}g_{\mu\nu}U^2 = 2\sigma u_\mu u_\nu - \sigma g_{\mu\nu}, \tag{4.9}$$

which is exactly equal to (4.8). Thus, we conclude that the tensor $E_{\mu\nu}$ [Eq. (3.3′)] vanishes. This implies that either $m = 0$ or

$$F_{\mu\nu} = A_{\mu,\nu} - A_{\nu,\mu} = 0$$

and, hence, A_μ is the gradient of a scalar

$$A_\mu = \frac{\partial \phi}{\partial x^\mu}.$$

From this it appears that a *massive vector meson* cannot be described by a perfect fluid in terms of a pressure and density only.

Now let us turn to the second possibility, viz., that the two vector fields u^μ and B^μ are orthogonal to each other, i.e.,

$$\tfrac{1}{2}U = u^\mu B_\mu = 0. \tag{4.10}$$

In this case, (3.4) reduces to

$$u_\mu u^\nu(\sigma^2 + \sigma B^2) + S^\alpha_\mu S^\nu_\alpha + B^2 S^\nu_\mu - X^\nu_\mu = \tfrac{1}{4} g^\nu_\mu[\sigma^2 + S^2 - B^2(\sigma + S) - X]. \quad (4.11)$$

Contracting this equation in turn with u^μ and B^μ leads to [cf. Eq. (3.6)]:

$$u_\mu \bar{b} = 0 \quad \text{or} \quad \bar{b} = 0, \quad (4.12a)$$

$$Y_\mu + \bar{d} B_\mu = 0, \quad (4.12b)$$

where $\bar{b}$ and $\bar{d}$ differ from their unbarred counterparts (3.7) by the fact that now $U = 0$. As before, it is again possible to generate additional equations by operation with S. For example, from (4.12b) we obtain

$$Z_\mu + \bar{d} X_\mu = 0. \quad (4.12c)$$

Since Y_μ is determined by B_μ and Z_μ by X_μ, it appears that we have again three arbitrary vector fields determining the structure of S^ν_μ. Nevertheless, we see that these are not independent. If we again assume for S^ν_μ the form (3.13),

$$S^\nu_\mu = \bar{\alpha} g^\nu_\mu + \bar{\beta} U^\nu_\mu + \bar{\gamma} \beta_\mu B^\nu + \bar{\delta} u_\mu u^\nu, \quad (4.13)$$

it follows from

$$S^\nu_\mu u_\nu = (\bar{\alpha} + \bar{\delta}) u_\mu + \bar{\beta} A_\mu = 0$$

that

$$\bar{\alpha} + \bar{\delta} = 0, \qquad \bar{\beta} = 0. \quad (4.14)$$

Also,

$$X_\mu = S^\nu_\mu B_\nu = (\bar{\alpha} + \bar{\gamma} B^2) B_\mu = \lambda B_\mu,$$

indicating that X_μ and B_μ are parallel.

Furthermore, from

$$Y_\mu = S^\nu_\mu X_\nu = \lambda^2 A_\mu$$

and (4.12b), we have the relation

$$\lambda^2 + \bar{d} = 0. \quad (4.15)$$

Additional conditions on the different functions are imposed by evaluating S:

$$S = 4\bar{\alpha} + \bar{\gamma} B^2 + \bar{\delta} = -(\sigma + B^2) \quad (4.15')$$

and

$$S^2 = (\sigma + B^2)^2 + 2\lambda^2 = 2\bar{\alpha}^2 + \lambda^2. \quad (4.15'')$$

It can be shown that there are no further independent conditions and that S^ν_μ [Eq. (4.13)] is completely determined by the two vectors u^μ and B^μ and two functions σ and B^2:

$$\bar{\alpha} = -(\sigma + B^2) = -P, \quad \bar{\beta} = 0, \quad \bar{\gamma} = 2P/B^2, \quad \bar{\delta} = -\bar{\alpha} = P, \quad (4.16)$$

which gives, for the energy-momentum tensor (3.2),

$$T^{\mu\nu} = \sigma u^\mu u^\nu + S^{\mu\nu} = (\sigma + P) u^\mu u^\nu - P g^{\mu\nu} + (2P/B^2) B^\mu B^\nu \quad (4.17)$$

with

$$P = \sigma + B^2. \quad (4.17')$$

This expression differs from that for a perfect fluid (3.1) in two important aspects. In the first place, in addition to the usual terms a third term appears which may be taken to describe the stresses, since even for a comoving "fluid" it does not vanish. Secondly, the vector potential B^μ (or A^μ) contributes to the effective pressure through B^2, although, of course, P [Eq. (4.17')] is not the actual pressure of the fluid.

5. CONSERVATION LAWS AND FIELD EQUATIONS

The vector A^μ is not arbitrary, but must be a solution of the field equations (2.7), while the energy-momentum tensor (4.17) has to satisfy the conservation laws[8]

$$D_\mu T^{\mu\nu} = 0. \quad (5.1)$$

(Of course, these are a consequence of Bianchi's identities and are identically satisfied once Einstein's equations are solved. Nevertheless, they are of considerable interest and determine the behavior of the velocity field u^μ.)

We find it convenient to introduce the unit vector

$$a^\mu = A^\mu/A = B^\mu/B, \quad (5.2)$$

so that

$$a^\mu a_\mu = 1, \quad (5.2')$$

and the scalar

$$C = \ln B. \quad (5.3)$$

In terms of this vector, (4.17) can now be written as

$$T^{\mu\nu} = u^\mu u^\nu(\sigma + P) - P g^{\mu\nu} + 2P a^\mu a^\nu. \quad (5.4)$$

From the conservation laws (5.1) we find

$$u_\mu u^\nu_{;\nu}(\sigma + P) + u_{\mu;\nu} u^\nu(\sigma + P) + u_\mu u^\nu(\sigma + P)_{,\nu} = P_{,\mu} - 2P_{,\nu} a^\nu a_\mu - 2P(a_{\mu;\nu} a^\nu + a_\mu a^\nu_{;\nu}).$$

Contracting this equation with u^μ and a^μ, respectively, on account of the orthogonality of these vectors we obtain

$$u^\nu_{;\nu}(\sigma + P) + u^\nu \sigma_{,\nu} = -2P\Phi \quad (5.5a)$$

and

$$2P a^\nu_{;\nu} + a^\nu P_{,\nu} = -(\sigma + P)\Psi, \quad (5.5b)$$

where we have introduced the two quantities

$$\Phi = a_{\mu;\nu} a^\nu u^\mu = -u_{\mu;\nu} a^\nu a^\mu,$$
$$\Psi = u_{\mu;\nu} u^\nu a^\mu = -a_{\mu;\nu} u^\mu u^\nu. \quad (5.5c)$$

[8] We limit our discussion to the energy-momentum tensor given by (4.17), although it is not difficult to treat the general case also.

Furthermore, from the subsidiary condition (2.5) we also find

$$a^{\mu}_{;\mu} + C_{,\mu}a^{\mu} = 0, \tag{5.6}$$

so that (5.5b) can be written as

$$2PC_{,\mu}a^{\mu} - P_{,\mu}a^{\mu} = (\sigma + P)u_{\mu;\nu}u^{\nu}a^{\mu}. \tag{5.5b'}$$

Since this equation is of the form $f_{\mu}a^{\mu} = 0$, it is suggestive to take as its solution $f_{\mu} = 0$. Substituting this result into (5.4) and contracting with u^{μ}, however, leads to $P = B^2$, which is only consistent with (4.17) if $\sigma = 0$. A more satisfactory solution is of the form

$$u_{\mu;\nu}u^{\nu}(\sigma + P) + P_{,\mu} - 2PC_{,\mu} = u_{\mu}f, \tag{5.7}$$

where f is an arbitrary function.

Substituting (5.7) into (5.4) and contracting with u^{μ} yields for f,

$$f = P_{,\mu}u^{\mu} - 2Pu^{\mu}C_{,\mu} = \sigma_{,\mu}u^{\mu} - 2\sigma u^{\mu}C_{,\mu}, \tag{5.7'}$$

and, after some simplification, for (5.7) yields

$$u_{\mu;\nu}u^{\nu}(\sigma + P) = (u_{\mu}u^{\alpha} - g^{\alpha}_{\mu})(P_{,\alpha} - 2PC_{,\alpha}). \tag{5.7''}$$

This expression is seen to be similar to the usual results of hydrodynamics[9] and differs from it mainly through the appearance of the extra term involving the electromagnetic potential. If we now also insert (5.7'') into (5.4), we obtain

$$Pa_{\mu;\nu}a^{\nu} = P\Phi u_{\mu} + (g^{\alpha}_{\mu} - a_{\mu}a^{\alpha} - u_{\mu}u^{\alpha})(P_{,\alpha} - PC_{,\alpha}), \tag{5.8}$$

which determines the streaming of the vector field a^{μ}. If a is a spacelike vector, so that $a^{\mu}a_{\mu} = -1$ (and hence B^2 negative), a similar calculation gives slightly modified results. Instead of (5.3), we find

$$T^{\mu\nu} = (\sigma + P)u^{\mu}u^{\nu} - Pg^{\mu\nu} - 2Pa^{\mu}a^{\nu} \tag{5.3'}$$

with $P = \sigma - |B|^2$.

The sign of the right-hand side of (5.5a) is now positive, but (5.5b) remains unchanged. Furthermore, (5.7'') remains the same, but (5.8) has to be replaced by

$$Pa_{\mu;\nu}a^{\nu} = P\Phi u_{\mu} + (u_{\mu}u^{\alpha} - a_{\mu}a^{\alpha} - g^{\alpha}_{\mu})(P_{,\alpha} - PC_{,\alpha}). \tag{5.8'}$$

It is easy to see that contracting this equation with u^{μ} and a^{μ} yields nothing new.

The behavior of the two vector fields is then determined by (5.7'') and (5.8) together with the subsidiary conditions (5.5a) and (5.6).

Furthermore, from the field equations (2.4),

$$D_{\alpha}D^{\alpha}B_{\mu} - R_{\mu\nu}B^{\nu} = m^2B_{\mu\nu},$$

upon introducing the vector a^{μ} and scalar C, we also obtain

$$(\Box C + C_{\alpha}C^{\alpha})a_{\mu} + 2C^{\beta}a_{\mu;\beta} + D_{\alpha}D^{\alpha}a_{\mu\nu} - R_{\mu\lambda}a^{\lambda} = m^2a_{\mu}, \tag{5.9}$$

where

$$\Box C + C_{,\alpha}C'^{\alpha} = \frac{\Box(P - \sigma)}{2(P - \sigma)} - \frac{1}{4}\frac{(P - \sigma)_{,\alpha}(P - \sigma)_{,\beta}g^{\alpha\beta}}{(P - \sigma)^2},$$

and, upon contracting,

$$(\Box C + C_{\alpha}C^{\alpha}) - a^{\mu}_{;\alpha}a^{\alpha}_{\mu;} - R_{\mu\lambda}a^{\lambda}a^{\mu} = m^2. \tag{5.9'}$$

These equations together with (5.7) completely determine a^{μ} and C.

This completes, then, the general analysis of the vector meson field including the subsidiary conditions. It is still necessary to solve Einstein's field equations (2.8). Since the general case is rather involved, we limit ourselves to special examples involving spherical symmetry and the expanding universe. These are given in a subsequent work.

[9] See, e.g., A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnetisme* (Mason et Cie., Paris, 1955).

Asymptotic Fields in Some Models of Quantum Field Theory. II

RAPHAEL HØEGH-KROHN
Institute of Mathematics, University of Oslo

(Received 19 September 1967)

A quantum field with nonlocal but translation-invariant interaction is considered. We prove that, with a proper smoothness condition on the interaction, the asymptotic limits of the annihilation–creation operators exists. The asymptotic limits are then used to prove that the state space decomposes as a tensor product of an incoming (outgoing) Fock space and a zero-particle space.

1. INTRODUCTION

In an earlier paper[1] we considered a fermion field with self-interaction and proved that, if the kernels describing the interaction were sufficiently smooth, then the asymptotic limits of the annihilation–creation operators did exist. It was also proved that they were free fermion fields with the same mass as the original free fields and that the total Hilbert space could be written in terms of these asymptotic fields as a tensor product of an asymptotic Fock space and a zero-particle space.

The object of this paper is to prove a similar result for the case of translation-invariant or momentum-conserving interaction, if the interaction contains no zero- or one-particle interaction. If, on the other hand, there is a zero- or one-particle interaction, then the asymptotic annihilation–creation operators can not be expected to exist, as we know from perturbation consideration (see, for instance, Refs. 2–4).

2. FERMION FIELD WITH MOMENTUM-CONSERVING INTERACTION

As in Paper I, here we consider a free fermion field with mass m in the Fock representation. This is a Hilbert space $\mathcal{H}$, where the elements in $\mathcal{H}$ are represented by sequences $f = \{f_n\}$ of antisymmetric functions $f_n = f_n(x_1, \cdots, x_n)$ in n variables $x_1, \cdots, x_n$, $x_i \in E_3$. The inner product in $\mathcal{H}$ is given by

$$(f, g) = \sum_{n=1}^{\infty} n! \int \cdots \int \bar{f}_n(x_1, \cdots, x_n) \times g_n(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$

and the annihilation operators $a(x)$ by

$$(a(x)f)_n(x_1, \cdots, x_n) = (n+1) f_n(x, x_1, \cdots, x_n).$$

The creation operator $a^*(x)$ is the adjoint of $a(x)$. The free-energy operator H_0 is given by

$$(H_0 f)_n = \sum_{i=1}^{n} \Omega_i f_n(x_1, \cdots, x_n),$$

and $(H_0 f)_0 = 0$, where Ω_i is the operator $\Omega = (-\Delta + m^2)^{\frac{1}{2}}$ applied to x_i. The mass m is strictly positive. The total energy operator is $H = H_0 + V$, where the interaction V is momentum conserving and contains no vacuum or one-particle interaction. That is,

$$V = \sum_{k,l=2} V_{k,l},$$

where V_{kl} is given in terms of the kernel

$$V_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l),$$

which is translation-invariant, i.e.,

$$V_{kl}(x_1 + z, \cdots, x_k + z \mid y_1 + z, \cdots, y_l + z) = V_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l)$$

and

$$V_{kl} = \int \cdots \int dx_1 \cdots dx_k\, dy_1 \cdots dy_l \times V_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l) \times a^*(x_1) \cdots a^*(x_k) a(y_1) \cdots a(y_l).$$

We recall that f_n is antisymmetric if $f_n = \operatorname{asym} f_n$, where

$$\operatorname{asym} f_n(x_1, \cdots, x_n) = \frac{1}{n!} \sum_{\sigma} (-1)^{\sigma} f_n(x_{\sigma(1)}, \cdots, x_{\sigma(n)}).$$

The reason for excluding terms of the form V_{k0}, V_{0l}, V_{k1}, and V_{1l}—i.e., vacuum and one-particle interaction—is that such terms leads to infinite vacuum renormalization and finite-mass renormalization respectively, and it is beyond the scope of the technique presented in this paper to deal with such problems.

It is obvious that H_0 is self-adjoint on its natural domain of definition D_0, and we assume V to be symmetric, i.e.,

$$V_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l) = \bar{V}_{lk}(y_1, \cdots, y_l \mid x_1, \cdots, x_k).$$

Let $a(h) = \int a(x)h(x)\, dx$ and

$$a^*(h) = \int a^*(x)h(x)\, dx.$$

[1] R. Høegh-Krohn, J. Math. Phys. **9**, 2075 (1968), referred to as Paper I.
[2] L. D. Faddeev, Dokl. Akad. Nauk SSSR **152**, 573 (1963) [Sov. Phys.—Dokl. **8**, 881 (1964)].
[3] K. O. Friedrichs, *Perturbation of Spectra in Hilbert Space* (American Mathematical Society, Providence, R.I., 1965).
[4] L. van Hove, Physica **21**, 901 (1955); **22**, 257 (1956).

Due to the anticommutation relations of the fermion field we have

$$a(h)a^*(\bar{h}) + a^*(\bar{h})a(h) = \int |h(x)|^2\, dx.$$

Observing that $a^*(\bar{h})$ is the adjoint of $a(h)$, we now obtain

$$\|a(h)\| = \|a^*(h)\| \leq \|h\|_2.$$

On the other hand, let ϕ_0 be the vacuum state $\phi_0 = \{1, 0, \cdots\}$. Then

$$\|a^*(h)\phi_0\| = \|h\|_2,$$

and this proves the following lemma.

Lemma 1: $\|a(h)\| = \|a^*(h)\| = \|h\|_2$.

For the subsequent development we also need the following lemma.

Lemma 2: Let $f \in D_0$ (domain of H_0); then

$$n\, \|f_n\|_2 \leq m^{-1}\, \|(H_0 f)_n\|_2,$$

where m is the mass of the free fermion field.

Proof:

$$(H_0 f)_n = \sum_{i=1}^{n} \Omega_i f_n$$

and, since $\Omega = (-\Delta + m^2)^{\frac{1}{2}}$, we have that $(h, \Omega h) \geq m(h, h)$, which proves the lemma.

Lemma 3: Let $\|u\|_0 = \max\{\|u\|_1, \|u\|_2\}$ and let A_{kl}, for $k, l \geq 1$, be the annihilation–creation operator with the kernel

$$A_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l) = \operatorname*{asym}_x \operatorname*{asym}_y u_1(x_1 - y_l) \cdots u_k(x_k - y_l) \times v_1(y_1 - y_l) \cdots v_{l-1}(y_{l-1} - y_l).$$

Then, for all $f \in D_0$, we have

$$\|A_{kl} f\| \leq C_{kl}\, \|u_1\|_0 \cdots \|u_k\|_0\, \|v_1\|_0 \cdots \|v_{l-1}\|_0 \cdot \|H_0 f\|,$$

where $C_{kl} = m^{-1}(|k - l| + 1)^{\frac{1}{2}}$.

Proof: Let $p = k - l + n$; then

$$\|A_{kl} f_n\|^2 = p! \int \cdots \int dx_1 \cdots dx_p \times \left| \binom{n}{l} l! \operatorname*{asym}_{x_1, \cdots, x_p} \int \cdots \int dy_1 \cdots dy_l \times u_1(x_1 - y_l) \cdots u_k(x_k - y_l) \times v_1(y_1 - y_l) \cdots v_{l-1}(y_{l-1} - y_l) \times f_n(y_1, \cdots, y_l, k_{k+1}, \cdots, x_p) \right|^2.$$

By the fact that asym is an orthogonal projection in L_2, we get

$$\|A_{kl} f_n\|^2 \leq p! \int \cdots \int dx_1 \cdots dx_p \times \left[\binom{n}{l} l! \int dy_l\, |u_1(x_1 - y_l)| \times \left| \operatorname*{asym}_{x_2, \cdots, x_p} \int \cdots \int dy_1 \cdots dy_{l-1} \times u_2(x_2 - y_l) \cdots u_k(x_k - y_l) \times v_1(y_1 - y_l) \cdots v_{l-1}(y_{l-1} - y_l) \times f_n(y_1, \cdots, y_l, x_{k+1}, \cdots, x_p) \right| \right]^2.$$

By the Schwarz inequality,

$$\|A_{kl} f_n\|^2 \leq p! \int \cdots \int dx_1 \cdots dx_p \int d\tilde{y}_l\, |u_1(x_1 - \tilde{y}_l)| \times \int dy_l\, |u_1(x_1 - y_l)| \times \left| \binom{n}{l} l! \operatorname*{asym}_{x_2, \cdots, x_p} \int \cdots \int dy_1 \cdots dy_{l-1} \times u_2(x_2 - y_l) \cdots u_k(x_k - y_l) \times v_1(y_1 - y_l) \cdots v_{l-1}(y_{l-1} - y_l) \times f_n(y_1, \cdots, y_l, x_{k+1}, \cdots, x_p) \right|^2$$

$$= p!\, \|u_1\|_1^2 \int \cdots \int dy_l\, dx_2 \cdots dx_p \times \left| \binom{n}{l} l! \operatorname*{asym}_{x_2, \cdots, x_p} \int \cdots \int dy_1 \cdots dy_{l-1} \times u_2(x_2 - y_l) \cdots u_k(x_k - y_l) \times v_1(y_1 - y_l) \cdots v_{l-1}(y_{l-1} - y_l) \times f_n(y_1, \cdots, y_l, x_{k+1}, \cdots, x_p) \right|^2.$$

By Lemma 1, this is bounded by

$$p!\, \|u_1\|_1^2 \frac{(n-1)!}{(p-1)!} \left[\binom{n}{l} l! \right]^2 \cdot \left[\binom{n-1}{l-1} (l-1)! \right]^{-2} \times \int dy_l\, \|u_2\|_2^2 \cdots \|u_k\|_2^2\, \|v_1\|_2^2 \cdots \|v_{l-1}\|_2^2 \times \int \cdots \int dy_1 \cdots dy_{l-1}\, dx_{k+1} \cdots dx_p \times |f_n(y_1, \cdots, y_l, x_{k+1}, \cdots, x_p)|^2$$

$$= p \cdot n \cdot n!\, \|u_1\|_1^2\, \|u_2\|_2^2 \cdots \|u_k\|_2^2 \times \|v_1\|_2^2 \cdots \|v_{l-1}\|_2^2\, \|f_n\|_2^2.$$

Recalling that $p = k - l + n$, by Lemma 2 we get

$$\|A_{kl} f_n\| \leq \frac{(k - l + n)n}{n^2} \frac{1}{m^2}\, \|u_1\|_1^2\, \|u_2\|_2^2 \cdots \|u_k\|_2^2 \times \|v_1\|_2^2 \cdots \|v_{l-1}\|_2^2\, \|H_0 f_n\|^2,$$

and this proves the lemma.

Lemma 4: Let $a(g) = \int a(x)g(x)\,dx$ and $a^*(g) = \int a^*(x)g(x)\,dx$ and $f \in D_0$ if $k, l \geq 2$; then

$$\|[A_{kl}, a(g)]f\| \quad \text{and} \quad \|[A_{kl}, a^*(g)]f\| \leq kC_{kl}\|g\|_\infty \|u_1\|_o \cdots \|u_k\|_o \|v_1\|_o \cdots \|v_{l-1}\|_o \|H_0 f\|.$$

Proof: Let $p = k - l + n - 1$; then

$$\begin{aligned}\|[A_{kl}, a(g)]f_n\|^2 &= p!\int\cdots\int dx_1\cdots dx_p \left| \int dz\, g(z)\binom{n}{l} l! \right. \\ &\quad \times \operatorname*{asym}_{x_1,\cdots,x_p} \int\cdots\int dy_1\cdots dy_l \\ &\quad \times \sum_{i=1}^{k} (-1)^i u_1(x_1 - y_l)\cdots u_i(z - y_l) \\ &\quad \times u_{i+1}(x_i - y_l)\cdots u_k(x_{k-1} - y_l) \\ &\quad \times v_1(y_1 - y_l)\cdots v_{l-1}(y_{l-1} - y_l) \\ &\quad \left. \times f_n(y_1, \cdots, y_l, x_k, \cdots, x_p)\right|^2.\end{aligned}$$

By the Schwarz inequality we get that this is equal to

$$\begin{aligned}&p!\, k\sum_{i=1}^{k}\int\cdots\int dx_1\cdots dx_p \\ &\quad \times \left|\binom{n}{l} l!\int dz\, g(z) \operatorname*{asym}_{x_1,\cdots,x_p} \int\cdots\int dy_1\cdots dy_l \right. \\ &\quad \times u_1(x_1 - y_l)\cdots u_i(z - y_l) \\ &\quad \times u_{i+1}(x_i - y_l)\cdots u_k(x_{k-1} - y_l) \\ &\quad \times v_1(y_1 - y_l)\cdots v_{l-1}(y_{l-1} - y_l) \\ &\quad \left. \times f_n(y_1\cdots y_l, x_k\cdots x_p)\right|^2.\end{aligned}$$

Since asym is a projection in L_2, the first term in the summation is bounded by

$$\begin{aligned}&\int\cdots\int dx_1\cdots dx_p \\ &\quad \times \left|\binom{n}{l} l!\iint dz\, dy_l\, u_1(z - y_l)g(z)u_2(x_1 - y_l) \right. \\ &\quad \times \operatorname*{asym}_{x_2,\cdots,x_p} \int\cdots\int dy_1\cdots dy_{l-1} \\ &\quad \times u_3(x_2 - y_l)\cdots u_k(x_{k-1} - y_l) \\ &\quad \times v_1(y_1 - y_l)\cdots v_{l-1}(y_{l-1} - y_l) \\ &\quad \left. \times f_n(y_1, \cdots, y_l, x_k, \cdots, x_p)\right|^2 \\ &\leq \|g\|_\infty^2 \|u_1\|_1^2 \int\cdots\int dx_1\cdots dx_p \\ &\quad \times \left[\binom{n}{l} l!\int dy_l\, |u_2(x_1 - y_l)| \right. \\ &\quad \times \left| \operatorname*{asym}_{x_2,\cdots,x_p} \int\cdots\int dy_1\cdots dy_{l-1} \right. \\ &\quad \times u_3(x_2 - y_l)\cdots u_k(x_{k-1} - y_l) \\ &\quad \times v_1(y_1 - y_l)\cdots v_{l-1}(y_{l-1} - y_l) \\ &\quad \left. \left. \times f_n(y_1, \cdots, y_l, x_k, \cdots, x_p)\right|\right]^2.\end{aligned}$$

By the Schwarz inequality, this is bounded by

$$\begin{aligned}&\|g\|_\infty^2 \|u_1\|_1^2 \|u_2\|_1 \int\cdots\int dx_1\cdots dx_p \int dy_l\, |u_2(x_1 - y_l)| \\ &\quad \times \left[\binom{n}{l} l! \left| \operatorname*{asym}_{x_2,\cdots,x_p} \int\cdots\int dy_1\cdots dy_{l-1} \right. \right. \\ &\quad \times u_3(x_2 - y_l)\cdots u_k(x_{k-1} - y_l) \\ &\quad \times v_1(y_1 - y_l)\cdots v_{l-1}(y_{l-1} - y_l) \\ &\quad \left. \left. \times f_n(y_1, \cdots, y_l, x_k, \cdots, x_p)\right|\right]^2 \\ &= \|g\|_\infty^2 \|u_1\|_1^2 \|u_2\|_1^2 \int dy_l \int\cdots\int dx_2\cdots dx_p \\ &\quad \times \left|\binom{n}{l} l! \operatorname*{asym}_{x_2,\cdots,x_p} \int\cdots\int dy_1\cdots dy_{l-1} \right. \\ &\quad \times u_3(x_2 - y_l)\cdots u_k(x_{k-1} - y_l) \\ &\quad \times v_1(y_1 - y_l)\cdots v_{l-1}(y_{l-1} - y_l) \\ &\quad \left. \times f_n(y_1, \cdots, y_l, x_k, \cdots, x_p)\right|^2.\end{aligned}$$

By Lemma 1, this is bounded by

$$\begin{aligned}&\|g\|_\infty^2 \|u_1\|_1^2 \|u_2\|_1^2 \frac{(n-1)!}{(p-1)!} \\ &\quad \times \left[\binom{n}{l} l!\right]^2 \left[\binom{n-1}{l-1}(l-1)!\right]^{-2} \\ &\quad \times \int dy_l\, \|u_3\|_2^2 \cdots \|u_k\|_2^2 \|v_1\|_2^2 \cdots \|v_{l-1}\|_2^2 \\ &\quad \times \int\cdots\int dy_1\cdots dy_{l-1}\, dx_k\cdots dx_p \\ &\quad \times |f_n(y_1, \cdots, y_l, x_k, \cdots, x_p)|^2 \\ &= \|g\|_\infty^2 \|u_1\|_1^2 \|u_2\|_1^2\, n \\ &\quad \times \frac{n!}{(p-1)!} \|u_3\|_2^2 \cdots \|u_k\|_2^2 \|v_1\|_2^2 \cdots \|v_{l-1}\|_2^2 \|f_n\|_2^2 \\ &\leq n \cdot \frac{n!}{(p-1)!} \|g\|_\infty^2 \|u_1\|_0^2 \cdots \|u_k\|_0^2 \\ &\quad \times \|v_1\|_0^2 \cdots \|v_{l-1}\|_0^2 \|f_n\|_2^2.\end{aligned}$$

By estimating the remaining terms in the summation in the same way, we get

$$\begin{aligned}\|[A_{kl}, a(g)]f_n\|^2 &\leq n \cdot pk^2 \|g\|_\infty^2 \\ &\quad \times \|u_1\|_0^2 \cdots \|u_k\|_0^2 \|v_1\|_0^2 \cdots \|v_{l-1}\|_0^2 \|f_n\|^2.\end{aligned}$$

Using Lemma 2, we get the first part of Lemma 4. The second part is proved in the same way.

We are now in position to give the smallness and smoothness conditions on the kernels

$$V_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l).$$

Assumption 1: The kernels

$$V_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l)$$

can be represented in the following way:

$$V_{kl}(x_1, \cdots, x_k \mid y_1, \cdots, y_l) = \sum_{n=1}^{\infty} A_{kl}^{(n)}(x_1, \cdots, x_k \mid y_1, \cdots, y_l),$$

such that

$$\sum_{nkl} C_{kl} \|u_1^{nkl}\|_0 \cdots \|u_k^{nkl}\|_0 \|v_1^{nkl}\|_0 \cdots \|v_{l-1}^{nkl}\|_0 < 1,$$

where $A_{kl}^{(n)}(x_1, \cdots, x_k \mid y_1, \cdots, y_l)$ is a kernel of the form given in Lemma 3, and $C_{kl} = m^{-1}(|k-l|+1)^{\frac{1}{2}}$. Moreover,

$$V_{kl} = 0 \text{ for } k = 0 \text{ or } 1, \text{ or } 1 = 0 \text{ or } 1.$$

Assumption 2:

$$\sum_{nkl} kC_{kl} \|u_1^{nkl}\|_0 \cdots \|u_k^{nkl}\|_0 \|v_1^{nkl}\|_0 \cdots \|v_{l-1}^{nkl}\|_0 < \infty.$$

Theorem 1: Under Assumption 1 we have that

$$H = H_0 + V,$$

with $V = \sum_{k,1=2}^{\infty} V_{kl}$, is a self-adjoint operator on the domain of H_0, i.e., on D_0.

Proof: From Assumption 1 and Lemma 3 we get that, for $f \in D_0$,

$$\|Vf\| \leq \lambda \|H_0 f\|$$

with

$$\lambda = \sum_{nkl} C_{kl} \|u_1^{nkl}\|_0 \cdots \|u_k^{nkl}\|_0 \|v_1^{nkl}\|_0 \cdots \|v_{l-1}^{nkl}\|_0 < 1.$$

It is a well-known result that this implies that $H = H_0 + V$ is self-adjoint on the same domain as H_0.

We observe that it follows from Theorem 1 that D_0 is invariant under e^{itH} as well as under e^{itH_0}, and that for $f \in D_0$, both $e^{itH}f$ and $e^{itH_0}f$ are strongly differentiable with respect to t.

Lemma 5: For $h \in D_\Omega \subset L_2(E_3)$ we have that $a(h)$, as well as $a^*(h)$, leaves D_0 invariant, where D_Ω is the domain of Ω in $L_2(E_3)$.

Proof: The proof of this lemma follows immediately from the definitions of $a(h)$ and $a^*(h)$ (see, for instance, Ref. 1).

We define

$$a_t(h) = e^{-itH}e^{itH_0}a(h)e^{-itH_0}e^{itH},$$
$$a_t^*(h) = e^{-itH}e^{itH_0}a^*(h)e^{-itH_0}e^{itH}$$

and observe that for $h \in D_\Omega$ both $a_t(h)$ and $a_t^*(h)$ leave D_0 invariant. Moreover, $a_t(h)f$ and $a_t^*(h)f$ for f in D_0 and h in D_Ω are both strongly differentiable in t, and by differentiation we get, for $h \in D_\Omega$, that

$$a_s(h) - a(h) = -i\int_0^s e^{-itH}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}\,dt,$$
$$a_s^*(h) - a^*(h) = -i\int_0^s e^{-itH}[V, e^{itH_0}a^*(h)e^{-itH_0}]e^{itH}\,dt. \tag{1}$$

For details see Ref. 1.

Theorem 2: Under Assumptions 1 and 2 we have that the strong limits of $a_t(h)$ and $a_t^*(h)$ as $t \to \pm\infty$ exist for all $h \in L_1(E_3)$. Denote

$$a_\pm(h) = \operatorname*{strong\,lim}_{t\to\pm\infty} a_t(h),$$
$$a_\pm^*(h) = \operatorname*{strong\,lim}_{t\to\pm\infty} a_t^*(h).$$

Then $a_\pm$ and $a_\pm^*$ satisfy the same anticommutation relations as do a and a^*:

$$\{a_\pm(h), a_\pm(g)\} = \{a_\pm^*(h), a_\pm^*(g)\} = 0,$$
$$\{a_\pm(h), a_\pm^*(g)\} = \int h(x)g(x)\,dx,$$

and the operators $a_\pm$, $a_\pm^*$, and H satisfy the same commutation relations as do a, a^*, and H_0; i.e., for $h \in D_\Omega$ we get

$$[H, a_\pm(h)] = a_\pm(-\Omega h), \quad [H, a_\pm^*(h)] = a_\pm^*(\Omega h),$$

and for $h \in D_\Omega$ we have that $a_\pm(h)$, as well as $a_\pm^*(h)$, leaves D_0 invariant.

Proof: The proof is similar to one given in Ref. 1 but for the fact that here strong limits substitute the norm limits in Ref. 1.

Since $\|a_t(h)\| = \|h\|_2$, we have that the mapping $a_t(h)f: L_2(E_3)x\mathcal{H} \to \mathcal{H}$ is uniformly bounded in t, so it is sufficient to prove that $a_t(h)f$ tends strongly to a limit for $h \in C_0^\infty$ and f in D_0. From (1) we get that this is equivalent to proving that the integrals on the right-hand side of (1) converge strongly at infinity. Consider, therefore,

$$\|e^{-itH_0}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}f\| = \|[V, a(h_t)]e^{itH}f\|,$$

where $h_t = e^{-it\Omega}h$.

By Lemma 4 and Assumptions 1 and 2, we get that this is bounded by

$$C\,\|h_t\|_\infty \cdot \|H_0 e^{itH}f\|,$$

where C is a constant depending only on V. According to a lemma proved in Ref. 1, we have that

$$\|h_t\|_\infty \leq C_1\,|t|^{-\frac{3}{2}}$$

for $h \in C_0^\infty$. Since H and H_0 have the same domain of definition, there exist two constants a and b, such that

$$\|H_0 f\| \leq a \|Hf\| + b \|f\|.$$

Using this, we get the new bound

$$C_2 |t|^{-\frac{3}{2}} (a \|Hf\| + b \|f\|),$$

which proves that the integrals on the right-hand side of (1) converge strongly.

That $a_\pm$ and $a_\pm^*$ satisfy the same anticommutation relations as do a and a^* follows from the strong convergence. Consider now, for $h \in D_\Omega$ and $f \in D_0$, the expression $a_t(h)f$. Since $a_t(h)$ leaves D_0 invariant, we have that $a_t(h)f \in D_0$ and thus

$$\begin{aligned} Ha_t(h)f &= [H, a_t(h)]f + a_t(h)Hf, \\ [H, a_t(h)]f &= e^{-itH}[H, e^{itH_0}a(h)e^{-itH_0}]e^{itH}f \\ &= e^{-itH}e^{itH_0}a(-\Omega h)e^{-itH_0}e^{itH}f \\ &\quad + e^{-itH}[V, e^{itH_0}a(h)e^{-itH_0}]e^{itH}f. \end{aligned}$$

We have already proved that the last term tends strongly to zero and the first term is $a_t(-\Omega h)f$, which we know converges strongly to $a_\pm(-\Omega h)f$. Using the fact that H is closed, we get that $a_\pm(h)f \in D_0$, i.e., is in the domain of H, and that $[H, a_\pm(h)]f = a_\pm(-\Omega h)f$. This proves the theorem.

Theorem 3: The Hilbert space $\mathcal{H}$ decompose in two ways as a tensor product of two Hilbert spaces

$$\mathcal{H} = \mathcal{F}_\pm \otimes \mathcal{H}_\pm^0.$$

$\mathcal{F}_\pm$ is the Fock representation of the incoming (outgoing) asymptotic fields. $\mathcal{H}_\pm^0$ is the incoming (outgoing) zero-particle space; i.e., the subspace of $\mathcal{H}$ of the form $\Omega_\pm \otimes \mathcal{H}_\pm^0$ is annihilated by all incoming (outgoing) annihilation operators.

In accordance with this decomposition, the energy operator H decomposes as a sum:

$$H = H_0^\pm \otimes 1 + 1 \otimes H_\pm^0,$$

where $H_0^\pm$ is the free-energy operator with mass m in $\mathcal{F}_\pm$ and $H_\pm^0$ is the restriction of H to $\Omega_\pm \otimes \mathcal{H}_\pm^0$. $\Omega_\pm$ is the vacuum of $\mathcal{F}_\pm$.

Proof: The proof of this theorem is the same as in Ref. 1, since only the fact that H was bounded below and the strict positivity of the mass m, together with the commutation relation between H and the asymptotic fields, were used in the proof of this theorem in Ref. 1.

Theorem 4: The Møller-wave operators $W_\pm$, defined as

$$W_\pm = \operatorname*{strong\,lim}_{t \to \pm\infty} e^{-itH}e^{itH_0},$$

exist under Assumptions 1 and 2 and define a unitary equivalence between H_0 in $\mathcal{H}$ and H in $\mathcal{H}_0^\pm$. Here $\mathcal{H}_0^\pm$ is the smallest closed subspace containing the vacuum element ϕ_0 of $\mathcal{H}$ and invariant under $a_\pm^*(h)$ for all h in $L_2(E_3)$.

Remark: Such a theorem was first proved by Chistyakov,[5] but for the sake of completeness we give a proof below. It is interesting to observe here how much more information one gets on the spectral structure of H from the existence of asymptotic annihilation–creation operators than from the existence of the wave operators. From the existence of the wave operators one may conclude that there is a subspace of $\mathcal{H}$ in which H acts as a free-energy operator. But as we have seen in Theorem 3, from the existence of the asymptotic annihilation–creation operators one may conclude that $\mathcal{H}$ decomposes as a tensor product, so that H acts as a free-energy operator in one of the factors.

Proof: Since ϕ_0 is cyclic in $\mathcal{H}$—relative to the set of operators $a^*(h)$, $h \in L_2(E_3)$—it is enough to prove that $e^{-itH}e^{itH_0}a^*(h_1) \cdots a^*(h_n)\phi_0$ converges strongly. But since H_0 and H both annihilate ϕ_0, this is equal to $a_t^*(h_1) \cdots a_t^*(h_n)\phi_0$, and this we know converges strongly.

This proves the theorem.

---

[5] A. L. Chistyakov, Dokl. Akad. Nauk SSSR **158**, 66 (1964).

[6] Y. Kato and N. Mugibayashi, Progr. Theoret. Phys. (Kyoto) **30**, 103 (1963).

Ground-State Many-Boson Problem with Repulsive Potentials

LEONARD EYGES
Air Force Cambridge Research Laboratories, Office of Aerospace Research
L. G. Hanscom Field, Bedford, Massachusetts

(Received 31 July 1968)

We extend to the N-body case previous techniques for solving three-body problems with repulsive interparticle potentials and periodic boundary conditions on each particle. For clarity, we begin with one-dimensional problems, although the techniques are not peculiar to them and can be generalized to three dimensions. We decompose the wavefunction Ψ into $N(N-1)/2$ parts, according to

$$\Psi = \sum_{i<j=1}^{N} \psi_{ij};$$

from the Schrödinger equation we find a basic equation for a typical ψ_{ij}. As a test of this equation, we apply it for $N = 4$ and the case of δ-function interactions, and solve it numerically to find close agreement between the energy per particle of a four-body system and a system of the same density in which $N \to \infty$. The numerical results are consistent with the analytic one that $E_g(\infty)$, which is the ground-state energy per particle of the system with an infinite number of particles, is related to $E_g(N)$, which is that energy for the N-body system of the same density, by $E_g(N) = E_g(\infty)[1 - (1/N)]$ for weak repulsion and by $E_g(N) = E_g(\infty)[1 - (1/N^2)]$ for very strong repulsion. We begin a similar comparison for the three-dimensional hard-sphere case by working out the problem of two hard spheres with periodic boundary conditions over a length L_3 such that the density is the same as for an N-body problem in volume $V = L_N^3$: $2/L_3^3 = N/L_N^3$. We find that we get the analog of the one-dimensional result, for which case $E_g(2) = E_g(\infty)/2$ for $N = 2$. That is, we find $E_g(2) = \pi a^3\hbar/mv$, which is just one-half of the many-body theory result. We also calculate the two-body correlation function and find good agreement with the many-body one, as calculated by Lee, Huang, and Yang.

I. INTRODUCTION

In previous papers[1,2] we have presented a new method for solving the quantum-mechanical problem of finding the ground-state energy and wavefunction of a system of three interacting particles. One of its essential novelties was the decomposition of the wavefunction into three parts, one for each pair of interparticle interactions. Among other applications,[3,4] one has been to the problem of particles with repulsive δ-function interactions[2] in one dimension. It is then natural to ask whether or not this work can be extended to more than three bodies, possibly even to the N-body problem.

In this paper, we give a partial answer to this question. In particular, we show that for repulsive interparticle interactions in one dimension, the equations can be generalized to the N-particle case; moreover, the derivation can be extended to three dimensions. As a test of the equations that we derive, we have applied them for $N = 4$ to one-dimensional δ-function interparticle interactions. This serves as a useful check, since we expect the numerical results (whose nature is discussed below) to lie between those that are at hand for $N = 3$ and those for very large N; in fact, they do. The problem can also be solved for $N = 2$, so that we now can compare results for $N = 2$, 3, 4 and $N \to \infty$.

The main result we are interested in is the energy per particle for systems with different N, but with the *same* linear density. The linear density is defined to be N/L_N, where L_N is the length of the "box" over which periodic boundary conditions are applied. We have already made this comparison[2] between the cases $N = 3$ and $N \to \infty$ and found surprisingly close agreement. In the present paper, we add the results for $N = 4$ and for $N = 2$; with them the general trend is quite striking; we find, for example, that the energies per particle for $N = 4$ and $N \to \infty$ usually agree (as a function of the strength of interaction) to within ten or fifteen per cent. Some analytic light is thrown on this numerical work by the fact that we can find formulas for the energy per particle as a function of N in the two limits of very weak and very strong interactions. For weak interactions, it turns out that this quantity differs from its value for $N \to \infty$ by the factor $[1 - (1/N)]$ and for very strong interactions by $[1 - (1/N^2)]$. Presumably, interactions of intermediate strength are bracketed by these limits. Thus, if we seek to approximate the many-body results by the few-body ones, the worst possible case is $N = 2$ and weak interactions for which case the answer is off by a factor two; but the accuracy increases markedly as either N or the strength of the interaction increases.

With these results for the one-dimensional problem,

[1] Leonard Eyges, Phys. Rev. **115**, 1643 (1959); Phys. Rev. **121**, 1744 (1961); J. Math. Phys. **6**, 1320 (1965).
[2] Leonard Eyges, J. Math. Phys. **7**, 938 (1966).
[3] Leonard Eyges and John R. Jasperse, J. Math. Phys. **9**, 805 (1968).
[4] J. R. Jasperse and M. H. Friedman, Phys. Rev. **159**, 69 (1967).

it is natural to ask if there are analogous ones for three-dimensional problems with repulsive potentials, of which the so-called hard-core potential is an example. For such repulsive potentials in three dimensions, the three, four, and many-body problems are, of course, not as tractable as they were for δ functions in one dimension; but we can solve the two-body problem. For we have shown previously in one dimension that the solution of the problem of *two particles* with repulsive interparticle interactions and periodic boundary conditions over length L on *each* coordinate is reducible to the *one-body* problem of a particle moving in a periodic lattice generated by putting one of the potentials at each point of a one-dimensional lattice of lattice spacing L.

A similar result holds in three dimensions, and we can show that the problem of two hard spheres with periodic boundary conditions on each coordinate is the same as that of a point particle (of reduced mass $M/2$) moving in the periodic field generated by putting one hard sphere at each lattice point of a simple cubic lattice with lattice constant L. Now this is a problem that can be solved, to the accuracy we seek, by the Wigner–Seitz method; and we do this. We get expressions for the energy per particle of this two-body system and for the two-body correlation function, which is the relative probability for finding the two particles within a given distance of each other.

The only available results for the N-body problem, with which we might compare these, are the perturbation ones for small hard spheres[5]; these presumably correspond to the weak-interaction case in one dimension. The comparison is quite interesting. First, for the energy we find essentially the same result here as we did in one dimension: The energy per particle for the two-body problem is just *half* that for the N-body problem. If the one-dimensional case is any guide, this result should improve as the strength of the interaction, i.e., the hard-sphere radius, increases. Secondly, we have compared the two-body correlation function (as derived in Ref. 5) for the N-body system and as calculated for the two-body system in the way we have indicated. The agreement between these two functions is surprisingly good. The function as calculated from the two-body problem serves as a quite respectable approximation to that derived from the N-body one.

II. N-BODY EQUATIONS IN ONE DIMENSION

We start with the problem of N identical particles of mass m, coordinates $x_1, x_2, \cdots, x_N$, and with *total* potential energy v_t given by

$$v_t = \sum_{i<j} v_{ij}, \tag{1}$$

where v_{ij} is shorthand for $v(|x_i - x_j|)$. The Schrödinger equation for the system is then

$$\left(-\sum_{n=1}^{N} \frac{\partial^2}{\partial x_n^2} + \sum_{i<j} v_{ij}\right)\Psi = E\Psi.$$

As in our previous work, it is convenient to introduce other sets of coordinates in place of the set $x_1, \cdots, x_N$. In fact, we shall introduce one such set for each potential v_{ij} and, therefore, $N(N-1)/2$ sets altogether. For the three-body problem, these sets were the so-called Jacobi coordinates, and these can be defined for the N-particle system as well; but it turns out that they are inconvenient in that they tend to camouflage the symmetry of the ground-state N-body wavefunction. This defect is not shared by the sets we now describe.

Consider then the following coordinates associated with the potential v_{ij}. We first write the original coordinates in their natural order:

$$x_1, x_2, x_3, \cdots, x_N.$$

For a given i and j, we then replace x_1 and x_2 by their relative and center-of-mass counterparts x_{ij} and X_{ij}:

$$x_{ij} = x_i - x_j, \quad X_{ij} = x_i + x_j;$$

then we replace x_i and x_j by x_1 and x_2, respectively, to give the ijth of the $N(N-1)/2$ sets

$$x_{ij}, X_{ij}, x_3, \cdots, x_{i-1}, x_1, \cdots, x_{j-1}, x_2, \cdots, x_N.$$

Given these sets, one of the basic ideas of our general method is to write the wavefunction Ψ as a sum of partial wavefunctions or two-body orbitals ψ_{ij},

$$\Psi = \sum_{i<j} \psi_{ij}, \tag{2}$$

wherein we take the ijth orbital to be a function of the set appropriate to it:

$$\psi_{ij} = \psi_{ij}(x_{ij}, X_{ij}, \cdots).$$

The functions ψ_{ij} can be taken to satisfy[6]

$$-\sum_{n=1}^{N} \frac{\partial^2 \psi_{ij}}{\partial x_n^2} + v_{ij}\Psi = E\psi_{ij}, \tag{3}$$

since, if we sum this over i and j and use (2), we see this equation is equivalent to the original Schrödinger equation.

Moreover, it is moderately clear (from the fact that Ψ is a symmetrical function of all its variables and from the symmetrical way of defining coordinates) that all the ψ_{ij} are identical functions of their respective variables; e.g., ψ_{34} is the same function of $x_{34}, X_{34} \cdots$

[5] T. D. Lee, K. Huang, and C. N. Yang, Phys. Rev. **106**, 1135 (1957).

[6] We set $\hbar = 2m = 1$.

as ψ_{29} is of $x_{29}, X_{29}\cdots$. This point is discussed in some detail in the references,[1,2] so we shall simply assume it here. With this, we see that we have only one unknown function to find and the index ij on ψ_{ij} does not label *different* functions but labels a *single function* but of different variables.

We turn our attention then to Eq. (3) for the orbital ψ_{ij}. Since all the orbitals are of the same form, we can be specific and consider the equation for ψ_{12}; and since this turns out to be more convenient for the bookkeeping, we shall assume this done:

$$\psi_{ij} \to \psi_{12}. \tag{4}$$

We want the wavefunction to be periodic with period L in each of the variables $x_1, x_2, \cdots, x_N$, and this will be insured if each of the orbitals is similarly periodic. We assume such periodicity and so expand ψ_{12} in Fourier series, i.e., with basis functions

$$\exp \frac{2\pi i}{L}(m_1x_1 + \cdots + m_Nx_N). \tag{5}$$

We introduce x_{12} and X_{12}, and let $m_1 - m_2 = n_1$, $m_1 + m_2 = n_2$, which implies that n_1 and n_2 are both even or both odd. We shall frequently express this condition by using the function $\Delta(s, t)$ we have introduced previously:

$$\Delta(s,t) = \begin{cases} 1; & s, t \text{ both even or both odd,} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

With this, the basis function (5) can be written

$$\exp\left[\frac{2\pi i}{2L}(n_1x_{12} + n_2X_{12} + 2m_3x_3 + \cdots + 2m_Nx_N)\right] \times \Delta(n_1, n_2). \tag{7}$$

We have one further requirement. We want Ψ to be a wavefunction for a system whose center of mass is at rest; i.e., we want it to be an eigenfunction of the center-of-mass operator $\hat{P}_{cm} = -i\sum_{i=1}^{N} \partial/\partial x_i$ with eigenvalue zero. This will be so if each basis function is a similar eigenfunction. Applying this condition to (5), we find the requirement

$$n_2 + M = 0,$$

where

$$M = m_3 + m_4 + \cdots + m_N.$$

We can then eliminate n_2 from (7) to get the final basis function appropriate to periodic boundary conditions and zero momentum of the center of mass:

$$\exp\left[\frac{2\pi i}{2L}(n_1x_{12} - MX_{12} + 2m_3x_3 + \cdots + 2m_Nx_N)\right] \times \Delta(n_1, M). \tag{8}$$

If then we sum[7] over all such functions with coefficients $C(n_1, m_3, m_4, \cdots, m_N)$, we can conclude that the general expression for the orbital ψ_{12} that yields a *periodic wavefunction with center of mass at rest* is

$$\psi_{12} = \sum_{n_1, m_3, \cdots, m_N} C(n_1, m_3, m_4, \cdots, m_N) \times \exp\left[\frac{2\pi i}{2L}(n_1x_{12} - MX_{12} + 2m_3x_3 + \cdots + 2m_Nx_N)\right]\Delta(n_1, M). \tag{9}$$

Since the ψ_{ij} orbital is the same function of $x_{ij}, X_{ij}, \cdots$ that ψ_{12} is of $x_{12}, X_{12}, \cdots$, we can write

$$\psi_{ij} = \sum_{n_1, m_3, \cdots, m_{N'}} C(n_1, m_3, m_4, \cdots, m_N) \times \exp\frac{2\pi i}{2L}(n_1x_{ij} - MX_{ij} + 2m_3x_3 + \cdots + 2m_ix_1 + \cdots + 2m_jx_2 + \cdots + 2m_Nx_N) \times \Delta(n_1, M). \tag{10}$$

There is one more point to make before we begin to reduce Eq. (3). As we have discussed previously, we must expand the potential v_{12} in Fourier series, that is, replace it by its periodic counterpart. We write then

$$v(x_{12}) = \sum_l U(l) \exp\left(\frac{i2\pi x_{12}l}{L}\right). \tag{11}$$

If then we put Eqs. (10) and (11) into Eq. (3), we get (on putting primes on the summation variables for convenience and remembering that $ij \to 12$) the following:

$$\sum_{n_1', m_3' \cdots m_N}\left[E - \left(\frac{2\pi}{L}\right)^2\left(\frac{n_1'^2 + M'^2}{2} + m_3'^2 + \cdots + m_N'^2\right)\right] \times C(n_1', m_3', m_4' \cdots m_N')\Delta(n_1', M') \times \exp\left[\frac{2\pi i}{2L}(n_1'x_{12} - M'X_{12} + 2m_3'x_3 + \cdots + 2m_N'x_N)\right]$$

$$= \sum_{l, n_1', m_3', \cdots, m_{N'}} U(l) \exp\left(\frac{i2\pi x_{12}l}{L}\right) C(n_1', m_3' \cdots m_N') \times \sum_{i<j=1}^{N} \exp\left[\frac{2\pi i}{2L}(n_1'x_{ij} - M'X_{ij} + 2m_3'x_3 + \cdots + 2m_i'x_1 + \cdots + 2m_j'x_2 + \cdots + 2m_N'x_N)\right] \times \Delta(n_1', M'). \tag{12}$$

[7] A sum simply written as $\sum_s$ will mean $\sum_{s=-\infty}^{\infty}$, and similarly for multiple sums.

Now we multiply this by

$$\exp\left[-\frac{2\pi i}{2L}(n_1x_{12} - MX_{12} + 2m_3x_3 + \cdots + 2m_Nx_N)\right]$$

and integrate over all variables from zero to $2L$. The left-hand side of Eq. (12) then becomes simply

$$\left[E - \left(\frac{2\pi}{L}\right)^2\left(\frac{n_1^2 + M^2}{2} + m_3^2 + \cdots + m_N^2\right)\right] \times C(n_1, m_3, \cdots, m_N)\Delta(n_1, M).$$

On the right-hand side, we must do the typical integral:

$$\int\cdots\int \exp\left[\frac{2\pi i}{2L}\{(2l - n_1)x_{12} + MX_{12} - 2m_3x_3 + \cdots - 2M_Nx_N + n_1'x_{ij} - M'X_{ij} + 2m_3'x_3 + \cdots + 2m_i'x_1 + \cdots + 2m_j'x_2 + \cdots + 2m_N'x_N\}\right] \times dx_{12}\,dx_3\cdots dx_N. \quad (13)$$

Although there are $N(N-1)/2$ such integrals, they break up into four different types depending on whether x_{ij} has both indices in common with x_{12}, one index in common, or none. That is, there are integrals of the following types:

I. one integral for which $x_{ij} = x_{12}$;
II. $N - 2$ integrals for which $x_{ij} = x_{1j}$;
III. $N - 2$ integrals for which $x_{ij} = x_{i2}$;
IV. $\frac{N^2 - 5N}{2} + 3$ integrals for which neither $i = 1$, nor $j = 2$.

The integral for which $x_{ij} = x_{12}$ is trivial and leads to the following term on the right-hand side of (12):

$$\sum_{n_1'} U\left(\frac{n_1 - n_1'}{2}\right)C(n_1', m_3, m_4, \cdots, m_N)\Delta(n_1', M).$$

Now consider a typical integral of type II. For convenience, we take the case $j = 3$; this is no real loss of generality, and it has the advantage of permitting more compactness in the writing. To do the integral, we must transform the expression in x_{13}, X_{13} variables in curly brackets, that is, in the exponent of the integrand of (13) to x_{12}, $X_{12}\cdots$ variables. On so doing, it becomes

$$n_1'\left(\frac{X_{12} + x_{12}}{2} - x_3\right) - M'\left(\frac{X_{12} - x_{12}}{2} + x_3\right) + 2m_3'\left(\frac{X_{12} - x_{12}}{2}\right) + 2m_4'x_4 + \cdots 2m_N'x_N,$$

and the integral we have to do becomes the (N-fold)-one:

$$\int\cdots\int \exp\left[\frac{2\pi i}{2L}\left\{x_{12}\left(2l - n_1 + \frac{n_1' + M'}{2} - m_3'\right) + X_{12}\left(M + \frac{n_1' - M'}{2} + m_3'\right) + x_3(-2m_3 - n_1' - M') + x_4(m_4 - m_4') + \cdots + x_N(m_N - m_N')\right\}\right] dx_{12}\,dX_{12}\,dx_3\cdots dx_N. \quad (14)$$

Each of the N integrations yields a δ function and we end up with a set of N simultaneous equations:

$$2l - n_1 + \frac{n_1' + M'}{2} - m_3' = 0, \quad (15a)$$

$$M + \frac{n_1' - M'}{2} + m_3' = 0, \quad (15b)$$

$$2m_3 + n_1' + M' = 0, \quad (15c)$$

$$m_4 = m_4', \quad (15d)$$

$$\vdots \qquad \vdots$$

$$m_N = m_N'. \quad (15N)$$

These equations can all be satisfied in the following way. Eqs. (15d) through (15N) imply that

$$M' = M - m_3 + m_3'.$$

With this Eq. (15c) becomes

$$-m_3' = m_3 + n_1' + M,$$

which enables us to eliminate a summation over m_3' in Eq. (12) in terms of one over n_1'. With the above results for m_3' and M', Eq. (15b) turns out to be the same as (15c), i.e., it is satisfied identically. Finally, Eq. (15a) becomes

$$l = (n_1 + n_1' + M)/2,$$

which enables us to eliminate l from the summation on the right-hand side of Eq. (12). Thus, the multiple sum there becomes simply a sum over n_1' and finally we see that, corresponding to the integral (14), there is a term on the right-hand side of Eq. (12) with the form

$$\sum_{n_1'} U\left(\frac{n_1 + n_1' + M}{2}\right) \times C(n_1', -m_3 - n_1' - M, m_4, m_5\cdots m_N).$$

Of course, we get a similar term for other values of j than $j = 3$, and, *mutatis mutandis*, similar results for integrals of type III. In fact, all the integrals of type

II and III yield terms for the right-hand side of Eq. (12) of the form

$$\sum_{s=3}^{N}\sum_{n_1'}\left[U\left(\frac{n_1+n_1'+M}{2}\right)+U\left(\frac{n_1-n_1'-M}{2}\right)\right] \times C(n_1', m_3, m_4\cdots m_{s-1}, -M-m_s-n_1', m_{s+1}\cdots m_N). \quad (16)$$

Now we consider integrals of type IV. Again for ease of writing we take a typical one, say that for $i=3$, $j=4$. On working it out, we find it leads to a set of equations analogous to (15), and by adding and subtracting a pair of them they reduce to

$$2l-n_1+m_3'-m_4'=0, \quad (17a)$$

$$M+m_3'+m_4'=0, \quad (17b)$$

$$M'+m_3+m_4=0, \quad (17c)$$

$$n_1'-m_3+m_4=0, \quad (17d)$$

$$m_5'=m_5, \quad (17e)$$

$$\vdots$$

$$m_N'=m_N. \quad (17N)$$

These are not all independent, since it is easy to see that if (17e) through (17N) are satisfied and (17b) is satisfied, then (17c) will be satisfied as well. Eq. (17d) enables us to eliminate n_1' from the summation, and so we are left with Eqs. (17a) and (17b) in the three unknowns l, m_3', and m_4'. We can choose to eliminate two of these in terms of the third. If we eliminate l and m_3' in favor of m_4', we get

$$m_3'=-M-m_4',$$

$$l=m_4'+[(n_1+M)/2].$$

This leads to a term in the right-hand side of Eq. (12) of the form

$$\sum_{m_4'}U\left(m_t'+\frac{n_1+M}{2}\right) \times C(m_3-m_4, -m_4'-M, m_4', m_5, m_6\cdots m_N).$$

In a similar way, we can calculate the other integrals of type IV and find that they contribute the following to the right-hand side of Eq. (12):

$$\sum_{t<s=3}^{N}\sum_{m_t'}U\left(m_t'+\frac{n_1+M}{2}\right) \times C(m_s-m_t, m_4, m_5\cdots m_{s-1}, -m_t'-M, m_{s+1}\cdots m_{t-1}, m_t', m_{t+1}\cdots m_N).$$

Thus finally, our basic set of equations, written out in full for the function $C(n_1, m_3, m_4\cdots m_N)$, is

$$\left[E-\left(\frac{2\pi}{L}\right)^2\left(\frac{n_1^2+M^2}{2}+m_3^2+\cdots m_N^2\right)\right]C(n_1, m_3, \cdots, m_N)\Delta(n_1, M)$$

$$=\sum_{n_1'}U\left(\frac{n_1-n_1'}{2}\right)C(n_1', m_3, m_4, \cdots, m_N)\Delta(n_1', M)$$

$$+\sum_{s=3}^{N}\sum_{n_1'}\left(U\left(\frac{n_1+n_1'+M}{2}\right)+U\left(\frac{n_1-n_1'-M}{2}\right)\right)C(n_1', m_3, m_4\cdots m_{s-1}, -M-m_s-n_1', m_{s+1}\cdots m_N)$$

$$+\sum_{t<s=3}^{N}\sum_{m_t'}U\left(m_t'+\frac{n_1+M}{2}\right)C(m_s-m_t, m_4, m_5\cdots m_{s-1}, -m_t'-M, m_{s+1}\cdots m_{t-1}, m_t', m_{t+1}\cdots m_N). \quad (18)$$

III. ONE-DIMENSIONAL FOUR-BODY PROBLEM WITH δ-FUNCTION INTERACTIONS; FEW-BODY APPROXIMATIONS TO THE MANY-BODY PROBLEM

In this section, we test and apply Eqs. (18) by first writing them for $N=4$, and then further specializing these equations to the case of δ-function interparticle interactions. We then solve the resulting equations numerically to find the ground-state energy of the four-body system as a function of δ-function strength. With this, with previous results for $N=3$ and 2, and with those of Lieb and Liniger[8] for $N\to\infty$, we can compare the ground-state energy per particle as a function of N, extending a previous comparison of this kind.

[8] E. Lieb and W. Liniger, Phys. Rev. 130, 1605 (1963).

We begin then by writing out Eqs. (18) for $N = 4$. With $M = m_3 + m_4$, they become

$$\left[E - \left(\frac{2\pi}{L}\right)^2\left(\frac{n_1^2 + M^2}{2} + m_3^2 + m_4^2\right)\right]C(n_1, m_3, m_4)\Delta(n_1, M)$$
$$= \sum_{n_1'} U\left(\frac{n_1 - n_1'}{2}\right)C(n_1', m_3, m_4)\Delta(n_1', M)$$
$$+ \sum_{n_1'}\left(U\left(\frac{n_1' - n_1 + 3m_3 + m_4}{2}\right) + U\left(\frac{n_1' + n_1 + m_4 + 3m_3}{2}\right)\right)C(n_1', -n_1' - 2m_3 - m_4, m_4)$$
$$+ \sum_{n_1'}\left(U\left(\frac{n_1' - n_1 + m_3 + 3m_4}{2}\right) + U\left(\frac{n_1' + n_1 + m_3 + 3m_4}{2}\right)\right)C(n_1', -n_1' - m_3 - 2m_4, m_3)$$
$$+ \sum_{m_t'} U\left(m_t' + \frac{n_1 + M}{2}\right)C(m_3 - m_4, -m_3 - m_4 - m_t', -m_t'). \tag{19}$$

For the special case of δ-function interactions, i.e., those for which

$$v_{ij} = t\delta(x_{ij}),$$

the Fourier coefficients $U(n)$ are constant:

$$U(n) = \frac{t}{L}.$$

Then Eq. (19) simplifies to

$$\left[E - \left(\frac{2\pi}{L}\right)^2\left(\frac{n_1^2 + M^2}{2} + m_3^2 + m_4^2\right)\right]C(n_1, m_3, m_4)\Delta(n_1, m_3 + m_4)$$
$$= \frac{t}{L}\sum_{n_1'} C(n_1', m_3, m_4) + \frac{2t}{L}\sum_{n_1'}(C(n_1', -n_1' - 2m_3 - m_4, m_4) + C(n_1', -n_1' - m_3 - 2m_4, m_3))$$
$$+ \frac{t}{L}\sum_{n} C(m_3 - m_4, n, -n - m_3 - m_4). \tag{20}$$

We see that for this case we can reduce the problem to one for a function $F(m_3, m_4)$ of two variables by the ansatz

$$C(n_1, m_3, m_4) = F(m_3, m_4)\Big/\left\{E - \left(\frac{2\pi}{L}\right)^2\left(\frac{n_1^2 + M^2}{2} + m_3^2 + m_4^2\right)\right\}. \tag{21}$$

Putting (21) into (20) leads to the equation for $F(m_3, m_4)$:

$$F(m_3, m_4)\Delta(n_1, m_3 + m_4)$$
$$= \frac{t}{L}F(m_3, m_4)\sum_{n_1'}\Delta(n_1', m_3 + m_4)\Big/\left\{E - \left(\frac{2\pi}{L}\right)^2\left[\frac{n_1'^2}{2} + \tfrac{1}{2}(m_3 + m_4)^2 + m_3^2 + m_4^2\right]\right\}$$
$$+ \frac{2t}{L}\sum_{n_1'}F(-n_1' - 2m_3 - 2m_4, m_4)\Big/\left\{E - \left(\frac{2\pi}{L}\right)^2\left[\frac{n_1'^2}{2} + \tfrac{1}{2}(n_1' + 2m_3)^2 + (n_1' + 2m_3 + m_4)^2 + m_4^2\right]\right\}$$
$$+ \frac{2t}{L}\sum_{n_1'}F(-n_1' - m_3 - 2m_4, m_3)\Big/\left\{E - \left(\frac{2\pi}{L}\right)^2\left[\frac{n_1'^2}{2} + \tfrac{1}{2}(n_1' + 2m_4)^2 + (n_1' + m_3 + 2m_4)^2 + m_3^2\right]\right\}$$
$$+ \frac{2t}{L}\sum_{n}F(n, -n - m_3 - m_4)\Big/\left\{E - \left(\frac{2\pi}{L}\right)^2\left[\frac{(m_3 + m_4)^2}{2} + \tfrac{1}{2}(m_3 + m_4)^2 + n^2 + (n + m_3 + m_4)^2\right]\right\}. \tag{22}$$

Now it is convenient to change some of the names of the summation variables. In the first sum on the right-hand side, we replace n_1' by n; in the other sums on that side, we replace $n_1' + 2m_3 + m_4$ by n. After some algebra, and with the definition

$$E = 2\left(\frac{\pi}{L}\right)^2 E',$$

we then get the equation

$$2\pi^2 F(m_3, m_4) = tLF(m_3, m_4) \sum_n \Delta(n, m_3 + m_4)/\{E' - [n^2 + (m_3 + m_4)^2 + 2m_3^2 + 2m_4^2]\}$$
$$+ tL \sum_n [2F(n, m_4) + 2F(n, m_3) + F(n, -n - m_3 - m_4)]/\{E' - 4[m_3^2 + m_4^2 + n^2 + m_3 m_4 + m_3 n + m_4 n]\}. \quad (23)$$

This is as far as we can go algebraically, and at this point we have solved Eq. (23) numerically on a digital computer. This equation really represents an infinite set of homogeneous ones which has a solution only if its determinant vanishes; the ground-state energy is found by searching for the smallest root of E for a given t. To solve Eqs. (23) in practice we have, of course, been forced to truncate them; in so doing, we have kept in mind the symmetry relationships of the coefficients $F(m, n)$ which follow directly from Eqs. (23). These are

$$F(m, n) = F(n, m),$$
$$F(-n, -m) = F(n, m).$$

To test the effect of truncation, we have solved the equations in different orders of truncation and have kept as few as four and as many as forty-nine of the coefficients. In Table I, we give some of the values thus found for the ground-state energy as a function of t. The numerical results we have obtained are accurate to more figures than we have thought necessary to put down here. As we shall shortly see, in the limit of infinite δ-function strength the quantity $E_g L^2/2\pi^2$ approaches the value 10, so that the larger values listed for this quantity can be considered to be for rather strong interactions. In Table II, we present an example of the numerical results for the coefficients $F(m_3, m_4)$.

With these results, we can now extend a comparison we have made previously for the three-body problem.

TABLE I. Ground-state energy E_g vs δ-function strength t for the four-body problem in one dimension.

$\frac{E_g L^2}{2\pi^2}$	tL
1.00	3.768
3.30	17.20
5.00	34.81
7.20	87.29
8.40	174.5

We take systems of different number of particles N but of the same linear density ρ and compare the energy per particle as a function of delta-function strength. We do this for $N = 2$, 3, 4, and $N \to \infty$. The case $N \to \infty$ is got from the work of Lieb and Liniger,[8] that for $N = 4$ from the present paper, and that for $N = 3$ and $N = 2$ from a previous one. For convenience, we plot $E_g/N\rho^2$ vs $t\rho/2$, which essentially compares energy per particle vs δ-function strength, since ρ is constant for the different systems. The results are given in Fig. 1.

In addition to these numerical results, it is worth looking analytically at two extremes of δ-function strength. First, for small t, we can get a perturbation result as follows. The ground-state wavefunction for periodic boundary conditions over length L_N is, for a single particle, just the constant function $1/(L_N)^{\frac{1}{2}}$, and for N particles is the product of N such functions.

TABLE II. The function $F(m_3, m_4)$ found by solving Eq. (23), for the case $E_g L^2/2\pi^2 = 5.00$. The number in parentheses in some boxes is the power of ten by which the quantity that stands to the left of it is to be multiplied.

m_4 \ m_3	0	1	2	3	4	5
5						9.21 (−3)
4					0.01456	5.63 (−3)
3				0.02650	9.34 (−3)	4.32 (−5)
2			0.06566	0.01902	5.27 (−6)	2.88 (−4)
1		0.09249	0.08838	0.01024	5.63 (−3)	3.54 (−3)
0	1.000	0.2655	0.06456	0.02230	0.01193	7.45 (−3)
−1		0.1169	0.01549	7.04 (−3)	4.14 (−3)	2.69 (−3)
−2			8.85 (−4)	4.46 (−4)	1.27 (−4)	2.02 (−4)
−3				8.60 (−5)	8.97 (−5)	1.25 (−4)
−4					4.89 (−5)	8.94 (−5)
−5						1.19 (−4)

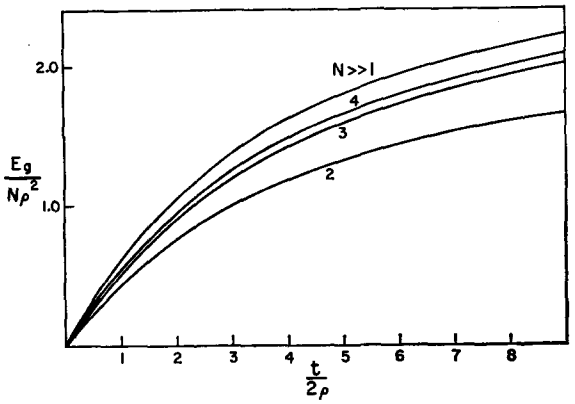


FIG. 1. Ground-state energy per particle E_g/N for systems of repulsive delta functions of strength t, with $N = 2$, 3, 4 and $N \gg 1$ particles. The linear density ρ is the same for all systems.

The first-order perturbation result for the interacting system is then

$$E_g = \frac{N(N-1)}{2}\int \psi^2 t\delta(x)\,dx = \frac{N(N-1)t}{2L_N}.$$

For an N-body system with density $\rho = N/L_N$ the *energy per particle* E_g/N is therefore

$$\frac{E_g}{N} = \left(1 - \frac{1}{N}\right)\frac{t\rho}{2}. \tag{24}$$

Thus, the energy per particle of the system with $N = 2$ is one-half that of the system of the same density for which $N \to \infty$, and for $N = 3$, $\frac{2}{3}$ of it, etc. Now, we consider the opposite limit, for infinite δ-function strength. Here the problem has been solved by Girardeau,[9] who gives the following formula[10]:

$$E_g = \frac{1}{6}\left(N - \frac{1}{N}\right)\frac{\pi^2\hbar^2N^2}{mL_N^2}.$$

For a given density ρ we find for the energy per particle, in our units,

$$\frac{E_g}{N} = \frac{\pi^2\rho^2}{3m}\left(1 - \frac{1}{N^2}\right). \tag{25}$$

For $N = 2$, this energy is $\frac{3}{4}$ of that for the infinite system, for $N = 3$ it is $\frac{8}{9}$, etc. In summary then, if we approximate the energy per particle of a many-body system by that for a few-body one, the error is at worst a factor of two (two-body system and weak potentials) and the error rapidly diminishes as the potential strength and/or the number of bodies are increased.

[9] M. Girardeau, J. Math. Phys. 1, 516 (1960).

[10] This result holds in fact only for odd n, but we shall use it in the spirit of an interpolation formula for even n as well.

IV. HARD SPHERES IN THREE DIMENSIONS AND A TWO-BODY APPROXIMATION TO THE N-BODY PROBLEM

The one-dimensional problems discussed above are not, of course, of direct physical interest. For treating a physical system, as for example the ground state of liquid helium, we must discuss three-dimensional problems. In the last analysis, we want to be able to treat the problem of N particles "in a box" of volume $V = L_N^3$, with a potential which acts between each pair of particles, and with the understanding that "in a box" means that the wavefunction is periodic over length L_N in each of its $3N$ coordinates. We can't, of course, solve this problem directly, but, motivated by the close approximation set out above between the N-body and few-body problems in one dimension, we can take a similar tack in three dimensions. We shall then begin by considering two particles which, in the above sense, are in a box of length L_2 and for which the density is the same as for the N-body problem, i.e., for which

$$\frac{2}{L_2^3} = \frac{N}{L_N^3} = \rho.$$

If the coordinates of the particles are x_1, y_1, z_1 and x_2, y_2, z_2, and the wavefunction is $\psi(x_1, y_1, z_1, x_2, y_2, z_2)$ or, as we shall write more succinctly, $\psi(\mathbf{r}_1, \mathbf{r}_2)$, this condition of periodicity means, for example,

$$\psi(x_1 + L_2, y_1, z_1, x_2, y_2, z_2) = \psi(x_1, y_1, z_1, x_2, y_2, z_2) \tag{26}$$

with a similar condition on the other five coordinates. The Schrödinger equation that must be solved, subject to (26), is

$$\left[-\frac{\hbar^2}{2m}(\nabla_1^2 + \nabla_2^2) + v(r_{12})\right]\psi(\mathbf{r}_1, \mathbf{r}_2) = E\psi(\mathbf{r}_1, \mathbf{r}_2), \tag{27}$$

where $v(r_{12})$ is the central potential, possibly with hard core, that acts between the particles.

Now we recall that in one dimension it was useful—in fact almost essential—to consider the two-body potential not as it was given, but rather as extended in a periodic way. This extension of the potential to the periodic counterpart v_p did not change its definition in the basic domain but simply added replicas to it outside that domain. The advantage of the periodic potential was that with it one could use Fourier series in solving the Schrödinger equation; there is the same advantage in three dimensions.

For the wavefunctions, we want each particle coordinate to satisfy periodic boundary conditions over a length L_2 in each of the three mutually perpendicular coordinate directions. Thus, we want to

introduce a periodic potential $v_p(\mathbf{r})$ which has the same periodicity properties, which are essentially the properties of a simple cubic lattice. We do this in the following way: We consider a simple cubic lattice with lattice basis vectors $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ that are just unit vectors along the coordinate axes

$$\mathbf{b}_1 = \frac{\mathbf{i}}{L_2}, \quad \mathbf{b}_2 = \frac{\mathbf{j}}{L_2}, \quad \mathbf{b}_3 = \frac{\mathbf{k}}{L_2}.$$

A general reciprocal lattice vector K_j is then defined by

$$\mathbf{K}_j = 2\pi(j_1\mathbf{b}_1 + j_2\mathbf{b}_2 + j_3\mathbf{b}_3),$$

where j_1, j_2, j_3 are integers. If the potential $v(r)$ has Fourier transform $w(\mathbf{k})$,

$$v(\mathbf{r}) = \int w(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\,d\mathbf{k},$$

then the periodic potential $v_p(\mathbf{r})$ that we introduce in its place is

$$v_p(\mathbf{r}) = \sum_{\mathbf{K}_j} w(\mathbf{K}_j)\exp(i\mathbf{K}_j\cdot\mathbf{r}).$$

With this, we consider the construction of the wavefunction of the system. Since this is periodic in $\mathbf{r}_1$ and $\mathbf{r}_2$, we can consider it as built up by summing over basis functions of the form

$$\exp[i(\mathbf{K}_j\cdot\mathbf{r}_1 + \mathbf{K}_s\cdot\mathbf{r}_2)]$$

with appropriate coefficients. Then, if in this basis function we introduce the relative and center-of-mass coordinates $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ and $\mathbf{R} = \mathbf{r}_1 + \mathbf{r}_2$ and proceed much as in the one-dimensional case, it is straightforward to show that a candidate for the wavefunction $\psi(\mathbf{r}, \mathbf{R})$, in that it is periodic in $\mathbf{r}_1$ and $\mathbf{r}_2$, is

$$\psi(\mathbf{r}, \mathbf{R}) = \sum_{\mathbf{K}_m}\sum_{\mathbf{K}_n} C(\mathbf{K}_m, \mathbf{K}_n)\exp i\left(\frac{\mathbf{K}_m\cdot\mathbf{r}}{2} + \frac{\mathbf{K}_n\cdot\mathbf{R}}{2}\right) \times \Delta(m_1, n_1)\Delta(m_2, n_2)\Delta(m_3, n_3), \quad (28)$$

where $C(\mathbf{K}_m, \mathbf{K}_n)$ is some arbitrary function. If we put (28) into the Schrödinger equation with $\Delta(\mathbf{m}, \mathbf{n}) = \Delta(m_1, n_1)\Delta(m_2, n_2)\Delta(m_3, n_3)$, we get

$$[E - \tfrac{1}{2}(\mathbf{K}_m^2 + \mathbf{K}_n^2)C(\mathbf{K}_m, \mathbf{K}_n)]\Delta(\mathbf{m}, \mathbf{n}) = \sum_{\mathbf{K}_m} w\left(\frac{\mathbf{K}_m - \mathbf{K}'_m}{2}\right)C(\mathbf{K}'_m, \mathbf{K}_n)\Delta(\mathbf{m}', \mathbf{n}). \quad (29)$$

As in one dimension, we shall assume that the ground-state solution corresponds to zero momentum of the center of mass. In this case, $C(\mathbf{K}_m, \mathbf{K}_n)$ takes the form of an arbitrary function of $\mathbf{K}_m$ times a delta function in $\mathbf{K}_n$:

$$C(\mathbf{K}_m, \mathbf{K}_n) = D(\mathbf{K}_m)\delta(\mathbf{K}_n). \quad (30)$$

If now we put (30) into (29) and transform back to position space, we conclude that to solve (29) with center of mass at rest is equivalent to solving the Schrödinger equation

$$\left(-\frac{\hbar^2}{m}\nabla^2 + v_p(\mathbf{r})\right)\psi_0(\mathbf{r}) = E\psi_0(\mathbf{r}), \quad (31)$$

where $\psi_0(\mathbf{r})$ is the Fourier transform of $D(\mathbf{K}_m)$.

Although we have sketched the formal derivation of (31), it is clear on other grounds that (31) is the equation we want. For if we start with Eq. (27) for $\psi(\mathbf{r}_1, \mathbf{r}_2)$ and introduce the coordinates $\mathbf{r}$ and $\mathbf{R}$ and the periodic potential $v_p(\mathbf{r})$, the wavefunction which nominally becomes a function of both $\mathbf{r}$ and $\mathbf{R}$ is in fact, for the center-of-mass at rest, a function of $\mathbf{r}$ only. We have then called this function $\psi_0(\mathbf{r})$ in (31). Thus, Eq. (31) is just Eq. (27) transformed and relabeled. The point that needs discussion now is to show how, using Eq. (31), we can satisfy the periodicity conditions on $\mathbf{r}_1$ and $\mathbf{r}_2$. To do this we begin by observing that Eq. (31) is like the equation of a single particle of reduced mass $m/2$ moving in a periodic potential. We know from the theory of this equation in solid-state physics that it has solutions of Bloch form, i.e., in which a phase factor $e^{i\boldsymbol{\lambda}\cdot\mathbf{r}}$ multiplies a function which is periodic in $\mathbf{r}$ with the periodicity of the lattice. The ground-state solution usually corresponds to $\boldsymbol{\lambda} = 0$, and we shall assume this here. This being the case, the ground-state solution of Eq. (31) is a periodic function of $\mathbf{r}$. That is, if $\boldsymbol{\rho}_s$ is a lattice vector, then

$$\psi_0(\mathbf{r} + \boldsymbol{\rho}_s) = \psi_0(\mathbf{r}).$$

But, since $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, this means

$$\psi_0((\mathbf{r}_1 + \boldsymbol{\rho}_s) - r_2) = \psi_0(\mathbf{r}_1 - \mathbf{r}_2),$$

which shows that ψ_0 is separately periodic in $\mathbf{r}_1$ (and, of course, equally in $\mathbf{r}_2$) and therefore satisfies the periodicity conditions we have imposed.

Let us then consider the solution of Eq. (31); it will suffice to use the Wigner–Seitz method. According to this method, we introduce a sphere of radius l which is centered at a given potential and whose volume is equal to that of the unit cell. In this case

$$4\pi l^3/3 = L_2^2 = 2u, \quad u = L_N^3/N. \quad (32)$$

Then we approximately satisfy the periodicity condition by requiring that the normal derivative of the wavefunction vanish on the surface of the sphere ($r = l$). We shall take the interparticle potential as corresponding to a hard core in the *relative variable* r of *radius* a, which is equivalent to the boundary condition that the wavefunction vanish at $r = a$. In terms of the two particles with which we began, this implies that they are hard spheres of *diameter* a.

To come back then to (31), we observe that a general form of the wavefunction in the region $a < r < l$ is

$$Aj_0(kr) + Bn_0(kr), \tag{33}$$

where j_0 and n_0 are spherical Bessel and Neumann functions and $k^2 = mE/\hbar^2$. On applying the two conditions stated above to (33), we find that the (unnormalized) wavefunction is

$$j_0(kr) + \tan(ka)n_0(kr), \tag{34}$$

and that k is determined from the equation

$$kl = \tan k(l - a). \tag{35}$$

We are interested in the case $a \ll l$, for which we get, on expanding (35), the result $k^2 \approx 3a/l^3$. With the expression in (32) for l^3, this yields $k^2 = 2\pi a/v$ or

$$E_g = 2\pi a\hbar^2/mv.$$

We must remember, however, that this result refers to the energy of the two-body system. The ground-state energy *per particle* of this system is then:

$$\text{Energy per particle} = \hbar^2\pi a/mv \quad \text{(two-body system)}.$$

This is to be compared with the analogous result for the N-body ($N \gg 1$) system:

$$\text{Energy per particle} = 2\hbar^2\pi a/mv \quad (N\text{-body system}).$$

In a way, this is a very gratifying result—for this is just the answer we got for one-dimensional δ functions in the perturbation theory limit, i.e., the energy per particle for the two-body system is just *half* that for the N-body ($N \gg 1$) system. This agreement between one dimension and three then encourages us in the hope that there may be similar and perhaps even better agreement away from the perturbation limit, and that in three dimensions, as in one, the two-body problem is a fair approximation to the N-body problem.

To look further into this, we can make another comparison between the two-body and N-body results: we can compare the respective correlation functions. By "correlation function," we mean the expression for the relative probability $D(r)$ that two particles are within a distance r of each other. For the two-body case this function is just the square of the wavefunction; i.e., from (34), on adding a subscript to D to label the two-body function, it is

$$D_2(r) = C[(j_0(kr) + \tan(ka)n_0(kr)]^2.$$

Here C is a normalization constant at our disposal. The corresponding function for the N-body case as given by Lee, Huang, and Yang[5] and corrected[11] by Bocchieri, Orzalesi, and Smith[12] is

$$D_N(r) = [1 + G(r)]^2 + [1 + F(r)]^2 - 1 - 2f[F(r) + G(r)], \tag{36}$$

where the functions $G(r)$, $F(r)$, and f are defined in the two papers just cited.

The correlation functions depend on a parameter which is essentially the ratio of volume of the hard core to the volume available per particle. We can characterize this by a quantity

$$\lambda = 8\pi a^3/v.$$

We have then calculated and compared the functions $D_2(r)$ and $D_N(r)$ for various values of λ. In calculating $D_N(r)$, we have evaluated by digital computer the integrals that enter the functions $F(r)$ and $G(r)$. We shall present results for the typical case $\lambda = 10^{-3}$.

TABLE III. Comparison of the correlation functions $D_2(r)$ and $D_N(r)$ for $8\pi a^3/v = 10^{-3}$. $D_2(r)$ is normalized arbitrarily to equal $D_N(r)$ at $r/a = 3.0$.

r/a	$D_2(r)$	$D_N(r)$
1.00	0.00	0.025
1.50	0.108	0.124
2.00	0.252	0.259
3.0	0.449	0.449
5.0	0.646	0.641
10.0	0.813	0.809
20.0	0.884	0.902
30.0	0.891	0.935
50.0		0.962
100.0		0.980
200.0		0.990

The agreement we have found is quite striking. In fact, it is so close that, were we to plot the two functions, they would be hard to distinguish over much of their range, for any reasonable graph size. We have, therefore, made a tabular comparison; this is set out in Table III, for which we add some words of explanation. The normalization of the N-body wavefunction is arbitrarily taken to be that $D_N(r) \to 1$ as $r \to \infty$. Now the two-body function is defined only within the unit cell, i.e., for $r < l$, so we cannot normalize it the same way; instead, we have chosen its normalization so that it coincides with $D_N(r)$ at the arbitrary point $r = 3a$. Again, for the reason that $D_2(r)$ is only defined for $r < l$, the function $D_N(r)$ is plotted over a

[11] The correction cited is that of changing the factor $4f$ in Ref. 5 to the factor $2f$ that appears in Eq. (36).

[12] P. Bocchieri, C. A. Orzalesi, and V. H. Smith, Jr., Nuovo Cimento **52**, 18 (1967).

more extended range than is $D_2(r)$.[13] Finally, we note that the function $D_2(r)$ is zero at $r = a$, which it must be to satisfy the boundary conditions exactly. The function $D_N(r)$, on the other hand, satisfies the boundary conditions only approximately at $r = a$—witness the fact that it is small but not strictly zero there.

As we have remarked, the agreement between the two functions is quite good, perhaps to such a degree that it is partly fortuitous. In this connection it should be borne in mind that Table III does not really compare D_2 with D_N, but compares an approximation to D_2 (Wigner–Seitz method) with an approximation to D_N (use of pseudopotentials). Also, the close agreement between the functions throughout the range over which they are mutually defined is not in necessary contradiction with the fact that the energy per particle differs by a factor of two for the two different cases; it may be the longer-range correlation in the N-body case that accounts for this.

[13] For the value of λ we have chosen l/a to be about 30, which is why the last entry in the table for $D_2(r)$ is for $r/a = 30$.

ACKNOWLEDGMENTS

I would like to thank Paul Tsipouras for programming the solution to the four-body equations, and Joseph Russell for doing likewise for the integrals involved in evaluating the function $D_N(r)$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Some Analytic Properties of Scattering Amplitudes for Long-Range Forces*

WILLIAM R. ROSS† AND JOHN R. TAYLOR
Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado

(Received 10 September 1968)

We examine the properties of the partial-wave amplitude $a(l, k)$ and the full amplitude $A(k, \cos\theta)$ for scattering by a long-range potential made up of a Coulomb part $2\alpha/r$ and a short-range part $V(r)$. The properties of $a(l, k)$ as an analytic function of l and k are shown to be quite similar to those of the usual short-range amplitude, except in the neighborhood of the threshold $k = 0$, which point we examine in detail. The full amplitude is treated as a function of $\cos\theta$ for fixed physical momentum k; using the Sommerfeld–Watson transformation, we show that $A(k, \cos\theta)$ is analytic in the cut plane of $\cos\theta$.

1. INTRODUCTION

Because of the photon's zero mass and the infinite range of the associated forces, it is still not clear whether electromagnetic forces can be incorporated into analytic S-matrix theory. The long range of the forces invalidates all usual definitions of the S matrix; and even if an S matrix can be defined, it will probably not satisfy some of the usual requirements in S-matrix theory and, in particular, the infrared problems associated with soft photons will presumably remain.[1]

It is in the hope of clarifying a few of these questions that we examine here some properties of the nonrelativistic scattering by a long-range potential

$$U(r) = 2\alpha/r + V(r)$$

consisting of a Coulomb interaction $2\alpha/r$ and a short-range potential $V(r)$. This problem has already received considerable attention. Dollard[2] has shown that, although the usual definitions fail when Coulomb forces are present, it is possible to define a satisfactory S matrix which has the normal interpretation and leads to the usual amplitude used in practical calculations. Cornille and Martin[3] and Mentovsky[4] have examined the partial-wave amplitude as an analytic function of momentum k for fixed physical angular momentum l. Klarsfeld[5] has considered the same amplitude as a function of l for fixed physical k.

In this paper, we consider both the partial-wave and

* Research supported in part by United States Air Force, Office of Scientific Research, Office of Aerospace Research, under Contract No. A.F.-A.F.O.S.R.-30-67.

† Present address: Physics Department, University of New Brunswick, Fredericton, New Brunswick, Canada.

[1] These questions have been discussed by several authors and widely differing conclusions have been expressed. G. F. Chew [Sci. Progr. **51**, 529 (1963)] argues that electromagnetic interactions can certainly not find a place in S-matrix theory. Papers expressing the opposite view include: A. O. Barut, Acta Phys. Austriaca Suppl. **2**, 162 (1966); A. O. Barut and R. A. Blade, Nuovo Cimento **39**, 331 (1965); T. T. Chow and M. Dresden, Rev. Mod. Phys. **39**, 143 (1967). The spirit of these papers is to assume that electromagnetic interaction *can* be included in S-matrix theory and that the usual properties—existence of S, cluster decomposition, analyticity—continue to hold.

[2] J. D. Dollard, J. Math. Phys. **5**, 729 (1964).
[3] H. Cornille and A. Martin, Nuovo Cimento **26**, 298 (1962).
[4] Yu. L. Mentovsky, Nucl. Phys. **65**, 673 (1965).
[5] S. Klarsfeld, Nuovo Cimento **48A**, 1059 (1967).

more extended range than is $D_2(r)$.[13] Finally, we note that the function $D_2(r)$ is zero at $r = a$, which it must be to satisfy the boundary conditions exactly. The function $D_N(r)$, on the other hand, satisfies the boundary conditions only approximately at $r = a$—witness the fact that it is small but not strictly zero there.

As we have remarked, the agreement between the two functions is quite good, perhaps to such a degree that it is partly fortuitous. In this connection it should be borne in mind that Table III does not really compare D_2 with D_N, but compares an approximation to D_2 (Wigner–Seitz method) with an approximation to D_N (use of pseudopotentials). Also, the close agreement between the functions throughout the range over which they are mutually defined is not in necessary contradiction with the fact that the energy per particle differs by a factor of two for the two different cases; it may be the longer-range correlation in the N-body case that accounts for this.

[13] For the value of λ we have chosen l/a to be about 30, which is why the last entry in the table for $D_2(r)$ is for $r/a = 30$.

ACKNOWLEDGMENTS

I would like to thank Paul Tsipouras for programming the solution to the four-body equations, and Joseph Russell for doing likewise for the integrals involved in evaluating the function $D_N(r)$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Some Analytic Properties of Scattering Amplitudes for Long-Range Forces*

WILLIAM R. ROSS† AND JOHN R. TAYLOR
Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado

(Received 10 September 1968)

We examine the properties of the partial-wave amplitude $a(l, k)$ and the full amplitude $A(k, \cos\theta)$ for scattering by a long-range potential made up of a Coulomb part $2\alpha/r$ and a short-range part $V(r)$. The properties of $a(l, k)$ as an analytic function of l and k are shown to be quite similar to those of the usual short-range amplitude, except in the neighborhood of the threshold $k = 0$, which point we examine in detail. The full amplitude is treated as a function of $\cos\theta$ for fixed physical momentum k; using the Sommerfeld–Watson transformation, we show that $A(k, \cos\theta)$ is analytic in the cut plane of $\cos\theta$.

1. INTRODUCTION

Because of the photon's zero mass and the infinite range of the associated forces, it is still not clear whether electromagnetic forces can be incorporated into analytic S-matrix theory. The long range of the forces invalidates all usual definitions of the S matrix; and even if an S matrix can be defined, it will probably not satisfy some of the usual requirements in S-matrix theory and, in particular, the infrared problems associated with soft photons will presumably remain.[1]

It is in the hope of clarifying a few of these questions that we examine here some properties of the nonrelativistic scattering by a long-range potential

$$U(r) = 2\alpha/r + V(r)$$

consisting of a Coulomb interaction $2\alpha/r$ and a short-range potential $V(r)$. This problem has already received considerable attention. Dollard[2] has shown that, although the usual definitions fail when Coulomb forces are present, it is possible to define a satisfactory S matrix which has the normal interpretation and leads to the usual amplitude used in practical calculations. Cornille and Martin[3] and Mentovsky[4] have examined the partial-wave amplitude as an analytic function of momentum k for fixed physical angular momentum l. Klarsfeld[5] has considered the same amplitude as a function of l for fixed physical k.

In this paper, we consider both the partial-wave and

* Research supported in part by United States Air Force, Office of Scientific Research, Office of Aerospace Research, under Contract No. A.F.-A.F.O.S.R.-30-67.

† Present address: Physics Department, University of New Brunswick, Fredericton, New Brunswick, Canada.

[1] These questions have been discussed by several authors and widely differing conclusions have been expressed. G. F. Chew [Sci. Progr. **51**, 529 (1963)] argues that electromagnetic interactions can certainly not find a place in S-matrix theory. Papers expressing the opposite view include: A. O. Barut, Acta Phys. Austriaca Suppl. **2**, 162 (1966); A. O. Barut and R. A. Blade, Nuovo Cimento **39**, 331 (1965); T. T. Chow and M. Dresden, Rev. Mod. Phys. **39**, 143 (1967). The spirit of these papers is to assume that electromagnetic interaction *can* be included in S-matrix theory and that the usual properties—existence of S, cluster decomposition, analyticity—continue to hold.

[2] J. D. Dollard, J. Math. Phys. **5**, 729 (1964).
[3] H. Cornille and A. Martin, Nuovo Cimento **26**, 298 (1962).
[4] Yu. L. Mentovsky, Nucl. Phys. **65**, 673 (1965).
[5] S. Klarsfeld, Nuovo Cimento **48A**, 1059 (1967).

the full amplitude. We first establish the properties of the partial-wave amplitude $a(l, k)$ as an analytic function of both l and k and then use these results to find the properties of the full amplitude $A(k, \cos\theta)$ as an analytic function of $\cos\theta$ for fixed physical k.

In Sec. 2, we examine $a(l, k)$ using a method developed by Froissart[6] for the non-Coulomb case. This method allows us to treat a very wide class of potentials, our only requirement being that the short-range part $V(r)$ satisfy

$$|r^{2-\eta}e^{\epsilon r}V(r)| < \text{const} \tag{1.1}$$

for some η and ϵ both greater than zero. We illustrate our results with the example of a Coulomb plus a pure Yukawa potential, for which the analytic properties are very similar to those of the non-Coulomb case; specifically, $a(l, k)$ is meremorphic in (l, k) except for branch points at

$$k = \pm in\mu/2$$

where $n = 0, 1, 2, \cdots$ and μ is the inverse range of the Yukawa potential. The most important difference between the Coulomb and non-Coulomb cases is the singularity at $k = 0$, which in the Coulomb case is an accumulation point of bound-state poles and is also the starting point of the left-hand cut. We examine this singularity in detail and show that, in spite of these difficulties, it is still possible to express the amplitude in the familiar N/D form.

In Sec. 3, we consider the full amplitude $A(k, \cos\theta)$ for the case where the short-range part of the potential is a Yukawa. Following the method of Regge[7] we use the Sommerfeld–Watson transformation to establish the analytic properties of $A(k, \cos\theta)$ from those of the partial-wave amplitude $a(l, k)$. We find that this transformation cannot be applied to the complete amplitude A but that, at least for physical k, it can be applied to the difference between A and the pure Coulomb amplitude A_c:

$$A'(k, \cos\theta) = A(k, \cos\theta) - A_c(k, \cos\theta).$$

This establishes that, for physical k, $A'(k, \cos\theta)$ is analytic in the whole plane of $\cos\theta$ except on the branch cut starting at

$$\cos\theta = 1 + \mu^2/2k^2.$$

The crucial part of the proof is the verification that, as $|l| \to \infty$ in any direction in the half-plane $\text{Re}\, l \geq -\frac{1}{2}$, the partial-wave amplitude $a(l, k)$ is sufficiently bounded to allow use of the Sommerfeld–Watson transformation. This we establish by examination of a Lippmann–Schwinger equation for $a(l, k)$. Our method breaks down when k becomes complex, and for this reason we can establish the properties of $A(k, \cos\theta)$ for real k only.

We conclude, therefore, that the analytic properties of the amplitude for scattering by long-range forces are a natural generalization of the corresponding properties of short-range amplitudes. This suggests that, at least as far as analyticity is concerned, there should be no obstacle to incorporating long-range forces into S-matrix theory.

Most of our proofs are rather long and complicated and we therefore omit some details, for which we refer the reader to the thesis of the first named author.[8]

2. THE PARTIAL-WAVE AMPLITUDE

A. Outline and Results

The partial-wave amplitude is defined in terms of solutions of the radial Schrödinger equation

$$\psi'' + \left[k^2 - \frac{\lambda^2 - \frac{1}{4}}{r^2} - \frac{2\alpha}{r} - V(r)\right]\psi = 0. \tag{2.1}$$

(We use units for which $\hbar = 2m = 1$ and introduce the angular-momentum variable $\lambda \equiv l + \frac{1}{2}$.) As usual, we define certain standard solutions of this equation: the "regular" solution

$$\phi(\lambda, k, r) \underset{r\to 0}{\sim} r^{l+1}, \tag{2.2}$$

and the "incoming" and "outgoing" solutions

$$\chi^{\pm}(\lambda, k, r) \underset{r\to\infty}{\sim} \exp \mp i[kr - (\alpha/k)\ln(2kr)]. \tag{2.3}$$

The Jost functions are defined as the coefficients in the expansion of ϕ in terms of[9] $\chi^{\pm}$:

$$\phi(r) = [f^{+}(\lambda, k)\chi^{-}(r) - f^{-}(\lambda, k)\chi^{+}(r)]/2ik \tag{2.4}$$

and the partial-wave S matrix is defined as the ratio of these coefficients,

$$S(\lambda, k) = f^{+}(\lambda, k)e^{i\pi l}/f^{-}(\lambda, k). \tag{2.5}$$

Finally, the partial-wave amplitude is defined as

$$a(\lambda, k) = [S(\lambda, k) - 1]/2ik. \tag{2.6}$$

Our starting point is the Schrödinger equation for a pure Coulomb potential [i.e., Eq. (2.1) with $V \equiv 0$] with corresponding solutions

$$\phi_c(\lambda, k, r) = (2ik)^{-l-1}M_{\nu,\lambda}(2ikr) \tag{2.7}$$

and

$$\chi_c^{\pm}(\lambda, k, r) = e^{-i\pi\nu/2}W_{\pm\nu,\lambda}(\pm 2ikr), \tag{2.8}$$

[6] M. Froissart, J. Math. Phys. **3**, 922 (1962).
[7] T. Regge, Nuovo Cimento **14**, 951 (1959).
[8] W. R. Ross, Ph.D. thesis, University of Colorado, 1968.
[9] Whenever possible without danger of confusion we drop the arguments λ and/or k from the functions $\phi(\lambda, k, r)$ etc.

where $M_{\nu,\lambda}(z)$ and $W_{\nu,\lambda}(z)$ are Whittaker functions[10] and

$$\nu \equiv i\alpha/k. \tag{2.9}$$

In terms of these pure Coulomb solutions, we obtain iterative expansions for the actual solutions ϕ and $\chi^{\pm}$ in powers of the short-range potential $V(r)$. These lead to expressions for the Jost functions $f^{\pm}$ as power series in $V(r)$ and it is from these power series that we deduce the analytic properties of $f^{\pm}$ and, hence, those of S or a.

Our method is a direct generalization of a technique developed by Froissart[6] for pure short-range potentials. The series expansions for $f^{\pm}$ are written in the form

$$f^{\pm} = f_c^{\pm} + \int_0^{\infty} dr\, \chi_c^{\pm}(r)V(r)\phi_c(r) + \int_0^{\infty} dr' \int_0^{r'} dr\, \chi_c^{\pm}(r')V(r')I(r, r')V(r)\phi_c(r), \tag{2.10}$$

where $I(r, r')$ is, of course, a series in powers of V. Assuming that $r^{2-\eta}e^{\epsilon r}V(r)$ is bounded, we show that this expression is analytic in a certain domain of (λ, k) but that the integrals diverge when (λ, k) moves outside of this region. Those parts of the integrals which diverge can be explicitly separated and expressed in terms of the Mellin and Laplace transforms of the potential $V(r)$ and the kernel $I(r, r')$. We examine the analytic properties of the transforms of $I(r, r')$ and find their singularities in terms of the singularities of the corresponding transforms of the potential $V(r)$. In this way we establish the analytic properties of $f^{\pm}(\lambda, k)$ for all (λ, k) and express all singularities in terms of the singularities of the Mellin and Laplace transforms of $V(r)$.

Our conclusion is that the Jost function $f^{+}(\lambda, k)$ is analytic for all λ and k except when λ or k is contained in one of the following sets of points[11]:

$$\lambda \in \left\{ -\frac{n}{2} + \frac{m+1}{2}(\mathcal{S} - 2) \right\}$$

or

$$k \in \left\{ 0, \frac{m\mathfrak{I}_0 + \mathfrak{I}_1}{2i}, \frac{m\mathfrak{I}_0 + 2\mathfrak{I}_1}{2i}, \frac{\mathfrak{I}_2}{2i} \right\}, \tag{2.11}$$

where m and $n = 0, 1, 2, \cdots$, and $\mathcal{S}$ and $\mathfrak{I}_n$ denote the sets of singularities of the transforms

$$u(\sigma) = \int_0^1 dr\, r^{\sigma-1}V(r) \tag{2.12}$$

and

$$\tilde{u}_n(q) = \int_1^{\infty} dr\, e^{-qr}r^{n\nu}V(r) \tag{2.13}$$

with $\nu = i\alpha/q$. The set of points $m\mathfrak{I}_0$, for example, is the set of all

$$q = q_1 + \cdots + q_m, \quad q_i \in \mathfrak{I}_0.$$

The singularities of $f^{-}(\lambda, k)$ are obtained by the substitution $k \to -k$ and the domain of meremorphy of S follows immediately from the definition $S = f^{+}e^{i\pi l}/f^{-}$.

For any given potential, this result immediately locates the singularities of $f^{\pm}$ and S. For the example of a pure Yukawa $V(r) = \gamma e^{-\mu r}/r$, the singularities of f^{+} are poles at $\lambda = -(n+1)/2$ and branch points at $k = in\mu/2$. This implies meremorphy of S except at $k = \pm in\mu/2$, as already mentioned in the introduction.

The nature of the singularity at $k = 0$ is examined in Sec. 2C and in Sec. 2D we show that the partial-wave amplitude can be written in the usual "N over D" form.

After this brief outline of our method and results we proceed to the main proof. As far as possible we refer the reader to the paper of Froisart[6]; for details of certain estimates we appeal to Ref. 8.

B. The Main Proof

We first note that one can easily show, using the definitions (2.2) to (2.4) that

$$f^{-}(\lambda, k) = e^{\pi\alpha/k}f^{+}(\lambda, ke^{-i\pi}) \tag{2.14}$$

and that it is therefore sufficient to establish analyticity of f^{+}, for which Eq. (2.4) implies the following expression:

$$f^{+} = W[\chi^{+}, \phi] = \lim_{r\to\infty} W[\chi_c^{+}, \phi], \tag{2.15}$$

where $W[u, v]$ denotes the Wronskian of the functions $u(r)$ and $v(r)$.

To make use of this expression for f^{+} we replace the differential equation (2.1) and the boundary conditions (2.2) for $\phi(r)$ by the integral equation

$$\phi(r) = \phi_c(r) + \int_0^r dr'\, G(r, r')V(r')\phi(r'), \tag{2.16}$$

where G is the appropriate Green's function

$$G(r, r') = [u(r')v(r) - u(r)v(r')]/W[u, v],$$

where u and v are any two independent pure Coulomb solutions (e.g., $\chi_c^{\pm}$). Substituting this integral equation into Eq. (2.15) we get

$$f^{+} = f_c^{+} + \int_0^{\infty} dr\, \chi_c^{+}(r)V(r)\phi(r), \tag{2.17}$$

[10] See E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge, England, 1962), 4th ed., Chap. 16.

[11] There are in addition fixed poles when λ is a negative half-integer but these always cancel out of $S = f^{+}e^{i\pi l}/f^{-}$. See Eq. (2.26) below.

where f_c^+ denotes the pure-Coulomb Jost function

$$f_c^+ \equiv W[\chi_c^+, \phi_c] = (2ik)^{-l}e^{\pi\alpha/k}\Gamma(2l+2)/\Gamma(l-\nu+1). \quad (2.18)$$

Finally we iterate Eq. (2.16) for ϕ and substituting the resulting series into Eq. (2.17) obtain the expression (2.10) for f^+ with

$$I(r, r') = \sum_{n=2}^{\infty} I_n(r, r'), \quad (2.19)$$

$$I_n(r, r') = \int_{r=r_1<\cdots<r_n=r'} \prod_{m=1}^{n-1} G(r_{m+1}, r_m) \prod_{m=2}^{n-1} V(r_m)\, dr_m, \quad (2.20)$$

and

$$I_2(r, r') = G(r', r).$$

By a method similar to that used by Froissart we can show that the kernel $I(r, r')$ is an entire function of λ and k and satisfies the following bound[12]:

$$|I(r, r')| < N_{\lambda,k}[F(r')/F(r)](r' - r), \quad (2.21)$$

where

$$F(r) = \begin{cases} r^{|\mathrm{Re}\,\lambda|-\frac{1}{2}}, & r \le 1, \quad |\mathrm{Re}\,\lambda| \ge \frac{1}{2}, \\ 1, & r \le 1, \quad |\mathrm{Re}\,\lambda| \le \frac{1}{2}, \\ e^{|\mathrm{Im}\,k|r}, & r > 1. \end{cases} \quad (2.22)$$

If we now assume that the short-range part of the potential satisfies condition (1.1), this bound for $I(r, r')$ implies that the two integrals in Eq. (2.10) for $f^+(\lambda, k)$ converge and define an analytic function for

$$\mathrm{Re}\,\lambda > \tfrac{1}{2} - \eta, \quad \mathrm{Im}\,k < \epsilon/2 \quad (2.23)$$

(except of course at $k = 0$ where the Coulomb functions have a branch point).

If λ moves outside of this region the integrals (2.10) diverge because of bad power behavior near $r = 0$; if k moves out the integrals diverge because of bad exponential behavior as $r \to \infty$. In order to continue outside of the region (2.23), we must, therefore, isolate those parts of the integrals which diverge and to this end we use the power series for ϕ_c and χ_c^+ to write

$$\phi_c(r) = \sum_{n=0}^{N} a_n r^{n+l+1} + \text{remainder} \quad (2.24)$$

with a similar expression for χ_c^+, the remainders being of order $r^{N+\lambda+\frac{1}{2}}$ and $r^{N-|\lambda|+\frac{1}{2}}$, respectively, for small r. Similarly for large r, we use the integral expansion[13]

$$\chi_c^+(r) = (2kr)^{\nu}\left[e^{-ikr} + \int_{ik}^{A} e^{-\nu r} F'(p/ik)\, dp/ik\right] + \text{remainder}, \quad (2.25)$$

where $F'(z)$ denotes the derivative of the hypergeometric function $F[-l-\nu, l-\nu+1, 1, (1-z)/z]$ and the remainder is of order $r^{\nu}e^{-Ar}$; with a similar expression for ϕ_c.

Much as in the paper of Froissart, we can now continue the two integrals in Eq. (2.10) for f^+ to all (λ, k). Before doing so we note that it is convenient to consider not f^+, but

$$f'^+ = f^+/\Gamma(2\lambda+1), \quad (2.26)$$

since this avoids the fixed pole of ϕ_c when $(2\lambda+1)$ is a negative integer. (And, of course, we can calculate S in terms of $f'^{\pm}$ just as well as $f^{\pm}$.)

To continue the single integral in Eq. (2.10) we split it into two parts, $\int_0^1$ and $\int_1^{\infty}$, and substitute the separation (2.24) for ϕ_c and χ_c^+ into the first and that of (2.25) into the second. By choosing N and A sufficiently large we can guarantee that, for any given (λ, k), the integrals involving the remainders are convergent and analytic. Thus the integral $\int_0^1$ has singularities, if any, coming from a finite sum of terms of the form

$$\int_0^1 r^{n+2\lambda+1} V(r)\, dr = u(n+2\lambda+2)$$

which is precisely the Mellin transform defined in Eq. (2.12). By definition, these are analytic except when

$$\lambda \in \{-n/2 + \tfrac{1}{2}(S-2)\}. \quad (2.27)$$

Similarly, the integral $\int_1^{\infty}$ is analytic except when[14]

$$k \in \{0, \mathfrak{T}_2/2i\}. \quad (2.28)$$

To isolate the divergent parts of the double integral in Eq. (2.10), we divide the range of integration into three parts: $0 < r < r' < 1$, $0 < r < 1 < r' < \infty$, and $1 < r < r' < \infty$. In the first case, for example, substitution of Eq. (2.24) for ϕ_c and χ_c^+ leads to potentially divergent integrals of the form

$$\int_0^1 dr' \int_0^{r'} dr\, r'^{n'+\lambda+\frac{1}{2}} r^{n+\lambda+\frac{1}{2}} V(r')V(r)I(r, r') = -(2\pi)^{-2}\int_{-i\infty}^{i\infty}\int A(n+\lambda-\sigma+\tfrac{3}{2}, n'+\lambda-\sigma'+\tfrac{3}{2}) \times u(\sigma)u(\sigma')\, d\sigma\, d\sigma', \quad (2.29)$$

[12] This corresponds to Eq. (8) of Froissart. For the proof see Ref. 8.

[13] For the proof of this integral relation, which corresponds to Eq. (17) of Froissart, see Ref. 8.

[14] The singularity at $k = 0$ comes from χ_c^+ which is singular at $k = 0$.

where $A(s, s')$ is the double Mellin transform of $I(r, r')$ and we have used the inverse of the Mellin transform (2.12) in writing the second line.

The integral (2.29) can be analytically continued except when the singularities of $u(\sigma)$ or $u(\sigma')$ pinch against those of A. These latter singularities we discuss in a moment. Clearly, once they are located we can find those of the integral (2.29) and hence all singularities of f'^{+} arising from small r and r' in the double integral of Eq. (2.10).

The remaining two parts of the double integral of Eq. (2.10)—r small, r' large and r and r' both large—can be treated similarly and their singularities located in terms of those of the Mellin–Laplace and double Laplace transforms of $I(r, r')$.[15]

It remains to establish the analytic properties of the three transforms of $I(r, r')$ which we can do much as in the paper of Froissart.[15] Thus the bound (2.21) guarantees analyticity of $A(s, s')$ when

$$\operatorname{Re} s > \max(|\operatorname{Re}\lambda| - \tfrac{1}{2}, 0), \quad \operatorname{Re} s' > 1 - \operatorname{Re} s.$$

One can continue beyond this region using the differential equation for $I(r, r')$ whose Mellin transform is[16]

$$\begin{aligned}&[(s-1)(s-2) - l(l+1)]A(s-2, s') \\ &\quad - 2\alpha A(s-1, s') + k^2 A(s, s') \\ &\quad - (2\pi i)^{-1}\int_{-i\infty}^{i\infty} A(s-\sigma, s')u(\sigma)\,d\sigma = 1/(s+s'-1)\end{aligned}$$

and a similar equation in s'. These equations allow a strip by strip continuation of $A(s, s')$ which turns out to be analytic except when[17]

$$s \in \{-1 - s' - n + m(\mathcal{S} - 2)\}$$

or

$$s \in \{\pm\lambda - \tfrac{1}{2} - n + m(\mathcal{S} - 2)\}.$$

Returning to Eq. (2.29) we find that the double integral for f'^{+} has singularities when

$$\lambda \in \{-\tfrac{1}{2}n + \tfrac{1}{2}(m+1)(\mathcal{S} - 2)\}. \tag{2.30}$$

A similar analysis for the other two parts of the double integral gives singularities when

$$k \in \left\{\frac{m\mathcal{T}_0 + \mathcal{T}_1}{2i}, \frac{m\mathcal{T}_0 + 2\mathcal{T}_1}{2i}\right\}, \tag{2.31}$$

but no new singularities in λ.

[15] See Ref. 6 or Ref. 8.

[16] This corresponds to Eq. (10a) of Froissart. Note that we have an extra term with coefficient α, coming from the Coulomb potential.

[17] Compare Eq. (12) of Froissart. Note that in our case the Coulomb potential causes some additional singularities.

Combining Eqs. (2.27), (2.28), (2.30), and (2.31) we arrive at the result quoted in Eq. (2.11).[18]

C. The Singularity at Threshold

The singularity at threshold for our long-range potential differs from that of the short-range case for two obvious reasons. First, since the Coulomb potential has an infinite number of bound states close to threshold, the point $k = 0$ is an accumulation point of singularities. Second, the left-hand cut, which for a Yukawa potential starts at $E = -\mu^2/4$ (or $k = i\mu/2$), in our case starts at $E = 0$. This makes the branch point at $E = 0$ considerably more complicated. In particular, as a function of k, the short-range partial-wave amplitude is analytic at $k = 0$ when l is physical, while for our case this is not so.

In order to examine the branching properties of the Jost function $f^{+}(\lambda, k)$ at $k = 0$ we note first that $\phi(\lambda, k, r)$ is clearly analytic at $k = 0$ and hence that the circuit relations for $f^{+} = W[\chi^{+}, \phi]$ are the same as those of χ^{+}. Since χ^{+} satisfies the integral equation

$$\chi^{+}(r) = \chi_c^{+}(r) - \int_r^{\infty} dr' G(r, r')V(r')\chi^{+}(r')$$

[analogous to Eq. (2.16)], where G is analytic at $k = 0$, it follows that the circuit relations for χ^{+} are in turn the same as those of the pure Coulomb solution χ_c^{+}. These latter can be derived from standard properties of Whittaker functions[10] and imply that

$$\begin{aligned}f^{+}(\lambda, ke^{-2\pi i}) &= f^{+}(\lambda, k)e^{-2\pi\alpha/k} \\ &\quad - f^{+}(\lambda, ke^{-\pi i})2\pi i e^{-2\pi\alpha/k}/\Gamma(-l-\nu)\Gamma(l-\nu+1).\end{aligned} \tag{2.32}$$

In the non-Coulomb limit, for which $\nu \equiv i\alpha/k = 0$, the second term on the right vanishes when l is physical and f^{+} has no branch point. Obviously, when $\alpha \neq 0$ this is not so.

The circuit relations for $S(\lambda, k)$ or $a(\lambda, k)$ follow simply from Eq. (2.32). It is easily seen that the branch point at $k = 0$ is still present for physical angular momenta.

D. Representation as N over D

In the case of short-range potentials, one can express the amplitude a/k^{2l} as the quotient of two functions N and D, where N has only the left-hand and D only the right-hand cut. In our case the corresponding expression is

$$\frac{a(\lambda, k)e^{\pi\alpha/2k}}{k^{2l}} = \frac{N(\lambda, k)}{D(\lambda, k)}. \tag{2.33}$$

[18] All singularities of f'^{+} are certainly included in Eq. (2.11). Some points of this set may not in fact be singular if some cancellation occurs.

The function N has a cut in E from 0 to $-\infty$ and is real for $E > 0$, λ real, while D has a cut from $E = 0$ to $+\infty$ and is real for $E < 0$, λ real.

It is clear from the definitions (2.5) and (2.6) that Eq. (2.33) can be satisfied by the choice

$$N(\lambda, k) = [f^+(\lambda, k)e^{i\pi l} - f^-(\lambda, k)]/2(ik)^{l+1} \quad (2.34)$$

and

$$D(\lambda, k) = (-ik)^l e^{-\pi\alpha/2k} f^-(\lambda, k). \quad (2.35)$$

That N and D defined in this way have the required properties follows from the following four relations, all of which can be checked by inspection of the asymptotic forms (2.2) and (2.3):

$$\begin{aligned} [\phi(\lambda^*, k^*, r)]^* &= \phi(\lambda, k, r), \\ \phi(\lambda, -k, r) &= \phi(\lambda, k, r), \\ [\chi^+(\lambda^*, k^*, r)]^* &= \chi^-(\lambda, k, r), \\ \chi^-(\lambda, ke^{i\pi}, r) &= e^{-\pi\alpha/k}\chi^+(\lambda, k, r). \end{aligned} \quad (2.36)$$

From these it follows easily that $D(\lambda, k)$ as defined in Eq. (2.35) is real when λ is real and k positive imaginary ($E < 0$). The desired analytic properties of D follow from the results of Sec. 2B. Similarly from Eq. (2.36) follows an identical equation for $f^{\pm}(\lambda, k)$ and from this it follows for real l and $k > 0$ that $N(\lambda, k)$ as defined in Eq. (2.34) is real. This completes the proof.

3. THE FULL AMPLITUDE

A. Outline and Results

As one would expect, the properties of the full amplitude $A(k, \cos\theta)$ can be derived from those of the partial-wave amplitude $a(\lambda, k)$ by means of the partial-wave series. For a purely short-range potential (such as the Yukawa) $a(\lambda, k)$ falls off exponentially as $\lambda \to +\infty$ and the partial-wave series converges very well. In the present case (short-range *plus* Coulomb) $a(\lambda, k)$ oscillates as $\lambda \to +\infty$ and the partial-wave series for A is of no use. However, the difference

$$a'(\lambda, k) = a(\lambda, k) - a_c(\lambda, k) \quad (3.1)$$

behaves much as the ordinary short-range amplitude and the series

$$\begin{aligned} A'(k, \cos\theta) &\equiv A(k, \cos\theta) - A_c(k, \cos\theta) \\ &\equiv \sum_{l=0}^{\infty} (2l + 1)a'(\lambda, k)P_l(\cos\theta) \end{aligned} \quad (3.2)$$

converges well. It is obviously sufficient to study the properties of A', since the Coulomb amplitude A_c is explicitly known.[19]

[19] See, for example, A. Messiah, *Quantum Mechanics* (John Wiley & Sons, New York, 1966), p. 430, Eq. (XI.55a). In particular, $A_c(k, \cos\theta)$ is analytic in $\cos\theta$ except on a cut from $\cos\theta = 1$ to $+\infty$.

In this section, we shall take as the short-range part of our potential a pure Yukawa

$$V(r) = \gamma e^{-\mu r}/r.$$

For this case we show that as $|\lambda| \to \infty$ anywhere in the right half-plane $\mathrm{Re}\,\lambda \geq 0$ (*including* the imaginary axis) $a'(\lambda, k)$ is bounded by

$$|a'(\lambda, k)| < \text{const} \times |\lambda^{2\nu - \frac{1}{2}} e^{-\alpha\lambda}| \quad (3.3)$$

as $|\lambda| \to \infty$ with $\mathrm{Re}\,\lambda \geq 0$ and k real, where

$$\alpha = \cosh^{-1}(1 + \mu^2/2k^2). \quad (3.4)$$

In the standard way[20] this bound guarantees that the partial-wave series (3.2) converges and defines an analytic function of $\cos\theta$ provided

$$\mathrm{Im}\,\theta < \alpha. \quad (3.5)$$

This condition confines $\cos\theta$ to the interior of the so-called Lehmann ellipse centered at the origin with semimajor axis $(1 + \mu^2/2k^2)$.

The bound (3.3) also allows us to continue beyond the Lehmann ellipse using the Sommerfeld–Watson transformation; that is, we can replace the sum (3.2) by the appropriate contour integral and then distort the contour to the imaginary λ axis to give the well-known expression[20]

$$A'(k, \cos\theta) = i\int_{-i\infty}^{i\infty} d\lambda \frac{\lambda a'(\lambda, k)P_l(-\cos\theta)}{\sin\pi l} + \text{Regge pole terms}, \quad (3.6)$$

where the usual Regge pole terms are, according to Eq. (3.3), finite in number. The expression (3.6) is convergent and defines an analytic function of $\cos\theta$ provided only

$$\mathrm{Re}\,\theta \neq 0;$$

that is, A' is analytic in the $\cos\theta$ plane cut from 1 to $+\infty$.

Combining this result with the analyticity in the Lehmann ellipse of Eq. (3.5), we conclude that, as one might expect, $A'(k, \cos\theta)$ is analytic for all $\cos\theta$ except on a cut from $\cos\theta = (1 + \mu^2/2k^2)$ to $+\infty$.

B. The Proof

In order to establish the bound (3.3) for $a'(\lambda, k)$, we express a' in terms of an off-shell T matrix satisfying a Fredholm equation of the Lippmann–Schwinger type. Our method is suggested by a paper of Brown *et al.*[21] on short-range potentials, modified by a trick described by Scadron, Weinberg, and Wright.[22]

[20] See, for example, E. J. Squires, *Complex Angular Momenta and Particle Physics* (W. A. Benjamin, Inc., New York, 1964), pp. 3–6.

[21] L. D. Brown, D. Fivel, B. W. Lee, and R. F. Sawyer, Ann. Phys. (N.Y.) **23**, 187 (1963).

[22] M. Scadron, S. Weinberg, and J. Wright, Phys. Rev. **135B**, 202 (1964).

Our starting point is the integral equation

$$\psi^+(r) = \psi_c(r) + \int_0^\infty dr' G^+(r, r')V(r')\psi^+(r') \quad (3.7)$$

for the scattering wavefunction $\psi^+(\lambda, k, r)$. In this equation G^+ is the Green's function

$$G^+(r, r') = -\chi_c^-(r_>)\phi_c(r_<)/f_c^-$$

and ψ_c is normalized as

$$\psi_c(r) = \beta\phi_c(r),$$

where

$$\beta = ie^{-\pi\alpha/2k}(2k)^{l+1} \times [\Gamma(l + \nu + 1)\Gamma(l - \nu + 1)]^{\frac{1}{2}}/\pi\Gamma(2\lambda + 1).$$

The scattering solution $\psi^+(\lambda, k, r)$ is, of course, proportional to the solution $\phi(\lambda, k, r)$.

The off-shell T matrix is defined as usual to be

$$T_\lambda^+(p, k) = \int_0^\infty dr\, \psi_c(\lambda, p, r)V(r)\psi^+(\lambda, k, r) \quad (3.8)$$

and comparison of the behavior of ψ^+ and ϕ for large r shows that the amplitude $a'(\lambda, k)$ is related to the on-shell T matrix $T_\lambda^+(k, k)$ as follows:

$$a'(\lambda, k) = T_\lambda^+(k, k)S_c(\lambda, k)/k^2,$$

where S_c denotes the pure Coulomb partial-wave S matrix

$$S_c(\lambda, k) = \Gamma(l + \nu + 1)/\Gamma(l - \nu + 1).$$

Since $S_c \sim \lambda^{2\nu}$ as $|\lambda| \to \infty$, this means that we can prove the bound Eq. (3.3) by showing that

$$|T_\lambda^+(k, k)| < \text{const} \times |\lambda^{-\frac{1}{2}}e^{-\alpha\lambda}| \quad (3.9)$$

as $|\lambda| \to \infty$, $\text{Re}\,\lambda \geq 0$, k real.

We obtain this bound on T^+ by studying the Lippmann–Schwinger equation which follows from the representation[23]

$$G^+(r, r') = \int_0^\infty dq \frac{\psi_c(q, r)\psi_c(q, r')}{k^2 - q^2 + i\epsilon}. \quad (3.10)$$

Substitution of this representation into Eqs. (3.7) and (3.8) gives the integral equation

$$T^+(p, k) = T_c(p, k) + \int_0^\infty dq \frac{T_c(p, q)T^+(q, k)}{k^2 - q^2 + i\epsilon}, \quad (3.11)$$

where T_c is defined by Eq. (3.8) with ψ^+ replaced by ψ_c.

[23] See Ref. 8.

As it stands, this integral equation is not L^2 for real k. It is, however, a simple matter to replace Eq. (3.11) by an equivalent equation for the operator

$$\tilde{T}^+ = V^{-\frac{1}{2}}T^+V^{-\frac{1}{2}}$$

and the L^2 norm of the latter equation is

$$\int_0^\infty dk' \int_0^\infty dk'' \frac{|T_c(k', k'')|^2}{(k^2 - k'^2 - i\epsilon)(k^2 - k''^2 + i\epsilon)}. \quad (3.12)$$

To show that this norm exists and has suitable properties as $|\lambda| \to \infty$ one must now assume some specific form for the short-range part of the potential. We shall consider just the case where V is a pure Yukawa, in which case T_c can be explicitly evaluated in terms of the hypergeometric function[24]

$$T_c(k', k'') = \gamma\beta(k')\beta(k'') \times \Gamma(b)s^{a'+a''-b}(s - 2ik')^{-a'}(s - 2ik'')^{-a''} \times F[a', a'', b, -4k'k''/(\mu^2 + [k' - k'']^2)],$$

where

$$a' = l + 1 - i\alpha/k',$$
$$b = 2\lambda + 1,$$
$$s = \mu + i(k' + k'').$$

Using standard properties of the hypergeometric function, it is easy to show that the integral (3.12) is convergent. Furthermore, from the asymptotic form of the hypergeometric function,[25] it follows that

$$|T_c(k', k'')| \leq f(k', k'')\lambda^{-\frac{1}{2}}e^{-\alpha\lambda},$$

where

$$\alpha = \cosh^{-1}[(\mu^2 + k'^2 + k''^2)/2k'k''] \quad (3.13)$$

and the precise form of the function $f(k', k'')$ is unimportant. Substitution of the bound (3.13) into the norm (3.12) shows the latter to be $O(\lambda^{-\frac{1}{2}})$ as $|\lambda| \to \infty$ anywhere in the right half-plane $\text{Re}\,\lambda > 0$. This means that the solution T^+ of Eq. (3.11) can be arbitrarily well approximated by T_c when λ is sufficiently large, which, in turn, implies that T^+ satisfies a bound of the same type as T_c for large $|\lambda|$, namely, Eq. (3.13). This is the desired result.

[24] The properties of the hypergeometric functions used in this section can be found in *Higher Transcendental Functions*, A. Erdelyi, Ed. (McGraw-Hill Book Co., New York, 1953), Vol. 1.

[25] See, in particular, Ref. 24, p. 77, Eq. (16). The bound (3.13) is analogous to the result obtained from Hobson's inequality for Legendre functions in Ref. 21.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Functional Methods in Statistical Mechanics. I. Classical Theory*

W. E. BRITTIN AND W. R. CHAPPELL
Department of Physics and Astrophysics and Joint Institute for Laboratory Astrophysics,† Boulder, Colorado

(Received 15 April 1968)

A statistical-mechanical theory of fields is developed. Since a field has an infinite number of degrees of freedom, it is natural and convenient to use functional methods for its description. The most general statistical-mechanical state for a field is represented by a distribution *functional* which satisfies a functional differential equation analogous to the Liouville equation. The functional Fourier transform (characteristic functional) is introduced and its properties are studied. Multitime functionals and various reduced distribution *functions* are also discussed. The formalism is applied to the free electromagnetic fields as well as to a system of charged particles (plasma) interacting via the electromagnetic field.

INTRODUCTION

The statistical theory of charged particles and electromagnetic fields has been for some time a problem of great interest. Usually, the statistics of the field enters only indirectly through relations between the field and the fluctuating charge and current densities due to the particles. However, in 1957 Brittin[1] and, independently, Harris,[2] developed an approach in which the fields expressed in terms of the oscillator variables were treated statistically on an equal footing with the particles by means of an extended phase space. This approach has been used by several people to derive kinetic equations.[3] However, it has some inherent difficulties[4] which can, to a large extent, be traced to the fact that for many cases the oscillator coordinates do not provide the most appropriate description of the electromagnetic field. In many ways it is more satisfactory to have a statistical theory in terms of the actual measurable electric and magnetic fields. Such a theory is presented here and it is shown to provide a very succinct and powerful way of describing such a system.

By its very nature, this description leads to the introduction of probability *functionals* because the corresponding "phase space" is a *function* space. Functional descriptions are not new in kinetic theory. They have been used to describe Brownian motion,[5] to re-express the BBKGY hierarchy,[6] and to treat the problem of turbulent fluids.[7,8] Of course, the methods of functional analysis are still not developed to a point where there are many techniques for solving explicit problems. However, in many problems, particularly in quantum field theory,[9] they have proved very useful. Thus, it is important to formulate certain problems in functional terms—especially problems involving fluctuating fields. Such problems are most naturally expressed in functional form. Several authors have considered the latter problem in the recent past in connection with charged particles with only longitudinal fields present.[10,11]

In this paper we consider the general problem of the description of statistically fluctuating fields. Such fields might be the hydrodynamic velocity field of fluids, the average one-particle phase-space distribution function for particles, or the exact microscopic phase-space distribution function for particles. We derive general expressions for the *functional* Liouville equation for systems described by such fields. We then introduce the characteristic functional which obeys an equation of motion derived from the Liouville equation. The characteristic functional is of particular interest in turbulence, because various moments of the field are given very simply in terms of the characteristic functional.[7,8] The theories of Hopf,[7] and Rosen,[8] and Nakayama and Dawson[10] are included within the general formalism presented here. A rigorous foundation for the work of Dupree[12] is also given.

I. DISTRIBUTION FUNCTIONALS

A field is thought of as a real function $\phi(x)$ defined on some underlying space whose points are denoted

* This work was supported in part by the U.S. Air Force Wright Air Development Division.

† JILA is operated jointly by the National Bureau of Standards and University of Colorado.

[1] W. E. Brittin, Phys. Rev. **106**, 843 (1957).

[2] E. G. Harris, Naval Research Laboratories Report 4944, 1957.

[3] Yu. L. Klimontovich, Zh. Eksp. Teor. Fiz. **37**, 735 (1959) [Sov. Phys.—JETP **10**, 524 (1960)].

[4] R. E. Aamodt, O. C. Eldridge, and N. Rostoker, Phys. Fluids **7**, 1952 (1964).

[5] S. G. Brush, Rev. Mod. Phys. **33**, 79 (1961).

[6] N. N. Bogoliubov, in *Studies in Statistical Mechanics, Vol. I*, J. de Boer and G. E. Uhlenbeck, Eds. (North-Holland Publ. Co., Amsterdam, 1962),

[7] E. Hopf, J. Ratl. Mech. Anal. **1**, 87 (1952); E. Hopf and E. W. Titt, *ibid.* **2**, 587 (1953).

[8] S. Rosen, Phys. Fluids **3**, 519 (1960).

[9] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill Book Co., New York, 1965).

[10] T. Nakayama and J. Dawson, J. Math. Phys. **8**, 553 (1967); T. Nakayama, Phys. Fluids **10**, 247 (1967).

[11] I. Hosokawa, J. Math. Phys. **8**, 221 (1967).

[12] T. H. Dupree, Phys. Fluids **6**, 1714 (1963).

by x (usually, x denotes a point in some Euclidean n-space E_n). The function ϕ may have any (finite) number r of components[13] ϕ_i. The field, in general, depends upon the time t, and is assumed to satisfy a partial differential equation which is of first order in time

$$\frac{\partial\hat{\phi}}{\partial t} = \Lambda[\hat{\phi}], \quad \text{i.e.,} \quad \frac{\partial\hat{\phi}_i}{\partial t} = \Lambda_i[\hat{\phi}], \quad i = 1, \cdots, r. \tag{1}$$

In general, Λ may be a nonlinear operator, but it does not depend upon time. The "caret" over ϕ indicates that $\hat{\phi}$ is a time-dependent field which satisfies Eq. (1). If ϕ is a field which does not depend upon time, we define $\dot{\phi}(x)$ as

$$\dot{\phi}(x) \equiv \Lambda[\phi]. \tag{2}$$

The statistical description of the field may be accomplished by the use of a field distribution functional $F = F[\phi, t]$. Roughly speaking, F is a "function" of the values $\phi(x), \phi(x'), \cdots$ of ϕ at all the points x in E_n. The functional F is the probability density for the field $\phi(x)$, which signifies that $F[\phi, t]\, d[\phi]$ is the probability that at time t the field is found to be $\phi(x)$ within the volume element $d[\phi]$ of function space.[14] We normalize F to unity:

$$\int F[\phi, t]\, d[\phi] = 1. \tag{3}$$

The distribution functional satisfies the continuity equation (conservation of probability):

$$\frac{\partial F}{\partial t} + \int dx \Lambda[x, \phi] \frac{\delta}{\delta\phi(x)} F = 0, \tag{4}$$

where we have assumed that $\delta\Lambda[x, \phi]/\delta\phi(x) = 0$.[15] The functional derivative is defined as the distribution[14] $\delta F/\delta\phi$ in the following expression

$$\frac{d}{d\lambda} F[\phi + \lambda\phi_1]|_{\lambda=0} = \left(\frac{\delta F}{\delta\phi}, \phi_1\right), \tag{5}$$

where the scalar product $(\delta F/\delta\phi, \phi_1)$ for real functions is defined by

$$\int dx \frac{\delta F[\phi]}{\delta\phi(x)} \phi_1(x).$$

[13] The formalism presented here is more general than the Hamiltonian formulation of fields which it includes as a special case. For example, we may think of ϕ as having "coordinate" and "momentum" components.

[14] See Appendix A for a fuller discussion of the various mathematical concepts used in this paper.

[15] For the problems considered in this paper we have $\delta\Lambda/\delta\phi = 0$. If this is not the case, some of the following development must be slightly modified.

Equation (4) may be written in operator form

$$\frac{\partial F}{\partial t} + LF = 0, \tag{6}$$

with the Liouville operator defined as

$$L \equiv \int dx \Lambda[x, \phi] \frac{\delta}{\delta\phi(x)}. \tag{7}$$

Average values $\langle G\rangle$ of functionals $G[\phi, t]$ of the field are computed through functional integration of the product of F and G:

$$\langle G[\phi]\rangle = \int F[\phi, t] G[\phi, t]\, d[\phi] \equiv \langle F, G\rangle. \tag{8}$$

In the above equation, $\langle F, G\rangle$ defines a scalar product for functionals.

The Liouville operator is skew-symmetric with respect to the scalar product $\langle F, G\rangle$:

$$\langle LF, G\rangle = -\langle F, LG\rangle, \tag{9}$$

where it is assumed that F and G are in the domain of L.

The formal solution of Eq. (6) is expressed by

$$F[\phi, t] = e^{-tL} F_0[\phi], \tag{10}$$

$F_0[\phi]$ being the value of F at time zero. Equation (10) may also be expressed in terms of the solution $\hat{\phi}(-t)$ of Eq. (1) which has the value $\phi(x)$ at time zero:

$$F[\phi, t] = F_0[\hat{\phi}(-t)], \quad \hat{\phi}(0) \equiv \phi. \tag{11}$$

Equation (11) follows from the fact that probability is conserved along the natural motion of the system. The equation of motion for average values of functionals of ϕ is obtained directly through

$$\begin{aligned} \frac{d}{dt}\langle G[\phi]\rangle &= \left\langle \frac{\partial F}{\partial t}, G \right\rangle \\ &= -\langle LF, G\rangle \\ &= \langle F, LG\rangle \equiv \langle F, \dot{G}\rangle, \end{aligned} \tag{12}$$

with $\dot{G} \equiv LG$. (If G has an explicit dependence upon time, we must define $\dot{G}$ as $LG + \partial G/\partial t$.) We may also introduce "Heisenberg" operators $\hat{G}(t)$ which carry all of the time evolution. That is,

$$\begin{aligned} \langle G\rangle &= \langle F(t), G\rangle = \langle e^{-tL} F_0, G\rangle \\ &= \langle F_0, e^{tL} G\rangle = \langle F_0, \hat{G}(t)\rangle, \end{aligned} \tag{13}$$

where $\hat{G}(t) \equiv e^{tL} G$. Thus, $\langle G\rangle_t = \langle \hat{G}(t)\rangle_0$ which means that all statistical averaging may be done at the initial time. In the above equations, G itself may depend explicitly upon time.

If the delta functional $\Delta[\phi]$ is introduced (see Appendix A) having the properties

$$\Delta[\phi] = 0, \quad \phi \not\equiv 0,$$
$$\int \Delta[\phi]\, d[\phi] = 1,$$
$$\int \Delta[\phi_1 - \phi_2] G[\phi_2]\, d[\phi_2] = G[\phi_1], \tag{14}$$

we may recover solutions of the exact field equations by choosing $F_0[\phi] = \Delta[\phi - \phi_0]$. In fact, for any functional $G[\phi]$, we have

$$\begin{aligned}\langle G[\phi]\rangle &= \int G[\phi] e^{-L(\phi)t} \Delta[\phi - \phi_0]\, d[\phi] \\ &= \int (e^{L(\phi)t} G[\phi]) \Delta[\phi - \phi_0]\, d[\phi] \\ &= e^{L(\phi_0)t} G[\phi_0] = G[e^{L(\phi_0)t}\phi_0] = G[\hat{\phi}(x, t)].\end{aligned} \tag{15}$$

In particular for $G = \phi$,

$$\langle \phi(x)\rangle = \hat{\phi}(x, t) = e^{tL(\phi_0)}\phi_0. \tag{16}$$

We note that $\hat{\phi}(x, t)$ is a functional of ϕ_0, the value of $\hat{\phi}$ at $t = 0$, and that the integration of Eq. (1) is equivalent to Eq. (16). For from Eq. (1), it follows that

$$\hat{\phi}[\phi_0, t] = \phi_0 + \int_0^t \Lambda[\hat{\phi}(\phi_0, t)]\, dt$$

and, from Eq. (15),

$$\begin{aligned}\hat{\phi}[\phi_0, t] &= \phi_0 + \int_0^t e^{tL(\phi_0)} \Lambda[\phi_0]\, dt \\ &= \phi_0 + \frac{e^{tL(\phi_0)} - 1}{L(\phi_0)} \Lambda[\phi_0] \\ &= \phi_0 + \frac{e^{tL(\phi_0)} - 1}{L(\phi_0)} L(\phi_0)\phi_0 \\ &= e^{tL(\phi_0)}\phi_0.\end{aligned} \tag{17}$$

Conversely, from $\hat{\phi}(t) = e^{tL(\phi_0)}\phi_0$, it follows that

$$\begin{aligned}\hat{\phi}(t) &= \phi_0 + (e^{tL(\phi_0)} - 1)\phi_0 \\ &= \phi_0 + \frac{e^{tL(\phi_0)} - 1}{L(\phi_0)} L(\phi_0)\phi_0 \\ &= \phi_0 + \int_0^t e^{\tau L(\phi_0)}\, d\tau\, \Lambda[\phi_0] \\ &= \phi_0 + \int_0^t d\tau\, \Lambda[e^{\tau L(\phi_0)}\phi_0] \\ &= \phi_0 + \int_0^t d\tau\, \Lambda[\hat{\phi}(\tau)],\end{aligned} \tag{18}$$

where we have used the fact that

$$\begin{aligned}L(\phi_0)\phi_0(x) &= \int dx'\, \Lambda[x', \phi_0] \frac{\delta}{\delta\phi_0(x')} \phi_0(x) \\ &= \Lambda[x, \phi_0] = \dot{\phi}_0.\end{aligned}$$

The distribution functional $F[\phi, t]$ may also be expressed in terms of an average of a delta functional, since

$$\begin{aligned}F[\phi, t] = F_0[\phi(-t)] &= \int F_0[\phi_1] \Delta[\phi_1 - \hat{\phi}(-t)]\, d[\phi_1] \\ &= \langle \Delta[\phi - \hat{\phi}(-t)]\rangle_0.\end{aligned} \tag{19}$$

This form for $F[\phi, t]$ is useful if we wish to consider averages of functionals at different times. For example, if we want the average value of $\hat{G}[\hat{\phi}(t_1), \hat{\phi}(t_2), \cdots, \hat{\phi}(t_n)]$, we may write

$$\begin{aligned}&\langle \hat{G}[\hat{\phi}(t_1), \cdots, \hat{\phi}(t_n)]\rangle \\ &\quad = \int F_n[\phi_1, t_1; \phi_2, t_2; \cdots; \phi_n, t_n] \\ &\qquad \times G[\phi_1, \phi_2 \cdots \phi_n]\, d[\phi_1] \cdots d[\phi_n],\end{aligned} \tag{20}$$

where we have introduced the multiple-time distribution functional $F_n[\phi_1, t_1; \cdots; \phi_n, t_n]$. However, we now show that, since all the statistical information is contained in $F[\phi, t] = F_1[\phi, t]$, we may express F_n in terms of F. In fact,

$$\begin{aligned}&\langle \hat{G}[\hat{\phi}(t_1), \cdots, \hat{\phi}(t_n)]\rangle \\ &\quad = \int F[\phi, 0] \hat{G}[\hat{\phi}(t_1), \hat{\phi}(t_2) \cdots \hat{\phi}(t_n)]\, d[\phi] \\ &\quad = \int F[\phi, 0] \int \Delta[\phi_1 - \hat{\phi}(t_1)] \\ &\qquad \times \Delta[\phi_2 - \hat{\phi}(t_2)] \cdots \Delta[\phi_n - \hat{\phi}(t_n)] \\ &\qquad \times G[\phi_1, \phi_2, \cdots, \phi_n]\, d[\phi_1] \cdots d[\phi_n]\, d[\phi]\end{aligned} \tag{21}$$

$$\begin{aligned}&\quad = \int F[\phi, 0] \exp\left\{-\sum_{i=1}^n L_i t_i\right\} \\ &\qquad \times \Delta[\phi_1 - \phi] \Delta[\phi_2 - \phi] \cdots \Delta[\phi_n - \phi] \\ &\qquad \times G[\phi_1, \cdots, \phi_n]\, d[\phi_1] \cdots d[\phi_n]\, d[\phi].\end{aligned} \tag{22}$$

Therefore,

$$\begin{aligned}&F_n[\phi_1, t_1; \phi_2, t_2; \cdots; \phi_n, t_n] \\ &\quad = \int d[\phi] F[\phi, 0] \exp\left\{-\sum_{i=1}^n L_i t_i\right\} \\ &\qquad \times \Delta[\phi_1 - \phi] \cdots \Delta[\phi_n - \phi] \\ &\quad = \int d[\phi] F[\phi, 0] \exp\left\{-\sum_{i=1}^n L_i t_i\right\} \\ &\qquad \times \Delta[\phi_1 - \phi] \Delta[\phi_2 - \phi_1] \cdots \Delta[\phi_n - \phi_1] \\ &\quad = \int d[\phi] F[\phi, 0] \Delta[\hat{\phi}_1(-t_1) - \phi] \\ &\qquad \times \exp\left\{-\sum_{i=1}^n L_i t_i\right\} \Delta[\phi_2 - \phi_1] \cdots \Delta[\phi_n - \phi_1] \\ &\quad = F[\phi_1 t_1] \exp\left\{-\sum_{i=1}^n L_i t_i\right\} \Delta[\phi_2 - \phi_1] \cdots \Delta[\phi_n - \phi_1] \\ &\quad = F[\phi_1 t_1] \Delta[\hat{\phi}_2(-t_2) - \hat{\phi}_1(-t_1)] \cdots \\ &\qquad \Delta[\hat{\phi}_n(-t_n) - \hat{\phi}_1(-t_1)],\end{aligned} \tag{23}$$

which shows that all multiple-time correlations may be expressed in terms of $F[\phi, t]$, the single-field distribution functional.

II. REDUCED DISTRIBUTION FUNCTIONS

The specification of $F[\phi, t]$ requires an enormous amount of statistical information which must be given at the same time, say time zero. Many questions do not require such a vast amount of information for their elucidation. For example, if G is an ordinary function of the field variables $G = G[\phi(x_1), \phi(x_2), \cdots, \phi(x_n)]$, we have for the average of G:

$$\begin{aligned}\langle G\rangle &= \Big\langle \int\cdots\int G(\phi_1, \cdots, \phi_n) \\ &\quad \times \delta(\phi_1 - \phi(x_1))\cdots\delta(\phi_n - \phi(x_n))\, d\phi_1\cdots d\phi_n \Big\rangle \\ &= \iint G(\phi_1, \cdots, \phi_n)\, d\phi_1\cdots d\phi_n \\ &\quad \times \langle \delta(\phi_1 - \phi(x_1))\cdots\delta(\phi_n - \phi(x_n))\rangle \\ &\equiv \iint d\phi_1\cdots d\phi_n G(\phi_1, \cdots, \phi_n) \\ &\quad \times f_n(\phi_1, x_1; \phi_2, x_2; \cdots \phi_n, x_n; t), \end{aligned} \tag{24}$$

where $\phi_1, \cdots, \phi_n$ are numerical-valued quantities and $f_n\, d\phi_1\cdots d\phi_n$ may be regarded as the probability that at time t the field has the value ϕ_1 at x_1, ϕ_2 at $x_2, \cdots$, and ϕ_n at x_n within the range $d\phi_1, \cdots, d\phi_n$. Of particular interest are the moments $\langle\phi(x_1)\cdots\phi(x_n)\rangle$ which are simply related to f_n:

$$\begin{aligned}&\langle\phi(x_1)\cdots\phi(x_n)\rangle \\ &\quad = \int \phi_1\phi_2\cdots\phi_n f_n(\phi_1, \cdots, \phi_n t)\, d\phi_1\cdots d\phi_n. \end{aligned} \tag{25}$$

The Heisenberg picture may be used to obtain a particularly useful form for f_n

$$\begin{aligned}&f_n(x_1\phi_1, \cdots, x_n\phi_n t) \\ &\quad = \Big\langle \prod_{i=1}^{n} \delta(\phi_i - \phi(x_i)) \Big\rangle \\ &\quad = \Big\langle e^{-tL}F_0, \prod_{i=1}^{n} \delta(\phi_i - \phi(x_i)) \Big\rangle \\ &\quad = \Big\langle F_0, \prod_{i=1}^{n} \delta(\phi_i - \hat{\phi}(x_i, t)) \Big\rangle \\ &\quad = \langle \delta(\phi_1 - \hat{\phi}(x_i t))\cdots\delta(\phi_n - \hat{\phi}(x_n t))\rangle_0. \end{aligned} \tag{26}$$

The latter form has been used by Dupree[12] in his treatment of radiation and plasmas.

If we consider an arbitrarily large volume V of the underlying space E_n, and divide V into an arbitrarily large but finite number M of cells $\Delta\tau_\alpha$, $\alpha = 1, \cdots, M$, $V = \sum_\alpha \Delta\tau_\alpha$, then we may "project" the function $\phi(x)$ upon the cells $\Delta\tau_\alpha$ and write an approximate expression:

$$\phi(x) \sim \sum_{\alpha=1}^{M} \bar{\phi}_\alpha E_\alpha(x), \tag{27}$$

with

$$\bar{\phi}_\alpha \equiv \frac{1}{\Delta\tau_\alpha}\int \phi(x)E_\alpha(x)\, dx, \tag{28}$$

where $E_\alpha(x)$ is the characteristic function for the cell $\Delta\tau_\alpha$ [$E_\alpha(x) = 1$ if $x \in \Delta\tau_\alpha$, and $E_\alpha(x) = 0$ otherwise]. The distribution functional, then, is approximately a function of the M variables $\bar{\phi}_\alpha$:

$$F[\phi] \sim F_M[\bar{\phi}_1\cdots\bar{\phi}_\alpha\cdots\bar{\phi}_M t] \tag{29}$$

and, apart from the normalization constant,

$$\begin{aligned}&\int F[\phi]\, d[\phi] \\ &\quad \sim \int F_M[\bar{\phi}_1, \cdots, \bar{\phi}_M t]\, d\bar{\phi}_1\, d\bar{\phi}_2\cdots d\bar{\phi}_M = 1. \end{aligned} \tag{30}$$

Further, if the $x_1, \cdots, x_n$ in Eq. (26) correspond to cells $\alpha_1, \alpha_2\cdots\alpha_n$, we have approximately

$$\begin{aligned}&f_n(x_1, \phi_1, \cdots, x_n\phi_n t) \\ &\quad \sim \int F_M[\bar{\phi}_1, \cdots, \bar{\phi}_n] \prod_{i=1}^{n} \delta(\phi_i - \bar{\phi}_{\alpha_i}) \prod_{\alpha=1}^{M} d\bar{\phi}_\alpha. \end{aligned} \tag{31}$$

The above argument also shows how, in a manner similar to that used in Wiener measure,[5] we can recover the full distribution functional by allowing n in f_n to become infinite in a suitable way such that $(\phi_1, \phi_2, \cdots, \phi_n) \to \phi(x)$.

The equations of motion for f_n form a hierarchy similar to the BBGKY hierarchy of ordinary statistical mechanics. These equations may be derived from the equation of motion for $F[\phi, t]$ or from that of ϕ using Eq. (26). We can also introduce multitime correlation functions, most simply defined by

$$\begin{aligned}&f_n(x_1t_1\phi_1; x_2t_2\phi_2; \cdots; x_nt_n\phi_n) \\ &\qquad = \Big\langle \prod_{i=1}^{n} \delta[\phi_i - \hat{\phi}(x_i t_i)] \Big\rangle_0, \end{aligned} \tag{32}$$

which find application to fluctuation phenomena.[12]

We next proceed to the equations satisfied by f_n, and to this end we introduce a generating functional $\mathfrak{G}[u(x, \phi)]$ for f_n by means of

$$\begin{aligned}\mathfrak{G}[u, t] &= 1 + \sum_{n=1}^{\infty} \frac{1}{n!}\int\cdots\int d1\cdots dn \\ &\quad \times f_n(1, 2, \cdots, n, t)u(1)u(2)\cdots u(n), \end{aligned} \tag{33}$$

where $1 \equiv x_1, \phi_1, 2 \equiv x_2, \phi_2$, etc., and $d1 = dx_1\, d\phi_1$, $\cdots$. We have

$$f_n(1, \cdots, n, t) = \left[\frac{\delta^n G[u]}{\delta u(1) \cdots \delta u(n)}\right]_{u=0} \tag{34}$$

From the definition (31) of f_n, it follows that

$$\mathcal{G}[u, t] = \sum_{n=0}^{\infty} \frac{1}{n!} \int \cdots \int d1 \cdots dn \times \left\langle \prod_{i=1}^{n} \delta(\phi_i - \phi(x_i)) \right\rangle u(1) \cdots u(n), \tag{35}$$

$$\mathcal{G}[u, t] = \left\langle \exp\left[\int d1\, \delta(\phi_1 - \phi(x_1))u(1)\right]\right\rangle = \int F[\phi, t] \exp\left[\int d1\, \delta(\phi_1 - \phi(x_1))u(1)\right] d[\phi]. \tag{36}$$

We set $E[\phi, u] = \exp\left[\int d1\, \delta(\phi_1 - \phi(x_1))u(1)\right]$ so that

$$\mathcal{G} = \int F \cdot E\, d[\phi] = \langle F, E\rangle \tag{37}$$

and

$$\frac{\partial \mathcal{G}}{\partial t} = \left\langle \frac{\partial F}{\partial t}, E \right\rangle = -\langle LF, E\rangle = \langle F, LE\rangle, \tag{38}$$

where $L = \int dx\, \Lambda[x, \phi]\delta/\delta\phi(x)$. We note that, since

$$\frac{\delta E}{\delta \phi(x)} = E \int d\phi_1 \delta(\phi_1 - \phi(x)) \frac{\partial u(1)}{\partial \phi_1} \tag{39}$$

and

$$\frac{\delta E}{\delta u(1)} = E\delta(\phi_1 - \phi(x_1)), \tag{40}$$

we have

$$\frac{\partial \mathcal{G}}{\partial t} = \left\langle F, \int dx \Lambda[x, \phi(\cdot)] \int d\phi_1 \delta(\phi_1 - \phi(x)) \frac{\partial u(x\phi_1)}{\partial \phi_1} E \right\rangle \tag{41}$$

$$= \left\langle F, \int dx\, \Lambda\left[x, \int \phi_1 \delta(\phi_1 - \phi(\cdot))\, d\phi_1\right] \times \int d\phi_2 \delta(\phi_2 - \phi(x)) \frac{\partial u(x\phi_2)}{\partial \phi_2} E \right\rangle \tag{42}$$

$$= \left\langle F, \int dx\, \Lambda\left[x, \int d\phi_1 \phi_1 \frac{\delta}{\delta u(\cdot\, \phi_1)}\right] \times \int d\phi_2 \frac{\partial u(x\phi_2)}{\partial \phi_2} \frac{\delta}{\delta u(x\phi_2)} E \right\rangle \tag{43}$$

or, finally,

$$\frac{\partial \mathcal{G}}{\partial t} = \iint dx\, d\phi \frac{\partial u(x\phi)}{\partial \phi} \frac{\delta}{\delta u(x\phi)} \times \Lambda\left[x, \int d\phi_1 \phi_1 \frac{\delta}{\delta u(\cdot\, \phi_1)}\right]\mathcal{G}. \tag{44}$$

For applications it is important to note that

$$\frac{\delta^2 \mathcal{G}}{\delta u(x, \phi_1)\delta u(x, \phi_2)} = \delta(\phi_1 - \phi_2) \frac{\delta \mathcal{G}}{\delta u(x, \phi_1)}. \tag{45}$$

The above equation expresses the fact that

$$f_2(x, \phi_1; x, \phi_2) = \delta(\phi_1 - \phi_2) f_1(x\phi_1), \quad \text{etc.} \tag{46}$$

We may obtain an expression for $\partial f_n/\partial t$ by differentiating Eq. (44) n times with respect to $u(1), \cdots, u(n)$ and setting $u \equiv 0$:

$$\frac{\partial f_n}{\partial t}(x_1, \phi_1; \cdots; x_n, \phi_n; t) = \frac{\partial}{\partial t} \frac{\delta^n \mathcal{G}}{\delta u(1) \cdots \delta u(n)}\bigg|_{u=0} = \iint dx\, d\phi \sum_{i=1}^{n} \delta(x - x_i) \frac{\partial \delta(\phi - \phi_i)}{\partial \phi} \times \Lambda\left[x \int d\phi_1 \phi_1 \frac{\delta}{\delta u(\cdot\, \phi_1)}\right] \frac{\delta^n}{\delta u(1) \cdots \delta u(n)} \mathcal{G}\bigg|_{u=0} \tag{47}$$

The general structure of the hierarchy (47) depends, of course, upon $\Lambda[x, \phi]$, but it is clear, in general, that Eq. (47) couples f_n to f_m, $m > n$.

As a special case we consider $\Lambda[x\phi] = A\phi(x) + D\phi(x)$, where A is a constant matrix and D is a matrix involving derivatives with respect to x. Then making use of Eq. (45), we find that

$$\frac{\partial f_n}{\partial t} + \sum_{k=1}^{n} A\phi_k \frac{\partial f_n}{\partial \phi_k} + \int d(n+1) \times \sum_{k=1}^{n} [D_k \delta(x_k - x_{n+1})]\phi_{n+1} \frac{\partial f_{n+1}}{\partial \phi_k} = 0. \tag{48}$$

A very simple example consists of a linear string. Here, $\phi = \binom{y(x)}{v(x)}$, where $y(x)$ is the transverse displacement of the string at the position x along the string, and $v(x)$ is the transverse velocity. The equations of motion are:

$$\dot{\phi} = \begin{pmatrix} \dot{y}(x) \\ \dot{v}(x) \end{pmatrix} = \begin{pmatrix} v(x) \\ c^2 y''(x) \end{pmatrix}, \tag{49}$$

so that $A\phi = \binom{v}{0}$ and

$$D\phi = c^2 \frac{d^2}{dx^2}\begin{pmatrix} 0 \\ y(x) \end{pmatrix}.$$

Therefore the hierarchy for the linear string is[16,17]

$$\frac{\partial f_n}{\partial t}(\mathbf{x}_1, \mathbf{y}_1, \mathbf{v}_1; \cdots; \mathbf{x}_n, \mathbf{y}_n, \mathbf{v}_n; t) + \sum_{k=1}^{n} \mathbf{v}_k \cdot \frac{\partial}{\partial \mathbf{y}_k} f_n + c^2 \sum_{k=1}^{n} \int \delta''(x_k - x_{n+1}) \mathbf{y}_{n+1} \cdot \frac{\partial f_{n+1}}{\partial v_k} = 0. \quad (50)$$

Equation (44) for $\mathfrak{G}$ can be expressed somewhat more simply if one introduces the functional Fourier transform of $\mathfrak{G}$:

$$\mathfrak{G}[u] = \int e^{i(u,v)} \mathcal{N}[v]\, d[v(x\phi)],$$

with $(u, v) \equiv \int u(x\phi)v(x\phi)\, dx\, d\phi$. Then f_n are expressed as moments of v with respect to $\mathcal{N}[v]$:

$$f_n(1, \cdots, nt) = \frac{\delta^u \mathfrak{G}}{\delta u(1) \cdots \delta u(n)}\bigg|_{u \equiv 0} = i^n \int v(1) \cdots v(n) \mathcal{N}[u]\, d[u] \equiv i^n \langle v(1) \cdots v(n) \rangle_{\mathcal{N}}. \quad (51)$$

The equation satisfied by $\mathcal{N}$ is obtained from Eq. (44) and reads

$$\frac{\partial \mathcal{N}}{\partial t} = \iint dx\, d\phi \frac{\partial v(x, \phi)}{\partial \phi} \Lambda\left[x, i \int d\phi_1 \phi_1 v(\cdot\, \phi_1)\right] \frac{\delta \mathcal{N}}{\delta v(x\phi)}. \quad (52)$$

It is important to note that, as a result of relation (45), we have

$$i^2 v(x\phi_1) v(x\phi_2) \mathcal{N} = i\delta(\phi_1 - \phi_2) v(x\phi_1) \mathcal{N}. \quad (53)$$

We now present another way of obtaining expressions for the functions f_n. From the definition Eq. (31), we have

$$f_n(1, \cdots, nt) = \int \prod_{i=1}^{n} \delta(\phi_1 - \phi(x_i)) F[\phi]\, d[\phi]. \quad (54)$$

Let

$$F[\phi] = \int e^{i(\phi, \phi_1)} G[\phi_1]\, d[\phi_1],$$

$$(\phi, \phi_1) \equiv \int \phi(x)\phi_1(x)\, dx, \quad (55)$$

then

$$f_n = \int e^{i(\phi, \phi_1)} \prod_{i=1}^{n} \delta(\phi_1 - \phi(x_1)) G[\phi_1]\, d[\phi]\, d[\phi_1] = \frac{1}{(2\pi)^{nr}} \iint e^{i(\phi, \phi_1)} \int \prod_{i=1}^{n} d\omega_i \times \exp\left\{ i \sum_{i=1}^{n} \omega_i (\phi_i - \phi(x)) \right\} G[\phi_1][d\phi_1][d\phi] = \frac{1}{(2\pi)^{nr}} \int \exp\left\{ i \sum_{i=1}^{n} \omega_i (\phi_i - \phi(x)) \right\} \times g_n(x_1, \omega_1; \cdots; x_n, \omega_n; t)\, d\omega_1 \cdots d\omega_n, \quad (56)$$

with

$$g_n \equiv \iint d[\phi]\, d[\phi_1] e^{i(\phi, \phi_1)} \times \exp\left\{ - i \int \sum_{i=1}^{n} \omega_i \delta(x - x_i) \phi(x)\, dx \right\} G[\phi_1] = G[\chi_n], \quad (57)$$

where $\chi_n(x) \equiv \sum_{i=1}^{n} \omega_i \delta(x - x_i)$. In order to obtain Eq. (57), we made use of the delta functional defined by Eq. (14). It is interesting to observe that the value of the functional Fourier transform G of F for the argument χ_n gives the ordinary Fourier transform g_n of f_n.

Expressions for the moments or field correlations,[18]

$$\langle \phi(x_1)^{n_1} \phi(x_2)^{n_2} \cdots \phi(x_s)^{n_s} \rangle \equiv \int f_s(x_1\phi_1, \cdots, x_s\phi_s) \phi_1^{n_1} \cdots \phi_s^{n_s}\, d\phi_1 \cdots d\phi_s,$$

may be obtained directly. Since

$$\frac{1}{(2\pi)^r} \int e^{i\omega\phi} \phi^n\, d\phi = (-i)^{|n|} \delta^{(n)}(\omega),$$

where $|n| \equiv n^1 + n^2 + \cdots + n^r$, Eq. (55), when used in the above expression for the moments, yields

$$\langle \phi(x_1)^{n_1} \phi(x_2)^{n_2} \cdots \phi(x_s)^{n_s} \rangle = (i)^{(|n_1| + |n_2| + \cdots + |n_s|)} \frac{\partial^{n_1 + \cdots + n_s}}{\partial \omega_1^{n_1} \cdots \partial \omega_s^{n_s}} G[\chi]\big|_{\omega = 0} \quad (58)$$

$$= (i)^{(|n_1| + |n_2| + \cdots + |n_s|)} \frac{\partial^{n_1 + \cdots + n_s}}{\delta \phi_1(x_1)^{n_1} \cdots \delta \phi_1(x_s)^{n_s}} \times G[\phi_1]\big|_{\phi_1 = 0}. \quad (59)$$

A special case of the above result was first reported by Hopf.[7]

[16] W. E. Brittin and W. R. Chappell, *Lectures in Theoretical Physics, Vol. VIII*, W. E. Brittin, Ed. (University of Colorado Press, Boulder, Colo., 1966); W. E. Brittin, *Lectures in Theoretical Physics, Vol. IX*, W. E. Brittin, Ed. (Gordon and Breach Science Publishers, New York, 1967).

[17] W. E. Brittin, W. R. Chappell, and A. Y. Sakakura, U.S.A.F. Aerospace Research Laboratories Technical Report 64-85, 1964; W. R. Chappell, JILA Report 35, 1965.

[18] The field ϕ has r components; thus, n is an r-component vector $(n^1, n^2, \cdots, n^r)$ and ϕ^n is $\phi_1^{n_1}\phi_2^{n_2} \cdots \phi_r^{n_r}$. For some problems it is convenient to set $K(2, 1) = 0$ for $t_2 < t_1$.

III. SUBSIDIARY CONDITIONS

It may happen that the components of ϕ are not independent but satisfy certain subsidiary conditions

$$S_k[\phi] = 0, \quad k = 1, \cdots, s. \tag{60}$$

We may incorporate these conditions by demanding that F be zero unless Eqs. (60) are satisfied[16]:

$$S_k[\phi]F[\phi, t] = 0, \quad k = 1, \cdots, s. \tag{61}$$

In terms of the functional Fourier transform G of F, these subsidiary conditions become

$$S_k\left[i\frac{\delta}{\delta\phi_1}\right]G[\phi_1, t] = 0, \quad k = 1, \cdots, s. \tag{62}$$

For example, in the case of the electromagnetic field the condition $\nabla \cdot \mathbf{B} = 0$ is reflected in the subsidiary condition

$$\nabla \cdot \frac{\delta}{\delta \mathbf{B}_1(\mathbf{r})} G[\mathbf{B}_1(\mathbf{r})] = 0, \tag{63}$$

which guarantees (by setting $\mathbf{B}_1$ equal to zero in Eq. 63), $\nabla \cdot \langle \mathbf{B}(\mathbf{r}) \rangle = 0$.

In certain cases, such as the electromagnetic field, subsidiary conditions on the field are satisfied by virtue of the field equations if they are satisfied initially. In these cases the reflected subsidiary conditions $S_k[\phi]F = 0$ are satisfied for all times $t > 0$, if $S_k[\phi]F = 0$ at time $t = 0$. We prove this assertion as follows: the condition for $S_k[\hat{\phi}]$ to be zero if it is zero at time $t = 0$, is

$$\begin{aligned}\frac{\partial S_k[\hat{\phi}(xt)]}{\partial t} &= \int dx\, \dot{\hat{\phi}}(xt) \frac{\delta S_k[\hat{\phi}]}{\delta \hat{\phi}(xt)} \\ &= \int dx\, \Lambda[x, \hat{\phi}] \frac{\delta S_k[\hat{\phi}]}{\delta \hat{\phi}(xt)} = 0\end{aligned} \tag{64}$$

or, for an arbitrary function $\phi(x)$,

$$\int dx\, \Lambda[x, \phi] \frac{\delta S_k[\phi]}{\delta \phi(x)} = 0. \tag{65}$$

However,

$$\begin{aligned}S_k[\phi]\frac{\partial F}{\partial t} &= \frac{\partial (S_k F)}{\partial t} \\ &= -S_k \int dx\, \Lambda[x, \phi] \frac{\delta F}{\delta \phi(x)} \\ &= -\int dx\, \Lambda[x, \phi] \frac{\delta (S_k F)}{\delta \phi(x)} \equiv -L(S_k F)\end{aligned}$$

by virtue of Eq. (65). Therefore,

$$S_k F(t) = e^{-tL}[S_k F(0)], \tag{66}$$

so that, if $S_k F = 0$ at time $t = 0$, $S_k[\phi]F[\phi, t] = 0$, $t > 0$. A similar statement applies to $S_k[i\delta/\delta\phi_1] \times G[\phi_1 t] = 0$.

IV. FORMAL SOLUTIONS

We have seen (Sec. I) that the solution of $\partial F/\partial t + LF = 0$ can be written formally as

$$\begin{aligned}F[\phi_2 t_2] &= e^{-(t_2 - t_1)L_2} F[\phi_2, t_1], \\ F[\phi_2 t_2] &= \int e^{-(t_2 - t_1)L_2} \Delta[\phi_2 - \phi_1] F[\phi_1 t_1]\, d[\phi_1]\end{aligned} \tag{67}$$

which suggests, as has been observed by Rosen,[8] that it is convenient to introduce a propagator $K[\phi_2 \phi_1 t_2 t_1]$ such that

$$F[\phi_2 t_2] = \int K[\phi_2, \phi_1, t_2, t_1] F[\phi_1 t_1]\, d[\phi_1]. \tag{68}$$

We observe directly that

$$\begin{aligned}K[\phi_2, \phi_1, t_2, t_1] &= e^{-(t_2 - t_1)L_2} \Delta[\phi_2 - \phi_1] \\ &= \Delta[\hat{\phi}_2(-t_2) - \hat{\phi}_1(-t_1)] \\ &= \Delta[\hat{\phi}_2(t_1 - t_2) - \phi_1] \\ &= \Delta[\phi_2 - \hat{\phi}_1(t_2 - t_1)],\end{aligned} \tag{69}$$

where $\hat{\phi}_i(t)$ satisfies the equation $\partial \hat{\phi}_i/\partial t = \Lambda[\hat{\phi}_i]$ and $\hat{\phi}_i(0) = \phi_i$, $i = 1, 2 \cdots$. The propagator K satisfies the equation[19]

$$\frac{\partial K[2, 1]}{\partial t_2} + L_2 K_2[2, 1] = 0 \tag{70}$$

and the initial condition

$$K[\phi_2, \phi_1, t_1, t_2] = \Delta[\phi_2 - \phi_1]. \tag{71}$$

In addition, it has the semigroup property

$$K[2, 1] = \int K[2, 3]\, d[\phi_3] K[3, 1]. \tag{72}$$

That the formal solution Eq. (69) has the semigroup property follows from the identity

$$\Delta[\phi_2 - \phi_1] = \int \Delta[\phi_2 - \phi_3]\, d[\phi_3] \Delta[\phi_3 - \phi_1]. \tag{73}$$

The solution given by Eq. (69) can also be obtained by solving Eqs. (70) and (71) in terms of path integrals.[6] The interval $[t_1, t_2]$ is split up into a large number, $N + 1$, of intervals $t_1 = t^0, t^1, \cdots, t^{N+1} = t_2$ with $\Delta t = t^{i+1} - t^i$, $i = 0, \cdots, N$. The relation given in Eq. (72) is then repeated N times to yield

$$\begin{aligned}K[2, 1] = \int \cdots \int K(N + 1, N)\, d[\phi^N] K[N, N - 1] \\ \times d[\phi^{N-1}] \cdots d[\phi^1] K[1, 0].\end{aligned} \tag{74}$$

---

[19] For some problems, it is convenient to set $K(2, 1) = 0$ for $t_2 < t_1$. In this case, Eq. (70) becomes $\partial K(2, 1)/\partial t_2 + L_2 K(2, 1) = \delta(t_2 - t_1)\Delta[\phi_2 - \phi_1]$.

We have from Eq. (69) that

$$K[i, i-1] = e^{-\Delta t L_i}\Delta[\phi^i - \phi^{i-1}] = \int e^{-\Delta t L_i} \exp\left\{i\int z_i(x)[\phi^i(x) - \phi^{i-1}(x)]\,dx\right\} d[z_i]. \tag{75}$$

However, since

$$L_i = \int dx_i\, \Lambda(x_i, \phi^i)\frac{\delta}{\delta\phi^i},$$

we may write

$$\begin{aligned} K[i, i-1] &= \int \exp\left\{-i\Delta t\int dx_i\,\Lambda[x_i\phi^i]\, z_i(x_i)\right\} \\ &\quad\times \exp\left\{i\Delta t\int\left[\frac{\phi^i(x) - \phi^{i-1}(x)}{\Delta t}\right] z_i(x)\,dx\right\} d[z_i(x)] \\ &= \int \exp\left\{i\Delta t\int dx\, z_i(x)\right. \\ &\quad\left.\times\frac{\phi^i(x) - \phi^{i-1}(x)}{\Delta t} - \Lambda[x\phi^i]\right\} d[z_i(x)]. \end{aligned} \tag{76}$$

Following Rosen,[8] we introduce continuous time-dependent functions $\phi(x, t)$, $z(x, t)$, such that $\phi(x, t_i) = \phi^i(x)$, $z(xt_i) = z_i(x)$. Then in the limit $N\to\infty$, Eq. (74) becomes the conditional path integral[8,20]

$$K[2, 1] = \iint\limits_{\substack{\phi(xt_2)=\phi_2\\ \phi(xt_1)=\phi_1}} d[z(xt)]\, d[\phi(xt)] \times \exp\left\{i\int_{t_1}^{t_2} dt\int dx\, z(x, t)\left(\frac{\partial\phi}{\partial t} - \Lambda[x, \phi]\right)\right\}. \tag{77}$$

The restrictions $\phi(xt_2) = \phi_2$, $\phi(xt_1) = \phi_1$ may be removed by introducing delta functionals[21]:

$$\begin{aligned} K[2, 1] &= \iint d[z(x, t)]\, d[\phi(xt)] \\ &\quad\times \exp\left\{i\int_{t_1}^{t_2} dt\, dx\, z(xt)\left(\frac{\partial\phi}{\partial t} - \Lambda[x\phi]\right)\right\} \\ &\quad\times \Delta[\phi_1 - \phi(xt_1)]\Delta[\phi_2 - \phi(xt_2)] \\ &= \iint d[z(xt)]\, d[\phi(xt)] \\ &\quad\times \exp\left\{i\int_{t_1}^{t_2} dt\, dx\, z(x, t)\left(\frac{\partial\phi}{\partial t} - \Lambda[x\phi]\right)\right\} \\ &\quad\times \exp\left\{-i\int z(xt_1)[\phi_1 - \phi(xt_1)]\,dx\right\} \\ &\quad\times \exp\left\{i\int z(xt_2)[\phi_2 - \phi(xt_2)]\,dx\right\}, \end{aligned} \tag{78}$$

where there are now no restrictions on the values of ϕ at t_1 and t_2. The $z(xt)$ integration can be performed immediately to obtain

$$K[2, 1] = \int d[\phi(xt)]\Delta\left(\frac{\partial\phi(xt)}{\partial t} - \Lambda[x, \phi]\right) \times \Delta[\phi_1 - \phi(xt_1)]\Delta[\phi_2 - \phi(xt_2)], \tag{79}$$

where Δ is the delta functional for functions $\phi(x, t)$ of x and t with $t_1 \le t \le t_2$.

The delta functional Δ in Eq. (79) shows that contributions to $K[2, 1]$ come only from exact paths $\partial\hat{\phi}/\partial t = \Lambda[\hat{\phi}]$. We therefore expand the delta functional[22] into a sum over exact trajectories

$$\Delta\left[\frac{\partial\phi}{\partial t} - \Lambda[\phi]\right] = \int \Delta[\phi - \hat{\phi}]N[\hat{\phi}]\, d[\hat{\phi}], \tag{80}$$

where the integral is now over a "spacelike" surface $\hat{\phi}[x, t(x)]$, (since a solution is determined by one point on the trajectory) and $N[\hat{\phi}]$ is to be determined. Substitution of Eq. (80) into Eq. (79) then yields the result

$$K[2, 1] = \int d[\hat{\phi}]N[\hat{\phi}]\Delta[\phi_1 - \hat{\phi}(x, t_1)]\Delta[\phi_2 - \hat{\phi}(x, t_2)] \tag{81}$$

and, since we may take the integral over $\hat{\phi}$ to be that over $\hat{\phi}(xt_1)$, say, we have

$$\begin{aligned} K[2, 1] &= N[\phi_1]\Delta[\phi_2 - \hat{\phi}(x, t_2)]\big|_{\hat{\phi}(xt_1)=\phi_1} \\ &= N[\phi_1]\Delta\left[\phi_2 - \phi_1 - \int_{t_1}^{t_2}\Lambda[\hat{\phi}_1(\tau)]\,d\tau\right]. \end{aligned} \tag{82}$$

When $t_2 = t_1$, $K[2, 1] = \Delta[\phi_2 - \phi_1]$, so $N[\phi_1] = 1$ and

$$K[2, 1] = e^{-(t_2-t_1)L_2}\Delta[\phi_2 - \phi_1]$$

which is just Eq. (69). As a byproduct we have the identity

$$\Delta\left[\frac{\partial\phi}{\partial t} - \Delta[\phi]\right] = \int \Delta[\phi - \hat{\phi}]\, d[\hat{\phi}], \tag{83}$$

which shows that all solutions $\hat{\phi}$ contribute the same weight. If $\phi(xt)$ is a solution $\hat{\phi}'$ say, Eq. (83) states that

$$\Delta[0] = \int \Delta[\hat{\phi}' - \hat{\phi}]\, d[\hat{\phi}], \tag{84}$$

which, at first glance, appears to yield unity on the rhs. However, this is not the case, since

$$\int d[\phi(xt)]\Delta[\phi(xt) - \phi_1] = 1$$

only when the integral goes over *all* space-time paths.

[20] J. Tarski, *Lectures in Theoretical Physics, Vol. X*, W. E. Brittin, Ed. (Gordon and Breach Science Publishers, New York, 1967).

[21] Some of these expressions appear strange at first glance. However, they can be demonstrated by dividing the interval (t_1, t_2) into small segments to obtain suitable approximations.

[22] Analogous to the expansion $\delta[f(x)] = \sum_i \delta(x - x_i)/|f'(x_i)|$, $f(x_i) = 0$.

The propagator $K[2, 1]$ may be Fourier transformed to yield the propagator $\tilde{K}[2, 1]$ for the characteristic functional $G[\psi]$:

$$G[\psi_2 t_2] = \int \tilde{K}[2, 1] G[\psi_1 t_1]\, d[\psi_1], \tag{85}$$

where

$$\tilde{K}[\psi_2 \psi_1 t_2 t_1] \equiv \tilde{K}[2, 1] = \iint e^{-i(\psi_2, \phi_2)} K[2, 1] e^{i(\psi_1, \phi_1)}\, d[\phi_1]\, d[\phi_2]. \tag{86}$$

The expression (76) for $K[2, 1]$ is then substituted into Eq. (86) to yield (after integration by parts of the $z\partial\phi/\partial t$ term):

$$\begin{aligned} K[2, 1] = \iint & d[z(xt)]\, d[\phi(x, t)]\, d[\phi_1]\, d[\phi_2] \\ & \times \exp[-i(\psi_2, \phi_2)] \\ & \times \exp\left\{-i \int_{t_1}^{t_2} \left[z(xt) \Lambda[x\phi] + \frac{\partial z}{\partial t} \phi(xt) \right] dx\, dt \right. \\ & \left. + i \int [z(xt_2)\phi(xt_2) - z(xt_1)\phi(xt_1)]\, dx \right\} \\ & \times \exp\left\{-i \int z(xt_1)[\phi_1(x) - \phi(xt_1)]\, dx\right\} \\ & \times \exp\left\{i \int z(xt_2)[\phi_2(x) - \phi(xt_2)]\, dx\right\} \\ & \times \exp[i(\psi_1, \phi_1)], \end{aligned} \tag{87}$$

$$\begin{aligned} \tilde{K}[2, 1] = \iint & d[z(xt)]\, d[\phi(xt)] \\ & \times \Delta[z(xt_2) - \psi_2(x)] \Delta[z(xt_1) - \psi_1(x)] \\ & \times \exp\left\{-i \iint_{t_1}^{t_2} dt\, dx \left\{ z(xt) \Lambda[x\phi] \right.\right. \\ & \left.\left. + \frac{\partial z(xt)}{\partial t} \phi(xt) \right\}\right\} \end{aligned} \tag{88}$$

or

$$\begin{aligned} \tilde{K}[2, 1] = & \iint_{\substack{z(xt_2)=\psi_2(x) \\ z(xt_1)=\psi_1(x)}} d[z(xt)]\, d[\phi(xt)] \\ & \times \exp\left\{-i \iint_{t_1}^{t_2} dt\, dx \left\{ z(xt) \Lambda[x\phi] + \frac{\partial z}{\partial t} \phi(xt) \right\}\right\}. \end{aligned} \tag{89}$$

The above conditional path integral for $\tilde{K}[2, 1]$ has been obtained previously by Rosen[8] using directly the procedure employed above for $K[2, 1]$. This form (89) for $\tilde{K}[2, 1]$ is particularly useful as Rosen has pointed out, if Λ is quadratic in $\phi(xt)$. In that case, the integration over ϕ can be performed. This is the case, for example, for the Navier–Stokes equation[8] and the Vlasov equation.

The formal solution Eq. (89) may be used for the evaluation of moments. We have, from Eq. (59),

$$\begin{aligned} & \langle \phi(x_1)^{n_1} \phi(x_2)^{n_2} \cdots \phi(x_r)^{n_r} \rangle_{t_2} \\ & = i^{|n|} \frac{\delta^{|n|}}{\delta\psi_2(x_1)^{n_1} \cdots \delta\psi_2(x_r)^{n_r}} \\ & \quad \times \int \tilde{K}(2, 1) G[\psi_1 t_1]\, d[\psi_1]\big|_{\psi_2=0} \end{aligned} \tag{90}$$

$$\begin{aligned} & = i^{|n|}(-1)^{|n|} \iint d\{z\}\, d\{\phi\}\, \frac{\delta^{|n|}}{\delta z(x_1 t_2)^{n_1} \cdots \delta z(x_r t_2)^{n_r}} \\ & \quad \times \Delta[z(xt_2) - \psi_2(x)] \\ & \quad \times \Delta[z(xt_1) - \psi_1(x)] \\ & \quad \times \exp\left\{-i \iint_{t_1}^{t_2} dt\, dx \left\{ z(xt) \Lambda[x\phi] + \frac{\partial z(xt)}{\partial t} \phi(xt) \right\}\right\} \\ & \quad \times G[\psi_1 t_1]\, d[\psi_1] \end{aligned} \tag{91}$$

$$\begin{aligned} & = \iint_{z(xt_2)=0} d\{z\}\, d\{\phi\} \phi(x_1 t_2)^{n_1} \cdots \phi(x_r t_2)^{n_r} \\ & \quad \times \exp\left\{-i \iint_{t_1}^{t_2} dt\, dx \left\{ z(xt) \Lambda[x\phi] + \frac{\partial z(xt)}{\partial t} \phi(xt) \right\}\right\} \\ & \quad \times G[z(xt_1), t_1]. \end{aligned} \tag{92}$$

Equation (92) generalizes to all orders, the results obtained by Rosen[8] for $\langle \phi(x_2) \rangle$ and $\langle \phi(x_1)\phi(x_2) \rangle$.

V. PARTICLES AS FIELDS IN PHASE SPACE

We now consider a system of N particles interacting with the field ϕ and introduce the microscopic phase-space density[23–27] $\hat{f}(\mathbf{r}, \mathbf{p}, t)$ defined by

$$\hat{f}(\mathbf{r}, \mathbf{p}, t) = \sum_{i=1}^{N} \delta(\mathbf{r} - \hat{\mathbf{r}}_i(t)) \delta(\mathbf{p} - \hat{\mathbf{p}}_i(t)). \tag{93}$$

The coupled equations of motion for the particle-field system may be taken to be

$$\frac{\partial \hat{f}(x, t)}{\partial t} = \Lambda_1[x, \hat{\phi}, \hat{f}] \tag{94}$$

and

$$\frac{\partial \hat{\phi}(\mathbf{r}, t)}{\partial t} = \Lambda_2[\mathbf{r}, \hat{\phi}, \hat{f}], \tag{95}$$

where x stands for the phase point $\mathbf{r}, \mathbf{p}$.

[23] Yu. L. Klimontovich, Zh. Eksp. Teor. Fiz. **33**, 982 (1957) [Sov. Phys.—JETP **6**, 753 (1958)].
[24] E. P. Gross, J. Nucl. Energy **C2**, 173 (1961).
[25] J. Dawson and T. Nakayama, Phys. Fluids **9**, 1881 (1966).
[26] W. R. Chappell, J. Math. Phys. **8**, 553 (1967).
[27] C. S. Wu, *Lectures in Theoretical Physics, Vol. IX*, W. E. Brittin, Ed. (Gordon and Breach Science Publishers, New York, 1967).

The statistical-mechanical treatment of this system may be accomplished by the introduction of a distribution functional $F[\phi, f, t]$ for the pair of "fields" $\phi(\mathbf{r})$ and $f(\mathbf{r}, \mathbf{p})$. The Liouville equation for F can be written

$$\frac{\partial F}{\partial t} + \int dx\, \Lambda_1[x, \phi, f] \frac{\partial F}{\partial f(x)} + \int d\mathbf{r}\, \Lambda_2[\mathbf{r}, \phi, f] \frac{\partial F}{\partial \phi(\mathbf{r})} = 0, \quad (96)$$

provided that

$$\int dx \frac{\delta}{\delta f(x)} \Lambda_1(x, \phi, f) + \int \frac{d\mathbf{r} \delta \Lambda_2[\mathbf{r}, \phi, f]}{\delta \phi(\mathbf{r})} = 0. \quad (97)$$

Just as in the cases previously treated, we may introduce the functional Fourier transform $G[\phi_1, f_1]$ in terms of which the various moments may be expressed, i.e.,

$$f_r(x_1 x_2 \cdots x_r t) \equiv \langle f(x_1) f(x_2) \cdots f(x_r) \rangle = i^r \frac{\delta^r}{\delta f_1(x_1) \cdots \delta f_1(x_r)} G|_{f_1=0, \phi_1=0}, \quad (98)$$

$$\langle \phi(\mathbf{r}_1) \phi(\mathbf{r}_2) \phi(\mathbf{r}_s) \rangle = i^s \frac{\delta^s}{\delta \phi_1(\mathbf{r}_1) \cdots \delta \phi_1(\mathbf{r}_s)} G|_{f_1=0, \phi_1=0}, \quad (99)$$

$$\langle \phi(\mathbf{r}_1) f(x_1) \phi(\mathbf{r}_2) \cdots \phi(\mathbf{r}_s) f(x_s) \rangle = i^{2s} \frac{\delta^{2s}}{\delta \phi_1(\mathbf{r}_1) \delta f_1(x_1) \cdots \delta \phi_1(\mathbf{r}_s) \delta f(x_s)} G|_{f_1=0, \phi_1=0},$$

etc. The quantity $f_r(x_1, \cdots, x_r t)$ is a reduced r-particle phase-space distribution function. However, f_r includes self correlations which can be eliminated by introducing the usual r-particle distribution functions $\tilde{f}_r$ through[10]

$$\begin{aligned} \tilde{f}_r(x_1 \cdots x_r t) &= i \frac{\delta}{\delta f_1(x_1)} \left(i \frac{\delta}{\delta f_1(x_1)} - \delta(1, 2) \right) \\ &\times \left(i \frac{\delta}{\delta f_1(x_3)} - \delta(1, 3) - \delta(2, 3) \right) \\ &\cdots \left(i \frac{\delta}{\delta f_1(x_r)} - \delta(1, r) \right. \\ &\left. - \delta(2, r) - \cdots - \delta(r-1, r) \right) G|_{f_1=0, \phi_1=0}. \end{aligned} \quad (100)$$

Correlation functions may be introduced through their generating functional H, i.e.,

$$G = e^H \quad (101)$$

which lead to Mayer-type cluster expansions.

VI. THE FREE ELECTROMAGNETIC FIELD

Maxwell's equations for the free electromagnetic field are:

$$\nabla \times \mathbf{B} - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} = 0, \quad \nabla \cdot \mathbf{B} = 0,$$

$$\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} = 0, \quad \nabla \cdot \mathbf{E} = 0. \quad (102)$$

Therefore, the distribution functional $F[\mathbf{E}, \mathbf{B}, t]$ for the field satisfies the equation of motion

$$\frac{1}{c} \frac{\partial F}{\partial t} + \int d\mathbf{r} \left\{ \nabla \times \mathbf{B} \cdot \frac{\delta F}{\delta \mathbf{E}} - \nabla \times \mathbf{E} \cdot \frac{\delta F}{\delta \mathbf{B}} \right\} = 0, \quad (103)$$

while G, the functional Fourier transform, satisfies the equation

$$\frac{1}{c} \frac{\partial G}{\partial t} + \int d\mathbf{r} \left\{ \nabla \times \mathbf{E}_1 \cdot \frac{\delta G}{\delta \mathbf{B}_1} - \nabla \times \mathbf{B}_1 \cdot \frac{\delta G}{\delta \mathbf{E}_1} = 0 \right. \quad (104)$$

which is exactly the same as Eq. (103) with $\mathbf{B} \to \mathbf{E}_1$ and $\mathbf{E} \to \mathbf{B}_1$. Since the initial-value problem for the free electromagnetic field can be solved, Eqs. (103) and (104) can be solved. To this end we introduce in Eq. (104) the complex field $\mathbf{F}_1 = \mathbf{B}_1 + i\mathbf{E}_1$, $\mathbf{F}_1^* = \mathbf{B}_1 - i\mathbf{E}_1$. Then Eq. (104) becomes

$$\frac{1}{c} \frac{\partial G}{\partial t} - i \int d\mathbf{r} \left\{ \nabla \times \mathbf{F}_1 \cdot \frac{\delta G}{\delta \mathbf{F}_1} - \nabla \times \mathbf{F}_1^* \cdot \frac{\delta G}{\delta \mathbf{F}_1^*} \right\} = 0. \quad (105)$$

Equation (105) may be solved directly[16] to yield

$$\begin{aligned} G[\mathbf{F}_1, \mathbf{F}_1^*, t] &= \exp \left\{ ict \int d\mathbf{r} \left[\nabla \times \mathbf{F}_1 \cdot \frac{\delta}{\delta \mathbf{F}_1} \right. \right. \\ &\left. \left. - \nabla \times \mathbf{F}_1^* \cdot \frac{\delta}{\delta \mathbf{F}_1^*} \right] \right\} G[\mathbf{F}_1, \mathbf{F}_1^*, 0] \\ &= G[e^{ict\nabla \times} \mathbf{F}_1, e^{-ict\nabla \times} \mathbf{F}_1^*, 0]. \end{aligned} \quad (106)$$

The solution (106) corresponds to the fact that the characteristic equation for $\hat{\mathbf{F}}_1$ is $\partial \hat{\mathbf{F}}_1 / \partial t = -ic \nabla \times \hat{\mathbf{F}}_1$, which has the solution $\hat{\mathbf{F}}_1(t) = e^{-ict\nabla \times} \hat{\mathbf{F}}_1(0)$. Thus, Eq. (106) is an example of the general result (11).

The subsidiary conditions $(\nabla \cdot \mathbf{E})F = (\nabla \cdot \mathbf{B})F = 0$ mean that we may consider the arguments $\mathbf{E}$, $\mathbf{B}$ of $F[\mathbf{E}, \mathbf{B}, t]$ as being transverse. Further, since $\int \mathbf{E}_1(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r})\, d\mathbf{r}$ projects out the transverse part of $\mathbf{E}_1(\mathbf{r})$, we may consider $G[\mathbf{E}_1(\mathbf{r}), \mathbf{B}_1(\mathbf{r}), t]$ to be a functional of the transverse parts of $\mathbf{E}_1$, $\mathbf{B}_1$. That is, we may consider as arguments of G, only those $\mathbf{E}_1$, $\mathbf{B}_1$ for which $\nabla \cdot \mathbf{E}_1 = \nabla \cdot \mathbf{B}_1 = 0$.

The solution $\hat{\mathbf{F}}_1(t) = e^{-ict\nabla\times}\hat{\mathbf{F}}_1(0)$ may be expressed simply in terms of the Pauli "D" function:

$$\hat{\mathbf{F}}_1(\mathbf{r}, t) = \left(\frac{1}{c}\frac{\partial}{\partial t} - i\nabla \times \right)\int D\{\mathbf{r} - \mathbf{r}', t\}\hat{\mathbf{F}}_1(\mathbf{r}', 0)\, d\mathbf{r}', \tag{107}$$

with[28]

$$D(\mathbf{r}, t) \equiv \frac{1}{(2\pi)^3}\int \exp(i\mathbf{k}\cdot\mathbf{r})\frac{\sin(c|\mathbf{k}|t)}{|\mathbf{k}|}\, d\mathbf{k}, \tag{108}$$

and $\nabla\cdot\hat{\mathbf{F}}_1(\mathbf{r}, 0) = 0$. The function D may also be expressed as $D(\mathbf{r}, t) = (4\pi r)^{-1}\{\delta(ct - r) - \delta(ct + r)\}$, which shows that the fields $\mathbf{E}_1$, $\mathbf{B}_1$ propagate with velocity c and then are mixed:

$$\hat{\mathbf{B}}_1(t) = \frac{1}{c}\frac{\partial}{\partial t} D_t\hat{\mathbf{B}}_1(0) + \nabla \times D_t\hat{\mathbf{E}}_1(0), \tag{109}$$

$$\hat{\mathbf{E}}_1(t) = \frac{1}{c}\frac{\partial}{\partial t} D_t\hat{\mathbf{E}}_1(0) - \nabla \times D_t\hat{\mathbf{B}}_1(0), \tag{110}$$

where $D_tF(\mathbf{r}) \equiv \int D(\mathbf{r} - \mathbf{r}', t)F(\mathbf{r}')\, d\mathbf{r}'$. Equations (109) and (110) correspond to the propagation of the free electromagnetic field expressed by

$$(\mathbf{E}(\mathbf{r}, t) + i\mathbf{B}(\mathbf{r}, t)) = \left(\frac{1}{c}\frac{\partial}{\partial t} - i\nabla \times \right)\int D(\mathbf{r} - \mathbf{r}', t)(\mathbf{E}_0(\mathbf{r}') + i\mathbf{B}_0(\mathbf{r}'))\, d\mathbf{r}' \tag{111}$$

or

$$\mathbf{E}(t) = \frac{1}{c}\frac{\partial}{\partial t} D_t\mathbf{E}_0 + \nabla \times D_t\mathbf{B}_0$$

and

$$\mathbf{B}(t) = \frac{1}{c}\frac{\partial}{\partial t} D_t\mathbf{B}_0 - \nabla \times D_t\mathbf{E}_0. \tag{112}$$

The functional $G[\mathbf{E}_1, \mathbf{B}_1 t]$ thus may be written

$$G[\mathbf{E}_1, \mathbf{B}_1, t] = G\left[-\frac{1}{c}\frac{\partial}{\partial t} D_{-t}\mathbf{E}_1 - \nabla \times D_{-t}\mathbf{B}_1, -\frac{1}{c}\frac{\partial}{\partial t} D_{-t}\mathbf{B}_1 + \nabla \times D_{-t}\mathbf{E}_1, 0\right], \tag{113}$$

$$G[\mathbf{E}_1, \mathbf{B}_1, t] = G\left[\frac{1}{c}\frac{\partial}{\partial t} D_t\mathbf{E}_1 + \nabla \times D_t\mathbf{B}_1, \frac{1}{c}\frac{\partial}{\partial t} D_t\mathbf{B}_1 - \nabla \times D_t\mathbf{E}_1, 0\right], \tag{114}$$

since $D_t = -D_{-t}$. The functional in Eq. (114) represents the most general statistical situation for the free (classical) electromagnetic field. From it one may directly obtain expressions for the correlations

$$\mathbf{E}(\mathbf{r}_1)\mathbf{E}(\mathbf{r}_2)\cdots\mathbf{E}(\mathbf{r}_k)\mathbf{B}(\mathbf{r}_{k+1})\cdots\mathbf{B}(\mathbf{r}_l)\rangle_t$$

in terms of

$$\langle\mathbf{E}(\mathbf{r}_1')\mathbf{E}(\mathbf{r}_2')\cdots\mathbf{E}(\mathbf{r}_s')\mathbf{B}(\mathbf{r}_{s+1}')\cdots\mathbf{B}(\mathbf{r}')\rangle_0, \quad s = 0, \cdots l,$$

since the correlations are functional derivatives of G evaluated at $\mathbf{E}_1 = \mathbf{B}_1 = 0$ [cf. Eq. (59)]. For example,

$$\langle\mathbf{E}(\mathbf{r})\rangle_t = i\frac{\delta G}{\delta\mathbf{E}_1(\mathbf{r})}\bigg|_{\mathbf{E}_1,\mathbf{B}_1=0} = \frac{1}{c}\frac{\partial}{\partial t} D_t\langle\mathbf{E}\rangle_0 + \nabla \times D_t\langle\mathbf{B}\rangle_0 \tag{115}$$

and

$$\langle\mathbf{B}(\mathbf{r})\rangle_t = i\frac{\delta G}{\delta\mathbf{B}_1(\mathbf{r})}\bigg|_0 = \frac{1}{c}\frac{\partial}{\partial t} D_t\langle\mathbf{B}\rangle_0 - \nabla \times D_t\langle\mathbf{E}\rangle_0 \tag{116}$$

as was anticipated. Another simple example is

$$\langle\mathbf{E}(\mathbf{r}_1)\mathbf{B}(\mathbf{r}_2)\rangle_t = i^2\frac{\delta^2 G}{\delta\mathbf{E}_1(\mathbf{r}_1)\delta\mathbf{B}(\mathbf{r}_2)}\bigg|_0 \tag{117}$$

or

$$\begin{aligned}\langle\mathbf{E}(\mathbf{r}_1)\mathbf{B}(\mathbf{r}_2)\rangle_t = {} & \frac{1}{c^2}\frac{\partial}{\partial t} D_t(1)\frac{\partial D_t(2)}{\partial t}\langle\mathbf{E}_0(\mathbf{r}_1)\mathbf{E}_0(\mathbf{r}_2)\rangle \\ & + \frac{1}{c}\nabla_1 \times D_t(1)\frac{\partial D_t(2)}{\partial t}\langle\mathbf{B}_0(\mathbf{r}_1)\mathbf{B}_0(\mathbf{r}_2)\rangle \\ & - \frac{1}{c}\frac{\partial}{\partial t} D_t(1)\nabla_2 \times D_t(2)\langle\mathbf{E}_0(\mathbf{r}_1)\mathbf{B}_0(\mathbf{r}_2)\rangle \\ & - \nabla_1 \times D_t(1)\nabla_2 \times D_t(2)\langle\mathbf{B}_0(\mathbf{r}_1)\mathbf{E}_0(\mathbf{r}_2)\rangle,\end{aligned} \tag{118}$$

using an obvious notation. Although the results shown here are quite simple and could have been obtained by other methods, we have included them in order to illustrate the power and generality of our methods.

VII. STATISTICAL MECHANICAL BASIS OF PLASMA PHYSICS

We consider a system composed of N particles each having charge e and mass m[29] interacting via the electromagnetic field. The system is described in terms of the electromagnetic field $\mathbf{E}$, $\mathbf{B}$ and the exact classical microscopic phase-space density function f. The quantities $\hat{\mathbf{E}}$, $\hat{\mathbf{B}}$, $\hat{f}$ are coupled through the exact microscopic Maxwell–Klimontovich equations[16,23–26]:

$$\frac{1}{c}\frac{\partial\hat{\mathbf{E}}}{\partial t} = +\nabla \times \hat{\mathbf{B}} - \frac{4\pi e}{mc}\int\mathbf{p}\hat{f}(\mathbf{r}, \mathbf{p})\, d\mathbf{p}, \tag{119}$$

$$\frac{1}{c}\frac{\partial\hat{\mathbf{B}}}{\partial t} = -\nabla \times \hat{\mathbf{E}}, \tag{120}$$

$$\frac{\partial\hat{f}}{\partial t} = -\frac{\mathbf{p}}{m}\cdot\nabla\hat{f} - e\left(\hat{\mathbf{E}} + \frac{\mathbf{p}}{mc} \times \hat{\mathbf{B}}\right)\cdot\frac{\partial\hat{f}}{\partial\mathbf{p}}, \tag{121}$$

$$\nabla\cdot\hat{\mathbf{E}} = 4\pi e\int\hat{f}(\mathbf{r}, \mathbf{p})\, d\mathbf{p}, \tag{122}$$

[28] S. T. Ma, Phys. Rev. **68**, 166 (1945).

[29] For simplicity we consider only one species, but the multicomponent system is easily handled.

and

$$\nabla \cdot \hat{\mathbf{B}} = 0. \tag{123}$$

It must be stressed that $\hat{f}$ described in the above system of equations must be considered as an implicit function of the exact positions and velocities ($\mathbf{p}_i \equiv m\mathbf{v}_i$) of all the particles, as well as a function of $\mathbf{r}$, $\mathbf{p}$:

$$f(\mathbf{r}, \mathbf{p}) = \sum_{k=1}^{N} \delta(\mathbf{r} - \hat{\mathbf{r}}_k)\delta(\mathbf{p} - \hat{\mathbf{p}}_k) \tag{124}$$

and, further, that Eqs. (119)–(123) are equivalent to the usual microscopic Maxwell–Lorentz equations. The point of view which we now adopt is that these equations form a closed system of coupled field equations for the three fields $\mathbf{E}$, $\mathbf{B}$, f. Then the general theory as outlined in Sec. 5 can be applied directly. We introduce the distribution functional $F[\mathbf{E}, \mathbf{B}, f, t]$ for the fields $\mathbf{E}(\mathbf{r})$, $\mathbf{B}(\mathbf{r})$, $f(\mathbf{r}, \mathbf{p})$. The functional F satisfies the following equation:

$$\frac{\partial F}{\partial t} + \int d\mathbf{r}\left\{\left[c\nabla \times \mathbf{B}(\mathbf{r}) - \frac{4\pi e}{m}\int d\mathbf{p}\, \mathbf{p} f(\mathbf{r}, \mathbf{p})\right] \cdot \frac{\partial F}{\partial \mathbf{E}(\mathbf{r})} - c\nabla \times \mathbf{E}(\mathbf{r}) \cdot \frac{\delta F}{\delta \mathbf{B}(\mathbf{r})}\right\} + \int d\mathbf{r}\, d\mathbf{p}\left\{\left[-\frac{\mathbf{p}}{m} \cdot \nabla f(\mathbf{r}, \mathbf{p})\right] - e\left[\mathbf{E}(\mathbf{r}) + \frac{\mathbf{p}}{mc} \times \mathbf{B}(\mathbf{r})\right] \cdot \frac{\partial f(\mathbf{r}, \mathbf{p})}{\partial \mathbf{p}}\right\} \frac{\partial F}{\partial f(\mathbf{r}, \mathbf{p})} = 0. \tag{125}$$

In addition, it must satisfy the subsidiary conditions,

$$\left\{\nabla \cdot \mathbf{E}(\mathbf{r}) - 4\pi e \int f(\mathbf{r}, \mathbf{p})\, d\mathbf{p}\right\} F = 0, \tag{126}$$

and

$$\{\nabla \cdot \mathbf{B}(\mathbf{r})\} F = 0. \tag{127}$$

Our goal is to obtain coupled equations of motion for various particle distribution functions

$$\langle f(x_1) \cdots f(x_r)\rangle \equiv f_r(x_1, \cdots, x_r)$$

and field-particle correlation functions

$$\langle \mathbf{E}(\mathbf{r}_1) \cdots \mathbf{E}(\mathbf{r}_s)\mathbf{B}(\mathbf{r}'_1) \cdots \mathbf{B}(\mathbf{r}'_t) f(x_1) \cdots f(x_r)\rangle.$$

(Actually, we are interested in the particle distributions $\tilde{f}_r$ which do not contain self-correlations, but, as was mentioned in Sec. V, a simple transformation on F allows the $\tilde{f}_r$ to be found.)

The functional Fourier transform $G[\mathbf{E}_1, \mathbf{B}_1, f_1]$ is introduced through

$$F[\mathbf{E}, \mathbf{B}, f] = \int \exp\{i[(\mathbf{E}, \mathbf{E}_1) + (\mathbf{B}, \mathbf{B}_1) + (f, f_1)]\} \times G[\mathbf{E}_1, \mathbf{B}_1, f_1]\, d[\mathbf{E}_1]\, d[\mathbf{B}_1]\, d[f_1]. \tag{128}$$

The equation of motion for G is found directly from Eq. (125) and is expressed by

$$\frac{\partial G}{\partial t} + \int d\mathbf{r}\left\{c\nabla \times \mathbf{E}_1 \cdot \frac{\delta G}{\delta \mathbf{B}_1} + c\nabla \times \mathbf{B}_1 \cdot \frac{\delta G}{\delta \mathbf{E}_1} + \frac{4\pi e}{m}\int d\mathbf{p}\, \mathbf{p} \cdot \mathbf{E}_1 \frac{\delta G}{\delta f_1}\right\} + \int d\mathbf{r}\, d\mathbf{p}\left\{-\frac{\mathbf{p}}{m} \cdot \nabla f_1 \frac{\delta G}{\delta f_1} - i \exp\left(\frac{\partial f_1}{\partial \mathbf{p}}\right) \cdot \left[\frac{\delta}{\delta \mathbf{E}_1} + \frac{\mathbf{p}}{mc} \times \frac{\delta}{\delta \mathbf{B}_1}\right] \frac{\delta G}{\delta f_1} = 0. \tag{129}$$

In addition, G must satisfy the subsidiary conditions

$$\left\{\nabla \cdot \frac{\delta}{\delta \mathbf{E}_1} - 4\pi e \int d\mathbf{p}\, \frac{\delta}{\delta f_1}\right\} G = 0 \tag{130}$$

and

$$\nabla \cdot \frac{\delta}{\delta \mathbf{B}_1} G = 0. \tag{131}$$

Equations (129)–(131) serve as a basis for the general statistical theory of particles interacting with the electromagnetic field. The resulting moment equations form a hierarchy similar to the usual BBKGY hierarchy of statistical mechanics, and are equivalent to the Maxwell–Lorentz theory with lack of information built in.[12,16,17] These equations differ from the customary equations in that the particle distribution functions contain self-correlations. That is,

$$f_2(x_1x_2) = \langle f(x_1)f(x_2)\rangle = i^2 \frac{\delta^2 G}{\delta f_1(x_1)\delta f_1(x_2)}\bigg|_{f_1=0} = \tilde{f}_2 + \delta(1, 2)\tilde{f}_1,$$

where $\tilde{f}_2$ has no self-correlation.

We have shown that these equations generate the usual hierarchy arising from the Liouville equation. However, our point of view is that these functional equations should be considered as the basic starting point for the statistical theory of charged particles interacting via the electromagnetic field. Especially for situations such as those arising in turbulence, where the familiar truncation procedures are not applicable, we may expect methods based on the use of the functional equations to lead to new results. For example, the method of Lewis,[30] which can be used to obtain closed equations for reduced distribution functions, can be applied directly to the functional Liouville equation. In this case the entropy is a functional of the distribution *functional* F: $S[F] = -k \int d[F] F \ln F$, which is to be maximized subject to constraints which require F to be a functional of those field quantities required for the macroscopic description of the system. For example, one might want to

[30] R. M. Lewis, J. Math. Phys. **8**, 1448 (1967).

describe the macroscopic behavior of the system in term of $\langle f(\mathbf{r}, \mathbf{p})\rangle$, $\langle \mathbf{E}(\mathbf{r})\rangle$, $\langle \mathbf{B}(\mathbf{r})\rangle$, and $\langle f(\mathbf{r}, \mathbf{p})\mathbf{E}(\mathbf{r})\rangle$. The results of this line of investigation will be published later.

APPENDIX A

Remarks on Functionals

A functional may be considered as a mapping which assigns to the function $\phi(x)$, a number $F[\phi]$. The functions $\phi(x)$ usually lie in some function space, e.g., Banach space, etc. We assume that the argument functions $\phi(x)$ are elements of a real Hilbert-space $\mathcal{H}$ with the scalar product $(\phi, \psi) = \int_{E_n} \phi(x)\psi(x)\, dx$, where E_n is the underlying n-dimensional Euclidean space of the argument x of the function $\phi(x)$. The norm $\|\phi\|$ of the function $\phi(x)$ is defined by $\|\phi\|^2 = (\phi, \phi)$. With the notion of the size of a function being given by its norm, we may apply many of the standard ideas of analysis to functionals. For example, a functional is continuous at the argument function ϕ if, for a given $\epsilon > 0$, there exists a δ_ϵ such that $|F(\phi) - F(\psi)| < \epsilon$ for all ψ for which $\|\phi - \psi\| < \delta_\epsilon$. If a functional $F[t, \phi]$ is a function of t as well as a functional of $\phi(x)$, derivatives and integrals with respect to t may be performed in the usual way, e.g., $\partial F/\partial t \equiv D[t, \phi]$ is again a functional of ϕ and a function of t.

We now introduce the idea of differentiation of a functional with respect to the *function* $\phi(x)$. This notion parallels the idea of the gradient $\nabla F(\mathbf{x})$ of a scalar function with vector argument. We observe that for such a function

$$\frac{d}{d\lambda} F(\mathbf{x} + \lambda\mathbf{y})|_{\lambda=0} = \mathbf{y} \cdot \nabla F, \tag{A1}$$

so that ∇F appears as vector argument in the scalar product $\mathbf{y} \cdot \nabla F$. In analogy to the above development we can form from the functional $F[\phi]$ the quantity

$$\frac{d}{d\lambda} F[\phi + \lambda\psi]|_{\lambda=0} \tag{A2}$$

whenever it exists. For our purposes we shall assume that the expression in Eq. (A2) is a linear bounded functional[31,32] of the function $\psi(x)$. Then there exists a function $\delta F/\delta\phi(x)$ [called the *functional derivative* of F with respect to the function $\phi(x)$], such that

$$\frac{d}{d\lambda} F[\phi + \lambda\psi]|_{\lambda=0} = \left(\psi, \frac{\delta F}{\delta\phi}\right). \tag{A3}$$

It is to be noted that, in general, $\delta F/\delta\phi(x)$ is a functional of ϕ as well as an explicit function of x. If we replace $\psi(x)$ by the δ distribution $\delta(x - x_0)$, we obtain

$$\frac{d}{d\lambda} F[\phi + \lambda\psi]|_{\substack{\lambda=0 \\ \psi=\delta(x-x_0)}} = \frac{\delta F}{\delta\phi(x_0)}. \tag{A4}$$

However, one must exhibit care with such procedures since $\delta F/\delta\phi(x)$, as defined by Eq. (A3), is in general a distribution itself, and therefore its value at a given point x may not make sense. [For example, if $F[\phi(x)] = \phi(x)$ we have $\delta\phi(x)/\delta\phi(x') = \delta(x - x')$.] Higher-order derivatives are defined in a similar way:

$$\frac{d^n}{d\lambda^n} F[\phi + \lambda\psi]|_{\lambda=0} = \left(\psi^n, \frac{\delta^n F}{\delta\phi^n}\right)_n \tag{A5}$$

$$\equiv \int_{E_n} \cdots \int_{E_n} dx_1\, dx_n \psi(x_1) \cdots \psi(x_n) \frac{\delta^n F[\phi]}{\delta\phi(x_1) \cdots \delta\phi(x_n)}. \tag{A6}$$

If the nth derivatives exist, and if $F[\phi + \lambda\psi]$ is analytic in λ for $\lambda = 1$, we have Taylor's theorem for functionals:

$$F[\phi + \psi] = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{E_n} \cdots \int_{E_n} dx_1 \cdots dx_n \times \frac{\delta^n F}{\delta\phi(x_1) \cdots \delta\phi(x_n)} \psi(x_1)\psi(x_2) \cdots \psi(x_n). \tag{A7}$$

A functional may also be thought of as a function of an infinite number of variables,[33] for if one introduces a fixed basis $\phi_1, \phi_2, \cdots$ in the Hilbert space of the functions $\phi(x)$, we have

$$F[\phi(x)] = F\left[\sum_n a_n \phi_n(x)\right] \equiv F(a_1, a_2, \cdots). \tag{A8}$$

Further, since for an orthonormal basis

$$a_n = \int_{E_n} \phi_n(x)\phi(x)\, dx,$$

$$\frac{\delta F}{\delta\phi(x)} = \sum_n \frac{\delta a_n}{\delta\phi(x)} \frac{\partial F}{\partial a_n} = \sum_n \frac{\partial F}{\partial a_n} \phi_n(x). \tag{A9}$$

We note that $(\phi_n, \delta F/\delta\phi(x)) = \partial F/\partial a_n$, which gives further insight into the relationship between functional derivative and derivative (gradient). Another way of looking at the functional derivative consists in splitting up the space E_n into a number of tiny cells Δx_i, $i = 1, 2, 3, \cdots$ such that

$$E_n = \bigcup_{i=1}^{\infty} \Delta x_i$$

and

$$\Delta x_i \cap \Delta x_j = 0$$

[31] If the quantity in Eq. (A3) exists, it is referred to as the Gâteaux differential. If the Gâteaux differential is a bounded linear functional, it is called a Fréchet differential. See Ref. 32.

[32] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* (American Mathematical Society, Providence, R.I., 1957), p. 109.

[33] V. Volterra, *Theory of Functionals* (Blackie and Sons, Ltd., London, 1930).

for $i \neq j$. Then the function $\phi(x)$ may be "projected"

$$\phi(x) \sim \sum_{i=1}^{\infty} \phi_i E_i(x), \tag{A10}$$

where $E_i(x)$ is the characteristic function for the cell Δx_i and

$$\phi_i = \frac{1}{|\Delta x_i|} \int \phi(x) E_i(x)\, dx \tag{A11}$$

($|\Delta_{x_i}|$ is the volume of the cell Δx_i).

A functional $F[\phi]$ may now be considered to be approximately a function of the variables $\phi_1, \phi_2, \cdots$:

$$F[\phi] = F[\phi_1, \phi_2, \cdots]. \tag{A12}$$

We then have

$$\delta F \sim \sum_{i=1}^{\infty} \frac{\partial F}{\partial \phi_i}\, \delta\phi_i$$

or

$$\delta F \sim \sum_{i=1}^{\infty} \frac{1}{|\Delta x_i|} \frac{\partial F}{\partial \phi_i}\, \delta\phi_i\, |\Delta x_i|. \tag{A13}$$

If we now allow each $|\Delta x_i|$ to approach zero, $\delta\phi_i$ is assumed to approach a smooth function $\delta\phi(x)$ and the sum in Eq. (A13) becomes an integral

$$\delta F \sim \int \frac{\delta F[\phi]}{\delta \phi(x)}\, \delta\phi(x)\, dx, \tag{A14}$$

where $\delta F[\phi]/\delta\phi(x)$ is the limit

$$\lim_{|\Delta x_i| \to 0} \frac{1}{|\Delta x_i|} \frac{\partial F}{\partial \phi_i}.$$

Here the functional derivative appears as a "derivative per unit volume" and the argument x in $\delta F/\delta\phi(x)$ is that point upon which the volume Δx_i shrinks to zero.

The integration of functionals over the Hilbert space of functions $\phi(x)$ presents considerable mathematical difficulty so we shall be content to give a heuristic treatment of functional integration.[34] If we represent the functional $F[\phi]$ as a function of its real Fourier components $a_n = (\phi_n, \phi)$, we may define the integral $\int F[\phi]\, d[\phi]$ as the limit

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{da_1}{(2\pi)^{\frac{1}{2}}} \cdots \frac{da_n}{(2\pi)^{\frac{1}{2}}} F\left[\sum_n a_n \phi_n\right], \tag{A15}$$

if it exists. The factors $(2\pi)^{\frac{1}{2}}$ are introduced in order to avoid infinite normalization constants in functional Fourier transforms. There are other methods[35] of introducing functional integration, but the above procedure will be sufficient for our needs. The integral in Eq. (A15) is invariant under orthogonal changes in basis and under fixed translation,

$$\int F[\phi + \phi_1]\, d[\phi] = \int F[\phi]\, d[\phi]. \tag{A16}$$

The functional Fourier expansion may be carried out as follows:

$$\begin{aligned} F[\phi] &= F[\textstyle\sum a_n \phi_n] \\ &= \lim_{n \to \infty} \int \exp\left(i \sum_{i=1}^{n} a_n b_n\right) \\ &\qquad \times G[b_1 \cdots b_n] \frac{db_1}{(2\pi)^{\frac{1}{2}}} \cdots \frac{db_n}{(2\pi)^{\frac{1}{2}}} \\ &= \int e^{i(\phi, \phi_1)} G[\phi_1]\, d[\phi_1], \end{aligned} \tag{A17}$$

where

$$\phi_1(x) = \sum_{i=1}^{\infty} b_n \phi_n(x).$$

Since

$$\begin{aligned} &G[b_1 \cdots b_n] \\ &= \int \exp\left(-i \sum_{i=1}^{n} a_i b_i\right) F[\textstyle\sum a_n \phi_n] \frac{da_1}{(2\pi)^{\frac{1}{2}}} \cdots \frac{da_n}{(2\pi)^{\frac{1}{2}}}, \end{aligned} \tag{A18}$$

we have

$$G[\phi_1] = \int e^{-i(\phi_1, \phi)} F[\phi]\, d[\phi]. \tag{A19}$$

In particular, the delta functional $\Delta[\phi]$ appears as the limit

$$\begin{aligned} \Delta[\phi] &= \lim_{n \to \infty} \int \cdots \int \exp\left(i \sum_{i=1}^{n} a_i b_i\right) \frac{db_1}{(2\pi)^{\frac{1}{2}}} \cdots \frac{db_n}{(2\pi)^{\frac{1}{2}}} \\ &= \lim_{N \to \infty} \prod_{n=1}^{N} [(2\pi)^{\frac{1}{2}} \delta(a_n)] \\ &= \int e^{i(\phi, \phi_1)}\, d[\phi_1]. \end{aligned} \tag{A20}$$

[34] A rigorous treatment of functional integration with physical applications and an extensive bibliography is given in Ref. 20.

[35] F. A. Berezin, *The Method of Second Quantization* (Academic Press Inc., New York, 1966), pp. 37–41.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Functional Integrals Representing Distribution Functions in Statistical Mechanics

CLAS BLOMBERG
Division of Theoretical Physics, Royal Institute of Technology, Stockholm, Sweden

(Received 28 November 1967)

We show how to obtain formal solutions of the chain of equations for distribution functions in classical statistical mechanics. These solutions are in the form of complex functional integrals. They are not unique, which fact is a fundamental property of the equations, and the different solutions are recognized by different integration paths in the complex function space. The different manners of integration correspond to different phases, of which some can be identified with the possible physical states. The treatment of the integrals in some cases is also discussed. They are closely related to generalizations of the molecular field approach to the problem. It is also shown that the functional integrals can be written as averages over an external field and that essentially the same form is valid in the quantum-mechanical case.

1. INTRODUCTION

The problem of solving a many-body problem in statistical mechanics is so very complicated that it is worthwhile to try new mathematical methods for the problem. For this purpose, some authors have used functional integration.[1] By the use of such methods we get formal, closed expressions for the quantities of interest. They also give valuable information about simplifications and new approximation methods for the solution of the problem. The drawback of this method is, of course, that only some types of functional integrals can be treated with mathematical rigor. In this work, we treat the integrals as limits of ordinary many-dimensional integrals, and merely assume that this limit exists. We remark that, at the moment, no method exists that can rigorously treat a general many-body problem in the thermodynamic limit. The functional integrals have the advantage that they contain in a closed form the physical facts and at least give hints as to how to proceed to a better understanding of the problem.

Functional integrals have essentially been used to represent the partition function Z. As is well known, the structure of Z becomes extremely complicated as the number of particles in the system increases. The function does not exist in the thermodynamic limit, nor has it a simple asymptotic form. However, certain quantities, defined as quotients of functions of this kind, have simple asymptotic properties and can be well defined in the limit. It is also known that these limiting quantities are not unique analytic functions, but correspond to different analytic functions in different regions of its parameters (e.g., temperature, chemical potential).

For this reason, a method has been developed by the author,[2] in a paper which is referred to as I, where distribution functions with well-defined limits are represented by functional integrals. These expressions are in fact the solutions of a chain of equations, rewritten as functional differential equations by the use of generating functionals. These equations were first derived by Bogoliubov.[3] The mentioned nonuniqueness of the limiting procedure shows up here in a nonuniqueness of the solutions of the equations. We get, in fact, a very large number of solutions, each characterized by its mode of integration.[4] We briefly recall the results of I in Sec. 2, where the equations and the solutions are presented. We also discuss the solutions of a much simpler equation which, in fact, has much in common with the general one and which gives valuable information about the more complicated integrals. We further discuss these solutions in Sec. 3 and also show how we get the equations of a molecular field theory by integrating over a saddle point. The integrals were defined in I for regular interaction potentials, but in Sec. 4, we show how to treat cases with a hard-core potential, especially a lattice gas (where the hard core is necessary). As problems of these kinds are the only ones which can be exactly solved by any methods, they are valuable to

[1] Reviews of such attempts are found in: M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience Publishers, Inc., New York, 1950), Chap. 4; I. M. Gel'fand and A. M. Yaglom, Uspekhi Mat. Nauk **9**, 77 (1956); [English transl.: J. Math. Phys. **1**, 48 (1960)].

Among the works which use these or similar methods, the following can be mentioned: R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); S. F. Edwards and R. E. Peierls, Proc. Phys. Soc. (London) **A224**, 24 (1954); I. M. Gel'fand and R. A. Minlos, Dokl. Akad. Nauk (SSSR) **97**, 209 (1954); R. Scalettar, Ann. Phys. (N.Y.) **38**, 238 (1966); S. F. Edwards and D. Sherrington, Proc. Phys. Soc. (London) **90**, 3 (1967).

[2] C. Blomberg, Acta Polytechnica Scandinavica, Ph 49, 1967.

[3] N. N. Bogoliubov, Zh. Fiz. (SSSR) **19**, 256 (1946) [English transl.: *Studies in Statistical Mechanics I*, edited by Uhlenbeck de Boer (North-Holland Publishing Company, Amsterdam, 1962)].

[4] We note that similar results are the aim in quite another approach to the problem: the algebraic method for treating infinitely large many-body systems. This method is, for example, developed in: H. Araki and E. J. Woods, J. Math. Phys. **4**, 637 (1963); G. Emch and M. Guenin, J. Math. Phys. **7**, 915 (1966); E. J. Verboven, Physica **32**, 2081 (1966).

study with this method. We also discuss the connection with Wiener integrals in Sec. 5. As these integrals are the only ones which are mathematically studied, this is very important. In some cases, particularly for the one-dimensional Coulomb field, it is easy to formulate the problem in terms of such integrals. However, in most cases it is not possible. Finally, in Sec. 6, we show briefly that the functional integrals can be written in a very general form, including the corresponding integrals in quantum-mechanical problems. In this form the distribution functions (or the corresponding quantities) are written as simple independent-particle solutions in an external field, averaged over all possible external fields with a certain measure.

2. GENERATING FUNCTIONALS FOR DISTRIBUTION FUNCTIONS

In this section we first state the problem and the conventions which are used in this work. We assume a system of particles acted on by an external potential $U(\mathbf{r})$ and an interaction potential $\varphi(\mathbf{r}-\mathbf{r}')$ between pairs of particles. We assume that $\varphi(\mathbf{r})$ is everywhere finite. In some cases it is convenient to assume that the Fourier transform of φ exists. This means, among other things, that $\int \varphi(\mathbf{r})\,d^3r$ is finite. None of these assumptions is fulfilled among physical potentials. Nevertheless, they are convenient to use in formal analysis and it seems improbable that they should imply any special properties of physical interest in a continuous system. In a discrete system such as the lattice gas, it is more important to assume a hard core, i.e., it must not be possible to have more than one particle at each point.

For convenience, we use a grand canonical ensemble throughout this paper. We also put $1/kT=\beta$ and $\alpha=\mu/kT$, where μ is the chemical potential.

A general distribution function is defined as

$$F_n(\mathbf{r}_1,\cdots,\mathbf{r}_n)=\sum_{N\geq n}\frac{N!}{(N-n)!}\times\int_V\cdots\int_V d^3r_{n+1}\cdots d^3r_N\rho(N;\mathbf{r}_1,\cdots,\mathbf{r}_N),\quad(1)$$

where

$$\rho(N;\mathbf{r}_1\cdots\mathbf{r}_N)=\frac{(2\pi mkT)^{\frac{3}{2}N}e^{N\alpha}}{Z\cdot N!}\times\exp\left\{-\beta\left[\sum_{k,j}\varphi(\mathbf{r}_k-\mathbf{r}_j)+\sum_k U(\mathbf{r}_k)\right]\right\}.\quad(2)$$

Z is the partition function, necessary to normalize the density function ρ.

The simplest way of writing the chain of equations is by introducing the generating functional

$$\mathcal{F}[u]=1+\sum_{n=1}^{\infty}\frac{1}{n!}\int_V\cdots\int_V d^3r_1\cdots d^3r_n\times F_n(\mathbf{r}_1,\cdots,\mathbf{r}_n)u(\mathbf{r}_1)\cdots u(\mathbf{r}_n).\quad(3)$$

The following equation was derived for a slightly different functional by Bogoliubov[5]:

$$\frac{\delta\mathcal{F}[u]}{\delta u(\mathbf{r}')}=\exp\{\alpha-\beta U(\mathbf{r}')-\Omega_0\}\times\mathcal{F}[(u(\mathbf{r})+1)\exp\{-\beta\varphi(\mathbf{r}-\mathbf{r}')\}-1],\quad(4)$$

where

$$\Omega_0=-\tfrac{3}{2}\log(2\pi mkT).\quad(5)$$

It was shown in I that (4) is a more restrictive equation than the chain of equations derived by Kirkwood and others.[6] That chain of equations is the most extensively used. It can be derived from (4), but the opposite is not possible.

The solution of (4) is greatly simplified if we first consider the following equation, treated in I:

$$\frac{df(x)}{dx}=e^a f(\{x+1\}e^{-b}-1).\quad(6)$$

This is, of course, a very simplified, but far from trivial version of (4). The general solution can be written as

$$I(x,b,d,c)=\text{const}\times\int_C\exp\left\{\frac{t^2}{2b}+(x+1)e^{t+d}\right\}dt=\text{const}_1\times\int_C\exp\left\{\frac{bs^2}{2}+(x+1)e^{bs+d}\right\}ds.\quad(7)$$

Here $d=a+b/2$. In I, only the first form was used. Equation (7) was derived by using an integral-transformation method in (6). C is a path in the complex s- (or t-) space such that the value of the integral is not changed by a translation parallel to the real axis. This means that C starts in some direction at infinity where the integrand is zero and ends in another, nonequivalent direction. We can always start or end in a "channel" parallel to the positive real axis at a distance of an odd number times π from it in the t-space. In these channels the double exponential part of (7) becomes zero. If $b>0$, we can also move parallel to the imaginary axis, and if $b<0$, we can move parallel to the negative real axis. Some symmetrical

[5] See Ref. 3.

[6] This chain of equation was first derived by: J. G. Kirkwood, J. Chem. Phys. 3, 300 (1935); J. Yvon, *Actualités scientifiques et industrielles* (Hermann et Cie., Paris, 1935). Since then, it has been used by a number of authors, e.g., N. N. Bogoliubov (Ref. 3).

paths and their nomenclature used in I are shown in Fig. 1.

The solutions of (6) with the condition $f(0) = 1$ are:

$$f(x) = I(x, b, d, C)/I(0, b, d, C). \tag{8}$$

In particular, we want to know the derivatives at $x = 0$:

$$f^{(n)}(0) = e^{n(d-b'^2)}I(0, b, d - nb, C)/I(0, b, d, C). \tag{9}$$

The solutions of (4) can be obtained by the same methods. As we aim at the most possible mathematical rigor, we first solve the equation in a discrete case where the continuous volume is divided into a finite number M of cells around points $\mathbf{r}_1, \cdots, \mathbf{r}_M$. These cells have volumes $\Delta_1, \cdots, \Delta_M$. Any function is written as a vector: $f(\mathbf{r}) \to \{f_1 = f(\mathbf{r}_1), f_2, \cdots, f_M\}$. Instead of (4) we get a system of partial differential equations

$$\frac{\partial F[u_1, \cdots, u_M]}{\partial u_n} = \exp(\alpha - \Omega_0 + \log \Delta_n - \beta U_n) \times F[(u_1 + 1)e^{-\beta\varphi_{1n}} - 1, (u_2 + 1)e^{-\beta\varphi_{2n}} - 1, \cdots, (u_M + 1)e^{-\beta\varphi_{Mn}} - 1], \tag{10}$$

whose solutions are

$$I_M(u_1, \cdots, u_M; S_M) = \text{const} \times \int_{S_M} \exp\left\{\sum_{n,n} s_n s_{n'} \frac{\beta\varphi_{nn'}}{2}\Delta_n\Delta_{n'} + \sum_n (u_n + 1)\Delta_n \times \exp\left(\sum_{n'} \beta\varphi_{nn'}s_{n'}\Delta_{n'} + B_n\right)\right\} ds_1 \cdots ds_M. \tag{11}$$

Here

$$B_n = \alpha - \Omega_0 - \beta U_n - \varphi_{nn}. \tag{12}$$

In the continuous limit we get

$$\mathfrak{J}[u(\mathbf{r}); S] = \lim_{\substack{M\to\infty \\ \text{all } \Delta_n \to 0}} I_M[u_1 \cdots u_M; S_M] = \text{const} \times \int_S \Phi[u, s] \prod_{\mathbf{r}} ds(\mathbf{r}), \tag{13}$$

where Φ is the functional

$$\Phi[u(\mathbf{r}), s(\mathbf{r})] = \exp\left\{\iint s(\mathbf{r})s(\mathbf{r}') \frac{\beta\varphi(\mathbf{r} - \mathbf{r}')}{2} d^3r\, d^3r' + \int [u(\mathbf{r}) + 1] \times \exp\left(\int \beta\varphi(\mathbf{r} - \mathbf{r}')s(\mathbf{r}')\, d^3r' + B(\mathbf{r})\right) d^3r\right\}; \tag{14}$$

$$B(\mathbf{r}) = \alpha - \Omega_0 - \beta U(\mathbf{r}) - \varphi(0). \tag{15}$$

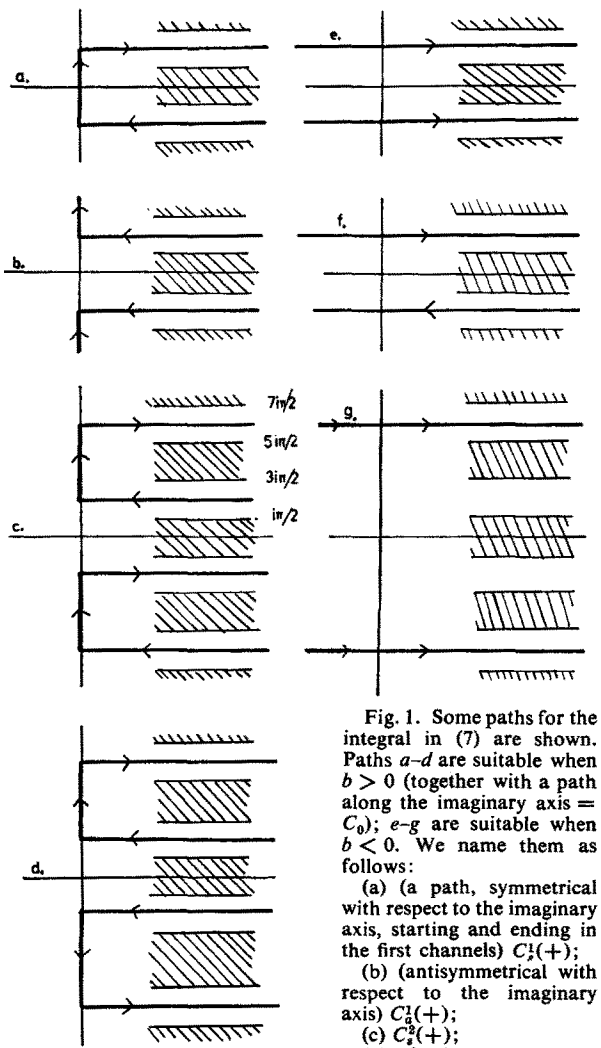


Fig. 1. Some paths for the integral in (7) are shown. Paths *a–d* are suitable when $b > 0$ (together with a path along the imaginary axis = C_0); *e–g* are suitable when $b < 0$. We name them as follows:
(a) (a path, symmetrical with respect to the imaginary axis, starting and ending in the first channels) $C_s^1(+)$;
(b) (antisymmetrical with respect to the imaginary axis) $C_a^1(+)$;
(c) $C_s^2(+)$;
(d) $C_a^2(+)$;
(e) (valid for $b < 0$, and symmetrical with respect to the real axis) $C_s^1(-)$;
(f) $C_a^1(-)$;
(g) $C_s^2(-)$.

The conditions upon the paths S_M are very much the same as these of the paths C in (7). In fact, they must be composed of the paths shown in Fig. 1. Each variable is integrated along a combination of these paths.

We always consider the functional integral as the limit (13). In this way, it is not necessary that the limits of the integrals exist, but that the limits of certain quotients exist. We always write the formal functional expressions and understand them as the respective limits. From a physical point of view, this is completely satisfactory. It also seems to be probable, although no mathematical proof has been obtained, that only functions with a finite (or possibly an

enumerable) number of discontinuities contribute to the physical properties. If this is the case, the integrals in (14) are proper Riemann integrals. (If it were not the case, the actual division of the space, which has no physical meaning, would be crucial.)

To get the generating functional, we require that

$$\mathcal{F}[0] = 1, \tag{16}$$

and for the distribution functions we get

$$\begin{aligned} &F_n(\mathbf{r}_1, \cdots, \mathbf{r}_n) \\ &= \left[\frac{\delta^{(n)}\mathcal{F}[u]}{\delta u(\mathbf{r}_1)\cdots\delta u(\mathbf{r}_n)}\right]_{u=0} \\ &= \int_S \Phi_n[0, s; \mathbf{r}_1, \cdots, \mathbf{r}_n] \prod ds(\mathbf{r}) \Big/ \int_S \Phi[0, s] \prod ds(\mathbf{r}), \end{aligned} \tag{17}$$

where

$$\begin{aligned} &\Phi_n[u, s; \mathbf{r}_1, \cdots, \mathbf{r}_n] \\ &= \exp\left\{n(\alpha - \Omega_0) - \beta\sum_{k=1}^{n} U(\mathbf{r}_k) - \beta\sum_{k,j=1}^{n}\varphi(\mathbf{r}_k - \mathbf{r}_j)\right\} \\ &\quad \times \exp\left\{\iint s(\mathbf{r})s(\mathbf{r}')\frac{\beta}{2}\varphi(\mathbf{r} - \mathbf{r}')\, d^3r\, d^3r'\right. \\ &\quad + \int [u(\mathbf{r}) + 1] \exp\left[\int \beta\varphi(\mathbf{r} - \mathbf{r}')s(\mathbf{r}')\, d^3r'\right. \\ &\quad \left.\left. + B_n(\mathbf{r}; \mathbf{r}_1, \cdots, \mathbf{r}_n)\right] d^3r\right\}. \end{aligned} \tag{18}$$

Here

$$B(\mathbf{r}; \mathbf{r}_1, \cdots, \mathbf{r}_n) = B(\mathbf{r}) - \beta\sum_{k=1}^{n}\varphi(\mathbf{r} - \mathbf{r}_k). \tag{19}$$

B is the same function as in (15).

3. SOME PROPERTIES OF THE INTEGRALS

One obvious way to try to calculate (14) [or the simpler integral (7)] is to develop the double exponential factor in the integrand into a power series. Then each term gives an integral over a Gaussian function which can be calculated. The series corresponds directly to the Ursell–Mayer series.[7] It is, however, easy to understand that this series can give a correct answer only in those cases where the Gaussian factor determines the behavior at infinity. This means that it can only be an asymptotic series for integrals parallel to the real axis. In particular, this series does not make any difference between the various channels and it is probably always divergent when the quadratic form

$$\int s(\mathbf{r})\varphi(\mathbf{r} - \mathbf{r}')s(\mathbf{r}')\, d^3r\, d^3r'$$

is not positively definite.

[7] Series expansions of this type were derived by: H. D. Ursell, Proc. Cambridge Phil. Soc. 23, 685 (1927); J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1940).

Another way to get an approximation of the integral is to look for saddle points of the integrand. In fact, we find those for the function $s(\mathbf{r})$, obeying the equation

$$\begin{aligned} &s_0(\mathbf{r}) + [u(\mathbf{r}) + 1] \\ &\quad \times \exp\left[\beta\int \varphi(\mathbf{r} - \mathbf{r}')s_0(\mathbf{r}')\, d^3r' + B(\mathbf{r})\right] = 0. \end{aligned} \tag{20}$$

If we put $u = 0$ and $s(\mathbf{r}) = -\rho(\mathbf{r})$, this becomes a molecular-field equation for an average density $\rho(\mathbf{r})$, as the second term gives the Bolzmann distribution from a potential $\int \varphi(\mathbf{r} - r')\rho(\mathbf{r}')\, d^3r'$ arising from this density.[8] If we expand the integral around the saddle point, we again obtain a Gaussian form, which represents the total integral if its range is small. In particular, its width must be smaller than the periods of trigonometric functions, arising from the complex values of the functions.

In order to get an easy understanding of the procedure, we first examine the simple integral (7). We use the second form and get a saddle point when

$$s_0(x) + (x + 1)e^{bs_0(x)+d} = 0. \tag{21}$$

The derivative of $s(x)$ is given by

$$\begin{aligned} s_0'(x) &= -e^{bs_0(x)+d}/[1 + b(x + 1)e^{bs_0(x)+d}] \\ &= \frac{s_0(x)/(1 + x)}{1 - bs_0(x)}. \end{aligned} \tag{22}$$

In the neighborhood of $s_0(x)$, the integrand of (7) becomes

$$\begin{aligned} \Phi(\sigma, x) \approx \exp\left\{\frac{s_0(x)}{2}[bs_0(x) - 2]\right\} \\ \times \exp\left\{\frac{b\sigma^2}{2}[1 - bs_0(x)]\right\}, \end{aligned} \tag{23}$$

where $s = s_0(x) + \sigma$. If this expression is used in (7) and (8), we get

$$\begin{aligned} f(x) \approx \exp\left\{\frac{b}{2}[s_0(x) - s_0(0)]\left[s_0(x) + s_0(0) - \frac{2}{b}\right]\right\} \\ \times \left|\frac{1 - bs_0(x)}{1 - bs_0(0)}\right|^{\frac{1}{2}}. \end{aligned} \tag{24}$$

The derivative of f at $x = 0$ is then obtained from (22) and (24):

$$f'(0) = -s_0(0) + \frac{b}{2}s_0(0)[1 - bs_0(0)]^{-2}. \tag{25}$$

[8] A general review of such approaches to various problems is found in R. Brout, *Phase Transitions* (W. A. Benjamin, Inc., New York, 1965).

An interesting case arises when b is negative ($= -\beta$). Then (21) has two real solutions if

$$d + \log\{\beta(x+1)\}$$

is less than one. One of these solutions is between 0 and $-1/\beta$, the other is smaller than $1/\beta$. If e^d is much smaller than 1 and β not too large, this is fulfilled and the saddle-point method gives an accurate result. We get the following solutions:

$$\text{(1)}\ s_0(x) \approx -(x+1)e^d,$$
$$\text{(2)}\ s_0(x) \approx -|d|/\beta - 1/\beta \log\{|d|/\beta(x+1)\}. \quad (26)$$

For $f'(o)$ this gives

$$\text{(1)}\ f'(o) \approx e^d(1+\beta/2) + O(e^{2d}),$$
$$\text{(2)}\ f'(o) \approx |d|/\beta + 1/\beta \log\{|d|/\beta\}. \quad (27)$$

We also see that (25) has a pole if $bs_0(0) = 1$. This merely expresses the fact that the width of the Gaussian approximation in (23) is large and that this method, strictly speaking, is inapplicable. The pole has no direct meaning, although it indicates an approximate site of a possible phase transition.

We treat the functional integrals in the same way. Before writing down the formal expressions, we draw some conclusions from (27) which are immediately applicable in the physical case. We get such expressions if we have a negative interaction and a small value of $\exp\{d\}$. The latter quantity is essentially the fugacity. The first expression in (27) for the derivatives of the generating functional gives the density, and we find that it is small for the first type of solutions; in fact, it is proportional to the fugacity, as it should be. The second solution, however, gives a very large density which increases with the fugacity. This is, of course, physically impossible. (It would mean that the pressure is negative.) Therefore, only the first solution is physically possible, which means that we have symmetrical paths of the type $C_s(-)$ in Fig. 1. The second type of solution, corresponding to antisymmetrical paths of the type $C_a(-)$, is not possible for small values of the fugacity. However, for larger values of the fugacity, it behaves in the same way as the first solution and can represent a physically possible state. It seems possible that this indeed is the liquid state. A rough estimate, made in I, shows that the transition to the liquid state really occurs in a region where the two solutions are equally possible.

The corresponding relations for the multidimensional case can be written in a formula similar to (24). We do it for the discrete case and use (11) for the integrand. The saddle point obeys the equation

$$s_n^0(u) + (u_n + 1)\exp\{\beta \sum \varphi_{nn'} s_{n'}^0(u)\Delta_{n'} + B_n\} = 0. \quad (28)$$

Then (11) gives

$$F[u_1 \cdots u_M] \approx \exp\left\{\sum_n \sum_{n'} [s_n^0(u) - s_0^0(o)]\right\} \times \left\{\sum_{n'} \frac{\beta}{2} \varphi_{nn'}[s_{n'}^0(u) + s_{n'}^0(o)]\Delta_{n'} - 1\right\} \times \exp\{-\tfrac{1}{2}\log[\det F(u)/\det F(o)]\}. \quad (29)$$

Here F is the matrix $\{F_{nn'}\}$:

$$F_{nn'} = \beta/2\left[\varphi_{nn'} - \beta \sum_{n''} \varphi_{nn''}\varphi_{n'n''}s_{n''}^0(u)\Delta_{n''}\right]. \quad (30)$$

The quotient of the determinants can, of course, be written as the quotients of the products of the eigenvalues when the logarithm gives a sum which, in the continuous limit, becomes an integral.

From (28) we can now evaluate derivatives $\delta s_n^0(u)/\delta u(\mathbf{r})$ which are used to get the distribution functions from (30).

4. THE PROBLEM OF HARD CORES

In the formulas (13)–(15), it has been necessary to assume that $\varphi(\mathbf{r})$ is always finite. This is probably not fulfilled for real potentials, but, as has already been mentioned, the actual behavior for small r is usually immaterial in physical problems.

As in earlier problems, we first study the behavior of formula (7) when b is very large. This will show the general method. If we integrate along the imaginary axis, we can use the aforementioned series in powers of e^d:

$$\begin{aligned} I(x) &= \text{const} \times \int_c \exp\left(\frac{bs^2}{2}\right)[1 + (x+1)e^{bs+a+b/2} \\ &\quad + \tfrac{1}{2}(x+1)^2 e^{2ba+2a+b} + \cdots] \\ &= \text{const} \times \int_c \left\{\exp\left(\frac{bs^2}{2}\right)\right. \\ &\quad + (x+1)e^a \cdot \exp\left[\frac{b(s+1)^2}{2}\right] \\ &\quad \left. + \tfrac{1}{2}e^{2a}\exp\left[\frac{b(s+2)^2}{2} - \frac{3b}{2}\right] + \cdots\right\} ds. \end{aligned} \quad (31)$$

When $b \to \infty$, we get the following relation from (8) when we integrate along the entire imaginary axis:

$$\lim_{b\to\infty} f(x) = 1 + \frac{x}{1+e^{-a}}. \quad (32)$$

If we integrate along an allowed channel parallel to the real axis, the integrand becomes

$$\Phi(x,\sigma) = \exp\left[\frac{b\sigma^2}{2} + i(2n+1)\pi\sigma - \frac{(2n+1)^2\pi^2}{b} - (x+1)e^{b\sigma+a+b/2}\right]. \quad (33)$$

Here $s = \sigma + i(2n+1)\pi$. It is easy to see that it is a monotonically decreasing function of σ. When σ is finite, it is certainly zero if b is infinitely large. Therefore, for the paths of the type $C_a(+)$ or $C_s(+)$ in Fig. 1, the integrals along the parts parallel to the real axis are zero. We thus have to consider only parts along the imaginary axis where (31) is valid.

We now consider two important cases. The first one is the discrete lattice gas, for which (11) is immediately valid with all $\Delta_n = 1$. However, we must have $\varphi_{nn} = \infty$, as otherwise we would allow several particles at the same point. We can do the same expansion as in (31). It is easy to see that the result (32), when integrating along the entire imaginary axis for all variables, is equivalent to

$$I_M(u_1, \cdots, u_M; C_0) = \text{const} \times \int_C \exp\left(\sum_{n,m} s_n s_m \frac{\beta\varphi_{nm}}{2}\right) \prod_n \left\{1 + (u_n + 1) \times \exp\left[\sum_m \beta\varphi_{nm} s_m + B_n^0\right]\right\} \prod_n ds_n. \quad (34)$$

In formula (34), all $\varphi_{nn} = 0$, and B_n^0 is given by (12) without φ_{nn}. C_0 means that all s^n are integrated along the same path. Formula (34) is in fact equivalent to the one derived by Siegert for the partition function.[9]

In principle, we can use this method to treat singular potentials with infinite $\varphi(0)$ or an infinite hard core. We can use the same formula, (34), when the hard core is taken away. This is, as before, only valid for paths parallel to the imaginary axis. We get, in the discrete case,

$$I_M(u_1 \cdots u_M; C_0) = \text{const} \times \int_C \exp\left\{\sum_{n,m} s_n s_m \frac{\beta\varphi_{nm}}{2}\right\} \times \prod_n \left[1 + \theta_n(u_n + 1) \exp\left\{\sum_m \beta\varphi_{nm} s_m + B_n^0\right\}\right] ds_n. \quad (35)$$

In (35) φ_{nm} is chosen in such a way that it is zero inside the hard core. θ^n is an operator, which obeys the following rules:

$\theta_n \cdot \theta_m = 0$, if n and m are inside the same core,
$= 1$, otherwise.

5. CONNECTION WITH WIENER INTEGRALS

The only type of functional integral which is extensively studied is the Wiener integral, which can be written as

$$\mathfrak{I} = \lim \int_a^b \cdots \int \exp\left[-\sum_n \frac{(s_{n+1} - s_n)^2}{\Delta_n}\right] \times F(s_1 \cdots s_n) \prod ds_n. \quad (36)$$

Such an integral can be properly defined.[10]

Formally, it can be written as

$$\mathfrak{I} = \int_a^b \cdots \int \exp -\left\{\int \left[\frac{ds(x)}{dx}\right]^2 dx\right\} F[s] \prod_x ds(x). \quad (37)$$

Our integrals cannot, in general, be written in this way. However, in the cases where all s-variables are integrated along the imaginary axis, it can be transformed to this form, when a and b are allowed to go to infinity. One obvious way is the case of an one-dimensional Coulomb potential which obeys the equation

$$\frac{d^2\varphi(x)}{dx^2} = \delta(x). \quad (38)$$

We make the following substitution in (14):

$$s(x) = i \frac{d^2}{dx^2} t(x). \quad (39)$$

The Gaussian factor is then

$$\int s(x)\varphi(x - x')s(x')\, dx\, dx' = -\int \left[\frac{dt(x)}{dx}\right]^2 dx. \quad (40)$$

The complete integral is

$$\mathfrak{I}(u) = \text{const} \times \int \exp\left\{-\int \left[\frac{dt(x)}{dx}\right]^2 dx + \int [u(x) + 1] \exp[it(x) + B(x)\, dx]\right\} \prod_x t(x), \quad (41)$$

where $\varphi(0) = 0$, which is why it causes no difficulty here. $B(x)$ would include a background change distribution. Equation (41) is a true Wiener integral, written in a formal way, and is equivalent to the integral derived by Lenard and Edwards.[11]

In fact, an integral of the type (41) can be written down for all positive-definite integrals. We note that the quadratic form can be written (here, for the sake

[9] A. J. Siegert, in *Statistical Physics; 1962 Brandeis Lectures* (W. A. Benjamin, Inc., New York, 1963), Vol. 3.

[10] See, for instance, the first two works of Ref. 1. This type of integral was first treated in the papers by: N. Wiener, J. Math. & Phys. **2**, 131 (1923); N. Wiener, Proc. London Math. Soc., Ser. 2, **22**, 454 (1924).

[11] S. F. Edwards and A. Lenard, J. Math. Phys. **3**, 778 (1962).

of simplicity, we consider the one-dimensional case):

$$\int s(x)\varphi(x-x')s(x')\,dx\,dx'$$
$$=\int [s_+^2(k)+s_-^2(k)]\varphi(k)\frac{dk}{2\pi}$$
$$=-\int [t_+^2(k)+t_-^2(k)]k^2\frac{dk}{2\pi}=\int -\left[\frac{dt(x)}{dx}\right]^2 dx. \quad (42)$$

Here

$$s_+(k)=\int s(x)\cos kx\,dx, \quad s_-(k)=\int s(x)\sin kx\,dx \quad (43)$$

and

$$t_+(k)=i[\varphi^{\frac{1}{2}}(k)/k]s_+(k). \quad (44)$$

This transformation gives the following integral:

$$\mathfrak{J}(u)=\text{const}\times\int\exp\left\{-\left[\frac{dx(x)}{dx}\right]^2 dx + \int [u(x)+1]\exp\left[-i\int\frac{d}{dx}\varphi^{\frac{1}{2}}(x-x')t(x')\,dx' + B(x)\right]dx\right\}\prod dt(x). \quad (45)$$

$\varphi^{\frac{1}{2}}(x)$ is the Fourier transform of $\varphi^{\frac{1}{2}}(k)$, which is a well-defined function.

6. THE EXTERNAL FIELD AVERAGE

The fact that the integrals are so closely related to the molecular-field approach suggests that the functional integral can be interpreted as an averaging procedure over possible external, complex densities. In fact, this appears to be a valuable approach, because a number of different physical problems can evidently be written by functional integrals in essentially the same way.

We first consider Eq. (4) without an interaction potential, but with an external density ρ:

$$\frac{\delta\mathcal{F}_0[u;\rho]}{\delta u(\mathbf{r}')}=\exp\left[\alpha-\beta U(\mathbf{r}') - \beta\int\varphi(\mathbf{r}'-\mathbf{r}'')\rho(\mathbf{r}'')\,d^3r''-\Omega_0\right]\mathcal{F}[u,\rho]. \quad (46)$$

This equation has the solution

$$\mathcal{F}_0[u;\rho]=\exp\left\{\int d^3r\,u(\mathbf{r}) \times\exp\left[-\beta\int\varphi(\mathbf{r}-\mathbf{r}')\rho(\mathbf{r}')\,d^3r'+B(\mathbf{r})\right]\right\}, \quad (47)$$

where B is the same as in (15). This means that (14) can be written in the following form:

$$\mathcal{F}[u]=\text{const}\times\int\exp\left[\int s(\mathbf{r})s(\mathbf{r}')\varphi(\mathbf{r}-\mathbf{r}')\,d^3r\,d^3r'\right]\times\mathcal{F}_0[u+1,-s]\prod_{\mathbf{r}}ds(\mathbf{r}). \quad (48)$$

This is, of course, an averaging over the external field $s(\mathbf{r})$.

We immediately see that the integral (34) is of the same type, since a nonnormalized generating functional for an external density ρ_m is

$$I_0[u_1\cdots u_M;\rho_1\cdots\rho_M]=\prod_n\{1+(u_n+1)\exp[-\textstyle\sum\beta\varphi_{nm}\rho_m+B_n^0]\}. \quad (49)$$

We can also get the same expression in the quantum-mechanical case. There, instead of distribution functions, we use the Green's functions

$$G_n(\mathbf{r}_1t_1,\cdots,\mathbf{r}_nt_n;\mathbf{r}_1't_1',\cdots,\mathbf{r}_n't_n')=\left(\frac{1}{i}\right)^n\langle T[\psi(\mathbf{r}_1t_1)\cdots\psi(\mathbf{r}_nt_n)\psi^*(\mathbf{r}_n't_n')\cdots\psi^*(\mathbf{r}_1't_1')]\rangle. \quad (50)$$

At the beginning we assume that the particles obey Bose statistics. As in the classical case, we use a generating functional here:

$$\mathcal{F}[u,v]=1+\sum_1^\infty\frac{1}{(n!)^2}\int\cdots\int G_n(\mathbf{r}_1t_1\cdots\mathbf{r}_nt_n;\mathbf{r}_1't_1'\cdots\mathbf{r}_n't_n') \times u(\mathbf{r}_1t_1)\cdots u(\mathbf{r}_nt_n) \times v(\mathbf{r}_1't_1')\cdots v(\mathbf{r}_n't_n')\,d^3r_1\,dt_1\cdots d^3r_n'\,dt_n'. \quad (51)$$

We thus obtain a chain of equations for the Green's functions, which can be written in terms of $\mathcal{F}$ as[12]

$$\left[i\frac{\partial}{\partial t}+\frac{\hbar^2}{2m}\nabla^2\right]\frac{\delta\mathcal{F}[u,v]}{\delta u(\mathbf{r},t)} = v(\mathbf{r},t)\mathcal{F}+U(\mathbf{r},t)\frac{\delta\mathcal{F}}{\delta u(\mathbf{r},t)} + i\int\varphi(\mathbf{r}-\mathbf{r}')\frac{\delta^3\mathcal{F}}{\delta u(\mathbf{r}',t)\delta v(\mathbf{r}'t)\delta u(\mathbf{r},t)}. \quad (52)$$

By using an integral transformation in the function space, a formal solution of (52) was obtained in I in the form of a functional integral:

$$\mathcal{F}[u,v]=\int_C\exp\left\{\int[p(\mathbf{r},t)u(\mathbf{r},t)+q(\mathbf{r},t)v(\mathbf{r},t)]\,d^3r\,dt\right\}\times\bar{\mathcal{F}}[p,q]\,dp\,dq. \quad (53)$$

[12] This equation was first derived in P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

C is a suitable path in the complex function space, and

$$\bar{\mathcal{F}} = \text{const} \times \exp\Big[-\int q(\mathbf{r},t)\Big(i\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\nabla^2\Big) p(\mathbf{r},t)\, d^3r\, dt + \beta \int U(\mathbf{r}) p(\mathbf{r},t) q(\mathbf{r},t)\, d^3r\, dt + \frac{i\beta}{2}\int p(\mathbf{r},t) q(\mathbf{r},t) \varphi(\mathbf{r}-\mathbf{r}') \times p(\mathbf{r}'t) q(\mathbf{r}'t)\, d^3r\, d^3r'\, dt \Big]. \tag{54}$$

We can rewrite (54) by using

$$\int \Pi\, ds \exp\Big\{ -\frac{i\beta}{2}\int [s(\mathbf{r},t) - p(\mathbf{r},t) q(\mathbf{r},t)] \times \varphi(\mathbf{r}-\mathbf{r}')[s(\mathbf{r}',t) - p(\mathbf{r}',t) q(\mathbf{r}',t)]\, dt\, d^3r\, d^3r' \Big\}$$
$$= \text{const}_1 \times = \int \Pi\, ds \exp\Big\{ -\frac{i\beta}{2}\int [s(\mathbf{r},t)\varphi(\mathbf{r}-\mathbf{r}') s(\mathbf{r}'t) + 2s(\mathbf{r},t)\varphi(\mathbf{r}-\mathbf{r}') p(\mathbf{r}'t) q(\mathbf{r}'t)]\, d^3r\, d^3r'\, dt \Big\}$$
$$\times \exp\Big\{ -\frac{i\beta}{2}\int p(\mathbf{r},t) q(\mathbf{r},t) \times \varphi(\mathbf{r}-\mathbf{r}') p(\mathbf{r}',t) q(\mathbf{r}'t)\, d^3r\, d^3r'\, dt \Big\}. \tag{55}$$

As in the classical case, we regard the functional integrals as limits of multidimensional integrals, in the sense that all expressions except those with physical meaning are purely formal. From (57), we see that the solution in the presence of an external density, but without interaction potential, is

$$\bar{\mathcal{F}}_0[p, q; \rho] = \text{const}_2 \times \exp\Big\{ -\int q(\mathbf{r},t)\Big(i\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\nabla^2\Big) p(\mathbf{r}t)\, d^3r\, dt + \beta \int \Big[U(\mathbf{r},t) + \int \varphi(\mathbf{r}-\mathbf{r}')\rho(\mathbf{r}',t)\, d^3r' \Big] \times p(\mathbf{r},t) q(\mathbf{r},t)\, d^3r\, dt \Big\}. \tag{56}$$

By the use of (55), (56) can be rewritten as

$$\bar{\mathcal{F}}[p, q] = \text{const}_3 \cdot \int \exp\Big\{ -\frac{i\beta}{2} \times \int s(\mathbf{r},t)\varphi(\mathbf{r}-\mathbf{r}') s(\mathbf{r}',t)\, d^3r\, d^3r'\, dt \Big\} \times \bar{\mathcal{F}}_0[p, q; -is]\, \Pi\, ds. \tag{57}$$

Together with (56), this expression is essentially the same as (48): an averaging over an external density.

In I, the form (53) was used only for Bose systems. It would be applicable to a system of fermions if the variables, u, v, p, and q were not ordinary functions, but belonged to an anticommuting algebra. This seems to be a complicated task, although the expressions have a formal meaning. However, (57) can be used directly for a Fermi system if $\mathcal{F}_0$ is the independent particle solution in an external field.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

SU_3 Algebra in the Mixed-Shell Space $(0p, 0d, 1s)$. I

TAKEHIKO ISHIDZU AND HIDEO KAWARADA
Department of Applied Physics, University of Tokyo, Bunkyoku, Tokyo, Japan
AND
MASACHIYO SATO
Department of Applied Mathematics, University of Yokohama, Minamiku, Yokohama, Japan

(Received 25 February 1968)

The operators of Elliott are extended in such a way that they also describe an SU_3 algebra in the mixed-shell space $\mathcal{M}^*$, where $\mathcal{M}^*$ is given as the direct sum of two spaces $\mathcal{M}(0p)$ and $\mathcal{M}(0d, 1s)$ spanned by the single-particle wavefunctions in the $(0p)$ and $(0d, 1s)$ shells of a harmonic-oscillator potential. The representation of SU_3 in this space $\mathcal{M}^*$ is investigated in detail by the aid of the weight diagram in a way analogous to that of Banerjee and Levinson. The basis is expressed in an explicit manner using the one-particle wavefunctions in the usual shell model. The states arising from two- and three-particle systems are classified according to the irreducible representations in this extended space.

1. INTRODUCTION

The SU_3-coupling scheme of Elliott[1] has made it possible to reveal the rotational character of some nuclear levels on the basis of single-particle shell model in a harmonic-oscillator potential, especially for the $(0d, 1s)$ nuclei with remarkable success. The structure and the representation of the SU_3 algebra, which constitute the mathematical foundation of the Elliott scheme, have been investigated by several authors[2–4] for the purpose of wider application to the theory of nuclear structure. Banerjee and Levinson[2] have established a formalism to treat the residual interactions within this scheme, while Moshinsky[3] and his collaborators have developed another formalism to construct the irreducible state-vectors in polynomial forms of creation operators, which is applicable not only in the SU_3 scheme but also in the seniority scheme, and then applied it to the analysis of the $(0d, 1s)$ nuclei.

Physically speaking, the operators of Elliott are composed of the transport operators which carry a nucleon from a state (nlm) to another state $(n'l'm')$ in a same energy shell; that is to say, the acting space of these operators is limited to the shell space of a definite energy. On the other hand, if one attempts to apply this scheme to the configurations extending over different shells, $(0p)^a(0d, 1s)^b$ for example, one makes the direct product[5,6] of the SU_3-irreducible state-vectors belonging to each configuration $(0p)^a$ or $(0d, 1s)^b$, and then decomposes it by the Clebsch–Gordan coefficients of the SU_3 group. The irreducible components thus formed are assumed as the eigenstates of the compound configuration in the Elliott scheme.

In such formulation, however, the operators of Elliott are reduced to simple sums of two kinds of the transport operators belonging to the $(0p)$ and $(0d, 1s)$ shell, respectively. As the result, the nucleons are carried separately in each shell by these operators, while the transport of nucleons into another shell does not take place at all. Because the effective interaction between nucleons should be derived from the second-order invariant (Casimir operator) constructed from the operators of Elliott, the matrix elements of the above interaction involve the direct integrals only but no exchange integrals between the states of different shells. Moreover, one meets with the same situation for the analysis of the particle–hole systems; if one applies the direct-product procedure to the configuration $(0p)^{-n}(0d, 1s)^n$, for example, one cannot expect any contribution from the exchange integrals between particle and hole, which are usually considered indispensable[7,8] for the evaluation of energy of the collective states in the particle–hole systems.

Thus, one way of amending the above failure in the direct-product procedure within the framework of the SU_3 scheme, would be to construct the generators of SU_3 algebra in the nine-dimensional mixed-shell space $\mathcal{M}^*(0p, 0d, 1s)$ as linear combinations of the transport operators acting not only within one shell but also over different shells. If it were achieved, the matrix elements of the Casimir operator would certainly involve the exchange integrals and, thus, a more satisfactory SU_3 scheme would be obtained also for the particle–hole systems.

[1] J. P. Elliott, Proc. Roy. Soc. (London) **A245**, 128, 562 (1958); **A272**, 557 (1963).
[2] K. M. Banerjee and C. A. Levinson, Phys. Rev. **130**, 1036 (1963).
[3] M. Moshinsky, "Group theory and the many body problem" in *Physics of Many Particle Systems*, E. Meeron Ed. (Gordon and Breach, Science Publishers, New York, 1964).
[4] R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[5] B. J. Verhaar, Nucl. Phys. **21**, 508 (1960).
[6] H. Horie and T. Yokozawa, Phys. Letters **7**, 145 (1963).
[7] L. S. Kisslinger and R. A. Sorensen, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **32**, No. 9 (1960).
[8] M. Baranger, Phys. Rev. **120**, 957 (1960).

In this article, we verify the existence of the generators of SU_3 algebra in the mixed-shell space $\mathcal{M}^*$ by extending the operators of Elliott and then proceed to investigate its representations in $\mathcal{M}^*$ following Banerjee and Levinson. It should be noticed, however, that the generators and the Casimir operator in our scheme do not conserve the parity, because they involve the transport operators carrying nucleons from one shell to another. Thus, the basis of the irreducible representations of our scheme cannot be expected to give the realistic physical states immediately; some procedures must be performed in order to project this basis into the states with definite parity.

In Sec. 2, we define nine operators $v_q^{(t)}$ ($t = 0, 1, 2$; $q = t, t-1, \cdots, -t$) as linear combinations of the transport operators in the mixed-shell space by modifying the forms of the operators of Elliott. The commutators among eight of these, $v_q^{(t)}$ ($t = 1, 2$), are shown to close among themselves, so that these eight operators may be considered as generators of a Lie algebra. Moreover, because the structure constants appearing in these commutation relations are identical with those of Elliott, one may conclude that this algebra should be SU_3. In Sec. 3, the weight diagram for this algebra is investigated in comparison with those of Banerjee and Levinson. The basis of our scheme is expressed in terms of one-body oscillator wavefunctions in the usual shell model. In Sec. 4, the states arising from two- and three-particle systems are classified according to the irreducible representations in $\mathcal{M}^*$ for the chain of subalgebras $U_n \supset SU_3 \supset R_3$.

2. GENERATORS OF SU_3 ALGEBRA

Before introducing our new operators, we summarize briefly the outline of the formalism of Elliott. As is well known, the quantum states of a single particle moving in a harmonic-oscillator potential are classified into energy shells $\mathcal{E}(N)$ corresponding to the energy values of $\hbar\omega(N + \frac{3}{2})$, where ω is the frequency of the oscillator and N is any positive integer or zero. Each energy shell $\mathcal{E}(N)$ consists of degenerate orbital states (nlm) with definite parity, where $l = N, N-2, \cdots, 1, 0$, and $n = \frac{1}{2}(N - l)$. Thus, the wavefunctions $\phi(nlm)$ of a single particle in $\mathcal{E}(N)$ span an s-dimensional vector space $\mathcal{M}_s$, where $s = \sum_l (2l + 1)$. Hereafter, for the sake of simplicity, we designate by μ a set of quantum numbers (nlm) of a single nucleon in a harmonic-oscillator potential.

At first, one introduces the transport operators $E_\mu^{\mu'}$ by the following expression:

$$E_\mu^{\mu'} = a_{\mu'}^+ a_\mu \tag{1}$$

where $a_{\mu'}^+$ and a_μ are the creation and annihilation operators of a single nucleon, obeying the commutation relations:

$$[a_{\mu'}^+, a_\mu]_+ = \delta_{\mu'\mu}, \quad [a_{\mu'}^+, a_\mu^+]_+ = [a_{\mu'}, a_\mu]_+ = 0. \tag{2}$$

If one restricts the states μ to those belonging to $\mathcal{E}(N)$, all the operators $E_\mu^{\mu'}$ commute with the Hamiltonian $\mathcal{H} = \hbar\omega(\sum_\mu a_\mu^+ a_\mu + \frac{3}{2})$. Thus, physically speaking, the nucleons are carried by these operators among the states within a definite energy shell. The commutation relations of $E_\mu^{\mu'}$ are calculated by the use of (2), giving the following result:

$$[E_{\mu_1}^{\mu_1'}, E_{\mu_2}^{\mu_2'}] = E_{\mu_1}^{\mu_2'}\delta_{\mu_2}^{\mu_1'} - E_{\mu_2}^{\mu_1'}\delta_{\mu_1}^{\mu_2'}, \qquad \mu_1, \mu_1', \mu_2, \mu_2' \in \mathcal{E}(N). \tag{3}$$

As is seen from the above formula, the commutators of $E_\mu^{\mu'}$ are expressed again in terms of linear combinations of $E_\mu^{\mu'}$; that is to say, they are closed among themselves with respect to the commutator product, so that these s^2 operators constitute[3] the generators of the unitary algebra U_s. The same holds also for an equivalent set of operators $u_q^{(t)}$ which were introduced by Elliott[1] in the forms of the irreducible tensors of rank t with respect to the three-dimensional rotation group R_3:

$$u_q^{(t)} = (2l+1)^{-\frac{1}{2}} \sum_{m,m'} (l'tm'g \mid l'tlm) E_{lm}^{l'm'},$$
$$t = l + l', l + l' - 1, \cdots, |l - l'|; \qquad q = t, t-1, \cdots, -t. \tag{4}$$

The commutation relations among the $u_q^{(t)}$ are expressed as follows[1]:

$$[u_q^{(t)}(ll'), u_p^{(s)}(kk')] = \sum_{r,v} (2r+1)^{\frac{1}{2}} (tsqp \mid tsrv) \times \{(-1)^{t+s-r}\delta(l'k)W(tslk'; rl')u_v^{(r)}(lk') - \delta(lk')W(tsl'k; rk')u_v^{(r)}(kl')\}. \tag{5}$$

Next, the operators of Elliott which describe the SU_3 algebra are introduced in terms of $u_q^{(t)}$ or, equivalently, in terms of the transport operators $E_\mu^{\mu'}$ acting in $\mathcal{M}_s$. They are defined as follows:

$$L_q = \sum_l [l(l+1)(2l+1)]^{\frac{1}{2}} u_q^{(1)}(ll), \quad q = 1, 0, -1,$$

$$Q_q = \sum_l \left[-(2N+3)\left\{\frac{l(l+1)(2l+1)}{(2l-1)(2l+3)}\right\}^{\frac{1}{2}} u_q^{(2)}(ll) + \left\{\frac{6(l+1)(l+2)(N-l)(N+l+3)}{(2l+3)}\right\}^{\frac{1}{2}} \times \{u_q^{(2)}(l, l+2) + u_q^{(2)}(l+2, l)\}\right], \qquad q = 2, 1, \cdots, -2. \tag{6}$$

For the sake of later convenience, we write down the operators L_q and Q_q for the $(0p)$ and $(0d, 1s)$ shells:

$$N = 1 \quad \begin{cases} L_q(0p) = (6)^{\frac{1}{2}} u_q^{(1)}(pp), \\ Q_q(0p) = -(30)^{\frac{1}{2}} u_q^{(2)}(pp), \end{cases} \tag{7}$$

$$N = 2 \quad \begin{cases} L_q(0d, 1s) = (30)^{\frac{1}{2}} u_q^{(1)}(dd), \\ Q_q(0d, 1s) = -(70)^{\frac{1}{2}} u_q^{(2)}(dd) + (40)^{\frac{1}{2}} \\ \qquad \times \{u_q^{(2)}(ds) + u_q^{(2)}(sd)\}. \end{cases} \tag{8}$$

All these operators (6) commute with the Hamiltonian $\mathcal{H}$ and satisfy the following commutation relations which are verified by the use of (4):

$$\begin{aligned} [L_q, L_{q'}] &= -(2)^{\frac{1}{2}}(11qq' \mid 111q + q')L_{q+q'}, \\ [Q_q, L_{q'}] &= -(6)^{\frac{1}{2}}(21qq' \mid 212q + q')Q_{q+q'}, \\ [Q_q, Q_{q'}] &= (90)^{\frac{1}{2}}(22qq' \mid 221q + q')L_{q+q'}. \end{aligned} \tag{9}$$

As is seen from (9), the commutators of these eight operators, L_q and Q_q, are closed among themselves, thus forming generators of a Lie algebra which was proved by Elliott[1] to be SU_3.

We now introduce our new operators in the direct-sum space $\mathcal{M}(0p \oplus \mathcal{M}(0d, 1s)$. As stated above, the spaces $\mathcal{M}(0p)$ and $\mathcal{M}(0d, 1s)$ are already the representation spaces of the SU_3 algebra generated by the operators of Elliott [Eqs. (7) and (8), respectively]. We define anew nine operators $v_q^{(t)}$ as follows:

$$\begin{aligned} v^{(0)} &= \sum_l (2l + 1)^{\frac{1}{2}} u^{(0)}(nl, nl), \\ v_q^{(t)} &= \tfrac{1}{2}[v_q^{(t)}(0p) + v_q^{(t)}(0d, 1s) + w_q^{(t)}(0p, 0d, 1s)], \\ & \qquad t = 1, 2, \end{aligned} \tag{10}$$

where $v_q^{(t)}$ $(0p)$ and $v_q^{(t)}$ $(0d, 1s)$ are identical with the operators of Elliott [Eqs. (7) and (8)] except the sign of $u_q^{(2)}$ $(0p, 0p)$, while the coupling operators $w_q^{(t)}$ $(0p, 0d, 1s)$ are defined as follows:

$$\begin{aligned} w_q^{(1)}(0p, 0d, 1s) &= i(10)^{\frac{1}{2}} u_q^{(1)}[0p, 0d] + i(8)^{\frac{1}{2}} u_q^{(1)}[0p, 1s], \\ w_q^{(2)}(0p, 0d, 1s) &= -i(90)^{\frac{1}{2}} u_q^{(2)}[0p, 0d], \end{aligned} \tag{11}$$

with the abbreviation

$$u_q^{(t)}[nl, n'l'] = u_q^{(t)}(nl, n'l') + u_q^{(t)}(n'l', nl).$$

The commutation relations among the operators $v_q^{(t)}$ $(t = 1, 2)$, give the following results (see Appendix):

$$\begin{aligned} [v_q^{(1)}, v_{q'}^{(1)}] &= -(2)^{\frac{1}{2}}(11qq' \mid 111q + q')v_{q+q'}^{(1)}, \\ [v_q^{(2)}, v_{q'}^{(1)}] &= -(6)^{\frac{1}{2}}(21qq' \mid 212q + q')v_{q+q'}^{(2)}, \\ [v_q^{(2)}, v_{q'}^{(2)}] &= (90)^{\frac{1}{2}}(22qq' \mid 221q + q')v_{q+q'}^{(1)}. \end{aligned} \tag{12}$$

Comparing (12) with (9), we find that the eight operators $v_q^{(t)}$ defined above are closed among themselves with respect to the commutator product, thus forming generators of a Lie algebra in the space $\mathcal{M}^*$. Moreover, we may conclude that our operators (10) generate an SU_3 algebra in the sum space $\mathcal{M}^*$, because the structure constants involved in (12) are identical with those of (9). In addition, the following remarks should be noticed: (i) The commutation relations (12) are not changed if we take the following antisymmetric combinations, instead of symmetric ones:

$$u_q^{(t)}\{nl, n'l'\} = u_q^{(t)}(nl, n'l') - u_q^{(t)}(n'l', nl). \tag{13}$$

It is necessary, however, to adopt the symmetric ones in order that the Hermiticity of $v_q^{(t)}$ is assured. (ii) $v^{(0)}$ is the number operator, which commutes with every one of $v_q^{(t)}$ $(t = 1, 2)$, but does not play any role as a generator of the algebra.

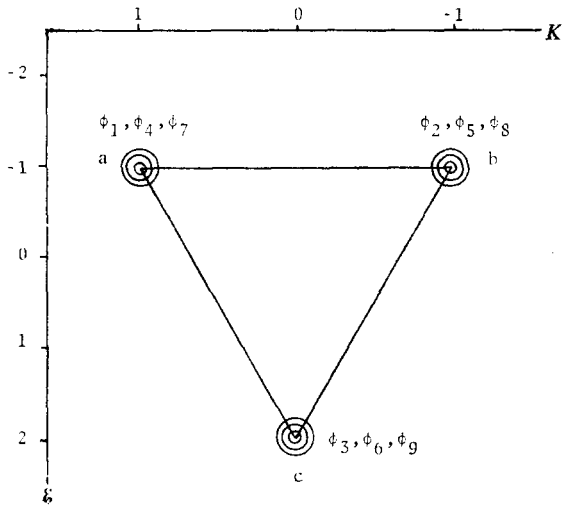


FIG. 1. Weight diagram of SU_3 in $\mathcal{M}^*$

3. REPRESENTATION IN THE MIXED-SHELL SPACE:

A. Weight Diagram of SU_3 in $\mathcal{M}^*$

As the SU_3 algebra is a Lie algebra of rank two, there exist two generators which commute with each other.[9] In our case, these commuting generators are $v_0^{(1)}$ and $v_0^{(2)}$. If we write their eigenvalues as K and $\mathcal{E}$, respectively, they are represented by a set of points on the $K\mathcal{E}$ plane, which is called the weight diagram. The weight diagram for our SU_3 representation in $\mathcal{M}^*$ is shown in Fig. 1. For the sake of comparison, we have shown also the weight diagram of the representation (10) and (01) in the scheme of Elliott, in Fig. 2, corresponding to one particle and one hole state in $(0p)$ shell. Although the triangle abc in Fig. 1

[9] G. Racah, "Group Theory and Spectroscopy," Institute for Advanced Study, Lecture notes, Princeton, New Jersey, 1951.

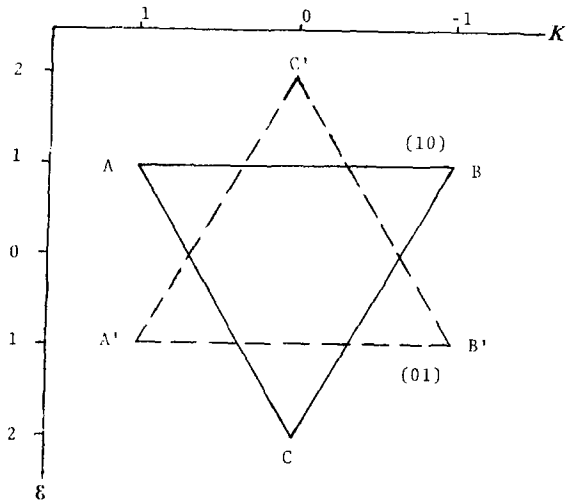


FIG. 2. Weight diagram in coupling scheme of Elliott.

is identical with that of ABC in Fig. 2, the former gives a three-fold degenerate representation.

In the following, the eigenstates corresponding to each point of the weight diagram are expressed by the single-particle wavefunctions of an oscillator in the usual shell model:

$$\begin{aligned}
&\phi_1 = d_2, && \phi_2 = d_{-2},\\
&\phi_3 = 2^{-\frac{1}{2}}(d_1 + ip_1), && \phi_4 = 2^{-\frac{1}{2}}(d_1 - ip_1),\\
&\phi_5 = 2^{-\frac{1}{2}}(d_{-1} + ip_{-1}), && \phi_6 = 2^{-\frac{1}{2}}(d_{-1} - ip_{-1}),\\
&\phi_7 = 6^{-\frac{1}{2}}(d_0 - 2^{\frac{1}{2}}s_0 - 3^{\frac{1}{2}}ip_0), &&\\
&\phi_8 = 6^{-\frac{1}{2}}(d_0 - 2^{\frac{1}{2}}s_0 + 3^{\frac{1}{2}}ip_0), &&\\
&\phi_9 = 3^{-\frac{1}{2}}(2^{\frac{1}{2}}d_0 + s_0), &&
\end{aligned} \tag{14}$$

where the abbreviation d_2, for example, expresses the normalized wavefunction of a nucleon with $l = 2$, $m = 2$ in an oscillator potential.

B. Representation in Many-Body Space

We now proceed to deal with a many-body system consisting of n nucleons in the $(0p, 0d, 1s)$ shell. The wavefunctions for this system form a complete set of antisymmetrical tensors of rank n in the space $\mathcal{M}_s$ spanned by the single-particle wavefunctions, so that they are classified by the aids of the irreducible representations of SU_3 algebra in the n-fold direct-product space $\bar{\mathcal{M}} = \mathcal{M}^* \times \mathcal{M}^* \times \cdots \times \mathcal{M}^*$.

According to the representation theory,[9] the irreducible representation of the SU_3 algebra is characterized by the highest weight, a two-dimensional vector, while the highest weight is expressed in terms of two fundamental vectors in the weight space, which are called the fundamental dominant weight (f.d.w.). The f.d.w. for the SU_3 algebra corresponds to the two points on the weight diagram, C ($\epsilon = 2$, $K = 0$) and A' ($\epsilon = 1$, $K = 1$) in Fig. 2. We express them as two-dimensional vectors as follows:

$$M^{(1)} = (2, 0), \quad M^{(2)} = (1, 1). \tag{15}$$

Then, the highest weight M of an irreducible representation is given by the following formula:

$$M = \lambda_1 M^{(1)} + \lambda_2 M^{(2)} = (2\lambda_1 + \lambda_2, \lambda_2), \tag{16}$$

where λ_1 and λ_2 are assured to be nonnegative integers.[9] Thus, the integers (λ_1, λ_2) are available in order to characterize the irreducible representations and also to classify the many-body states.

The irreducible representations may be specified also by the eigenvalues of Casimir operator which is defined by the second-order invariant derived from the generators. We define first the many-body operators $V_q^{(t)}$ as follows:

$$V_q^{(t)} = \sum_{i=1}^{n} v_q^{(t)}(i). \tag{17}$$

The fundamental properties of $v_q^{(t)}$ concerning to the commutation relations derived in the previous section are shared also by the many-body operators. Then, we define the Casimir operator for our system as follows:

$$C = 3(V^{(1)} \cdot V^{(1)}) + (V^{(2)} \cdot V^{(2)}). \tag{18}$$

The eigenvalues of C with respect to the SU_3-irreducible states (λ_1, λ_2) are given by the following relation:

$$C(\lambda_1, \lambda_2) = 4\{(\lambda_1 + \lambda_2)(\lambda_1 + \lambda_2 + 3) - \lambda_1\lambda_2\}. \tag{19}$$

This operator C is taken as the effective interaction between nucleons which gives rise to the SU_3 coupling scheme.

C. Classification of the Many-Body States

We are now ready to classify the states arising from the configurations $(0p, 0d, 1s)^n$ by the aids of the irreducible representations of the SU_3 algebra. Table I shows the possible values of the Casimir

TABLE I. Values of the Casimir operator and SU_3-irreducible states for n-particle configurations.

n	(λ_1, λ_2)	$\langle C \rangle$
1	(10)	16
2	(20)	40
	(10)	16
3	(30)	72
	(11)	36
	(00)	0

TABLE II. Classification of states arising from $(0p, 0d, 1s)^n$.

n	$[f]$	dim.	(λ_1, λ_2)	dim.	L
1	[1]	9	$(10)^3$	9	0, 1, 2
	[2]	45	$(20)^6$	36	0^2, 1, 2^3, 3, 4
			$(10)^3$	9	0, 1, 2
2			$(20)^3$	18	1^2, 2, 3
	[11]	36	$(10)^6$	18	1^2, 2, 3
			$(30)^{10}$	100	0^2, 1^2, 2^4, 3^3, 4^3, 5, 6
	[3]	165	$(11)^8$	64	0^2, 1^4, 2^4, 3^3, 4
			$(00)^1$	1	0
			$(30)^8$	80	0, 1^3, 2^4, 3^3, 4^2, 5
3	$[21]^2$	480	$(11)^{19}$	152	0^3, 1^8, 2^9, 3^6, 4^3, 5
			$(00)^8$	8	1, 2
			$(30)^1$	10	1, 3
	[111]	84	$(11)^8$	64	0^2, 1^4, 2^4, 3^3, 4
			$(00)^{10}$	10	1, 3

operator as well as (λ_1, λ_2) arising from one-, two-, and three-particles configurations.

Moreover, the generators of the unitary algebra U_s defined by (4) may be extended to the mixed-shell space and, then, also to the many-body space in an analogous way as above. Thus, the states are classified by the irreducible representations of the chain of subalgebras; U_s, SU_3, and R_3. The final results of classification are given in Table II.

4. CONCLUDING REMARKS

The SU_3 scheme, which we have developed in the mixed-shell space $(0p, 0d, 1s)$, is expected to be useful for the analysis of the levels of nuclei in the neighborhood of O^{16}, because the excitation of the core should be taken into account in this region so that one must treat them as many-body systems in the mixed-shell space $\mathcal{M}^*$.

The purpose of extending the SU_3 scheme to the mixed-shell space was to take into account the exchange integrals which are indispensable in order to reveal the collective character of nuclear levels on the basis of the shell model. Although the first aim has been achieved in our scheme, one is met with another difficulty: the nonconservation of parity. In order to obtain a realistic physical result, one must project the above-obtained irreducible states into those with a definite parity. We shall discuss this problem in detail in a forthcoming paper together with the reformulation of our scheme which is more adequate for the treatment of particle–hole systems.

APPENDIX A

In order to derive the commutation relations, which may be applied to arbitrary operators given as linear combinations of $u_q^{(t)}$, we consider the operators of the following form:

$$v_q^{(t)} = \sum_l a(l) u_q^{(t)}(ll) + \sum_{ll'} b(ll')\{u_q^{(t)}(ll') + u_q^{(t)}(l'l)\}$$

where $a(l)$ and $b(ll')$ are arbitrary constants. The commutation relations are derived by successive application of (5), giving the following results:

(I) Among the operators of the same rank t:

$$[v_q^{(t)}, v_p^{(t)}] = \sum_{r\ \mathrm{odd}} (-)2(2r+1)^{\frac{1}{2}}(ttqp \mid ttrv) \times \left[\sum_l A_l^{(r)} u_v^{(r)}(ll) + \sum_{ll'} B_{ll'}^{(r)}\{u_v^{(r)}(ll') + u_v^{(r)}(l'l)\}\right] \tag{A1}$$

where

$$A_l^{(r)} = a^2(l) W(ttll; rl) + \sum_{ll'} b^2(ll') W(ttll; rl'),$$

$$B_{ll'}^{(r)} = b(ll')\{a(l)W(ttll'; rl) + a(l')W(ttll'; rl')\} + \sum_{l''} b(ll'')b(l''l')W(ttll'; rl''). \tag{A1'}$$

(II) Among the operators of different ranks t and s:

$$[v_q^{(t)}, v_p^{(s)}] = \sum_r (2r+1)^{\frac{1}{2}}(tsqp \mid tsrv) \times \left[\sum_l A_l^{(r)} u_v^{(r)}(ll) + \sum_{ll'} B_{ll'}^{(r)} \times \{(-)^{\Delta-1} u_v^{(r)}(ll') + u_v^{(r)}(l'l)\}\right], \tag{A2}$$

where

$$A_l^{(r)} = \{(-)^\Delta - 1\}\left\{a(l)c(l)W(tsll; rl) + \sum_{l'} b(ll')d(ll')W(tsll; rl')\right\},$$

$$B_{ll'}^{(r)} = (-)\{d(ll')a(l)W(tsll'; rl) + (-)^{\Delta-1}d(ll')a(l')W(tsl'l; rl') + (-)^{\Delta-1}b(ll')c(l)W(tsl'l; rl) + b(ll')c(l')W(tsll'; rl')\} - \sum_{l''}\{b(ll'')d(l''l')W(tsll'; rl'') + (-)^{\Delta-1}b(l'l'')d(l''l)W(tsl'l; rl'')\},$$

$$\Delta = t + s - r.$$

These formulas are conveniently used in deriving the commutation relations (9) and (12).

Although the summation with respect to the index

r in (A1) and (A2) is extended over the values limited by the triangular condition among the indices (t, s, r), the operators $u_q^{(3)}$ and $u_q^{(4)}$ do not appear in (9). This is the most important features of the operators L_q and Q_q or $v_q^{(1)}$ and $v_q^{(2)}$ in order that they may close among themselves with respect to the commutator product. This peculiar character owes partly to the accidental vanishing of the corresponding Racah coefficients. Similar situation was found also in deriving the subgroup G_2 in the classification of the states arising from the configurations f^n of electrons.[10]

[10] G. Racah, Phys. Rev. 76, 1352 (1949).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Multiple Scattering of Electromagnetic Waves in an Underdense Plasma

KENNETH M. WATSON*
Institute of Geophysics and Planetary Physics, University of California, La Jolla

(Received 10 September 1968)

Scattering of electromagnetic waves by an extended underdense plasma is studied. The analysis begins with expressions for multiple scattering of waves. An explicit account of coherent scatterings leads to modified equations. These modified equations are used to derive a transport equation for the intensity (a tensor expressed in polarization components). It is shown that this transport equation may be applied to the calculation of radar backscatter.

1. INTRODUCTION

An extensive literature is available for the description of electromagnetic wave scattering from a plasma, calculated in the first Born approximation. The elementary phenomenological theory expresses this in terms of the plasma density correlation function.[1–4] More detailed theories have taken account from first principles of plasma fluctuations.[3,5–7]

Extension of the theory beyond the first Born approximation has not proven simple, although conditions for the validity of this approximation may be given in some detail.[8] For certain applications, a fluctuating dielectric constant, having assigned statistical properties, may be used in Maxwell's equations.[9] Except when simplifying conditions obtain, the solution of these equations is apt to be formidable. Also, relating the dielectric constant to the plasma properties is not generally trivial.

In an interesting paper Ruffine and de Wolf[10] have analyzed the second Born approximation, with particular emphasis on cross polarization in backscatter. Their results are useful because this is the lowest order in which cross polarization of backscatter occurs and because of the insight gained into the failure of the first Born approximation. Experience with the second Born approximation has unfortunately indicated that this ordinarily has too limited a range of validity to be of much practical use.

In the present paper we discuss plasma scattering of electromagnetic waves, beginning with the scattering from individual electrons. Scattering from the electrons of the plasma is expressed in terms of coupled multiple-scattering equations. It is shown that separation of the scattered waves into coherent and incoherent parts arises when statistical averages are taken of the electron coordinates. Coherently interfering wavelets propagate in a refracting medium, the refractive index being expressed in terms of the plasma correlation functions and density. The incoherent scattering is, in turn, described by a "classical-like" transport equation. Under suitable conditions this transport equation reduces to a strictly classical equation, such as one would write down for massive particles.

The theory developed here represents the classical version of a quantum-mechanical theory of multiple

* Presently on leave from the Department of Physics, University of California, Berkeley.

[1] H. G. Booker, J. Geophys. Res. 64, 2164 (1959).
[2] A. D. Wheelon, J. Res. Natl. Bur. Std. 63D, 205 (1959).
[3] E. E. Salpeter, Phys. Rev. 120, 1528 (1960).
[4] L. Van Hove [Phys. Rev. 95, 249 (1954)] seems to have given the first derivation using the time-dependent correlation function. Van Hove's derivation is quantum mechanical, so of course includes classical scattering as a special case.
[5] M. Rosenbluth and N. Rostoker, Phys. Fluids 5, 776 (1962).
[6] O. Buneman, J. Geophys. Res. 67, 2050 (1962).
[7] D. T. Farley, J. Geophys. Res. 71, 4091 (1966).
[8] E. E. Salpeter and S. B. Treiman, J. Geophys. Res. 69, 869 (1964).
[9] V. I. Tatarski, *Wave Propagation in a Turbulent Medium* (McGraw-Hill Book Co., Inc., New York, 1961).
[10] R. S. Ruffine and D. A. de Wolf, J. Geophys. Res. 70, 4313 (1965).

r in (A1) and (A2) is extended over the values limited by the triangular condition among the indices (t, s, r), the operators $u_q^{(3)}$ and $u_q^{(4)}$ do not appear in (9). This is the most important features of the operators L_q and Q_q or $v_q^{(1)}$ and $v_q^{(2)}$ in order that they may close among themselves with respect to the commutator product. This peculiar character owes partly to the accidental vanishing of the corresponding Racah coefficients. Similar situation was found also in deriving the subgroup G_2 in the classification of the states arising from the configurations f^n of electrons.[10]

[10] G. Racah, Phys. Rev. 76, 1352 (1949).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Multiple Scattering of Electromagnetic Waves in an Underdense Plasma

KENNETH M. WATSON*
Institute of Geophysics and Planetary Physics, University of California, La Jolla

(Received 10 September 1968)

Scattering of electromagnetic waves by an extended underdense plasma is studied. The analysis begins with expressions for multiple scattering of waves. An explicit account of coherent scatterings leads to modified equations. These modified equations are used to derive a transport equation for the intensity (a tensor expressed in polarization components). It is shown that this transport equation may be applied to the calculation of radar backscatter.

1. INTRODUCTION

An extensive literature is available for the description of electromagnetic wave scattering from a plasma, calculated in the first Born approximation. The elementary phenomenological theory expresses this in terms of the plasma density correlation function.[1–4] More detailed theories have taken account from first principles of plasma fluctuations.[3,5–7]

Extension of the theory beyond the first Born approximation has not proven simple, although conditions for the validity of this approximation may be given in some detail.[8] For certain applications, a fluctuating dielectric constant, having assigned statistical properties, may be used in Maxwell's equations.[9] Except when simplifying conditions obtain, the solution of these equations is apt to be formidable. Also, relating the dielectric constant to the plasma properties is not generally trivial.

In an interesting paper Ruffine and de Wolf[10] have analyzed the second Born approximation, with particular emphasis on cross polarization in backscatter. Their results are useful because this is the lowest order in which cross polarization of backscatter occurs and because of the insight gained into the failure of the first Born approximation. Experience with the second Born approximation has unfortunately indicated that this ordinarily has too limited a range of validity to be of much practical use.

In the present paper we discuss plasma scattering of electromagnetic waves, beginning with the scattering from individual electrons. Scattering from the electrons of the plasma is expressed in terms of coupled multiple-scattering equations. It is shown that separation of the scattered waves into coherent and incoherent parts arises when statistical averages are taken of the electron coordinates. Coherently interfering wavelets propagate in a refracting medium, the refractive index being expressed in terms of the plasma correlation functions and density. The incoherent scattering is, in turn, described by a "classical-like" transport equation. Under suitable conditions this transport equation reduces to a strictly classical equation, such as one would write down for massive particles.

The theory developed here represents the classical version of a quantum-mechanical theory of multiple

* Presently on leave from the Department of Physics, University of California, Berkeley.

[1] H. G. Booker, J. Geophys. Res. 64, 2164 (1959).
[2] A. D. Wheelon, J. Res. Natl. Bur. Std. 63D, 205 (1959).
[3] E. E. Salpeter, Phys. Rev. 120, 1528 (1960).
[4] L. Van Hove [Phys. Rev. 95, 249 (1954)] seems to have given the first derivation using the time-dependent correlation function. Van Hove's derivation is quantum mechanical, so of course includes classical scattering as a special case.
[5] M. Rosenbluth and N. Rostoker, Phys. Fluids 5, 776 (1962).
[6] O. Buneman, J. Geophys. Res. 67, 2050 (1962).
[7] D. T. Farley, J. Geophys. Res. 71, 4091 (1966).
[8] E. E. Salpeter and S. B. Treiman, J. Geophys. Res. 69, 869 (1964).
[9] V. I. Tatarski, *Wave Propagation in a Turbulent Medium* (McGraw-Hill Book Co., Inc., New York, 1961).
[10] R. S. Ruffine and D. A. de Wolf, J. Geophys. Res. 70, 4313 (1965).

scattering developed by the author.[11,12] This was used to derive a transport equation for the description of incoherent scattering.[13] Since classical mechanics represents a limiting case in quantum mechanics, the above theories could, in principle, be applied to our present problem.[14] We choose not to do this, but to give a direct derivation within the framework of classical mechanics. The derivation given here is much simpler and more satisfying to the physical intuition than the corresponding quantum-mechanical development, with its reliance on formal operator algebra.

In the next section we define our problem and introduce some of the notation to be used. The multiple-scattering equations and their coherent-incoherent decomposition will be given in Sec. 3, followed by a brief discussion of these in Sec. 4. The transport equation is derived in Sec. 5. Some comments on the use of this are made in Sec. 6. The special problem associated with radar backscatter is described in Sec. 7, where the appropriate transport theory is given.

2. DESCRIPTION OF THE PROBLEM

We suppose a plane-polarized electromagnetic wave to be incident on a plasma, whose properties are described below. At some point $\mathbf{r}$ the electric-field vector associated with this incident wave is expressed as

$$\mathbf{E}_I(\mathbf{r}) = \hat{\mathbf{e}}(1)E_I(\mathbf{r}),$$
$$E_I(\mathbf{r}) = E_0 e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{2.1}$$

Here $\mathbf{k}$ is 2π times the wavenumber vector and $\hat{\mathbf{e}}(1)$ is the polarization vector of this incident wave.[15]

If the wave (2.1) were scattered by a single electron at a point $\mathbf{z}$, the scattered field vector (in the wave zone) would be

$$\mathbf{E}_{sc}(\mathbf{r}) = \sum_{j=1}^{2} \hat{\mathbf{e}}_{\hat{\mathbf{p}}}(j)E_{sc}(j), \tag{2.2a}$$

$$E_{sc}(j) = R^{-1}e^{ikR}f_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}})E_I(\mathbf{z}). \tag{2.2b}$$

Here $\mathbf{R} \equiv \mathbf{r} - \mathbf{z}$, $\hat{\mathbf{p}}$ is a unit vector parallel to $\mathbf{R}$, and $\hat{\mathbf{k}} = \mathbf{k}/k$, etc. Unit vectors $\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(1)$ and $\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(2)$ are defined as follows:

$$\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(2) = \frac{\hat{\mathbf{p}} \times \hat{\mathbf{k}}}{|\hat{\mathbf{p}} \times \hat{\mathbf{k}}|},$$
$$\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(1) = \hat{\mathbf{e}}_{\hat{\mathbf{p}}}(2) \times \hat{\mathbf{k}}. \tag{2.3a}$$

[11] K. M. Watson, Phys. Rev. **89**, 575 (1953); **105**, 1388 (1957).
[12] A review of the quantum theory of scattering from composite systems is given by M. L. Goldberger and K. M. Watson, *Collision Theory* (John Wiley & Sons, Inc., New York, 1964).
[13] K. M. Watson, Phys. Rev. **118**, 886 (1960).
[14] An illustration of this is given on p. 772 of Ref. 12.
[15] The description for circularly and elliptically polarized waves may, of course, be obtained from a superposition of plane-polarized waves. The theory for elliptic polarization is given in Sec. 6.

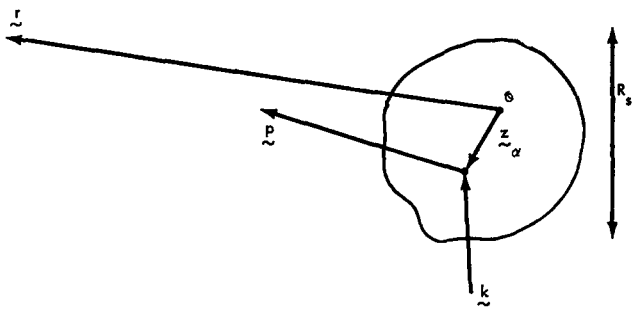


FIG. 1. Ray paths for single scattering.

The Thomson scattering amplitude is

$$f_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}) = -r_0 \hat{\mathbf{e}}_{\hat{\mathbf{p}}}(j) \cdot \hat{\mathbf{e}}(1), \tag{2.3b}$$

where $r_0 = e^2/mc^2 = 2.8 \times 10^{-13}$ cm is the classical electron radius.

If the wave (2.1) illuminates a plasma containing N electrons at respective positions $\mathbf{z}_1, \cdots, \mathbf{z}_N$, the field $\mathbf{E}_{sc}(\mathbf{r})$ will contain wavelets scattered from all of these. In the first Born approximation the rescattering of once-scattered waves is neglected, and, instead of (2.2b), we have

$$E_{sc}(j) = \sum_{\alpha=1}^{N} \frac{e^{ikR_\alpha}}{R_\alpha} f_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}})E_I(\mathbf{z}_\alpha). \tag{2.3c}$$

Here r and the z_α's are measured from a point O in the plasma (as illustrated in Fig. 1), $\mathbf{R}_\alpha \equiv \mathbf{r} - \mathbf{z}_\alpha$, and $\hat{\mathbf{p}}$ is taken parallel to $\mathbf{r}$. We have assumed here that $r \gg R_s$, the "size" of the plasma. (Scattering from the positive ions is neglected in our analysis.)

When the first Born approximation is not valid, we must replace $E_I(\mathbf{z}_\alpha)$ in Eq. (2.3) by the *total* field (incident plus scattered) impinging on electron α. We do this explicitly in the next section.

To calculate average scattered power we must specify the statistical properties of the plasma. We do this heuristically as follows. First, we assume the plasma properties to be stationary. The probability that the N electrons are at $\mathbf{z}_1, \cdots, \mathbf{z}_N$ within the volume elements $d^3z_1, \cdots, d^3z_N$ is written as

$$P_N(\mathbf{z}_1, \cdots, \mathbf{z}_N)\, d^3z_1 \cdots d^3z_N.$$

The single-electron distribution function is

$$P_1(\mathbf{z}_1) = \int P_N\, d^3z_2 \cdots d^3z_N. \tag{2.4}$$

We suppose that $P_1(\mathbf{z}_1)$ vanishes for $z_1 \gg R_s$, a characteristic plasma dimension. The pair-distribution function is

$$P_2(\mathbf{z}_1, \mathbf{z}_2) = \int P_N\, d^3z_3 \cdots d^3z_N. \tag{2.5}$$

It is convenient to re-express this in terms of the *pair*-correlation function $g(\mathbf{z}_1, \mathbf{z}_2)$:

$$P_2(\mathbf{z}_1, \mathbf{z}_2) = P_1(\mathbf{z}_1)P_1(\mathbf{z}_2)[1 + g(\mathbf{z}_1, \mathbf{z}_2)]. \quad (2.6)$$

We shall suppose that $g(\mathbf{z}_1, \mathbf{z}_2)$ is characterized by a parameter R_c, called the "correlation range," such that g vanishes when $|\mathbf{z}_1 - \mathbf{z}_2| \gg R_c$. (For studying turbulent flow, the use of a "correlation range" may be artificial. In such cases a "correlation range" $R_c \approx k^{-1}$ is selected by the scattering process.)

Continuing the process of reduction begun with Eqs. (2.4)–(2.6), we can define $P_3, P_4 \cdots$ and a corresponding hierarchy of correlation functions. For example, we can write

$$P_3(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3) = P_1(\mathbf{z}_1)P_1(\mathbf{z}_2)P_1(\mathbf{z}_3)\{1 + g(\mathbf{z}_1, \mathbf{z}_2) + g(\mathbf{z}_2, \mathbf{z}_3) + g(\mathbf{z}_3, \mathbf{z}_1) + g_3(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)\}, \quad (2.7)$$

where g_3 is the triple-correlation function which we assume to vanish unless $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{z}_3$ are all separated by distances not much greater than R_c.

For calculating averages it is convenient to introduce a set of conditional probabilities with the equations

$$\begin{aligned} P_N(\mathbf{z}_1, \cdots, \mathbf{z}_N) &\equiv P_1(\mathbf{z}_1)P^{(1)}_{N-1}(\mathbf{z}_2, \cdots, \mathbf{z}_N; \mathbf{z}_1) \\ &\equiv P_2(\mathbf{z}_1, \mathbf{z}_2)P^{(2)}_{N-2}(\mathbf{z}_3, \cdots, \mathbf{z}_N; \mathbf{z}_1, \mathbf{z}_2) \\ &\equiv \cdots. \end{aligned} \quad (2.8)$$

In terms of these we adopt the following notation for averages of some function $A(\mathbf{z}_1, \cdots, \mathbf{z}_N)$:

$$\langle A\rangle_\alpha \equiv \int \prod_{\sigma(\neq\alpha)=1}^{N} d^3z_\sigma \times AP^{(1)}_{N-1}(\mathbf{z}_1, \cdots, \mathbf{z}_{\alpha-1}, \mathbf{z}_{\alpha+1}, \cdots, \mathbf{z}_N; \mathbf{z}_\alpha), \quad (2.9a)$$

$$\langle A\rangle_{\alpha\beta} \equiv \int \prod_{\sigma(\neq\alpha,\beta)=1}^{N} d^3z_\sigma \times AP^{(2)}_{N-2}(\mathbf{z}_1, \cdots, \mathbf{z}_{\alpha-1}, \mathbf{z}_{\alpha+1}, \cdots, \mathbf{z}_{\beta-1}, \mathbf{z}_{\beta+1}, \cdots, \mathbf{z}_N; \mathbf{z}_\alpha, \mathbf{z}_\beta), \quad (2.9b)$$

etc. Thus, for example,

$$\begin{aligned} \langle A\rangle &\equiv \int P_N A\, d^3z_1 \cdots d^3z_N \\ &= \int \langle A\rangle_{\alpha\beta} P_2(\mathbf{z}_\alpha, \mathbf{z}_\beta)\, d^3z_\alpha\, d^3z_\beta. \end{aligned} \quad (2.9c)$$

The description of "Doppler shifts" in the scattered frequencies would require the introduction of time-dependent correlation functions. We shall not consider this generalization in the present paper.

To prepare for and illustrate the notation to be used later, we now write down the expression for the scattered power in first Born approximation. If a receiver is placed at a large distance $\mathbf{r}$ from the plasma and detects waves with a plane polarization $\hat{\mathbf{e}}$, the received power per unit area is[16]

$$\begin{aligned} \mathfrak{F}_{rec} &= c\left\langle \frac{|\hat{\mathbf{e}} \cdot \mathbf{E}_{sc}(\mathbf{r})|^2}{8\pi} \right\rangle \\ &= \frac{\mathfrak{F}_0}{r^2} |f(\hat{\mathbf{e}})|^2 \sum_{\alpha,\beta=1}^{N} \int e^{i\Delta\mathbf{k}\cdot(\mathbf{z}_\alpha - \mathbf{z}_\beta)} P_2(\mathbf{z}_\alpha, \mathbf{z}_\beta)\, d^3z_\alpha\, d^3z_\beta. \end{aligned} \quad (2.10)$$

Here $\mathfrak{F}_0 = (c/8\pi)\,|E_0|^2$ is the incident power flux,

$$\Delta\mathbf{k} \equiv \mathbf{p} - \mathbf{k}, \quad (2.11)$$

where $\mathbf{p} = k\hat{\mathbf{p}}$, and

$$f(\hat{\mathbf{e}}) \equiv \sum_{j=1}^{2} \hat{\mathbf{e}} \cdot \hat{\mathbf{e}}_{\hat{\mathbf{p}}}(j) f_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}). \quad (2.12)$$

If we assume that $N \gg 1$ and write

$$\rho(\mathbf{x}) \equiv NP_1(\mathbf{x}) \quad (2.13)$$

for the electron density, we can use Eq. (2.6) to express (2.10) in the form

$$\begin{aligned} \mathfrak{F}_{rec} &= \frac{\mathfrak{F}_0}{r^2} |f(\hat{\mathbf{e}})|^2 \left| \int \rho(\mathbf{z}) e^{i\Delta\mathbf{k}\cdot\mathbf{z}}\, d^3z \right|^2 \\ &\quad + \frac{\mathfrak{F}_0}{r^2} |f(\hat{\mathbf{e}})|^2 \int d^3z\, d^3z' \rho(\mathbf{z})\rho(\mathbf{z}')g(\mathbf{z}, \mathbf{z}')e^{i\Delta\mathbf{k}\cdot(\mathbf{z}-\mathbf{z}')}, \end{aligned} \quad (2.14)$$

where we have neglected the terms corresponding to $\alpha = \beta$ in Eq. (2.10). The first term above represents the *coherently* scattered power. It may also be expressed as[17]

$$(\mathfrak{F}_{rec})_{coh} = \frac{c}{8\pi} |\langle \hat{\mathbf{e}} \cdot \mathbf{E}_{sc}\rangle|^2. \quad (2.15)$$

The second term in Eq. (2.14) then represents the *incoherent* scattering, or

$$(\mathfrak{F}_{rec})_{inc} = \frac{c}{8\pi} [\langle |\hat{\mathbf{e}} \cdot \mathbf{E}_{sc}|^2\rangle - |\langle \hat{\mathbf{e}} \cdot \mathbf{E}_{sc}\rangle|^2]. \quad (2.16)$$

The incoherent differential scattering cross section is seen from (2.14) to be[18]

$$\sigma(\hat{\mathbf{p}}, \hat{\mathbf{k}}) = |f(\hat{\mathbf{e}})|^2 \int d^3z\, d^3z'\, \rho(\mathbf{z})\rho(\mathbf{z}')g(\mathbf{z}, \mathbf{z}')e^{i\Delta\mathbf{k}\cdot(\mathbf{z}-\mathbf{z}')}. \quad (2.17a)$$

[16] In MKS units, the constant c in Eq. (2.10) is $c = 4\pi(\epsilon_0/\mu_0)^{\frac{1}{2}}$. In unrationalized Gaussian units, c is the velocity of light.

[17] An excellent review of coherence and incoherence in the scattering of waves has been given by M. Lax, Rev. Mod. Phys. 23, 289 (1951).

[18] We define differential scattering cross section as scattered power per unit solid angle per unit flux. The total cross section is the integral of σ over all solid angles. The conventional radar backscatter cross section is then $4\pi\sigma(-\hat{\mathbf{k}}, \hat{\mathbf{k}})$ in the notation of Eq. (2.17a).

We may characterize the order of magnitude of this as

$$\sigma \approx |f(\hat{\mathbf{e}})|^2 N[R_c^3 \rho]. \tag{2.17b}$$

The neglected terms corresponding to $\alpha = \beta$ in Eq. (2.10) would give a contribution

$$\sigma_d = N |f(\hat{\mathbf{e}})|^2 \tag{2.18}$$

to the incoherent scattering.[19] This is evidently negligible when

$$R_c^3 \rho \gg 1. \tag{2.19}$$

We are now ready to impose certain restrictions on the plasma which will let us make corresponding simplifying assumptions in our analysis. Certain of these that are *basic* to our analysis are described first. Others that could be relaxed at the cost of some added complexity are then given.

The "basic" assumptions made in this work are:

(B1): $N \gg 1$, where N is the total number of electrons. This condition is so well satisfied in practice that it is not worth keeping corrections, which would amount to bookkeeping on individual electrons.

(B2): $lk \gg 1$, where l is the scattering mean free path [defined in Eq. (2.23) below]. We interpret this to mean that a rescattering always occurs in the wave zone of a previous scattering. Although our multiple scattering equations and coherent decomposition can easily be written down without this assumption,[20] the simple intuitive interpretation of our equations would be modified and the equations themselves would be more cumbersome to handle.

(B3): The plasma is underdense. Again, our formal theory may even be applied to the overdense case. The treatment of overdense scattering is so different from that of underdense scattering that separate discussion of these two cases is warranted.

(B4): $kR_s \gg 1$, where R_s is a parameter characterizing the size of the plasma. This means that we can neglect coherent scattering outside the diffraction cone, which has an angular aperture of order $(kR_s)^{-1}$. We shall actually interpret this condition as implying that coherent scattering occurs only in the forward direction (that is, parallel to $\hat{\mathbf{k}}$).

If the condition (B4) were violated, our discussion would be trivial. From (B2) we have $kR_s \gg R_s/l$. If kR_s were of order unity, we would have $l \gg R_s$ and a valid first Born approximation.

(B5): $l \gg R_c$, where R_c is the "correlation range." This condition implies that subsequent rescatterings will tend to occur in regions of plasma uncorrelated to regions of previous scatterings. Thus, expansions in successively higher orders of correlation functions can be made.

(B6): $|\nabla \ln n| \ll k$, where $n(\mathbf{r})$ is the refractive index of the plasma at a point $\mathbf{r}$. This condition implies that we can use the eikonal approximation for the coherent propagation of a wave between incoherent scatterings. Expressed in more physical terms, this means that the waves will propagate along ray paths, as is assumed in geometrical optics.

The following assumptions are made for "convenience." They permit us to make analytical simplifications in our equations, but could be relaxed with only minor changes in our analysis.

(C1): $R_c |\nabla \ln \rho| \ll 1$. This permits us to treat the mean electron density ρ as a constant over the correlation range. We also assume that the above inequality holds when ρ is replaced by other plasma parameters, such as temperature.

(C2): The correlation function $g(\mathbf{z}, \mathbf{z}')$ is a spherically symmetric function of $(\mathbf{z} - \mathbf{z}')$. Thus

$$g(\mathbf{z}, \mathbf{z}') \equiv g(\mathbf{z}; |\mathbf{z} - \mathbf{z}'|) \cong g(\mathbf{z}'; |\mathbf{z} - \mathbf{z}'|), \tag{2.20}$$

where in the last form we have made use of (C1). This condition implies that the refractive index $n(\mathbf{r})$ depends on position only and not on the direction of propagation.

Relaxation of the conditions (C1) and (C2) requires only a trivial generalization of our treatment.

(C3): $|n - 1| \ll 1$. This condition will be interpreted as implying that we can replace eikonal *ray paths* by straight line paths. There is no great complication of our formal analysis if this restriction is dropped. The principal simplification resulting from condition (C3) (other than the obvious geometrical one) is that no rotation of the polarization vector occurs along ray paths. This means that the Green's function describing coherent propagation is a scalar rather than a tensor quantity.

(C4): When $|\mathbf{z}_{\alpha'} - \mathbf{z}_\alpha| = O(R_c)$, we assume that

$$S(\alpha', \beta) \equiv \int_{z_\beta}^{z_{\alpha'}} n(\mathbf{x})\, ds \cong S(\alpha, \beta) + n(\mathbf{z}_\alpha)\hat{\mathbf{R}}_{\alpha\beta} \cdot (\mathbf{z}_{\alpha'} - \mathbf{z}_\alpha),$$

where $R_{\alpha\beta} = |\mathbf{z}_\alpha - \mathbf{z}_\beta| = O(l)$ [see Eq. (3.23) etc.]. The restriction (C4) permits us to simplify certain exponential phases.

[19] Sometimes the terms "quasicoherent" and "strictly incoherent" are applied to respective contributions (2.17) and (2.18), to the scattering. Since we shall neglect the "strictly incoherent" scattering, we can apply the term "incoherent" to (2.16), etc., without confusion.

[20] The derivation of the Lorentz–Lorenz formula on p. 772 of Ref. 12 was given without the introduction of assumption (B2).

(C5): We assume that the inequality (2.19) is valid. It seems that this condition will usually be met and it saves carrying a few extra terms in our equations.[21]

The assumptions listed above permit us to simplify the incoherent scattering cross section (2.17a) and to obtain a simple expression for the mean free path l. First, let us average over initial and sum over final polarization directions to obtain the Thomson differential cross section:

$$\sigma_T(\hat{\mathbf{p}}\cdot\hat{\mathbf{k}}) = \tfrac{1}{2}\sum_{\hat{\mathbf{e}}(1),\hat{\mathbf{e}}}[f(\hat{\mathbf{e}})]^2 = \tfrac{1}{2}r_0^2[1+(\hat{\mathbf{p}}\cdot\hat{\mathbf{k}})^2]. \quad (2.21)$$

Then, using Eq. (2.20), Eq. (2.17a) becomes

$$\sigma = \int d^3 z\rho^2(\mathbf{z})\int d^3 R g(\mathbf{z};R)e^{i\Delta\mathbf{k}\cdot\mathbf{R}}\sigma_T(\hat{\mathbf{p}}\cdot\hat{\mathbf{k}}), \quad (2.22)$$

on summing and averaging our polarizations, as indicated. The mean free path is the reciprocal of the total scattering cross section per unit volume, which from (2.22) is seen to be expressed as

$$\frac{1}{l(z)} = \rho^2(\mathbf{z})\int d\Omega_{\hat{\mathbf{p}}}\int d^3 R g(\mathbf{z};R)e^{i\Delta\mathbf{k}\cdot\mathbf{R}}\sigma_T(\hat{\mathbf{p}}\cdot\hat{\mathbf{k}}), \quad (2.23)$$

where $d\Omega_{\hat{\mathbf{p}}}$ is an element of solid angle about the direction of $\hat{\mathbf{p}}$. A more accurate expression for l will be obtained with Eq. (4.5). From (2.23) we obtain the estimate

$$R_c/l = O[(r_0R_c^2\delta\rho)^2] \ll 1, \quad (2.24)$$

where $\delta\rho^2$ is the mean-square electron density fluctuation.

3. MULTIPLE SCATTERING

We turn now to the task of correcting the first Born approximation (2.2) for the scattered field $\mathbf{E}_{sc}$. The electric field vector for a wave arriving at a point $\mathbf{z}_\alpha$ will be the sum of the incident field plus that for the waves scattered from all the electrons. Thus we can write this in the form

$$\mathbf{E}(\mathbf{z}_\alpha) = \mathbf{E}_I(\mathbf{z}_\alpha) + \sum_{\beta(\neq\alpha)=1}^{N}\sum_{j=1}^{2}\hat{\mathbf{e}}_{\alpha\beta}(j)F_{\alpha\beta}(\mathbf{z}_\alpha, j). \quad (3.1)$$

To explain this equation we first recall that $\mathbf{E}_I$ is just the incident field (2.1). The polarization vectors are defined as follows:

$$\hat{\mathbf{e}}_{\alpha\beta}(2) = \frac{\hat{\mathbf{k}}_{\alpha\beta}\times\hat{\mathbf{k}}}{|\hat{\mathbf{k}}_{\alpha\beta}\times\hat{\mathbf{k}}|}, \qquad \hat{\mathbf{e}}_{\alpha\beta}(1) = \hat{\mathbf{e}}_{\alpha\beta}(2)\times\hat{\mathbf{k}}_{\alpha\beta}, \quad (3.2)$$

where

$$\hat{\mathbf{k}}_{\alpha\beta} = \frac{\mathbf{z}_\alpha - \mathbf{z}_\beta}{|\mathbf{z}_\alpha - \mathbf{z}_\beta|}. \quad (3.3)$$

The quantity $F_{\alpha\beta}(\mathbf{z}_\alpha, j) \equiv F_{\alpha\beta}(j)$ represents the component along $\hat{\mathbf{e}}_{\alpha\beta}(j)$ of the electric field vector of that wave scattered from an electron at $\mathbf{z}_\beta$ to the point $\mathbf{z}_\alpha$. The quantity $\mathbf{z}_\alpha$ in (3.1) will sometimes be taken to represent just a point in space and sometimes to represent an electron coordinate. In the former case the restriction $\beta \neq \alpha$ represents no real restriction on the summation, since the sum then runs over all electron coordinates. In the latter case we omit the wave scattered by electron α, so $\mathbf{E}(\mathbf{z}_\alpha)$ represents the field *incident* on electron α. In any case, this restriction can be ignored when we replace the sums by integrals in performing averages such as those of Eqs. (2.9) [because of assumption (B1)].

To obtain the total scattered radiation we set $\mathbf{z}_\alpha = \mathbf{r}$, a point far from the plasma.

We of course require equations to determine the $F_{\alpha\beta}$. It is intuitively evident that these are

$$F_{\alpha\beta}(\mathbf{z}_\alpha, i) = G^0_{\alpha\beta}f_{i1}(\alpha\beta, \beta 0)E_I(\mathbf{z}_\beta) + \sum_{\sigma(\neq\beta)=1}^{N}\sum_{j=1}^{2}G^0_{\alpha\beta}f_{ij}(\alpha\beta, \beta\sigma)F_{\beta\sigma}(\mathbf{z}_\beta, j). \quad (3.4)$$

Here

$$G^0_{\alpha\beta} = e^{ikR_{\alpha\beta}}/R_{\alpha\beta}, \quad (3.5)$$

with

$$\mathbf{R}_{\alpha\beta} \equiv \mathbf{z}_\alpha - \mathbf{z}_\beta, \quad (3.6)$$

and

$$f_{ij}(\alpha\beta, \beta\sigma) = -r_0\hat{\mathbf{e}}_{\alpha\beta}(i)\cdot\hat{\mathbf{e}}_{\beta\sigma}(j) \quad (3.7)$$

is the Thomson scattering amplitude for scattering a wave from the direction $\hat{\mathbf{k}}_{\beta\sigma}$ to $\hat{\mathbf{k}}_{\alpha\beta}$. Finally,

$$f_{i1}(\alpha\beta, 0) = -r_0\hat{\mathbf{e}}_{\alpha\beta}(i)\cdot\hat{\mathbf{e}}(1), \quad (3.8)$$

describing the *first* scatterings of the incident wave.

Equation (3.4) represents the fundamental relation for studying the multiple scattering of waves. Expressions of this type seem to have been first proposed by Foldy[22] for study of the scattering of sound waves. Various applications were discussed by Lax.[17] A precise derivation from the wave equation was given by Watson.[11] Here it was shown that equations having the structure of Eqs. (3.4) represent *exact* solutions to the wave equation (in our case, Maxwell's equations for scattering by N electrons). Two approximations are made, however, in the form given by Eqs. (3.4). First, corrections to the Thomson amplitudes f_{ij} obtained from quantum electrodynamics are

[21] Assumption (C5) was not made in Ref. 13.

[22] L. L. Foldy, Phys. Rev. 67, 107 (1945).

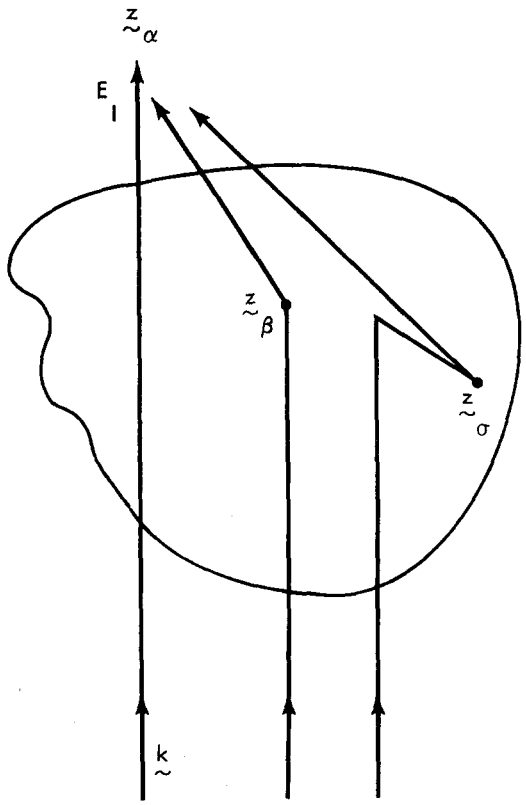


FIG. 2. Graphical illustration of terms in Eq. (3.9).

omitted. For our applications these are very small and justifiably neglected. Second, we have used assumption (B2) to treat each scattering as occurring in the wave zone of a previous scattering. This lets us use the Thomson amplitude and the scalar Green's function (3.5) rather than the exact tensor Green's function of Maxwell's equations. We emphasize that correct equations which include near-zone scattering are known[11,12] and have the form of Eqs. (3.4), but with the use of the exact Green's function. At the cost of some added algebraic complexity, the applications of this paper can be carried through for these corrected equations.

The structure of Eqs. (3.1) and (3.4) can perhaps be seen more clearly if we sequentially substitute the right-hand side into the left-hand side of (3.4) to obtain the series

$$\mathbf{E}(\mathbf{z}_\alpha) = \mathbf{E}_I(\mathbf{z}_\alpha) + \sum_{\beta(\neq\alpha)=1}^{N} \Bigg\{ \sum_{j=1}^{2} \hat{\mathbf{e}}_{\alpha\beta}(j) G^0_{\alpha\beta} f_{j1}(\alpha\beta, \beta 0) E_I(\mathbf{z}_\beta) + \sum_{j=1}^{2} \hat{\mathbf{e}}_{\alpha\beta}(j) \sum_{\sigma(\neq\beta)=1}^{N} \sum_{i=1}^{2} G^0_{\alpha\beta} f_{ji}(\alpha\beta, \beta\sigma) \times G^0_{\beta\sigma} f_{i1}(\beta\sigma, \sigma 0) E_I(\mathbf{z}_\sigma) + \cdots \Bigg\}. \tag{3.9}$$

The first term, of course, represents just the incident wave, while the second represents once-scattered waves. The third term describes twice-scattered waves, etc. These terms are illustrated graphically in Fig. 2.

For our applications a modified form of Eq. (3.4) will prove convenient. This modification results from a separate treatment of coherent and incoherent scatterings. In the quantum theory this is done with certain projection operators. In the present classical theory we accomplish this with a careful ordering of statistical averages, as in Eqs. (2.15) and (2.16).

What we wish ultimately to calculate is the power flux at some point $\mathbf{r}$, that is,

$$\mathcal{F}_{\text{rec}} = c \frac{\langle \hat{\mathbf{e}} \cdot \mathbf{E}^*(\mathbf{r}) \hat{\mathbf{e}} \cdot \mathbf{E}(\mathbf{r}) \rangle}{8\pi}, \tag{3.10}$$

where the notation of Eqs. (2.9) for the statistical average has been used. Now, each factor of $\hat{\mathbf{e}} \cdot \mathbf{E}$ in (3.10) can be thought to consist of a sum of terms describing sequences of scatterings, as in (3.9). Thus $\hat{\mathbf{e}} \cdot \mathbf{E}$ has the form

$$\hat{\mathbf{e}} \cdot \mathbf{E}(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{\alpha_1, \cdots, \alpha_n} Q_n(\mathbf{r}; \mathbf{z}_{\alpha_1}, \cdots, \mathbf{z}_{\alpha_n}).$$

A typical term in Eq. (3.10) will thus have the form

$$\begin{aligned} \mathcal{F}_{mn} &\equiv \langle Q^*_m(\mathbf{r}; \mathbf{z}_{\beta_1}, \cdots, \mathbf{z}_{\beta_m}) Q_m(\mathbf{r}; \mathbf{z}_{\alpha_1} \cdots \mathbf{z}_{\alpha_n}) \rangle \\ &= \int P_{m+n} Q^*_m Q_n \, d^3 z_{\beta_1} \cdots d^3 \mathbf{z}_{\alpha_n}. \end{aligned} \tag{3.11}$$

Our assumptions (B1) and (C5) let us suppose that the $\mathbf{z}_\alpha$'s are all distinct from the $\mathbf{z}_\beta$'s.

The probability distribution P_N may be developed in a series of correlation functions following the procedure of Eqs. (2.4)–(2.7). Because of assumption (B5) we expect the principal contributions to come from low-order correlations. For any given term of this series the $\mathbf{z}_{\beta_1} \cdots \mathbf{z}_{\alpha_n}$ will be grouped into clusters of correlated coordinates, there being no correlation between $\mathbf{z}$'s of different clusters.

The simplest example of this is provided by Eqs. (2.10)–(2.16) for the first Born terms. The next simplest example is given by the interference term resulting from a first times a second Born contribution. Reference to (3.9) shows that the resulting expression (3.11) is of the form

$$\mathcal{F}_{12} = \langle Q^*_1(\mathbf{r}; \mathbf{z}_1) Q_2(\mathbf{r}; \mathbf{z}_2, \mathbf{z}_3) \rangle. \tag{3.12}$$

The expression (2.7) for P_3 lets us write this as

$$\begin{aligned} \mathcal{F}_{12} = &\int P_1(\mathbf{z}_1) P_1(\mathbf{z}_2) P_1(\mathbf{z}_3) g_3(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3) \\ &\times Q^*_1 Q_2 \, d^3 z_1 \, d^3 z_2 \, d^3 z_3 + \langle Q^*_1 \rangle \langle Q_2 \rangle \\ &+ \int d^3 z_1 \, d^3 z_2 \, P_1(\mathbf{z}_1) P_1(\mathbf{z}_2) g_2(\mathbf{z}_1, \mathbf{z}_2) Q^*_1(\mathbf{r}; \mathbf{z}_1) \\ &\times \left[\int d^3 z_3 \, P_1(\mathbf{z}_3) Q_2(\mathbf{r}; \mathbf{z}_2, \mathbf{z}_3) \right] \\ &+ [\text{"2" replaced by "3" in } g_2]. \end{aligned} \tag{3.13}$$

The first term above corresponds to scattering with all three electrons within one correlation range of each other. According to assumption (B5), we expect this to be relatively small. The second term represents a contribution to the coherent scattering, which, by assumption (B2), is important only in the forward direction. The third and fourth terms are of the form which will be of principal interest to us, containing both *coherent* and *incoherent* contributions. For example, the average over $\mathbf{z}_3$ in the third term will be called a *coherent part* of the scattering, since $\mathbf{z}_3$ is uncorrelated with $\mathbf{z}_1$. The average over $\mathbf{z}_1$ and $\mathbf{z}_2$ will be called *cross correlated*, since it involves a correlation between coordinates in both Q's.

This classification can be applied to any term in (3.11). For example, an average over a correlated cluster $\mathbf{z}_{\alpha_a} \cdots \mathbf{z}_{\alpha_b}$ in Q_n, and thus not correlated with any of the coordinates $\mathbf{z}_{\beta_1} \cdots \mathbf{z}_{\beta_m}$ of Q_m, will be called a "coherent part" of the average. An average over a correlated cluster $\mathbf{z}_{\alpha_a} \cdots \mathbf{z}_{\alpha_b}$ invoking coordinates of both Q's will be called "cross correlated." Evidently, each average over a correlated cluster of coordinates will fall into one or the other of these classifications.

Our technique of separating coherent from incoherent contributions is first to evaluate all the coherent parts contributing to $F_{\alpha\beta}$ directly from Eqs. (3.4). This will lead to a modified set of multiple-scattering equations. In evaluating (3.10) we must then, of course, keep only *cross-correlated* averages.

As a first step in doing this, we note that coherent parts give rise to forward scattering only. This follows from assumptions (B4), (B6), and (C3).[23] For example,

$$\int P_1(\mathbf{z}_\beta) G^0_{\alpha\beta} f_{ij}(\alpha\beta, \beta\sigma) F_{\beta\sigma}(\mathbf{z}_\beta, j)\, d^3z_\beta \cong \delta_{ij} f_0 \int P_1(\mathbf{z}_\beta) G^0_{\alpha\beta} F_{\beta\sigma}(\mathbf{z}_\beta, j)\, d^3z_\beta, \quad (3.14)$$

where

$$f_0 = f_{ii}(\alpha\beta, \alpha\beta) = -r_0 \quad (3.15)$$

is the Thomson amplitude for forward scattering. The result (3.14) follows since the straight line path from $\mathbf{z}_\sigma$ to $\mathbf{z}_\alpha$ represents the path of *stationary phase* for the exponential

$$\exp\,[ik(R_{\alpha\beta} + R_{\beta\sigma})],$$

as is illustrated in Fig. 3. The verification of (3.14) may also be seen from our conclusion with Eq. (3.21) that the coherent parts give rise to the eikonal ray paths, which do not show deflections at individual scatterings.

As a first approximation, let us consider a sum over coherent parts in (3.4), neglecting any correlations between electrons. For propagation from $\mathbf{z}_\sigma$ to $\mathbf{z}_\alpha$ we obtain the series

$$G^{(1)}_{\alpha\sigma} \equiv G^0_{\alpha\sigma} + f_0 \sum_\beta \int G^0_{\alpha\beta} G^0_{\beta\sigma} P_1(\mathbf{z}_\beta)\, d^3z_\beta + f_0^2 \sum_{\beta_1,\beta_2} \int G^0_{\alpha\beta_1} G^0_{\beta_1\beta_2} G^0_{\beta_2\sigma} P_1(\mathbf{z}_{\beta_1}) \times P_1(\mathbf{z}_{\beta_2})\, d^3z_{\beta_1}\, d^3z_{\beta_2} + \cdots. \quad (3.16)$$

This may be summed into the form of an integral equation for $G^{(1)}_{\alpha\sigma}$:

$$G^{(1)}_{\alpha\sigma} = G^0_{\alpha\sigma} + f_0 \sum_\beta \int G^0_{\alpha\beta} G^{(1)}_{\beta\sigma} P_1(\mathbf{z}_\beta)\, d^3z_\beta,$$

or

$$G^{(1)}_{\alpha\sigma} = G^0_{\alpha\sigma} + f_0 \int G^0_{\alpha\beta} G^{(1)}_{\beta\sigma} \rho(\mathbf{z}_\beta)\, d^3z_\beta, \quad (3.17)$$

where we have introduced the electron density ρ defined by Eq. (2.13). We have also used assumption (B1) to neglect the restrictions implied by Eqs. (3.14) on the sums over electron coordinates.

To solve Eq. (3.17), we first observe that $G^0_{\alpha\sigma}$ satisfies the differential equation

$$[\nabla^2_\alpha + k^2] G^0_{\alpha\sigma} = -4\pi\delta(\mathbf{z}_\alpha - \mathbf{z}_\sigma).$$

Operating on both sides of Eq. (3.17) with $[\nabla^2_\alpha + k^2]$ and using this relation, we obtain

$$[\nabla^2_\alpha + k^2 n_1^2(\mathbf{z}_\alpha)] G^{(1)}_{\alpha\sigma} = -4\pi\delta(\mathbf{z}_\alpha - \mathbf{z}_\sigma), \quad (3.18)$$

where

$$n_1^2(\mathbf{z}) = 1 + (4\pi f_0/k^2)\rho(\mathbf{z}). \quad (3.19)$$

Thus, $G^{(1)}_{\alpha\sigma}$ satisfies the wave equation in a medium of refractive index $n_1(\mathbf{z})$. Assumption (B6) permits us to write $G^{(1)}_{\alpha\sigma}$ in the eikonal form

$$G^{(1)}_{\alpha\sigma} = e^{ikS_{\alpha\sigma}}/R_{\alpha\sigma}, \quad (3.20)$$

where the eikonal is

$$S_{\alpha\sigma} = \int_{\mathbf{z}_\sigma}^{\mathbf{z}_\alpha} n_1(\mathbf{x})\, ds, \quad (3.21)$$

and the integral is taken from $\mathbf{z}_\sigma$ to $\mathbf{z}_\alpha$ along a ray

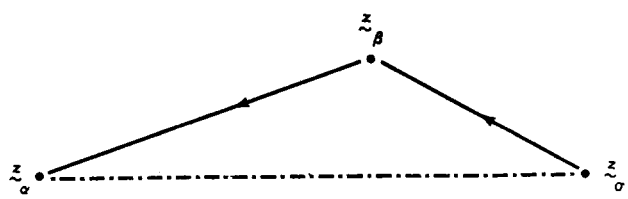


FIG. 3. Integration path for Eq. (3.14).

[23] As noted earlier, assumption (C3) is not necessary, but permits us to neglect rotation of the polarization vector along a ray path.

path.[24,25] Our assumption (C3) lets us take this ray path as a straight line. This conclusion provides a justification for Eq. (3.14).

We may evidently generalize Eq. (3.16) keeping pair correlations, etc. From the structure of the series (3.4) we see that this leads to the Green's function

$$G_{\alpha\sigma} = G^0_{\alpha\sigma} + f_0 \sum_\beta \int G^0_{\alpha\beta} G_{\beta\sigma} P_1(\mathbf{z}_\beta)\, d^3 z_\beta + \sum_{j=1}^{2} \sum_{\beta_1,\beta_2} \int G^0_{\alpha\beta_1} f_{ij}(\alpha\beta_1, \beta_1\beta_2) G_{\beta_1\beta_2} \times f_{ji}(\beta_1\beta_2, \beta_2\sigma) G_{\beta_2\sigma} P_1(\mathbf{z}_{\beta_1}) P_1(\mathbf{z}_{\beta_2}) \times g(\mathbf{z}_{\beta_1}, \mathbf{z}_{\beta_2})\, d^3 z_{\beta_1}\, d^3 z_{\beta_2} + \cdots, \quad (3.22)$$

where "$\cdots$" refers to contributions from higher-order correlations. In writing this equation we have anticipated [as in Eq. (3.14)] that $\mathbf{R}_{\alpha\beta_1}$ and $\mathbf{R}_{\beta_2\sigma}$ will be parallel, since coherent parts give rise to forward scattering only. (This may again be justified by direct calculation or from our final result.) Since there is no change of direction for the two scatterings, there will be no change of polarization. It is for this reason that the initial and final polarization indices "i" are taken to be the same [the isotropy assumption (C2) for the medium has also been invoked here].

We now address ourselves to the solution of Eq. (3.22), dropping the contribution of "higher-order-than-pair" correlations. To anticipate our final result, $G_{\alpha\sigma}$ will again have the form (3.20) and (3.21), except with a more accurate approximation to the refractive index. Given this anticipation and the fact that $\mathbf{z}_{\beta_1}$ and $\mathbf{z}_{\beta_2}$ are correlated (and thus separated by a distance comparable to R_c), we have

$$G_{\beta_2\sigma} \cong \exp\,[in_1(\mathbf{z}_{\beta_1})\mathbf{k}_{\beta_1\sigma} \cdot \mathbf{R}_{\beta_2\beta_1}] G_{\beta_1\sigma} \quad (3.23)$$

and

$$G_{\beta_1\beta_2} = (1/R_{\beta_1\beta_2}) \exp\,[in_1(\mathbf{z}_{\beta_1}) kR_{\beta_1\beta_2}]. \quad (3.24)$$

Here we have used the notation of Eqs. (3.3) and (3.6) and have set $\mathbf{k}_{\beta_1\sigma} \equiv k\hat{\mathbf{k}}_{\beta_1\sigma}$. We have also used assumptions (C1) and (C4) to simplify the eikonal phases. Finally, we have replaced the exact refractive index n by its first approximation n_1, the relative error being of order (R_c/l) [see Eq. (4.7)].

The approximations (3.23) and (3.24) permit us to rewrite Eq. (3.22) in the form

$$G_{\alpha\sigma} = G^0_{\alpha\sigma} + \int G^0_{\alpha\beta} \gamma(\mathbf{z}_\beta) G_{\beta\sigma}\, d^3 z_\beta, \quad (3.25)$$

where

$$\gamma(\mathbf{z}) = f_0\rho(\mathbf{z}) + \rho(\mathbf{z}) \int d^3x\, \rho(\mathbf{x}) g(\mathbf{z}, \mathbf{x}) \times \frac{e^{in_1(\mathbf{z})kR}}{R} e^{-in_1(\mathbf{z})k\hat{\mathbf{q}}\cdot\mathbf{R}} \sum_{j=1}^{2} [f_{ji}(\hat{\mathbf{R}}, \hat{\mathbf{q}})]^2. \quad (3.26a)$$

Here we have introduced the unit vector

$$\hat{\mathbf{q}} \equiv \mathbf{R}_{\beta_1\sigma}/R_{\beta_1\sigma},$$

written $\mathbf{R} \equiv \mathbf{z} - \mathbf{x}$, and have returned to the notation (2.4) for the Thomson amplitude. Approximations (C1) and (C2) allow us to rewrite $\gamma(\mathbf{z})$ as

$$\gamma(\mathbf{z}) = f_0\rho(\mathbf{z}) + \rho^2(\mathbf{z}) \int d^3R\, g(\mathbf{z}; R) \times \frac{1}{R} \exp\,[in_1(\mathbf{z})(kR - k\hat{\mathbf{q}} \cdot \mathbf{R})] \times \sum_{j=1}^{2} [f_{ji}(\hat{\mathbf{R}}, \hat{\mathbf{q}})]^2. \quad (3.26b)$$

This expression is independent of the vector $\hat{\mathbf{q}}$ and the polarization index i.

On following the steps which led from Eq. (3.17) to (3.18), we see that $G_{\alpha\sigma}$ satisfies the equation

$$[\nabla^2_\alpha + k^2 n^2(\mathbf{z}_\alpha)] G_{\alpha\sigma} = -4\pi\delta(\mathbf{z}_\alpha - \mathbf{z}_\sigma), \quad (3.27)$$

where

$$n^2(\mathbf{z}) = 1 + (4\pi/k^2)\gamma(\mathbf{z}). \quad (3.28)$$

Therefore, in the eikonal approximation we have

$$G_{\alpha\sigma} = (1/R_{\alpha\sigma}) e^{ikS_{\alpha\sigma}},$$
$$S_{\alpha\sigma} \equiv S(\mathbf{z}_\alpha, \mathbf{z}_\sigma) = \int_{\mathbf{z}_\sigma}^{\mathbf{z}_\alpha} n(\mathbf{x})\, ds = S_{\sigma\alpha}, \quad (3.29)$$

as in Eqs. (3.20) and (3.21). Were we to drop the restriction (C3), the path integral here would, in general, follow a curved ray path and $G_{\alpha\sigma}$ would be a tensor in the polarization indices.

Making use of our assumption (C3) that $n \cong 1$ and our conclusion that (3.26b) does not depend on the polarization index i, we can re-express (3.28) in the form

$$n(\mathbf{z}) \cong 1 + (2\pi f_0/k^2)\rho(\mathbf{z}) + (2\pi/k^2)\rho^2(\mathbf{z}) \int d^3R\, g(\mathbf{z}; R) \sigma_T(\hat{\mathbf{R}} \cdot \hat{\mathbf{q}}) \times (1/R) \exp\,[in_1(\mathbf{z}) k(R - \hat{\mathbf{q}} \cdot \mathbf{R})], \quad (3.30)$$

where σ_T is defined by Eq. (2.21).

The procedure used to derive Eqs. (3.29) and (3.30) can evidently be generalized to include the contribution from higher-order correlations in the refractive index. We do not do this explicitly here.

[24] See, for example, M. Born and E. Wolf, *Principles of Optics* (The Macmillan Company, New York, 1964).

[25] For ray paths which are sufficiently curved, the factor $R^{-1}_{\alpha\sigma}$ in (3.20) must be modified. For the correct expression, see J. Chen and K. M. Watson, Phys. Rev. **174**, 152 (1968).

We are now ready to rewrite Eqs. (3.1) and (3.4), taking due account of contributions from "coherent parts." The new equations are

$$\mathbf{E}(\mathbf{z}_\alpha) = \mathbf{E}_c(\mathbf{z}_\alpha) + \sum_{\beta(\neq\alpha)=1}^{N} \sum_{j=1}^{2} \hat{\mathbf{e}}_{\alpha\beta}(j) E_{\alpha\beta}(j) \quad (3.31)$$

and

$$E_{\alpha\beta}(i) = G_{\alpha\beta} f_{i1}(\alpha\beta, \beta\sigma) E_c(\mathbf{z}_\beta) + \sum_{\sigma(\neq\beta)=1}^{N} \sum_{j=1}^{2} G_{\alpha\beta} f_{ij}(\alpha\beta, \beta\sigma) E_{\beta\sigma}(j). \quad (3.32)$$

The *coherent wave* $\mathbf{E}_c(\mathbf{z}_\alpha)$ is defined by the relation [we consider $\mathbf{z}_\alpha$ as not being an electron coordinate in the defining Eq. (3.33)]:

$$\mathbf{E}_c(\mathbf{z}_\alpha) \equiv \langle \mathbf{E}(\mathbf{z}_\alpha) \rangle \equiv \langle \mathbf{E}(\mathbf{z}_\alpha) \rangle_\alpha = e^{ikS_\alpha} E_0 \hat{\mathbf{e}}(1) \equiv \hat{\mathbf{e}}(1) E_c(\mathbf{z}_\alpha), \quad (3.33)$$

where the eikonal is

$$S_\alpha \equiv S(\mathbf{z}_\alpha) = \int_{-\infty}^{\mathbf{z}_\alpha} [n(\mathbf{x}) - 1]\, ds + \hat{\mathbf{k}} \cdot \mathbf{z}_\alpha, \quad (3.34)$$

with the path integral taken along a straight line parallel to $\hat{\mathbf{k}}$ and terminating at $\mathbf{z}_\alpha$. The constant of integration in (3.34) is chosen to be consistent with that of Eq. (2.1).

In using these equations to calculate a power flux, etc., as in Eq. (3.10), we must of course keep only *cross correlations*—and not twice take account of coherent parts.

Let us first verify Eq. (3.33). Using Eqs. (3.1) and (3.4), we have

$$\begin{aligned} E_c(\mathbf{z}_\alpha) &= E_I(\mathbf{z}_\alpha) + f_0 \sum_\beta \int P_1(\mathbf{z}_\beta) G^0_{\alpha\beta} E_I(\mathbf{z}_\beta)\, d^3 z_\beta \\ &\quad + \sum_{\beta_1,\beta_2} f_0^2 \int P_1(\mathbf{z}_{\beta_1}) P_1(\mathbf{z}_{\beta_2}) G^0_{\alpha\beta_1} G^0_{\beta_1\beta_2} \\ &\quad \times E_I(\mathbf{z}_{\beta_2})\, d^3 z_{\beta_1}\, d^3 z_{\beta_2} + \cdots \\ &= E_I(\mathbf{z}_\alpha) + \int G_{\alpha\beta\gamma}(\mathbf{z}_\beta) E_I(\mathbf{z}_\beta)\, d^3 z_\beta. \end{aligned} \quad (3.35)$$

Use of Eqs. (2.1) and (3.27) shows that E_c satisfies the equation

$$[\nabla_\alpha^2 + k^2 n^2(\mathbf{z}_\alpha)] E_c(\mathbf{z}_\alpha) = 0. \quad (3.36)$$

This and the boundary condition implied by Eq. (2.1) lead to Eqs. (3.33) and (3.34).

To derive Eqs. (3.32), we return to Eqs. (3.4). Let $\mathbf{z}_\alpha$ and $\mathbf{z}_\beta$ be coordinates at which two successive incoherent (i.e., to be cross correlated) scatterings occur. We sum over all possible coherent scatterings between these two points to obtain the series (3.22), and thus obtain $G_{\alpha\beta}$ as the Green's function describing propagation between these two scatterings. When all scatterings in the series (except that at $\mathbf{z}_\beta$) are coherent, we obtain the first term on the right-hand side of Eq. (3.32).

4. APPLICATIONS OF THE MULTIPLE-SCATTERING EQUATIONS

In Sec. 5 we shall derive a transport equation for the scattered intensity from Eqs. (3.31) and (3.32). In this section we describe only the simplest application of the multiple-scattering equations.

Relatively little effort has been made to solve Eqs. (3.1) and (3.4) or (3.31) and (3.32) directly, except when the number N of scatterers is small. Available variational principles probably have only a very limited applicability, but do not seem to have been explored. Approximations in which the "scatterers" are volume cells or correlation cells chosen so that N is effectively small are possible, but also not explored. Restriction to a finite set of possible ray paths suggests another class of approximations to the multiple-scattering equations.

We make no attempt to explore methods for solving the multiple-scattering equations here. To illustrate the difference between the set (3.1)–(3.4) and the set (3.31)–(3.33), we shall discuss the single-scattering approximation to the latter. In the quantum theory of scattering this is usually called the "distorted-wave Born approximation" (or DWBA). The DWBA expression is obtained by substituting the first term on the right-hand side of Eq. (3.32) into (3.31). This gives the incoherently scattered field at a large distance $\mathbf{r}$ as

$$\begin{aligned} \mathbf{E}_{\rm sc}(\mathbf{r}) &= \sum_{j=1}^{2} \hat{\mathbf{e}}_{\hat{\mathbf{p}}}(j) E_{\rm sc}(j), \\ E_{\rm sc}(j) &= \sum_{\alpha=1}^{N} \frac{1}{r} \exp\{ik[S(\mathbf{r}, \mathbf{z}_\alpha) + S(\mathbf{z}_\alpha)]\} f_{j1}(\hat{\mathbf{p}}, \hat{\mathbf{k}}) E_0, \end{aligned} \quad (4.1)$$

where the notation of Eqs. (2.2) and (2.3) has been used. The expression for the received power flux is now

$$\begin{aligned} \mathfrak{F}_{\rm rec} &= c \left\langle \frac{|\hat{\mathbf{e}} \cdot \mathbf{E}_{\rm sc}(\mathbf{r})|^2}{8\pi} \right\rangle \\ &= (\mathfrak{F}_0 / r^2) \sigma, \end{aligned}$$

where the scattering cross section in the DWBA is

$$\sigma = |f(\hat{\mathbf{e}})|^2 \int d^3 z_1\, d^3 z_2\, \rho(\mathbf{z}_1) \rho(\mathbf{z}_2) g(\mathbf{z}_1, \mathbf{z}_2) \times \exp\{ik[S(\mathbf{r}, \mathbf{z}_1) - S^*(\mathbf{r}, \mathbf{z}_2) + S(\mathbf{z}_1) - S^*(\mathbf{z}_2)]\}. \quad (4.2)$$

When the refractive index is unity, this, of course, reduces to Eq. (2.17a).

Assumptions (C1) and (C4) permit us to write

$$S(\mathbf{r}, \mathbf{z}_2) \cong S(\mathbf{r}, \mathbf{z}_1) + n_1(\mathbf{z}_1)\hat{\mathbf{p}} \cdot \mathbf{R},$$
$$S(\mathbf{z}_2) \cong S(\mathbf{z}_1) - n_1(\mathbf{z}_1)\hat{\mathbf{k}} \cdot \mathbf{R}, \tag{4.3}$$

where $\mathbf{R} \equiv \mathbf{z}_1 - \mathbf{z}_2$. This lets us write

$$\sigma = |f(\hat{\mathbf{e}})|^2 \int d^3z\, \rho^2(\mathbf{z}) \times \exp\{-2k[\operatorname{Im} S(\mathbf{r}, \mathbf{z}) + \operatorname{Im} S(\mathbf{z})]\} \times d^3R\, g(\mathbf{z}; R) \exp\{in_1(\mathbf{z})(\mathbf{p} - \mathbf{k}) \cdot \mathbf{R}\}. \tag{4.4}$$

If we sum this over final polarizations $\hat{\mathbf{e}}$ and scattering directions $\hat{\mathbf{p}}$, and average over initial polarizations $\hat{\mathbf{e}}(1)$, for the scattering per unit volume we obtain

$$\frac{1}{l(\mathbf{z})} = \rho^2(\mathbf{z}) \int d\Omega_{\hat{\mathbf{p}}} \int d^3R g(\mathbf{z}; R) \times \exp\{in_1(\mathbf{z})(\mathbf{p} - \mathbf{k}) \cdot \mathbf{R}\} \times \sigma_T(\hat{\mathbf{p}} \cdot \hat{\mathbf{k}}). \tag{4.5}$$

This differs from the expression (2.23) only by the appearance of $n_1(\mathbf{z})$ in the exponential. Equation (4.5) provides the expression for the scattering mean-free-path in the DWBA.

The expression above may be compared with the quantity $2k \operatorname{Im} n(\mathbf{z})$ obtained from Eq. (3.30). This is

$$2k \operatorname{Im} n(\mathbf{z}) = 4\pi\rho^2(\mathbf{z}) \int d^3R g(\mathbf{z}; R) \sigma_T(\hat{\mathbf{R}} \cdot \hat{\mathbf{q}}) \times \frac{\sin n_1 kR}{kR} e^{-in_1 \mathbf{R} \cdot \hat{\mathbf{q}}}, \tag{4.6}$$

since $\sigma_T(\hat{\mathbf{R}} \cdot \hat{\mathbf{q}})$ is even in $\hat{\mathbf{R}}$. Now because we have assumed successive scatterings to lie in the wave zone, we have

$$\frac{4\pi \sin n_1 kR}{n_1 kR} \sigma_T(\hat{\mathbf{R}} \cdot \hat{\mathbf{q}}) \cong \int d\Omega_{\hat{\mathbf{p}}}\, e^{in_1 \mathbf{p} \cdot \mathbf{R}} \sigma_T(\hat{\mathbf{p}} \cdot \hat{\mathbf{q}}).$$

Recalling that $n_1 \cong 1$, we may use this to rewrite (4.6) as

$$2k \operatorname{Im} n(\mathbf{z}) \cong \rho^2(\mathbf{z}) \int d\Omega_{\hat{\mathbf{p}}} \int d^3 R g(\mathbf{z}; R) \sigma_T(\hat{\mathbf{p}} \cdot \hat{\mathbf{q}}) \times \exp[in_1(\mathbf{p} - k\hat{\mathbf{q}}) \cdot \mathbf{R}] = 1/l(\mathbf{z}), \tag{4.7}$$

where $l(\mathbf{z})$ is given by Eq. (4.5). This result is, of course, expected from more general considerations.

The relation (4.7) lets us put (4.4) into an intuitively more plausible form:

$$\sigma = |f(\hat{\mathbf{e}})|^2 \int d^3\, z \rho^2(\mathbf{z}) \exp\left[-\int_{-\infty}^{z} \frac{ds}{l(x)} - \int_{z}^{r} \frac{ds}{l(\mathbf{x})}\right] \times \int d^3\, R g(\mathbf{z}; R) l^{i(\mathbf{p}-\mathbf{k}) \cdot \mathbf{R} n_1(\mathbf{z})}, \tag{4.8}$$

where the path integrals are taken along the incident and outgoing ray paths. This shows that scattering in the interior of the plasma is reduced by absorption of the incident beam. The simple incoherent scattering cross section is further reduced by absorption of the outgoing beam. This latter reduction would, of course, be compensated if we were to include more than one incoherent scattering.[26]

5. USE OF A TRANSPORT EQUATION

Instead of attempting to solve the multiple scattering equations (3.1) and (3.4) [or (33.1) and (3.32)] directly, one can use these to derive a transport equation for the power flux. In the present section we show how to do this. The simplifying assumptions of Sec. 2 is accepted, so no attempt at great generality is made at present.

We first introduce the energy density[27]

$$U(\mathbf{z}_\alpha; \hat{\mathbf{e}}) \equiv \left\langle \frac{|\hat{\mathbf{e}} \cdot \mathbf{E}(\mathbf{z}_\alpha)|^2}{8\pi} \right\rangle_\alpha \tag{5.1}$$

and the coherent density

$$U_c(\mathbf{z}_\alpha; \hat{\mathbf{e}}) \equiv \frac{|\hat{\mathbf{e}} \cdot \mathbf{E}_c(\mathbf{z}_\alpha)|^2}{8\pi} = U_c(\mathbf{z}_\alpha)[\hat{\mathbf{e}} \cdot \hat{\mathbf{e}}(1)]^2. \tag{5.2}$$

Here $\mathbf{z}_\alpha$ is considered to be some point in space, rather than an electron coordinate. If we take $\mathbf{z}_\alpha = \mathbf{r}$, a point far from the plasma, then $cU(\mathbf{r}, \hat{\mathbf{e}})$ is the scattered point flux received at $\mathbf{r}$ with linear polarization $\hat{\mathbf{e}}$. (We recall that when $\mathbf{z}_\alpha$ is not an electron coordinate, we need not distinguish between the averages $\langle \cdots \rangle_\alpha$ and $\langle \cdots \rangle$.)

Substitution of Eq. (3.31) into (5.1) gives

$$U(\mathbf{z}_\alpha; \hat{\mathbf{e}}) = U_c(\mathbf{z}_\alpha; \hat{\mathbf{e}}) + \sum_{\beta, \beta'} \sum_{i,j=1}^{2} \left\langle \hat{\mathbf{e}} \cdot \hat{\mathbf{e}}_{\alpha\beta}(i) \hat{\mathbf{e}} \cdot \hat{\mathbf{e}}_{\alpha\beta'}(j) \frac{[E^*_{\alpha\beta}(i) E_{\alpha\beta'}(j)]}{8\pi} \right\rangle_\alpha, \tag{5.3}$$

where the cross terms vanish. That is,

$$\langle E_{\alpha\beta'}(j) \rangle_\alpha = 0, \tag{5.4}$$

because $E_{\alpha\beta'}$ must contain cross correlations. [This particular conclusion (5.4) is obvious from the

[26] It is shown in Ref. 12 that the term $[\int_z^r ds/l(\mathbf{x})]$ in the exponent in Eq. (4.8) does not occur when one calculates the total scattering in the DWBA.

[27] The units chosen for Eqs. (5.1), (5.2), etc., are irrelevant, since our final equations are homogeneous in either energy density or in power flux. We have taken the normalization to be consistent with unrationalized Gaussian units, which are usually encountered in atomic physics. The photon density might also seem natural in a transport theory. Also, instead of the expressions (5.1) and (5.2), one might prefer to develop the theory of this section for the full energy momentum and stress tensor.

definition (3.33) of $\mathbf{E}_c$.] The second term in Eq. (5.3) can be written in the form

$$\int d^3z_\beta\, d^3z_{\beta'}\, \hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_{\alpha\beta}(i)\hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_{\alpha\beta'}(j)\rho(\mathbf{z}_\beta)\rho(\mathbf{z}_{\beta'}) \times [1 + g(\mathbf{z}_\beta, \mathbf{z}_{\beta'})]\left\langle \frac{E^*_{\alpha\beta}(i)E_{\alpha\beta'}(j)}{8\pi}\right\rangle_{\alpha\beta\beta'}$$
$$\cong \int d^3z_\beta\, \hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_{\alpha\beta}(i)\hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_{\alpha\beta}(j)\rho(\mathbf{z}_\beta) \times \int d^3z_{\beta'}\, \rho(\mathbf{z}_{\beta'})g(\mathbf{z}_\beta, \mathbf{z}_{\beta'})\left\langle \frac{E^*_{\alpha\beta}(i)E_{\alpha\beta'}(j)}{8\pi}\right\rangle_{\alpha\beta\beta'}. \quad (5.5)$$

To understand the above equation we first note that the averages above must contain no *coherent parts*. The "one" term in $[1 + g(\mathbf{z}_\beta, \mathbf{z}_{\beta'})]$ was dropped because this corresponds to a coherent part, except for possible higher-order correlations involving $\mathbf{z}_\beta, \mathbf{z}_{\beta'}$, and at least one other electron coordinate. Such higher-order correlations may be neglected because of assumption (B5).[28] To see this, we note that the leading one of these neglected terms is of the form [here use set $\mathbf{z}_\alpha = \mathbf{r}$, a distant point, and replace the f's in Eq. (3.32) by r_0]:

$$\frac{\mathfrak{I}_0}{r^2}\int d^3z_\beta\, d^3z_{\beta'}\, d^3z_\sigma\, \rho(\mathbf{z}_\beta)\rho(\mathbf{z}_{\beta'})\rho(\mathbf{z}_\sigma) \times g_3(\mathbf{z}_\beta, \mathbf{z}_{\beta'}, \mathbf{z}_\sigma)[r_0 e^{-i(\mathbf{p}-\mathbf{k})\cdot\mathbf{z}_\beta}]^* \times \left[r_0\frac{e^{ikR_{\beta'\sigma}}}{R_{\beta'\sigma}}r_0 e^{i(\mathbf{k}\cdot\mathbf{z}_\sigma - \mathbf{p}\cdot\mathbf{z}_{\beta'})}\right]$$
$$\approx (\mathfrak{I}_0/r^2)[r_0^2(\rho R_c^3)N] \times [r_0 R_c^2\rho].$$

In obtaining this, we have set all oscillating exponentials equal to unity, which in general provides a gross overestimate. The inequality (2.24) shows that this is less than the "leading terms," as characterized by Eqs. (2.14) and (2.17), by a factor of $\mathcal{O}(R_c/l)^{\frac{1}{2}}$—or smaller, if the neglected exponentials are included.

We have also replaced $\hat{\mathbf{e}}_{\alpha\beta'}(j)$ by $\hat{\mathbf{e}}_{\alpha\beta}(j)$ in Eq. (5.5). This follows from assumption (B5), since for our applications $R_{\alpha\beta} = \mathcal{O}(l)$, while $R_{\beta'\beta} = \mathcal{O}(R_c)$.

The form of Eq. (5.5) suggests that we define

$$U_{ij}(\alpha, \beta) = \int d^3z_\gamma\, d^3z_\sigma\, \delta[\tfrac{1}{2}(\mathbf{z}_\gamma + \mathbf{z}_\sigma) - \mathbf{z}_\beta] \times \rho(\mathbf{z}_\gamma)\rho(\mathbf{z}_\sigma)g(\mathbf{z}_\gamma, \mathbf{z}_\sigma)\left\langle \frac{E^*_{\alpha\gamma}(i)E_{\alpha\sigma}(j)}{8\pi}\right\rangle_{\alpha\gamma\sigma}, \quad (5.6)$$

where $\mathbf{z}_\beta$ is now considered to be a point in space rather than an electron coordinate. When this is inserted into (5.5) and then into (5.3), we get

$$U(z_\alpha; \hat{\mathbf{e}}) = U_c(\mathbf{z}_\alpha; \hat{\mathbf{e}}) + \sum_{i,j=1}^{2}\int d^3z_\beta \hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_{\alpha\beta}(i)\hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_{\alpha\beta}(j) \times U_{ij}(\alpha, \beta). \quad (5.7a)$$

Thus, the integrand in the second term above represents the density of energy at $\mathbf{z}_\alpha$, having polarization $\hat{\mathbf{e}}$, arriving from within a unit volume at $\mathbf{z}_\beta$.

We can convert this into a flux as follows: let $\hat{\mathbf{p}}$ be a fixed unit vector and let

$$I_{ij}(\mathbf{z}_\alpha, \hat{\mathbf{p}}) \equiv I_c(\mathbf{z}_\alpha)\delta_{i1}\,\delta_{j1}\,\delta_{\hat{\mathbf{k}},\hat{\mathbf{p}}} + c\int_{-\hat{\mathbf{p}}} R_{\alpha\beta}^2\, dR_{\alpha\beta}U_{ij}(\alpha, \beta). \quad (5.8a)$$

Here we have introduced

$$I_c(\mathbf{z}_\alpha) = cU_c(\mathbf{z}_\alpha),$$

the flux of coherent power, and the Dirac delta function $\delta_{\mathbf{k},\hat{\mathbf{p}}} = \delta_{\hat{\mathbf{p}},\hat{\mathbf{k}}}$, defined by

$$\int d\Omega_{\hat{\mathbf{p}}}A(\hat{\mathbf{p}})\delta_{\hat{\mathbf{k}},\hat{\mathbf{p}}} = A(\hat{\mathbf{k}}),$$

where $A(\hat{\mathbf{p}})$ is some function of the unit vector $\hat{\mathbf{p}}$ continuous at $\hat{\mathbf{k}} = \hat{\mathbf{p}}$.[29] The integral in (5.8a) is taken along the straight line defined by $-\hat{\mathbf{p}} = \mathbf{R}_{\beta\alpha}/R_{\alpha\beta}$.

Equation (5.8a) can be written more concisely as the matrix equation

$$I(\mathbf{z}_\alpha, \hat{\mathbf{p}}) = \mathbf{I}_c(\mathbf{z}_\alpha)\delta_{\hat{\mathbf{k}},\hat{\mathbf{p}}} + c\int_{\hat{\mathbf{p}}} R_{\alpha\beta}^2\, dR_{\alpha\beta}\,\mathbf{U}(\alpha, \beta), \quad (5.8b)$$

where we consider $\mathbf{I}$, $\mathbf{I}_c$, and $\mathbf{U}$ to be column matrices with four elements.

On comparing Eqs. (5.7a) and (5.8a), we see that

$$U(\mathbf{z}_\alpha; \hat{\mathbf{e}}) = \frac{1}{c}\int d\Omega_{\hat{\mathbf{p}}}\sum_{i,j=1}^{2}\hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_{\alpha\beta}(i)\hat{\mathbf{e}}\cdot\hat{\mathbf{e}}_{\alpha\beta}(j)I_{ij}(\mathbf{z}_\alpha, \hat{\mathbf{p}}). \quad (5.7b)$$

Evidently,

$$I_{ii}(\mathbf{z}_\alpha, \hat{\mathbf{p}})\, d\Omega_{\hat{\mathbf{p}}}$$

represents the power per unit area of radiation having polarization "i" and propagating parallel to $\hat{\mathbf{p}}$, within the cone of solid angle $d\Omega_{\hat{\mathbf{p}}}$.

We are now ready to derive an equation for the flux $\mathbf{I}(\mathbf{z}, \hat{\mathbf{p}})$. When $R_{\alpha'\alpha} = \mathcal{O}(R_c)$ and $R_{\alpha\beta} = \mathcal{O}(l)$, assumptions (C1) and (C4) imply that, as in Eq. (3.23),

$$G_{\alpha'\beta} \cong \exp[in_1(\mathbf{z}_\alpha)\mathbf{k}_{\alpha\beta}\cdot\mathbf{R}_{\alpha'\alpha}] \times G_{\alpha\beta}, \quad (5.9a)$$

where $\mathbf{k}_{\alpha\beta} = k\hat{\mathbf{k}}_{\alpha\beta}$ [Eq. (3.3)] and we have again replaced n by n_1. On the other hand, when $R_{\beta'\beta} = \mathcal{O}(R_c)$ and $R_{\alpha\beta} = \mathcal{O}(l)$, we have

$$G_{\alpha\beta'} \cong \exp[-in_1(\mathbf{z}_\beta)\mathbf{k}_{\alpha\beta}\cdot\mathbf{R}_{\beta'\beta}] \times G_{\alpha\beta}. \quad (5.9b)$$

[28] An exception occurs for the singular case of radar backscatter. This actually causes no difficulty, as is shown in Sec. 7.

[29] In spherical coordinates $\delta(\mathbf{k} - \mathbf{p}) = (1/k^2)\delta(k - p)\delta_{\hat{\mathbf{k}},\hat{\mathbf{p}}}$. A representation for $\delta_{\hat{\mathbf{k}},\hat{\mathbf{p}}}$ is given on p. 346 of Ref. 12.

Use of these expressions in Eqs. (3.32) and (3.33) permits us to conclude that when $R_{\alpha'\alpha} = \mathcal{O}(R_c)$,

$$E_{\alpha'\beta}(i) \cong \exp\,[in_1(\mathbf{z}_\alpha)\mathbf{k}_{\alpha\beta}\cdot\mathbf{R}_{\alpha'\alpha}] \times E_{\alpha\beta}(i). \quad (5.10a)$$

When $R_{\beta'\beta} = \mathcal{O}(R_c)$, we have instead

$$\begin{aligned} E_{\alpha\beta'}(j) \cong{}& \exp\,[in_1(\mathbf{z}_\beta)(\mathbf{k}-\mathbf{k}_{\alpha\beta})\cdot\mathbf{R}_{\beta'\beta}] \\ &\times G_{\alpha\beta}f_{j1}(\alpha\beta,\beta\sigma)E_c(\mathbf{z}_\beta) \\ &+ \sum_{\sigma(\neq\beta)=1}^{N}\sum_{t=1}^{2}\exp\,[in_1(\mathbf{z}_\beta)(\mathbf{k}_{\beta\sigma}-\mathbf{k}_{\alpha\beta})\cdot\mathbf{R}_{\beta'\beta}] \\ &\times G_{\alpha\beta}f_{jt}(\alpha\beta,\beta\sigma)E_{\beta\sigma}(t). \end{aligned} \quad (5.10b)$$

Here we have used assumption (B5) to set

$$f_{jt}(\alpha\beta',\beta'\sigma) \cong f_{jt}(\alpha\beta,\beta\sigma).$$

We have also appropriately replaced the restriction $\sigma \neq \beta'$ by $\sigma \neq \beta$ when $R_{\alpha\beta'}^{-1}$ was replaced by $R_{\alpha\beta}^{-1}$ in G.

Now, let us substitute the right-hand side of Eq. (3.32) for each factor of E in (5.6). If we make use of Eqs. (5.10), we have

$$\begin{aligned} U_{ij}(\alpha,\beta) = {}& \int d^3z_\gamma\, d^3z_{\gamma'}\,\delta[\tfrac{1}{2}(\mathbf{z}_\gamma+\mathbf{z}_{\gamma'})-\mathbf{z}_\beta] \\ &\times \rho(\mathbf{z}_\gamma)\rho(\mathbf{z}_{\gamma'})g(\mathbf{z}_\gamma,\mathbf{z}_{\gamma'})\,|G_{\alpha\beta}|^2 \\ &\times \Big\{[f_{i1}(\alpha\beta,\beta o)f_{j1}(\alpha\beta,\beta o)]U_c(\mathbf{z}_\beta) \\ &\times \exp\,[in_1(\mathbf{z}_\beta)(\mathbf{k}_{\alpha\beta}-\mathbf{k})\cdot(\mathbf{R}_{\gamma\beta}-\mathbf{R}_{\gamma'\beta})] \\ &+ \sum_{t,t'=1}^{2}\int d^3z_\sigma\, d^3z_{\sigma'}\,\rho(\mathbf{z}_\sigma)\rho(\mathbf{z}_{\sigma'})g(\mathbf{z}_\sigma,\mathbf{z}_{\sigma'}) \\ &\times [f_{it}(\alpha\beta,\beta\sigma)f_{jt'}(\alpha\beta,\beta\sigma)] \\ &\times \exp\,[in_1(\mathbf{z}_\beta)(\mathbf{k}_{\alpha\beta}-\mathbf{k}_{\beta\sigma})\cdot\mathbf{R}_{\gamma\gamma'}] \\ &\times \left\langle\frac{E^*_{\beta\sigma}(t)E_{\beta\sigma'}(t')}{8\pi}\right\rangle_{\beta\sigma\sigma'} + \text{cross terms}\Big\}, \end{aligned} \quad (5.11)$$

where the "cross terms" are

$$\begin{aligned} &\frac{1}{8\pi}\int d^3z_\gamma\, d^3z_{\gamma'}\,\delta[\tfrac{1}{2}(\mathbf{z}_\gamma+\mathbf{z}_{\gamma'})-\mathbf{z}_\beta] \\ &\quad\times \rho(z_\gamma)\rho(\mathbf{z}_{\gamma'})g(\mathbf{z}_\gamma,\mathbf{z}_{\gamma'})\sum_{\sigma,t}E^*_c(\mathbf{z}_\gamma) \\ &\quad\times G^*_{\alpha\gamma}G_{\alpha\gamma'}f_{i1}(\alpha\beta,\beta o)\langle f_{jt}(\alpha\beta,\beta\sigma)E_{\gamma'\sigma}(t)\rangle_{\alpha\gamma\gamma'} \\ &\quad+ \text{complex conjugate}. \end{aligned} \quad (5.12)$$

According to assumption (B5), we expect $\mathbf{z}_\sigma$ to be uncorrelated with $\mathbf{z}_{\gamma'}$ or $\mathbf{z}_\gamma$. The average above thus contains "coherent parts." We thus omit the "cross terms" from Eq. (5.11).

To simplify Eq. (5.11) let us define the 4×4 matrix:

$$\begin{aligned} (ij|\,M(\alpha\beta,\beta\sigma)\,|tt') \equiv{}& [f_{it}(\alpha\beta,\beta\sigma)f_{jt'}(\alpha\beta,\beta\sigma)] \\ &\times \int d^3z_\gamma\, d^3z_{\gamma'}\,\delta[\tfrac{1}{2}(\mathbf{z}_\gamma+\mathbf{z}_{\gamma'})-\mathbf{z}_\beta] \\ &\times \rho(\mathbf{z}_\gamma)\rho(\mathbf{z}_{\gamma'})g(\mathbf{z}_\gamma,\mathbf{z}_{\gamma'}) \\ &\times \exp\,[in_1(\mathbf{z}_\beta)(\mathbf{k}_{\alpha\beta}-\mathbf{k}_{\beta\sigma})\cdot\mathbf{R}_{\gamma\gamma'}]. \end{aligned} \quad (5.13)$$

This and the expression (5.6) allow us to write (5.11) as

$$\begin{aligned} U_{ij}(\alpha,\beta) = {}& |G_{\alpha\beta}|^2\,\{(ij|\,M(\alpha\beta,\beta o)\,|11)U_c(\mathbf{z}_\beta)\} \\ &+ \sum_{t,t'=1}^{2}\int d^3z_\sigma\,(ij|\,M(\alpha\beta,\beta\sigma)\,|tt')U_{tt'}(\beta,\sigma). \end{aligned} \quad (5.14a)$$

In the more compact matrix notation of Eq. (5.8b) this is

$$\begin{aligned} \mathbf{U}(\alpha,\beta) = |G_{\alpha\beta}|^2\Big\{&\frac{1}{c}\,\mathbf{M}(\alpha\beta,\beta o)\mathbf{I}_c(\mathbf{z}_\beta) \\ &+ \int d^3z_\sigma\,\mathbf{M}(\alpha\beta,\beta\sigma)\mathbf{U}(\beta,\sigma)\Big\}. \end{aligned} \quad (5.14b)$$

To express (5.14) in terms of the more convenient flux $\mathbf{I}$ of Eq. (5.8), we introduce two unit vectors

$$\begin{aligned} \hat{\mathbf{p}} &= \mathbf{R}_{\alpha\beta}/R_{\alpha\beta}, \\ \hat{\mathbf{p}}' &= \mathbf{R}_{\beta\sigma}/R_{\beta\sigma}. \end{aligned} \quad (5.15)$$

Then, on multiplying both sides of Eq. (5.14b) by $cR_{\alpha\beta}^2\,dR_{\alpha\beta}$ and integrating along the line parallel to $(-\hat{\mathbf{p}})$, we obtain (after some simplification)

$$\begin{aligned} \mathbf{I}(\mathbf{z}_\alpha,\hat{\mathbf{p}}) = {}& \mathbf{I}_c\delta_{\hat{\mathbf{k}},\hat{\mathbf{p}}} + \int_{(-\hat{\mathbf{p}})} R_{\alpha\beta}^2\,dR_{\alpha\beta}\,|G_{\alpha\beta}|^2 \\ &\times \int d\Omega_{\hat{\mathbf{p}}'}\,\mathbf{M}(\hat{\mathbf{p}},\hat{\mathbf{p}}')\mathbf{I}(\mathbf{z},\hat{\mathbf{p}}'). \end{aligned} \quad (5.16)$$

Here we have set

$$\mathbf{M}(\alpha\beta,\beta\sigma) \equiv \mathbf{M}(\hat{\mathbf{p}},\hat{\mathbf{p}}'), \quad (5.17)$$

an obvious change in notation in view of Eqs. (5.15).

At this point it is convenient to drop the α and β labels in Eq. (5.16), replacing (say) $\mathbf{z}_\alpha$ by $\mathbf{x}$ and $\mathbf{z}_\beta$ by $\mathbf{z}$. We may also use Eqs. (3.29) and (4.7) to set

$$|G_{\alpha\beta}|^2 = \frac{1}{R^2}\exp\left[-\int_{\mathbf{z}}^{\mathbf{x}}\frac{ds'}{l(\mathbf{x}')}\right], \quad (5.18)$$

where the path integral is taken along the straight line from $\mathbf{z}$ to $\mathbf{x}$. The final form of our integral equation is then seen to be

$$\begin{aligned} \mathbf{I}(\mathbf{x},\hat{\mathbf{p}}) = {}& \mathbf{I}_c(\mathbf{x})\delta_{\hat{\mathbf{k}},\hat{\mathbf{p}}} + \int_{(-\hat{\mathbf{p}})} dR\,\exp\left[-\int_{\mathbf{z}}^{\mathbf{x}}\frac{ds'}{l(\mathbf{x}')}\right] \\ &\times \int d\Omega_{\hat{\mathbf{p}}'}\,\mathbf{M}(\hat{\mathbf{p}},\hat{\mathbf{p}}')\mathbf{I}(\mathbf{z},\hat{\mathbf{p}}'), \end{aligned} \quad (5.19)$$

where $\hat{\mathbf{p}}$ is parallel to $(\mathbf{x} - \mathbf{z})$ and the R integral is taken along the line through $\mathbf{x}$ extending in the direction of $(-\hat{\mathbf{p}})$.

6. DISCUSSION AND EXTENSION TO THE CASE OF ELLIPTICAL POLARIZATION

The derivation of the transport equation (5.19) has been one of the principal objectives of this paper. We shall not discuss here the most obvious generalizations of this equation, which include the use of curved ray paths and a series of higher-order correlations in $n(\mathbf{x})$ and $\mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{p}}')$.

Differentiation of Eq. (5.19) along a path element δs parallel to $\hat{\mathbf{p}}$ leads to the integro-differential equation

$$\frac{d\mathbf{I}(\mathbf{x}, \hat{\mathbf{p}})}{ds} = -\frac{1}{l(\mathbf{x})}\mathbf{I} + \int d\Omega_{p'}\, \mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{p}}')\mathbf{I}(\mathbf{x}, \hat{\mathbf{p}}'). \quad (6.1)$$

This has the form of the familiar radiation transport equation,[30] except for the specific quantities $l(\mathbf{x})$ and $\mathbf{M}(\hat{\mathbf{p}}, \hat{\mathbf{p}}')$. The advantage of our derivation from the wave equation lies in the inclusion of some wave interference effects through the structure of $\mathbf{M}$ and l and in the ease of generalizing Eq. (6.1).

Approximations (C1) and (C2) permit us to simplify Eq. (5.13) for $\mathbf{M}$:

$$\begin{aligned}(ij|\, M(\hat{\mathbf{p}}, \hat{\mathbf{p}}')\, |st) &= \frac{3}{8\pi}\sigma_T[\rho(\mathbf{z})]^2[\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(i)\cdot\hat{\mathbf{e}}_{\hat{\mathbf{p}}'}(s)\hat{\mathbf{e}}_{\hat{\mathbf{p}}}(j)\cdot\hat{\mathbf{e}}_{\hat{\mathbf{p}}'}(t)] \\ &\quad\times\int d^3R\, g(\mathbf{z}; R)\exp[in_1(\mathbf{z})k(\hat{\mathbf{p}}' - \hat{\mathbf{p}})\cdot\mathbf{R}], \quad (6.2)\end{aligned}$$

where we have returned to the notation of Eqs. (2.2) and (2.3) for the $\hat{\mathbf{e}}$'s and

$$\sigma_T = \frac{8\pi}{3}r_0^2$$

is the Thomson cross section. We note from Eq. (6.2) that $\mathbf{M}$ is *real* and that

$$(ij|\, M(\hat{\mathbf{p}}, \hat{\mathbf{p}}')\, |st) = (ji|\, M(\hat{\mathbf{p}}, \hat{\mathbf{p}}')\, |ts). \quad (6.3)$$

In terms of $\mathbf{M}$ we can write the "scattering cross section of a correlation cell" as

$$\begin{aligned}\sum(\hat{\mathbf{p}}\cdot\hat{\mathbf{p}}') &= \frac{1}{2\rho}\sum_{i,t}(ii|\, M(\hat{\mathbf{p}}, \hat{\mathbf{p}}')\, |tt) \\ &= \rho(\mathbf{z})\sigma_T(\hat{\mathbf{p}}\cdot\hat{\mathbf{p}}')\int d^3R\, g(\mathbf{z}; R) \\ &\quad\times\exp[in_1(\mathbf{z})k(\hat{\mathbf{p}}' - \hat{\mathbf{p}})\cdot\mathbf{R}], \quad (6.4)\end{aligned}$$

where $\sigma_T(\hat{\mathbf{p}}\cdot\hat{\mathbf{p}}')$ is defined by Eq. (2.21). Comparison with Eq. (4.7) shows that

$$\rho(\mathbf{z})\int d\Omega_{\hat{\mathbf{p}}'}\sum(\hat{\mathbf{p}}\cdot\hat{\mathbf{p}}') = \frac{1}{l(z)}, \quad (6.5)$$

as is to be expected. Indeed, this result shows us that Eq. (6.1) is satisfied by a uniform distribution of randomly polarized light. That is,

$$I_{ij}(\mathbf{x}, \hat{\mathbf{p}}) = \tfrac{1}{2}\delta_{ij}I_0,$$

where I_0 is a constant, provides a solution to the transport equation in the absence of a source. (Had we allowed for energy dissipation in the plasma, this conclusion would not, of course, be valid.)

Equation (5.19) is easily generalized to describe the incident radiation that has circular or elliptical polarization. To see this, let us introduce the unit vector $\hat{\mathbf{e}}(2)$, which is perpendicular to both $\hat{\mathbf{k}}$ and $\hat{\mathbf{e}}(1)$. Then, instead of Eq. (2.1), we write

$$\hat{\mathbf{e}}(j)\cdot\mathbf{E}_I(\mathbf{r}) = E_0(j)e^{i\mathbf{k}\cdot\mathbf{r}}, \quad j = 1, 2. \quad (6.6)$$

Instead of Eq. (3.33), we have now

$$\hat{\mathbf{e}}(j)\cdot\mathbf{E}_c(\mathbf{x}) = E_0(j)e^{ikS(\mathbf{x})}. \quad (6.7)$$

The coherent intensity tensor is thus

$$\begin{aligned}I_{cij}(\mathbf{x}) &= e^{-2k\,\mathrm{Im}\,S(\mathbf{x})}\frac{E_0^*(i)E_0(j)}{8\pi} \\ &\equiv e^{-2k\,\mathrm{Im}\,S(\mathbf{x})}I_{ij}^{(0)}, \quad (6.8)\end{aligned}$$

where before we had $I_{ij}^{(0)} = \delta_{i1}\delta_{j1}\mathcal{I}_0$. Evidently,

$$[I_{ij}^{(0)}]^* = I_{ji}^{(0)}. \quad (6.9)$$

For the special case that $\hat{\mathbf{e}}(1)$ and $\hat{\mathbf{e}}(2)$ are parallel to the principal axes of the polarization ellipse, the phases of $E_0(1)$ and $E_0(2)$ differ by $\pi/2$. In this case $I_{12}^{(0)}$ is pure imaginary.

Because $\mathbf{M}$ is real, Eq. (5.19) consists in general of four coupled *real* equations, even though our I_{ij}'s are complex. It follows from Eqs. (6.9) and (6.3) that

$$[I_{ij}(\mathbf{x}, \hat{\mathbf{p}})]^* = I_{ji}(\mathbf{x}, \hat{\mathbf{p}}). \quad (6.10)$$

Thus, four real quantities are required fully to specify the radiation. A common choice are the *Stokes parameters*,[31] defined as

$$\begin{aligned}I &\equiv I_{11} + I_{22}, \\ Q &\equiv I_{11} - I_{22}, \quad (6.11) \\ \tfrac{1}{2}(U - iV) &\equiv I_{12}.\end{aligned}$$

(For the case discussed originally of plane polarized

[30] See, for example, S. Chandrasekhar, *Radiative Transfer* (Dover Publications, Inc., New York, 1960).

[31] See, for example, Chap. I of Ref. 30 for a description of these four parameters.

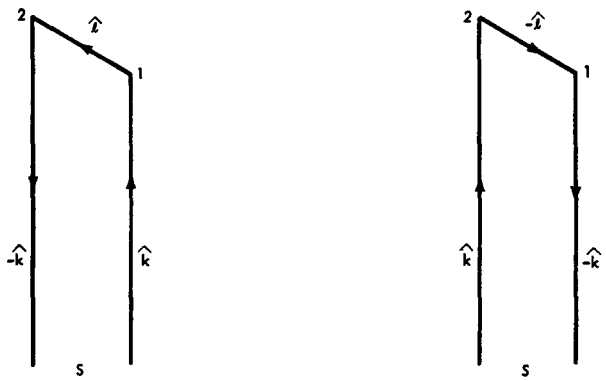


FIG. 4. Example of coherently interfering terms for backscatter.

light, I_{12} is real and $V = 0$.) Evidently, Eq. (5.19) can be written as four coupled, real equations for the four Stokes parameters.

Since Eq. (5.19) is a linear integral equation with real kernel, we can express its solutions in terms of a real *transfer matrix* [or resolvent kernel] $(ij|\, T\,|st)$:

$$I_{ij}(\mathbf{x}, \hat{\mathbf{p}}) = \sum_{s,t=1}^{2} (ij|\, T\,|st) I_{st}^{(0)}. \tag{6.12}$$

It follows from Eqs. (6.9) and (6.10) that

$$(ij|\, T\,|st) = (ji|\, T\,|ts). \tag{6.13}$$

It is evident that in the single-scattering approximation to Eq. (5.19), in which $\mathbf{I}(\mathbf{z}, \hat{\mathbf{p}}')$ is replaced by $\mathbf{I}_c \delta_{\hat{\mathbf{k}},\hat{\mathbf{p}}'}$ within the integrand,the scattering is given in the DWBA, and the expression (4.8) is again obtained for the scattering cross section.

7. RADAR BACKSCATTER

For the calculation of radar backscatter using the transport equation (5.19), a difficulty is encountered.[32] To see how this arises and how it is to be resolved we must return to the original multiple scattering, Eqs. (3.32).

We first note that, for our definition of the polarization vectors,

$$\hat{\mathbf{e}}_{-\hat{\mathbf{q}}}(j) = (-1)^{j-1} \hat{\mathbf{e}}_{\hat{\mathbf{q}}}(j), \quad j = 1, 2. \tag{7.1}$$

From this it follows that [see Eq. (3.7)]

$$f_{ji}(-\hat{\ell}, -\hat{\mathbf{q}}) = (-1)^{i+j} f_{ij}(\hat{\mathbf{q}}, \hat{\ell}). \tag{7.2}$$

To illustrate our problem, we consider the double-scattering approximation to the radar backscatter, using Eqs. (3.31) and (3.32). The source and receiver are supposed to be colocated at a large distance r from the plasma, and scattering takes place from electrons $\mathbf{z}_1$ and $\mathbf{z}_2$. Figure 4 illustrates two sequences of scatterings that contribute. For initial polarization component s and final polarization component t, these are

$$Q_2(t, s) = \sum_j \frac{e^{ikS_{r2}}}{r} f_{tj}(-\hat{\mathbf{k}}, \hat{\ell}) \frac{e^{ikS_{21}}}{R_{21}} f_{js}(\hat{\ell}, \hat{\mathbf{k}}) e^{ikS_{1r}} \tag{7.3a}$$

and

$$\tilde{Q}_2(t, s) = \sum_j \frac{e^{ikS_{r1}}}{r} f_{tj}(-\hat{\mathbf{k}}, -\hat{\ell}) \frac{e^{ikS_{12}}}{R_{12}} f_{js}(-\hat{\ell}, \hat{\mathbf{k}}) e^{ikS_{2r}}. \tag{7.3b}$$

Here, for convenience, we have chosen the eikonal for the incident wave to vanish at the source [absorbing a phase factor into $E_0(s)$] and have taken the incident amplitude equal to unity. Evidently Q_2 and $\tilde{Q}_2$ differ only by a reversal of propagation vectors.

On making use of Eq. (7.2), we see from the above that

$$\tilde{Q}_2(t, s) = (-1)^{t+s} Q_2(s, t). \tag{7.4}$$

Now, on referring to Fig. 5, we see that, for a sequence of n scatterings, there again exists a second set obtained by reversing all propagation vectors (only for backscatter can this be done). The corresponding contributions, corresponding to Eqs. (7.3), are

$$Q_n(t, s) = \sum_{j_1, \cdots j_{n-1}} \frac{e^{ikS_{rn}}}{r} f_{tj_{n-1}}(-\hat{\mathbf{k}}, \hat{\ell}_{n-1}) \cdots f_{j_1 s}(\hat{\ell}_1, \hat{\mathbf{k}}) e^{ikS_{1r}}; \tag{7.5a}$$

$$\tilde{Q}_n(t, s) = \sum_{j_1, \cdots j_{n-1}} \frac{e^{ikS_{r1}}}{r} f_{tj_1}(-\hat{\mathbf{k}}, -\hat{\ell}_1) \cdots \times f_{j_{n-1} s}(-\hat{\ell}_{n-1}, \hat{\mathbf{k}}) e^{ikS_{nr}}. \tag{7.5b}$$

Again, using (7.2), we see that

$$\tilde{Q}_n(t, s) = (-1)^{t+s} Q_n(s, t). \tag{7.6}$$

Evidently for the single scattering approximation, the two sequences of paths giving $\tilde{Q}_1$ and Q_1 are *not*

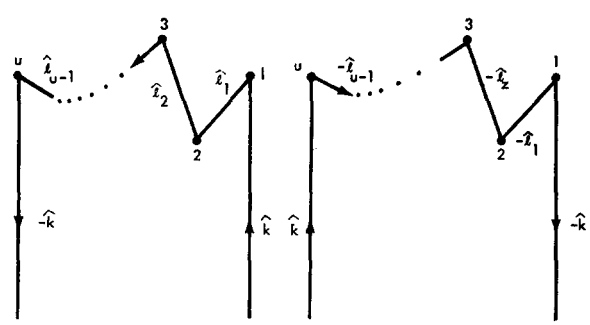


FIG. 5. Illustration of Eqs. (7.5a) and (7.5b).

[32] I am indebted to Dr. R. Ruffine for pointing out this difficulty to me.

distinct. For all higher sequences of scatterings, however, $\tilde{Q}_n$ and Q_n are distinct.

Now, to obtain the backscattered intensity at large distance r, we introduce the backscatter transfer matrix $\mathcal{C}$:

$$I_{ij}(\mathbf{r}, -\hat{\mathbf{k}}) = \sum_{s,t} (ij|\ \mathcal{C}\ |st) I_{st}^{(0)}. \tag{7.7}$$

Reference to Eqs. (3.10) and (3.11) shows that

$$(ij|\ \mathcal{C}\ |st) = \tfrac{1}{2} \sum_{n=2}^{\infty} N^{2n} \langle [Q_n(i, s) + \tilde{Q}_n(i, s)]^* [Q_n(j, t) + \tilde{Q}_n(j, t)] \rangle + N^2 \langle Q_1^*(i, s) Q_1(j, t) \rangle, \tag{7.8}$$

disregarding numerical factors and suppressing reference to electron coordinates. [Note that cross terms of the form $Q_r^* Q_n$ appearing in Eq. (3.11) do not contribute, since the Q's now refer to Eqs. (3.32) rather than (3.4).] The factor of one-half appears in (7.8) because, on integrating over all coordinates and keeping both sets of paired graphs [as in Fig. (5)], we count every scattering twice. Also, in Eq. (7.8) we have explicitly taken account of the observation that, for one scattering only, the duplicate graph does not occur.

The transfer matrix $\mathbf{T}$ of Eq. (6.13) (defined for other than backscatter) is evidently expressed as

$$(ij|\ T\ |st) = \sum_{n=1}^{\infty} N^{2n} \langle Q_n^*(i, s) Q_n(j, t) \rangle. \tag{7.9}$$

This and the relations (7.6) let us write (7.8) in the form

$$\begin{aligned}(ij|\ \mathcal{C}\ |st) = (ij|\ T\ |st) &+ \tfrac{1}{2}[(-1)^{i+s}(sj|\ T\ |it) \\ &+ (-1)^{j+t}(it|\ T\ |sj)] \\ &- \tfrac{1}{2}[(-1)^{i+s}(sj|\ T(\text{s.c.})\ |it) \\ &+ (-1)^{j+t}(it|\ T(\text{s.c.})\ |sj)].\end{aligned} \tag{7.10}$$

Here $\mathbf{T}$(s.c.) is the expression for $\mathbf{T}$ in the approximation that only a single scattering is kept. From Eq. (5.19) we see that this is (for backscatter)

$$\mathbf{T}(\text{s.c.}) = \int_{\hat{\mathbf{k}}} dR \exp\left[-2\int_z^r \frac{ds'}{l(\mathbf{x}')}\right] \times \mathbf{M}(-\hat{\mathbf{k}}, \hat{\mathbf{k}}). \tag{7.11}$$

For calculating radar backscatter we must use the transfer matrix $\mathcal{C}$ of Eq. (7.10). For calculating scattered radiation in all directions, except within the cone with axis parallel to $\hat{\mathbf{k}}$ and cone angle of order $(kR_c)^{-1}$, we use the transfer matrix (6.12). For non-backscatter within this excluded cone, Eq. (7.10) must be modified by certain diffraction form factors.

The use of curved-ray paths and the extension to higher-order correlations will be described in subsequent publications.

Before concluding, we note the work of Keller[33] on wave propagation in media having a fluctuating refractive index. The fluctuations are assumed to have amplitudes characterized by a small parameter, and ensemble averages are formed to calculate both the wave amplitude and two-point correlation function. A transport equation is derived for the latter.

ACKNOWLEDGMENTS

I would like to acknowledge conversations with Dr. R. S. Ruffine which stimulated the present work and, in particular, pointed out the interesting problem arising in connection with radar backscatter. I would also like to thank Dr. Adolph Hochstim for several discussions of this problem.

This work was supported in part by the Jason Division of the Institute for Defense Analyses through a contract from the Advanced Research Projects Agency and in part by National Science Foundation Grant GA–849.

[33] J. B. Keller, *Proceedings of the Seventeenth Symposium in Applied Mathematics* (American Mathematical Society, Providence, R.I., 1964); J. Math. Phys. 5, 537 (1964).

Combinatorial Approach to the N-Representability of P-Density Matrices

MARK L. YOSELOFF AND HAROLD W. KUHN
Department of Mathematics, Princeton University, Princeton, New Jersey

(Received 22 July 1968)

This paper considers the determination of N-representability (for diagonal elements) of p-density matrices restricted to certain finite-dimensional subspaces of l_2 of the configuration space of N identical antisymmetric particles. In particular, an arbitrary set of $N + p$ spin orbitals is selected and one considers the $\binom{N+p}{p}$-dimensional subspace generated by all possible Slater determinants of the spin orbitals being considered. Applying a combinatorial approach to the problem, a necessary and sufficient set of conditions is determined; previous work has dealt only with necessary conditions, except in the 1-matrix case. The paper concludes by presenting a probabilistic interpretation of these conditions which seems of particular interest for the 2-matrix case. The conditions presented here in combination with the Pauli principle give a probabilistic view of the expected occupation of p-tuples of spin orbitals in terms of the expected occupations of lower-order-tuples of spin orbitals.

I. INTRODUCTION

According to well-known physical theory,[1] given a system of N identical antisymmetric particles with coordinates $x_1, x_2, \cdots, x_N$, where each x_i is a combination of a space coordinate r_i from R^3 and a spin coordinate s_i from Z_2, the system's physical situation is described by the wavefunction

$$\Psi(x_1, x_2, x_3, \cdots, x_N).$$

This function is antisymmetric and normalizable.

If Ω_{op} is a Hermitian operator representing a physical quantity associated with the system, it may be expanded as

$$\Omega_{op} = \Omega_{(0)} + \sum_i \Omega_i + \frac{1}{2!}\sum_{ij}{}' \Omega_{ij} + \frac{1}{3!}\sum_{ijk}{}' \Omega_{ijk} + \cdots, \tag{1}$$

where the nth term is an $(n-1)$-particle operator and the prime on the summation indicates that only terms in which all indices are distinct are summed.

In order to determine the average value of this quantity $\langle \Omega_{op} \rangle_{av}$, in the situation given by a normalized wavefunction Ψ, density matrices may be introduced. The p-density matrix or, more simply, the p matrix, is defined by

$$\Gamma^{(p)}(x_1'x_2' \cdots x_p' \mid x_1x_2 \cdots x_p) = \binom{N}{p} \int \Psi^*(x_1', x_2', \cdots, x_p', x_{p+1}, \cdots, x_N) \times \Psi(x_1, \cdots, x_N)\, dx_{p+1} \cdots dx_N. \tag{2}$$

Then,

$$\langle \Omega_{op} \rangle_{av} = \Omega_{(0)} + \int \Omega_1 \Gamma^{(1)}(x_1' \mid x_1)\, dx_1 + \int \Omega_{12}\Gamma^{(2)}(x_1'x_2' \mid x_1x_2)\, dx_1\, dx_2 + \cdots. \tag{3}$$

By expanding the operator in this way, we can greatly simplify a computation of an approximation of its value.

In general, density matrices are bounded linear operators, of trace class, which satisfy the following conditions:

(i) they are Hermitian;
(ii) they are antisymmetric;
(iii) they satisfy the equation

$$\Gamma^{(p-1)}(x_1'x_2' \cdots x_{p-1}' \mid x_1x_2 \cdots x_{p-1}) = \frac{p}{N+1-p} \times \int \Gamma^{(p)}(x_1'x_2' \cdots x_{p-1}'x_p \mid x_1x_2 \cdots x_{p-1}x_p)\, dx_p. \tag{4}$$

However, in order for the expansion given by (3) to have physical significance, the matrices used in the computation must, in addition, satisfy the condition of N-representability.[2,3] The purpose of this paper is to derive complete conditions for the N-representability of the diagonal elements of projections of p matrices into certain finite-dimensional subspaces of l_2 of the configuration space of a fermion system.

II. N-REPRESENTABILITY

A p matrix $\Gamma^{(p)}(x_1'x_2' \cdots x_p' \mid x_1x_2 \cdots x_p)$ is said to be N-representable if there exists some normalized antisymmetric wavefunction Ψ of N particles, such that Γ and Ψ satisfy Eq. (2).

As several authors have described,[1,3,4] in order to facilitate an investigation of density matrices, a countable set of single particle functions $f_i(x)$, $i = 1, 2, 3, \cdots$, may be selected. This set is complete, orthonormal, and is such that any normalizable

[1] For a more detailed discussion of this, see P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955).

[2] C. Garrod and J. Percus, J. Math. Phys. **5**, 1756 (1964).

[3] E. B. Wilson and F. Weinhold, J. Chem. Phys. **47**, 2298 (1967).

[4] E. B. Wilson and F. Weinhold, J. Chem. Phys. **46**, 2752 (1967).

single-particle function $f(x)$ may be expanded as

$$f(x) = \sum_i f_i(x)c_i, \quad c_i = \int f(x_1)f_i^*(x_1)\,dx_1.$$

Using these "spin orbitals," as such functions are called, any normalizable wavefunction for N particles $\Psi(x_1, x_2, \cdots, x_N)$ may be written as

$$\Psi = \sum_K C_K D_K, \tag{5}$$

where K runs over all possible sets of N indices chosen from the natural numbers and, when $K = \{k_1, k_2, \cdots, k_N\}$,

$$C_K = \int \Psi(x_1, x_2, \cdots, x_N) \times f_{k_1}^*(x_1)\cdots f_{k_N}^*(x_N)\,dx_1\,dx_2\cdots dx_N \tag{6}$$

and D_K is the Slater determinant

$$(N!)^{-\frac{1}{2}}\begin{vmatrix} f_{k_1}(x_1) & f_{k_1}(x_2) & \cdots & f_{k_1}(x_N) \\ f_{k_2}(x_1) & f_{k_2}(x_2) & \cdots & f_{k_2}(x_N) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ f_{k_N}(x_1) & f_{k_N}(x_2) & \cdots & f_{k_N}(x_N) \end{vmatrix}. \tag{7}$$

In addition, the normalization condition

$$\sum_K |C_K|^2 = \int |\Psi|^2\,dx_1\,dx_2\cdots dx_N$$

is satisfied.

For purposes of actually carrying out a computation as in (3), it is usually necessary to select a finite set of M spin orbitals. Löwdin [1] discusses the problem of choosing the set of M spin orbitals which make the expression given by (5) best approximate the full expansion of the wavefunction, where the sets K are now restricted to the indices of the functions selected.

If f_i, $i = 1, 2, 3, \cdots, M$, is the set of spin orbitals chosen, then the N-representability problem can be asked for p matrices and wavefunctions expanded only on the subspace of l_2 of the configuration space spanned by all possible Slater determinants of these finitely many functions. From the defining properties of density matrices (4), they may actually be written as matrices on such finite-dimensional subspaces. For a p matrix $\Gamma^{(p)}$, Löwdin[5] considers the diagonal elements

$$L_{i_1 i_2 \cdots i_p} = \int g_{i_1\cdots i_p}^* \Gamma^{(p)} g_{i_1\cdots i_p}\,dx_1\cdots dx_p\,dx_1'\cdots dx_p', \tag{8}$$

where $g_{i_1\cdots i_p}$ is the $p \times p$ Slater determinant of the spin orbitals $f_{i_1}, \cdots, f_{i_p}$ and the variables $x_1, \cdots, x_p$. He shows that, if Ψ is an N-particle wavefunction, expanded as in (5), satisfying (6) and (7), and Γ is the p matrix associated with Ψ by Eq. (2), then

$$L_{i_1\cdots i_p} = \sum_{K(i_1,\cdots,i_p \in K)} |C_K|^2. \tag{9}$$

Conversely, given a p matrix Γ, if one could choose a set of C_K such that for all $i_1, \cdots, i_p$, condition (9) is satisfied and, in addition, the normalization condition

$$\sum_K |C_K|^2 = 1 \tag{10}$$

holds, then the diagonal elements of Γ and the wavefunction Ψ constructed from these C_K using (5) should satisfy (2). Thus, the N-representability problem for diagonal elements is to find a set of C_K satisfying (9) and (10).

III. COMBINATORIAL FORMULATIONS OF N-REPRESENTABILITY

One may now consider the following combinatorial problem:

Let K be an arbitrary subset of N distinct indices $\{i_1, \cdots, i_N\}$ chosen from $\{1, 2, 3, \cdots, M\}$. Let the constants

$$y_{i_1\cdots i_p} \geq 0, \quad \text{for all} \quad 1 \leq i_1 < i_2 < \cdots < i_p \leq M,$$

and

$$\sum_{1\leq i_1 < \cdots < i_p \leq M} y_{i_1\cdots i_n} = 1 \tag{11}$$

be given. Under what conditions do the equations

$$\sum_{K(i_1,\cdots,i_p \in K)} t_K = y_{i_1\cdots i_p}, \quad 1 \leq i_1 < \cdots < i_p \leq M \tag{12}$$

have a nonnegative solution for the t_K?

By setting

$$y_{i_1\cdots i_p} = \binom{N}{p}^{-1} L_{i_1\cdots i_p}$$

and

$$t_K = \binom{N}{p}^{-1} |C_K|^2, \tag{13}$$

it is clear that this problem is equivalent to the problem of the N-representability of diagonal elements as presented at the end of the last section.

There are several other combinatorial problems which are equivalent to the solvability problem presented above.[6] However, this formulation was chosen because of its comparative ease of solution.

[5] In Löwdin's notation, $L_{i_1\ldots i_p}$ is denoted by $\Gamma(i_1, \cdots, i_p)$.

[6] For other equivalent formulations and additional discussion of the 1-matrix case, see H. W. Kuhn, Proc. Sym. Appl. Math. **10**, 141 (1960).

IV. A SOLUTION FOR THE SOLVABILITY PROBLEM

In this section, necessary and sufficient conditions will be given for the solvability problem with $M = N + p$. In order to find these conditions, we use the theorem that a system of linear equations either has a nonnegative solution or there is a vector in the polar cone of the cone generated by the system which makes a positive inner product with the vector of values of the original system of equations.[7]

For the problem at hand, this may be stated as:

Either the system of equations (12) has a nonnegative solution, or there exists a vector $W = (w_{i_1 \cdots i_p})$, $1 \leq i_1 < \cdots < i_p \leq M$, such that

$$\sum_{i_1, \cdots, i_p \in K} w_{i_1 \cdots i_p} \leq 0, \quad \text{for all } K, \tag{14}$$

and

$$\sum_{1 \leq i_1 < \cdots < i_p \leq M} w_{i_1 \cdots i_p} y_{i_1 \cdots i_p} > 0. \tag{15}$$

The system of $\binom{M}{p}$ inequalities given by (14), defined on the $\binom{M}{p}$ variables $w_{i_1 \cdots i_p}$, defines a cone C. If we can find, for each of these inequalities, a vector W^K that satisfies that inequality strictly and the other inequalities as equations, then these are a complete set of extreme rays for the cone.

In the following, $|K|$ denotes the order of the set K.

Lemma: If

$$w_{p-j} = (-1)^{j+1}(p-j)! \prod_{k=0}^{j-1} (M - 2p + k), \quad \text{for } j = 0, 1, \cdots, p, \tag{16}$$

then if $0 \leq h \leq q < p$,

$$A(-1)^{h-1} \prod_{j=0}^{q-h+1} (M - 2p - j) \binom{q-1}{h-2} \prod_{k=1}^{h-2} (M - 2p + k) + \sum_{i=h-1}^{q} \binom{M-p-q}{p-i} \binom{q}{i} w_{p-i}$$
$$= A(-1)^h \prod_{j=0}^{q-h} (M - 2p - j) \binom{q-1}{h-1} \prod_{k=1}^{h-1} (M - 2p + k) + \sum_{i=h}^{q} \binom{M-p-q}{p-i} \binom{q}{i} w_{p-i},$$

where

$$A = \prod_{r=0}^{p-q-1} (M - p - q - r).$$

[7] For one proof of this, see D. Gale, *The Theory of Linear Economic Models* (McGraw-Hill Book Co., New York, 1960), Chap. 2.

Proof: The expression on the left above is equal to

$$A(-1)^{h-1} \prod_{j=0}^{q-h+1} (M - 2p - j) \binom{q-1}{h-2} \prod_{k=1}^{h-2} (M - 2p + k) + \binom{M-p-q}{p-h+1} \binom{q}{h-1} w_{p-(h-1)} + \sum_{i=h}^{q} \binom{M-p-q}{p-i} \binom{q}{i} w_{p-i}$$
$$= A\Bigg[(-1)^{h-1} \prod_{j=0}^{q-h+1} (M - 2p - j) \binom{q-1}{h-2} \times \prod_{k=1}^{h-2} (M - 2p + k) + (-1)^h \prod_{j=0}^{q-h} (M - 2p - j) \binom{q}{h-1} \times \prod_{k=0}^{h-2} (M - 2p + k)\Bigg] + \sum_{i=h}^{q} \binom{M-p-q}{p-i} \binom{q}{i} w_{p-i}$$
$$= A(-1)^h \prod_{j=0}^{q-h} (M - 2p - j) \times \frac{(q-1)(q-2)\cdots(q-h+2)}{(h-1)!} \times \prod_{k=1}^{h-2} (M - 2p + k) \times \{q(M-2p) - (h-1)[M - 2p - (q-h+1)]\} + \sum_{i=h}^{q} \binom{M-p-q}{p-i} \binom{q}{i} w_{p-i},$$

which is equal to the expression on the right in the statement of the lemma. Q.E.D.

Theorem 1: If w_{p-j} is defined as in (16), then the extreme rays of C are given by $W^K = (w^K_{i_1 \cdots i_p})$, $1 \leq i_1 < \cdots < i_p \leq M$, where

$$w^K_{i_1 \cdots i_p} = w_{p-j} \quad \text{if} \quad |\{i_1, \cdots, i_p\} \cap K| = p - j. \tag{17}$$

Proof: Let K' be an arbitrary set of N indices. Then, if $K' = K$,

$$\sum_{i_1, \cdots, i_p \in K'} w^K_{i_1 \cdots i_p} = \binom{M-p}{p} w_p = -\binom{M-p}{p} (p!) < 0.$$

If $K' \neq K$, then, $|K' \cap K| = M - p - q$ where $q > 0$. Taking the sum we have

$$\sum_{i_1, \cdots, i_p \in K'} w^K_{i_1 \cdots i_p} = \sum_{i=0}^{q} \binom{M-p-q}{p-i} \binom{q}{i} w_{p-i}.$$

But, by applying the lemma q times, we have

$$\sum_{i_1,\cdots,i_p\in K'} w^K_{i_1\cdots i_p} = A(-1)^q(M-2p) \times \prod_{k=1}^{q-1}(M-2p+k) + \binom{M-p-q}{p-q} w_{p-q}$$

$$= A\left[(-1)^q(M-2p)\prod_{k=1}^{q-1}(M-2p+k) + (-1)^{q+1}\prod_{k=0}^{q-1}(M-2p+k)\right] = 0.$$

Thus, for each W^K,

$$\sum_{i_1,\cdots,i_p\in K} w^K_{i_1\cdots i_p} < 0$$

and, for all $K' \neq K$,

$$\sum_{i_1,\cdots,i_p\in K'} w^K_{i_1\cdots i_p} = 0.$$

Therefore, the W^K are a complete set of extreme rays. Q.E.D.

Since the W^K are a complete set of extreme rays for C, we have

$$C = \left\{W \mid W = \sum_K a_K W^K, \quad \text{where} \quad a_K \geq 0\right\}.$$

Then, for a vector $Y = (y_{i_1\cdots i_p})$, $1 \leq i_1 < \cdots < i_p \leq M$, and $y_{i_1\cdots i_p} \geq 0$, for all $i_1, \cdots, i_p$, then $WY \leq 0$ for all $W \in C$ if and only if $W^KY \leq 0$ for all K. So we have the following:

Theorem 2: With K as in the statement of the solvability problem in Sec. III and Y satisfying (11), the system of equations (12) has a nonnegative solution if and only if

$$\sum_{1\leq i_1<\cdots<i_p\leq M} w^K_{i_1\cdots i_p} y_{i_1\cdots i_p} \leq 0, \quad \text{for all } K, \quad (18)$$

where $w^K_{i_1\cdots i_p}$ is defined as in (17).

Proof: By the theorem quoted at the beginning of this section, the system of equations given by (12) has a nonnegative solution if and only if for all $W \in C$, $WY \leq 0$. But, this is if and only if $W^KY \leq 0$ for all K, which is the condition given by (18). Q.E.D.

V. CONCLUSIONS

We may now return to the N-representability problem as it is formulated at the end of Sec. II and apply the results obtained in Sec. IV. By applying (13), given a p matrix $\Gamma^{(p)}$, and applying (8), if $M = N + p$, then the diagonal elements of $\Gamma^{(p)}$ are N-representable if and only if the following conditions hold:

$$\sum_{1\leq i_1<\cdots<i_p\leq M} w^K_{i_1\cdots i_p} L_{i_1\cdots i_p} \leq 0, \quad \text{for all } K, \quad (19a)$$

$$\sum_{1\leq i_1<\cdots<i_p\leq M} L_{i_1\cdots i_p} = \binom{N}{p}, \quad (19b)$$

where $w^K_{i_1\cdots i_p}$ is as in (17).

This provides a complete solution for the N-representability of the diagonal elements of a p-density matrix restricted to an $\binom{N+p}{p}$-dimensional subspace of l_2 of the configuration space of the particles. These subspaces are those generated by an arbitrary set of $N + p$ spin orbitals.

One interpretation of conditions (19) can be given in terms of particles filling spin orbitals. From this point of view, for a p matrix, there are N particles and $N + p$ positions, each of which can hold one particle. Then, $L_{i_1\cdots i_p}$ is viewed as the product of the total number of p-tuples of particles and the probability that the spin orbitals $i_1, \cdots, i_p$ are all filled by a particular p-tuple of particles. More simply, $L_{i_1\cdots i_p}$ is just the expected number of p-tuples of particles which lie in the p-tuple of spin orbitals $i_1, \cdots, i_p$. The Pauli principle states that this value lies between zero and one.

As an example of the form that conditions (19) now take, we consider the 2-matrix case (i.e., $p = 2$). Then (19b) becomes

$$\sum_{1\leq i_1<i_2\leq N+2} L_{i_1i_2} = \binom{N}{2},$$

which simply states that the sum, over all pairs of spin orbitals, of the expected number of pairs of particles lying in a pair of spin orbitals is equal to the total number of pairs of particles.

By applying (19b), and (iii) of Sec. I [Eq. (4)], condition (19a) may be written as

$$L_{i_1i_2} \geq L_{i_1} + L_{i_2} - 1,$$

for all i_1, i_2. That is, the expected number of pairs of particles lying in a pair of spin orbitals is greater than or equal to the sum of the expected number of particles in each separate spin orbital minus one. This condition means that if both members of a pair have a large expectation of being filled by individual particles, then the pair of spin orbitals has a large expectation of being filled by a pair of particles. Similar interpretations in terms of expected occupancy can be derived for all p matrices.

If one chooses $N + k$ spin orbitals, where $0 \leq k < p$, and asks the N-representability question for a p matrix, some results have already been obtained. The full solution for these cases seems well within reach using the techniques applied to the case $M = N + p$. We plan to present a detailed solution to all of these cases in the near future.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Level Density of a Fermi System: Nonperiodic Perturbations of the Energy-Level Scheme*

PETER B. KAHN†
State University of New York at Stony Brook, Stony Brook, New York and Weizmann Institute of Science, Rehovoth, Israel

AND

N. ROSENZWEIG
Argonne National Laboratory, Argonne, Illinois

(Received 29 May 1968)

The level density of a degenerate Fermi system is modified by the perturbation in the position of a single-particle level. To study this effect, miscellaneous exact relations between the level densities of the perturbed and unperturbed system are derived. For the special case of the perturbed *uniform* model, these connections become a set of recursion relations which lead to a complete solution of the problem. Results are also obtained in the saddle-point approximation, and these have a simple interpretation in terms of the usual Fermi occupation probabilities. If a single-particle level is deleted from the scheme, the resultant diminution in the level density persists indefinitely with increasing excitation energy. Information about the adequacy of the saddle-point approximation is obtained by comparison with some exact solutions.

1. INTRODUCTION

Out of Bethe's[1] original estimate of the nuclear level density on the basis of the Fermi-gas model arose a now classical problem in mathematical physics which in its simplest form may be characterized as follows: Given a set of single-particle levels ϵ_0, $\epsilon_1, \epsilon_2, \cdots$, which may be called the "structure" of the system, how does the density of levels depend upon the details of this structure?

Van Lier and Uhlenbeck[2] showed that asymptotically the most important parameter pertaining to the structure is the average density of single-particle levels in the vicinity of the Fermi energy. The most comprehensive treatment of the problem is in the work of Bloch,[3] where it is shown that the saddle-point approximation is quite adequate for deriving the dependence of the level density on finer details of the structure, and many interesting results are obtained on the basis of a level scheme of considerable generality.

A useful supplement to Bloch's study has been to consider more specialized energy-level schemes for which the saddle-point approximation yields results in a closed form.[4–6] Occasionally the study of these schematic models has led to at least a partial understanding of the effects arising from various distinctive features of the independent-particle model of the nucleus, such as shell structure, gaps, etc.[7–10]

One class of models for which a particularly neat result could be obtained consists of the "periodic" single-particle level schemes, i.e., schemes in which the pattern of levels repeats over and over again.[6] If a periodic pattern is modified in such a way that the resulting level scheme is again periodic with the density of single-particle levels kept fixed, then the net effect on the level density of the system merely amounts to an additive correction to the excitation energy in the standard formulas. As a result, the dependence of the level density on the details of the periodic pattern, while significant up to considerable energies, eventually disappears asymptotically.

In this paper we illustrate the effects of a *nonperiodic* perturbation of a single-particle level scheme. For this purpose it is sufficient to study in detail the effects produced by the shift in the position of *one* single-particle level, although the methods are applicable also in more general cases. We find that the effects produced by a nonperiodic perturbation usually go beyond producing merely an additive correction to the energy. Furthermore, the resultant modifications in the level may persist indefinitely with increasing excitation energy.

* Research supported by the Weizmann Memorial Foundation, the U.S. Atomic Energy Commission, and the National Science Foundation. A preliminary account of this work was contained in contributed paper 8.125 at the International Conference in Nuclear Structure, Tokyo, 7–13 September 1967.

† Senior Weizmann Fellow 1967–68.

[1] H. Bethe, Phys. Rev. **50**, 332 (1936).

[2] C. van Lier and G. E. Uhlenbeck, Physica **4**, 531 (1937).

[3] C. Bloch, Phys. Rev. **93**, 104 (1954).

[4] N. Rosenzweig, Phys. Rev. **108**, 817 (1957).

[5] N. Rosenzweig, Nuovo Cimento **43**, 227 (1966).

[6] P. B. Kahn and N. Rosenzweig, Phys. Letters **22**, 307 (1966).

[7] N. Rosenzweig, L. M. Bollinger, L. L. Lee, and J. P. Schiffer, *Proceedings of the Second United Nations International Conference on Peaceful Uses of Atomic Energy* (United Nations, Geneva, 1958).

[8] H. W. Newton and J. H. Gibbons, in *Fast Neutron Physics*, J. B. Marion and J. L. Fowler, Eds. (Interscience Publishers, Inc., New York, 1963), Part II.

[9] M. Blann and G. Merkel, Nucl. Phys. **52**, 673 (1964).

[10] A. Chatterjee, Phys. Rev. **134**, B374 (1964); Nucl. Phys. **60**, 273 (1964).

In Sec. 2 we derive some exact relations between the level densities for the unperturbed and perturbed level schemes. In the case of the perturbed *uniform* model, these considerations lead to an exact solution of the problem by means of recursion relations. In Sec. 3 the level density is discussed in the usual saddle-point approximation, the adequacy of which may be judged by comparison with the exact results obtained for the perturbed uniform model. Section 4 deals with a generalization in which the independent-particle levels may be occupied by more than one particle. Concluding remarks are contained in Sec. 5.

2. EXACT RELATIONS

A. Definitions and Notation

We introduce an unperturbed (reference) system consisting of N independent fermions which occupy the single-particle level scheme $\epsilon_0, \epsilon_1, \epsilon_2, \cdots$. The Fermi level is given by ϵ_{N-1}, the zero-point energy is denoted by E_0, an arbitrary energy value of the combined system of particles by E, and the excitation energy $Q \equiv E - E_0$. Let the number of states of energy E be denoted by $C(N, E)$.

Next we define a perturbed system in which the N independent fermions occupy a level scheme that differs from the above reference scheme in that one single-particle level, say ϵ_i, is shifted to an arbitrary position ϵ_s (which may or may not coincide with a level of the reference scheme). It should be noted that the Fermi level and the value of the zero-point energy will only in certain special cases remain the same as in the reference system. This dependence on the details of the perturbation $\epsilon_i \to \epsilon_s$ will play a considerable role in what follows. In the perturbed-level scheme of the N-particle system, let $C(N, E; \epsilon_i, \epsilon_s)$ denote the number of states of energy E.

A particular case of a level shift of the above type is one in which a level of the reference scheme is deleted altogether. Such a deletion creates a gap in the reference scheme. The number of states for this case is denoted by $C(N, E; \epsilon_i, \infty)$. Another situation is one in which a level is added to the reference scheme, and the number of states is denoted by $C(N, E; \infty, \epsilon_s)$.

B. Arbitrary Single-Particle Schemes

We now develop connections between the density of states in the reference system and that in the perturbed system. The results facilitate discussion of a variety of schemes. We use the method of generating functions to enumerate the number of states; explicitly for the reference system, one has[1]

$$\sum_{N',E'} C(N', E')x^{N'}y^{E'} = \prod_j (1 + xy^{\epsilon_j}). \tag{1}$$

Similarly for the perturbed system, the number of states is given by

$$\sum_{N',E'} C(N', E'; \epsilon_i, \epsilon_s)x^{N'}y^{E'} = \prod_j (1 + xy^{\epsilon_j}) \cdot \frac{(1 + xy^{\epsilon_s})}{(1 + xy^{\epsilon_i})} \tag{2}$$

and

$$\sum_{N',E'} C(N', E'; \epsilon_i, \infty)x^{N'}y^{E'} = \prod_j (1 + xy^{\epsilon_j}) \cdot \frac{1}{(1 + xy^{\epsilon_i})}. \tag{3}$$

Then one can establish connections between the various systems by, for example, multiplying Eq. (3) through by $(1 + xy^{\epsilon_i})$, comparing like terms, and observing that

$$C(N, E; \epsilon_i, \infty) + C(N - 1, E - \epsilon_i, \epsilon_i, \infty) = C(N, E). \tag{4}$$

Similarly, multiplying Eq. (3) by $(1 + xy^{\epsilon_s})$ yields

$$C(N, E; \epsilon_i, \infty) + C(N - 1, E - \epsilon_s; \epsilon_i, \infty) = C(N, E; \epsilon_i, \epsilon_s). \tag{5}$$

C. Degenerate Systems

For the remainder of this work we restrict our discussion to degenerate systems, i.e., to excitation energies $Q < |\epsilon_{N-1} - \epsilon_0|$. For a degenerate system, a little reflection shows that the density of states for a given level scheme can depend only on the position of the Fermi level in the scheme and on the excitation energy. We express this property of degenerate systems, for example for the reference system, by introducing the function σ defined by

$$\sigma(\epsilon_{N-1}, Q) = C(N, E). \tag{6}$$

A similar relationship holds, of course, for the perturbed system, and we write

$$\sigma(\tilde{\epsilon}_F, \tilde{Q}; \epsilon_i, \epsilon_s) = C(N, E; \epsilon_i, \epsilon_s). \tag{7}$$

In the above, $\tilde{\epsilon}_F$ and $\tilde{Q}$ denote respectively the Fermi level and excitation energy of the perturbed system which correspond to the values of N and E. Let E_0 denote the zero-point energy of the perturbed system; then $Q \equiv E - E_0$.

We shall now illustrate how Eq. (4) of Sec. 2B may, for degenerate systems, be written in terms of the σ function. Careful attention must be paid to the dependence of ϵ_F and E_0 on ϵ_i and ϵ_s. Consider, for example, the transcription of Eq. (4) for the case $\epsilon_i \leq \epsilon_{N-1}$, $\epsilon_s = \infty$. For a system perturbed by the

deletion of one single-particle level,

$$\begin{aligned}\text{number of states} &= C(N, E; \epsilon_i, \infty),\\ \text{Fermi level} &= \epsilon_N,\\ \text{zero-point energy} &= E_0 + \epsilon_N - \epsilon_i,\\ \text{excitation energy} &= (Q + E_0) - (E_0 + \epsilon_N - \epsilon_i)\\ &= Q - \epsilon_N + \epsilon_i.\end{aligned}$$

For the system of $N - 1$ particles,

$$\begin{aligned}\text{number of states} &= C(N - 1, E - \epsilon_i; \epsilon_i, \infty),\\ \text{Fermi level} &= \epsilon_{N-1},\\ \text{zero-point energy} &= E_0 - \epsilon_i,\\ \text{excitation energy} &= (Q + E_0 - \epsilon_i) - (E_0 - \epsilon_i)\\ &= Q.\end{aligned}$$

For the reference system,

$$\begin{aligned}\text{number of states} &= C(N, E),\\ \text{Fermi level} &= \epsilon_{N-1},\\ \text{zero-point energy} &= E_0,\\ \text{excitation energy} &= Q.\end{aligned}$$

In view of Eqs. (6) and (7) and the above information, Eq. (4) may be written

$$\sigma(\epsilon_N, Q - \epsilon_N + \epsilon_i; \epsilon_i, \infty) + \sigma(\epsilon_{N-1}, Q; \epsilon_i, \infty) = \sigma(\epsilon_{N-1}, Q), \quad \epsilon_i \leq \epsilon_{N-1}. \tag{8}$$

Similarly, one can show that

$$\sigma(\epsilon_{N-1}, Q; \epsilon_i, \infty) + \sigma(\epsilon_{N-2}, Q + \epsilon_{N-1} - \epsilon_i; \epsilon_i, \infty) = \sigma(\epsilon_{N-1}, Q), \quad \epsilon_i > \epsilon_{N-1}. \tag{9}$$

Many other relations of this kind may be obtained in a similar fashion.

D. Perturbed Uniform Model

Let us now consider a further specialization in which the reference system consists of the uniform scheme of single-particle levels $0, 1, 2, \cdots$. For such a degenerate system, the density of states can be shown to be independent of the position of the Fermi level and depends only on the excitation energy. That is, we have the translational invariance

$$\sigma_0(\epsilon_{N-1}, Q) = \sigma_0(\epsilon_{N-1} - D, Q) \tag{10}$$

in which D is an integer and the subscript zero serves as a reminder that one is dealing with the uniform scheme.

Similarly, if we consider a perturbation of the uniform model, the density of states can depend only on the excitation energy and on relative displacements of ϵ_i and ϵ_s from the Fermi level. That is, one has the translational invariance

$$\sigma(\tilde{\epsilon}_F, \tilde{Q}; \epsilon_i, \epsilon_s) = \sigma(\tilde{\epsilon}_F - D, \tilde{Q}; \epsilon_i - D, \epsilon_s - D). \tag{11}$$

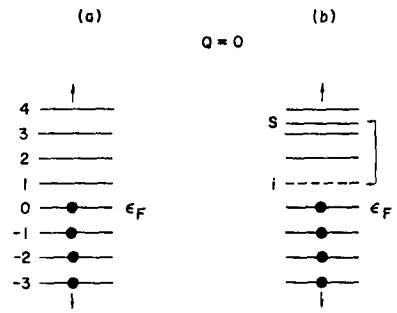


FIG. 1. Illustration of the single-particle level schemes used in the derivation of the density of states. Part (a) illustrates the "uniform model" in which we have a sequence of equally spaced levels, which at zero excitation energy are occupied to the Fermi level ϵ_F. Part (b) illustrates the "perturbed uniform model" in which the uniform model has been altered by shifting a level from position ϵ_i of the uniform scheme to a new position ϵ_s, which may or may not coincide with an existing level. The position of its Fermi level will be altered, for example, if $\epsilon_i < \epsilon_F$ and $\epsilon_s > \epsilon_F$.

In what follows, it is convenient to redefine the single-particle energy-level scheme so that the Fermi level always has the value zero. Thus, as shown in Fig. 1, energy levels below the Fermi level are denoted by the negative integers, whereas the levels above the Fermi level consist of the positive integers. At this point we introduce the symbol ρ to represent the density with this standardization:

$$\begin{aligned}\rho(Q; \infty, \infty) &= \rho_0(Q) \equiv \sigma_0(0, Q),\\ \rho(Q; i, s) &\equiv \sigma(0, Q; \epsilon_i, \epsilon_s),\end{aligned} \tag{12}$$

where $\rho(Q; i, s) \equiv 0$ for $Q \leq 0$.

With the help of the translational invariance expressed by (10) and (11) and the definitions (12), the relations (8) and (9) of Sec. 2C become

$$\rho_0(Q) = \rho(Q + i - 1; i - 1, \infty) + \rho(Q; i, \infty), \quad \text{for } i < 0, \tag{13}$$

$$\rho_0(Q) = \rho(Q - i; i + 1, \infty) + \rho(Q; i, \infty), \quad \text{for } i > 0, \tag{14}$$

$$\rho_0(Q) = \rho(Q - 1; -1, \infty) + \rho(Q; 1, \infty). \tag{15}$$

One must be careful to ensure that the excitation energy Q and the gap at i are measured relative to the same zero value. For example, on the left-hand side of Eq. (8) one term has its Fermi level at ϵ_N and the other at ϵ_{N-1}. After a shift of the position of the Fermi level to zero, the missing level at position $i - 1$ in the N-particle term is to be compared with a missing level at position i in the $(N - 1)$-particle term, and the excitation energy $Q - 1$ is to be compared with an excitation energy Q.

One can now enumerate all the recursion formulas necessary to cover shifts, gaps, and extra levels. One uses as basic quantities $\rho_0(Q)$, the values of which are

tabulated[11] and the values of $\rho(Q; 1, \infty)$ which we have calculated for $1 \leq Q \leq 50$ (Table I). The analysis for all cases proceeds in parallel to the transition from Eqs. (8) and (9) to Eqs. (13) to (15). The level densities for systems with shifted levels are related by means of Eq. (5) to the level density for the uniform model and the level density for a system with a gap. All systems with a gap are related to the uniform model as indicated by Eqs. (13) to (15), then recursively to $\rho_0(Q)$ and $\rho_0(Q; 1, \infty)$. Proceeding in an exactly analogous manner to Eq. (3) the level density for system with an additional level can be related to the uniform model. In summary, therefore, one has explicitly the following relationships:

$$\rho(Q, i, s) = \rho(Q; i, \infty) + \rho(Q - s; i + 1, \infty), \quad \text{for } i, s \geq 1, \quad (16)$$

$$\rho_0(Q) = \rho(Q; \infty, s) - \rho(Q - |s| - \delta_s; \infty, \infty), \quad \delta_s = \begin{cases} 1, & \text{for } s \leq 0, \\ 0, & \text{for } s > 0, \end{cases} \quad (17)$$

$$\rho(Q; -1, s) = \rho(Q - s; 1, \infty) + \rho(Q; -1, \infty), \quad \text{for } s > 0, \quad (18)$$

$$\rho(Q; i, s) = \rho(Q; i, \infty) + \rho(Q - s; i + 1, \infty), \quad \text{for } i < -1, s \geq 1, \quad (19)$$

$$\rho(Q; i, s) = \rho(Q; i, \infty) + \rho(Q + s - 1; i - 1, \infty), \quad \text{for } i, s \leq 0, \quad (20)$$

$$\rho(Q; i, s) = \rho(Q - 1 + s; i - 1, \infty) + \rho(Q; i, \infty), \quad \text{for } i > 1, s < 0, \quad (21)$$

$$\rho(Q; 1, s) = \rho(Q; 1, \infty) + \rho(Q - 2 + s; 2, \infty), \quad \text{for } s < 0, \quad (22)$$

$$\rho(Q; 2, \infty) \equiv \rho(Q; -1, \infty). \quad (23)$$

These relationships can be used to measure the adequacy of the saddle-point approximation, as it is discussed in Sec. 3D.

3. LEVEL DENSITIES IN THE SADDLE-POINT APPROXIMATION

A. Asymptotic Formulas for Degenerate Systems

By means of the Darwin–Fowler integral, we write the density of states for the perturbed uniform model as

$$C(N, E; \epsilon_i, \epsilon_s) = \frac{1}{(2\pi i)^2} \oint\oint \prod_{j=0}^{\infty} \frac{(1 + xy^j)}{x^{N+1}y^{E+1}} \frac{(1 + xy^{\epsilon_s})}{(1 + xy^{\epsilon_i})} dx\, dy, \quad (24)$$

where the product is taken over all integers.

[11] M. Gupta "Partitions," *Royal Society of London—Mathematical Tables* (Cambridge University Press, London, 1958), Vol. 4.

TABLE I. The exact-counting values of $\rho(Q; +1, \infty)$, the density of states when a gap is produced by deleting the state immediately above the Fermi level ϵ_F.

Q	$\rho(Q; +1, \infty)$	Q	$\rho(Q; +1, \infty)$	Q	$\rho(Q; +1, \infty)$
3	2	19	260	35	7808
4	3	20	334	36	9432
5	4	21	420	37	11338
6	6	22	532	38	13631
7	8	23	664	39	16326
8	12	24	835	40	19544
9	16	25	1034	41	23316
10	23	26	1288	42	27806
11	30	27	1588	43	33054
12	42	28	1962	44	39273
13	54	29	2404	45	46534
14	73	30	2953	46	55096
15	94	31	3598	47	65076
16	124	32	4392	48	76808
17	158	33	5328	49	90446
18	206	34	6466	50	106426

The integrand has one and only one saddle point on the positive real axes. If one introduces the exponential transformation

$$x = e^{\alpha}, \quad y = e^{-\beta} \quad (25)$$

and performs the integrations in the usual way, the resulting asymptotic density of states is

$$\rho(Q; i, s) = \frac{e^{f(\alpha,\beta)}}{2\pi[f_{\alpha\alpha}f_{\beta\beta} - (f_{\alpha\beta})^2]^{\frac{1}{2}}}, \quad (26)$$

where

$$f(\alpha, \beta) = E\beta - N\alpha + \sum_{j=0}^{\infty} \ln(1 + e^{\alpha}e^{-\beta j}) - \ln(1 + e^{\alpha}e^{-\beta\epsilon i}) + \ln(1 + e^{\alpha}e^{-\beta\epsilon_s}). \quad (27)$$

All of the quantities in formula (27) must be evaluated at the saddle point (α_0, β_0) which is determined by the solution of the simultaneous equations

$$f_{\alpha} = f_{\beta} = 0.$$

Use of the Euler–MacLaurin expansion[4,12] to evaluate the sums in (27) yields

$$f(\alpha, \beta) = E\beta - N\alpha + \beta^{-1}(\tfrac{1}{6}\pi^2 + \tfrac{1}{2}\alpha^2) + \tfrac{1}{2}\alpha + \tfrac{1}{12}\beta - \ln(1 + e^{\alpha}e^{-\beta\epsilon i}) + \ln(1 + e^{\alpha}e^{-\beta\epsilon_s}), \quad (28)$$

where terms of order $e^{-\alpha}$ and $\beta^2e^{-\alpha}$ have been neglected. Accordingly one can expect accurate results only in the limit $\alpha \to \infty$, β sufficiently small (or equivalently, $N \to \infty$, Q sufficiently large).

To facilitate discussion of the level density as it is obtained in the saddle-point approximation, it is

[12] K. Husimi, Proc. Phys. Soc. Japan **20**, 912 (1938).

convenient to introduce the new quantities

$$\Delta x \equiv \alpha/\beta - (N - \tfrac{1}{2}), \tag{29}$$

$$\Delta\epsilon_i \equiv \epsilon_i - (N-1), \quad \Delta\epsilon_s \equiv \epsilon_s - (N-1), \tag{30}$$

$$\Delta x_i \equiv \{1 + \exp[\beta(\Delta\epsilon_i - \tfrac{1}{2} - \Delta x)]\}^{-1},$$

$$\Delta x_s \equiv \{1 + \exp[\beta(\Delta\epsilon_s - \tfrac{1}{2} - \Delta x)]\}^{-1}. \tag{31}$$

Thus our saddle-point equations become

$$f_\alpha = 0 = \Delta x - \Delta x_i + \Delta x_s,$$

$$f_\beta = 0 = (-\pi^2/6\beta^2) + Q + \Delta(i,s) - \tfrac{1}{24} - \tfrac{1}{2}(\Delta x)^2 - \tfrac{1}{2}\Delta x + \Delta\epsilon_i\Delta x_i - \Delta\epsilon_s\Delta x_s. \tag{32}$$

The resulting level density is then

$$\rho(Q; i, s) = \left(\frac{1-\Delta x_i}{1-\Delta x_s}\right)\frac{\exp\{2\beta[Q + \Delta(i,s) - \tfrac{1}{24} - \tfrac{1}{4}\Delta x + \tfrac{1}{2}(\Delta x_i\Delta\epsilon_i - \Delta x_s\Delta\epsilon_s)]\}}{(48)^{\frac{1}{2}}(\pi^2/6\beta^2)D^{\frac{1}{2}}}, \tag{33}$$

where

$$D = 1 - \beta[\Delta x_i(1-\Delta x_i) - \Delta x_s(1-\Delta x_s)] - (3\beta^3/\pi^2)[\Delta x_i(1-\Delta x_i)(\Delta x - \Delta\epsilon_i + \tfrac{1}{2})^2 - \Delta x_s(1-\Delta x_s)(\Delta x - \Delta\epsilon_s + \tfrac{1}{2})^2] - (3\beta^4/\pi^2) \times [\Delta x_i(1-\Delta x_i)\Delta x_s(1-\Delta x_s)(\Delta\epsilon_i - \Delta\epsilon_s)^2]. \tag{34}$$

The expression (33) for the level density $\rho(Q; i, s)$ is not the result of a solution of the saddle-point equations, but rather a very useful reduction which can serve as a starting point for the consideration of special cases, numerical solutions, etc. That is, for a given Q, ϵ_i, ϵ_s one must still solve Eqs. (31) and (32) for the unknown quantities β, Δx, Δx_i, Δx_s. We shall sometimes denote the level density as $\rho_{\text{trans}}(Q; i, s)$ to bring attention to the fact that the transcendental equation (31) must be solved to obtain a solution. The zero-point correction $\Delta(i, s)$ can be evaluated explicitly as

$$\Delta(i,s) = \delta_i(1-\delta_s) - \delta_i\Delta\epsilon_i + \delta_s\Delta\epsilon_s,$$

where

$$\delta_i = \begin{cases} 1, & \text{for } i \le 0, \\ 0, & \text{for } i > 0, \end{cases} \qquad \delta_s = \begin{cases} 1, & \text{for } s \le 0, \\ 0, & \text{for } s > 0. \end{cases} \tag{35}$$

B. Special Cases and Linear Approximation

In order to analyze the general formula $\rho(Q; i, s)$ as given by Eq. (33), we note the following limiting cases:

Case (i): $\epsilon_s \to \infty$. Then $\Delta x_s = 0$ and $\Delta x_s\Delta\epsilon_s = 0$. [The term $\ln(1 + e^{\alpha}e^{-\beta\epsilon_s})$ is effectively absent from Eq. (28).] One is then considering the asymptotic level density for a system in which the uniform model is perturbed by the absence of a level at position ϵ_i in the single-particle level scheme. In particular, if $\Delta\epsilon_i = +1$, the saddle-point equations can be solved *explicitly*, since one sees that the solution of Eq. (31) is $\Delta x_i = \frac{1}{2}$, and Eq. (32) yields $\beta = (\pi/\sqrt{6})(Q + \frac{1}{12})^{-\frac{1}{2}}$. The level density is then written

$$\rho(Q; 1, \infty) = \frac{\frac{1}{2}\exp\{\pi[\frac{2}{3}(Q + \frac{1}{12})]^{\frac{1}{2}}\}}{(48)^{\frac{1}{2}}(Q + \frac{1}{12})(1 - \frac{1}{4}\beta)^{\frac{1}{2}}}. \tag{36}$$

Case (ii): If $\epsilon_i = \epsilon_s$ or if $\Delta\epsilon_i$, $\Delta\epsilon_s \to \infty$ so that $\Delta x = 0$, one can again solve for the saddle point explicitly and obtain

$$\beta = (\pi/\sqrt{6})(Q - \tfrac{1}{24})^{-\frac{1}{2}}$$

and

$$\rho(Q; a, a) = \rho(Q; \infty, \infty) \equiv \rho_0(Q) = \frac{\exp\{\pi[\frac{2}{3}(Q - \frac{1}{24})]^{\frac{1}{2}}\}}{(48)^{\frac{1}{2}}(Q - \frac{1}{24})}, \tag{37}$$

which is the known result for the uniform model.[4–6]

Case (iii): $\Delta\epsilon_i$, $\Delta\epsilon_s \ll Q$ (linear approximation). One now returns to the expression for $f(\alpha, \beta)$ as given by Eq. (27) and retains terms only through $O(\alpha^2)$ and $O(\beta)$. One calls this approximation the linear approximation and the corresponding level density is denoted by ρ_{lin}. One solves the appropriate saddle-point equations *explicitly* and finds

$$\rho_{\text{lin}}(Q; i, s) = \frac{\exp[\pi(\frac{2}{3}Q^*)^{\frac{1}{2}}]}{(48)^{\frac{1}{2}}Q^*},$$

where

$$Q^* = Q - \tfrac{1}{24} + \Delta(i,s) + \tfrac{1}{2}\Delta\epsilon_i - \tfrac{1}{2}\Delta\epsilon_s. \tag{38}$$

Case (iv): $\Delta\epsilon_i \ll Q$, $\Delta\epsilon_s \to \infty$ (linear approximation). The linear approximation is again valid if one omits the term $\ln(1 + e^{\alpha}e^{-\beta\epsilon_s})$ in Eq. (27) and proceeds as in (iii). Then one finds

$$\rho_{\text{lin}}(Q; i, \infty) = \frac{\exp[\pi(\frac{2}{3}Q^*)^{\frac{1}{2}}]}{2(48)^{\frac{1}{2}}Q^*}, \tag{39}$$

with

$$Q^* = Q - (\tfrac{5}{12}) + \Delta(i, \infty) + \tfrac{1}{2}\Delta\epsilon_i$$

and

$$\Delta(i, \infty) = \delta_i(1 - \Delta\epsilon_i).$$

If $\Delta\epsilon_i = +1$, $\Delta\epsilon_s \to \infty$, one recovers formula (36), except for the absence of the relatively unimportant factor $(1 - \frac{1}{4}\beta)^{\frac{1}{2}}$ in the denominator.

Case (v): $\Delta\epsilon_s \ll Q$, $\Delta\epsilon_i \to \infty$ (linear approximation). Proceeding exactly as above, except that one now has an additional level at ϵ_s, one finds

$$\rho_{\text{lin}}(Q; \infty, s) = 2 \exp [\pi(\tfrac{2}{3}Q^*)^{\frac{1}{2}}]/(48)^{\frac{1}{4}}Q^*, \quad (40)$$

with

$$Q^* = Q + \tfrac{1}{12} + \Delta(\infty, s) - \tfrac{1}{2}\Delta\epsilon_s$$

and

$$\Delta(\infty, s) = \delta_s \Delta\epsilon_s .$$

In the limit of large excitation energy Q, one sees that the level density obtained from Eq. (39) is only *half* that for the uniform model. Similarly, the level density indicated by Eq. (40) is *twice* that for the uniform model. This behavior, characteristic of the nonperiodic nature of the perturbation, is discussed in the following section.

C. Illustration of Results

The novel feature of the nonperiodic perturbation is the presence of the multiplicative factor $(1 - \Delta x_i)/(1 - \Delta x_s)$ in the level-density formula (33). It is evident from Eq. (31) that Δx_i varies with $\Delta\epsilon_i$ so as to become negligible when the latter becomes comparable to Q and becomes $\sim\frac{1}{2}$ when $\Delta\epsilon_i \ll Q$ (or when $\Delta\epsilon_i \to \infty$ and $\Delta\epsilon_s \ll Q$), as shown in Fig. 2. The same statement holds with subscripts i and s interchanged.

To study the origin of this behavior, the density of states for a system with a missing level will be directly compared with that for a system in which a particular level is unoccupied. A little reflection shows that the relation

$$C(N, E; \epsilon, \infty) = [1 - f(\epsilon, Q)]C_0(N, E), \quad \text{for } \epsilon > \epsilon_F, \quad (41)$$

holds *exactly*. Here $f(\epsilon, Q)$ denotes the fraction of all states of the unperturbed system for which the single-particle level ϵ is occupied. If the system is degenerate (i.e., if $Q \lesssim \epsilon_F$), then (41) may also be written

$$\rho(Q; \epsilon, \infty) = [1 - f(\epsilon, Q)]\rho_0(Q). \quad (42)$$

On the other hand, from Eq. (33) we infer that for sufficiently high excitation energy Q, the density of states is

$$\rho(Q; \epsilon, \infty) \approx [(1 - \Delta x_\epsilon)]\rho_0(Q). \quad (43)$$

From (42) and (43) we may conclude that, asymptotically,

$$f(\epsilon, Q) \sim \Delta x_\epsilon \sim \{1 + \exp [\beta(\epsilon - \epsilon_F)]\}^{-1}. \quad (44)$$

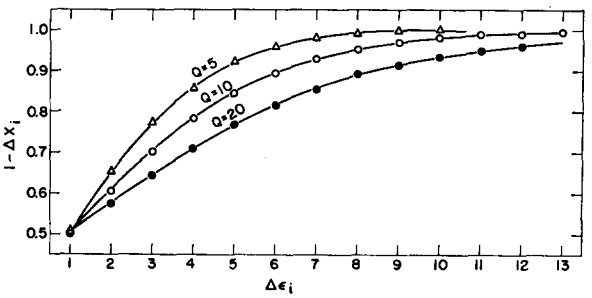


FIG. 2. Illustration of the change in the Fermi factor $(1 - \Delta x_i)$ as a function of $\Delta\epsilon_i$ which defines the position of the deleted level relative to the Fermi level. The values of Δx_i indicated by circles and triangles were obtained by a numerical solution of the saddle-point equations (32), with $\Delta\epsilon_s \to \infty$, $\Delta x_s = 0$.

Thus Δx_ϵ represents (at least approximately) the occupation probability, and the asymptotic expression (44) is precisely the familiar expression from the theory of the ideal Fermi gas, for which case

$$\beta \approx \pi(6Q)^{-\frac{1}{2}}(kT)^{-1}.$$

This result is not restricted to the uniform model and is valid for any single-particle level scheme in which a level has been shifted from position ϵ_i to ϵ_s. If the initial position of the level is below the Fermi level, one must relate the ground state of the perturbed system to that of the unperturbed system in such a way that a proper correspondence can be made between contributing configurations.

It also follows that if n single-particle levels are missing in the neighborhood of the Fermi level, the level density will be diminished by a factor 2^n relative to that of the corresponding unperturbed model. These modifications do not disappear with increasing excitation energy.

D. Comparison Between Exact Counting and Asymptotic Solutions

One can find an explicit value for $\rho(Q; i, s)$ for any given Q, i, s by a numerical solution of Eqs. (31) and (32). This can then be compared with the results of exact counting—thereby ascertaining the general validity of the asymptotic results ($Q \gg 1$). Figures 3 and 4 provide illustrations of typical agreement. We compare the ratios $[\rho(Q; i, s)/\rho(Q, \infty, \infty)]_{\text{trans}}$ and $[\rho(Q; i, s)/\rho_0(Q)]_{\text{exact}}$ rather than $\rho_{\text{trans}}(Q; i, s)$ and $\rho_{\text{exact}}(Q; i, s)$ directly since we want to judge the effectiveness of our treatment of the perturbation caused by the shifts $\epsilon_i \to \epsilon_s$. A comparison of $\rho_{\text{trans}}(Q; \infty, \infty)$ with $\rho_{\text{exact}}(Q; \infty, \infty)$ has already been given in Ref. 4.

All of the recursion relations of Sec. 2D are *approximately* maintained in the saddle-point approximation as given by the numerical solution of Eqs. (31)

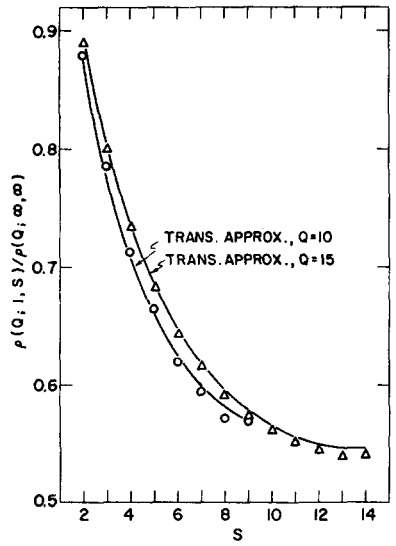


FIG. 3. The ratio of perturbed to unperturbed level density as a function of the level shift illustrated for two values of the excitation energy Q. The single-particle level originally located at one unit above the Fermi level is shifted to a position s units above the Fermi level. The solid lines correspond to the ratio obtained by a numerical solution of the saddle-point equations, whereas the circles and triangles correspond to the results of exact counting.

and (32). In particular, for $Q \geq 5$, a typical discrepancy is $\leq 3\%$. However, the particular identity given by Eq. (23), namely,

$$\rho(Q; 2, \infty) = \rho(Q; -1, \infty)$$

is maintained *exactly* by the saddle-point treatment. [That is, one observes that

$$\Delta x_{-1} = 1 - \Delta x_{+2}, \quad \beta_{-1} = \beta_{+2},$$

and from these identical expressions for the level density $\rho(Q, i, s)$ follow.]

This same identity is found to hold in the linear approximation, as one sees immediately from Eq. (39). This gives one additional confidence in the consistency of the approximation procedures.

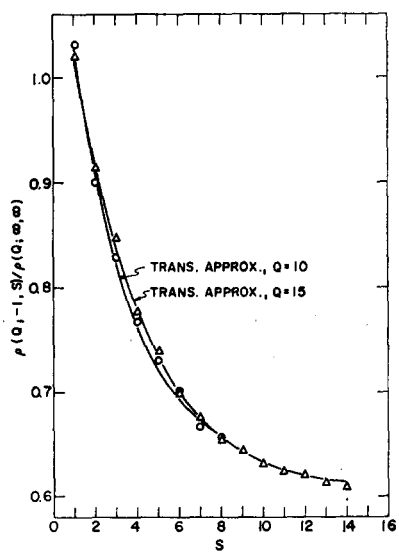


FIG. 4. Same as Fig. 3 except that the single-particle level shift creates a gap one unit *below* an additional level s units *above* the Fermi level of the uniform level scheme.

4. PERTURBED UNIFORM MODEL OF DEGENERACY g IN THE SADDLE-POINT APPROXIMATION

A. Reduction for Degenerate Systems

The results of Sec. 3 can be extended to the study of a perturbed uniform model of degeneracy g, shown schematically in Fig. 5. One proceeds by a development parallel to that of Sec. 3A, so that, corresponding to Eq. (24), one has

$$C(N, E; \epsilon_i, g_i; \epsilon_s, g_s) = \frac{1}{(2\pi i)^2} \oint\oint \prod_{m=0}^{\infty} \frac{(1 + xy^m)^g}{x^{N+1} y^{E+1}} \cdot \frac{(1 + xy^{\epsilon_s})^{g_s}}{(1 + xy^{\epsilon_i})^{g_i}} \, dx \, dy, \tag{45}$$

where the level at ϵ_i has its degeneracy reduced by g_i and the level at ϵ_s has its degeneracy enhanced by g_s. In the general case in which $g_i \neq g_s$, one is effectively considering situations in which one has either shifted some levels to infinity or brought some additional levels from infinity. Corresponding to Eq. (28), we now have

$$f(\alpha, \beta) = E\beta - N\alpha + (g/\beta)(\tfrac{1}{2}\alpha^2 + \tfrac{1}{6}\pi^2) + g(\tfrac{1}{2}\alpha + \tfrac{1}{12}\beta) - g_i \ln(1 + e^{\alpha} e^{-\beta\epsilon_i}) + g_s \ln(1 + e^{\alpha} e^{-\beta\epsilon_s}). \tag{46}$$

To facilitate the discussion of the saddle-point equations, one now introduces the new quantities

$$\Delta x = g(\alpha/\beta) - (N - \tfrac{1}{2}g), \tag{47}$$

$$\Delta\epsilon_i \equiv \epsilon_i - g^{-1}(N - n); \quad \Delta\epsilon_s \equiv \epsilon_s - g^{-1}(N - n), \tag{48}$$

$$\Delta x_i \equiv (1 + \exp\{\beta[g\Delta\epsilon_i - (n - \tfrac{1}{2}g) - \Delta x]\})^{-1}, \tag{49a}$$

$$\Delta x_s = (1 + \exp\{\beta[g\Delta\epsilon_s - (n - \tfrac{1}{2}g) - \Delta x]\})^{-1}, \tag{49b}$$

where n particles occupy the Fermi level in the ground

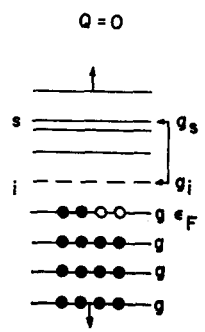


FIG. 5. A perturbed uniform model of degeneracy $g = 4$. One starts with a uniform level sequence in which one has g particles per level and $n = 2$ particles occupying the Fermi level in the ground-state configuration. The level at position ϵ_i has its degeneracy reduced by g_i, where $0 \leq g_i \leq g$; and the level at position ϵ_s has its degeneracy enhanced by g_s. In general, $g_i \neq g_s$.

state. Then our saddle-point equations are

$$f_\alpha = 0 = \Delta x - g_i\Delta x_i - g_s\Delta x_s,$$
$$f_\beta = 0 = -\pi^2 g/6\beta^2 + Q + \Delta(i, g_i; s, g_s) - \tfrac{1}{2}g^{-1}(n - \tfrac{1}{2}g)^2 + \tfrac{1}{12}g - \tfrac{1}{2}g^{-1}(\Delta x)^2 + \Delta x(\tfrac{1}{2} - ng^{-1}) + (g_i\Delta x_i\Delta\epsilon_i - g_s\Delta x_s\Delta\epsilon_s), \quad (50)$$

where

$$\Delta(i; g_i, s; g_s) \equiv -g_i\delta_i\Delta\epsilon_i + g_s\delta_s\Delta\epsilon_s + g^{-1}(n - \tfrac{1}{2}g)(g_i\delta_i - g_s\delta_s) + \tfrac{1}{2}g^{-1}(g_i\delta_i - g_s\delta_s)^2. \quad (51)$$

One can now write an expression for the level density, which like Eq. (33) does not result from a solution of the saddle-point equations but is rather a useful transposition. This expression is

$$\rho(Q; i, g_i; s; g_s) = \frac{(1 - \Delta x_i)^{g_i}}{(1 - \Delta x_s)^{g_s}} \frac{\exp(2\beta Q^*)}{[48D]^{\frac{1}{2}}(g\pi^2/6\beta^2)}, \quad (52)$$

with

$$Q^* = Q + \Delta - \tfrac{1}{2}g^{-1}(n - \tfrac{1}{2}g)^2 + \tfrac{1}{12}g + \tfrac{1}{2}\Delta x(\tfrac{1}{2} - ng^{-1}) + \tfrac{1}{2}(g_i\Delta x_i\Delta\epsilon_i - g_s\Delta x_s\Delta\epsilon_s)$$

and

$$D = 1 - \beta[g_i\Delta x_i(1 - \Delta x_i) - g_s\Delta x_s(1 - \Delta x_s)] - (3\beta^3/\pi^2 g) \times [g_i\Delta x_i(1 - \Delta x_i)(\Delta x - g\Delta\epsilon_i + n - \tfrac{1}{2}g)^2 - g_s\Delta x_s(1 - \Delta x_s)(\Delta x - g\Delta\epsilon_s + n - \tfrac{1}{2}g)^2] - (3\beta^4/\pi^2 g^2)g_i\Delta x_i(1 - \Delta x_i) \times g_s\Delta x_s(1 - \Delta x_s)(\Delta\epsilon_i - \Delta\epsilon_s)^2. \quad (53)$$

B. Special Cases and the Linear Approximation

One can find explicit solutions of the saddle-point equations, for two cases, corresponding to those considered in Cases (i) and (ii) of Sec. 3B.

Case (a): A gap at $\Delta\epsilon_i = 1$, and the number of missing states equal to the degeneracy ($n = g$). Then for $g_i = g$, $g_s = 0$, $\Delta\epsilon_s \to \infty$, the density of states is

$$\rho(Q; i, g_i; 0, \infty) = \frac{1}{2^g} \frac{\exp\{\pi[\tfrac{2}{3}g(Q + g/12)]^{\frac{1}{2}}\}}{(48)^{\frac{1}{2}}(Q + g/12)(1 - \tfrac{1}{4}\beta g)^{\frac{1}{2}}}. \quad (54)$$

Case (b): The uniform model of degeneracy g. Then for $g_i = g_s = 0$, the expression is

$$\rho(Q; 0, \infty; 0, \infty) = \frac{\exp\{\pi[\tfrac{2}{3}g(Q - \tfrac{1}{24})]^{\frac{1}{2}}\}}{(48)^{\frac{1}{2}}(Q - \tfrac{1}{24})}. \quad (55)$$

Case (c): For the remainder of this section we restrict ourselves to the linear approximation. Then corresponding to Eq. (38), one has

$$\rho_{\text{lin}}(Q; i, g_i; s, g_s) = \frac{\exp\{\pi[\tfrac{2}{3}gQ^*]^{\frac{1}{2}}\}}{2^{g_i - g_s}(48)^{\frac{1}{2}}Q^*}$$

with

$$Q^* = Q + \tfrac{1}{12}g - \tfrac{1}{2}g^{-1}(n - \tfrac{1}{2}g)^2 - \tfrac{1}{2}g^{-1}[\tfrac{1}{2}(g_i - g_s)]^2 - g^{-1}(n - \tfrac{1}{2}g)[\tfrac{1}{2}(g_i - g_s)] + \Delta(i, g_i; s, g_s) + \tfrac{1}{2}g_i\Delta\epsilon_i - \tfrac{1}{2}g_s\Delta\epsilon_s \quad (56)$$

and with Δ as given by Eq. (51).

Case (d): The linear approximation can also be used to treat the problem of a uniform model in which the degeneracy of the levels varies in the neighborhood of the Fermi level but has the constant value g everywhere else. If $N_i = \sum n_i g_i$ effective levels are omitted and $N_s = \sum n_s g_s$ effective levels are added, and if p_i of these are removed from below the Fermi level and q_s are added below the Fermi level, then in the linear approximation one finds

$$\rho_{\text{lin}}[Q; N_i, N_s] = \frac{\exp[\pi(\tfrac{2}{3}gQ^*)^{\frac{1}{2}}]}{2^{N_i - N_s}(48)^{\frac{1}{2}}Q^*}, \quad (57)$$

with

$$Q^* = Q + \frac{g}{12} - \frac{1}{2g}\left(n - \frac{g}{2}\right)^2 - \frac{1}{2g}\left(\frac{N_i - N_s}{2}\right)^2 - \frac{1}{g}\left(n - \frac{g}{2}\right)\left(\frac{N_i - N_s}{2}\right) + \Delta + \tfrac{1}{2}\sum g_i\Delta\epsilon_i - \tfrac{1}{2}\sum g_s\Delta\epsilon_s. \quad (58)$$

The factor $2^{N_i - N_s}$ effectively measures the deviation from the uniform model with degeneracy g, since $Q^* \to Q$ in the limit of large excitation energy. One has thereby been able to discuss a perturbed uniform model with degeneracy g even though the pattern in the neighborhood of the Fermi level has been altered in a most general manner.

5. CONCLUDING REMARKS

In this work we have studied how the level density was affected by perturbing the position of a level in an independent-particle energy-level scheme. There are two main conclusions: (1) we find that the calculation in the saddle-point approximation constitutes an adequate treatment of the problem, and (2) we obtain the interesting result that a sizable level shift gives rise to effects that persist to very high excitation energies. An extreme example of this is given by the deletion of a level from the scheme, in which case the asymptotic level density is twice that for the unperturbed energy-level scheme.

Whenever a problem of this kind is solved in the saddle-point approximation, one finds that some

properties of the exact solution are preserved whereas oher properties are not rigorously maintained. Well-known examples of properties that are preserved include the fact that the level density is independent of the number of particles for a degenerate system,[2] the symmetry between particles and holes for uniform schemes of degeneracy g,[4] and the fact that the level density is independent of the definition of the "phase" for a periodic scheme of levels.[6] In this work we also establish certain properties of the exact solution, namely recursion relations (13)–(23). It turns out that these connections are not maintained rigorously by the saddle-point solutions but only to a very good approximation. There is one curious exception to this general statement, namely relation (23) is rigorously maintained in the saddle-point approximation.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the assistance of J. Abdale in connection with these calculations and to thank Dr. F. E. Throw for his valuable comments on the manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Sufficiency Condition for the Validity of the WKB Approximation

F. H. NORTHOVER
Department of Mathematics, Carleton University, Ottawa, Canada

(Received 16 January 1968)

In spite of the many applications—in theoretical physics and elsewhere—of this well-known method of approximating to solutions of the differential equation $W'' + fW = 0$, when f is slowly varying through a region where $f \neq 0$, no simple *sufficient* condition for its validity appears yet to have been given. In the present paper, such a condition is derived. Also, the connection formulas establishing the relation between the constants in the WKB approximation to a given solution in the various regions of the complex plane delineated by the "Stokes lines," which converge on a simple zero of f, are rederived, as existing arguments are open to criticism on at least two major grounds. Finally, a sufficient condition is given for the existence of a common region of validity for the approximation based on the series solutions of the differential equation around a zero of f and the WKB approximation valid sufficiently far from this zero.

1. INTRODUCTION

Jeffrey's (WKB) method is a procedure for finding approximate solutions to a second-order differential equation of the form

$$W'' + f(z)W = 0 \tag{1}$$

through some region R in which $f(z)$ is analytic, which contains no zeros (or limit points of sets of zeros) of $f(z)$, and in which $f(z)$ varies sufficiently slowly. (By this we mean that f' and as many of the subsequent derivatives as may be required have to be sufficiently small through R.)

The method is of importance in many physical applications and the literature is quite extensive; a review of it and a detailed consideration of the problem itself can be found in Heading's book.[1] However, in spite of the numerous investigations which have been made, there still does not seem to be available just what is needed most of all to give precision to the practical applications of the method—namely, a simple *sufficient* condition for its validity with a statement of the order of magnitude of the error.

The WKB approximation fails sufficiently near a zero of f, but, of course, the solutions of (1) can then be expressed as power series in $z - c$, where c is the zero. These power-series solutions converge within the largest circle which can be drawn round c without enclosing a singularity of f. It can sometimes happen that the WKB approximation is still good and useful at points sufficiently near the zero c for the first few terms of these power series, or, less stringently, those containing only $f'(c)$, to provide also a good approximation. In such a case it is obviously of interest to know how the WKB approximate forms connect with the series solutions of (1) and this problem will be investigated at the end of the paper.

The WKB method is often described as an "asymptotic approximation valid in regions where $f(z)$ is sufficiently large." This is misleading; it would not, for example, be true for a function f such that $f'/f^{\frac{3}{2}}$

[1] J. Heading, *Phase Integral Methods* (John Wiley & Sons, Inc., New York, 1961).

properties of the exact solution are preserved whereas oher properties are not rigorously maintained. Well-known examples of properties that are preserved include the fact that the level density is independent of the number of particles for a degenerate system,[2] the symmetry between particles and holes for uniform schemes of degeneracy g,[4] and the fact that the level density is independent of the definition of the "phase" for a periodic scheme of levels.[6] In this work we also establish certain properties of the exact solution, namely recursion relations (13)–(23). It turns out that these connections are not maintained rigorously by the saddle-point solutions but only to a very good approximation. There is one curious exception to this general statement, namely relation (23) is rigorously maintained in the saddle-point approximation.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the assistance of J. Abdale in connection with these calculations and to thank Dr. F. E. Throw for his valuable comments on the manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Sufficiency Condition for the Validity of the WKB Approximation

F. H. NORTHOVER
Department of Mathematics, Carleton University, Ottawa, Canada

(Received 16 January 1968)

In spite of the many applications—in theoretical physics and elsewhere—of this well-known method of approximating to solutions of the differential equation $W'' + fW = 0$, when f is slowly varying through a region where $f \neq 0$, no simple *sufficient* condition for its validity appears yet to have been given. In the present paper, such a condition is derived. Also, the connection formulas establishing the relation between the constants in the WKB approximation to a given solution in the various regions of the complex plane delineated by the "Stokes lines," which converge on a simple zero of f, are rederived, as existing arguments are open to criticism on at least two major grounds. Finally, a sufficient condition is given for the existence of a common region of validity for the approximation based on the series solutions of the differential equation around a zero of f and the WKB approximation valid sufficiently far from this zero.

1. INTRODUCTION

Jeffrey's (WKB) method is a procedure for finding approximate solutions to a second-order differential equation of the form

$$W'' + f(z)W = 0 \tag{1}$$

through some region R in which $f(z)$ is analytic, which contains no zeros (or limit points of sets of zeros) of $f(z)$, and in which $f(z)$ varies sufficiently slowly. (By this we mean that f' and as many of the subsequent derivatives as may be required have to be sufficiently small through R.)

The method is of importance in many physical applications and the literature is quite extensive; a review of it and a detailed consideration of the problem itself can be found in Heading's book.[1] However, in spite of the numerous investigations which have been made, there still does not seem to be available just what is needed most of all to give precision to the practical applications of the method—namely, a simple *sufficient* condition for its validity with a statement of the order of magnitude of the error.

The WKB approximation fails sufficiently near a zero of f, but, of course, the solutions of (1) can then be expressed as power series in $z - c$, where c is the zero. These power-series solutions converge within the largest circle which can be drawn round c without enclosing a singularity of f. It can sometimes happen that the WKB approximation is still good and useful at points sufficiently near the zero c for the first few terms of these power series, or, less stringently, those containing only $f'(c)$, to provide also a good approximation. In such a case it is obviously of interest to know how the WKB approximate forms connect with the series solutions of (1) and this problem will be investigated at the end of the paper.

The WKB method is often described as an "asymptotic approximation valid in regions where $f(z)$ is sufficiently large." This is misleading; it would not, for example, be true for a function f such that $f'/f^{\frac{3}{2}}$

[1] J. Heading, *Phase Integral Methods* (John Wiley & Sons, Inc., New York, 1961).

took large values at some points in the region in question. The essential point is that the first two derivatives of f have to be sufficiently small throughout the region in question, with $f \neq 0$. Also, it is not correct to refer to it as an "asymptotic approximation," for we shall see in the present paper that it is actually the first term of a *convergent* series. The impression seems to have arisen because of the fact that the WKB approximation usually agrees with the dominant term of asymptotic series representations of solutions of (1) when the latter exist.

Although it is true in many practical cases that $|f| \gg 1$, there is nothing to be gained mathematically by writing (as do some authors) $f(z) = k^2 g(z)$, where k is large and $g(z) = 0(1)$ as $k \to \infty$. For $g(z)$, in general, contains k in some unknown fashion and so attempts at approximation in, say, series of inverse powers of k (for k large) are not likely to lead anywhere if a completely rigorous treatment is being insisted upon. In the special case when $g(z)$ is independent of k, it is possible to formally satisfy (1) by the expression

$$\exp\left\{ik\int^z (g)^{\frac{1}{2}}\,d\zeta - \tfrac{1}{4}\ln g + \sum_1 k^{-i}\alpha_i(z)\right\},$$

where the first two terms in the exponent *give the WKB approximation*. The functions $\alpha_i(z)$ in the higher approximation terms are determined in succession by difference–differential equations of increasing complexity. But it does not seem possible to prove even that this series is an asymptotic series in k.

It may prove useful, both to set the background for the more careful investigation to follow and to provide information about the method for readers not already familiar with it, to begin by giving a descriptive nonrigorous derivation of the WKB approximate forms for the solution of (1).

Writing

$$W = e^{iF}, \tag{2}$$

then

$$F'^2 = f + iF''. \tag{3}$$

Suppose it is possible to obtain a solution of (3) which makes $|F''| \ll |f|$ throughout R; then, in that case,

$$F \doteq (f)^{\frac{1}{2}} \tag{4}$$

in R [If f is complex, $f^{\frac{1}{2}}$ is made one-valued by an appropriate system of cuts leaving the zeros (if any); $f^{\frac{1}{2}}$ is then of course analytic in R], i.e., by integration over a finite domain for z in R,

$$F \doteq \int_{z_0}^z f^{\frac{1}{2}}\,d\zeta. \tag{5}$$

If (4) were exact, we could differentiate to obtain F'', but owing to the possibility that the error, even though always small, could vary rapidly at some points, it does not follow that $F'' \doteq f'/2(f)^{\frac{1}{2}}$. However, *assuming* that the error in (4) is a sufficiently slowly varying function for the differentiated form of (4) to be also approximately true, we have $F'' \doteq f'/2(f)^{\frac{1}{2}}$: consistency with the hypothesis $|F''| \ll |f|$ then requires that $|f'/f^{\frac{3}{2}}| \ll 1$. Then (3) gives

$$F' \doteq f^{\frac{1}{2}}\left(1 + \frac{1}{2}\frac{iF''}{f}\right)^{\frac{1}{2}}$$
$$\doteq f^{\frac{1}{2}} + \tfrac{1}{4}if'/f \tag{6}$$

on using the ordinary binomial theorem approximation for index $\frac{1}{2}$.

And so, by (2), we then have

$$W \doteq f^{-\frac{1}{4}}e^{iF}, \tag{7}$$

where

$$F(z) = \int_{z_0}^z [f(\zeta)]^{\frac{1}{2}}\,d\zeta. \tag{7'}$$

This is the Jeffrey's (WKB) approximation.

It will be observed that the approximation (7) becomes exact when $f(z)$ is a *nonzero* constant. It might, therefore, be expected that the approximation would be a valid one in a given region R for which $f(z)$ was nonvanishing and sufficiently slowly varying. The above rough analysis shows already that we have to have $|f'/f^{\frac{3}{2}}| \ll 1$ to apply the method, with a further condition to ensure that the error in (4) varies sufficiently slowly for its differentiated form $F'' \doteq f'/2(f)^{\frac{1}{2}}$ to be also approximately valid. The analysis to follow shows that this further condition is that $|f''/f|$ should be small compared with unity.

We now investigate, along more rigorous lines, the circumstances under which the WKB forms $f^{-\frac{1}{4}}e^{\pm iF}$ can provide a good approximation to a given solution of (1).

Writing

$$W = Pe^{iF}, \tag{8}$$

we find, after some analysis (the details to be given in the author's book[2]), that

$$W(z) = W_1(z) - [f(z)]^{-\frac{1}{4}}\int_{z_0}^z \sin\,[F(z) - F(\zeta)] \times W(\zeta)\{[F(\zeta)]^{-\frac{1}{4}}\}''\,d\zeta, \tag{9}$$

where

$$W_1(z) = \{f(z)\}^{-\frac{1}{4}}\{Ae^{iF(z)} + Be^{-iF(z)}\} \tag{9'}$$

and where $F(z)$ is, of course, given explicitly by (7').

[2] F. H. Northover, *Applied Diffraction Theory* (Academic Press Inc., New York, to be published).

The function $W_1(z)$ is the WKB approximation to the particular solution $W(z)$ of (1) under consideration, which now appears as the solution of a Volterra type of linear integral equation; the error $W - W_1$ incurred by the use of the WKB approximation being expressed by the integral term.

Let $U_{(+)}$ and $U_{(-)}$ be those fundamental[3] solutions of (1) whose WKB approximations are $f^{-\frac{1}{4}}e^{iF}$ and $f^{-\frac{1}{4}}e^{-iF}$, respectively.

Then the pair of integral equations expressing $U_{(\pm)}$ in terms of their respective WKB approximations are obtained as special cases of (8) and may conveniently be summarized as

$$U_{(\pm)}(z) = [f(z)]^{-\frac{1}{4}}\Big\{e^{\pm iF(z)} - \int_{z_0}^{z} \sin [F(z) - F(\zeta)] \cdot U_{(\pm)}(\zeta)\{[f(\zeta)]^{-\frac{1}{4}}\}'' \, d\zeta\Big\}. \quad (10)$$

From (10), the function $AU_{(+)} + BU_{(-)}$ satisfies the same integral equation [i.e., (9)] as does $W(z)$. In reference to (9), $W_1(z)$ is, of course, a *given* function *not zero* and we know that the corresponding solution of the Volterra type linear integral equation, if one exists, will be unique.

Hence,

$$W(z) = AU_{(+)}(z) + BU_{(-)}(z). \quad (11)$$

Accordingly, in estimating the error between $W(z)$ and its WKB approximation $W_1(z)$, it is sufficient to estimate the error between $U_{(\pm)}(z)$ and their respective WKB approximations $f^{-\frac{1}{4}}e^{\pm iF}$.

For by (9′) and (11) we have

$$W - W_1 = A(U_{(+)} - f^{-\frac{1}{4}}e^{iF}) + B(U_{(-)} - f^{-\frac{1}{4}}e^{-iF}). \quad (12)$$

2. ON THE ORDER OF MAGNITUDE OF THE ERROR

A. On an Upper Bound for the Error

We are interested in the circumstances under which the relative error made in using a WKB approximation can be made small; i.e., writing

$$U_{(\pm)} = \Omega_{(\pm)}f^{-\frac{1}{4}}e^{\pm iF}, \quad (13)$$

we are interested in the circumstances under which $\Omega_{(\pm)} \doteq 1$, through some region R wherein $f(z)$ does not vanish.

Substituting in (10), we find that the equations for $\Omega_{(\pm)}$ are summarized as

$$\Omega_{(\pm)}(z) = 1 \pm \frac{i}{2}\int_{z_0}^{z}\{1 - e^{\pm 2i[F(\zeta)-F(z)]}\} \times \Omega_{(\pm)}(\zeta)[f(\zeta)]^{-\frac{1}{4}}\{[f(\zeta)]^{-\frac{1}{4}}\}'' \, d\zeta. \quad (14)$$

[3] If they were not fundamental solutions of (1), then $U_{(+)} = kU_{(-)}$ where k is a constant, which from (9) would imply $ke^{-iF} = e^{iF}$.

Since, by hypothesis, $f(z) \neq 0$ in R and since $F'(z) = f(z) \neq 0$ in R, it follows that the inverse $z(F)$ of the function $F(z)$ exists in R. To proceed with the investigation, it is found to be highly desirable to change the independent variable in (14) from z to F and, correspondingly, in the integral therein from ζ to T, where $T = F(\zeta)$. However, before we can do this, we must first be sure that this inverse function $z(F)$ is *one-valued*. Since this is not necessarily the case [e.g., $z(F)$ will not be one-valued if R is a ring-shaped region surrounding an isolated simple zero of $f(z)$, even if $F(z)$ is made one-valued by a cut] for all allowable values of F [corresponding to the requirement that z satisfy $f(z) \neq 0$], we must, if we wish to make use of this transformation in (14), restrict the variable z to a subregion R_1 of the region R where $f(z) \neq 0$.

In applications of the WKB method we are usually concerned with a ring-shaped region R encircling an isolated simple zero c, say, of $f(z)$. Throughout such a region $F(z)$ and its inverse $z(F)$ are certainly not one-valued, even if we make $F(z)$ one-valued by a suitable cut extending from c. However, it appears (see Appendix A) that we can always find a subregion R_1 of it, wherein $z(F)$ *is* one-valued. Further, it will transpire that this restricted region R_1 is sufficiently extensive to permit, through deductive processes we now explain, a valid investigation by this means of the accuracy of the WKB method, for the complete region R. For, as will appear, the constants (A, B) in the WKB approximation $W_1(z)$ of $W(z)$, take on different values in four wedge-shaped regions into which R is divided by a certain quartet of nonintersecting curves which start out from the zero $z = c$ of $f(z)$. In the discussion we have termed these four subregions of R the regions I, II, III, and IV. (See also Fig. 1.) It is found that the constants (A, B) in all the regions are uniquely determined once their values in any particular one of them are known. It proves

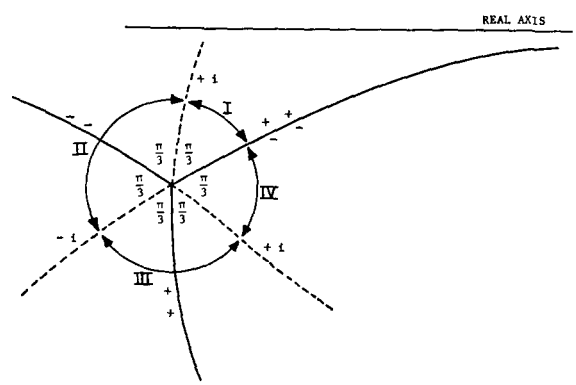


FIG. 1. Stoke's regions near a zero of $f(z)$.

convenient in the theory to take region I as this "basic reference region." It can be shown that the constants (A, B) for I are determined once the particular solution $W(z)$ of (1), to which we wish to approximate, has been chosen, and then, as has been just pointed out, the values of (A, B) for the other regions follow at once. In this way, a knowledge of the WKB approximation function in the region I is sufficient to determine it in all the other regions. Furthermore, since this region turns out to be within the region R_1 for which the inverse function $z(F)$ is one-valued, the error can be properly discussed by the proposed transformation for this particular region and then the error is determined, by the continuation property just explained, for the other regions as well. In this way the error $W - W_1$ can be estimated by the proposed transformation, for the whole region R_1, even though $z(F)$ may not be one-valued therein.

Hence, making the desired transformation of the independent variable in (14) from z to F and understanding that all analysis based thereon shall, in the first place, be restricted to regions where $z(F)$ is one-valued, we have

$$\Omega_{(\pm)}(F) = 1 \pm \frac{i}{2}\int^{F} [1 - e^{\pm 2i(T-F)}] \times \Omega_{(\pm)}(T)[f(\zeta)]^{-\frac{3}{4}}\{[f(\zeta)]^{-\frac{1}{4}}\}'' \, d\zeta, \quad (15)$$

where F is written for $F(z)$, T is written for $F(\zeta)$, but the primes still refer to differentiation with respect to ζ.

Before using this equation to obtain information about the error of the WKB approximation, it will be very useful to recapitulate briefly some relevant theory concerning the standard Volterra linear integral equation. This may be written as

$$\phi(x) = f(x) + \lambda\int_a^x K(x, \xi)\phi(\xi)\, d\xi \quad (16)$$

and we suppose that $K(x, \xi)$ and $f(x)$ are continuous.

The process of successive approximation (iteration) applied to the standard equation (16) gives the series

$$S(x) = f(x) + \lambda\int_a^x K(x, \xi)f(\xi)\, d\xi + \sum_{m=2}^{\infty} \lambda^m K_m(x), \quad (17)$$

where

$$K_m(x) = \int_a^x\int_a^{\xi_1}\int_a^{\xi_2}\cdots\int_a^{\xi_{m-1}} K(x, \xi_1) \times K(\xi_1, \xi_2)\cdots K(\xi_{m-1}, \xi_m), f(\xi_m)\, d\xi_1\, d\xi_2 \cdots d\xi_m. \quad (17')$$

Writing M, M', respectively, for the maximum moduli of $K(x, \xi)$ and $f(x)$ in the range, the modulus of the general term of the series does not exceed $|\lambda|^m M^m M' |x - a|^m$, and so the series converges uniformly when

$$|\lambda| < M^{-1}|x - a|^{-1}. \quad (18)$$

The sum function $S(x)$ defined by it is then a continuous function of x, and it is easy to verify, by actual substitution, that it satisfies the integral equation (16).

For the application to (14), $f(x) \equiv 1$, $\lambda = 1$ and, by (17), the error in approximating the solution of (16) by $S(x) \equiv 1$, i.e., $S(x) - 1$, satisfies the condition

$$|S(x) - 1| < \frac{M|x - a|}{1 - M|x - a|}. \quad (19)$$

So the relative error in using the corresponding WKB forms of the exact solutions $U_{(\pm)}(z)$ is summarized for these two cases by

$$|\Omega_{(\pm)}(z) - 1| < \frac{M_{(\pm)}(z)|F(z)|}{1 - M_{(\pm)}(z)|F(z)|}, \quad (20)$$

where $M(z)$ is now the maximum modulus of

$$\tfrac{1}{2}[1 - e^{\pm 2i(T-F)}]\{f(\zeta)\}^{-\frac{3}{4}}\{[f(\zeta)]^{-\frac{1}{4}}\}'' \quad (20')$$

upon the path of integration joining $T = 0$, $\zeta = z_0$, to $T = F$, $\zeta = z$.

B. The Behavior of the WKB Expression in the Region I[4]

As already mentioned, we are usually concerned in applications with the problem of approximating a solution of (1) by this method in a ring-shaped region R where $f(z) \neq 0$ surrounding an isolated, simple zero of this function. Now, an arbitrary change in the lower limit z_0 of the integral (7′) defining $F(z)$ only alters its value by a constant and so only changes the WKB functions $f^{-\frac{1}{4}} \exp(\pm iF)$ by constant multipliers [which can thereupon be absorbed in the arbitrary constants A, B of the general WKB function $W_1(z)$ of (9′)].

Therefore, we say, without loss of generality, take

$$F(z) = \int_c^z [f(\zeta)]^{\frac{1}{2}}\, d\zeta, \quad (21)$$

where c is an isolated simple zero of $f(z)$, provided, of course, that $f(z)$ is made one-valued by means of a suitable cut extending from the branch point c. The cut we have made (see Fig. 1) separates the region I ($0 \leq \arg F \leq \pi/2$) from the region IV ($5\pi/2 \leq \arg F \leq 3\pi$); on the region I side of it, $\arg F = 0$ and on the region IV side, $\arg F = 3\pi$. Also, $\arg f$ increases by 2π in making a counterclockwise circuit

[4] The case $z \to +\infty$, often important in applications, will be found to be a special case of "$z \to \infty$ in region I."

around $z = c$ from a point on the region I side of this cut to the corresponding point on the region IV side.

The definition (21) for $F(z)$ is the most convenient for analytical purposes and is therefore the one we adopt. Note carefully, however, that we may *not* alter the arbitrary constant z_0 to a *zero* c of $f(z)$ in the integral term of integral equation (9), nor in the corresponding integral terms of the equations leading up to it. This is because of the singularity which occurs in these integrands at such zeros.

The region I can be mapped out by curves starting from the branch point c along which $\arg F = k$, where k is a constant for which $0 \leq k \leq \pi/2$. These curves do not intersect after leaving the branch point $z = c$ and, if s denotes arc length measured from this point along one of them,

$$[f(z)]^{\frac{1}{2}} = F'(z) = e^{(k-\theta)i}(d/ds)\,|F(z)|, \qquad (22)$$

where θ is the angle made with Ox by the tangent to the curve at "z."

Since $f(z) \neq 0$ (except at c), $d\,|F(z)|/ds \neq 0$ (except at c). Since $F(c) = 0$ and, necessarily, $|F(z)| \geq 0$, it follows[5] that

$$d\,|F(z)|/ds > 0 \qquad (23)$$

along a k curve.

Hence, $|F(z)|$ is monotonically increasing along a k curve and so tends to a limit or to $+\infty$. To see which possibility is correct, we remark that in physical applications we are most concerned with solutions of (1) which exhibit oscillatory (wavelike) behavior as $z \to +\infty$. Assuming for the present that we can secure sufficient conditions for the validity of the WKB approximations in region I sufficiently far from $z = c$, we have to have $z \to +\infty$ implying that $\operatorname{Re} F(z) \to +\infty$ with $\operatorname{Im} F(z) > 0$ and *bounded.* Note that this makes the positive real axis asymptotic (from the region I side) to the cut $\arg F = (0, 3\pi)$.

Thus the possibility that $F(z)$ tends to a limit as $s \to \infty$ along a k curve is eliminated. The simplest sufficient condition on $f(z)$ to secure[6] this is that $f(z) \not\to 0$ as $z \to \infty$ in region I; and *we assume that $f(z)$ satisfies this condition.*

To summarize, if $f(z)$ is analytic at, and in a region of sufficient extent around, a simple zero $z = c$ and if $f(z) \not\to 0$ as $z \to \infty$ in the region defined by

$$0 \leq \arg \int_c^z [f(\zeta)]^{\frac{1}{2}}\, d\zeta \leq \pi/2,$$

then

$$|F(z)| \to \infty \quad \text{as} \quad z \to \infty$$

in this region, and (in general) $f^{-\frac{1}{4}}e^{iF}$ becomes exponentially small while $f^{-\frac{1}{4}}e^{-iF}$ becomes exponentially large. The case $z \to +\infty$ is an important subcase of this, and for it, $\operatorname{Im} F$ remains bounded while $\operatorname{Re} F \to +\infty$. The WKB forms are, in that case, oscillatory.

We are now in a position to examine the validity of the WKB approximation in region I.

C. Examination of the WKB Approximation in the Region I

The general WKB form, corresponding to any particular solution $W(z)$ of (1), is $W_1(z)$, where $W_1(z)$ is as given by (9′). We write, therefore,

$$W(z) = \Omega(z)W_1(z), \qquad (24)$$

so that $\Omega(z) - 1$ is the relative error incurred in using $W_1(z)$ instead of $W(z)$. Then we are interested in all the circumstances under which $\Omega(z) - 1$ can be small.

Substituting in (9) to obtain the integral equation for $\Omega(z)$ and then changing the independent variable from z to F [valid where the inverse function to $F(z)$ is one-valued], we have

$$\Omega(F) = 1 - \int_{F_0}^{F} \sin(F - T)\,\frac{Ae^{iT} + Be^{-iT}}{Ae^{iF} + Be^{-iF}} \times \Omega(T)\{f(\zeta)\}^{-\frac{3}{4}}[\{f(\zeta)\}^{-\frac{1}{4}}]''\, dT. \qquad (25)$$

As mentioned already, this equation is valid in region I because the inverse function to $F(z)$ has the necessary one-valuedness there. The path of integration must not, however, include $\zeta = c$ $(T = 0)$ owing to $f(c) = 0$.

Case 1: $z \to \infty$, $Im\ F \to \infty$: Certainly, we may here take $\operatorname{Im} F(z_0) < \operatorname{Im} F(z)$, $|F(z_0)| < |F(z)|$, and it is easy to construct a path of integration for the above integral whereon $\operatorname{Im} T$ is steadily increasing.

For, on a given k curve, $\arg T = k$, $|T|$ continually increases from 0 to $+\infty$ as we move out on it from the branch point c; hence, so does $\operatorname{Im} T$ $(= |T| \sin k)$. Thus $\operatorname{Im} T = \operatorname{Im} F_0$ for exactly one point, $X(k)$ say, of this k curve. As k is varied from $\arg F(z_0)$ to $\arg F(z)$, $X(k)$ goes from z_0 to the corresponding single point $X\{\arg F(z)\}$ upon the k curve $\arg T(\zeta) = \arg F(z)$ through $\zeta = z$.

We now take the map of the path of integration for (25) on the ζ plane as follows: from $\zeta = z_0$ along curve $\operatorname{Im} T = \operatorname{Im} F_0$ to its single intersection with k curve $\arg T = \arg F$; thence along this k curve up to $\zeta = z$ $(T = F)$. The corresponding path in the T plane

[5] Since $d\,|F(z)|/ds$ is a continuous function of s which never vanishes, it must necessarily either be always positive or always negative.

[6] This is true because the limit possibility would imply $d\,|F(z)|/ds \to 0$ as $s \to \infty$ along a k curve; i.e., by (22), that $f(z) \to 0$ as $s \to \infty$ along a k curve, contradicting the hypothesis made on $f(z)$.

is, of course, the two straight line segments $\operatorname{Im} T = \operatorname{Im} F_0$, $\arg T = \arg F$, which join $T = F_0$ to $T = F$. For the present case,

$$|e^{-2iF}| \gg |A/B| \tag{26}$$

and, in particular,

$$|e^{-2iF_0}| \gg |A/B|, \tag{27}$$

so that

$$\sin(F - T)\frac{Ae^{iT} + Be^{-iT}}{Ae^{iF} + Be^{-iF}} \simeq \frac{i}{2}[1 - e^{2i(F-T)}]. \tag{28}$$

On the part of the integration path for which $\operatorname{Im} T = \operatorname{Im} F_0$, $|e^{2i(F-T)}| = e^{-2\operatorname{Im}(F-F_0)}$ and is thus exponentially small. On the final part (where $\arg T = \arg F$), $e^{2i(F-T)}$ is small except near $T = F$, where it approximates to unity. Then, however, the right of (28) is $O(T - F)$,giving a contribution to the integral in question from this last-mentioned part $O(T - F)^2$.

Hence we shall take[7]

$$\left|\sin(F - T)\frac{Ae^{iT} + Be^{-iT}}{Ae^{iF} + Be^{-iF}}\right| \doteqdot \frac{1}{2} \tag{29}$$

in integral (25) for the case at present under study.

Hence, for the application of the theory of the Volterra integral equation—summarized from (16) to (19)—to the integral equation (25):

$$M < \epsilon/8, \tag{30}$$

where

$$\epsilon = \max\left(\frac{f''}{f^2} - \frac{5}{4}\frac{f'^2}{f^3}\right). \tag{31}$$

Application of (20) then shows that the relative error $\Omega(z) - 1$ in using the WKB approximation satisfies the inequality

$$|\Omega(z) - 1| < \frac{\epsilon l}{8 - \epsilon l}, \tag{32}$$

where l is the length of the integration path in (25) and the bound in (30) is taken for that path. But $0 \leq \arg T \leq \pi/2$ for the integration path, since, by hypothesis, this is in region I and $|F_0| < |F|$; hence, $l < |F_0| + |F| < 2|F|$.

Hence, (32) becomes

$$|\Omega(z) - 1| < \epsilon|F|/(4 - \epsilon|F|). \tag{33}$$

For this to be small, say $< N^{-1}$, through a subregion S, say, of region I, we must have, for z in S,

$$\epsilon < \frac{4}{N + 1}\frac{1}{|F|}. \tag{34}$$

[7] The modulus of the left of (28) is actually less than unity for the whole path of integration. This property will be used in the discussion of the next case, namely, $z \to +\infty$, Im F bounded.

Since f' and f'' are independent functions of z, it is desirable to have a sufficiency condition on each of these functions separately. Obviously that condition will be

$$\max\left\{\left|\frac{f''}{f}\right|, \left|\frac{f'}{f^{\frac{3}{2}}}\right|\right\} < \frac{4}{3}\left[\frac{1}{(N + 1)|F|}\right]^{\frac{1}{2}}, \tag{35}$$

for the number on the right is the largest possible *common* upper bound for the two functions on the left to guarantee satisfaction of the condition (34).

The condition (35) certainly fails sufficiently close to the zero c of f, since f is then $O(z - c)$ while F is $O(z - C)^{\frac{3}{2}}$. If we go far enough away from the zero, we have already seen that $|F|$ becomes as large as we please and so there is a possibility of the condition failing again then.

However, if f is sufficiently slowly varying in a neighborhood of c we can guarantee the existence of an annular region around c where (35) will be satisfied.

Let $K(N)$ be the upper bound of values of $z - c$ for which

$$\begin{aligned} f(z) &\doteqdot (z - c)f'(c), \\ f'(z) &\doteqdot f'(c), \\ f''(z) &\doteqdot f''(c) \end{aligned} \tag{36}$$

are good approximations with relative error less than N^{-1}. Take $z - c < K(N)$.

Now, from the Taylor expansion of $f'(z)$ about $z = c$, it is easy to see that, necessarily, $|z - c| < |f'(c)/f''(c)|$. For if not, $|z - c| > K(N)$. Therefore, (35) then becomes

$$\left|\frac{f'}{f^{\frac{3}{2}}}\right| < \frac{4}{3}\left[\frac{1}{(N + 1)|F|}\right]^{\frac{1}{2}}, \tag{37}$$

which by (36) requires that

$$K(N) > |z - c| > [\tfrac{3}{8}(N + 1)]^{\frac{2}{3}}|f'(c)|^{-\frac{1}{3}}, \tag{38}$$

which gives a lower bound on F, so that

$$|F| > \tfrac{1}{4}(N + 1) \tag{39}$$

and then (37) is equivalent to

$$\left|\frac{f'}{f^{\frac{3}{2}}}\right| > \frac{8}{3(N + 1)}. \tag{40}$$

This shows that, when the WKB approximation is a reasonable one ($N > 5$, say) the ratio between the fundamental WKB forms $\exp(-2iF)$ can become quite large, since by (39) $2|F| > \frac{1}{2}(N + 1)$ for all $|z - c| > \{\frac{3}{8}(N + 1)\}^{\frac{2}{3}}|f'(c)|^{-\frac{1}{3}}$. This property will be used in the next section to obtain continuation relations connecting the WKB approximations to a given

solution of (1) in the various "Stokes" regions I, II, III, and IV already noted, which surround the zero $z = c$.

If

$$K(N) < [\tfrac{3}{8}(N+1)]^{\frac{2}{3}} |f'(c)|^{-\frac{1}{3}}, \tag{41}$$

validity of the WKB method can only be guaranteed at points (if any such exist) which satisfy simultaneously the conditions $|z - c| > K(N)$ and (35). Note that a necessary (but not sufficient) condition for the existence of annulus (38) is

$$|f'(c)|^{\frac{4}{3}}/|f''(c)| > [\tfrac{3}{8}(N+1)]^{\frac{2}{3}}. \tag{42}$$

Case 2: $z \to \infty$, Im F Bounded: Here the forms $f^{-\frac{1}{4}} \exp(\pm iF)$ are of the same order of magnitude so we may estimate the relative error involved in using either.

For instance, the ratio Ω of the solution of (1) corresponding to $f^{-\frac{1}{4}} \exp(-iF)$, to this form itself is given as a special case of (24) by taking $A = 0$ and, hence, satisfies the equation

$$\Omega(F) = 1 - \frac{i}{2}\int_{F_0}^{F} [1 - e^{2i(F-T)}] \times \Omega(T)[f(\zeta)]^{-\frac{3}{4}}\{[f(\zeta)]^{-\frac{1}{4}}\}'' \, dT. \tag{43}$$

Using the same path of integration as before, we have thereon

$$|1 - e^{2i(F-T)}| \leq 2, \tag{44}$$

so that now, following through the same steps as before,

$$|\Omega(z) - 1| \leq \epsilon l/(4 - \epsilon l), \tag{45}$$

where ϵ is as defined before, and l is the length of the integration path. But as $\arg F$ and $\arg F_0$ are in this case both small, then $l \doteq |F| - |F_0| < |F|$.

Hence, we obtain exactly the same upper bound for $|\Omega(z) - 1|$ as before [see (33)] and, therefore, *exactly the same sufficiency condition.*

3. THE CONNECTION FORMULAS

If $f(z)$ has zeros, we may, by means of suitable cuts starting from these zeros, make $(f)^{\frac{1}{2}}$ and F one-valued.

In order to see how a cut is to be made, note that, e.g., near the zero c of $f(z)$, we have

$$F(z) \doteq \tfrac{2}{3}\{f'(c)\}^{\frac{1}{2}}(z - c)^{\frac{3}{2}} \tag{46}$$

and so the condition that $\arg(z - c)$ is not to increase by 2π implies that $\arg F$ must not increase by 3π.

We define the cut C from $z = c$ by requiring that, sufficiently near c,

$$-\tfrac{1}{3}\arg f'(c) < \arg z < 2\pi - \tfrac{1}{3}\arg f'(c), \tag{47}$$

in order to have, for convenience, the values of $\arg F$ on each side of the cut, independent of $\arg f'(c)$. We then have $\arg F = 0$ on one side of the cut and $\arg F = 3\pi$ upon the other.

The cut C is shown in Fig. 1 by a thickened curve.

In Fig. 1, the continuous lines indicate where F is purely real and the dotted ones where F is purely imaginary. The arguments of F on these curves are marked; also the symbols "$\pm i$" against the curves mean that, thereon, $\text{Re } F = 0$, $\text{Im } F \gtrless 0$, respectively; while the symbols "$\pm$" mean that, on these curves, $\text{Im } F = 0$, $\text{Re } F \gtrless 0$, respectively.

The dotted lines, upon which F is pure imaginary, divide the z plane round c into regions I, II, III, and IV, defined by

$$\begin{aligned} &\text{region I:} \quad 0 < \arg F < \pi/2, \\ &\text{region II:} \quad \pi/2 < \arg F < 3\pi/2, \\ &\text{region III:} \quad 3\pi/2 < \arg F < 5\pi/2, \\ &\text{region IV:} \quad 5\pi/2 < \arg F < 3\pi. \end{aligned} \tag{48}$$

These dotted lines are examples of what are sometimes called "Stokes lines," and it will be found that the constants A, B in the WKB approximation (9′) to a given solution $W(z)$ of (1) are different for these different regions, and so can be regarded as changing discontinuously as we cross a "Stokes line." The phenomenon is caused by the fact that a function completely analytic through R [i.e., the solution $W(z)$ of (1)] is being represented by a function formula having a line singularity through R, namely, the cut C across which $\arg F$ changes discontinuously. (It is often loosely stated that the WKB formula is many-valued, but this, of course, is not so in view of the cut.)

Let A_1, B_1; A_2, B_2; A_3, B_3; and A_4, B_4 denote the values of the constants in the WKB approximation (9′) to a given solution $W(z)$ of (1), in the regions I, II, III, and IV, respectively. Then we have to express (A_2, B_2), (A_3, B_3), and (A_4, B_4) in terms of (A_1, B_1), and we proceed to derive the formulas which effect this.

Let $u(z)$ be that solution whose WKB approximation in region II is $f^{-\frac{1}{4}}e^{iF}$, and $v(z)$ the solution whose approximation in II is $f^{-\frac{1}{4}}e^{-iF}$.

Then u and v are independent[8] solutions of (1). Therefore, any other solution of this equation is a linear combination of u and v, and so, in particular, are those solutions whose WKB approximations in region I are $f^{-\frac{1}{4}}e^{iF}, f^{-\frac{1}{4}}e^{-iF}$.

Hence, constants $\lambda_1, \hat{\lambda}_1, \mu_1, \hat{\mu}_1$ exist such that

$$\begin{aligned} \lambda_1 u(z) + \hat{\lambda}_1 v(z) &\sim f^{-\frac{1}{4}}e^{iF}, \\ \mu_1 u(z) + \hat{\mu}_1 v(z) &\sim f^{-\frac{1}{4}}e^{-iF}, \end{aligned} \quad \text{in region I.} \tag{49}$$

[8] For, by (9), $u = kv$ implies $e^{iF} = ke^{-iF}$.

Now by hypothesis, the expressions

$$f^{-\frac{1}{4}}(A_2e^{iF} + B_2e^{-iF}), \quad \text{in region II,} \tag{50}$$

and

$$f^{-\frac{1}{4}}(A_1e^{iF} + B_1e^{-iF}), \quad \text{in region I,} \tag{51}$$

are the WKB approximations to the *same* solution of (1) in the regions quoted. But (50) is the approximation, in region II, of the exact solution $A_2u(z) + B_2v(z)$ while, from (49), (51) is the approximation, in region I, of

$$A_1[\lambda_1u(z) + \hat{\lambda}_1v(z)] + B_1[\mu_1u(z) + \hat{\mu}_1v(z)].$$

Hence,

$$A_1[\lambda_1u(z) + \hat{\lambda}_1v(z)] + B_1[\mu_1u(z) + \hat{\mu}_1v(z)] = A_2u(z) + B_2v(z), \tag{52}$$

so that, since u and v are fundamental solutions of (1), we have, equating coefficients

$$\begin{aligned} A_2 &= \lambda_1A_1 + \mu_1B_1, \\ B_2 &= \hat{\lambda}_1A_1 + \hat{\mu}_1B_1. \end{aligned} \tag{53}$$

[If (53) were not true, (52) would imply $u/v = \text{const.}$]

Clearly, the λ's and μ's are independent of the A's and B's because they were defined in the above argument before the latter constants were introduced.

Similarly we can prove that each constant in any given region is linearly dependent on the pair for any other.

The Determination of the Constants in the Connection Relations

Consider the transition across the first Stokes line (the I–II boundary line, $\arg F = \frac{1}{2}\pi$) at distances sufficiently far from the zero $z = c$ of $f(z)$ to be in the region of validity for these WKB approximations, and also sufficiently far from c to make $|F| \gg 1$.

Then, $|e^{-iF}|$ is very large upon this line and $|e^{iF}|$ very small; therefore, from continuity considerations across this boundary between regions I and II,

$$B_2 = B_1 \tag{54}$$

while, by the preceding section,

$$A_2 = \lambda_1A_1 + \mu_1B_1, \tag{55}$$

where λ_1 and μ_1 are independent of A_1 and B_1.

Similarly, for a transition across the second Stokes line (the region II–III boundary, $\arg F = 3\pi/2$), we have

$$\begin{aligned} A_3 &= A_2, \\ B_3 &= \lambda_2A_2 + \mu_2B_2, \end{aligned} \tag{56}$$

where λ_2, μ_2 are independent of A_2 and B_2 and so[9] independent of A_1 and B_1.

[9] For we have $B_3 = \lambda_2\lambda_1A_1 + (\mu_2 + \lambda_2\mu_1)B_1$ and we know, by the preceding section, that $\lambda_2\lambda_1$; $(\mu_2 + \lambda_2\mu_1)$ will be independent of A_1 and B_1. But λ_1 and μ_1 were independent of A_1 and B_1. Hence, so are λ_2 and μ_2.

Similarly, for the transition across the third Stokes line (region III–IV boundary, $\arg F = 5\pi/2$), we have

$$B_4 = B_3 \tag{57}$$

and, by the preceding section,

$$A_4 = \lambda_1A_3 + \mu_1B_3, \tag{58}$$

the constants here being the same as in the corresponding formula (55) above for the region I–II transition because the circumstances of the (III–IV) transition for the small exponential near this boundary are identical with those of the (I–II) transition for the small exponential at this boundary. (The difference between $\arg F$ at the two boundaries is 2π.)

Lastly, by continuity considerations for the exact solution being represented by these WKB forms for *a transition across the cut C* at distances from c sufficient to make $|F| \gg 1$ and to justify their use, and on considering a point $z = \xi$, say, just upon the region I side of the cut, we have

$$[f(\xi)]^{-\frac{1}{4}}[A_1e^{iF(\xi)} + B_1e^{-iF(\xi)}] = -i[f(\xi)]^{-\frac{1}{4}}[A_4e^{-iF(\xi)} + B_4e^{iF(\xi)}], \tag{59}$$

since (as can be seen by considering the form of the functions near $z = c$) $\arg f$ increases by 2π and $\arg F$ by 3π when z encircles the branch point c in the positive sense once, starting from the point ξ and ending up at the contiguous point just the other side of the cut.

Hence, by (59),

$$\begin{aligned} -iA_4 &= B_1, \\ -iB_4 &= A_1. \end{aligned} \tag{60}$$

These equations give

$$\begin{aligned} \lambda_1 &= \omega, \quad \mu_1 = i/\omega, \\ \lambda_2 &= i/\omega, \quad \mu_2 = \omega, \end{aligned} \tag{61}$$

where ω is one of the three cube roots of unity.

The only way the author could see of deciding which value of ω was the correct one for the WKB representation problem was to compare the above theory for the solution of (1) with a differential equation whose solutions are well known. The details appear in the author's forthcoming book,[2] and the result is that we must take $\omega = 1$ in the above.

4. CONNECTION OF THE WKB APPROXIMATION WITH THE SERIES APPROXIMATION VALID SUFFICIENTLY NEAR A ZERO OF $f(z)$

The requirements to be satisfied to ensure that the regions of validity of the two types of approximation should overlap (or at least adjoin) are rather involved.[2]

However, a necessary condition is that $f''(c)$ be small enough to make

$$\frac{|f_0'|^{\frac{4}{3}}}{|f_0''|} > \max\left\{\frac{N}{40}\left[\frac{3}{8}(N+1)\right]^3; \frac{N}{24}\left[\frac{3}{8}(N+1)\right]^{\frac{7}{3}}\right\}. \tag{62}$$

When this situation exists, it appears[2] that one formula can be used to approximate $W(z)$ within the *entire circle* surrounding $z = c$ which just contains the annulus (38) (wherein the WKB type forms are valid).

This single formula is

$$W(z) \doteq \left(\frac{1}{3\pi}\right)^{\frac{1}{2}}[f'(c)]^{\frac{1}{3}}(z-c)^{\frac{1}{2}} \times (A_1 e^{5i\pi/12}H_{\frac{1}{3}}^{(1)}\{\tfrac{2}{3}[f'(c)]^{\frac{1}{2}}(z-c)^{\frac{3}{2}}\} + B_1 e^{-(5i\pi/12)}H_{\frac{1}{3}}^{(2)}\{\tfrac{2}{3}[f'(c)]^{\frac{1}{2}}(z-c)^{\frac{3}{2}}\}). \tag{63}$$

An improved form of (63), which approximates to the given solution $W(z)$ of (1) better when z is too far from c for a linear approximation to $f(z)$ to be of any use (i.e., when z is outside the above-mentioned circle), has been suggested by Langer[10] and has been further refined by the present author.[11] It is

$$(F/F')^{\frac{1}{2}}[A_1 H_{\frac{1}{3}}^{(1)}(F) + B_1 H_{\frac{1}{3}}^{(2)}(F)], \tag{64}$$

where $F(z)$ is as we have defined it.

This approximation will be found to satisfy the equation

$$W'' + \left(Q^2 + \frac{5}{36}\frac{Q^2}{F^2} - \frac{3}{4}\frac{Q'^2}{Q^2} + \frac{1}{2}\frac{Q''}{Q}\right)W = 0, \tag{65}$$

where

$$Q = [f(z)]^{\frac{1}{2}}. \tag{65'}$$

APPENDIX A: ON THE REQUIREMENTS FOR THE ONE-VALUEDNESS OF THE INVERSE FUNCTION TO $F(z)$

Let R be the radius of the largest circle that can be drawn round the zero $z = c$ of $f(z)$ without enclosing any singularities, or other zeros, of f.

Then, from consideration of the Taylor expansion of f, we see that, provided the elementary function $(z-c)^{\frac{1}{2}}$ is made one-valued by a cut drawn from c and extending to ∞, $f^{\frac{1}{2}}$ can be expanded in the following form:

$$f^{\frac{1}{2}} = \{f'(c)\}^{\frac{1}{2}}(z-c)^{\frac{1}{2}}\left\{1 + \sum_{n=1}^{\infty} a_n(z-c)^n\right\}, \tag{A1}$$

and the power series is convergent in $|z-c| < R$. This being a Taylor series (A1) therefore defines $f^{\frac{1}{2}}$ inside the cut circle.

Integrating (A1) term by term, which is valid in view of the uniformity of convergence of the series inside its circle of convergence, we have an expression for $F(z)$ which can be written in the form

$$F(z) = \tfrac{2}{3}[f'(c)]^{\frac{1}{2}}(z-c)^{\frac{3}{2}}\left[1 + \sum_{n=1}^{\infty} b_n(z-c)^n\right], \tag{A2}$$

the power series converging inside the same circle; and (A2) then defines $F(z)$ as a one-valued function inside the same cut circle, analytic therein except on the cut and at the branch point c.

We have, in the paper, chosen this cut so that $\arg F = 0$ on one side of it and $\arg F = 3\pi$ on the other. In this cut plane, $0 < \arg F < 3\pi$.

We show that the subregion R_1 of $|z-c| < R$, wherein the inverse $z(F)$ of $F(z)$ is one-valued, is defined by the condition $0 < \arg F < 2\pi$ and so comprises the regions I, II, and part of III.

This is obvious enough for z sufficiently near c, since there F is also arbitrarily small and

$$z - c \doteq (\tfrac{3}{2})^{\frac{2}{3}}[f'(c)]^{-\frac{1}{3}}F^{\frac{2}{3}}, \tag{A3}$$

and the only way to secure the one-valuedness of $F^{\frac{2}{3}}$ is to prevent $\arg F$ increasing or decreasing by 2π or more. This, of course, is accomplished by making a cut in the F plane from the branch point $F = 0$, and, bearing in mind that the cut plane must include the desired reference region I of the text (for which $0 < \arg F < \pi/2$), we see that the cut must be made (as in the usual way for such a function) along the positive real F axis; we have then, in this cut plane, $0 < \arg F < 2\pi$.

Turning now to the general case, recall that $F(z)$ is given by (A2) with z in the cut interior of the circle $|z-c| < R$, the cut being made by a curve drawn from c. In this cut region $0 < \arg F < 3\pi$; on the cut $\arg F = 0$ on one side and $\arg F = 3\pi$ upon the other.

Since

$$1 + \sum_{n=1}^{\infty} b_n(z-c)^n$$

has, by the hypothesis on $f(z)$, no zeros in $|z-c| < R$ and is analytic in this circle,

$$\left[1 + \sum_{n=1}^{\infty} b_n(z-c)^n\right]^{-\frac{2}{3}}$$

can be developed into the Taylor series

$$1 + \sum_{n=1}^{\infty} B_n(z-c)^n,$$

and the radius of convergence of this last series is also R.

[10] R. E. Langer, Phys. Rev. **51**, 669 (1937).
[11] F. H. Northover, Can. J. Phys. **33**, 241 (1955).

For brevity writing

$$A = (\tfrac{3}{2})^{\frac{2}{3}}[f'(c)]^{-\frac{1}{2}}, \tag{A4}$$

we find that (A2) becomes

$$z - c = A\Lambda\left[1 + \sum_{n=1}^{\infty} B_n(z-c)^n\right], \tag{A5}$$

where

$$\Lambda = F^{\frac{2}{3}}. \tag{A6}$$

In (A5), z is in the interior of the circle $|z - c| = R$, cut as described (so that therein $0 < \arg F < 3\pi$).

Now, throughout the circle $|z - c| < R$, the inverse function to Λ, $z(\Lambda)$ exists. For $\Lambda'(z) = \frac{2}{3}F^{-\frac{2}{3}}f^{\frac{1}{2}} \neq 0$.

But (A5) can be formally satisfied by

$$z - c = A\Lambda\left(1 + \sum_{n=1}^{\infty} C_n\Lambda^n\right), \tag{A7}$$

the coefficients C_n of the Taylor series in (7A) being uniquely determined (in succession) by

$$\begin{aligned} C_1 &= B_1A, \\ C_2 &= B_1AC_1 + B_2A^2, \\ C_3 &= B_1AC_2 + 2B_2A^2C_1 + B_3A^3, \\ &\cdots \end{aligned} \tag{A8}$$

the formal manipulations involved (multiplying series by themselves, equating coefficients, etc.) being justified within the radii of convergence of the power series involved.

As just pointed out, we know that the inverse function $z(\Lambda)$ exists for $|z - c| < R$. Since by (A7) it is representable by a Taylor series, it must, therefore, be a *unique* function of Λ, analytic within the circle of convergence of that Taylor series. The radius of convergence of this series for $z(\Lambda)$ will have to be K, where K is the least upper bound of the set of positive numbers k for which the inequality

$$|A|\,k\left|1 + \sum_{n=1}^{\infty} C_n k^n e^{ni\theta}\right| < R \tag{A9}$$

is true for *all* θ in $0 < \theta < 2\pi$.

Thus the inverse of $\Lambda(z)$, namely $z(\Lambda)$, is a one-valued, analytic function of Λ for $|\Lambda| < k$, $0 \leq \Lambda \leq 2\pi$, and therefore, *in particular*, it has this property in the sector $|\Lambda| < k$, $0 < \arg\Lambda < 4\pi/3$. But, by (A6), the points of this sector are in (1, 1) correspondence with the set of values of F for which $|F| < K^{\frac{3}{2}}$, $0 < \arg F < 2\pi$. Hence, $z - c$ must be a *one-valued* function of F for values of F for which $|F| < K^{\frac{3}{2}}$, $0 < \arg F < 2\pi$, i.e., for values of z for which $0 < \arg F < 2\pi$, $|z - c| < R$ (for $|F| < K^{\frac{3}{2}}$ implies $|\Lambda| < K$ which implies $|z - c| < R$), where R is the radius of the greatest circle round the zero c of $f(z)$ containing no other zeros or singularities of f.

APPENDIX B: CRITICISMS OF PREVIOUS DERIVATIONS OF THE WKB CONNECTION FORMULAS

It appears that previous analysis is open to criticism on two counts:

(i) It is assumed without *proof* that the relation between the constants in the connection formulas for two different "Stokes regions" has necessarily to be linear.

(ii) The fact that there are *three* solutions to the equations giving the coefficients in the linear connection relations appears to have been overlooked—and so also the consequent necessity of giving reasons why two of these solutions should be rejected.

Linear Inequalities for Density Matrices

ERNEST R. DAVIDSON*
Chemistry Department, University of Washington, Seattle, Washington

(Received 20 November 1968)

A set of linear inequalities for the reduced two-body density matrix is presented. The general problem of finding all such inequalities is discussed. It is shown that all of the linear inequalities found so far for ensemble representability may be written as a single family of inequalities.

INTRODUCTION

It is well known[1] that the average energy of any system involving only one- and two-body forces may be expressed in terms of the two-body reduced density matrix Γ defined by

$$\Gamma(1,2;1',2') = N(N-1)\int \Psi(1,2,3,\cdots,N) \times \Psi^*(1',2',3,\cdots,N)\,dX_3\cdots dX_N. \quad (1)$$

Thus, if

$$H = h_0 + \sum_{i=1}^{N} h_1(i) + \sum_{i\neq j}^{N} h_2(i,j), \quad (2)$$

where h_0, h_1, and h_2 are zero-, one-, and two-particle operators, respectively, H may be rewritten as

$$H = \sum_{i\neq j}^{N} g(i,j), \quad (3)$$

where

$$g(i,j) = N^{-1}(N-1)^{-1}h_0 + 2^{-1}(N-1)^{-1}[h_1(i) + h_1(j)] + h_2(i,j). \quad (4)$$

The operator g is then called the reduced two-body Hamiltonian to H. The average energy $\bar{E}$ is then given by

$$\bar{E} = \int [g(1,2)\Gamma(1,2;1',2')]_{\substack{1'=1\\2'=2}}\,dX_1\,dX_2, \quad (5)$$

which is more usually written as

$$\bar{E} = \mathrm{Tr}\,(g\Gamma). \quad (6)$$

This equation for $\bar{E}$ only holds provided Ψ is totally symmetric (bosons) or totally antisymmetric (fermions).

Similarly, if H involves only one-body forces, H may be written as

$$H = \sum_{i=1}^{N} h(i), \quad (7)$$

$$h(i) = h_1(i) + N^{-1}h_0. \quad (8)$$

The average energy may then be expressed in terms of the one-body reduced density γ, defined by

$$\gamma(1;1') = N\int \Psi(1,2,\cdots,N) \times \Psi^*(1',2,\cdots,N)\,dX_2\cdots dX_N, \quad (9)$$

as

$$\bar{E} = \mathrm{Tr}\,(h\gamma) \equiv \int [h(1)\gamma(1;1')]_{1'=1}\,dX_1. \quad (10)$$

In addition to these formulas for a single system, formulas for the average energy (per system of N particles) in an ensemble are easily obtained. If the set $\{\Psi_i\}$ of orthonormal functions are the possible states of the system and if ω_i ($\omega_i \geq 0$, $\sum_i \omega_i = 1$) is the probability in the ensemble of finding the system in state Ψ_i, then

$$\Gamma = \sum \omega_i\Gamma_i, \quad (11)$$

$$\gamma = \sum \omega_i\gamma_i, \quad (12)$$

where Γ_i and γ_i are the reduced densities for Ψ_i, and Eq. (6) or Eq. (10) still holds for the energy. Further, if the ensemble consists of systems with varying numbers of particles so the set $\{\Psi_i\}$ includes states of various N, Eqs. (11) and (12) are still valid, but

$$\bar{E} = h_0 + \mathrm{Tr}\,(h_1\gamma) + \mathrm{Tr}\,(h_2\Gamma) \quad (13)$$

must be used to obtain $\bar{E}$.

It would be very nice if the average energy formulas could be used by making direct guesses to Γ (or γ) without reference to any wavefunction. In order to obtain sensible results in such a procedure, it is necessary that the guessed Γ be derivable, in the manner indicated above, from wavefunctions. This problem of determining the boundary conditions on Γ is called the representability problem. If Γ is to correspond to one Ψ, the boundary conditions are the pure-state N-representability conditions. If Γ corresponds to an ensemble, the conditions should either be the ensemble N-representability condition or just the ensemble representability conditions.

Many such conditions are known. From its definition, Γ is obviously a positive Hermitian kernel. Further

$$\Gamma(1,2;1',2') = \pm\Gamma(2,1;1',2'), \quad (14)$$

* Alfred P. Sloan Research Fellow.
[1] P.-O. Löwdin, Phys. Rev. 97, 1474 (1955).

where + is used for bosons and − for fermions. In addition, an upper bound on the eigenvalues of Γ is known for the pure-state case.[2] There are also the so-called G-matrix conditions,[3] Q-matrix conditions,[3] and the Weinhold inequalities.[4]

For the purpose of doing calculations on the ground state of a molecule, it is only necessary to find the conditions on Γ such that Eq. (6) will give an upper bound to the ground-state fermion energy of the Schrödinger equation. This problem has not been solved, and the conditions presented in this paper are undoubtedly stronger than necessary to guarantee "variational stability." *Variational stability will be used in this context to mean that Eq.* (6) *or* (13) *gives a permitted energy of the system.*

Clearly a necessary set of representability conditions on Γ is that

$$E_{\min} \leq \bar{E} \leq E_{\max}, \tag{15}$$

where $E_{\min}$ and $E_{\max}$ are the minimum and maximum eigenvalues of H. *The whole set of these variational stability inequalities for all H provides a solution to the representability problem.*[3,5] It is hoped, however, that a subset of them would suffice for the Schrödinger Hamiltonian H_s.

I. CONDITIONS ON THE 1-MATRIX

Suppose H is of the form of Eq. (7). Then, if $\{\phi_i\}_{i=1}^r$ are the eigenfunctions of h with eigenvalues μ_i ($\mu_i \leq \mu_{i+1}$) in a truncated space of dimension r,

$$h = \sum_{i=1}^{r} \mu_i |\phi_i\rangle\langle\phi_i|. \tag{16}$$

The eigenvalues of H are now

$$E_{(i_1, \cdots, i_N)} = \sum_{j=1}^{N} \mu_{i_j}, \tag{17}$$

with wavefunctions

$$\Psi = \det |\phi_{i_1}(1), \cdots, \phi_{i_N}(N)|. \tag{18}$$

Hence,

$$E_{\max} = \sum_{i=r-N+1}^{r} \mu_i \tag{19}$$

and

$$E_{\min} = \sum_{i=1}^{N} \mu_i. \tag{20}$$

If

$$\lambda_i = \langle\phi_i|\, \gamma \,|\phi_i\rangle \equiv \int \phi_i^*(1)\gamma(1\,;\,1')\phi_i(1')\, dX_1\, dX_1', \tag{21}$$

then

$$\bar{E} = \sum_{i=1}^{r} \lambda_i \mu_i. \tag{22}$$

The fermion conditions on γ are now easily obtained if Eq. (15) is to hold for every choice of the ϕ_i and μ_i. First, $\bar{E}$ must be real, so the λ_i must be real for every ϕ_i. Hence γ must be a Hermitian kernel. Now if $\mu_r = 1$ and $\mu_i = 0$, $i < r$, then

$$0 \leq \lambda_r \leq 1 \tag{23}$$

must hold for every ϕ_r. Hence the eigenvalues of γ must all lie between zero and one. Further, if all $\mu_i = 1$, then

$$N \leq \sum_{i=1}^{r} \lambda_i \leq N,$$

so

$$\mathrm{Tr}\,(\gamma) = \sum_{i=1}^{r} \lambda_i = N. \tag{24}$$

These conditions will now be shown to be sufficient:

$$\sum_{i=1}^{r} \mu_i\lambda_i = \sum_{i=1}^{N} \mu_i - \sum_{i=1}^{N} \mu_i(1 - \lambda_i) + \sum_{i>N}^{r} \mu_i\lambda_i;$$

but by Eqs. (23) and (24), $(1 - \lambda_i)$ and λ_i are positive, so

$$\sum_{i=1}^{r} \mu_i\lambda_i \geq E_{\min} - \mu_N \sum_{i=1}^{N} (1 - \lambda_i) + \mu_{N+1} \sum_{i>N}^{r} \lambda_i.$$

But by Eq. (24),

$$\sum_{i=1}^{N} (1 - \lambda_i) = \sum_{i>N}^{r} \lambda_i = \Delta \geq 0,$$

so

$$\sum_{i=1}^{r} \mu_i\lambda_i \geq E_{\min} + (\mu_{N+1} - \mu_N)\Delta \geq E_{\min}.$$

A similar proof holds for the upper bound, so Eqs. (23) and (24) are sufficient for fermion-ensemble N-representability.

For an ensemble with variable N, Eq. (23) is still true, and Eq. (24) becomes

$$0 \leq \mathrm{Tr}\,(\gamma) \leq r, \tag{25}$$

since

$$E_{\max} = \max_{(N)} \sum_{i=r-N+1}^{r} \mu_i$$

and

$$E_{\min} = \min_{(N)} \sum_{i=1}^{N} \mu_i.$$

These conditions are sufficient for ensemble representability. The general conditions for pure-state N-representability may not be derived in this manner and are still unknown.

[2] F. Sasaki, Phys. Rev. **138**, 1338 (1965).
[3] C. Garrod and J. K. Percus, J. Math. Phys. **5**, 1756 (1964).
[4] F. Weinhold and E. B. Wilson, Jr., J. Chem. Phys. **47**, 2298 (1967).
[5] H. Kummer, J. Math. Phys. **8**, 2063 (1967).

For bosons,

$$E_{\max} = N\mu_r$$

and

$$E_{\min} = N\mu_1,$$

so the ensemble N-representability conditions are

$$0 \leq \lambda_i \leq N \tag{26}$$

and

$$\mathrm{Tr}\,(\gamma) = \sum_{i=1}^{r} \lambda_i = N. \tag{27}$$

For variable N, Eqs. (26) and (27) should be replaced by

$$0 \leq \mathrm{Tr}\,(\gamma) \leq \infty,$$
$$0 \leq \lambda_i \leq \infty,$$

since there is no upper bound to the number of bosons which may occupy a one-particle state.

These conditions on γ are, of course, not new[6] and are merely presented here as an illustration of the results obtainable from Eq. (15). The simplicity of these results is in direct contrast with the complexity of the results obtained in the next section for the 2-matrix.

II. CONDITIONS ON THE 2-MATRIX

A. General Considerations for Ensemble N-Representability

Let us consider operators H of the form of Eq. (2) on a finite linear vector space $\mathcal{R}$. Then every H, except multiples of the identity, may be mapped by a transformation into a new operator H' which has $E'_{\max} = \delta$ and $E'_{\min} = \epsilon$. Thus, if

$$H' = \alpha(H - \beta\mathbf{1}),$$

where α and β are real numbers and $\alpha > 0$, the choice

$$\beta = E_{\max} - \alpha^{-1}\delta, \tag{28}$$

$$\alpha = (\delta - \epsilon)/(E_{\max} - E_{\min}) > 0 \tag{29}$$

gives the correct H'. Further, if Γ is correctly normalized so that

$$\mathrm{Tr}\,(\Gamma) = N(N-1), \tag{30}$$

the inequality (15) is equivalent to

$$\epsilon \leq \bar{E}' \leq \delta. \tag{31}$$

This may be seen from the fact that

$$g' = \alpha g - \alpha\beta N^{-1}(N-1)^{-1}$$

so

$$\bar{E}' = \mathrm{Tr}\,(g'\Gamma) = \alpha\,\mathrm{Tr}\,(g\Gamma) - \alpha\beta.$$

Equation (31) then becomes

$$\alpha^{-1}\epsilon + \beta \leq \mathrm{Tr}\,(g\Gamma) \leq \alpha^{-1}\delta + \beta.$$

But if Eqs. (28) and (29) are solved for $E_{\max}$ and $E_{\min}$,

$$E_{\max} = \alpha^{-1}\delta + \beta,$$
$$E_{\min} = \alpha^{-1}\epsilon + \beta.$$

So Eq. (31) implies Eq. (15). Thus it is sufficient to use Eq. (30) and the inequalities resulting from those H with $E_{\max} = \delta$ and $E_{\min} = \epsilon$.

Now consider the sets of operators

$$\mathfrak{S} = \left\{H \mid E_{\max} = \delta, E_{\min} = \epsilon, H = \sum_{i \neq j} g(i,j)\right\} \tag{32}$$

and

$$\mathfrak{K} = \left\{H \mid E_{\max} \leq \delta, E_{\min} \geq \epsilon, H = \sum_{i \neq j} g(i,j)\right\}. \tag{33}$$

Since $\mathfrak{S} \subset \mathfrak{K}$, satisfying Eq. (31) for all elements of $\mathfrak{K}$ is equivalent to satisfying it for all elements of $\mathfrak{S}$, even though for the interior points of $\mathfrak{K}$ stronger inequalities could be written. But for elements of $\mathfrak{K}$, a linear combination $H = \sum X_i H_i$ with $X_i > 0$ and $\sum_i X_i = 1$ satisfies $E_{\max} \leq \delta$ and $E_{\min} \geq \epsilon$, so H is in $\mathfrak{K}$. Further, if Eq. (31) is satisfied for all the H_i, so that

$$\epsilon \leq \mathrm{Tr}\,(g_i\Gamma) \leq \delta,$$

then

$$\epsilon = \sum X_i\epsilon \leq \mathrm{Tr}\,(g\Gamma) \leq \sum X_i\delta = \delta.$$

So Eq. (31) is satisfied for H and is a weaker inequality for H than for some of the H_i. Thus $\mathfrak{K}$ is a convex set and the set $\mathfrak{S}$ is part of its boundary. Among the boundary points of $\mathfrak{K}$, there are some, called extreme points, which are not linear combinations of any others. These provide the strongest inequalities, and all other elements of $\mathfrak{K}$ (and hence all other inequalities) are linear combinations of these. Hence one solution to the representability problem would be to find the extreme points of $\mathfrak{K}$. Further, if only calculations involving a particular H_s were to be done, it would be sufficient to know enough of the extreme points K_i to allow the expansion of H'_s (the mapping of H_s in $\mathfrak{S}$) in the K_i:

$$H'_s = \sum X_i K_i.$$

If the inequality (31) were satisfied for just these extreme points, $\bar{E}_s$ would necessarily be an allowed energy of H_s. Although this argument, strictly speaking, holds only for bounded operators, it applies to any truncated representation of H_s in a finite vector space $\mathcal{R}$ for which all matrix elements of H_s are bounded. Thus these conditions would be sufficient for calculations as they are normally performed.

[6] A. J. Coleman, Rev. Mod. Phys. 35, 668 (1963).

The extreme points of $\mathcal{H}$ are not yet known. Some progress can be made, however, by considering for each H in $\mathcal{H}$ the sets

$$\mathcal{K}_H = \{\Psi \mid H\Psi = \delta\Psi\} \tag{34}$$

and

$$\mathcal{L}_H = \{\Psi \mid H\Psi = \epsilon\Psi\}, \tag{35}$$

where the Ψ are elements of the truncated vector space $\mathcal{R}$. Now if H is not an extreme point of $\mathcal{H}$, it may be written as the average of two other members of $\mathcal{H}$, say J and K,

$$H = xJ + yK, \quad x > 0, \quad y > 0, \quad x + y = 1.$$

But then

$$\langle\Psi| H |\Psi\rangle = x\langle\Psi| J |\Psi\rangle + y\langle\Psi| K |\Psi\rangle,$$

so

$$\langle\Psi| H |\Psi\rangle = \delta$$

only if $\Psi \in \mathcal{K}_J$ and $\Psi \in \mathcal{K}_K$. Thus

$$\mathcal{K}_H = \mathcal{K}_J \cap \mathcal{K}_K, \tag{36}$$

$$\mathcal{L}_H = \mathcal{L}_J \cap \mathcal{L}_K. \tag{37}$$

Further, if for some H, $\mathcal{K}_H \cup \mathcal{L}_H$ is empty, then $H \pm \alpha\mathbf{1}$ is still in $\mathcal{H}$ for α sufficiently small and $H = \frac{1}{2}(H + \alpha\mathbf{1}) + \frac{1}{2}(H - \alpha\mathbf{1})$. Hence if $\mathcal{K}_H \cup \mathcal{L}_H$ is empty, H is not extreme. Those extreme points of $\mathcal{H}$ required to express the inequalities in $\mathcal{G}$ can be seen from Eqs. (32), (34), and (35) to have neither $\mathcal{K}_H$ nor $\mathcal{L}_H$ empty. There are some extreme points of $\mathcal{H}$ with either $\mathcal{K}_H$ or $\mathcal{L}_H$ empty, however. For example, by Eqs. (36) and (37) it follows that $\delta\mathbf{1}$ and $\epsilon\mathbf{1}$ are extreme points of $\mathcal{H}$.

For a different choice of ϵ and δ, a new set

$$\mathcal{H}' = \{H' \mid E_{\max} \leq \delta' \text{ and } E_{\min} \geq \epsilon'\}$$

results. If the elements of $\mathcal{H}'$ are mapped into the elements of $\mathcal{H}$ by the transformation

$$H = \alpha(H' - \beta\mathbf{1}),$$

with

$$\alpha = (\delta - \epsilon)/(\delta' - \epsilon')$$

and

$$\beta = (\epsilon'\delta - \delta'\epsilon)/(\delta - \epsilon),$$

the linear combinations $H' = \sum X_i H_i'$ map into $H = \sum X_i H_i$ for $X_i > 0$, $\sum X_i = 1$, where each H_i' maps into the corresponding H_i. So the inequalities in $\mathcal{H}'$ are equivalent to their images in $\mathcal{H}$, and the extreme points of $\mathcal{H}'$ are mapped into extreme points of $\mathcal{H}$. Thus every choice of ϵ and δ gives an equivalent set of extreme points. In order to simplify the proofs, it is convenient to choose $\delta = -\epsilon = 1$ so that H and $-H$ are both in $\mathcal{H}$, and $-H$ is extreme if H is extreme. With this choice, only half of the inequalities need be considered since those for H and $-H$ are identical. Hence, let us specialize to the set

$$\mathcal{H}^* = \left\{H \mid E_{\max} \leq 1, E_{\min} \geq -1, H = \sum_{i \neq j} g(i,j)\right\}. \tag{38}$$

With this set $\mathcal{L}_H = \mathcal{K}_{-H}$ and $\mathcal{K}_H = \mathcal{L}_{-H}$. Also, if there is an element H of $\mathcal{H}^*$ with $\mathcal{K}_H \cup \mathcal{L}_H$ empty, then αH is still in H^* for some values of $\alpha > 1$.

Now suppose H and J are elements of $\mathcal{H}^*$ with

$$\mathcal{K}_H \subseteq \mathcal{K}_J$$

and

$$\mathcal{L}_H \subseteq \mathcal{L}_J.$$

Then, for every $1 \geq \epsilon \geq 0$,

$$P = (1 - \epsilon)H + \epsilon J \in \mathcal{H}^*.$$

Further, $Q = (1 + \epsilon)H + \epsilon(-J)$ is in $\mathcal{H}^*$ for $\epsilon > 0$ sufficiently small even though $1 + 2\epsilon > 1$. This may be seen from the fact that

$$\begin{aligned}\langle\Psi| Q |\Psi\rangle &= 1, \quad \Psi \in \mathcal{K}_H, \\ &= -1, \quad \Psi \in \mathcal{L}_H, \\ &= (1 + \epsilon)\langle\Psi| H |\Psi\rangle - \epsilon\langle\Psi| J |\Psi\rangle, \\ &\qquad \Psi \notin \mathcal{K}_H \cup \mathcal{L}_H.\end{aligned}$$

But in this last case, $\langle\Psi| Q |\Psi\rangle$ will be less than 1 if

$$\epsilon < (1 - \langle\Psi| H |\Psi\rangle)/(1 + \langle\Psi| H |\Psi\rangle)$$

and greater than -1 if

$$\epsilon < (1 + \langle\Psi| H |\Psi\rangle)/(1 - \langle\Psi| H |\Psi\rangle).$$

Thus if the least value of the right-hand side of these inequalities is selected over all $\Psi \notin \mathcal{K}_H \cup \mathcal{L}_H$, an $\epsilon > 0$ will be found such that $Q \in \mathcal{H}^*$. But then

$$H = \tfrac{1}{2}P + \tfrac{1}{2}Q,$$

so H is not an extreme point.

If for any H in $\mathcal{H}^*$ there exists another operator J in $\mathcal{H}^*$ with

$$\mathcal{K}_H \subseteq \mathcal{K}_J$$

and

$$\mathcal{L}_H \subseteq \mathcal{L}_J,$$

then H is not an extreme point of $\mathcal{H}^*$. Conversely, by Eqs. (36) and (37), *if there is no such operator J, then H must be extreme.*

It follows at once from this that $\mathbf{1}$ is the only extreme point with $\mathcal{L}$ empty and $-\mathbf{1}$ is the only extreme point with $\mathcal{K}$ empty. For any other selection of $\mathcal{L}$ and $\mathcal{K}$, there are three possibilities:

(i) No H of the form of Eq. (2) exists in $\mathcal{H}^*$. In this case there are no extreme points H for which

$\mathfrak{L} \subseteq \mathfrak{L}_H$ and $\mathfrak{K} \subseteq \mathfrak{K}_H$ and some Ψ must be dropped from $\mathfrak{L}$ or $\mathfrak{K}$ in order to obtain an extreme point.

(ii) Exactly one H of the form of Eq. (2) exists in $\mathcal{K}^*$. In this case H is an extreme point.

(iii) Several H_i of the form of Eq. (2) exist in $\mathcal{K}^*$. In this case $\mathfrak{L}$ or $\mathfrak{K}$ is not large enough, since by the construction of Q given above it is possible to find an operator in $\mathcal{K}^*$ for which

$$\mathfrak{L} \subseteq \mathfrak{L}_Q$$

and

$$\mathfrak{K} \subseteq \mathfrak{K}_Q,$$

and equality does not hold in both cases.

Clearly for the case that $\mathfrak{L} \cup \mathfrak{K}$ is the whole truncated vector space, there is only one operator H for each choice of $\mathfrak{K}$. If this can be written in the form of Eq. (2), it is an extreme point of $\mathcal{K}^*$; if not, there is no extreme point of that form. In any case, if the set Ψ_i spans the space $\mathfrak{K}$, the set ϕ_i spans the space $\mathfrak{L}$, and the set χ_i spans the residual of the finite vector space $\mathfrak{R}$, it can always be written as

$$H = \sum_i |\Psi_i\rangle\langle\Psi_i| - \sum_i |\phi_i\rangle\langle\phi_i| + \sum_{ij} \mu_{ij} |\chi_i\rangle\langle\chi_j|, \quad (39)$$

where μ is a Hermitian matrix with eigenvalues less than 1 in absolute value. The question, then, is whether there is a choice of the μ_{ij} for which H can also be written in the form of Eq. (2). It is unfortunately true that the extreme points of $\mathcal{K}^*$ depend on the truncated vector space $\mathfrak{R}$. From Eq. (39) it is clear that an extreme point of $\mathcal{K}^*(\mathfrak{R})$ will not generally be an extreme point for $\mathcal{K}^*(\mathfrak{R}')$.

All of the above discussion applies equally well to the ensemble representability problem for variable N, if Eq. (13) is used for the average energy and $E_{\max}$ is replaced by $\max_{(N)} E_{\max}(N)$ and $E_{\min}$ is replaced by $\min_{(N)} E_{\min}(N)$. In any expression such as $H\Psi$ it is to be understood that $H(1, \cdots, N)$ is used if Ψ involves N particles. The space $\mathfrak{R}$ would contain wavefunctions Ψ of various N.

For the special case of $N = 2$, the extreme points of $\mathcal{K}^*(\mathfrak{R})$ may be written by inspection. Consider every distinct partitioning of $\mathfrak{R}$ into two subspaces $\mathfrak{K}$ and $\mathfrak{L}$, so that $\mathfrak{K} \cup \mathfrak{L} = \mathfrak{R}$. Then, if the $\Psi_i(1, 2)$ span $\mathfrak{K}$ and the $\phi_i(1, 2)$ span $\mathfrak{L}$,

$$H = g(1, 2) + g(2, 1),$$
$$g(1, 2) = \tfrac{1}{2}\{\sum |\Psi_i\rangle\langle\Psi_i| - \sum |\phi_i\rangle\langle\phi_i|\}.$$

A complete set of ensemble 2-representability conditions in $\mathfrak{R}$ then is

$$\mathrm{Tr}\,(\Gamma) = 2,$$
$$-2 \leq \sum_i \langle\Psi_i|\,\Gamma\,|\Psi_i\rangle - \sum_i \langle\phi_i|\,\Gamma\,|\phi_i\rangle \leq 2,$$

where

$$\langle\Psi|\,\Gamma\,|\Psi\rangle = \int \Psi^*(1, 2)\Gamma(1, 2; 1', 2') \times \Psi(1', 2')\, dX_1\, dX_1'\, dX_2\, dX_2'.$$

For a different choice of ϵ and δ, i.e., $\epsilon = 0$, $\delta = 1$, these conditions may be written in the more transparent form

$$0 \leq \sum_i \langle\Psi_i|\,\Gamma\,|\Psi_i\rangle \leq 2,$$

which is true provided

$$\langle\Psi_i|\,\Gamma\,|\Psi_i\rangle \geq 0, \quad \text{for all } \Psi_i,$$

and

$$\mathrm{Tr}\,(\Gamma) = 2.$$

That is, for $N = 2$, it is necessary and sufficient that Γ be positive-definite and properly normalized. For pure-state representability, it is obvious that

$$|\langle\Psi_i|\,\Gamma\,|\Psi_j\rangle|^2 = \langle\Psi_i|\,\Gamma\,|\Psi_i\rangle\langle\Psi_j|\,\Gamma\,|\Psi_j\rangle,$$

which is a much stronger requirement, since it is equivalent to saying Γ has only one nonzero eigenvalue.

An alternative convex set may be defined. Every H of the form of Eq. (2) can be mapped into the set

$$\mu = \left\{H \mid E_{\min} = 0, H = \sum_{i \neq j} g(i, j)\right\} \quad (40)$$

by the mapping used previously. Since the upper bound is not specified, the mapping is not unique. This set is part of the boundary of the set

$$v = \left\{H \mid E_{\min} \geq 0, H = \sum_{i \neq j} g(i, j)\right\}. \quad (41)$$

Since every upper-bound condition on H is equivalent to a lower-bound condition on $-H$, it is sufficient to consider lower bounds only. Since every condition in μ is also in v, the inequalities

$$\mathrm{Tr}\,(g\Gamma) \geq 0, \quad \text{for} \quad H \in v, \quad (42)$$

contain all the information in Eq. (15) if $\mathrm{Tr}\,(\Gamma) = N(N-1)$. The set v is a convex cone since every H of the form

$$H = \sum X_i H_i, \quad X_i > 0, \quad \text{and} \quad H_i \in v$$

is also in v (regardless of $\sum_i X_i$). Further, if the inequalities [Eq. (42)] hold for the H_i, they hold, and are weaker, for H itself. The set v may then be characterized by its extreme rays—those operators H in v which are not linear combinations of other operators in v with positive coefficients, but which are arbitrary within a positive multiplicative constant.

Now consider for each H in v the set $\omega_H = \{\Psi \mid H\Psi = 0\}$. From the previous discussion [see

Eq. (37)], it is obvious that if J, $K \in v$ and $H = xJ + yK$ $(x, y > 0)$, then

$$\omega_H = \omega_J \cap \omega_K. \tag{43}$$

Further, if ω_H is empty, H is not extreme in v because

$$H = \tfrac{1}{2}(H + \alpha\mathbf{1}) + \tfrac{1}{2}(H - \alpha\mathbf{1}),$$

and $H \pm \alpha\mathbf{1}$ is in v for α sufficiently small. Also if H and J are in v and

$$\omega_H \subseteq \omega_J,$$

then $H \pm \epsilon J$ is in v for $\epsilon > 0$ and sufficiently small. For $H - \epsilon J$, it is only necessary to choose

$$\langle\Psi| H |\Psi\rangle - \epsilon \langle\Psi| J |\Psi\rangle > 0, \quad \text{for all} \quad \Psi \notin \omega_H.$$

There is a maximum nonzero choice of ϵ which satisfies this equation. But then

$$H = \tfrac{1}{2}(H + \epsilon J) + \tfrac{1}{2}(H - \epsilon J),$$

so H is not extreme.

If for any H in v there exists another operator J in v with $\omega_H \subseteq \omega_J$, then H is not on an extreme ray of v. Conversely, by Eq. (43), *if there is no such operator J, then H is extreme.*

Hence, for a given ω spanned by the set of functions Ψ_i, the residual of the vector space $\mathfrak{R}$ may be spanned by a set χ_i and H will have the form

$$H = \sum \mu_{ij} |\chi_i\rangle\langle\chi_j|,$$

where μ is a positive-definite Hermitian matrix. The question, then, is whether this H may also be expressed in the form of Eq. (2) for some choice of μ, and if so, is the choice of μ unique (within a multiplicative constant).

For the special case of $N = 2$, ω may be chosen to be the whole space $\mathfrak{R}$ orthogonal to one element χ. Then

$$H = g(1, 2) + g(2, 1),$$

where

$$g = \tfrac{1}{2} |\chi\rangle\langle\chi|$$

is on an extreme ray of v with the corresponding inequality

$$\langle\chi| \Gamma |\chi\rangle > 0.$$

Further, by the theorems given above, all extreme rays are of this form. Together with the trace condition

$$\mathrm{Tr}\,(\Gamma) = 2,$$

this provides a complete set of conditions on Γ for $N = 2$. It will be noticed that v is more convenient in this case than $\mathfrak{N}^*$ and has fewer extreme rays than $\mathfrak{N}^*$ has extreme points.

B. Subclass of Conditions on the 2-Matrix

Consider a fermion system of N particles:

$$H = \sum_{i \neq j}^{N} g(ij),$$

where

$$g(ij) = \sum_{k,l}^{r} \alpha_{kl} |\phi_k(i)\phi_l(j)\rangle\langle\phi_k(i)\phi_l(j)| \tag{44}$$

and

$$S = \{\Phi_I \mid \Phi_I = (N!)^{-\frac{1}{2}} \det |\phi_{i_1}(1), \cdots, \phi_{i_N}(N)|, \quad 1 \leq i_1 < i_2 \cdots < i_N \leq r,\ I = \{i_1, \cdots, i_N\}\}. \tag{45}$$

Since the identity operator is in this set

$$g = N^{-1}(N-1)^{-1} \sum_{k,l}^{r} |\phi_k\phi_l\rangle\langle\phi_k\phi_l|,$$

all of the previous theorems remain true for this subset of operators. But for this simplified form, the Φ_I are the eigenfunctions of H in S with eigenvalues

$$E_I = \sum_{j \neq k} \alpha_{i_j i_k}. \tag{46}$$

Hence, the requirement that H be in $\mathfrak{N}^*$ is simply

$$-1 \leq E_I \leq 1, \quad \text{for all } I, \tag{47}$$

and the requirement that H be in v is just

$$E_I \geq 0, \quad \text{for all } I. \tag{48}$$

Further, this set of operators is not a trivial subset. If

$$\lambda_{ij} = \langle\phi_i\phi_j| \Gamma |\phi_i\phi_j\rangle = \int \phi_i^*(1)\phi_j^*(2)\Gamma\phi_i(1')\phi_j(2')\, dX_1\, dX_1'\, dX_2\, dX_2', \tag{49}$$

then

$$\lambda_i = (N-1)^{-1} \sum_j \lambda_{ij}.$$

The following examples are of interest (all α_{ij} not specified are zero):

$$\alpha_{kk} = 1 \Rightarrow 0 \leq \lambda_{kk} \leq 0, \tag{50}$$

$$\alpha_{ij} = -\alpha_{ji} = 1 \Rightarrow 0 \leq \lambda_{ij} - \lambda_{ji} \leq 0, \tag{51}$$

$$\alpha_{ij} = 1\ (i \neq j) \Rightarrow 0 \leq \lambda_{ij} \leq 1, \tag{52}$$

$$\alpha_{ij} = 1 \quad \text{for} \quad j = 1, \cdots, r \Rightarrow 0 \leq \lambda_i \leq 1, \tag{53}$$

$$\alpha_{ij} = 1 \quad \text{all} \quad i, j \Rightarrow N(N-1) \leq \sum_{i,j}^{r} \lambda_{ij} \leq N(N-1) \tag{54}$$

$$\Rightarrow N \leq \sum_{i}^{r} \lambda_i \leq N, \tag{55}$$

$$\alpha_{ik} = -1, \quad k = 1, \cdots, r,$$

except

$$\alpha_{ij} = N - 2 \Rightarrow \lambda_{ij} \leq \lambda_i, \tag{56}$$

$$\alpha_{ik} = -1, \alpha_{kj} = -1, k = 1, \cdots, r,$$

except

$$\alpha_{ij} = N - 3 \Rightarrow 0 \leq 1 - \lambda_i - \lambda_j + \lambda_{ij}, \tag{57}$$

$$\alpha_{ip} = -1, p = 1, \cdots, r,$$

except

$$\alpha_{ij} = N - 2 = \alpha_{ik},$$

$$\alpha_{kj} = -(N-1) \Rightarrow \lambda_{ij} + \lambda_{ik} \leq \lambda_i + \lambda_{kj}, \tag{58}$$

$$\alpha_{ip} = \alpha_{j,p} = \alpha_{p,k} = -1, p = 1, \cdots, r,$$

except

$$\alpha_{ik} = \alpha_{jk} = \alpha_{ij} = N - 3, \alpha_{ji} = 0,$$

$$\Rightarrow 0 \leq 1 - \lambda_i - \lambda_j - \lambda_k + \lambda_{ik} + \lambda_{jk} + \lambda_{ij}. \tag{59}$$

This list includes all of the Weinhold linear inequalities.[4]

In fact, the restricted set of H

$$\tilde{H} = \left\{ H \mid E_{\max} \leq 1, E_{\min} \geq -1, H = \sum_{i \neq j} g(i,j), \right.$$
$$\left. g = \sum \alpha_{kl} |\phi_k \phi\rangle\langle\phi_k \phi| \right\},$$

or

$$\tilde{v} = \left\{ H \mid E_{\min} \geq 0, H = \sum_{i \neq j} g(i,j), \right.$$
$$\left. g = \sum \alpha_{kl} |\phi_k \phi_l\rangle\langle\phi_k \phi_l| \right\},$$

can generate all of the linear inequalities which can be expressed using only the λ_{ij}. This follows because $\mathrm{Tr}(g\Gamma)$ would involve other quantities if g had nondiagonal terms. If the orbitals ϕ_i are subjected to a unitary transformation $\phi_i = \sum U_{ji}\phi'_j$, then similar inequalities must hold for every λ'_{ij}. Even the requirement that all these inequalities hold simultaneously for every unitary transformation is not a complete set of restrictions, however.[5] For instance, for every extreme ray in v there must be some function $\Psi \in \omega$ and some other element $\Psi' \notin \omega$ (or else $g \equiv 0$). For the extreme points of $\mathcal{K}^*$ there must be at least one $\Psi \in \mathfrak{K}$ and one $\Psi' \in \mathfrak{L}$. In either case, specification of Ψ and Ψ' specifies $2\binom{r}{N} - 3$ free parameters if Ψ and Ψ' are an orthonormal pair. But if this H corresponding to Ψ and Ψ' could be written in terms of the extreme points of $\tilde{\mathcal{K}}$ or $\tilde{v}$, it would imply

$$\Psi = \sum C'_I \Phi'_I, \quad I \text{ in } \mathfrak{K} \text{ (or } I \in \omega),$$
$$\Psi' = \sum C'_I \Phi'_I, \quad I \text{ in } \mathfrak{L} \text{ (or } I \notin \omega),$$

with

$$\Phi'_I = (N!)^{-\frac{1}{2}} \det |\phi'_{i_1}(1) \cdots \phi'_{i_N}(N)|.$$

But this last form contains only $\binom{r}{N} - 2$ free coefficients C'_I, and $\binom{r}{2}$ free coefficients U_{ij}. Hence, in general, an element of $\mathfrak{S}$ or μ is not in any of the sets $\tilde{\mathcal{K}}'$ or $\tilde{v}'$.

For example, for $N = 2$, the complete set of extreme points of $\tilde{\mathcal{K}}$ are given by partitioning the Slater determinants from $\mathfrak{S}$ into two sets $\mathfrak{K}$ and $\mathfrak{L}$. This gives the inequalities

$$\sum_{i,j} \lambda_{ij} = 2,$$
$$-2 \leq \sum_{ij} \sigma_{ij}\lambda_{ij} \leq 2,$$

where $\sigma_{ij} = (\pm 1)_{ij}$, and there are 2^z extreme points [where $z = \binom{r}{2}$]. The extreme rays of $\tilde{v}$ give the inequalities

$$\lambda_{ij} \geq 0,$$

and there are only $\binom{r}{2}$ extreme rays. For every matrix $\boldsymbol{\omega}$, however,

$$\Psi_i = \sum \omega_{Ii}\Phi_I,$$

there are $\binom{r}{2}$ extreme rays of v, and the total number of extreme rays of v is infinite and nondenumerable.

Now let us deal only with Γ of the correct fermion symmetry, so $\lambda_{ij} = \lambda_{ji}$ and $\lambda_{ii} = 0$. In this case it is sufficient to consider

$$g(i,j) = \sum_{l>k}^{r} \alpha_{kl} |\phi_k \phi_l\rangle\langle\phi_k \phi_l|,$$

for which

$$E_I = \sum_{l>k}^{N} \alpha_{i_k i_l}.$$

The extreme points of the set $\tilde{\mathcal{K}}$ may then be found by finding H such that

$$\mathfrak{K} = \left\{ I \mid I = \{i_1, \cdots, i_N\}, \right.$$
$$\left. 1 \leq i_1 < i_2 \cdots < i_N \leq r, \sum_{l>k}^{N} \alpha_{i_k i_l} = 1 \right\},$$

$$\mathfrak{L} = \left\{ I \mid I = \{i_1, \cdots, i_N\}, \right.$$
$$\left. 1 \leq i_1 < i_2 \cdots < i_N \leq r, \sum_{l>k}^{N} \alpha_{i_k, i_l} = -1 \right\}.$$

This leads to the set of linear equations

$$\sum_{l>k}^{N} \alpha_{i_k i_l} = 1, \qquad I \in \mathfrak{K},$$

$$\sum_{l>k}^{N} \alpha_{i_k i_l} = -1, \quad I \in \mathfrak{L},$$

and the inequalities

$$\left| \sum_{l>k}^{N} \alpha_{i_k i_l} \right| \leq 1, \quad I \notin \mathfrak{K} \cup \mathfrak{L}.$$

These may be solved by a systematic procedure. Since there are $\binom{r}{2}$ variables α_{ij}, there are only $\binom{r}{2}$ linearly independent sums in the set of $\binom{r}{N}$ possible sums. If $\binom{r}{2}$ independent sums are picked from the $\binom{r}{N}$ possible, and partitioned between $\mathcal{K}$ and $\mathcal{L}$ in every possible way, there results

$$2^{\binom{r}{2}}\binom{\binom{r}{N}}{\binom{r}{2}}$$

sets of equations, each one of which may correspond to an extreme point of $\tilde{\mathcal{K}}$. If the solution to one of these sets of equalities does not satisfy the inequalities, there is no extreme point with $\mathcal{K} \subseteq \mathcal{K}_H$ and $\mathcal{L} \subseteq \mathcal{L}_H$. If the inequalities are satisfied, then the solution is an extreme point of $\tilde{\mathcal{K}}$. Since generally $\mathcal{K} \subset \mathcal{K}_H$ and $\mathcal{L} \subset \mathcal{L}_H$, many of the sets $\mathcal{K}$, $\mathcal{L}$ will generate the same extreme point.

The extreme rays of $\tilde{v}$ may be generated in a similar manner. The extreme rays satisfy the conditions

$$\sum_{l>k}^{N} \alpha_{i_k i_l} = 0, \quad I \in \omega,$$

and

$$\sum_{l>k}^{N} \alpha_{i_k i_l} \geq 0, \quad I \notin \omega.$$

The set of equalities has a unique solution (within an arbitrary constant) only if the rank is $\binom{r}{2} - 1$. Thus, if $\binom{r}{2} - 1$ sums are picked out of the $\binom{r}{N}$ possible, there results

$$\binom{\binom{r}{N}}{\binom{r}{2} - 1}$$

sets of equations, each one of which may give an extreme ray of $\tilde{v}$. Again, if the solution to one of these systems of equalities does not satisfy the inequalities, there is no extreme point of $\tilde{v}$ with $\omega \subseteq \omega_H$. If all the inequalities are satisfied, then the solution is an extreme ray of $\tilde{v}$. Usually $\omega \subset \omega_H$, so many of the sets of equations generate the same extreme ray.

This calculation has been carried out for $\tilde{v}$, for $N = 3$ and $r = 5$ and 6. The results are the same as the inequalities given by Kuhn.[7] These may be written in the notation of Weinhold and Wilson as

$$\begin{aligned} r = 5, N = 3, \quad & 1 - \lambda_i - \lambda_j + \lambda_{ij} \geq 0, \\ r = 6, N = 3, \quad & 1 - \lambda_i - \lambda_j + \lambda_{ij} \geq 0, \\ & \lambda_i - \lambda_{ij} \geq 0, \\ & 1 - \lambda_i - \lambda_j - \lambda_j + \lambda_{ij} + \lambda_{ik} + \lambda_{jk} \geq 0, \\ & \lambda_{ij} \geq 0. \end{aligned}$$

It will be noticed that both of these examples lead to Weinhold inequalities. This raised the immediate question whether the Weinhold conditions contained all of the extreme rays in general. This was unlikely, since the Weinhold conditions were generated originally (from a completely different viewpoint) by considering operators H which were idempotent. Clearly there should be some extreme rays of $\tilde{v}$ with more than two distinct eigenvalues, and these would give new conditions on Γ. Two examples are easily found:

$$1 - \sum_{i=1}^{q} \lambda_i + \sum_{i<j}^{q} \lambda_{ij} \geq 0, \quad q = 1, 2, 3, \cdots, \tag{60}$$

$$\lambda_{q+1} - \sum_{i=1}^{q} \lambda_{i,q+1} + \sum_{i<j}^{q} \lambda_{ij} \geq 0, \quad q = 1, 2, \cdots. \tag{61}$$

These are Weinhold conditions for $q = 1$, 2 and are new conditions for $q > 2$. They have been verified in some cases to be extreme rays of $\tilde{v}$ for r sufficiently larger than $N + q$. For large r and N, the sets of linear equations for the extreme rays become too numerous, so it is impractical to carry out the calculations.

For the example of g, given by Eq. (44), a second approach is possible which is more practical, since the results are independent of r and N. This new set of inequalities may be generated by considering the general form of the Weinhold and Wilson results,

$$f = \alpha + \sum_{i=1}^{q} \beta_i \lambda_i + \sum_{i<j}^{q} \gamma_{ij} \lambda_{ij}.$$

This f is positive-semidefinite, provided that it is positive or zero for every Ψ_I in $\mathcal{S}$. This follows because f may be generated from Tr $(g\Gamma)$; so, for a legitimate Γ, we have

$$f \geq E_{\min} = \min_I \operatorname{Tr}(g\Gamma_I) = \min_I f_I.$$

If one lets (i) denote the subset of Slater determinants (for $r \geq N \geq q$) with ϕ_i occupied, $(\bar{i})$ the set with ϕ_i empty, and $(i\bar{j})$ the set with ϕ_i occupied and ϕ_j empty, etc., then the Ψ_I may be split into 2^q subsets according to the occupancy of q of the orbitals. The expression f gives the same value for all Ψ_I in one of these subsets. So if $\mathcal{J}$ is the set of all these subsets for which f is zero,

[7] H. W. Kuhn, Proc. Sym. Applied Math. **10**, 141 (1960).

TABLE I. Some conditions on the 2 matrix.

		p, s, t
$q = 1$	$1 - \lambda_i \geq 0$	1, 0, 1
	$\lambda_i \geq 0$	0, 1, 0
$q = 2$	$1 - \lambda_i - \lambda_j + \lambda_{ij} \geq 0$	2, 0, 1
	$\lambda_i - \lambda_{ij} \geq 0$	1, 1, 0
	$\lambda_{ij} \geq 0$	0, 2, −1
$q = 3$	$1 - \lambda_i - \lambda_j - \lambda_k + \lambda_{ij} + \lambda_{ik} + \lambda_{jk} \geq 0$	3, 0, 1
	$\lambda_i + \lambda_{kj} - \lambda_{ij} - \lambda_{ik} \geq 0$	2, 1, 0
	+ all permutations of those for $q = 2$	
$q = 4$	$1 - \lambda_i - \lambda_j - \lambda_k - \lambda_l + \lambda_{ij} + \lambda_{ik} + \lambda_{il} + \lambda_{jk} + \lambda_{jl} + \lambda_{kl} \geq 0$	4, 0, 1
	$\lambda_i + \lambda_{kj} + \lambda_{lj} + \lambda_{kl} + \lambda_{ij} - \lambda_{ik} - \lambda_{il} \geq 0$	3, 1, 0
	$\lambda_i + \lambda_j + \lambda_{ij} + \lambda_{kl} - \lambda_{ik} - \lambda_{il} - \lambda_{jk} - \lambda_{jl} \geq 0$	2, 2, −1
	$1 + 2\lambda_i - \lambda_j - \lambda_k - \lambda_l - \lambda_{ij} - \lambda_{ik} - \lambda_{il} + \lambda_{jk} + \lambda_{jl} + \lambda_{jk} \geq 0$	1, 3, −2
	$3 - 2\lambda_i - 2\lambda_j - 2\lambda_k - 2\lambda_l + \lambda_{ij} + \lambda_{ik} + \lambda_{il} + \lambda_{jk} + \lambda_{jl} + \lambda_{kl} \geq 0$	0, 4, −3
	+ all permutations of those for $q = 2, 3$	

then f generates an independent inequality, provided

$$\mathfrak{d} \nsubseteq \mathfrak{d}'$$

for some other f'. Following the procedure given previously, if $\binom{q+1}{2}$ of the 2^q subsets are selected and the corresponding f_I is set equal to zero, an f may be found. If it is unique, and positive on the other subsets, it is an independent condition for this q. In this way a maximum of

$$\binom{2^q}{\binom{q+1}{2}}$$

inequalities may be found. This has been done for $1 \leq q \leq 4$, and the results are given in Table I. These conditions are a sufficient set of ensemble representability conditions for $r = q$ (within the subspace $\tilde{v}$). They will not be complete for N representability, however, if there exists some f which is excluded from the list because it is negative only on subsets with more than N occupied orbitals. For example,

$$\lambda_i + \lambda_j - \lambda_{ij} - \lambda_{ik} - \lambda_{jl} \geq 0$$

is an extreme ray of $\tilde{v}$ for $r = 10$ and $N = 3$ but is not in Table I for $q = 4$.

Although larger values of q would require too much computer time, the results obtained so far may be generalized. For example, all of the conditions in Table I and in Eqs. (60) and (61) can be generated from the single inequality

$$\tfrac{1}{2}t(t+1) - t\sum_{i=1}^{p}\lambda_{n_i} + (t+1)\sum_{j=1}^{s}\lambda_{m_j} + \sum_{i<j}^{p}\lambda_{n_i n_j} + \sum_{i<j}^{s}\lambda_{m_i m_j} - \sum_{i,j}^{p,s}\lambda_{n_i m_j} \geq 0, \quad (62)$$

where $(n_1, \cdots, n_p)$ and $(m_1, \cdots, m_s)$ are disjoint subsets of $(1, \cdots, r)$, and t is an integer. This may easily be seen to be true. If, for some Ψ_I, j of the ϕ_{n_i} and k of the ϕ_{m_i} are occupied, the left-hand side takes the value

$$\tfrac{1}{2}\{[j - k - t - \tfrac{1}{2}]^2 - \tfrac{1}{4}\},$$

which is nonnegative for all integers j, k, t. Table I gives the values of p, s, and t which generate the inequalities given there. It is conceivable that Eq. (62) would generate all of the ensemble representability conditions which may be expressed using only the λ_i and λ_{ij}. Table I has the obvious pattern

$$(p, s, t) = (q - k, k, 1 - k), \quad k = 0, 1, \cdots, q.$$

Since (p, s, t) and $(s, p, -t - 1)$ give the same inequality, $k = 2, 3$ are not needed for $q = 3$.

III. RELATION TO SOME PREVIOUS CONDITIONS

As shown above, the Weinhold inequalities may easily be derived by these methods and stronger inequalities produced. In fact, it has been shown by Erdahl[8] and Kummer[9] that many of the Weinhold

[8] R. Erdahl, private communication, 1968.
[9] H. Kummer, private communication, 1968.

conditions are extreme points for the less stringent case of fermion ensemble representability (variable N). The nonnegativity of the G matrix and Q matrix[3] generally imposes conditions on the matrix elements of Γ other than λ_i and λ_{ij}, and hence these matrices contain some restrictions from v not contained in $\tilde{v}$. For some particular examples of the G-matrix conditions which involve only the λ_{ij}, however, the Weinhold inequalities are easily shown to be stronger conditions.

The Sasaki upper bound[2] on the eigenvalue of Γ could easily be derived by considering

$$g = |\omega(i,j)\rangle\langle\omega(i,j)|,$$

where

$$\langle\omega(i,j)\,|\,\omega(i,j)\rangle = 1.$$

Then

$$0 \leq \langle\omega|\,\Gamma\,|\omega\rangle \leq E_{\max}$$

and the maximum E for any ω gives the upper bound to the eigenvalues of Γ. The solution of this is known. For even N and r, for example,

$$\omega = \binom{2}{r}^{\frac{1}{2}} \sum_{i \text{ odd}}^{r} \frac{1}{\sqrt{2}} \det|\phi_i\phi_{i+1}|$$

and

$$\Psi = \mathcal{A}[\omega(1,2)\cdots\omega(N-1,N)]$$

gives

$$E_{\max} = N - r^{-1}N(N-2).$$

The general problem of finding the extreme points of $\mathcal{K}^*$ or v is very difficult. First of all, the number of subsets $\mathcal{K}$ and $\mathfrak{L}$ (or ω) which may be selected from $\mathcal{R}$ is not denumerable. The examples of the restricted $\tilde{\mathcal{K}}$ and $\tilde{v}$ were solvable only because the restrictions on H made the number of ways of selecting $\mathcal{K}$ and $\mathfrak{L}$ (or ω) finite. Further, most selections of ω will not lead to an operator H in v. This may be seen from the fact that if $\mathcal{R}$ is $\mathcal{S}$, the space of $\binom{r}{N}$ Slater determinants, then an arbitrary wavefunction in $\mathcal{R}$ contains $\binom{r}{N}$ free parameters while g contains only r^4 [really only $\binom{z}{2}$, where $z = \binom{r}{2}$, if permutation symmetry is built into Γ and g]. It is easy to write a set of $\binom{r}{N}$ linear equations for the r^4 matrix elements g_{ijkl}, such that

$$g = \sum g_{ijkl}\,|\phi_i\phi_j\rangle\langle\phi_k\phi_l|,$$

expressing the fact that

$$H\Psi = E\Psi.$$

For most selections of Ψ these equations have no solution. From this, one might conclude that for many of the extreme rays of v, ω contains only one function Ψ. Because the ways of selecting ω are not denumerable, the number of extreme rays may be infinite, and no systematic searching procedure may be devised to discover them. This is equally true for the sets $\mathcal{K}$ and $\mathfrak{L}$ so the same difficulty exists in $\mathcal{K}^*$.

From these dimensional considerations, one may conjecture that, except for a set of measure zero, all the elements of μ are on extreme rays of v for $\binom{r}{N} \gg r^4$. If this is true, then an arbitrary operator such as the Schrödinger operator H_s almost always maps onto an extreme ray of v when it is mapped into μ. Thus, probably, the condition for the variational stability of $\mathrm{Tr}\,(g_s\Gamma)$,

$$E_{\max}(H_s) \geq \mathrm{Tr}\,(g_s\Gamma) \geq E_{\min}(H_s),$$

is independent of all other ensemble N-representability conditions. Thus, if this conjecture is true, it will always be impossible to calculate $E_{\min}$ by variation of Γ subject to previously known N-representability conditions.

The possibility still appears to exist that sufficient nonlinear inequalities might be found to guarantee variational stability. That this is probably impossible may be seen from the fact that all of the linear inequalities may be replaced by the single nonlinear inequality

$$\min_{g} \{\mathrm{Tr}\,(g\Gamma) - E_{\min}(g)\} \geq 0.$$

This is nonlinear because the g^* for which the expression is minimum is a function of Γ. For a fixed Γ, this single inequality is necessary and sufficient for ensemble N-representability. If $\mathrm{Tr}\,(g_s\Gamma)$ were minimized subject to this constraint, however, g^* would usually turn out to be g_s, and $E_{\min}(g_s)$ would still have to be calculated by other methods.

ACKNOWLEDGMENTS

Most of the work reported here was carried out at the 1968 Density Matrix Conference held at Queen's University, Kingston, Ontario, under the auspices of the Air Force Office of Scientific Research. The author wishes to express his thanks to A. J. Coleman for inviting him to this conference. The author is also thankful to Bob Erdahl and Hans Kummer for several stimulating discussions.

First-Order Approximation to a Spherically Symmetric Solution of Einstein's Equations

R. STABELL
Department of Astronomy, University of Illinois, Urbana, Illinois

(Received 29 May 1968)

In analogy with the metrics of Schwarzschild and de Sitter, a first-order approximation to the spherically symmetric solution of Einstein's field equations is found where the deviation from the flat Minkowski space-time is given by potentials of the gravitational force and the cosmical Λ force. It is shown that in the Newtonian approximation the forces acting per unit mass of the fluid are the assumed gravitational self-attraction and cosmical Λ force.

1. INTRODUCTION

Schlüter[1] has shown that the cosmical constant Λ in Einstein's field equations may be interpreted as a force parameter; empty space acts as an additional source of gravitational potential. In the case of spherical symmetry, Schlüter found in the Newtonian approximation the following expression for the radial force per unit mass apart from the pressure gradient force:

$$F = -GM_r/r^2 + \tfrac{1}{3}\Lambda r, \tag{1}$$

where G is the gravitational constant and M_r is the mass inside a sphere of radius r. With the definition

$$L \equiv \Lambda/4\pi, \tag{2}$$

Eq. (1) may be written

$$F = -GM_r/r^2 + LV_r/r^2, \tag{3}$$

where V_r is the volume of the sphere with radius r. When Schlüter's idea about empty space is maintained, it is seen that L may be interpreted as a gravitational constant associated with space.

2. THE METRICS OF SCHWARZSCHILD AND DE SITTER

The analog of Eq. (1) in the case of a central body with mass M surrounded by empty space would be

$$F = -GM/r^2 + LV_r/r^2, \tag{4}$$

where F now is the force per unit mass of a test particle at distance r. The complete Schwarzschild metric describing the geometry outside a central body may be written in the following form[2]:

$$ds^2 = \left(1 - \frac{2m}{r} - \frac{\Lambda r^2}{3c^2}\right) dt^2 - \frac{1}{c^2}\left(\frac{dr^2}{1 - 2m/r - \Lambda r^2/3c^2} + r^2\, d\omega^2\right), \tag{5}$$

where $d\omega = d\theta^2 + \sin^2\theta\, d\phi^2$, c is the velocity of light, and $m = GM/c^2$. The coefficients of this metric differ from those of the Minkowski space-time by the two quantities $-2m/r$ and $-\Lambda r^2/3c^2$. If the sum of these two expressions is regarded as a dimensionless potential Φ, we may write

$$\begin{aligned} F_1 = -\frac{\partial \Phi}{\partial r} &= \frac{\partial}{\partial r}\left(\frac{2m}{r}\right) + \frac{\partial}{\partial r}\left(\frac{\Lambda r^2}{3c^2}\right) \\ &= -\frac{2m}{r^2} + \frac{2}{3}\frac{\Lambda r}{c^2} \\ &= \frac{2}{c^2}\left(-\frac{GM}{r^2} + \frac{LV_r}{r^2}\right). \end{aligned} \tag{6}$$

Hence, the geometry of the Schwarzschild (exterior) space-time is determined by the (dimensionless) potential of the Newtonian forces acting on a test particle.

In the case of empty space (no central body), Eq. (4) reduces to

$$F = LV_r/r^2.$$

The corresponding de Sitter's metric may be written as[3]

$$ds^2 = \left(1 - \frac{\Lambda r^2}{3c^2}\right) dt^2 - \frac{1}{c^2}\left(\frac{dr^2}{1 - \Lambda r^2/3c^2} + r^2\, d\omega^2\right). \tag{7}$$

Also in this case the deviation from flat Minkowskian space-time is given by the potential of the force acting on a test particle of unit mass.

3. THE CASE OF A UNIFORM MASS DISTRIBUTION

In analogy with the metrics of Schwarzschild and de Sitter, it may now be possible to determine the metric for the general case of a spherical symmetric mass distribution. From Schlüter's result, Eq. (3),

[1] A. Schlüter, Astrophys. J. **60**, 141 (1955).
[2] R. C. Tolman, *Relativity Thermodynamics and Cosmology* (Oxford University Press, London, 1934), p. 204.
[3] Reference 2, p. 349.

it is found that the potential Φ of the forces is given by

$$\Phi = -\int_0^r F\,dr = \int_0^r \frac{GM_r}{r^2}\,dr - \int_0^r \frac{LV_r}{r^2}\,dr.$$

From the assumption that the density is a function of time alone it follows that

$$\Phi = \frac{GM_r}{2r} - \frac{LV_r}{2r}.$$

Strict analogy with the former cases now requires Φ to be multiplied by a factor of $2/c^2$. Hence, a metric for a uniform mass distribution may be written

$$ds^2 = \left(1 + \frac{GM_r}{c^2 r} - \frac{LV_r}{c^2 r}\right) dt^2 - \frac{1}{c^2}\left(\frac{dr^2}{1 + GM_r/c^2 r - LV_r/c^2 r} + r^2\,d\omega^2\right). \quad (8)$$

This metric, however, may be considered only as an approximation to the general metric of an orthogonal space–time, spherically symmetric about the point $r = 0$. Hence, it has to be proved that in the Newtonian approximation the forces acting on a test particle in the space–time described by (8) are, in fact, given by Eq. (3). It is assumed that L/G differs from ρ by a few orders of magnitude at the most. To write out Einstein's equations, the following form of the metric is used:

$$ds^2 = e^{\nu}\,dt^2 - (1/c^2)(e^{\mu}\,dr^2 + r^2\,d\omega^2). \quad (9)$$

A constant κ is defined by $\kappa \equiv 8\pi G/c^2$. Since this constant is a small quantity, we may write, to the first order of κ,

$$e^{\nu} = e^{-\mu} = 1 + \kappa\psi. \quad (10)$$

In Dingle's notation, we have

$$D = e^{\nu}, \quad A = e^{-\nu}/c^2, \quad B = e^{-\nu}r^2/c^2, \quad C = e^{-\nu}r^2\sin^2\theta/c^2.$$

These expressions are now substituted into Einstein's equations.[4] Inspection of the formulas for T^{12}, T^{13}, T^{23}, T^{24}, and T^{34} shows that all these components of the energy tensor $T^{\alpha\beta}$ are identically zero. The matter is assumed to be a perfect fluid with the energy tensor given by

$$T^{\alpha\beta} = (\rho + p/c^2)u^{\alpha}u^{\beta} - g^{\alpha\beta}p/c^2, \quad (11)$$

where u^{α} is the velocity 4-vector, $g^{\alpha\beta}$ is the contravariant form of the metrical tensor, ρ is the density, and p the pressure.

[4] G. C. McVittie, *General Relativity and Cosmology* (Chapman and Hall Ltd., London, 1965), 2nd ed., p. 71.

After some calculations, it is found that the surviving Einstein equations can be written, to first order in κ, as

$$\kappa\left[\left(\rho + \frac{p}{c^2}\right)(u^4)^2 - \frac{e^{-\nu}p}{c^2}\right]e^{\nu} = -\kappa\left(\frac{\psi}{r^2} + \frac{\psi_r}{r^2}\right) - \frac{\Lambda}{c^2}, \quad (12)$$

$$\kappa\left[\left(\rho + \frac{p}{c^2}\right)(u^1)^2 + \frac{e^{\nu}p}{c^2}\right]e^{-\nu} = \kappa\left(\frac{c^2\psi}{r^2} + \frac{c^2\psi_r}{r}\right) + \Lambda, \quad (13)$$

$$\kappa p = \kappa\left(\psi_{tt} + \tfrac{1}{2}c^2\psi_{rr} + \frac{c^2\psi_r}{r}\right) + \kappa^2\left(\tfrac{1}{2}c^2\psi_r^2 + \tfrac{1}{2}c^2\psi\psi_{rr} + \frac{c^2\psi\psi_r}{r}\right) + \Lambda, \quad (14)$$

$$\kappa(\rho + p/c^2)u^4u^1 = -\kappa\psi_t/r. \quad (15)$$

The subscripts denote partial derivatives with respect to r and t. By inspection, it is found that two new equations arise from Eqs. (13) and (14) which make the Newtonian approximation unnecessary:

$$\psi_r + \psi/r + \Lambda/(8\pi G) = 0, \quad (13')$$

$$\psi_{rr} + 2\psi_r/r + \Lambda/(4\pi G) = 0. \quad (14')$$

These terms are of order c^2 times that of any other term in (12)–(15). The solution that satisfies both these differential equations is

$$\psi = -\frac{C}{r} - \frac{\Lambda r^2}{24\pi G},$$

where C is a constant of integration. Whence, we have

$$\kappa\psi = -\frac{K}{r} - \frac{\Lambda r^2}{3c^2}, \quad (16)$$

where K is a constant. It is thus shown that, of the metrics (9) with the first-order approximation (10), only those with a function ψ given by (16) are compatible with Einstein's equations. It is seen that the Schwarzschild metric (5), $K = 2m$, and the de Sitter metric (7), $K = 0$, follow from this investigation.

This result may be interesting in itself, but shows that the metric (8) cannot be used as a representation of a spherically symmetric space–time with a uniform mass distribution.

4. FIRST APPROXIMATION TO THE ISOTROPIC FORM OF THE METRIC FOR SPHERICAL SYMMETRY

It may still be possible to find a metric for a spherical mass distribution where the deviations from flat Minkowskian space–time are proportional to potentials of the Λ force and the gravitational force acting

per unit mass of the fluid. It is required that the absolute magnitudes of the constants of proportionality are of the same order of magnitude and that in the Newtonian approximation of the field equations the forces acting on a fluid element are the Λ force and the gravitational self-attraction of the matter.

Thus, giving up the *strict* analogy with the Schwarzschild and the de Sitter cases, the isotropic form of the metric for spherical symmetry is used:

$$ds^2 = e^{\nu}\,dt^2 - (e^{\mu}/c^2)(dr^2 + r^2\,d\omega^2). \tag{17}$$

The explicit forms of Einstein's equations to which this metric gives rise are worked out in detail by McVittie.[5]

Before proceeding it is worth noticing what happens when the exact solution, with nonzero cosmical constant for the region outside a central mass M given by Eq. (5), is transformed to the isotropic form. A new radial coordinate $\bar{r}$ is introduced in the metric (5) by

$$r = \left(1 + \frac{m}{2\bar{r}} - \frac{\Lambda\bar{r}^2}{24c^2}\right)^2 \bar{r}.$$

In the first-order approximation, we find

$$dr = \left(1 + \frac{m}{2\bar{r}} - \frac{\Lambda\bar{r}^2}{24c^2}\right)\left(1 - \frac{m}{2\bar{r}} - \frac{5\Lambda\bar{r}^2}{24c^2}\right) d\bar{r}$$

and, hence, the metric becomes

$$ds^2 = \left(1 - \frac{2m}{\bar{r}} - \frac{\Lambda\bar{r}^2}{3c^2}\right) dt^2 - \left(1 + \frac{2m}{\bar{r}} - \frac{\Lambda\bar{r}^2}{6c^2}\right)[d\bar{r}^2 + \bar{r}^2\,d\omega^2]. \tag{5'}$$

It is important to notice that this *does not* have the form

$$ds^2 = (1 - \kappa\psi)\,dt^2 - (1 + \kappa\psi)[dr^2 + r^2\,d\omega^2]. \tag{5''}$$

It is therefore not to be expected that the internal solution which we are seeking reduces in first approximation to the form (5''). Therefore, we begin by modifying the coefficients of the metric by introducing a function Ω. This function is assumed to be of the order c^2. The coefficients of the metric (17) are now given by

$$e^{\nu} = 1 + \kappa(2\Omega/c^2 - \psi), \quad e^{\mu} = 1 + \kappa\psi, \tag{18}$$

where $\psi = (n_1\rho + n_2 L/G)r^2$. Here n_1 and n_2 are constants. It is assumed again that L/G does not differ from ρ by more than a few orders of magnitude. When the function Ω is introduced in this way, the approximation to first order in κ is retained. This procedure is analogous to one used by McVittie.[6]

After the Newtonian approximation (see the appendix), the surviving Einstein equations are reduced to

$$\rho q^2 + p = -\psi_{TT} + \Lambda\psi - \tfrac{1}{2}\Lambda r\psi_r - 2\pi G\psi_r^2, \tag{19}$$

$$p = -\psi_{TT} + \Lambda\psi + \tfrac{1}{2}\Lambda r\psi_r + 2\pi G\psi_r^2 + 2\pi G r^2\left(\frac{\Lambda}{8\pi G}\right)^2, \tag{20}$$

$$\rho = -\nabla^2\psi - \frac{\Lambda}{8\pi G}, \tag{21}$$

$$\rho q = \psi_{rT}. \tag{22}$$

The subscripts r and T indicate partial derivative with respect to the radial distance r and Newtonian absolute time T, respectively. The operator ∇^2 is defined by

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}, \tag{23}$$

and q is the fluid velocity.

In the course of the calculations in the Appendix the function Ω is determined to be

$$\Omega = -\tfrac{1}{4}\Lambda r^2/\kappa, \tag{24}$$

whence

$$2\kappa\Omega/c^2 = -\tfrac{1}{2}\Lambda r^2/c^2.$$

This means that the terms in the coefficients of the metric are still of the same type. It remains to determine the two constants n_1 and n_2. Since the coefficients (18) are only a first-order approximation to the general case of spherical symmetry investigated by Schlüter, it also remains to show that the Newtonian forces acting per unit mass of the fluid are in fact given by Eq. (3).

By elimination of ρ, p, and q from Eqs. (19)–(22), the consistency relation is found to be

$$\psi_{rT}\left(\nabla^2\psi + \frac{\Lambda}{8\pi G}\right)^{-1} = \Lambda r\psi_r + 4\pi G\psi_r^2 + 2\pi G r^2\left(\frac{\Lambda}{8\pi G}\right)^2. \tag{25}$$

Thus, the four equations (19)–(22) are equivalent to (25) together with

$$p = -\psi_{TT} + \Lambda\psi + \tfrac{1}{2}\Lambda r\psi_r + 2\pi G\psi_r^2 + 2\pi G r^2\left(\frac{\Lambda}{8\pi G}\right)^2, \tag{26}$$

$$q = -\psi_{rT}\left(\nabla^2\psi + \frac{\Lambda}{8\pi G}\right)^{-1}, \tag{27}$$

$$\rho = -\nabla^2\psi - \frac{\Lambda}{8\pi G}. \tag{28}$$

[5] Reference 4, p. 74, Eqs. (4.419)–(4.422).

[6] Reference 4, p. 119.

With the help of the equation of continuity, the equation of motion for a perfect fluid in the case of spherical symmetry may be written as[6]

$$\rho F = \frac{\partial}{\partial T}(\rho q) + \frac{\partial}{\partial r}(\rho q^2 + p) + \frac{2}{r}(\rho q^2), \quad (29)$$

where F is the radial force per unit mass of the fluid, apart from the pressure gradient force. When the expressions for p, q, and ρ are substituted into Eq. (29), it is found that

$$F = 4\pi G\psi_r + \tfrac{1}{2}\Lambda r. \quad (30)$$

If the mass inside a surface of radius r is defined as

$$M_r = 4\pi\int_0^r \rho r^2\, dr, \quad (31)$$

then, by means of Eq. (28), it is found that with $n_1 = -\frac{1}{6}$ and $n_2 = -\frac{1}{12}$, the force per unit mass of the fluid is

$$F = -GM_r/r^2 + LV_r/r^2, \quad (32)$$

where M_r and V_r are the mass and the volume inside a sphere of radius r, respectively. Equation (32) is equivalent to Eq. (3). Hence, all the requirements for the metric are fulfilled. The coefficients of the metric (17) are given by

$$e^{\nu} = 1 + \frac{GM_r}{c^2 r} - \frac{LV_r}{c^2 r}, \quad (33)$$

$$e^{\mu} = 1 - \frac{GM_r}{c^2 r} - \frac{LV_r}{2c^2 r}. \quad (34)$$

The metric describes a (bounded or unbounded) spherically symmetric mass distribution where in general the density is a function of both radius and time. It is in accordance with the one describing a spherically symmetric static mass distribution with $\Lambda = 0$.[7]

In the special case where the density is a function of time alone the metric may be used for cosmological investigations. However, it is assumed in the calculations that the deviations from flat Minkowskian space-time are small. Hence, the following inequality must be satisfied:

$$\frac{\kappa}{6}\left(\rho - \frac{L}{G}\right)r^2 \ll 1.$$

The magnitude of L/G can at most be of the order of the mean density of the universe. An upper limit of the radius r is thus given by

$$r^2 \ll 6/(\kappa\rho).$$

If the minimum value of the mean density,[8] $\rho = 3.10^{-31}$ g/cm^3, is inserted, the upper limit of the radius is defined by $r \ll 10^{29}$ cm.

A first-order approximation to the isotropic form of the metric for spherical symmetry has been found where, in analogy with the cases of Schwarzschild and de Sitter, the deviation from flat Minkowskian space-time is given by the potentials of the gravitational force and the Λ force. It is shown that in the Newtonian approximation the forces acting per unit mass of the fluid are the Λ force and the gravitational self-attraction.

ACKNOWLEDGMENT

I wish to thank Professor G. C. McVittie for valuable suggestions.

APPENDIX

Inspection of Einstein's equations for the metric (17) shows that the components T^{12}, T^{13}, T^{23}, T^{24}, and T^{34} of the energy-momentum tensor $T^{\alpha\beta}$ are all identically zero. This implies that the two components u^2 and u^3 of the velocity 4-vector $u^{\alpha} = dx^{\alpha}/ds$ are zero and, hence, the 4-vector satisfies the following equation[5]:

$$1 = e^{\nu}(u^4)^2 - e^{-\nu}(u^1)^2/c^2. \quad (A1)$$

The surviving Einstein equations are worked out in detail by McVittie.[5] When the expressions for μ and ν given by (18) are substituted on the right-hand sides, the following equations are obtained to first order in κ:

$$\kappa\left[\left(\rho + \frac{p}{c^2}\right)(u^1)^2 + e^{-\mu}p\right]e^{\mu} = \kappa\left[\frac{2\Omega_r}{r} - \psi_{tt} + 2\pi G\psi_r^2 + 4\pi G\psi_r\left(\frac{2\Omega}{c^2} - \psi\right)_r\right] + \Lambda, \quad (A2)$$

$$\kappa p = \kappa\left[\Omega_{rr} + \frac{\Omega_r}{r} - \psi_{tt} + 2\pi G\left(\frac{2\Omega_r}{c^2} - \psi_r\right)^2\right] + \Lambda, \quad (A3)$$

$$\kappa\left[\left(\rho + \frac{p}{c^2}\right)(u^4)^2 - e^{-\nu}\frac{p}{c^2}\right]e^{\nu} = -\kappa\left(\psi_{rr} + \frac{2\psi_r}{r}\right) - \frac{\Lambda}{c^2}, \quad (A4)$$

$$\kappa\left(\rho + \frac{p}{c^2}\right)u^1u^4 = \kappa\psi_{rt}. \quad (A5)$$

[7] Reference 4, p. 67.

[8] J. H. Oort, Solvay Conference on *La Structure et l'évolution de l'univers*, R. Stoops, Ed. (Brussels, Belgium, 1958), p. 180.

It is seen that the terms $2\kappa\Omega_r/r + \Lambda$ in (A2) and $\kappa(\Omega_{rr} + \Omega_r/r) + \Lambda$ in (A3) are of magnitude c^2 times that of any other term in the equations. For a Newtonian approximation to be possible, these terms must vanish identically. Hence, it follows that

$$2\kappa\Omega_r/r + \Lambda = 0,$$
$$\kappa(\Omega_{rr} + \Omega_r/r) + \Lambda = 0.$$

Both equations are satisfied by the solution

$$\Omega = -\tfrac{1}{4}\Lambda r^2/\kappa, \tag{A6}$$

which determines the function Ω to be of a form in agreement with the requirements for the metric.

A Newtonian approximation is now carried out in which c is identified with an infinite velocity. The coordinate time t then becomes the Newtonian absolute time T and the coordinate r becomes the radial distance from the origin in absolute Euclidean space. The fluid velocity q is now defined by $q = u^1/u^4$.

The approximation of (A1) gives $u^4 = 1$. It is assumed that velocities are small, $q^2 \ll c^2$. The Newtonian forms of (A2) to (A5) then become

$$\rho q^2 + p = -\psi_{TT} + \Lambda\psi - \tfrac{1}{2}\Lambda r\psi_r - 2\pi G\psi_r^2, \tag{A7}$$

$$p = -\psi_{TT} + \Lambda\psi + \tfrac{1}{2}\Lambda r\psi_r + 2\pi G\psi_r^2 + 2\pi G r^2\left(\frac{\Lambda}{8\pi G}\right)^2, \tag{A8}$$

$$\rho = -\psi_{rr} - \frac{2\psi_r}{r} - \frac{\Lambda}{8\pi G}, \tag{A9}$$

$$\rho q = \psi_{rT}. \tag{A10}$$

The constants n_1 and n_2 can also be determined by solving Eq. (A9). The solution is found to be

$$\psi = -\tfrac{1}{6}\rho r^2 - \frac{1}{12}\frac{L}{G}r^2,$$

in accordance with the previous results.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Algebraic Theory of Ray Representations of Finite Groups*

WILLIAM G. HARTER
Department of Physics, University of California, Irvine, California

(Received 5 March 1968)

A theory of characters of ray representations of finite groups, that does not use any reference to a covering group, is derived by defining two generalizations of the concept of a group's class. Orthogonality relations are obtained over one of these generalized classes. This theory is used to discuss subduction and induction of ray representations while the Frobenius reciprocity theorem and generalizations thereof are proved. The theory provides a more efficient method of deriving and treating ray representations of finite groups for a given factor system than has previously been made available.

1. INTRODUCTION

Since ray representations of point groups have become useful in the theory of nonsymmorphic space groups,[1] methods of obtaining them have been given by Döring.[2] Tables of them for particular factor systems have been tabulated for the 32 point groups. The theory of double groups is really a special case of ray representation theory and tables of these have long been available for the point groups.[3] Recently this theory was extended to space groups.[4] Most of these treatments build on those given by Schur in his three papers.[5]

In all these theories, one deals with a so-called covering group whose order is, in general, some multiple of the order of the group which is being ray-represented. (Usually it is twice as large.) The following theory will show that this is unnecessarily laborious.

* Research supported by Number 9-443984-21112-3, NSF-GP No. 5844.

[1] G. Lyabarski, *The Application of Group Theory in Physics* (Pergamon Press, Inc., New York, 1960), p. 95.

[2] Z. Döring, Z. Naturforsch. **14**, 343 (1959).

[3] W. Opechowski, Physica **7**, 552 (1940).

[4] M. Glück, Y. Gur, and J. Zak, J. Math. Phys. **8**, 787 (1967).

[5] I. Schur, J. Reine Angew. Math. **127**, 20 (1904); **132**, 85 (1907); **139**, 155 (1911).

It is seen that the terms $2\kappa\Omega_r/r + \Lambda$ in (A2) and $\kappa(\Omega_{rr} + \Omega_r/r) + \Lambda$ in (A3) are of magnitude c^2 times that of any other term in the equations. For a Newtonian approximation to be possible, these terms must vanish identically. Hence, it follows that

$$2\kappa\Omega_r/r + \Lambda = 0,$$
$$\kappa(\Omega_{rr} + \Omega_r/r) + \Lambda = 0.$$

Both equations are satisfied by the solution

$$\Omega = -\tfrac{1}{4}\Lambda r^2/\kappa, \tag{A6}$$

which determines the function Ω to be of a form in agreement with the requirements for the metric.

A Newtonian approximation is now carried out in which c is identified with an infinite velocity. The coordinate time t then becomes the Newtonian absolute time T and the coordinate r becomes the radial distance from the origin in absolute Euclidean space. The fluid velocity q is now defined by $q = u^1/u^4$.

The approximation of (A1) gives $u^4 = 1$. It is assumed that velocities are small, $q^2 \ll c^2$. The Newtonian forms of (A2) to (A5) then become

$$\rho q^2 + p = -\psi_{TT} + \Lambda\psi - \tfrac{1}{2}\Lambda r\psi_r - 2\pi G\psi_r^2, \tag{A7}$$

$$p = -\psi_{TT} + \Lambda\psi + \tfrac{1}{2}\Lambda r\psi_r + 2\pi G\psi_r^2 + 2\pi G r^2\left(\frac{\Lambda}{8\pi G}\right)^2, \tag{A8}$$

$$\rho = -\psi_{rr} - \frac{2\psi_r}{r} - \frac{\Lambda}{8\pi G}, \tag{A9}$$

$$\rho q = \psi_{rT}. \tag{A10}$$

The constants n_1 and n_2 can also be determined by solving Eq. (A9). The solution is found to be

$$\psi = -\tfrac{1}{6}\rho r^2 - \frac{1}{12}\frac{L}{G}r^2,$$

in accordance with the previous results.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Algebraic Theory of Ray Representations of Finite Groups*

WILLIAM G. HARTER
Department of Physics, University of California, Irvine, California

(Received 5 March 1968)

A theory of characters of ray representations of finite groups, that does not use any reference to a covering group, is derived by defining two generalizations of the concept of a group's class. Orthogonality relations are obtained over one of these generalized classes. This theory is used to discuss subduction and induction of ray representations while the Frobenius reciprocity theorem and generalizations thereof are proved. The theory provides a more efficient method of deriving and treating ray representations of finite groups for a given factor system than has previously been made available.

1. INTRODUCTION

Since ray representations of point groups have become useful in the theory of nonsymmorphic space groups,[1] methods of obtaining them have been given by Döring.[2] Tables of them for particular factor systems have been tabulated for the 32 point groups. The theory of double groups is really a special case of ray representation theory and tables of these have long been available for the point groups.[3] Recently this theory was extended to space groups.[4] Most of these treatments build on those given by Schur in his three papers.[5]

In all these theories, one deals with a so-called covering group whose order is, in general, some multiple of the order of the group which is being ray-represented. (Usually it is twice as large.) The following theory will show that this is unnecessarily laborious.

* Research supported by Number 9-443984-21112-3, NSF-GP No. 5844.

[1] G. Lyabarski, *The Application of Group Theory in Physics* (Pergamon Press, Inc., New York, 1960), p. 95.

[2] Z. Döring, Z. Naturforsch. **14**, 343 (1959).

[3] W. Opechowski, Physica **7**, 552 (1940).

[4] M. Glück, Y. Gur, and J. Zak, J. Math. Phys. **8**, 787 (1967).

[5] I. Schur, J. Reine Angew. Math. **127**, 20 (1904); **132**, 85 (1907); **139**, 155 (1911).

It will be shown how one may quickly construct and use a concise square character table and a set of representations straightaway from any given factor system. In short, the following theory will make it possible for one to use already tabulated results more easily and quickly to obtain irreducible ray representations for other finite groups or factor systems which might be important in other areas of interest.[6] But what is most important, an understanding of the possible structures of these mathematical objects is obtained systematically; this in turn sheds more light on the structure of representations of groups themselves.

A systematic treatment of ray representations of groups was begun by Rudra in a series of three papers.[7] Unfortunately, many of his results are not valid for factor systems that are not equivalent to the trivial one. In order to include the nontrivial cases, it will be necessary to begin the following treatment by starting with the basic definitions of a ray representation.

2. BASIC DEFINITIONS

If one nonsingular n by n matrix $\mathcal{R}(R)$ is assigned to each element R of a group $\mathcal{G} = \{\cdots R, S, T, \cdots\}$ such that

$$\mathcal{R}(R)\mathcal{R}(S) = \omega_{R,S}\mathcal{R}(RS)$$

(where $\omega_{R,S}$ = complex number) for all R and S in $\mathcal{G}$, then that set $\{, \mathcal{R}(R), \mathcal{R}(S), \mathcal{R}(T), \cdot\}$ of matrices is called an nth *degree ray representation* of $\mathcal{G}$, and the constants ω_{RS} make up what is called a *factor system* $\{\omega\}$ over $\mathcal{G}$.

Now if all the matrices of this ray representation are multiplied by different constants, the resulting set of matrices $\{\mathcal{R}'(R) \equiv C_R\mathcal{R}(R),\ \mathcal{R}'(S) \equiv C_S\mathcal{R}(S), \cdots\}$ is also a ray representation but with a new factor system $\{\omega'\}$ as is shown below:

$$\begin{aligned}\mathcal{R}'(R)\mathcal{R}'(S) &= C_RC_S\mathcal{R}(R)\mathcal{R}(S)\\ &= (C_RC_S/C_{RS})\omega_{R,S}\mathcal{R}'(RS)\\ &\equiv \omega'_{R,S}\mathcal{R}'(RS),\end{aligned} \tag{1}$$

where $\omega'_{R,S} = (C_RC_S/C_{RS})\omega_{R,S}$. If two factor systems $\{\omega\}$ and $\{\omega'\}$ are related in this manner, they are said to be *projective equivalent* or *p-equivalent*[8] or in the same *class*.[9]

If the $\mathcal{R}$ are $n \times n$ matrices, one may let C_R equal any nth root of $\det \mathcal{R}(R)$ for all R in $\mathcal{G}$ and obtain the following:

$$\det[\mathcal{R}'(R)\mathcal{R}'(S)] = \det[\omega'_{R,S}\mathcal{R}'(RS)],$$

$$C_R^n[\det \mathcal{R}(R)]C_S^n[\det \mathcal{R}(S)] = (\omega'_{R,S})^nC_{RS}^n[\det \mathcal{R}'(RS)],$$

$$1 = (\omega'_{R,S})^n. \tag{2}$$

Hence every factor system associated with an nth degree ray representation is p-equivalent to one in which all the factors are nth roots of unity; i.e., $(\omega'_{R,S})^n = 1$ and $\omega^*_{R,S}\omega_{R,S} = 1$ for all R and S in $\mathcal{G}$. If $n = 1$, then $\omega'_{R,S} = 1$ and $\{\omega\}$ is p-equivalent to the trivial factor system, which is sometimes called the *vector* factor system.

3. ASSOCIATIVITY AND THE REGULAR REPRESENTATION

It is helpful to imagine an abstract set of g elements $\{a_R, a_S, a_T\}$ (one for each element R, S, or T of a group $\mathcal{G}$) that obey the relations

$$a_Ra_S = \omega_{R,S}a_{RS}, \tag{3}$$

where the $\omega_{R,S}$ belong to a given factor system $\{\omega\}$. Such a collection forms what is defined as a ring or associative algebra. (But it is not a group unless $\omega_{R,S} \equiv 1$.) I call the set of all linear combinations of the elements a_R (using complex coefficient) the *ray algebra* $\mathcal{A}(\mathcal{G}, \omega)$. The elements a_R are here called a system of *base elements*, since any element in $\mathcal{A}(\mathcal{G}, \omega)$ can be written as a unique linear combination of them. Obviously there are other bases for which the same is true.

This definition is motivated by more general theories of rings and algebras. There the most general algebra $\mathcal{A}$ with n-base elements $\{a_1, a_2, a_3, \cdots, a_n\}$ has the following multiplicative structure:

$$a_ia_j = \sum_{k=1}^{k=n} c_{ij}^k a_k. \tag{4}$$

Here the complex numbers c_{ij}^k are called structure constants and can be thought of as components of $n \times n$ matrices that will represent the algebra $\mathcal{A}$. For if one lets $c_{ij}^k = \mathcal{R}_{ik}(a_j)$, then the relations

$$\sum_{B=1}^{B=n} \mathcal{R}_{\alpha\beta}(a_i)\mathcal{R}_{\beta\gamma}(a_j) = \sum_{k=1}^{k=n} c_{ij}^k\mathcal{R}_{\alpha\gamma}(a_k) \tag{5}$$

follow directly from an expansion of the associativity relations $(a_ia_j)a_k = a_i(a_ja_n)$ which one demands of any ring or any set of matrices.

[6] A. O. Barut, J. Math. Phys. 7, 1908 (1966).
[7] P. Rudra, J. Math. Phys. 6, 1273, 1278 (1965); 7, 935 (1966).
[8] J. S. Lomont, *Applications of Finite Groups* (Academic Press, London, 1959), p. 729.
[9] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1962), p. 462.

These n n-by-n matrices $\mathcal{R}(a_j)$ form what is called the *regular representation* of $\mathcal{A}$ in basis $\{a_j\}$. Now the regular representation of $\mathcal{A}(\mathcal{G}, \omega)$ in the basis $\{\cdots a_R, a_S \cdots\}$ will be made of g-by-g matrices that have a very simple form. In particular, $\mathcal{R}_{RT}(a_S) = \omega_{R,S}$ if and only if $RS = T$. Otherwise $\mathcal{R}_{RT}(a_S) = 0$.

An easy way to construct these matrices is first to arrange a multiplication table of the base elements with the elements a_R written in some order across the top, and the "inverses" $a_{R^{-1}}$ appearing in the same order down the side:

$$\begin{array}{c|ccccc} & \cdots & a_R & \cdots & a_T & \cdots \\ \hline \vdots & & \vdots & & \vdots & \\ a_{R^{-1}} & \cdots & \omega_{R^{-1},R}a_1 & \cdots & \omega_{R^{-1},T}a_{R^{-1}T} & \cdots \\ \vdots & & \vdots & & \vdots & \\ a_{T^{-1}} & \cdots & \omega_{T^{-1},R}a_{T^{-1}R} & \cdots & \omega_{T^{-1},T}a_1 & \cdots \\ \vdots & & \vdots & & \vdots & \end{array} \tag{6}$$

Then, since $\mathcal{R}_{RT}(a_S) = \omega_{R,S}$ if and only if $R^{-1}T = S$, you obtain matrices of the form

$$\mathcal{R}(a_S) = \begin{vmatrix} \cdots\cdots\cdots\ \omega_{1,S}\ \cdots\cdots\cdots \\ \cdots\ \omega_{2,S}\ \cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\ \omega_{3,S}\ \cdots \\ \omega_{4,S}\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots \\ \cdots\cdots \end{vmatrix} \tag{7}$$

by simply inspecting the table. (Note that there is one and only one entry in each column and row.) Incidentally, the associativity of multiplication implies that

$$(a_R\, a_S)a_T = a_R(a_S a_T),$$

$$\omega_{R,S}\omega_{RS,T}a_{RST} = \omega_{S,T}\omega_{R,ST}a_{\omega ST},$$

or

$$\omega_{R,S} = \frac{\omega_{R,ST}\omega_{S,T}}{\omega_{RS,T}} \tag{8}$$

for all R, S, and T in $\mathcal{G}$.

The regular representation of the unit element 1 obtained by the preceding construction is

$$\mathcal{R}(a_1) = \begin{vmatrix} \omega_{11} & 0 & 0 & \cdots \\ 0 & \omega_{21} & 0 & \cdots \\ 0 & 0 & \omega_{31} & \cdots \end{vmatrix}. \tag{9}$$

But the associativity relations demand that

$$\omega_{R,1} = \frac{\omega_{R,1}\omega_{1,1}}{\omega_{R1}} = \omega_{1,1} \tag{10}$$

for all R in $\mathcal{G}$. (Similarly $\omega_{1,S} = \omega_{1,1}$.) Clearly no loss of generality will occur if we assume from now on that

$$\omega_{R,1} = \omega_{1,1} = \omega_{1,S} = 1, \tag{11}$$

so that $\mathcal{R}(a_1)$ is a unit matrix.

Now the trace of $\mathcal{R}(a_1)$ is g, the order of group $\mathcal{G}$. From this one sees that $\omega_{R^{-1},R} = \omega_{R,R^{-1}}$, since

$$\operatorname{Tr} \mathcal{R}(a_R)\mathcal{R}(a_{R^{-1}}) = \operatorname{Tr} \mathcal{R}(a_{R^{-1}})\mathcal{R}(a_R) \tag{12}$$

implies that

$$\omega_{R,R^{-1}} \operatorname{Tr} \mathcal{R}(a_1) = \omega_{R^{-1},R} \operatorname{Tr} \mathcal{R}(a_1). \tag{13}$$

Note, however, that the trace of $\mathcal{R}(a_S)$ (for $S \neq 1$) is zero.

Finally, it is possible to show that each matrix $\mathcal{R}(a_S)$ is unitary. Since $\mathcal{R}_{RT}(a_S) = \omega_{R,S}$ (assuming now that $RS = T$), one has that $\mathcal{R}_{TR}(a_{S^{-1}}) = \omega_{T,S^{-1}}$. Now the inverse of matrix $\mathcal{R}(a_S)$ is

$$[\mathcal{R}(a_S)]^{-1} = (\omega_{S^{-1},S})^{-1}\mathcal{R}(a_{S^{-1}}),$$

while the Hermitian conjugate of the matrix $\mathcal{R}(a_S)$ is $\mathcal{R}^{+}(a_S)$, where

$$\begin{aligned} \mathcal{R}^{+}_{TR}(a_S) = \omega^{*}_{R,S} = \frac{1}{\omega_{R,S}} &= \omega_{RS,S^{-1}}/(\omega_{R,S}\omega_{RS,S^{-1}}) \\ &= \omega_{RS,S^{-1}}/(\omega_{R,SS^{-1}}\omega_{S,S^{-1}}) \\ &= \omega_{R,S^{-1}}/\omega_{S,S^{-1}} \\ &= \mathcal{R}_{RT}(a_{S^{-1}})/\omega_{S,S^{-1}} \\ &= [\mathcal{R}_{TR}(a_S)]^{-1}. \end{aligned} \tag{14}$$

The associativity relations (8) have been used along with the conventions (11) $\omega_{R,1} = 1$ and (2)

$$\omega_{R,S^{*}}\omega_{R,S} = 1.$$

4. NILPOTENTS, IDEMPOTENTS, AND OTHER ALGEBRAIC CONCEPTS

In this section it is necessary to make some statements without proof, since such proofs are still quite lengthy. The references (10, 11, 12) which contain these proofs are concerned with a type of algebra or ring, of which the ray algebras are a special case.

[10] H. V. McIntosh, "Abelian Groups with Operators," RIAS Technical Report 57-2, 1958.

[11] C. L. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Interscience Publishers, Inc., New York, 1962).

[12] W. G. Harter, Doctoral dissertation, University of California at Irvine, 1967.

It is convenient to define a *contagiously nilpotent element* η of algebra $\mathcal{A} = \{\cdots a, b, c, \cdots\}$ to be a nonzero element that transforms every element a of $\mathcal{A}$ into a nilpotent $n = \eta a$ or into zero $\eta a = 0$. (A nilpotent n is a nonzero element that goes to zero if raised to some power, i.e., $n^k = nnn\cdots n = 0$, for some $k = 2, 3, 4, \cdots$.) If it can be shown that contagiously nilpotent elements do not exist in a given $\mathcal{A}$, then[9–11] any element a of algebra $\mathcal{A}$ can be written as a linear combination of a certain convenient set of base elements

$$\{P_{11}^{(\alpha)}P_{12}^{(\alpha)}\cdots P_{1l(\alpha)}^{(\alpha)}P_{21}^{(\alpha)}P_{22}^{(\alpha)}\cdots P_{l(\alpha)l(\alpha)}^{(\alpha)}P_{11}^{(\beta)}P_{12}^{(\beta)}\cdots\},$$

which will be called *unit dyads.*

$$a = \sum_{(\alpha)}\sum_{L}\sum_{m}\mathfrak{D}_{Lm}^{(\alpha)}(a)P_{Lm}^{(\alpha)}. \tag{15}$$

These unit dyads have a simple multiplicative structure something like their naming would imply:

$$P_{ij}^{(\alpha)}P_{nL}^{(\beta)} = \begin{matrix} 0 & \text{if } (\alpha) \neq (\beta) \quad \text{or} \quad j \neq k, \\ P_{iL}^{(\alpha)} & \text{if } \alpha = \beta \quad \text{and} \quad j = k. \end{matrix} \tag{16}$$

Note that $P_{ij}^{(\alpha)}$ is idempotent if $i = k$, i.e., $P_{jj}^{(\alpha)}P_{jj}^{(\alpha)} = P_{jj}^{(\alpha)}$. By construction[10–12] these idempotents all share the following property. There is an a in $\mathcal{A}$ such that $P_{ii}^{(\alpha)}aP_{jj}^{(\alpha)}$ is nonzero and

$$P_{ii}^{(\alpha)}bP_{jj}^{(\alpha)} = \mu P_{ii}^{(\alpha)}aP_{jj}^{(\alpha)} = \nu P_{ij}^{(\alpha)} \tag{17}$$

for any b in $\mathcal{A}$, where μ and ν are constants. The meaning and application of these elements is further explored shortly.

First it must be shown that this expansion (15) is valid for ray algebras, by showing that there are no contagiously nilpotent elements in $\mathcal{A}(\mathfrak{G}, \omega)$. This we do by showing that no nonzero element

$$\eta = \sum_{R=1}^{R=g}\eta_R a_R \quad \text{of} \quad \mathcal{A}(\mathfrak{G}, \omega)$$

makes *every* element a of $\mathcal{A}(\mathfrak{G}, \omega)$ into a nilpotent $\eta a = n$ or into zero $\eta a = 0$. If

$$a = \sum_{S}\frac{\eta_S^* a_{S^{-1}}}{\omega_{S^{-1},S}},$$

then one has the following:

$$\eta a = \sum_{R,S}\eta_R\eta_S^* a_R\frac{a_{S^{-1}}}{\omega_{S^{-1},S}}.$$

The regular representation of this is

$$\mathcal{R}(\eta a) = \sum_{R,S}\eta_R\eta_S^*\mathcal{R}(a_R)\mathcal{R}^+(a_S),$$

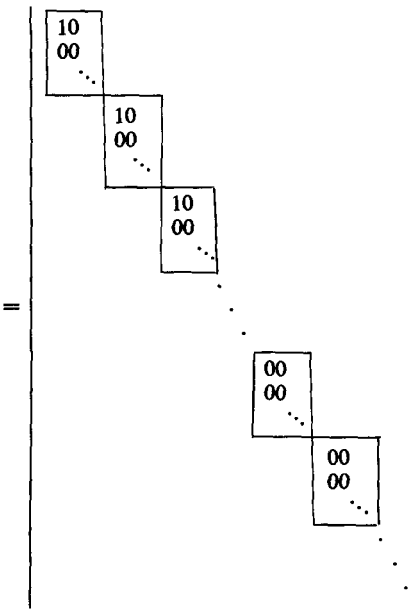
which is a Hermitian matrix since the $\mathcal{R}(a_R)$ are unitary. Also one sees that

$$\operatorname{Tr}\mathcal{R}(\eta a) = \sum_{R,S}\eta_R\eta_S^*\operatorname{Tr}\frac{\mathcal{R}(a_R a_{S^{-1}})}{\omega_{S^{-1}S}} = \sum_{R}|\eta_R|^2 g,$$

which shows $\mathcal{R}(\eta a)$ is a diagonalizable matrix with a nonzero trace. Clearly ηa is not nilpotent, and the basis of unit dyads is valid.

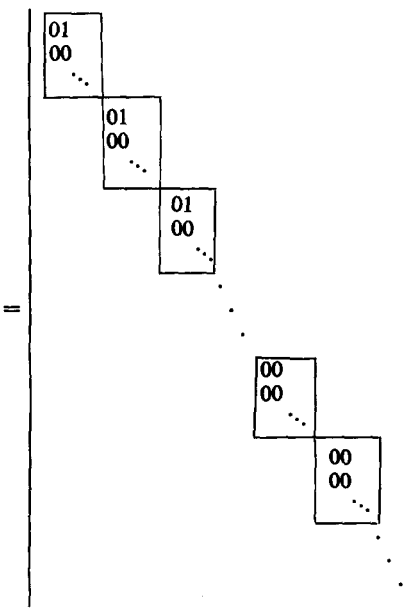
Now imagine the construction of the regular representation $\mathcal{R}^P$ of the unit dyads *using the unit dyads as a basis* and relations (16). For example, one has the following:

$\mathcal{R}^P(P_{11}^{(\alpha)}) =$



(18)

$\mathcal{R}^P(P_{12}^{(\alpha)}) =$



(19)

and

$$\mathcal{R}^P(P_{11}^{(\beta)}) = \begin{vmatrix} \boxed{\begin{matrix}00\\00\end{matrix}}\ddots & & & & \\ & \boxed{\begin{matrix}00\\00\end{matrix}}\ddots & & & \\ & & \boxed{\begin{matrix}00\\00\end{matrix}}\ddots & & \\ & & & \ddots & \\ & & & & \boxed{\begin{matrix}10\\00\end{matrix}}\ddots & \\ & & & & & \boxed{\begin{matrix}10\\00\end{matrix}}\ddots \\ & & & & & & \ddots \end{vmatrix} \quad (20)$$

This regular representation can represent any element a of $\mathcal{A}(\mathfrak{G}, \omega)$ according to (15). It is equivalent to the regular representation (7) over the basis $\{\cdot\, a_R \cdot\}$ constructed in Sec. 3, but it has the following form:

$$\mathcal{R}^P(a) = \begin{vmatrix} \boxed{\mathfrak{D}^{(\alpha)}(a)} & & & & \\ & \boxed{\mathfrak{D}^{(\alpha)}(a)} & & & \\ & & \boxed{\mathfrak{D}^{(\alpha)}(a)} & & \\ & & & \ddots & \\ & & & & \boxed{\mathfrak{D}^{(\beta)}(a)} & \\ & & & & & \boxed{\mathfrak{D}^{(\beta)}(a)} \\ & & & & & & \ddots \end{vmatrix} \quad (21)$$

Here the $\mathfrak{D}^{(\alpha)}(a)$ are matrices of the coefficients $\mathfrak{D}^{(\alpha)}_{Lm}(a)$ in (15). These $\mathfrak{D}^{(\alpha)}$ are in fact the matrices of the irreducible representations of $\mathcal{A}(\mathfrak{G}, \omega)$. Furthermore, the frequency of $\mathfrak{D}^{(\alpha)}$ on the diagonal of $\mathcal{R}^P$ is equal to the dimension $l^{(\alpha)}$. Also it is true that

$$g = l^{(\alpha)^2} + l^{(\beta)^2} + \cdots, \quad (22)$$

which will be called the *diophantine solution* of $\mathcal{A}(\mathfrak{G}, \omega)$.

By definition the unit dyads are some linear combination of the old base elements $\{a_1 a_2 \cdots a_R \cdots a_g\}$:

$$P_{ij}^{(\beta)} = \sum_T P_{ij}^{(\beta)}(a_T) a_T. \quad (23)$$

We solve for the constants $P_{ij}^{(\beta)}(a_T)$. We find that

$$P_{ij}^{(\beta)}(a_T) = (g\omega_{T^{-1},T})\ \mathrm{Tr}\ \mathcal{R}(P_{ij}^{(\beta)} a_{T^{-1}}),$$

since Tr $\mathcal{R}(a_T)$ is zero if T is not 1, but is g if $T = 1$. Next, we obtain the result that

$$P_{ij}^{(\beta)}(a_T) = (g\omega_{T^{-1},T})^{-1}\ \mathrm{Tr}\ \mathcal{R}\left(P_{ij}^{(\beta)}\left\{\sum_k \sum_m \sum_n \mathfrak{D}^{(\alpha)}_{mn}(a_{T^{-1}}) P^{(\alpha)}_{mn}\right\}\right) = \frac{l^{(\beta)}}{g\omega_{T^{-1},T}} \mathfrak{D}^{(\beta)}_{ji}(a_{T^{-1}}),$$

which gives finally

$$P_{ij}^{(\beta)} = \frac{l^{(\beta)}}{g} \sum_T \frac{\mathfrak{D}^{(\beta)}_{ji}(a_{T^{-1}})}{\omega_{T^{-1},T}} a_T. \quad (24)$$

These $\mathfrak{D}^{(\alpha)}$, like the regular representation, can be chosen unitary,[13] where

$$[\mathfrak{D}^{(\beta)}(a_{T^{-1}})/\omega_{T^{-1},T}] = [\mathfrak{D}^{(\beta)}(a_T)]^+. \quad (25)$$

We shall assume that the $\mathfrak{D}^{(\alpha)}$ are unitary so that

$$P_{ij}^{(\beta)} = \frac{l^{(\beta)}}{g} \sum_L \mathfrak{D}^{(\beta)*}_{ij}(a_T) a_T. \quad (26)$$

It is now a simple matter to derive representation orthogonality relations and the Schur's Lemmas.

Theorem 1[14]*:*

$$\frac{l^{(\gamma)}}{g} \sum_T \frac{1}{\omega_{T^{-1},T}} \mathfrak{D}^{(\gamma)}_{mn}(a_{T^{-1}}) \mathfrak{D}^{(\beta)}_{Lj}(a_T) = \delta^{\gamma\beta}\delta_{Ln}\delta_{jm}.$$

Proof: By inspecting (18), (19), and (20), one may write

$$\mathfrak{D}^{(\beta)}_{Lj}[P^{(\gamma)}_{nm}] = \delta^{(\gamma)\beta}\delta_{Lm}\delta_{jm}. \quad (27)$$

Now expanding $P^{(\gamma)}_{nm}$ according to (24) yields the desired relation.

Theorem 2[14]*:* The only (square) $l^{(\alpha)} \times l^{(\alpha)}$ matrix m that commutes with all the matrices of an irreducible ray representation $\{\cdots \mathfrak{D}^{(\alpha)}(a_T) \cdots\}$ of group $\mathfrak{G}$ is a multiple of the unit matrix.

Proof: Let m be an $l^{(\alpha)} \times l^{(\alpha)}$ matrix that satisfies $m\mathfrak{D}^{(\alpha)}(a_T) = \mathfrak{D}^{(\alpha)}(a_T)m$ for all T. This implies that $m\mathfrak{D}^{(\alpha)}[P_{ij}^{(\alpha)}] = \mathfrak{D}^{(\alpha)}[P_{ij}^{(\alpha)}]m$ for all i and j. From (27) one sees that m is a multiple of the unit matrix.

Theorem 3[14]*:* The only (in general rectangular) $l^{(\alpha)} \times l^{(\beta)}$ matrix m that satisfies the relations

$$m\mathfrak{D}^{(\alpha)}(a_T) = \mathfrak{D}^{(\beta)}(a_T)m$$

for all a_T, when $\alpha \neq \beta$, is the zero matrix.

[13] P. Rudra, J. Math. Phys. 6, 1273 (1965), p. 1274.
[14] Reference 13, p. 1275.

Proof: Define the elements

$$P^{(\alpha)} = \sum_{j=1}^{j=l^{(\alpha)}} P_{jj}^{(\alpha)} \tag{28}$$

and note that $\mathfrak{D}^{(\alpha)}[P^{(\alpha)}]$ is the unit matrix, while $\mathfrak{D}^{(\beta)}[P^{(\alpha)}]$ is the zero matrix if $\alpha \neq \beta$ using (27). Clearly

$$m\mathfrak{D}^{(\alpha)}[P^{(\alpha)}] = \mathfrak{D}^{(\beta)}[P^{(\alpha)}]m$$

or

$$m = 0.$$

5. THE CENTER OF A RAY ALGEBRA

The *center* C of $\mathcal{A}(\mathfrak{G}, \omega)$ is the maximal set of elements in $\mathcal{A}(\mathfrak{G}, \omega)$ that commute with every element of $\mathcal{A}(\mathfrak{G}, \omega)$. C is a subalgebra of $\mathcal{A}(\mathfrak{G}, \omega)$, and it would contain the idempotents $P^{(\alpha)}$ defined in (28). We now set out to determine how many $P^{(\alpha)}$ there may be for a general $\mathcal{A}(g, \omega)$ and learn how to obtain them for a particular ray algebra.

Suppose that $c = \sum_R \gamma_R a_R$ is in C. Then

$$ca_S = a_S c \tag{29}$$

for all a_S in $\mathcal{A}(\mathfrak{G}, \omega)$. Also one has that

$$(a_{S^{-1}} c a_S)/\omega_{S^{-1},S} = c. \tag{30}$$

From this one sees that

$$\sum_{S=1}^{S=g} \frac{a_{S^{-1}} c a_S}{\omega_{S^{-1},S}} = \sum_S c = gc, \tag{31}$$

where g is the order of group $\mathfrak{G}$. One is assured that any element in the center must have the form

$$c = \frac{1}{g} \sum_{2=1}^{s=1} a_{S^{-1}} \left[\left(\sum_R \gamma_R a_R \right) \Big/ \omega_{S^{-1},S} \right] a_S$$

$$= \sum_R \frac{\gamma_R}{g} \left\{ \sum_S \frac{a_{S^{-1}} a_R a_S}{\omega_{S^{-1},S}} \right\}.$$

So every element of the center must be some linear combination of some of the g elements c_R defined by

$$c_R = \sum_{s=1}^{s=g} \frac{a_{S^{-1}} a_R S_S}{\omega_{S^{-1}S}} = \sum_{s=1}^{s=g} \frac{\omega_{S^{-1},R}\omega_{S^{-1}R,S}}{\omega_{S^{-1}S}} a_{S^{-1}RS}. \tag{32}$$

We must now find which c_R, if any, are linearly dependent. For the number of linearly independent c_R is exactly equal to the number of inequivalent IR of $\mathcal{A}(\mathfrak{G}, \omega)$.

Now c_R is the linear combination of $a_R, a_{R'}, \cdots$, where $R, R', \cdots$, are in the class of R in $\mathfrak{G}$. Suppose

$$c_R = (\alpha a_R + \beta a_{R'} \cdots), \tag{33}$$

where all terms have been collected. If in group $\mathfrak{G}$ it happens that an element T transforms R into $R'(T^{-1}RT = R')$ so that

$$a_{T^{-1}} a_R a_T = \omega_{T^{-1},RT}\omega_{R,T} a_{R'}, \tag{34}$$

then

$$a_{T^{-1}} c_R a_T/\omega_{T^{-1},T}$$

$$= (\omega_{T^{-1},T})^{-1}(\alpha a_{T^{-1}} a_R a_T + \cdots)$$

$$= (\omega_{T^{-1},T})^{-1}(\alpha\omega_{T^{-1},RT}\omega_{R,T} a_{R'} + \cdots), \tag{35}$$

where $(a_{T^{-1}} c_R a_T)/\omega_{T^{-1}T} = c_R$ by (30). Equating the coefficients of the first terms of (33) and (35) (which is correct since $a_{R'}$ appears only once in either series), one has

$$\beta = \alpha[(\omega_{T^{-1},RT}\omega_{R,T})/\omega_{T^{-1},T}].$$

It will be conventional to let $\alpha = \nu_R$ and

$$\epsilon_{R'} \equiv \beta/\alpha = (\omega_{T^{-1},RT}\omega_{R,T})/\omega_{T^{-1},T} \tag{36}$$

so that

$$c_R = \nu_R(a_R + \epsilon_{R'} a_{R'} + \epsilon_{R''} a_{R''} + \cdots), \tag{37}$$

where the $\epsilon_{R''}$ is a root of unity,

$$\epsilon_{R''}^* \epsilon_{R''} = 1,$$

and ν_R is some number whose possible values will be discovered shortly.

Now define the character of an irreducible representation $\mathfrak{D}^{(\alpha)}$ of $\mathcal{A}(\mathfrak{G}, \omega)$ to be $\chi^{(\alpha)}(a_R) \equiv \operatorname{Tr} \mathfrak{D}^{(\alpha)}(a_R)$. Now note that $\chi^{(\alpha)}(a_{R'})$ is not necessarily equal to $\chi^{(\alpha)}(a_R)$ even if R' is in the class of R in $\mathfrak{G}$. Instead one may take the trace of the equation [where it is assumed that $R' = T^{-1}RT$ as in (34)]

$$[\mathfrak{D}^{(\alpha)}(a_T)]^{-1}\mathfrak{D}^{(\alpha)}(a_R)\mathfrak{D}^{(\alpha)}(a_T) = \frac{\omega_{T^{-1},RT}\omega_{R,T}}{\omega_{T^{-1},T}} \mathfrak{D}^{(\alpha)}(a_{R'})$$

to obtain

$$\chi^{(\alpha)}(a_R) = \frac{\omega_{T^{-1},RT}\omega_{R,T}}{\omega_{T^{-1},T}} \chi^{(\alpha)}(a_{R'}). \tag{38}$$

If one substitutes $a = c_R$ in Eq. (15), there results

$$c_R = \sum_\beta \sum_L \sum_\kappa \mathfrak{D}_{L\kappa}^{(\beta)}(c_R) P_{L\kappa}^{(\beta)}.$$

But since c_R is in the center, Schur's Lemmas demand that

$$\mathfrak{D}_{Ln}^{(\beta)}(c_R) = \begin{cases} 0 & \text{if } L \neq \kappa, \\ \dfrac{1}{l^{(\beta)}} \chi^{(\beta)}(c_R) & \text{if } L = \kappa, \end{cases}$$

whence

$$c_R = \sum_\beta \sum_L \frac{1}{l^{(\beta)}} \chi^{(\beta)}(c_R) P_{LL}^{(\beta)} = \sum_\beta \frac{1}{l^{(\beta)}} \chi^{(\beta)}(c_R) P^{(\beta)}. \tag{39}$$

Of course one must compute $\chi^{(\beta)}(c_R)$. From (37) one has

$$\chi^{(\beta)}(c_R) = \nu_R[\chi^{(\beta)}(a_R) + \epsilon_{R'}\chi^{(\beta)}(a_{R'}) + \cdots].$$

Using (36) and (38), one obtains

$$\chi^{(\beta)}(c_R) = O(C_R)\nu_R\chi^{(\beta)}(a_R), \tag{40}$$

where $O(C_R)$ is the order of the class of R in $\mathfrak{G}$. Hence we obtain

$$c_R = O(C_R)\nu_R \sum_\beta \frac{\chi^{(\beta)}(a_R)}{l^{(\beta)}} P^{(\beta)}. \tag{41}$$

Similarly, if $R' = T^{-1}RT$, then

$$c_{R'} = O(C_R)\nu_{R'} \sum_\beta \frac{\chi^{(\beta)}(a_{R'})}{l^{(\beta)}} P^{(\beta)},$$

since $O(C_{R'}) \equiv O(C_R)$. Now, using (38) and assuming that c_R is not zero, one obtains

$$c_{R'} = \left(\frac{\omega_{T^{-1},T}}{\omega_{T^{-1},RT}}\omega_{R,T}\frac{\nu_{R'}}{\nu_R}\right)c_R \equiv \gamma(T)c_R. \tag{42}$$

So, if c_R is not zero, all the other $c_{R'}$ (corresponding to elements R' of class C_R) are simply proportional to c_R. We now evaluate $\gamma(T)$ in (42).

Substituting c_R for c in (31) and using (37), we obtain

$$gc_R = \nu_R\left(\sum_S \frac{a_{S^{-1}}a_R a_S}{\omega_{S^{-1}S}} + \epsilon_{R'}\sum_S \frac{a_{S^{-1}}a_{R'}a_S}{\omega_{S^{-1},S}} + \cdots\right)$$
$$= \nu_R(c_R + \epsilon_{R'}c_{R'} + \cdots),$$

which by (42) becomes

$$gc_R = \nu_R\left(c_R + \epsilon_{R'}\frac{\omega_{T^{-1},T}}{\omega_{T^{-1},RT}}\omega_{R,T}\frac{\nu_{R'}}{\nu_R}c_R + \cdots\right),$$

or

$$gc_R = (\nu_R + \nu_{R'} + \cdots)c_R, \tag{43}$$

where (36) gave the last result. The series on the right has $o(C_R)$ terms $\nu_{R'}$. The highest value that any ν_R or $\nu_{R'}$ can have is $(g/o(C_R))$, since this is the number of elements T, U, V, $\cdots$ in $\mathfrak{G}$ that transform R into $R' = T^{-1}RT = U^{-1}RU = V^{-1}RV = \cdots$, where we assume all factors are modulus unity [see (2)]. Clearly, by (43), either each of the ν_R, $\nu_{R'}$, etc., is exactly $(g/o(C_R))$, or else $c_R = c_{R'} = \cdots = 0$. In the latter case $\nu_R = \nu_{R'} = \cdots = 0$.

Theorem 1: Given

$$c_R = \sum_{S=1}^{S=S} \frac{\omega_{S^{-1},R}\omega_{S^{-1}R,S}}{\omega_{S^{-1},S}} a_{S^{-1}RS}$$

for all R in group $\mathfrak{G}$, there are no more linearly independent c_R than there were classes in $\mathfrak{G}$. Furthermore, c_R is zero (nonzero) if and only if all $c_{R'}$ are zero (nonzero) for all R' in the class C_R. Finally, if c_R is nonzero, then

$$c_R = g/o(C_R)(a_R + \epsilon_{R'}a_{R'} + \cdots) = \epsilon_{R'}c_{R'}, \tag{44}$$

where $\epsilon_{R'}$ is given by (36).

Definition: If the elements $\{a_R, a_{R'}, \cdots\}$ correspond to a class C_R in $\mathfrak{G}$ that gives nonzero c_R, they will be said to belong to the *ray class* of a_R in $\mathcal{A}(G, \omega)$. If $c_R = 0$, they will be said to belong to the *zeroing class* of a_R.

The term "zeroing class" will hereafter be abbreviated to *zass*, and a "ray class" will be called a *rass*.

Definition: If c_R is nonzero, then the element

$$\kappa_R = (a_R + \epsilon_{R'}a_{R'} + \cdots) = \frac{1}{\nu_R}c_R \tag{45}$$

[where $\epsilon_{R'}$ is given by (36)] is called the *rass sum* of a_R.

From (28), (26), and (38) one obtains

$$P^{(\alpha)} = \sum_{j=1}^{j=l^{(\alpha)}} P^{(\alpha)}_{jj} = \frac{l^{(\alpha)}}{g}\sum_R \chi^{(\alpha)*}(a_R)a_R. \tag{46}$$

This can be written as a linear combination of independent rass sums, since it is in the center:

$$\sum_s \frac{a_{s^{-1}}P^{(\alpha)}a_s}{\omega_{s^{-1},s}} = gP^{(\alpha)} = \frac{l^{(\alpha)}}{g}\sum_{R=1}^{R=g}\chi^{(\alpha)}(a_R)c_R,$$

$$P^{(\alpha)} = \frac{l^{(\alpha)}}{g^2}\sum_{\substack{\text{independent}\\ \text{rasses } j}}(\chi^{(\alpha)*}(a_{R_j})c_{R_j} + \chi^{(\alpha)*}(a_{R_j'})c_{R_j'} + \cdots). \tag{47}$$

In the series above, there is one term for each element of the jth rass of $\mathcal{A}(\mathfrak{G}, \omega)$. Assuming, as before, that $R_j' = T^{-1}R_jT$ and using (44) and (38), there results

$$P^{(\alpha)} = \frac{l^{(\alpha)}}{g^2}\sum_{\substack{\text{rasses}\\ j}}\left(\chi^{(\alpha)}(a_{R_j})c_{R_j} + \frac{1}{\epsilon_{R_j}^*\epsilon_{R_j}}\chi^{(\alpha)}(a_{R_j}),c_R + \cdots\right),$$
$$= \frac{l^{(\alpha)}}{g^2}\sum_{\substack{\text{rasses}\\ j}} O(C_{R_j})\chi^{(\alpha)*}(a_{R_j})c_{R_j},$$
$$= \frac{l^{(\alpha)}}{g}\sum_{\substack{\text{rasses}\\ j}}\chi^{(\alpha)*}(a_{R_j})\kappa_{R_j}, \tag{48}$$

where κ_{R_j} is given by (45).

Finally, (41) and (45) give

$$\kappa_{R_j} = O(C_{R_j})\sum_{\substack{\text{irreducible}\\ \text{representative}\\ (\beta)}}\frac{\chi^{(\beta)}(a_{R_j})}{l^{(\beta)}}P^{(\beta)}. \tag{49}$$

6. CHARACTER THEORY INVOLVING RASSES

The following two theorems show that only the rasses need be considered in manipulations of characters, while the rest of the ray algebra (the zasses) can be ignored.

Theorem 2: If $\chi^{(\alpha)}(a_R) = 0$ for all (α), then $c_R = 0$.

Proof: This follows immediately from (41).

Theorem 3: If $c_{R_k} = 0$, then $\chi^{(\beta)}(a_{R_k}) = 0$ for all (β).

Proof: The expressions (47) for $P^{(\alpha)}$ do not include any elements from a zass. The expressions (46) for $P^{(\alpha)}$ show then that $\chi^{(\alpha)}(a_R)$ is identically zero for an element a_R in a zass.

Now suppose $\{\mathcal{R}(a_1)\ \mathcal{R}(a_2) \cdots \mathcal{R}(a_g)\}$ is a representation of $\mathcal{A}(\mathfrak{G}, \omega)$, i.e., a ray representation of $\mathfrak{G}$ with factor system $\{\omega\}$. An immediate consequence of the preceding formalism is that if $\mathcal{R}$ is not equivalent to one of the $\mathfrak{D}^{(\alpha)}$, it must be equivalent to a direct sum of them. Let

$$\mathcal{R} \sim \oplus \sum_{(\alpha)} f^{(\alpha)}(\mathcal{R}) \mathfrak{D}^{(\alpha)},$$

where the integer $f^{(\alpha)}(\mathcal{R})$ is the *frequency* of $\mathfrak{D}^{(\alpha)}$ in $\mathcal{R}$, or the (α) *contents* of $\mathcal{R}$. Now one has

$$f^{(\alpha)}(\mathcal{R}) = \operatorname{Tr} \mathcal{R}(P_{ii}^{(\alpha)}) = \frac{1}{l^{(\alpha)}} \operatorname{Tr} \mathcal{R}(P^{(\alpha)}),$$

which by (48) expands to

$$f^{(\alpha)}(R) = \frac{1}{g} \sum_{\substack{\text{rasses} \\ j}} o(C_{R_j}) \chi^{(\alpha)*}(a_{R_j}) \operatorname{Tr} \mathcal{R}(a_{R_j}). \quad (50)$$

It is an easy matter to obtain rass–IR orthogonality relations. Substituting (49) into (48) and comparing coefficients yields

$$\frac{1}{g} \sum_{\text{rasses}} o(C_{R_j}) \chi^{(\alpha)}(a_{R_j}) \chi^{(\beta)}(a_{R_j}) = \delta^{\alpha\beta}, \quad (51)$$

while substitution of (48) in (49) yields

$$\frac{o(C_{R_L})}{g} \sum_{(\beta)} \chi^{(\beta)}(a_{R_L}) \chi^{(\beta)}(a_{R_j}) = \delta_{Lj}. \quad (52)$$

7. CALCULATING CHARACTERS: AN EXAMPLE

The following is a table [see (6)] of $\mathcal{A}(D_6\omega)$ derived from a spinor representation $\mathfrak{D}^{(\frac{1}{2})}$ of $O(3)$:

	1	r^2	r^4	ρ_1	ρ_2	ρ_3	r^3	r^5	r	ρ_1'	ρ_2'	ρ_3'
1	1	r^2	r^4	1	2	3	r^3	r^5	r	$1'$	$2'$	$3'$
r^4	r^4	-1	$-r^2$	3	-1	-2	$-r$	$-r^3$	r^5	$-3'$	$-1'$	$2'$
r^2	r^2	r^4	-1	2	3	-1	r^5	$-r$	r^3	$2'$	$-3'$	$1'$
ρ_1	1	-3	-2	-1	r^4	r^2	$-1'$	$-3'$	$-2'$	r^3	r	r^5
ρ_2	2	1	-3	$-r^2$	-1	r^4	$-2'$	$-1'$	$3'$	r^5	r^3	$-r$
ρ_3	3	2	1	$-r^4$	$-r^2$	-1	$3'$	$-2'$	$1'$	$-r$	r^5	$-r^3$
r^3	r^3	r^5	r	$1'$	$2'$	$3'$	-1	$-r^2$	r^4	-1	-2	3
r	r	r^3	r^5	$3'$	$1'$	$2'$	r^4	-1	r^2	3	-1	2
r^5	r^5	$-r$	$-r^3$	$2'$	$-3'$	$-1'$	$-r^2$	$-r^4$	-1	-2	-3	-1
ρ_1'	$1'$	$3'$	$-2'$	$-r^3$	$-r$	r^5	1	-3	2	-1	r^4	$-r^2$
ρ_2'	$2'$	$1'$	$3'$	$-r^5$	$-r^3$	$-r$	2	1	3	$-r^2$	-1	$-r^4$
ρ_3'	$3'$	$-2'$	$-1'$	$-r$	r^5	r^3	-3	-2	1	r^4	r^2	-1

(The notation a_R has been replaced by R for each element of D_6.) The element r corresponds to a rotation by 60° around the six-fold axis, while $\rho_1\rho_2$, ρ_3, ρ_1', ρ_2', and ρ_3' correspond to various 180° rotations around the axis shown in Fig. 1. (These are indicated by the numbers 1, 2, 3, 1′, 2′, and 3′ in the table.)

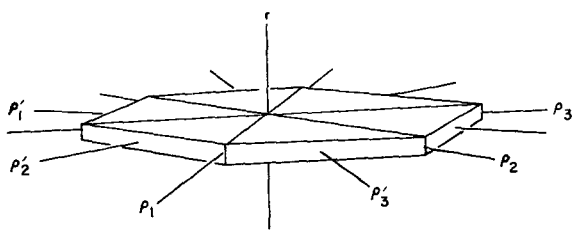


FIG. 1. The group of rotations in $D_6 = D_3 \times C_2$.

There are three rasses and three zasses. The rass sums are $k_1 = 1$, $k_2 = r^2 - r^4$ and $k_3 = r - r^5$. These three form the following algebra:

	k_1	k_2	k_3
k_1	k_1	k_2	k_3
k_2	k_2	$2k_1 - k_2$	k_3
k_3	k_3	k_3	$2k_1 + k_2$

The minimal equations satisfied by k_2 and k_3 are the following:

$$(k_2)^2 + k_2 - 2k_1 = 0, \quad (k_3)^3 - 3k_1 = 0,$$

with the roots $(1, -2)$ for k_2, and $(\sqrt{3}, -\sqrt{3}, 0)$ for

k_3. These roots will serve (sometimes repeatedly[15]) as coefficients $O(C_R)\chi^{(\beta)}(a_R)/l^{(\beta)} = a_R^{(\beta)}$ in (49):

$$\kappa_R = O(C_R)\sum_\beta \frac{\chi^{(\beta)}(a_R)}{l^{(\beta)}} P^{(\beta)} = \sum_\beta a_R^{(\beta)} P^{(\beta)}. \quad (49')$$

The coefficients are tabulated below:

	k_1	k_2	k_3
$a^{(\alpha)}$	1	1	$\sqrt{3}$
$a^{(\beta)}$	1	1	$-\sqrt{3}$
$a^{(\gamma)}$	1	−2	0

The character table is completed once the values of $l^{(\alpha)}$, $l^{(\beta)}$, and $l^{(\gamma)}$ are obtained. The relation (51) gives the following simple formula for these:

$$\left(\frac{1}{l^{(\alpha)}}\right)^2 = \frac{1}{g}\sum_{\substack{\text{rasses}\\ j}} \frac{1}{oC_{R_j}} \left(\frac{oC_{R_j}\chi^{(\alpha)}(a_{R_j})}{l^{(\alpha)}}\right)\left(\frac{oC_{R_j}\chi^{(\alpha)}(a_{R_j})}{l^{(\alpha)}}\right)$$

$$= \frac{1}{g}\sum_j \frac{1}{oC_{R_j}} |a_{R_j}^{(\alpha)}|^2.$$

The resulting ray character table of D_6 is shown below:

	1	r^2	r
$\chi^{(\alpha)}$	2	1	$\sqrt{3}$
$\chi^{(\beta)}$	2	1	$-\sqrt{3}$
$\chi^{(\gamma)}$	2	−2	0

This $\mathcal{A}(D_6, \omega)$ has diophantine solution (22) $2^2 + 2^2 + 2^2 = 12$. The characters of other rass elements r^4 and r^5 follow from (38). The characters of zass elements are zero.

Other "double value" representations of $O(3)$ subduce ray representations of D_6. The contents of these are easily obtained by writing "down" the traces of them for 1, r^2, and r, using the formula

$$\chi^{(J/2)}(\omega) = [\sin(J+1)\omega/2]/(\sin \omega/2)$$

(where $\omega = 2\pi/3$ for r^2 and $\omega = \pi/3$ for r) and then applying (50).

8. CANONICAL RASSES

So far it has been assumed that one treated a given factor system without transformation, provided it was in a "unimodular" form corresponding to (2) and (11).

However, the theory of Sec. 5 inclines one to transform to a (generally) different factor system $\{\omega'\}$ that is p-equivalent to $\{\omega\}$ but more convenient to deal with. For if

$$\kappa_R = a_R + \epsilon_{R'} a_{R'} + \cdots \quad (45)$$

is a rass sum, one may write a new element

$$b_{R'} \equiv \epsilon_{R'} a_{R'} \quad (53)$$

for each element $a_{R'}$ in the sum where $\epsilon_{R'}$ is given by (36) as

$$\epsilon_{R'} = \frac{\omega_{T^{-1},RT}\omega_{R,T}}{\omega_{T^{-1},T}}$$

if $R' = T^{-1}RT$. Now if all these new elements $b_{R'}$ for all the rass sums in $\mathcal{A}(\mathfrak{G}, \omega)$ are collected along with the a_R (which we label $b_R \equiv a_R$ for notational convenience) and the members of zasses a_z (which we label $b_z \equiv a_z$), the resulting set $\{b_R \cdots b_S \cdots b_T \cdots\}$ is a basis of a new ray algebra $\mathcal{A}(\mathfrak{G}, \omega')$, which is said to have *canonical rasses.*

One advantage of $\mathcal{A}(\mathfrak{G}, \omega')$ is that if $\{\cdots M(b_R)\cdots M(b_s)\cdots M(b_T)\cdots\}$ is a representation of it, then one has

$$\text{Tr } M(b_{R'}) = \epsilon_{R'} \text{ Tr } M(a_{R'})$$

$$= \epsilon_{R'} \frac{\omega_{T^{-1},T}}{\omega_{T^{-1},RT}\omega_{R,T}} \text{ Tr } M(a_R),$$

or

$$\text{Tr } M(b_{R'}) = \text{Tr } M(a_R) = \text{Tr } M(b_R), \quad (54)$$

provided $b_{R'}$ and b_R are in the same rass. If they are both in a *zass*, then the identity

$$\text{Tr } M(b_{R'}) = 0 = \text{Tr } M(b_R) \quad (55)$$

holds in any case.

Furthermore, in this $\mathcal{A}(\mathfrak{G}, \omega')$ it is true that

$$b_v^{-1} b_{R'} b_v = \omega'_{v^{-1},v} b_{v^{-1}Rv'} \quad (56)$$

for any base element b_v, provided $b_{R'}$ is in a rass.

9. INDUCTION INVOLVING RAY ALGEBRAS

The process of *induction* in groups is well described by Coleman[16] and Bradley.[17] A slight modification of their derivation allows us to perform the same operation with ray algebras without involving a covering group.

The idea is to obtain a representation [labeled $\mathcal{R} = M \uparrow \mathcal{A}(\mathfrak{G}, \omega)$] of ray algebra $\mathcal{A}(\mathfrak{G}, \omega)$ if you know a representation $\{\cdots M(a_H)\cdots\}$ of a subalgebra $\mathcal{A}(\mathcal{H}, \omega)$ of $\mathcal{A}(\mathfrak{G}, \omega)$ corresponding to a subgroup $\mathcal{H}$ of $\mathfrak{G}$.

[15] But the roots themselves (of the minimal equation) will never be repeated. If they were, a nilpotent η in the center could be constructed, and this would be a contagious nilpotent in $\mathcal{A}(\mathfrak{G}, \omega)$, since $\eta^n a^n = 0 = (\eta a)(\eta a)\cdots(\eta a) = (\eta a)^n$, contrary to Theorem 1.

[16] A. J. Coleman, *Induced Representations with Applications to S_n and GL(n), Queens Papers No. 4* (Queen's University, Kingston, Ontario, Canada, 1966).

[17] C. J. Bradley, J. Math. Phys. 7, 1146 (1966).

Proceeding in the manner of Bradley,[17] one constructs left cosets of $\mathcal{H}$ in $\mathcal{G}$,

$$\mathcal{G} = \sum_R R\mathcal{H}, \tag{57}$$

while singling out one element R from each coset to be the Rth *coset leader.*

Now, returning to the corresponding ray algebras, suppose m vectors $\{{}^1\Psi_1\ {}^1\Psi_2 \cdots {}^1\psi_m\}$ form a basis of representation M,

$$a_H\ {}^1\Psi_i = \sum_{i=1}^{i=m} {}^1\Psi_i M_{ij}(a_H). \tag{58}$$

Let m new vectors $\{{}^R\Psi_1\ {}^R\Psi_2 \cdots {}^R\Psi_m\}$ be defined for each coset leader in (57):

$$a_R\ {}^1\Psi_j \equiv {}^R\Psi_j. \tag{59}$$

Then, if a_g is any element of $\mathcal{A}(\mathcal{G}, \omega)$, one has that

$$a_g\ {}^S\Psi_j = a_g a_S\ {}^1\Psi_j = \left(\frac{a_R a_{R^{-1}}}{\omega_{R^{-1},R}}\right) a_g g_S\ {}^1\Psi_j. \tag{60}$$

If you pick a leader R such that $a_{R^{-1}} a_g a_S$ is an element of $\mathcal{A}(\mathcal{H}, \omega)$, (60) becomes

$$a_g\ {}^S\Psi_j = \frac{a_R}{\omega_{R^{-1},R}} \sum_{i=1}^{i=m} {}^1\Psi_i M_{ij}(a_{R^{-1}} a_g a_S) \tag{61}$$

by (58), which in turn becomes

$$a_g\ {}^S\Psi_j = \sum_{i=1}^{i=m} {}^R\Psi_i M_{ij}(a_{R^{-1}} a_g a_S)/\omega_{R^{-1},R}$$

by (59). The vectors $\{{}^1\psi_1 \cdots {}^R\psi_j \cdots\}$ form the basis of the *induced ray representation* $M \uparrow \mathcal{A}(\mathcal{G}, \omega) \equiv \mathcal{R}$. One matrix $\mathcal{R}(a_g)$ of this is depicted in Fig. 2.

We now assume that $\mathcal{A}(\mathcal{G}, \omega)$ has canonical rasses and proceed to compute the traces of the induced

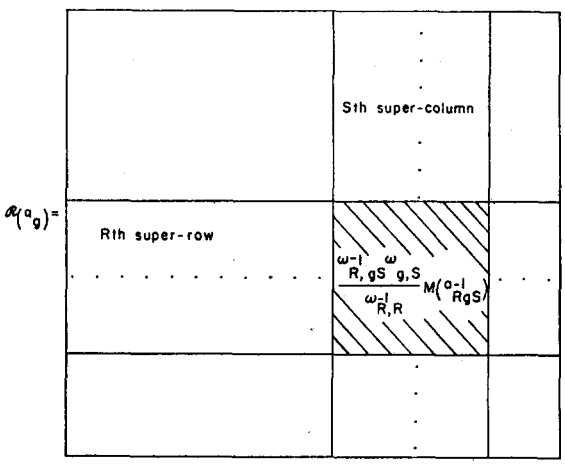


FIG. 2. The matrix $M \uparrow \mathcal{A}(\mathcal{G}, \omega) \equiv \mathcal{R}(a_g)$ induced by representation M of ray-subalgebra $\mathcal{A}(\mathcal{H}, \omega)$.

representation. One has

$$\operatorname{Tr}(M \uparrow \mathcal{A}(\mathcal{G}, \omega))(a_g) = \sum_{\substack{\text{leaders} \\ R}} \operatorname{Tr} M(a_{R^{-1}} a_g a_R)/\omega_{R^{-1},R}\{\delta(\mathcal{H}, R^{-1}gR)\},$$

where δ is defined by

$$\delta(\mathcal{H}, R^{-1}gS) \equiv \begin{cases} 1, & \text{if } R^{-1}gR \in \mathcal{H}, \\ 0, & \text{if } R^{-1}gR \notin \mathcal{H}. \end{cases}$$

Using (56), one obtains

$$\operatorname{Tr}(M \uparrow \mathcal{A})(a_g) = \sum_{\substack{\text{leaders} \\ R}} \operatorname{Tr} M(a_{R^{-1}gR})\{\delta(\mathcal{H}, R^{-1}gR\}, \tag{62}$$

which can be converted to a sum over all g using the following lemmas with (54) and (55).

Lemma 1: If a leader R transforms g into $h \in \mathcal{H}$ ($R^{-1}sR = L \in \mathcal{H}$), then all the members of coset $R\mathcal{H}$ will transform g into the class C_h of h in $\mathcal{H}$.

Proof: If $R^{-1}gR = h$ is in $\mathcal{H}$, and if h' is in $\mathcal{H}$, then $h'^{-1}(R^{-1}gR)h' = h'^{-1}hh'$. This becomes

$$(Rh')^{-1}g(Rh') = h'^{-1}hh'.$$

Lemma 2: If a leader R fails to transform g into an $h \in \mathcal{H}$, then no member of coset $R\mathcal{H}$ can transform g into $\mathcal{H}$.

Proof: Assume $R^{-1}gR = G$ is not in $\mathcal{H}$. Then $(Rh')^{-1}g(Rh') = h'^{-1}Gh'$, and clearly this is not in $\mathcal{H}$.

The result is

$$\operatorname{Tr}(M \uparrow \mathcal{A})(a_g) = \frac{1}{h} \sum_{\substack{\text{all} \\ \text{elements} \\ G \in \mathcal{G}}} \operatorname{Tr} M(a_{G^{-1}} a_g a_G)\delta(\mathcal{H}, G^{-1}gG). \tag{63}$$

Now assume that g is in class C_j of $\mathcal{G}$ and that a_g is in the jth rass of $\mathcal{A}(\mathcal{G}, \omega)$. Then it follows that

$$\operatorname{Tr}(M \uparrow \mathcal{A})(a_g) = \operatorname{Tr}(M \uparrow \mathcal{A})_j = \frac{g}{ho(C_j)} \sum_{\substack{\text{all } h \\ \text{in } C_j \cap \mathcal{H}}} \operatorname{Tr} M(a_h). \tag{64}$$

Assuming that this jth class C_j of $\mathcal{G}$ contains classes $C_1^j, C_2^j, C_3^j, \cdots, C_{n_j}^j$ of $\mathcal{H}$, one has

$$\operatorname{Tr}(M \uparrow \mathcal{A})_j = \frac{g}{ho(C_j)} \sum_{n=1}^{n=n_j} o(C_n^j) \operatorname{Tr}(M)_n^j, \tag{65}$$

using the notation $\operatorname{Tr}(M)_n^j = \operatorname{Tr}(M(a_n))$ where a_n is any member of rass[18] C_n^j of $\mathcal{A}(\mathcal{H}, \omega)$.

[18] One should note that there cannot be a zass of $\mathcal{A}(\mathcal{H}, \omega)$ within a rass of $\mathcal{A}(\mathcal{G}, \omega)$.

The formalism is now set up to prove a Frobenius reciprocity theorem for ray algebras. Using Bradley's notation ($\downarrow$) for subduction, letting $D^{(A)}$ denote an irreducible representation of $\mathcal{A}(\mathcal{K})$ and $\mathfrak{D}^{(\alpha)}$, an irreducible representation of $\mathcal{A}(\mathcal{G})$, one obtains the following theorem:

Theorem 5:

$$f^{(\alpha)}(D^{(A)} \uparrow \mathcal{A}(\mathcal{G})) = f^{(A)}(\mathfrak{D}^{(\alpha)} \downarrow \mathcal{A}(\mathcal{K})).$$

Proof: Using (50) and (65), one has

$$f^{(\alpha)}(D^{(A)} \uparrow \mathcal{A}(\mathcal{G})) = \frac{1}{g} \sum_{\substack{\text{rasses } j \\ \text{of } \mathcal{A}(\mathcal{G})}} o(C_j)\chi_j^{(\alpha)*}\left\{\frac{g}{h_0(C_j)} \sum_{\substack{\text{rasses } n \\ \text{of } \mathcal{A}(\mathcal{K}) \\ \text{in } C_j}}^{n=n_j} o(c_n^j)X^{(A)}\binom{j}{n}\right\},$$

where $X^{(A)}\binom{j}{n}$ is a character for rass c_n^j of $\mathcal{A}(\mathcal{K})$. Rearranging this equation, one obtains

$$f^{(\alpha)}(D^{(\alpha)} \uparrow (\mathcal{G})) = \frac{1}{h} \sum_{\substack{\text{all rasses} \\ \binom{j}{n} \text{ of } \mathcal{A}(\mathcal{K})}} o(c_n^j)\chi_j^{(\alpha)*}X^{(A)}\binom{j}{n},$$

which, by (50) and the fact that $f^{(\alpha)*} = f^{(\alpha)}$, proves the theorem.

With this established, it is an easy matter to prove the following two generalizations [Eqs. (66) and (67) below]:

Theorem 6: If $\mathcal{A}(\mathcal{K}')$ and $\mathcal{A}(\mathcal{K}'')$ are two ray subalgebras of $\mathcal{A}(\mathcal{G})$, $D^{(A)'}$ is an IR of $\mathcal{A}(\mathcal{K}')$, $D^{(B)''}$ is an IR of $\mathcal{A}(\mathcal{K}'')$, and $\mathfrak{D}^{(\alpha)}$ is an IR of $\mathcal{A}(\mathcal{G})$, then one has the following relations:

$$f^{(A)'}(D^{(B)''} \uparrow \mathcal{A}(\mathcal{G})) \downarrow \mathcal{A}(\mathcal{K}')) = f^{(B)''}((D^{(A)'} \uparrow \mathcal{A}(\mathcal{G})) \downarrow \mathcal{A}(\mathcal{K}'')) \quad (66)$$

and

$$f^{(A)'}(D^{(B)''} \downarrow \mathcal{A}(\mathcal{K}' \cap \mathcal{K}'')) \uparrow \mathcal{A}(\mathcal{K}') = f^{(B)''}(D^{(A)'} \downarrow \mathcal{A}(\mathcal{K}' \cap \mathcal{K}'')) \uparrow \mathcal{A}(\mathcal{K}''). \quad (67)$$

Proof: The proof of (67) is virtually identical to the proof of (66), which is now given. Assuming, as in Sec. 6, that

$$(D^{(A)'} \uparrow \mathcal{A}(\mathcal{G})) \sim \oplus \sum_{(\alpha)} f^{(\alpha)}(D^{(A)'} \uparrow \mathcal{A}(\mathcal{G}))\mathfrak{D}^{(\alpha)}$$

and

$$\mathfrak{D}^{(\alpha)} \downarrow \mathcal{A}(\mathcal{K}'') \sim \oplus \sum_{(B)''} f^{(B)''}(\mathfrak{D}^{(\alpha)} \downarrow \mathcal{A}(\mathcal{K}''))D^{(B)''},$$

one obtains the following expression for the right-hand side of (66):

$$f^{(B)''}(D^{(A)'} \uparrow \mathcal{A}(\mathcal{G})) \downarrow \mathcal{A}(\mathcal{K}'')) = \sum_{(\alpha)} f^{(\alpha)}(D^{(A)'} \uparrow \mathcal{A}(\mathcal{G}))f^{(B)''}(\mathfrak{D}^{(\alpha)} \downarrow \mathcal{A}(\mathcal{K}'')).$$

Applying Theorem 5, one has

$$f^{(B)''}(D^{(A)'} \uparrow \mathcal{A}(\mathcal{G}) \downarrow \mathcal{A}(\mathcal{K}'')) = \sum_{(\alpha)} f^{(\alpha)}(D^{(A)'} \uparrow \mathcal{A}(\mathcal{G}))f^{(\alpha)}(D^{(B)''} \uparrow \mathcal{A}(\mathcal{G})),$$

which is clearly equal to the left side of (66).

10. OBTAINING RAY REPRESENTATION EXAMPLES

The reciprocity theorems allow the extension of the recursive techniques of Seitz[19] and Boerner[20] (for finding the IR of solvable groups) to the corresponding ray algebras. This will be described here without detailed proof, and examples will be treated.

Suppose $\mathcal{A}(\mathcal{K})$ is a ray subalgebra of $\mathcal{A}(\mathcal{G})$ corresponding to normal subgroup $\mathcal{K}$ of $\mathcal{G}$ with prime index p [$O(\mathcal{G}/\mathcal{K}) = p$]. Then the IR $\{\mathfrak{D}^{(\alpha)}, \mathfrak{D}^{(\beta)}, \cdots\}$ of $\mathcal{A}(\mathcal{G})$ will be obtained from the IR $\{D^{(A)}, D^{(B)}, \cdots\}$ of $\mathcal{A}(\mathcal{K})$ by the process of induction $\mathfrak{D}^{(\alpha)} = D^{(B)} \uparrow \mathcal{A}(\mathcal{G})$, as described in Sec. 9, and by a process called *extension* $\mathfrak{D}^{(\beta)} = D^{(B)} \rightarrow \mathcal{A}(\mathcal{G})$, which is described now.

Extension is used when $D^{(B)} \uparrow \mathcal{A}(\mathcal{G})$ is not an irreducible representation of $\mathcal{A}(\mathcal{G})$. This can be checked by using Eqs. (65) and (51). In this case, and under the conditions listed above, there is an IR $\mathfrak{D}^{(\beta)}$ of $\mathcal{A}(\mathcal{G})$ such that $\mathfrak{D}^{(\beta)} \downarrow \mathcal{A}(\mathcal{K}) = D^{(B)}$. Furthermore, the representation $D^{(B)}G$, defined by

$$D^{(B)}G(a_H) \equiv \frac{D^{(B)}(a_G{}^{-1}a_H a_G)}{\omega_{G^{-1},G}} \quad (68)$$

for a given G in $\mathcal{G}$ and all a_H in $\mathcal{A}(\mathcal{K})$, is equivalent to $D^{(B)}$; i.e.,

$$D^{(B)}G = \mathcal{C}^{-1}D^{(B)}\mathcal{C} \quad (69)$$

for some matrix $\mathcal{C}$. One may obtain all the solutions $\mathcal{C}$ to (67) by using the unit dyads of $\mathcal{A}(\mathcal{K})$. Clearly then,

$$\mathfrak{D}^{(\alpha)}(a_G) = \mu_G^{(\alpha)}\mathcal{C}, \quad (70)$$

where the allowed constants $\mu^{(\alpha)}$, $\mu^{(\beta)}$, $\cdots$, are obtained by inspecting $\mathcal{A}(\mathcal{G})$. A number of inequivalent IR $\{\mathfrak{D}^{(\alpha)}, \mathfrak{D}^{(\beta)}, \cdots\}$ of $\mathcal{A}(\mathcal{G})$ will result, corresponding to the number of distinct allowed constants ($\mu^{(\alpha)}$, $\mu^{(\beta)}$, $\cdots$).

A. Double-Valued Representations of D_n

The group D_n has a cyclic subgroup C_n of index 2, which is generated by an element R which satisfies the relation

$$R^n = 1. \quad (71)$$

The ray algebra $\mathcal{A}(D_n)$ derived from spinors has a ray subalgebra $\mathcal{A}(C_n)$, which is generated by an element

[19] C. M. Seitz, Ann. Math. 37, 17 (1936).

[20] H. Boerner, *Representations of Groups* (Interscience Publishers, Inc., New York, 1963), p. 95.

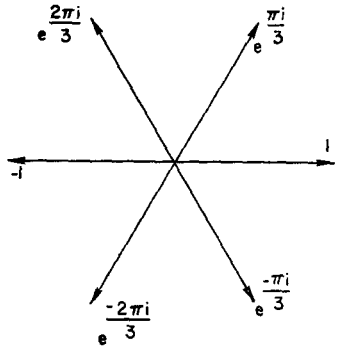


FIG. 3. Irreducible representations of the generator of C_6.

a_R which satisfies the relation below:

$$(a_R)^n = -a_1. \tag{72}$$

Now there are as many IR (labeled as $\{D^{(0)}_{(R)} D^{(1)}_{(R)} \cdots D^{(A)}_{(R)} \cdots\}$) of C_n as there are nth roots of unity. Similarly, there are as many IR $\{\cdots D^{(A)}(a_R) \cdots\}$ of $\mathcal{A}(C_n)$ as there are roots of minus one. In fact, for C_n one has $D^{(A)}(R) =$ [an nth root of (1)], while for $\mathcal{A}(C_n)$ one has $D^{(A)}(R) =$ [an nth root of (-1)].

Now the IR $\mathfrak{D}^{(\alpha)}$ of $D_n(\mathcal{A}(D_n))$ are obtained from the IR $D^{(A)}$ of $C_n(\mathcal{A}(C_n))$ by induction or extension, depending on whether case (a) or case (b), shown below, is relevant.

Case (a): $D^{(A)}$ not equal to $D^{(A)*}$. In this case

$$D^{(A)} \uparrow D_n \sim D^{(A)*} \uparrow D_n [D^{(A)} \uparrow \mathcal{A}(D_n) \sim D^{(A)*} \uparrow \mathcal{A}(D_n)]$$

and induction yields one two-dimensional IR of D_n [of $\mathcal{A}(D_n)$] for each conjugate pair $D^{(A)}$ and $D^{(A)*}$.

Case (b): $D^{(A)}$ equals $D^{(A)*}$. In this case extension yields two inequivalent one-dimensional IR of D_n [of $\mathcal{A}(D_n)$].

Examination of the following tables for $n = 2, 3, 4$, and 6 should make this clear, and the results for arbitrary n should thereby be transparent.

$n = 6$. The IR of C_6 and $\mathcal{A}(C_6)$ are indicated by vectors in complex plane in Figs. 3 and 4, respectively.

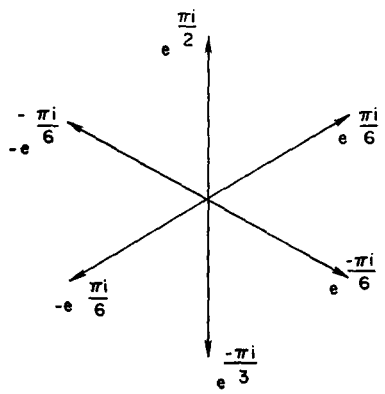


FIG. 4. Irreducible representations of the generator of $\mathcal{A}(C_6, \omega)$.

These yield IR of D_6 and $\mathcal{A}(D_6)$, whose character tables are drawn below:

	1	R^2	ρ	R^3	R^5	ρ'
(extended)	1	1	1	1	1	1
(extended)	1	1	−1	1	1	−1
(induced)	2	−1	0	2	−1	0
(extended)	1	1	1	−1	−1	−1
(extended)	1	1	−1	−1	−1	1
(induced)	2	−1	0	−2	1	0

	1	R^2	R^5
(induced)	2	1	$\sqrt{3}$
(induced)	2	1	$-\sqrt{3}$
(induced)	2	−2	0

$n = 4$. The IR of C_4 and $\mathcal{A}(C_4)$ are indicated in Figs. 5 and 6, respectively. These yield IR of D_4 and $\mathcal{A}(D_4)$

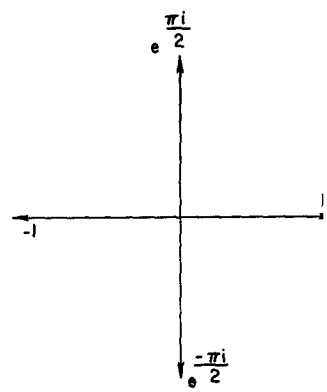


FIG. 5. Irreducible representations of the generator of C_4.

with the following character tables:

	1	R^2	R^3	C_1	C_2
(extended)	1	1	1	1	1
(extended)	1	1	−1	−1	1
(extended)	1	1	−1	1	−1
(extended)	1	1	1	−1	−1
(induced)	2	−2	0	0	0

	1	R
(induced)	2	$\sqrt{2}$
(induced)	2	$-\sqrt{2}$

$n = 3$. The IR of C_3 and $\mathcal{A}(C_3)$ are indicated in Figs. 7 and 8, respectively. These yield the following

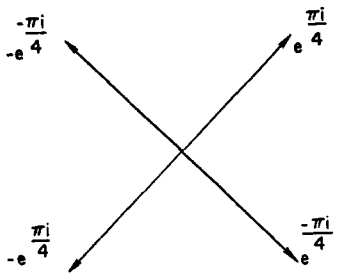


FIG. 6. Irreducible representations of the generator of $\mathcal{A}(C_4, \omega)$.

character tables for the IR of D_3 and $\mathcal{A}(D_3)$:

	a_R		
(extended)	1	1	1
(extended)	1	1	−1
(induced)	2	−1	0

(extended)	1	−1	i
(extended)	1	−1	$-i$
(induced)	2	1	0

Note that $\mathcal{A}(D_3)$ is p-equivalent to D_3.

$n = 2$. The IR of C_2 and $\mathcal{A}(C_2)$ are indicated in Figs. 9 and 10.

The resulting IR of D_2 and $\mathcal{A}(D_2)$ are characterized as follows:

	1	R	R'	R''	
(extended)	1	1	1	1	
(extended)	1	−1	1	−1	(induced) 1
(extended)	1	−1	1	−1	2
(extended)	1	−1	−1	1	

B. The Groups T and D

The tetrahedral group T is of order 12 and contains the group D_2 discussed in the preceding section. The representations of a nontrivial ray algebra $\mathcal{A}(T)$, derived from spinors, are characterized below along

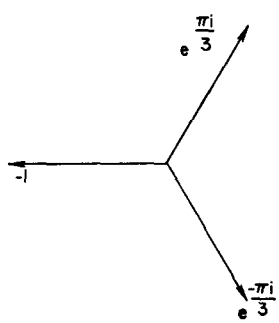


FIG. 7. Irreducible representations of the generator of C_3.

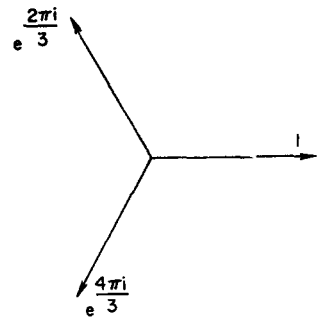


FIG. 8. Irreducible representations of the generator of $\mathcal{A}(C_3, \omega)$.

with the representations of the group T itself:

T	1	r	r^2	$R(2)$
1	1	1	1	1
1	1	$\exp\left(\frac{2\pi i}{3}\right)$	$\exp\left(\frac{-2\pi i}{3}\right)$	1
1	1	$\exp\left(\frac{-2\pi i}{3}\right)$	$\exp\left(\frac{2\pi i}{3}\right)$	1
3	3	0	0	−1

$\mathcal{A}(T)$	a_1	a_r	a_{r_2}
2	2	1	1
2	2	$\exp\left(\frac{2\pi i}{3}\right)$	$\exp\left(\frac{-2\pi i}{3}\right)$
2	2	$\exp\left(\frac{-2\pi i}{3}\right)$	$\exp\left(\frac{2\pi i}{3}\right)$

Here r and r^2 indicate the classes (rasses) of 120° rotations around the points and faces of a tetrahedron,

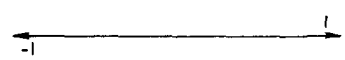


FIG. 9. Irreducible representations of the generator of C_2.

while the $R(2)$ indicates the class $\{R, R', R''\}$ of 180° rotations about the tetrahedron edges. [Clearly, $a_{R(2)}$ is in a zass of $\mathcal{A}(T)$.] Note that all the representations of $\mathcal{A}(T)$ are extensions of the one IR of $\mathcal{A}(D_2)$.

FIG. 10. Irreducible representations of the generator of $\mathcal{A}(C_2, \omega)$.

The octahedral (cubic) group O is of order 24 and has T as a subgroup. The representations of O and a nontrivial ray algebra $\mathcal{A}(O)$ are characterized below. Here class r of O contains classes r and r^2 of T. $R(4)$ and $R(2)$ are classes of 90° and 180° rotations, respectively, around cube faces. $i(2)$ is the class of 180° rotations around cube edges.

	0	1	r	$R_{(4)}$	$R_{(2)}$	$i_{(2)}$
(extended)	1	1	1	1	1
(extended)	1	1	−1	1	−1
(induced)	2	−1	0	2	0
(extended)	3	0	−1	−1	1
(extended)	3	0	1	−1	−1

$\mathcal{A}(0)$	a_1	a_r	$a_{R(4)}$
(extended)	2	1	$\sqrt{2}$
(extended)	2	1	$-\sqrt{2}$
(induced)	4	−1	0

The first two IR of $\mathcal{A}(0)$ are simply extensions of the one real IR of $\mathcal{A}(\mathbf{T})$. The third one can be induced by either of the complex IR of $\mathcal{A}(\mathbf{T})$.

The *diophantine solution* of $\mathcal{A}(0)$ is $2^2 + 2^2 + 4^2 = 24$. One sees from Theorem 1 that the solution $2^2 + 2^2 + 2^2 + 2^2 + 2^2 + 2^2 = 24$ could never exist for any ray algebra of 0.

11. OUTER PRODUCTS OF RAY REPRESENTATIONS

Suppose one has a number n of different non-p-equivalent factor systems defined over a group $\mathfrak{G}$, such that

$$\{\omega^{(1)}_{R,S} = 1\}\{\omega^{(2)}_{R,S}\cdots\}\{\cdots\omega^{(3)}_{R,S}\cdots\}\cdots\{\cdots\omega^{(n)}_{R,S}\cdots\}.$$

(Here $\omega^{(1)}$ is the trivial factor system.) Suppose also that no other factor system can be defined over $\mathfrak{G}$ that is not p-equivalent to one of these.

Denote by $\{\mathfrak{D}^{(\alpha)}, \mathfrak{D}^{(\alpha)'}, \cdots\}$ the set of all IR of $\mathcal{A}(\mathfrak{G}, \omega^{(\alpha)})$ and by $\{\mathfrak{D}^{(\beta)}, \mathfrak{D}^{(\beta)}, \cdots\}$ the IR of $\mathcal{A}(\mathfrak{G}, \omega^{(b)})$. Thus,

$$\mathfrak{D}^{(\alpha)}(a_R)\mathfrak{D}^{(\alpha)}(a_S) = \omega^{(a)}_{R,S}\mathfrak{D}(a_{RS}),$$
$$\mathfrak{D}^{(\beta)}(a_R)\mathfrak{D}^{(\beta)}(a_S) = \omega^{(b)}_{R,S}\mathfrak{D}(a_{RS}). \tag{73}$$

Consider the outer (tensor) products of two of these representations:

$$\{\cdots\mathfrak{D}^{(\alpha)}(a_R)\otimes\mathfrak{D}^{(\beta)}(a_R), \cdots, \mathfrak{D}^{(\alpha)}(a_S)\otimes\mathfrak{D}^{(\beta)}(a_S), \cdots\}.$$

These form a ray representation of $\mathfrak{G}$:

$$[\mathfrak{D}^{(\alpha)}(a_R)\otimes\mathfrak{D}^{(\beta)}(a_R)][\mathfrak{D}^{(\alpha)}(a_S)\otimes\mathfrak{D}^{(\beta)}(a_S)]$$
$$= \omega^{(a)}_{R,S}\omega^{(b)}_{R,S}[\mathfrak{D}^{(\alpha)}(a_{RS})\otimes\mathfrak{D}^{(\beta)}(a_{RS})]. \tag{74}$$

But by (74) the factor system of this representation is (in general) p-equivalent to some other factor system $\omega^{(c)}$. Therefore $(\mathfrak{D}^{(\alpha)}\otimes\mathfrak{D}^{(\beta)})$ can be reduced by similarity transformation $\mathfrak{C}$, and projective transformations $\{\cdots c_R\cdots c_S\cdots\}$, to a direct sum of the IR $\{\mathfrak{D}^{(\gamma)}, \mathfrak{D}^{(\gamma)'}\}$ of $\mathcal{A}(\mathfrak{G}, \omega^{(c)})$:

$$\mathfrak{C}^{-1}[c_R\mathfrak{D}^{(\alpha)}(a_R)\otimes\mathfrak{D}^{(\beta)}(a_R)]\mathfrak{C}$$
$$= \begin{vmatrix} \boxed{\mathfrak{D}^{(\gamma)}(a_R)} & & \\ & \boxed{\mathfrak{D}^{(\gamma)'}(a_R)} & \\ & & \ddots \end{vmatrix}. \tag{75}$$

One may use the unit dyads to compute the constants c_R and the matrices $\mathfrak{C}$ indicated in (75). The components of the $\mathfrak{C}$ matrix are *Clebsch–Gordan coefficients.*

One should note the difference between the preceding theory and that of Rudra. For example, in Eq. (21) of his first paper[7] he defines the direct product of two representations Γ_μ and Γ_ν belonging to a particular factor system $\{\omega_{R,S}\}$ by

$$\Gamma_{\mu\otimes\nu}(R)_{ik,jL} \equiv (\Gamma_\mu(R)\otimes\Gamma_\mu(R))_{ik,jL}$$
$$= \left[\prod_{P\in\mathfrak{G}}\omega_{R,P}\omega_{P,R}\right]^{-\frac{1}{2}g}\Gamma_\mu(R)_{ij}\Gamma_\nu(R)_{kL}$$

and supposedly obtains a representation $\Gamma_{\mu\otimes\nu}$ that belongs to that *same* factor system. But we have shown that this is impossible if $\{\omega_{R,S}\}$ is not p-equivalent to the trivial factor system.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Random Walks on Lattices. III. Calculation of First-Passage Times with Application to Exciton Trapping on Photosynthetic Units*

ELLIOTT W. MONTROLL
Department of Physics and Astronomy, University of Rochester, Rochester, New York and Instituut-Lorentz, Leiden, Netherlands

(Received 1 September 1968)

The following statistical problem arises in the theory of exciton trapping in photosynthetic units: Given an infinite periodic lattice of unit cells, each containing N points of which $(N-1)$ are chlorophyll molecules and one is a trap; if an exciton is produced with equal probability at any nontrapping point, how many steps on the average are required before the exciton reaches a trapping center for the first time? It is shown that, when steps can be taken to near-neighbor lattice points only, as $N \to \infty$, our required number of steps is

$$\langle n \rangle \sim \begin{cases} N^2/6, & \text{linear chain,} \\ \pi^{-1} N \log N, & \text{square lattice,} \\ 1.5164N, & \text{single cubic lattice.} \end{cases}$$

The correction terms for medium and relatively small N are obtained for a number of lattices.

1. INTRODUCTION

Photosynthesis in plants is the employment of light energy for the production of sugars and carbohydrates from CO_2 and H_2O, according to the chemical reaction

$$nCO_2 + nH_2O \to (CH_2O)_n + nO_2.$$

The detailed conversion occurs in an object called a photosynthetic unit which, among other things, contains a network or lattice of chlorophyll molecules. The long chain of events which starts with the absorption of a photon and ends with the production of the sugar can be separated into three main processes:

(a) A photon is absorbed by a chlorophyll molecule, exciting the molecule.

(b) The excitation or "exciton" is transferred from one chlorophyll molecule to a neighboring one on the lattice with the original excited molecule reverting to the ground state. This process continues, the exciton enjoying a random walk through the lattice.

(c) The exciton finally reaches a trapping center, triggering a detailed chemical reaction which is involved in the formation of sugar.

We shall be concerned with the statistical aspects of process (b) in this paper. A detailed survey of the current understanding of the various processes can be found in the excellent review by Duysens.[1]

Investigations are usually carried out on pigmented algae or bacteria suspended in a liquid medium. One type of experiment involves the radiation of the organism with light of known intensity and frequency, observing the rate of O_2 production. As early as 1932 Emerson and Arnold[2] probed the time lapse between processes (a) and (c) by radiating the specimens with light pulses of very short duration separated by much longer time intervals and observing the response of the system through O_2 formation. One factor contributing to the length of time between the original excitation and the formation of the O_2 molecule is the duration of the random walk discussed in (b) before trapping.

A maximum "flash yield" of oxygen per photon is obtained when the duration of the pulse is very short (say, $< 10^{-5}$ sec) and the interval between pulses is very long (say, 0.05 sec). Various experiments[1] indicate that eight to ten photons are required to liberate an O_2 molecule and the maximum O_2 yield is about one O_2 molecule per 200–500 chlorophyll molecules. It has been suggested that there must be about 250 chlorophyll molecules per trap in order to obtain the observed O_2 yields.

The model[1] which evolves from such experiments (and a statistical problem which it suggests) is as follows: The chlorophyll molecules form some kind of regular lattice in which is imbedded, in a regular way, one trap for every N chlorophyll molecules, N being in the range 250–500. A photon may be absorbed with equal probability by any chlorophyll molecule and then, by an exciton transfer mechanism, walk randomly through the lattice until it reaches a trapping center for the *first time*, at which moment it is absorbed and process (c) starts.

* This work was partially supported by the U.S. Office of Naval Research.

[1] L. N. M. Duysens, Progr. Biophys. 14, 1 (1964).

[2] R. Emerson and W. Arnold, J. Gen. Physiol. 15, 391 (1932).

The statistical problem which is then to be discussed in this paper is: Given an infinite lattice of unit cells, each cell containing N points of which $(N-1)$ are chlorophyll molecules and 1 is a trap; if an arbitrary chlorophyll molecule is excited, how many steps must be taken by the exciton in a random walk before it is trapped?

Actually, we consider the equivalent problem of a finite lattice of N points with a trap at the origin and allow the random walk to take place on the torus which is formed by connecting the opposite ends of the unit cell. Practically all the information and ideas required to discuss this problem have been given in the first two papers of this series.[3,4]

Pearlstein[5,6] and Robinson[7] have already discussed this problem to some extent for a square lattice, and Knox[8] has treated it for square and triangular lattices by direct machine inversion of the transfer matrix. Here an analytic solution for arbitrary lattices will be obtained in terms of asymptotic series in $1/N$. In some special cases in which there is an equal probability for a random walker to go to any nearest-neighboring point on a given step, we find the average number of steps $\langle n\rangle$ before trapping to be:

(a) one-dimensional chain:

$$\langle n\rangle = N(N+1)/6;$$

(b) square lattice:

$$\langle n\rangle = \pi^{-1}N\ln N + 0.195056N - 0.1170 - 0.051N^{-1} + O(N^{-2});$$

(c) simple cubic lattice:

$$\langle n\rangle = 1.5164N + O(N^{\frac{1}{2}}).$$

The two-dimensional result seems to give the best agreement with light flash experiments.[5]

2. GENERAL FORMALISM

Let $F_n(s)$ be the probability that a lattice walker which starts at the origin arrives at lattice point s *for the first time* after n steps. Also define

$$F(s,z) = \sum_{n=1}^{\infty} z^n F_n(s) \tag{1}$$

to be the generating function of the set $\{F_n(s)\}$. It can be shown that[3,4]

$$F(s,z) = [P(s,z) - \delta_{s,0}]/P(0,z), \tag{2}$$

where $P(s,z)$ is the generating function

$$P(s,z) = \sum_{n=0}^{\infty} z^n P_n(s), \tag{3}$$

with $P_n(s)$ being the probability that a walker starting from the origin arrives at s for the first time after n steps, independently of how many previous visits he already had at s.

Suppose that the origin is a trapping point and that at time $t=0$ a walker has the same probability of being at any nontrapping point on the lattice. Then the generating function for the probability that a walker will be trapped in a given number of steps is

$$G_k(z) = \frac{1}{N-1}\sum_{s\neq 0} F(s,z) = \frac{1}{N-1}\left\{\sum_s F(s,z) - F(0,z)\right\} \tag{4}$$

on a d-dimensional lattice with $m\times m\times\cdots\times m = m^d = N$ lattice points and periodic boundary conditions (i.e., walk on a d torus). $N-1$ is the number of nontrapping points. Then, from (2),

$$G_k(z) = \frac{1}{N-1}\left\{\sum \frac{[P(s,z)-\delta_{s,0}]}{P(0,z)} - 1 + \frac{1}{P(0,z)}\right\} = \frac{1}{N-1}\left\{-1 + \sum_s P(s,z)/P(0,z)\right\}. \tag{5}$$

However, since $\sum_s P_n(s) = 1$, we find

$$\sum_s P(s,z) = \sum_n z^n \sum_s P_n(s) = \sum_n z^n = 1/(1-z),$$

so that

$$G_d(z) = \frac{1}{N-1}\{[(1-z)P(0,z)]^{-1} - 1\}. \tag{6}$$

Note that $P(0,z)$ is the generating function for all walks which start and end at the origin.

The average number of steps required to reach the origin for the first time (i.e., to be trapped) is

$$\langle n\rangle = \partial G_d/\partial z|_{z=1} = \frac{1}{N-1}\frac{\partial}{\partial z}\left\{\frac{1}{(1-z)P(0,z)}\right\}_{z=1}. \tag{7}$$

This is our basic formula.

To obtain an idea of how (7) is to be applied, consider a ring of N equally spaced lattice points. Suppose that when a walker is at any of these points, the probability is $\frac{1}{2}$ that his next step will be to either one of his neighboring lattice points. Then it is easy to show that the generating function for all walks

[3] E. W. Montroll, in *Proceedings of the Symposium on Applied Mathematics, Vol. 16* (American Mathematical Society, Providence, R.I., 1964), p. 193.
[4] E. W. Montroll and G. Weiss, J. Math. Phys. **6**, 167 (1965).
[5] R. M. Pearlstein, Ph.D. dissertation, University of Maryland (1966).
[6] R. M. Pearlstein, Brookhaven Natl. Lab. Symp. **19**, 19 (1967).
[7] G. W. Robinson, Brookhaven Natl. Lab. Symp. **19**, 16 (1967).
[8] R. S. Knox, J. Theoret. Biol. **21**, 244 (1968).

which start and end at the origin is[3]

$$P(0, z) = \frac{1}{N}\sum_{k=0}^{N-1} 1/[1 - z\cos(2\pi k/N)]$$
$$= \left\{\frac{1 + x^N}{1 - x^N}\right\}\frac{1}{(1 - z^2)^{\frac{1}{2}}} \tag{8}$$

(the summation is carried out in Appendix A), where

$$x = [1 - (1 - z^2)^{\frac{1}{2}}]/z$$
$$= 1 - 2^{\frac{1}{2}}(1 - z)^{\frac{1}{2}} + (1 - z) - (\tfrac{3}{4})\sqrt{2}(1 - z)^{\frac{3}{2}}$$
$$+ (1 - z)^2 - \cdots, \tag{9}$$

so that, as $z \to 1$,

$$\frac{1}{(1 - z)P(0, z)}$$
$$= N - \frac{N(N^2 - 1)(1 - z)}{6} + O(1 - z)^2. \tag{10}$$

An immediate application of (7) with $d = 1$ then yields the following expression for the average number of steps required for a walker to be trapped:

$$\langle n\rangle = (N - 1)^{-1}N(N^2 - 1)/6$$
$$= N(N + 1)/6. \tag{11}$$

One has to work a little harder to establish similar formulas for two- and three-dimensional lattices. We consider first the class of walks in which the probability of a walker making a step represented by a given displacement vector is the same at every step. The cases of walks on two-dimensional square and triangular lattices are of this class, as is the case of a simple cubic lattice. The two-dimensional hexagonal lattice is not of this class. It corresponds to the situation in which one set of transition probabilities is appropriate for even-step numbers and another for odd-step numbers.

Let $p(s)$ be the probability that at any step a random walker makes a displacement s. Then

$$\sum_s p(s) = 1. \tag{12}$$

If one defines the "structure function" $\lambda(\theta)$ for the walk by the summation over all lattice points

$$\lambda(\theta) = \sum_s p(s)\exp i\theta \cdot s, \tag{13}$$

then, on a d-dimensional simple hypercubic lattice of $(m \times m \times \cdots \times m = m^d = N)$ lattice points, the generating function for those walks which start and end at the origin is[3,4] (the r's being integral components of a lattice vector r)

$$P(0, z) = m^{-d}\sum_{k_1=0}^{m-1}\sum_{k_2=0}^{m-1}\cdots\sum_{k_d=0}^{m-1}[1 - z\lambda(2\pi k/m)]^{-1}. \tag{14}$$

One finds for structure function of a walk on a two-dimensional square lattice in which the transition probabilities are $p(s) = 0$, except

$$p(1, 0) = p(-1, 0) = p(0, 1) = p(0, -1) = \tfrac{1}{4}, \tag{15a}$$
$$\lambda(\theta) = (e^{i\theta_1} + e^{-i\theta_1} + e^{i\theta_2} + e^{-i\theta_2})/4 = \tfrac{1}{2}(c_1 + c_2), \tag{15b}$$

where $\theta = (\theta_1, \theta_2) = \mathbf{i}\theta_2 + \mathbf{j}\theta_2$ and $c_j = \cos\theta_j$.

The case exhibited in Fig.1(a) with the weight $\frac{1}{6}$ for steps to any of the four nearest-neighbor points as well as to the identified two next-nearest neighbors (out of the possible four next nearest neighbor), i.e., $p(s) = 0$, except for

$$p(1, 0) = p(-1, 0) = p(0, 1) = p(0, -1)$$
$$= p(1, -1) = p(-1, 1) = \tfrac{1}{6}, \tag{16a}$$

yields the structure function

$$\lambda(\theta) = [\cos\theta_1 + \cos\theta_2 + \cos(\theta_1 - \theta_2)]/3. \tag{16b}$$

This example is of interest because, by shearing the square lattice of Fig. 1(a), it can be made into the triangular lattice of Fig.1(b). Since there is a one-to-one correspondence between every walk which starts and ends at the origin of the sheared lattice and every walk which starts and ends at the origin of the original lattice, the statistics of such walks on the triangular lattice which have equal probabilities of a walker stepping to any of the 6 nearest-neighbor points can be discussed in terms of walks on a square lattice with the structure function in (16b).

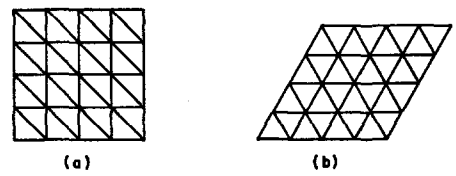


FIG. 1. See text for explanation.

Structure functions (15b) and (16b) are special cases of a more general class of random walks on a square lattice for which $p(s) = 0$, unless

$$p(1, 0) = p(-1, 0) = p(0, 1) = p(0, -1) = p, \tag{17a}$$
$$p(1, 1) = p(-1, -1) = q_1;$$
$$p(1, -1) = p(-1, 1) = q_2, \tag{17b}$$

with $4p + 2q_1 + 2q_2 = 1$. Then

$$\lambda(\theta) = 2p(\cos\theta_1 + \cos\theta_2) + 2q_1\cos(\theta_1 + \theta_2) + 2q_2\cos(\theta_1 - \theta_2). \quad (18)$$

Equation (15b) corresponds to $q_1 = q_2 = 0$ and $p = \frac{1}{4}$; (16b) corresponds to $q_1 = 0$, $q_2 = p = \frac{1}{6}$. A walk on which the probability of a jump to any second neighbor is a fraction α of that to a nearest neighbor is characterized by

$$q_1 = q_2 = \alpha p, \quad \text{so that} \quad 4p(1 + \alpha) = 1. \quad (19)$$

The structure function of a three-dimensional simple cubic lattice is

$$\lambda(\theta) = (\tfrac{1}{3})(\cos\theta_1 + \cos\theta_2 + \cos\theta_3). \quad (20)$$

We shall now employ (7) and (14) to find the average number of steps required for a walker to be trapped on a simple cubic lattice when the number of lattice points per trap is very large. From (14),

$$P(0, z) = \frac{1}{m^3}\sum_{k_1=0}^{m-1}\sum_{k_2=0}^{m-1}\sum_{k_3=0}^{m-1}[1 - \tfrac{1}{3}z(c_1 + c_2 + c_3)]^{-1} = \{m^3(1 - z)\}^{-1} + \phi(0, z). \quad (21)$$

Here

$$\phi(0, z) = m^{-3}\sum{}'[1 - z(c_1 + c_2 + c_2)/3]^{-1} \quad (22)$$

is the sum (21) with the origin $k_1 = k_2 = k_3 = 0$ omitted. The function $\phi(0, z)$ has no singularity at $z = 1$; indeed, as $m \to \infty$, we set $2\pi k/m = \theta$ and $2\pi/m = d\theta$ to obtain

$$\begin{aligned}\phi(0, 1) &\to \frac{1}{(2\pi)^3}\int_0^{2\pi}\!\!\int\!\!\int[1 - \tfrac{1}{3}(\cos\theta_1 + \cos\theta_2 + \cos\theta_3)]^{-1}\,d\theta_1\,d\theta_2\,d\theta_3 \\ &= \frac{4}{3\pi^2}[18 + 12\sqrt{2} + 10\sqrt{3} - 7\sqrt{6}] \times K^2[(2 - \sqrt{3})(\sqrt{3} - \sqrt{2})] \\ &= 1.51638\ 60591, \end{aligned} \quad (23)$$

where $K(k)$ is the complete elliptic integral of the first kind. This integral was first calculated by Watson.[9] Then, from (7),

$$\begin{aligned}\langle n\rangle &= \lim_{z\to 1}\frac{\partial}{\partial z}\left\{\frac{1}{m^3(1 - z)P(0, z)}\right\} \\ &= \lim_{z\to 1}\frac{\partial}{\partial z}\{1 - m^3(1 - z)\phi(0, 1) + O(1 - z)^{\frac{3}{2}}\} \\ &\sim N\phi(0, 1) = 1.51638\ 60591N + O(N^{+\frac{1}{2}}), \end{aligned} \quad (24)$$

where $N = m^3$, the total number of lattice points per trap.

[9] G. N. Watson, Quart. J. Math. Oxford, **10**, 266 (1939).

It is interesting to note that this result for the average number of steps required for trapping could have been deduced from the formula[3,4]

$$\langle n'\rangle/N \sim \phi(0, 1) - (\tfrac{3}{2}\pi l) + O(1/l^2) \quad (25)$$

for the average number $\langle n'\rangle$ of steps required by a random walker who starts from the origin to reach the point (l_1, l_2, l_3) [with $l = (l_1^2 + l_2^2 + l_3^2)^{\frac{1}{2}}$] for the first time on a simple cubic lattice with N lattice points. A remarkable property of this formula is that, when $l > 25$, the ratio of average number of steps required to go to a given lattice point for the first time to the total number of lattice points becomes independent of the location of the lattice point. As the lattice becomes larger and larger, this applies to a larger fraction of lattice points. Now the average number of steps to go from the origin to l for the first time is the same as the average number of steps to go from l to a trap at the origin if $p(s) = p(-s)$ for all s [$p(s)$ being defined above Eq. (12)]. Hence, if we average over all lattice points where a walker might have started on his way to a trap at the origin, as $N \to \infty$ we see that

$$\langle n\rangle \sim \langle n'\rangle \sim N\phi(0, 1),$$

which is exactly (24).

This reasoning can also be employed to find $\langle n\rangle$ for a two-dimensional lattice. The formula[3,4] which corresponds to (25) for the average number of steps required to reach (l_1, l_2) for the first time is

$$\langle n'\rangle \sim (N/\pi\sigma_1\sigma_2)\log\lambda, \quad (26a)$$

where

$$\lambda^2 = (l_1^2/\sigma_1^2) + (l_2^2/\sigma_2^2) \quad (26b)$$

and

$$\sigma_j^2 = \sum s_j^2 p(s)$$

and it is postulated that

$$\sum s_1 s_2 p(s) = 0.$$

This vanishing correlation postulate is not satisfied for the case (17) when $q_1 \neq q_2$ for

$$\sum s_1 s_2 p(s) = q_1 - q_2. \quad (26c)$$

For the present we limit our discussion to the case $\sigma_1^2 = \sigma_2^2$. For square lattice walks with symmetrical nearest-neighbor and next-nearest-neighbor step probabilities, $q_1 = q_2 = q$ and

$$\sigma_1^2 = \sigma_2^2 = 2p + 4q = (1 - 2p),$$

so that

$$\frac{\langle n'\rangle}{N} \sim \frac{\log[(l_1^2 + l_2^2)^{\frac{1}{2}}/(1 - 2p)]}{\pi(1 - 2p)}. \quad (27)$$

This formula is good except for those values of (l_1, l_2) close to the origin. However, as $N \to \infty$ this

represents a vanishingly small fraction of the lattice points. Now $l \equiv (l_1^2 + l_2^2)^{\frac{1}{2}}$ on the average is some fraction, say f of $m = \sqrt{N}$; i.e., an average lattice point is some fraction of the distance to the boundary of the unit cell and f is not too far from $\frac{1}{2}$. However, since this average $l = fm = f(N)^{\frac{1}{2}}$ appears in the logarithm in (27), as $N \to \infty$ the result is insensitive to f and we again find that as $N \to \infty$

$$\langle n \rangle \sim \langle n' \rangle \sim \frac{N \log [(fm)/(1-2p)]}{\pi(1-2p)} = \frac{N \log N}{2\pi(1-2p)} + O(N). \quad (28)$$

When the jumps are to nearest-neighbor points only, $p = \frac{1}{4}$ and

$$\langle n \rangle \sim \frac{1}{\pi} N \log N + O(N). \quad (29)$$

The above reasoning is not applicable to the one-dimensional case since the formula $\langle n' \rangle = s(N-s)$ depends sensitively on the lattice point s to which the walker is going (from which it starts in the case of the walk which starts at s and ends at the origin). All one can say without the detailed averaging process which was carried out in the derivation of the exact formula (11), $\langle n \rangle = N(N+1)/6$, is that an average s is $O(N)$, so that $\langle n \rangle \sim \langle n' \rangle = O(N^2)$.

While Eq. (29) is good for very large N, we see that when $N \sim 250$, which is the range of interest in the photosynthesis application, $\log N$ is not so large that the corrections of $O(N)$ can be neglected. Their calculation is a straightforward application of (7) but requires good estimates of $\phi(0, 1)$ for not too large values of N. This is discussed in detail in Appendix B, whose results can be summarized as follows: as $z \to 1$,

$$P(0, z) = \frac{1}{N(1-z)} + \{c_1 \log N + c_2 + c_3/N + c_4/N^2 + \cdots\} + O(1-z)^{\frac{1}{2}}. \quad (30)$$

Hence, from (7),

$$\langle n \rangle = \{c_1 N \log N + c_2 N + c_3 + c_4/N + \cdots\} N/(N-1), \quad (31)$$

where the constants $c_1, c_2, \cdots$ depend on the lattice and are listed in Eq. (B38) of Appendix B. The first one, c_1, has the value

$$c_1 = \left(\frac{1}{4\pi}\right)\{(p + 2q_1)(p + 2q_2)\}^{-\frac{1}{2}}, \quad (32)$$

where the q's and p are discussed in Eqs. (17–19). In particular, on a square lattice and on a triangular lattice on which only nearest-neighbor jumps are allowed and are given equal weight,

$$c_1 = \begin{cases} 1/\pi, & \text{square lattice;} \\ \sqrt{3}/(2\pi), & \text{triangular lattice.} \end{cases} \quad (33)$$

The numerical analysis of $\langle n \rangle$ for the complete range of N going from 4 to ∞ is presented in the next section.

While (32) is not valid for a hexagonal lattice, (31) still is. It is shown in Sec. 4 that for a hexagonal lattice

$$c_1 = 3^{\frac{3}{2}}/4\pi. \quad (33')$$

3. NUMERICAL RESULTS FOR SQUARE AND TRIANGULAR LATTICES

The constants c_2, c_3, and c_4 for square lattices are [see Eq. (B38) Appendix B, with $r = 1$]:

$$c_2 = (6-12p)^{-1} + \frac{2}{\pi}[\gamma + \log(2/\pi)] - \pi^{-1}\log(1+\eta) + \frac{4}{\pi}[e^{-2\pi} + \tfrac{3}{2}e^{-4\pi} + \tfrac{4}{3}e^{-6\pi} + \cdots], \quad (34)$$

$$c_3 = \tfrac{1}{2}(2-4p)^{-1}\{(7-36p)/6 + \pi(3\eta - 1)/36 + (4\pi/3)e^{-2\pi}(1 + 3e^{-2\pi} + 4e^{-4\pi} + \cdots) - 4\pi(5p-1)(1 + 6e^{-2\pi} + 12e^{-4\pi} + \cdots)/(1-2p)\}, \quad (35)$$

$$c_4 = \frac{7\pi^3}{21600} + \frac{11\pi^3}{1440} + \frac{9\pi^3}{4}\{(\tfrac{43}{10} - 5\pi)e^{-2\pi} + (\tfrac{387}{10} - 90\pi + 4\pi^2)e^{-4\pi} + (\tfrac{602}{5} - 420\pi + 12\pi^2)e^{-6\pi} + \cdots\}, \quad (36)$$

$$\eta = (6p-1)/(1-2p), \quad (37)$$

$$\gamma = \lim\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n\right) = 0.577\,215\,665.$$

Numerically, when only steps to nearest-neighbor points are allowed and all of these are given equal weight, i.e., $p = \frac{1}{4}$ and $\eta = r = 1$, we have

$$c_1 = 0.318\,309\,886, \quad c_3 = -0.1169\,6481,$$
$$c_2 = 0.195\,056\,166, \quad c_4 = -0.0514\,5650. \quad (38)$$

The accuracy of Eq. (31) as a function of N (using four terms) for this simple square lattice case can be investigated by comparing results based on (31) with exact results for $N = 4$, 9, 16, 36, 64, and 144.

The input for these exact results is the list of cosines:

$$\cos 0° = -\cos 180° = 1; \quad \cos 90° = \cos 270° = 0,$$
$$\cos 30° = \cos 330° = -\cos 150° = -\cos 210° = \tfrac{1}{2}\sqrt{3},$$
$$\cos 45° = \cos 315° = -\cos 135° = -\cos 225° = \tfrac{1}{2}\sqrt{2},$$
$$\cos 60° = \cos 300° = -\cos 120° = -\cos 240° = \tfrac{1}{2}.$$

Since $\cos 15° = \sqrt{2}(\sqrt{3}+1)/4$ and $\cos 75° = \sqrt{2} \times (\sqrt{3}-1)/4$, the functions of all the angles required for $N = 576$ are available in a form from which an exact number for $\langle n \rangle$ can be derived. The manner in which this list is employed for the values of N given above can be seen by considering the case $m = 4$, i.e., $N = 16$. The required function $P(0, z)$ is

$$P(0,z) = \tfrac{1}{16}\sum_{r_1=0}^{3}\sum_{r_2=0}^{3}\{1-(Z/2)[\cos(2\pi r_1/4)+\cos(2\pi r_2/4)]\}^{-1}. \tag{39}$$

The relevent angles in this case are $\theta = 2\pi r/4 = 0$, 90°, 180°, and 270°, which result successively by letting $r = 0, 1, 2, 3$. The values of the sum of the two cosines which correspond to all pairs (r_1, r_2) are exhibited in the table below

r_1/r_2	0	1	2	3
0	2	1	0	1
1	1	0	−1	0
2	0	−1	−2	−1
3	1	0	−1	0

from which it is clear that

$$P(0,z) = \frac{1}{16}\left\{\frac{1}{1-z}+\frac{4}{1-\frac{1}{2}z}+\frac{4}{1+\frac{1}{2}z}+\frac{1}{1+z}+6\right\}$$
$$= \frac{1}{16}\left\{\frac{1}{1-z}+\frac{103}{6}+O(1-z)\right\}.$$

Hence

$$1/(1-z)P(0,z) = 16\{1 - 103\,(1-z)/6 + O(1-z)^2\},$$

and from (7)

$$\langle n\rangle = \frac{16}{15}\frac{\partial}{\partial z}\{1-103(1-z)/6+O(1-z)^2\}_{z=1}$$
$$= \tfrac{16}{15}\left(\tfrac{103}{6}\right) = (17.16666\cdots)\tfrac{16}{15}.$$

The cases $m = 3$, 6, 8, and 12 (i.e., $N = 9$, 36, 64, 144) follow in the same manner. The results are summarized in Table I. Notice that, even when N is as small as 4, the error resulting from the use of Eq. (31) is only 8 parts in 250, while when $N = 144$, which is approaching the range of practical interest, it is only one part in 50 000.

TABLE I. Exact values of $\langle n \rangle$ compared with those based on Eq. (31) for various numbers N of lattice points per trap on a square lattice. $\langle n \rangle$ is the average number of steps before a walker is trapped when only jumps to nearest-neighbor lattice points are possible and all are given equal weight.

N	$\langle n\rangle(N-1)/N$ (exact)	Based on Eq. (31)
4	5/2 = 2.50	2.42
9	8 = 8.00	7.92
16	103/6 = 17.17	17.12
36	3,359/70 = 47.98	47.97
64	69,329/714 = 97.099	97.091
144	7,680,923/30,030 = 255.755	255.770

We can proceed in the same way to discuss walks on a triangular lattice for which at each step the walker has an equal probability of stepping to any of the six nearest-neighbor points. In this case $p = q_2 = \frac{1}{6}$ and $q_1 = 0$. Also $r = \frac{1}{2}\sqrt{3}$ and $\eta = \frac{1}{3}$. Then, from (B38) and (B33), if we let $\alpha = \pi\sqrt{3}$, we have

$$c_1 = \tfrac{1}{2}\sqrt{3} = 0.275\,664\,448, \tag{40a}$$

$$c_2 = \tfrac{1}{4} + \sqrt{3}(\gamma - \log\pi + \tfrac{1}{2}\log 3)/\pi - (2/\pi)\sqrt{3}(e^{-\alpha} - \tfrac{3}{2}e^{-2\alpha} + \tfrac{4}{3}e^{-3\alpha} - \cdots)$$
$$= 0.235\,214\,021, \tag{40b}$$

$$c_3 = -\frac{1}{4} - \frac{2\sqrt{3}}{9}(e^{-\alpha} - 3e^{-2\alpha} + 4e^{-3\alpha}\cdots)$$
$$= -0.251\,407\,596. \tag{40c}$$

Exact results for triangular lattices are given in Table II. These are to be compared with the numbers in the last column which are obtained from Eq. (31) with the c_j's given by (40).

4. LATTICES WITH MORE THAN ONE LATTICE POINT PER UNIT CELL

The derivation of our basic equation (7) has to be extended somewhat to cover lattices which have more than one lattice point per unit cell. The special case of interest here will be the hexagonal lattice, but we can go rather far with the general problem.

TABLE II. Exact values of $\langle n \rangle$ for triangular lattice compared with those based on Eq. (31). Walker steps to nearest-neighbor points only, and all six possible steps are given some weight, $\frac{1}{6}$.

N	$\langle n\rangle(N-1)/N$ (exact)	Based on Eq. (31)
4	9/4 = 2.25	2.22
9	22/3 = 7.33	7.32
16	63/4 = 15.75	15.74
36	2627/60 = 43.78	43.78
64	2469/28 = 88.179	88.176
144	318,962,951/1,381,380 = 230.902	230.900

The unit cells will be identified by the vectors $\{s\}$, which are the same as those used earlier in this paper. The numbers $\{j\}$ will be used to identify various points in a given unit cell.

The first statistical question to be considered is: Let a random walker be at point (j, s) initially; what is the probability that after n steps he will be at (j', s') when the various transition probabilities are the same at each step? We define

$$P_n(j', s'; j, s)$$

as the required probability. When there are two lattice points per unit cell, it is convenient to introduce the matrix

$$P_n(s', s) = \begin{bmatrix} P_n(1, s'; 1, s) & P_n(1, s'; 2, s) \\ P_n(2, s'; 1, s) & P_n(2, s'; 2, s) \end{bmatrix}, \quad (41)$$

the generalization to cases with more than two lattice points per unit cell being obvious. The transition probabilities which will apply at every step are $\{p_{jk}(s)\}$; the exhibited member of the set represents the probability that a walker at the kth lattice point in an arbitrary unit cell will on his next step arrive at the jth lattice point of a unit cell which is displaced by the vector s from the originally occupied unit cell. Clearly,

$$\sum_{j,s} p_{j,k}(s) = 1 \quad (42)$$

for all k since the walker arrives with certainty at some point after his step is taken. Also,

$$\sum_{j',s'} P_n(j', s'; j, s) = 1 \quad (43)$$

for all $n \geq 0$ and (j, s), and

$$P_0(j', s'; j, s) = \delta_{jj'}\delta_{ss'}. \quad (44)$$

Moreover,

$$P_{n+1}(s', s) = \sum_{s''} p(s' - s'')P_n(s'', s), \quad (45)$$

where

$$p(s) \equiv \begin{pmatrix} p_{11}(s) & p_{12}(s) \\ p_{21}(s) & p_{22}(s) \end{pmatrix}. \quad (46)$$

It is useful to construct the generating function

$$P(s', s; z) = \sum_{n=0} z^n P_n(s', s), \quad (47)$$

so that from (45)

$$P(s', s; z) = P_0(s', s) + z \sum_{s''} p(s' - s'')P(s'', s; z). \quad (48)$$

This equation can be solved for $P(s', s; z)$ in terms of a "known" matrix

$$\Lambda(\theta) \equiv \sum_s p(s)e^{i\theta\cdot s} \equiv \begin{pmatrix} \lambda_{11}(\theta) & \lambda_{12}(\theta) \\ \lambda_{21}(\theta) & \lambda_{22}(\theta) \end{pmatrix}, \quad (49a)$$

where

$$\lambda_{jk}(\theta) = \sum_s p_{jk}e^{i\theta\cdot s}. \quad (49b)$$

Then, from (42),

$$\sum_j \lambda_{jk}(0) = 1 \quad (50)$$

for $k = 1, 2$.

The mechanics of effecting the solution of (48) is provided through the function

$$U(2\pi r/m, z) \equiv \sum_{s_1'=0}^{m_1-1} \sum_{s_2'=0}^{m_2-1} P(s', s; z) \times \exp\{2\pi i(r_1 s_1'/m_1 + r_2 s_2'/m_2)\}. \quad (51)$$

When one employs periodic boundary conditions on a lattice with $m_1 \times m_2$ unit cells (m_1 in the horizontal and m_2 in the vertical directions), we have used the notation

$$r/m \equiv (r_1/m_1, r_2/m_2). \quad (52)$$

By multiplying (48) by $\exp(2\pi i r_1 s'/m)$ and summing over all unit cells s_1', we find

$$U(2\pi r/m, z) = z\Lambda(2\pi r/m)U(2\pi r/m, z) + \sum_{s'} P_0(s', s)e^{2\pi i r\cdot s'/m},$$

so that

$$U(2\pi r/m, z) = [1 - z\Lambda(2\pi r/m)]^{-1}\sum_{s''} P_0(s'', s)e^{2\pi i r\cdot s''/m}. \quad (53)$$

Then, by applying the standard Fourier inversion formula to the combination of (51) and (53), we find

$$P(s', s; z) = \frac{1}{m_1 m_2} \sum_r [I - z\Lambda(2\pi r/m)]^{-1} \times \sum_{s''} P_0(s'', s)e^{2\pi i r\cdot(s''-s')/m}. \quad (54)$$

The generalization to the case of more than two points per unit cell is again obvious. In the two points per unit cell case, one finds that

$$[I - z\Lambda(\theta)]^{-1} = [d(\theta, z)]^{-1}\begin{pmatrix} 1 - z\lambda_{22}(\theta) & z\lambda_{12}(\theta) \\ z\lambda_{21}(\theta) & 1 - z\lambda_{11}(\theta) \end{pmatrix}, \quad (55a)$$

where

$$d(\theta, z) \equiv [1 - z\lambda_{11}(\theta)][1 - z\lambda_{22}(\theta)] - z^2\lambda_{12}(\theta)\lambda_{21}(\theta). \quad (55b)$$

We are, of course, mainly interested in the probability $F_n(j', s'; j, s)$ of a walker starting at (j, s) and arriving at (j', s') for the *first time* after n steps. These F's are related to the P's through the relation [when $(j, s) \neq (j', s')$]

$$P_n(j', s'; j, s) = \sum_{k=1}^{n} P_{n-k}(j', s'; j', s')F_k(j', s'; j, s), \quad (56)$$

which is a consequence of the fact that the independent way of going from (j, s) to (j', s') in n steps is to go to (j', s') for the first time in k steps and then to return again to (j', s') in the remaining $(n - k)$ steps for $k = 1, 2, \cdots, n$. Now multiply both sides of (56) by $z^n \equiv z^{n-k}z^k$ and sum from $n = 1$ to ∞. Then, if we define the generating function

$$F(j', s'; j, s; z) = \sum_{n=1}^{\infty} z^n F_n(j', s'; j, s), \quad (57a)$$

we see that

$$f(j', s'; j, s; z) = P(j', s'; j, s; z)/P(j', s'; j', s'; z). \quad (57b)$$

The average number of steps required to go from (j, s) to (j', s') is

$$\langle n(j', s'; j, s)\rangle = \frac{\partial}{\partial z}\left\{\frac{P(j', s'; j, s; z)}{P(j', s'; j', s'; z)}\right\}_{z=1}. \quad (58)$$

From this we can immediately derive the generalization of the basic formula (7).

Let us suppose that initially our walker has the same probability, $1/(N - 1)$, of being at any of the $(N - 1)$ nontrapping points. Then the average number of steps required to go to (j', s') for the first time is [a prime in summation indicates that terms with $(j', s') \equiv (j, s)$ are omitted]

$$\langle n\rangle = \frac{1}{(N-1)}\sum_{j,s}'\frac{\partial}{\partial z}\left\{\frac{P(j', s'; j, s; z)}{P(j', s'; j', s'; z)}\right\}_{z=1} = \frac{1}{(N-1)}\frac{\partial}{\partial z}\sum_{j,s}'\left\{\frac{P(j', s'; j, s; z)}{P(j', s'; j', s'; z)} - 1\right\}_{z=1}. \quad (59)$$

We choose the point (j', s'), the location of the trap, to be $(1, 0)$. Then, from (54) and (55),

$$\sum_{j,s} P(1, 0; j, s; z) = \frac{1}{m_1 m_2}\sum_r\sum_s\left\{\frac{1 - z\lambda_{22}(2\pi r/m) + z\lambda_{12}(2\pi r/m)}{d(2\pi r/m, z)}\right\}e^{2\pi i r\cdot s/m} = \frac{1 - z\lambda_{22}(0) + z\lambda_{12}(0)}{[1 - z\lambda_{11}(0)][1 - z\lambda_{22}(0)] - z^2\lambda_{12}(0)\lambda_{21}(0)} \quad (60)$$

and, letting $\theta \equiv 2\pi r/m$,

$$P(1, 0; 1, 0; z) = \frac{1}{m_1 m_2}\sum_r\frac{1 - z\lambda_{22}(\theta)}{[1 - z\lambda_{11}(\theta)][1 - z\lambda_{22}(\theta)] - z^2\lambda_{12}(\theta)\lambda_{21}(\theta)}. \quad (61)$$

Hence, to calculate $\langle n\rangle$, we substitute (60) and (61) into

$$\langle n\rangle = \frac{1}{N-1}\frac{\partial}{\partial z}\sum_{j,s}'\left\{\frac{P(1, 0; j, s; z)}{P(1, 0; 1, 0; z)}\right\}_{z=1}. \quad (62)$$

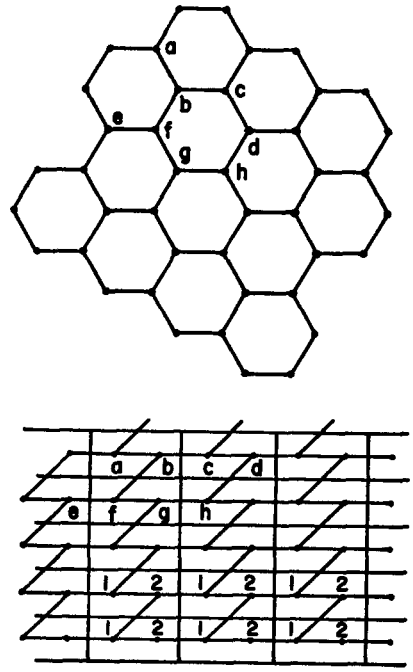


FIG. 2. See text for explanation.

Now let us analyze the special case of the hexagonal lattice on which our walker steps to one of its three nearest-neighbor points only, with probability $\frac{1}{3}$. The equivalence between a hexagonal lattice and a square lattice with two lattice points per unit cell is indicated in Fig. 2. Notice that the hexagonal lattice can be deformed and put on a square lattice so that there is a one-to-one correspondence between any path connecting two points on the hexagonal lattice and the corresponding path on the square lattice. Hence, any lattice statistics problem on the hexagonal lattice can be translated into one on the square lattice.

It is clear from Fig. 2 that the only nonvanishing transition probabilities $p_{jk}(s) \equiv p_{jk}(s_1, s_2)$ [as defined above Eq. (42)] are

$$p_{21}(0, 0) = p_{21}(0, 1) = p_{21}(-1, 0) = \tfrac{1}{3}, \quad (63a)$$

$$p_{12}(0, 0) = p_{12}(0, -1) = p_{12}(1, 0) = \tfrac{1}{3}, \quad (63b)$$

so that the structure matrix (49) has the matrix elements

$$\lambda_{11}(\theta) = \lambda_{22}(\theta) = 0, \quad (64a)$$

$$\lambda_{12}(2\pi k_1/m_1, 2\pi k_2/m_2) = \tfrac{1}{3}\{1 + \exp(-2\pi i k_2/m_2) + \exp(2\pi i k_1/m_1)\}, \quad (64b)$$

$$\lambda_{21}(2\pi k_1/m, 2\pi k_2/m_2) = \tfrac{1}{3}\{1 + \exp(2\pi i k_2/m_2) + \exp(-2\pi i k_1/m_1)\}, \quad (64c)$$

the matrix $\Lambda(\theta)$ being Hermitian. Then

$$\lambda_{12}\lambda_{21} = \tfrac{1}{9}\{3 + 2\cos\theta_1 + 2\cos\theta_2 + 2\cos(\theta_1 + \theta_2)\} = \tfrac{1}{9}\{3 + 2c_1 + 2c_2 + 2c_1c_2 - 2s_1s_2\}, \quad (65)$$

where, as usual,

$$c_j = \cos(2\pi k_j/m_j) \quad \text{and} \quad s_j = \sin(2\pi k_j/m_j). \tag{66}$$

On this basis $\lambda_{12}\lambda_{21} \to 1$ as $\theta_1, \theta_2 \to 0$, so that expression (60) becomes $1/(1-z)$ and (62) has the form

$$\langle n \rangle = \frac{1}{N-1}\frac{\partial}{\partial z}\left\{\frac{1}{(1-z)P(1,0;1,0;z)}\right\} \tag{67a}$$

with

$$P(1,0;1,0;z) = \frac{1}{m_1 m_2}\sum_{k_1=0}^{m_1-1}\sum_{k_2=0}^{m_2-1}[1-(z^2/9) \times \{3 + 2c_1 + 2c_2 + 2c_1c_2 - 2s_1s_2\}]^{-1}, \tag{67b}$$

which is the direct generalization of the basic formula (7). An alternative form for $P(1,0;1,0;z)$ is

$$P(1,0;1,0;z) = \frac{1}{m_1 m_2(1-\frac{1}{3}z^2)} \times \sum_{k_1=0}^{m_1-1}\sum_{k_2=0}^{m_2-1}\left\{1 - \frac{z_1}{6}(2c_1 + 2c_2 + 2c_1c_2 - 2s_1s_2)\right\}^{-1}, \tag{68a}$$

where

$$z_1 = 2z^2/(3-z^2).$$

The double sum is exactly the same form as that discussed in Appendix B when $m_1 = m_2 = m$. The general case can be discussed in essentially the same manner as this special one. We restrict ourselves here to the special case so that the results of Appendix B can be applied directly. When the total number of lattice point per trap $N - 1 = 2m_1m_2 - 1$ is large, the results should become independent of m_1/m_2 as long as $m_1 = O(N)$ and $m_2 = O(N)$. Hence the constant c_1 in (31) is correct. c_2 may require a correction, but not by a large factor, while c_3 would undoubtedly have to be changed.

In terms of the p's and q's of Eq. (B2) in Appendix B, $p = \frac{1}{6}$, $q_1 = \frac{1}{6}$, and $q_2 = 0$. Also, $2m_1m_2 = N = 2m^2$. The quantity

$$(1-\tfrac{1}{3}z^2)P(1,0;1,0;z)$$

has the same value $P(0,z)$ for the triangular lattice if one replaces m^2 by $N/2$ and z in $P(0,z)$ by z'. On this basis we find

$$\begin{aligned}(1-\tfrac{1}{3}z^2)P(1,0;1,0;z) &= [m^2(1-z_1)]^{-1} \\ &\quad + [(\sqrt{3}/2\pi)\log m^2 + c_2 + c_3/m^2 + \cdots] \\ &\quad + O(1-z_1),\end{aligned} \tag{69}$$

where the c_2 and c_3 have the same values as those appropriate for the triangular lattice (40b) and (40c):

$$c_2 = 0.235\,214\,021; \quad c_3 = -0.2514\,0796. \tag{70}$$

On this basis,

$$\begin{aligned}P(1,0;1,0;z) &= \frac{1}{N(1-z)} + \frac{1}{2N} + \frac{3}{2}\left\{\frac{\sqrt{3}}{2\pi}\log N \right. \\ &\quad \left. - \frac{\sqrt{3}}{2\pi}\log 2 + c_2 + \frac{2c_3}{N} + \cdots\right\} + O(1-z) \\ &= [N(1-z)]^{-1} + \{c_1^H \log N + c_2^H + c_3^H/N + \cdots\} \\ &\quad + O(1-z),\end{aligned} \tag{71}$$

where

$$\begin{aligned}c_1^H &= 3^{\frac{3}{2}}/2\pi = 0.413\,496\,672, \\ c_3^H &= 3c_3 + \tfrac{1}{2} = -0.25422\,27888, \\ c_2^H &= \tfrac{3}{2}[c_2 - (\sqrt{3}/2\pi)\log 2] = 0.0662\,0698.\end{aligned} \tag{72}$$

If (71) is substituted into (7), on a hexagonal lattice we find that

$$\langle n \rangle = \{c_1^H N \log N + c_2^H N + c_3^H + \cdots\}N/(N-1),$$

where the constants are given by (72).

ACKNOWLEDGMENTS

The author wishes to thank Dr. R. Pearlstein, Professor R. Knox and Professor L. Duysens for a number of informative discussions. He also wishes to thank Mr. F. T. Lee for checking many of the calculations in this paper and for finding several numerical errors.

APPENDIX A. EVALUATION OF $S(w_1, w_2)$

We define

$$S(w_1, w_2) = \frac{1}{m}\sum_{k=0}^{m-1}\{1 - w_1\cos(2\pi k/m) - w_2\sin(2\pi k/m)\}^{-1}. \tag{A1}$$

Let

$$w_1 + iw_2 = \rho e^{i\phi} \tag{A2}$$

and assume that $0 < \rho < 1$. Clearly,

$$w_1 = \rho\cos\phi, \quad w_2 = \rho\sin\phi, \quad \text{and} \quad \rho^2 = w_1^2 + w_2^2. \tag{A3}$$

Hence, if we let $\zeta(k) = (2\pi k/m) - \phi$,

$$\begin{aligned}S &= \frac{1}{m}\sum_{k=0}^{m-1}\{1 - \rho\cos\zeta(k)\}^{-1} \\ &= -\frac{2}{\rho m}\sum_{k=0}^{m-1} e^{i\zeta(k)}/\{(e^{i\zeta(k)} - x)(e^{i\zeta(k)} - x^{-1})\},\end{aligned} \tag{A4}$$

where x is the smaller root of the equation

$$x^2 - 2x/\rho + 1 = 0; \tag{A5}$$

i.e., since we postulated $0 < \rho < 1$,

$$0 < x = \{1 - (1 - \rho^2)^{\frac{1}{2}}\}/\rho < 1. \tag{A6}$$

The summand of S can be rewritten using partial fractions:

$$S = -\frac{2}{\rho m(x - x^{-1})}\left\{\sum_{k=0}^{m-1}(1 - xe^{i\phi}e^{-2\pi ik/m})^{-1} + \sum_{k=0}^{m-1} xe^{-i\phi}e^{2\pi ik/m}(1 - xe^{-i\phi}e^{2\pi ik/m})^{-1}\right\}.$$

Since

$$\frac{1}{m}\sum_{k=0}^{m-1} e^{2\pi ikl/m} = \begin{cases}1 & \text{for } l = 0, \pm m, \pm 2m \cdots \text{etc.},\\ 0 & \text{for other integral values of } l,\end{cases}$$

we find that the first sum in S is

$$\frac{1}{m}\sum_{k=0}^{m-1}\{1 + xe^{i\phi}e^{-2\pi ik/m} + x^2e^{2i\phi}e^{-2\cdot 2\pi ik/m} + \cdots\}$$
$$= 1 + x^me^{i\phi m} + x^{2m}e^{2\phi m} + \cdots = 1/(1 - x^me^{i\phi m}).$$

The second sum in S is evaluated in a similar way. Finally, we find that

$$S = (1 - \rho^2)^{-\frac{1}{2}}\left\{\frac{1}{1 - x^me^{im\phi}} + \frac{x^me^{-im\phi}}{1 - x^me^{-im\phi}}\right\}$$
$$= \frac{1}{(1 - \rho^2)^{\frac{1}{2}}}\left\{\frac{1 - x^{2m}}{1 - 2x^m\cos m\phi + x^{2m}}\right\}. \tag{A7}$$

APPENDIX B. ON THE CALCULATION OF

$$P(0, z) = \frac{1}{m^2}\sum_{k_1=0}^{m-1}\sum_{k_2=0}^{m-1}[1 - z\lambda(2\pi k_1/m, 2\pi k_2/m)]^{-1}$$

We have

$$P(0, z) = \frac{1}{m^2}\sum_{k_1=0}^{m-1}\sum_{k_2=0}^{m-1}[1 - z\lambda(2\pi k_1/m, 2\pi k_2/m)]^{-1}, \tag{B1}$$

when m is large and z is close to 1 and $\lambda(\theta_1, \theta_2)$ is defined by

$$\lambda(\theta_1, \theta_2) = 2p(c_1 + c_2) + 2(q_1 + q_2)c_1c_1 + 2(q_2 - q_1)s_1s_2 \tag{B2}$$

with

$$c_j = \cos\theta_j, \quad s_j = \sin\theta_j \quad \text{and} \quad \theta_j = 2\pi k_j/m, \tag{B3}$$

while

$$4p + 2(q_1 + q_2) = 1. \tag{B4}$$

Clearly,

$$P(0, z) = \frac{1}{m}\sum_{k_1=0}^{m-1}[1 - 2pzc_1]^{-1}f(z, \theta_1), \tag{B5a}$$

where

$$f(z, \theta_1) = \frac{1}{m}\sum_{k_2=0}^{m-1}(1 - w_1c_2 - w_2c_2)^{-1} \tag{B5b}$$

with

$$w_1 = 2z[p + (q_1 + q_2)c_1]/(1 - 2pzc_1), \tag{B6a}$$
$$w_2 = 2z(q_1 - q_2)s_1/(1 - 2pzc_1). \tag{B6b}$$

The sum $f(z, \theta_1)$ is just the $S(w_1, w_2)$ evaluated in Appendix A. Incorporating (A7) into (B5),

$$P(0, z) = \frac{1}{m}\sum_{k_1=0}^{m-1}(1 - 2pzc_1)^{-1} \times \left\{\frac{1 - x^{2m}}{1 - 2x^m\cos\phi m + x^{2m}}\right\}\frac{1}{(1 - \rho^2)^{\frac{1}{2}}}, \tag{B7}$$

where x is given by (A6), $p^2 = w_1^2 + w_2^2$, and ϕ by (A2).

We first evaluate the term which corresponds to $k_1 = 0$. Then $w_2 = 0$, $l = 0$, and

$$\rho = w_1 = 2z(p + q_1 + q_2)/(1 - 2pz) = z(1 - 2p)/(1 - 2pz) \tag{B8}$$

and the $k_1 = 0$ term is

$$m^{-1}[(1 + x^m)/(1 - x^m)][(1 - 2pz)^2 - \rho^2(1 - 2pz)^2]^{-\frac{1}{2}} = m^{-1}[(1 + x^m)/(1 - x^m)]/\alpha^{-1} \times \{2[1 - 2p][1 - \beta^2(1 - 4)]\}^{-\frac{1}{2}}, \tag{B9}$$

where

$$\alpha = (1 - z)^{\frac{1}{2}} \quad \text{and} \quad \beta = \alpha[2(1 - 2p)]^{-\frac{1}{2}}. \tag{B10}$$

The quantity x was defined in Eq. (A5) to be

$$x = \{1 - (1 - \rho^2)^{\frac{1}{2}}\}/\rho = [1 - 2p + 2p\alpha^2 - \alpha\{2(1 - 2p) + \alpha^2(4p - 1)\}^{\frac{1}{2}}]/(1 - \alpha^2)(1 - 2p) = 1 - 2\beta + 2\beta^2 + (4p - 3)\beta^3 + \cdots. \tag{B11}$$

Hence,

$$x^m = 1 - 2m\beta + 2m^2\beta^2 - 4m\beta^3(m^2 - 3p + \tfrac{5}{4})/3 + \cdots, \tag{B12}$$

and our $k = 0$ term becomes

$$[m^2(1 - z)]^{-1} + (m^2 - 1)/6m^2(1 - 2p) + O(1 - z)^{\frac{1}{2}}. \tag{B13}$$

For other values of k_1 as $z \to 1$

$$(1 - 2pzc_1)(1 - \rho^2)^{\frac{1}{2}} \to \{(1 - 2pc_1)^2 - 4[p + (q_1 + q_2)c_1]^2 - 4(q_1 - q_2)^2s_1^2\}^{\frac{1}{2}} = 4[(p + 2q_1)(p + 2q_2)]^{\frac{1}{2}} \times [1 + \eta\sin^2\pi k_1/m]^{\frac{1}{2}}\sin\pi k_1/m \tag{B14}$$

with

$$\eta = \frac{p(1 - 2p)}{(p + 2q_1)(p + 2q_2)} - 1. \tag{B14'}$$

In two special cases of interest [which we shall henceforth designate as cases (a) and (b)], we have

(a) square lattice: $p = \frac{1}{4}$, $q_1 = q_2 = 0$, and $\eta = 1$;

(b) triangular lattice: $p = q_2 = \frac{1}{6}$, $q_1 = 0$, and $\eta = \frac{1}{3}$.

On this basis we combine (B7), (B13), and (B14) to find

$$P(0, z) = \frac{1}{m^2(1-z)} + \frac{1}{2(1-2p)} \times \left\{\frac{m^2-1}{3m^2} + \frac{\phi(0,1)}{r}\right\} + O(1-z)^{\frac{1}{2}}, \quad \text{(B15)}$$

where

$$\phi(0, 1) = S_1 + S_2 + S_3$$

and

$$r = \left\{\frac{4(p+2q_1)(p+2q_2)}{(1-2p)^2}\right\}^{\frac{1}{2}} \quad \text{(B16a)}$$

with

$$S_1 = \frac{1}{m}\sum_{k=1}^{m-1} 1/\sin(\pi k/m), \quad \text{(B16b)}$$

$$S_2 = \frac{1}{m}\sum_{k=1}^{m-1} \{[1 + \eta \sin^2(\pi k/m)]^{-\frac{1}{2}} - 1\}/\sin(\pi k/m), \quad \text{(B16c)}$$

$$S_3 = \frac{1}{m}\sum_{k=1}^{m-1} \frac{1}{\sin(\pi k/m)} \left\{\frac{2x^N(\cos\phi N - x^N)}{1 - 2x^N\cos N\phi + x^{2N}}\right\} \times \frac{1}{[1 + \eta\sin^2(\pi k/N)]^{\frac{1}{2}}}. \quad \text{(B16d)}$$

It will be helpful to employ the Euler–McLauren summation formula:

$$\frac{1}{m}\sum_{k=1}^{m-1} g(\pi k/m) = \frac{1}{\pi}\int_0^\pi g(\theta)\,d\theta - \frac{1}{2m}[g(0) + g(\pi)] + \frac{\pi}{12m^2}[g'(\pi) - g'(0)] - \frac{\pi^3}{720m^4}[g'''(\pi) - g'''(0)] + \cdots. \quad \text{(B17)}$$

The sum S_1 can be rewritten as

$$S_1 = \frac{2}{\pi}\left[\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{m-1}\right] + \sum_{k=1}^{m-1}\left[\frac{1}{\sin\pi k/m} - \frac{1}{\pi k/m} - \frac{1}{(m-k)\pi/m}\right]\frac{1}{m}. \quad \text{(B18)}$$

Since

$$\sum_{k=1}^{m-1} k^{-1} = \log m + \gamma - \frac{1}{2m} - \frac{1}{12m^2} + \frac{1}{120m^4} - \frac{1}{252m^6} + \cdots \quad \text{(B19)}$$

with

$$\gamma = \lim_{n\to\infty}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n\right) = 0.57721\ 56649\cdots, \quad \text{(B20)}$$

and since Eq. (B17) implies that the second sum in (B18) has the form

$$\frac{1}{\pi}\int_0^\pi \left\{\frac{1}{\sin\theta} - \frac{1}{\theta} - \frac{1}{(\pi-\theta)}\right\} d\theta + \frac{1}{\pi m} + \frac{\pi}{6m^2}\left(\frac{1}{\pi^2} - \frac{1}{6}\right) - \frac{\pi^3}{360m^4}\left(\frac{6}{\pi^4} - \frac{7}{60}\right) + \cdots = \frac{2}{\pi}\log\frac{2}{\pi} + \frac{1}{\pi m} + \frac{\pi}{6m^2}\left(\frac{1}{\pi^2} - \frac{1}{6}\right) - \frac{\pi^3}{360m^4}\left(\frac{6}{\pi^4} - \frac{7}{60}\right) + \cdots, \quad \text{(B21)}$$

we have

$$S_1 = \frac{2}{\pi}\{\log m + [\gamma + \log(2/\pi)] - \pi^2/72m^2 + 7\pi^4/43200m^4 \cdots\}. \quad \text{(B22)}$$

The sum S_2 also has an asymptotic expansion when developed according to the Euler–McLauren formula:

$$S_2 = \frac{1}{\pi}\int_0^\pi \{(1 + \eta\sin^2\theta)^{-\frac{1}{2}} - 1\}\,d\theta/\sin\theta + \pi\eta/12m^3 + \pi^3\eta(1 + \tfrac{9}{2}\eta)/720m^4 + \cdots. \quad \text{(B23)}$$

The integral has the value

$$\frac{2}{\pi}\int_0^{\pi/2}\left\{-\frac{\eta}{2}\sin\theta + \frac{1\cdot 3}{2\cdot 4}\eta^2\sin^3\theta - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\eta^3\sin^5\theta + \cdots\right\} d\theta = -\frac{1}{\pi}\left(\eta - \frac{\eta^2}{2} + \frac{\eta^3}{3} - \cdots\right) = \frac{-1}{\pi}\log(1+\eta),$$

so that

$$S_2 = -\frac{1}{\pi}\log(1+\eta) + \frac{\eta\pi}{12m^2} + \frac{\eta\pi^3}{720m^4}(1 + \tfrac{9}{2}\eta) + \cdots. \quad \text{(B24)}$$

While the sum S_3 contributes very little to $P(0, z)$, it is somewhat more involved. To calculate it, the first step is to obtain from (A3), (A5), and (B6)

$$x = \rho^{-1}[1 - (1-\rho^2)^{\frac{1}{2}}] \quad \text{with} \quad \rho^2 = w_1^2 + w_2^2. \quad \text{(B25)}$$

After a certain amount of elementary algebra, using (B14), one finds

$$x = 1 - 2rs + 2r^2s^2 - \frac{s^3}{r} \times \left\{\frac{4p(1-2r^2)}{1-2p} + r^2(4r^2-1)\right\} + \cdots, \quad \text{(B26)}$$

where $s \equiv \sin \pi k/N$ and

$$r = \left\{\frac{4(p+2q_1)(p+2q_2)}{(1-2p)^2}\right\}^{\frac{1}{2}} = \begin{cases} 1, & \text{if } q_1 = q_2 = q; \text{ i.e., } 2p + 4q = 1 - 2p, \text{ case (a)}, \\ \frac{1}{2}\sqrt{3}, & \text{if } q_1 = 0 \text{ and } q_2 = p = \frac{1}{6}, \text{ case (b)}. \end{cases} \quad \text{(B27)}$$

One finds

$$x^m \sim e^{-2\pi rk}\left\{1 + \frac{1}{m^2}\left(\frac{4\pi^3k^3}{3r}\right)\frac{(2r^2-1)}{(1-2p)} \times (3p - r^2 + 2r^2p) + O(1/m^4)\right\}. \quad \text{(B28)}$$

The other quantity needed for the calculation of S_3 is $\varphi = \sin^{-1}(w_2/\rho)$. From Eq. (B6), $w_2 \equiv 0$ in case (a) with $q_1 = q_2$. In case (b)

$$w_1 = (1 - s^2)/(1 + s^2)$$

and

$$w_2 = -s(1-s^2)^{\frac{1}{2}}(1+s^2),$$
$$\rho = (1-s)^{\frac{1}{2}}/(1+s^2) \quad \text{with} \quad s = \sin \pi k/m,$$
$$w_2/\rho = \sin(-\pi k/m), \quad \text{so that} \quad \varphi = -\pi k/m.$$

Hence the combination which appears in S_3 is

$$\frac{2x^m(\cos \varphi m - x^m)}{1 - 2x^m \cos \varphi m + x^{2m}} = \begin{cases} 2x^m/(1-x^m), & \text{case (b) with } k \text{ even}, \\ & \text{case (a) as always}; \\ -2x^m/(1+x^m), & \text{case (b) with } k \text{ odd}. \end{cases} \quad \text{(B29)}$$

We see from (B25) that when $k \simeq m/2$, s is close to 1 and c close to zero. In this range

$$x \sim \tfrac{1}{2}\cos(\pi k/m) + O(\cos^2 \pi k/m) \sim \pi(k - m/2)/m,$$

which is a very small number when m is large. Furthermore, when $m \to \infty$, x^m vanishes exponentially. Hence the main contribution to S_3 is from values of k near 1 and m. From symmetry considerations we rewrite S_3 as double the sum from $k = 1$ to $[m/2]$ and use the formula (B28) for x^m since only small integral k's contribute. Then, for case (a), since $r = 1 = \eta$, we have

$$\frac{2x^m}{1-x^m} = 2\left(\frac{e^{-2\pi k}}{1-e^{-2\pi k}}\right) \times \left\{1 + \frac{4\pi^3k^3/3}{m^2(1-e^{-2\pi k})}\left(\frac{5p-1}{1-2p}\right) + O(1/m^4)\right\}$$

$$[\sin \pi k/m]^{-1}[1 + \eta^2 \sin^2 \pi k/m]^{-\frac{1}{2}} \sim \frac{m}{\pi k}\left[1 - \frac{1}{3}\frac{\pi^2k^2}{m^2} + O(1/m^4)\right],$$

so that

$$S_3 = \frac{4}{\pi}\sum_{k=1}^{[m/2]}\frac{1}{k}\left(\frac{e^{-2\pi k}}{1-e^{-2\pi k}}\right)\left\{1 - \frac{\pi^2k^2}{3m^2} \times \left(1 - \frac{4k}{(1-e^{-2\pi k})}\left[\frac{5p-1}{1-2p}\right]\right) + O(1/m^4)\right\}$$
$$= \frac{4}{\pi}\{e^{-2\pi} + \tfrac{3}{2}e^{-4\pi} + \tfrac{4}{3}e^{-6\pi} + \cdots\} - \frac{4\pi e^{-2\pi}}{3m^2}\left\{(1 + 3e^{-2\pi} + 4e^{-4\pi} + \cdots) - \frac{4\pi(5p-1)}{(1-2p)}(1 + 6e^{-2\pi} + 12e^{-4\pi} + \cdots)\right\} + O(m^{-4}). \quad \text{(B30)}$$

The term of $O(1/m^4)$ has the following value for the case $p = \frac{1}{4}$:

$$\frac{4\pi^3}{9m^4}\sum_{k=1}^{\infty}k^3\left(\frac{e^{-2\pi k}}{1-e^{-2\pi k}}\right) \times \left[\frac{43}{10} - \frac{5k\pi}{(1-e^{-2\pi k})} + \frac{4(k\pi)^2e^{-2\pi k}}{(1-e^{-2\pi k})^2}\right]$$
$$= \frac{4\pi^3}{9m^4}\{(\tfrac{43}{10} - 5\pi)e^{-2\pi} + (\tfrac{387}{10} - 90\pi + 4\pi^2)e^{4-\pi} + (\tfrac{802}{5} - 420\pi + 12\pi^2)e^{-6\pi} + \cdots\}. \quad \text{(B31)}$$

In case (b) the combination (B29) which appears in S_3 is

$$-\frac{2e^{-2\pi rk}}{e^{-2\pi k} - (-1)^k}\left\{1 + \frac{4\pi^3k^3}{3m^2r}\frac{(2r^2-1)}{(1-2p)} \times \frac{(3p - r^2 + 2pr^2)}{[1 - (-1)^k e^{-2\pi rk}]} + O(1/m^4)\right\},$$

while

$$(\sin \pi k/m)^{-1}[1 + \eta^2 \sin^2(\pi k/m)]^{-\frac{1}{2}} \sim \frac{m}{\pi k}\left\{1 + \frac{\pi^2k^2}{2m^2}(\tfrac{1}{3} - \eta^2) + \cdots\right\},$$

so that

$$S_3 = -\frac{4}{\pi}\sum_{k=1}^{[m/2]}\frac{1}{k}\left(\frac{e^{-2r\pi k}}{e^{-2r\pi k}-(-1)^k}\right)\left\{1+\frac{\pi^2k^2}{m^2}\left(\frac{1}{6}-\frac{\eta^2}{2}+\frac{4\pi k(2r^2-1)(3p-r^2+2pr^2)}{3r(1-2p)[1-(-1)^k e^{-2\pi rk}]}\right)+O\left(\frac{1}{m^4}\right)\right\}$$

$$= -\frac{4}{\pi}(e^{-2\pi r}-\tfrac{3}{2}e^{-4\pi r}+\tfrac{4}{3}e^{-6\pi r}\cdots)$$

$$-\frac{4}{m^2}\left(\frac{1}{6}-\frac{\eta^2}{2}\right)(e^{-2\pi r}-3e^{-4\pi r}+4e^{-6\pi r}\cdots)$$

$$+\frac{4\pi^2(2r^2-1)(3p-r^2+2pr^2)}{3r(1-2p)}$$

$$\times(e^{-2\pi r}-6e^{-4\pi r}+12e^{-6\pi r}-\cdots). \quad \text{(B32)}$$

Finally, when jumps to nearest neighbors on a triangular lattice all have the same probability $p=\frac{1}{6}$, $\eta=\frac{1}{3}$, and $r=\frac{1}{2}\sqrt{3}$, we have

$$S_3 = -\frac{4}{\pi}(e^{-2\pi r}-\tfrac{3}{2}e^{-4\pi r}+\tfrac{4}{3}e^{-6\pi r}-\cdots)$$

$$-\frac{4}{9m^2}(e^{-2\pi r}-3e^{-4\pi r}+4e^{-6\pi r}-\cdots). \quad \text{(B33)}$$

In either (B30) or (B32) one can write

$$S_3 = S_3^{(0)} + m^{-2}S_3^{(1)} + m^{-4}S_3^{(2)} + \cdots. \quad \text{(B34)}$$

The results of this appendix can then be summarized as follows [see Eq. (B15)]:

$$P(0,z) = \frac{1}{N(1-z)} + \frac{1}{2(1-2p)}\times\left\{\frac{1}{3}-\frac{1}{3N}+\frac{1}{r}(S_1+S_2+S_3)\right\} + O(1-z)^{\frac{1}{2}}, \quad \text{(B35)}$$

where $N \equiv m^2$, and (B22) and (B24) are used:

$$S_1 = \frac{1}{\pi}\log N + \frac{2}{\pi}\{[\gamma+\log(2/\pi)] - \pi^2/72N + 7\pi^4/43200N^2 + \cdots\}, \quad \text{(B36a)}$$

$$S_2 = -\frac{1}{\pi}\log(1+\eta) + \eta\pi/12N + \eta\pi^3(1+\tfrac{9}{2}\eta)/720N^2 + \cdots, \quad \text{(B36b)}$$

and S_3 is given by (B30) or (B32), depending on the case.

It is more useful to rearrange these terms so that one writes

$$P(0,z) = [N(1-z)]^{-1} + c_1\log N + c_2 + c_3/N + c_4/N^2 + \cdots + O(1-z)^{\frac{1}{2}}, \quad \text{(B37)}$$

where

$$c_1 = [2\pi r(1-2p)]^{-1} = \left(\frac{1}{4\pi}\right)\{(p+2q_1)(p+2q_2)\}^{-\frac{1}{2}}, \quad \text{(B38a)}$$

$$c_2 = \frac{1}{2(1-2p)}\left\{\frac{1}{3}+\frac{1}{r\pi}[2\gamma+2\log(2/\pi) - \log(1+\eta)] + S_3^{(0)}/r\right\}, \quad \text{(B38b)}$$

$$c_3 = \frac{1}{2(1-2p)}\{-\tfrac{1}{3}+\pi(3\eta-1)/36r + S_3^{(1)}/r\}, \quad \text{(B38c)}$$

$$c_4 = \frac{1}{2r(1-2p)}\left\{\frac{7\pi^3}{21600}+\frac{\eta\pi^3}{720}(1+\tfrac{9}{2}\eta)+S_3^{(2)}\right\}, \quad \text{(B38d)}$$

and the various $S_3^{(3)}$ are given by Eqs. (B30), (B31), and (B32).

Algebraic Classification of Four-Dimensional Riemann Spaces

JAMES THOMPSON AND GLEN SCHRANK
University of California, Santa Barbara, California

(Received 1 August 1968)

An algebraic classification of four-dimensional Riemann spaces with signature -2 is given. The possible Riemann and Ricci tensors are listed and inversion formulas from one to the other are obtained. Classifications are given for the two special spaces $R_{ij} = 0$ and $R^l_k R^k_m = 0$.

I. INTRODUCTION

The algebraic classification of general Riemann spaces was discussed in some detail by Ruse.[1,2] An extensive algebraic classification of Einstein spaces according to the eigenbivectors of the Riemann tensor has been given by Petrov.[3,4] Newman[5] has used a vierbein formalism to give an algebraic and differential classification of special Einstein spaces.

In this paper, we use a vierbein formalism to give an algebraic classification for the general Riemann space. The possible Riemann and Ricci tensors are classified. The inversion relations are found which give the possible Riemann tensors corresponding to a given Ricci tensor. Newman's classification of the special Einstein space $R_{ij} = 0$ is given in terms of our basis. Finally, we classify the Ricci null space ($R^l_k R^k_m = 0$).

The vierbein formalism is a completely general technique. In our classification we have kept the vector-basis tetrad as general as possible, specializing it only for the Ricci null spaces.

II. CLASSIFICATION OF THE RIEMANN AND RICCI TENSORS

We assume a four-dimensional Riemannian metric with signature $(+---)$. Any four linearly independent vectors can serve as the basis tetrad for the space. If we distinguish between the various tetrads by the norms of the vectors, there are 15 possibilities, summarized in Table I.

The level of computational difficulty in a vierbein formalism varies approximately as the square of the last column in Table I. Thus tetrad types 4, 12, 13, 14, and 15 are the simplest, computationally, to use; particular problems might use any of the other types to advantage.

TABLE I. Possible tetrads for metric signature $(+---)$.

Norms of basis vectors	Minimum number of nonzero scalar products ($V^m_{(i)} V_{(j)m} \neq 0$) $i, j \neq 0$ (Max no. = 6)	Nonzero norms	Total of nonzero norms plus minimum number of nonzero scalar products
1. + + + +	6	4	10
2. + + + −	3	4	7
3. + + − −	1	4	5
4. + − − −	0	4	4
5. − − − −	1	4	5
6. 0 + + +	6	3	9
7. 0 0 + +	6	2	8
8. 0 0 0 +	6	1	7
9. 0 0 0 0	6	0	6
10. 0 + + −	3	3	6
11. 0 0 + −	3	2	5
12. 0 0 0 −	3	1	4
13. 0 + − −	1	3	4
14. 0 0 − −	1	2	3
15. 0 − − −	1	3	4

We denote the four linearly independent vectors by ξ_i, χ_j, ζ_i, ρ_i. Then we form the following six independent bivectors

$$Q_{ij} = \xi_{[i}\zeta_{j]}, \quad M_{ij} = \rho_{[i}\xi_{j]},$$
$$P_{ij} = \chi_{[i}\zeta_{j]}, \quad N_{ij} = \chi_{[i}\rho_{j]},$$
$$L_{ij} = \zeta_{[i}\rho_{i]}, \quad S_{ij} = \xi_{[i}\chi_{j]},$$

where

$$\xi_{[i}\zeta_{j]} \equiv \tfrac{1}{2}(\xi_i\zeta_j - \zeta_i\xi_j)$$

Since these six bivectors are independent, they span the four-dimensional bivector space. We next form all possible independent algebraically possible Riemann tensors. That is, we form all possible four-index tensors which have the algebraic properties of the Riemann tensor, namely,

$$R_{ijkl} = -R_{jikl} = -R_{ijlk} = R_{klij},$$
$$R_{ijkl} + R_{iklj} + R_{iljk} = 0.$$

Since there are 20 independent components of the Riemann tensor (in four dimensions), there are 20

[1] H. S. Ruse, Proc. Roy. Soc. (Edinburgh) **62A**, 64 (1943–44).
[2] H. S. Ruse, Proc. London Math. Soc. **50**, 75 (1944).
[3] A. Z. Petrov, Sci. Not. Kazan State Univ. **114**, 55 (1954).
[4] A. Z. Petrov, "Einsteinian Space," Thesis, Kazan State University, U.S.S.R., 1956.
[5] E. Newman, J. Math. Phys. **2**, 324 (1961).

such tensors and we label them as follows:

$$
\begin{aligned}
(R_{ijkl})_{\mathrm{I}} &= Q_{ij}Q_{kl},\\
\mathrm{II} &= M_{ij}M_{kl},\\
\mathrm{III} &= P_{ij}P_{kl},\\
\mathrm{IV} &= N_{ij}N_{kl},\\
\mathrm{V} &= L_{ij}L_{kl},\\
\mathrm{VI} &= S_{ij}S_{kl},\\
\mathrm{VII} &= Q_{ij}M_{kl} + M_{ij}Q_{kl},\\
\mathrm{VIII} &= P_{ij}N_{kl} + N_{ij}P_{kl},\\
\mathrm{IX} &= Q_{ij}L_{kl} + L_{ij}Q_{kl},\\
\mathrm{X} &= M_{ij}S_{kl} + S_{ij}M_{kl}\\
\mathrm{XI} &= Q_{ij}P_{kl} + P_{ij}Q_{kl},\\
\mathrm{XII} &= M_{ij}N_{kl} + N_{ij}M_{kl},\\
\mathrm{XIII} &= P_{ij}L_{kl} + L_{ij}P_{kl},\\
\mathrm{XIV} &= N_{ij}S_{kl} + S_{ij}N_{kl},\\
\mathrm{XV} &= Q_{ij}S_{kl} + S_{ij}Q_{kl},\\
\mathrm{XVI} &= M_{ij}L_{kl} + L_{ij}M_{kl},\\
\mathrm{XVII} &= P_{ij}S_{kl} + S_{ij}P_{kl},\\
\mathrm{XVIII} &= N_{ij}L_{kl} + L_{ij}N_{kl},\\
\mathrm{XIX} &= Q_{ij}N_{kl} + N_{ij}Q_{kl} + L_{ij}S_{kl} + S_{ij}L_{kl},\\
\mathrm{XX} &= M_{ij}P_{kl} + P_{ij}M_{kl} + S_{ij}L_{kl} + L_{ij}S_{kl}.
\end{aligned}
$$

The above set of 20 tensors span the space of algebraically possible Riemann tensors. Any Riemann tensor can be written in the form

$$R_{ijkl} = \sum_{m=\mathrm{I}}^{\mathrm{XX}} C_m (R_{ijkl})_m .$$

The expansion coefficients C_m are, in general, functions of position.

Next, we form the basis for the algebraically possible Ricci tensors. Since the only algebraic property of the Ricci tensor is $R_{ij} = R_{ji}$, the basis spans the space of all symmetric second-rank tensors. A symmetric second-rank tensor in four dimensions has 10 independent components; thus there are 10 independent basis tensors and we label them as follows:

$$
\begin{aligned}
(R_{ij})_{\mathrm{I}} &= \xi_i\xi_j, & (R_{ij})_{\mathrm{II}} &= \chi_i\chi_j,\\
\mathrm{III} &= \zeta_i\zeta_j, & \mathrm{IV} &= \rho_i\rho_j,\\
\mathrm{V} &= \xi_{(i}\chi_{j)}, & \mathrm{VI} &= \xi_{(i}\zeta_{j)},\\
\mathrm{VII} &= \xi_{(i}\rho_{j)}, & \mathrm{VIII} &= \chi_{(i}\zeta_{j)},\\
\mathrm{IX} &= \chi_{(i}\rho_{j)}, & \mathrm{X} &= \zeta_{(i}\rho_{j)},
\end{aligned}
$$

where

$$\xi_{(i}\zeta_{j)} \equiv \tfrac{1}{2}(\xi_i\zeta_j + \zeta_i\xi_j).$$

III. INVERSION FORMULAS

We introduce the notation

$$
\begin{aligned}
(R^i_{jki})_{\mathrm{I}} &\equiv \mathrm{I}_c,\\
(R_{jk})_{\mathrm{I}} &\equiv \mathrm{I},\\
\xi_i\xi^i \equiv \xi^2 \quad \xi_i\zeta^i &\equiv \xi\cdot\zeta.
\end{aligned}
$$

Then, the contractions of the 20 algebraically possible Riemann basis tensors in terms of the above 10 algebraically possible Ricci basis tensors are

$$
\begin{aligned}
(R^i_{jki})_{\mathrm{I}} \equiv \mathrm{I}_c &= \tfrac{1}{2}\xi\cdot\zeta(R_{jk})_{\mathrm{VI}} - \tfrac{1}{4}\xi^2(R_{jk})_{\mathrm{III}} - \tfrac{1}{4}\zeta^2(R_{jk})_{\mathrm{I}}\\
&\equiv \tfrac{1}{2}\xi\cdot\zeta\mathrm{VI} - \tfrac{1}{4}\xi^2\mathrm{III} - \tfrac{1}{4}\zeta^2\mathrm{I},\\
\mathrm{II}_c &= \tfrac{1}{2}\rho\cdot\xi\mathrm{VII} - \tfrac{1}{4}\rho^2\mathrm{I} - \tfrac{1}{4}\xi^2\mathrm{IV},\\
\mathrm{III}_c &= \tfrac{1}{2}\chi\cdot\zeta\mathrm{VIII} - \tfrac{1}{4}\chi^2\mathrm{III} - \tfrac{1}{4}\zeta^2\mathrm{II},\\
\mathrm{IV}_c &= \tfrac{1}{2}\chi\cdot\rho\mathrm{IX} - \tfrac{1}{4}\chi^2\mathrm{IV} - \tfrac{1}{4}\rho^2\mathrm{II},\\
\mathrm{V}_c &= \tfrac{1}{2}\zeta\cdot\rho\mathrm{X} - \tfrac{1}{4}\zeta^2\mathrm{IV} - \tfrac{1}{4}\rho^2\mathrm{III},\\
\mathrm{VI}_c &= \tfrac{1}{2}\xi\cdot\chi\mathrm{VI} - \tfrac{1}{4}\xi^2\mathrm{II} - \tfrac{1}{4}\chi^2\mathrm{I},\\
\mathrm{VII}_c &= \tfrac{1}{2}[\xi^2\mathrm{X} + \zeta\cdot\rho\mathrm{I} - \xi\cdot\rho\mathrm{VI} - \xi\cdot\zeta\mathrm{VII}],\\
\mathrm{VIII}_c &= -\tfrac{1}{2}[\chi^2\mathrm{X} + \zeta\cdot\rho\mathrm{II} - \chi\cdot\rho\mathrm{VIII} - \chi\cdot\zeta\mathrm{IX}],\\
\mathrm{IX}_c &= \tfrac{1}{2}[\zeta^2\mathrm{VII} + \rho\cdot\xi\mathrm{III} - \zeta\cdot\xi\mathrm{X} - \zeta\cdot\rho\mathrm{VI}],\\
\mathrm{X}_c &= \tfrac{1}{2}[\xi^2\mathrm{IX} + \chi\cdot\rho\mathrm{I} - \xi\cdot\rho\mathrm{V} - \xi\cdot\chi\mathrm{VII}],\\
\mathrm{XI}_c &= -\tfrac{1}{2}[\zeta^2\mathrm{V} + \chi\cdot\xi\mathrm{III} - \zeta\cdot\xi\mathrm{VIII} - \chi\cdot\zeta\mathrm{VI}],\\
\mathrm{XII}_c &= \tfrac{1}{2}[\rho^2\mathrm{V} + \chi\cdot\xi\mathrm{IV} - \rho\cdot\xi\mathrm{IX} - \chi\cdot\rho\mathrm{VII}],\\
\mathrm{XIII}_c &= \tfrac{1}{2}[\zeta^2\mathrm{IX} + \chi\cdot\rho\mathrm{III} - \zeta\cdot\rho\mathrm{VIII} - \chi\cdot\zeta\mathrm{X}],\\
\mathrm{XIV}_c &= \tfrac{1}{2}[\chi^2\mathrm{VII} + \xi\cdot\rho\mathrm{II} - \chi\cdot\rho\mathrm{V} - \chi\cdot\xi\mathrm{IX}],\\
\mathrm{XV}_c &= -\tfrac{1}{2}[\xi^2\mathrm{VIII} + \chi\cdot\zeta\mathrm{I} - \xi\cdot\zeta\mathrm{V} - \xi\cdot\chi\mathrm{VI}],\\
\mathrm{XVI}_c &= \tfrac{1}{2}[\rho^2\mathrm{VI} + \zeta\cdot\xi\mathrm{IV} - \rho\cdot\xi\mathrm{X} - \zeta\cdot\rho\mathrm{VII}],\\
\mathrm{XVII}_c &= \tfrac{1}{2}[\chi^2\mathrm{VI} + \xi\cdot\zeta\mathrm{II} - \chi\cdot\zeta\mathrm{V} - \chi\cdot\xi\mathrm{VIII}],\\
\mathrm{XVIII}_c &= -\tfrac{1}{2}[\rho^2\mathrm{VIII} + \zeta\cdot\chi\mathrm{IV} - \rho\cdot\chi\mathrm{X} - \zeta\cdot\rho\mathrm{IX}],\\
\mathrm{XIX}_c &= \tfrac{1}{2}[2\xi\cdot\rho\mathrm{VIII} + 2\zeta\cdot\chi\mathrm{VII} - \chi\cdot\xi\mathrm{X}\\
&\quad - \zeta\cdot\rho\mathrm{V} - \zeta\cdot\xi\mathrm{IX} - \rho\cdot\chi\mathrm{VI}],\\
\mathrm{XX}_c &= -\tfrac{1}{2}[2\rho\cdot\chi\mathrm{VI} + 2\xi\cdot\zeta\mathrm{IX} - \rho\cdot\zeta\mathrm{V}\\
&\quad - \xi\cdot\chi\mathrm{X} - \zeta\cdot\chi\mathrm{VII} - \rho\cdot\xi\mathrm{VIII}].
\end{aligned}
$$

Since the above 20 contractions are given as linear combinations of the 10 algebraically possible Ricci basis tensors, only 10 of the 20 contractions are independent; the other 10 equations reduce to relationships between the contractions. For each different choice of normalization for the four basis vectors of the tetrad, there is, in general, a different set of 10 independent contractions. For example, if we choose normalization type 4 from Table I, with $\chi^2 = \rho^2 = \zeta^2 = -1$, $\xi^2 = 1$, all scalar products $= 0$, then we can choose the following as our 10 independent

equations:

$$\begin{aligned} I_c &= \tfrac{1}{4}[I - III], & IX_c &= -\tfrac{1}{2}VII, \\ II_c &= \tfrac{1}{4}[I - IV], & X_c &= \tfrac{1}{2}IX, \\ III_c &= \tfrac{1}{4}[II + III], & XI_c &= \tfrac{1}{2}V, \\ V_c &= \tfrac{1}{4}[III + IV], & XV_c &= -\tfrac{1}{2}VIII, \\ VII_c &= \tfrac{1}{2}X, & XVI_c &= -\tfrac{1}{2}VI. \end{aligned}$$

And our remaining 10 equations reduce to the following relationships between the contractions:

$$\begin{aligned} IV_c &= I_c - II_c + III_c, & XIV_c &= IX_c, \\ VI_c &= II_c - III_c + V_c, & XVII_c &= XVI_c, \\ VIII_c &= VII_c, & XVIII_c &= -XV_c, \\ XII_c &= -XI_c, & XIX_c &= 0, \\ XIII_c &= -X_c, & XX_c &= 0. \end{aligned}$$

To go the other way, from a given algebraic Ricci tensor to the algebraic Riemann tensor whose contraction is the algebraic Ricci tensor, is more complicated. We introduce the notation

$$I \equiv (R_{jk})_I, \quad I_r \equiv (R_{ijkl})_I.$$

In order to accomplish this inversion, we had to choose some minimum normalization. We chose the following:

$$\chi^2 = \rho^2 = -1,$$
$$\chi \cdot \rho = \xi \cdot \chi = \xi \cdot \rho = \zeta \cdot \chi = \zeta \cdot \rho = 0.$$

This corresponds to types 3, 4, 5, 13, 14, or 15 of Table I. With this choice of normalization, we obtained the inversion formulas listed below, where the algebraic Ricci tensor on the left of the colon results from the contraction of the algebraic Riemann tensor on the right. [Thus, for example, $(R_{kj})_I$ results from the contraction over i and l of $2[(R_{ijkl})_{II} + (R_{ijkl})_{VI} + \xi^2(R_{ijkl})_{IV}]$, or in our shorthand notation, I: $2(II_r + VI_r + \xi^2 IV_r)$]:

$$\begin{aligned} I&: 2(II_r + VI_r + \xi^2 IV_r), \\ \xi^2 II&: 2(II_r - VI_r + \xi^2 IV_r), \\ \zeta^2 II&: 2(-III_r + V_r + \zeta^2 IV_r), \\ \xi \cdot \zeta II&: XVII_r - XVI_r + 2\xi \cdot \zeta IV_r, \\ III&: 2(III_r + V_r + \zeta^2 IV_r), \\ \xi^2 IV&: 2(VI_r + \xi^2 IV_r - II_r), \\ \zeta^2 IV&: 2(III_r - V_r + \zeta^2 IV_r), \\ \xi \cdot \zeta IV&: XVI_r - XVII_r + 2\xi \cdot \zeta IV_r, \\ V&: -2XII_r, \\ VI&: -XVI_r - XVII_r + 2\xi \cdot \zeta IV_r, \\ VII&: -2XIV_r, \\ VIII&: 2XVIII_r, \\ \xi^2 IX&: 2X_r, \\ \zeta^2 IX&: 2XIII_r, \\ \xi \cdot \zeta IX&: -2XIX_r, \\ X&: 2VIII_r, \\ 0&: [(\xi \cdot \zeta)^2 - \xi^2\zeta^2]IV_r - I_r - \tfrac{1}{2}\xi \\ &\quad \cdot \zeta(XVI_r + XVII_r) - \tfrac{1}{2}\xi^2(III_r + V_r) \\ &\quad - \tfrac{1}{2}\zeta^2(II_r + VI_r), \\ 0&: -VII_r + \xi^2 VIII_r + \xi \cdot \zeta XIV_r, \\ 0&: -IX_r - \zeta^2 XIV_r - \xi \cdot \zeta VIII_r, \\ 0&: -XI_r + \zeta^2 XII_r + \xi \cdot \zeta XVIII_r, \\ 0&: -XV_r - \xi^2 XVIII_r - \xi \cdot \zeta XII_r, \\ 0&: -XX_r + 2XIX_r. \end{aligned}$$

The last six combinations of algebraic Riemann tensors give zero when contracted. The inversion equations given above are not unique. Indeed, since we are going from a 10-dimensional space to a 20-dimensional space, the inversion has a 10-parameter degree of freedom. When a particular normalization of the basis vector tetrad is chosen, the above 20 equations reduce to 10 inversion relations plus 10 combinations of algebraic Riemann tensors whose contractions give zero. To any of the 10 inversion relations we can add to the right side any arbitrary sum of the 10 combinations, since, when contracted, this sum is zero. This arbitrary choice represents our 10-parameter degree of freedom.

IV. NEWMAN'S CLASSIFICATION FOR $R_{ij} = 0$

Newman[5] found the 10-dimensional subspace of Riemann tensors whose contractions are zero Ricci tensors. This would correspond in our inversion equations above to the 10 combinations of algebraic Riemann tensors whose contractions give zero. Newman worked with a particularly interesting set of basis vectors. He chose a basis of type 14 from Table I with the further restriction that the six bivectors formed from the four basis vectors are each other's duals. Newman's normalization is

$$\xi^2 = \chi^2 = 0, \quad \xi \cdot \chi = 1,$$
$$\zeta^2 = \rho^2 = -1, \quad \zeta \cdot \rho = 0,$$
$$\xi \cdot \zeta = \xi \cdot \rho = \chi \cdot \zeta = \chi \cdot \rho = 0,$$

and

$$M_{ij} = Q^*_{ij},$$
$$N_{ij} = P^*_{ij},$$
$$S_{ij} = L^*_{ij},$$

where the dual is defined by

$$Q^*_{ij} \equiv \tfrac{1}{2}\sqrt{-g}\,\epsilon_{ijkl}Q^{kl}$$

The restriction $R_{ij} = 0$ is 10 independent equations between the 20 independent components of R_{ijkl}. This defines a 10-dimensional subspace, and, in our system, Newman has found the following 10 tensors which span the subspace:

$$\begin{aligned}
\mathrm{I}_N &= \mathrm{I}_r - \mathrm{II}_r, & \mathrm{I}_N^* &= \mathrm{VII}_r,\\
\mathrm{II}_N &= \mathrm{III}_r - \mathrm{IV}_r, & \mathrm{II}_N^* &= \mathrm{VIII}_r,\\
\mathrm{III}_N &= \mathrm{IX}_r - \mathrm{X}_r, & \mathrm{III}_N^* &= \mathrm{XV}_r + \mathrm{XVI}_r,\\
\mathrm{IV}_N &= \mathrm{XIII}_r - \mathrm{XIV}_r, & \mathrm{IV}_N^* &= \mathrm{XVII}_r + \mathrm{XVIII}_r,\\
\mathrm{V}_N &= \mathrm{XI}_r - \mathrm{XII}_r + 2(\mathrm{V}_r - \mathrm{VI}_r), & &\\
\mathrm{V}_N^* &= \mathrm{XIX}_r + \mathrm{XX}_r, & &
\end{aligned}$$

where

$\mathrm{I}_N \equiv$ Newman's basis-tensor number I,

$\mathrm{I}_r = (R_{ijkl})_{\mathrm{I}}$.

Every algebraic Riemann tensor which has a zero Ricci tensor is therefore a linear combination of these 10 basis tensors.

V. CLASSIFICATION OF RICCI NULL SPACES

We call a Riemann manifold for which $R_{im}R^m_j = 0$, $R_{ij} \neq 0$, a Ricci null manifold. Below we find all possible algebraic Ricci tensors which are Ricci null. The inversion formulas, in Sec. III, then give all possible algebraic Riemann tensors whose contractions are Ricci null.

We again assume the minimum normalization

$$\chi^2 = \rho^2 = -1,$$
$$\chi \cdot \rho = \xi \cdot \chi = \xi \cdot \rho = \zeta \cdot \chi = \zeta \cdot \rho = 0.$$

Let any arbitrary algebraic Ricci tensor be given by

$$R_{ij} = \sum_{m=\mathrm{I}}^{\mathrm{X}} a_m (R_{ij})_m .$$

Then R_{ij} is Ricci null if

$$\begin{aligned}
R_{mi}R^m j &= \left[\sum_{n=\mathrm{I}}^{\mathrm{X}} a_n (R_{mi})_n\right]\left[\sum_{n=\mathrm{I}}^{\mathrm{X}} a_n (R^m_j)_n\right]\\
&= \sum_{k=\mathrm{I}}^{\mathrm{X}} b_k (R_{ij})_k = 0,
\end{aligned}$$

where

$$\begin{aligned}
b_{\mathrm{I}} &= \xi^2 a_{\mathrm{I}}^2 + \xi \cdot \zeta a_{\mathrm{I}} a_{\mathrm{VI}} - \tfrac{1}{4} a_{\mathrm{V}}^2 + \tfrac{1}{4}\zeta^2 a_{\mathrm{VI}}^2 - \tfrac{1}{4} a_{\mathrm{VII}}^2,\\
b_{\mathrm{II}} &= -a_{\mathrm{II}}^2 + \tfrac{1}{4}\xi^2 a_{\mathrm{V}}^2 + \tfrac{1}{2}\xi \cdot \zeta a_{\mathrm{V}} a_{\mathrm{VIII}}\\
&\quad + \tfrac{1}{4}\zeta^2 a_{\mathrm{VIII}}^2 - \tfrac{1}{4} a_{\mathrm{IX}}^2,\\
b_{\mathrm{III}} &= \zeta^2 a_{\mathrm{III}}^2 + \xi \cdot \zeta a_{\mathrm{III}} a_{\mathrm{VI}} + \tfrac{1}{4}\xi^2 a_{\mathrm{VI}}^2\\
&\quad - \tfrac{1}{4} a_{\mathrm{VIII}}^2 - \tfrac{1}{4} a_{\mathrm{X}}^2,\\
b_{\mathrm{IV}} &= -a_{\mathrm{IV}}^2 + \tfrac{1}{4}\xi^2 a_{\mathrm{VII}}^2 + \tfrac{1}{2}\xi \cdot \zeta a_{\mathrm{VII}} a_{\mathrm{X}}\\
&\quad - \tfrac{1}{4} a_{\mathrm{IX}}^2 + \tfrac{1}{4}\zeta^2 a_{\mathrm{X}}^2,\\
b_{\mathrm{V}} &= \xi^2 a_{\mathrm{I}} a_{\mathrm{V}} + \xi \cdot \zeta a_{\mathrm{I}} a_{\mathrm{VIII}} - a_{\mathrm{II}} a_{\mathrm{V}} + \tfrac{1}{2}\xi \cdot \zeta a_{\mathrm{V}} a_{\mathrm{VI}}\\
&\quad + \tfrac{1}{2}\zeta^2 a_{\mathrm{VI}} a_{\mathrm{VIII}} - \tfrac{1}{2} a_{\mathrm{VII}} a_{\mathrm{IX}},\\
b_{\mathrm{VI}} &= 2\xi \cdot \zeta a_{\mathrm{I}} a_{\mathrm{III}} + \xi^2 a_{\mathrm{I}} a_{\mathrm{VI}} + \zeta^2 a_{\mathrm{III}} a_{\mathrm{VI}}\\
&\quad - \tfrac{1}{2} a_{\mathrm{V}} a_{\mathrm{VIII}} + \tfrac{1}{2}\xi \cdot \zeta a_{\mathrm{VI}}^2 - \tfrac{1}{2} a_{\mathrm{VII}} a_{\mathrm{X}},\\
b_{\mathrm{VII}} &= \xi^2 a_{\mathrm{I}} a_{\mathrm{VII}} + \xi \cdot \zeta a_{\mathrm{I}} a_{\mathrm{X}} - a_{\mathrm{IV}} a_{\mathrm{VII}} - \tfrac{1}{2} a_{\mathrm{V}} a_{\mathrm{IX}}\\
&\quad + \tfrac{1}{2}\xi \cdot \zeta a_{\mathrm{VI}} a_{\mathrm{VII}} + \tfrac{1}{2}\zeta^2 a_{\mathrm{VI}} a_{\mathrm{X}},\\
b_{\mathrm{VIII}} &= -a_{\mathrm{II}} a_{\mathrm{VIII}} + \zeta^2 a_{\mathrm{III}} a_{\mathrm{VIII}} + \tfrac{1}{2}\xi^2 a_{\mathrm{V}} a_{\mathrm{VI}}\\
&\quad + \tfrac{1}{2}\xi \cdot \zeta a_{\mathrm{VI}} a_{\mathrm{VIII}} - \tfrac{1}{2} a_{\mathrm{IX}} a_{\mathrm{X}} + \xi \cdot \zeta a_{\mathrm{III}} a_{\mathrm{V}},\\
b_{\mathrm{IX}} &= -a_{\mathrm{II}} a_{\mathrm{IX}} - a_{\mathrm{IV}} a_{\mathrm{IX}} + \tfrac{1}{2}\xi^2 a_{\mathrm{V}} a_{\mathrm{VII}}\\
&\quad + \tfrac{1}{2}\xi \cdot \zeta a_{\mathrm{V}} a_{\mathrm{X}} + \tfrac{1}{2}\xi \cdot \zeta a_{\mathrm{VII}} a_{\mathrm{VIII}} + \tfrac{1}{2}\zeta^2 a_{\mathrm{VIII}} a_{\mathrm{X}},\\
b_{\mathrm{X}} &= \xi \cdot \zeta a_{\mathrm{III}} a_{\mathrm{VII}} + \zeta^2 a_{\mathrm{III}} a_{\mathrm{X}} - a_{\mathrm{IV}} a_{\mathrm{X}}\\
&\quad + \tfrac{1}{2}\xi^2 a_{\mathrm{VI}} a_{\mathrm{VII}} + \tfrac{1}{2}\xi \cdot \zeta a_{\mathrm{VI}} a_{\mathrm{X}} - \tfrac{1}{2} a_{\mathrm{VIII}} a_{\mathrm{IX}}.
\end{aligned}$$

The $(R_{ij})_{\mathrm{I}}, \cdots, (R_{ij})_{\mathrm{X}}$ are linearly independent. Hence, the coefficient of each $(R_{ij})_k$ in our defining equation must be zero, i.e., $b_k = 0$. This gives us a set of 10 quadratic equations for the 10 factors $a_{\mathrm{I}}, \cdots, a_{\mathrm{X}}$. For any given normalization, these equations can be solved by the usual algebraic techniques. We have solved them for normalization types 4 and 14. The results follow below.

A. Normalization Type 4

We choose

$$\xi^2 = 1, \quad \zeta^2 = \rho^2 = \chi^2 = -1,$$
and all scalar products = 0.

The complete set of real solutions for the a_i's is: a_{V}, a_{VI}, a_{VII} arbitrary (at least one must be nonzero):

$$a_{\mathrm{I}} = \pm\tfrac{1}{2}(a_{\mathrm{V}}^2 + a_{\mathrm{VI}}^2 + a_{\mathrm{VII}}^2)^{\frac{1}{2}},$$

$$a_{\mathrm{II}} = \frac{a_{\mathrm{V}}^2}{4a_{\mathrm{I}}}, \qquad a_{\mathrm{III}} = \frac{a_{\mathrm{VI}}^2}{4a_{\mathrm{I}}}, \qquad a_{\mathrm{IV}} = \frac{a_{\mathrm{VII}}^2}{4a_{\mathrm{I}}},$$

$$a_{\mathrm{VIII}} = \frac{a_{\mathrm{V}} a_{\mathrm{VI}}}{2a_{\mathrm{I}}}, \quad a_{\mathrm{IX}} = \frac{a_{\mathrm{V}} a_{\mathrm{VII}}}{2a_{\mathrm{I}}}, \quad a_{\mathrm{X}} = \frac{a_{\mathrm{VI}} a_{\mathrm{VII}}}{2a_{\mathrm{I}}}.$$

B. Normalization Type 14

We choose

$$\xi^2 = \zeta^2 = 0, \qquad \chi^2 = \rho^2 = -1,$$
$$\xi \cdot \zeta = 1, \qquad \xi \cdot \chi = \chi \cdot \rho = \chi \cdot \zeta = \xi \cdot \rho = \zeta \cdot \rho = 0.$$

Then,

$a_{\mathrm{I}} \neq 0$, arbitrary; all other a_i's $= 0$ is a solution.

Also,

$a_{\mathrm{III}} \neq 0$, arbitrary; all other a_i's $= 0$ is a solution.

All other solutions are given by the following set:

$$a_{\mathrm{V}}, a_{\mathrm{VII}}, a_{\mathrm{VIII}}, a_{\mathrm{X}}$$

any solutions of

$$a_{\mathrm{V}}a_{\mathrm{X}} = a_{\mathrm{VII}}a_{\mathrm{VIII}}, \quad a_{\mathrm{V}}a_{\mathrm{VIII}} + a_{\mathrm{VII}}a_{\mathrm{X}} > 0,$$

and

$$a_{\mathrm{VI}} = \pm[\tfrac{1}{2}(a_{\mathrm{V}}a_{\mathrm{VIII}} + a_{\mathrm{VII}}a_{\mathrm{X}})]^{\frac{1}{2}},$$

$$a_{\mathrm{I}} = \frac{a_{\mathrm{V}}^2 + a_{\mathrm{VII}}^2}{4a_{\mathrm{VI}}}, \qquad a_{\mathrm{III}} = \frac{a_{\mathrm{VIII}}^2 + a_{\mathrm{X}}^2}{4a_{\mathrm{VI}}},$$

$$a_{\mathrm{II}} = a_{\mathrm{VI}}\frac{a_{\mathrm{V}}^2}{a_{\mathrm{V}}^2 + a_{\mathrm{VII}}^2}, \quad a_{\mathrm{IV}} = a_{\mathrm{VI}}\frac{a_{\mathrm{VII}}^2}{a_{\mathrm{V}}^2 + a_{\mathrm{VII}}^2},$$

$$a_{\mathrm{IX}} = 2a_{\mathrm{VI}}\frac{a_{\mathrm{V}}a_{\mathrm{VII}}}{a_{\mathrm{V}}^2 + a_{\mathrm{VII}}^2}.$$

Each of the above sets of solutions for normalization types 4 and 14 is exhaustive; that is, every Ricci null tensor is contained in the solutions.

One immediate application of our classification is to prove that all Ricci null tensors have zero scalar curvature; that is, that

$$R_{mi}R_j^m = 0 \Rightarrow R \equiv R_m^m = 0.$$

This is proved simply by taking either normalization, say type 4, and calculating R:

$$R \equiv R_m^m = \sum_{i=\mathrm{I}}^{\mathrm{X}} a_i(R_m^m)_i = a_{\mathrm{I}} - a_{\mathrm{II}} - a_{\mathrm{III}} - a_{\mathrm{IV}} = 0.$$

(Of course, the fact that $R_{mi}R_j^m = 0 \Rightarrow R_m^m = 0$ can also be proved by elementary matrix theory by an application of Sylvester's law of degeneracy.)

Starting again from the general classification we have also shown, in a similar manner, that all algebraically possible spaces for which $R_k^l R_m^k = 0$ are not necessarily $R_{lm}R^{klmn} = 0$ spaces, and all those for which $R_{lm}R^{klmn} = 0$ are not necessarily $R_k^l R_m^k = 0$ spaces. It is to be noted that these statements may not hold for bona fide Riemann and Ricci curvature tensors, although it is not unreasonable to conjecture that they do. These contractions occur in one approach to the elementary particle problem which uses Riemannian geometry.[6,7]

[6] G. Schrank, "Space, Time, and the Elementary Particles. I" (to be published).

[7] G. Schrank and J. Thompson, "Space, Time, and the Elementary Particles. II" (to be published).

Projection of Exact Spin Eigenfunctions*

VEDENE H. SMITH, JR.†
Quantum Chemistry Group, University of Uppsala, Uppsala, Sweden
AND
FRANK E. HARRIS‡
United Aircraft Research Laboratories, East Hartford, Connecticut

(Received 20 March 1967)

The relationships defining spin projectors are reviewed and the explicit formulations of the resulting spin-dependent (Sanibel) coefficients are shown to be consistent. Recurrence relations among Sanibel coefficients are derived. It is shown how the formal results here presented are useful in the calculation of matrix elements of spin-free and spin-dependent operators.

I. INTRODUCTION

For many reasons it is useful to have many-electron wavefunctions which are constructed to be eigenfunctions of one component of the spin angular momentum and of the square of the total spin angular momentum. Such eigenfunctions were first produced by the coupling of one-electron spin eigenfunctions by methods which related the eigenfunctions of a given spin system to those of the systems containing fewer spins.[1] These so-called vector-coupling methods are applicable to systems more general than the assemblies of spin-$\frac{1}{2}$ systems to be considered here and do not lead directly to highly convenient formulations of multi-electron spin problems.

An alternative approach is to introduce a projection operator[2] to obtain from an arbitrary multi-electron spin function an eigenfunction of the desired spin quantum numbers. This approach has turned out to have the practical advantage of ease of formal manipulation, particularly when, as is usually the case, we actually desire matrix elements rather than the spin eigenfunctions themselves. Various ways of introducing spin projectors have been described.[2,3] As we shall see, these are all equivalent, both in that they describe the same projection (namely, to the entire spin subspace of the given quantum numbers) and that they give numerically consistent results.

This paper is particularly concerned with developing the formulas needed for efficient use of the projection-operator method of describing spin eigenfunctions. Accordingly, we define a spin projector and discuss the set of coefficients needed to give it explicit implementation. Different expressions which have been obtained for these coefficients by different methods are then related to each other, making use of the theory of generalized hypergeometric series. We next discuss the recurrence relations satisfied by the coefficients. Then we review the numerical methods which are available for coefficient construction. Finally, we survey the application of the formal mathematics to problems involving matrix elements of spin-free and spin-dependent operators.

II. PROJECTION OPERATORS

We consider a system of N spin-$\frac{1}{2}$ particles. Individual particles are characterized by a two-component spin space; we use α and β to denote single-particle spin eigenfunctions with s_z eigenvalues $+\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$, respectively. The most general N-particle spin function is a linear combination of the 2^N different simple-product functions obtained by assigning states α and β to individual electrons in all possible ways. It is convenient to regard these 2^N simple-product functions as spanning a vector space; any N-spin function is a vector in this space.

The simple-product spin functions are eigenfunctions of $\mathcal{S}_z$, the operator for the z component of the total spin. If a product contains n_α α spins and n_β β spins, the eigenvalue is $\frac{1}{2}(n_\alpha - n_\beta)\hbar$. Writing this eigenvalue as $M\hbar$ and writing $\frac{1}{2}N = n$ (note that n may be either integral or half-integral), we have

$$\begin{aligned} n_\alpha &= n + M, \\ n_\beta &= n - M. \end{aligned} \tag{1}$$

On the other hand, a product function is not usually an eigenfunction of $\mathcal{S}^2$, the operator for the square of the total spin. However, since $\mathcal{S}^2$ and $\mathcal{S}_z$ commute, it

* Supported in part at Uppsala University by the King Gustaf VI Adolf's 70-Year Fund for Swedish Culture, Knut and Alice Wallenberg's Foundation, by the Aeronautical Research Laboratory, OAR, through the European office, Aerospace Research, U.S. Air Force, and at Stanford University by National Science Foundation Grant GP-5555.

† Present address: Department of Chemistry, Queen's University, Kingston, Ontario, Canada.
‡ Present address: Department of Physics, University of Utah, Salt Lake City, Utah 84112, U.S.A.

[1] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959).
[2] P.-O. Löwdin, Phys. Rev. 97, 1509 (1955). Includes explicit formulas for $S = M$.
[3] J. K. Percus and A. Rotenberg, J. Math. Phys. 3, 928 (1962). Includes explicit formulas for $S = M$, $S = 0$.

is possible to form linear combinations of product functions which are simultaneous eigenfunctions of both $\mathcal{S}^2$ and $\mathcal{S}_z$. We, therefore, may span the 2^N-dimensional N-spin space by the set of such simultaneous eigenfunctions and regard it as the union of a set of disjoint subspaces, each subspace characterized by its eigenvalues of $\mathcal{S}^2$ and $\mathcal{S}_z$. Since the eigenvalues are of the form $S(S+1)\hbar^2$ and $M\hbar$, each subspace is identified by its quantum numbers S and M.

We introduce the projection operator $\mathcal{O}_{SM}$, and define it to have the property that it projects an arbitrary N-spin function onto the subspace characterized by S and M. Formally, we may write $\mathcal{O}_{SM}$ as a partial resolution of the identity:

$$\mathcal{O}_{SM} = \sum_k |\theta_{SMk}\rangle\langle\theta_{SMk}|, \tag{2}$$

where θ_{SMk} is the kth eigenfunction of quantum numbers S and M, and the summation is over an orthonormal set of θ_{SMk} spanning the S, M subspace. In other words, the θ_{SMk} are an orthonormal set of all the linearly independent eigenfunctions for the given S and M. In view of the foregoing, it is clear that

$$\sum_{S,M} \mathcal{O}_{SM} = I, \tag{3}$$

where I is the identity operator. Moreover, from Eq. (2) and the orthonormality of the θ_{SMk},

$$\mathcal{O}_{SM}^2 = \mathcal{O}_{SM}, \tag{4}$$

$$\mathcal{O}_{SM}^\dagger = \mathcal{O}_{SM}, \tag{5}$$

verifying that θ_{SM} is a projector and showing that it is self-adjoint.

The projector introduced in Eq. (2) is not the only projector which produces an eigenfunction of the given S and M.[4] A projection onto any portion of the S, M subspace obviously would also produce such an eigenfunction and different eigenfunctions (but with the same S and M values) would result from projections onto different parts of the S, M subspace. The reason for using the complete subspace, as in Eq. (2), is the desire to have $\mathcal{O}_{SM}$ commute with all permutation operators.

To show that $\mathcal{O}_{SM}$ commutes with all permutation operators, we proceed as follows. Let P stand for an arbitrary permutation operator. Because P, $\mathcal{S}^2$, and $\mathcal{S}_z$ all commute, $P\theta_{SMk}$ must still be an eigenfunction with quantum numbers S and M, i.e., in general a linear combination of the θ_{SMk}:

$$P\theta_{SMk} = \sum_j V_{jk}(P)\theta_{SMj}. \tag{6}$$

It can be shown that these coefficients $V_{jk}(P)$ define an irreducible unitary matrix representation of the permutation group.[5] Now, referring to Eq. (2), we see that (noting that P is a unitary operator):

$$\mathcal{O}_{SM}P = \sum_k |\theta_{SMk}\rangle\langle P^{-1}\theta_{SMk}|$$
$$= \sum_{k,j} V_{jk}^*(P^{-1})\,|\theta_{SMk}\rangle\langle\theta_{SMj}|, \tag{7a}$$

$$P\mathcal{O}_{SM} = \sum_k |P\theta_{SMk}\rangle\langle\theta_{SMk}|$$
$$= \sum_{k,j} V_{jk}(P)\,|\theta_{SMj}\rangle\langle\theta_{SMk}|. \tag{7b}$$

Interchanging dummy subscripts k and j in Eq. (7b) and writing $V_{jk}^*(P^{-1}) = V_{kj}(P)$, we see that $\mathcal{O}_{SM}$ and P commute. Because the $V_{jk}(P)$ describe an irreducible representation, there exists no subset of the θ_{SMk} for which this analysis can be carried out for all P.

It is possible to use the $V_{jk}(P)$ to form an explicit representation of $\mathcal{O}_{SM}$. In particular,

$$\mathcal{O}_{SM} = \frac{d}{N!}\sum_P \sum_{k=1}^{d} V_{kk}^*(P)P, \tag{8}$$

where the P summation is over all $N!$ permutations of N spins, and d is the dimension of the S, M subspace (the number of linearly independent θ_{SMk}). Equation (8) shows that $\mathcal{O}_{SM}$ can be characterized by the characters of $\mathbf{V}(P)$.

Neither Eq. (2) nor Eq. (8) is an entirely convenient way of characterizing $\mathcal{O}_{SM}$ for practical use. It is more useful to give the result of the application of $\mathcal{O}_{SM}$ to each member of a basis of the N-spin space. Following Löwdin,[2] we seek to describe the action of $\mathcal{O}_{SM}$ on a simple product of α and β spin functions. This course of action has the great advantage that in actuality the analysis need be carried out only for one member of the basis; results for the remaining basis functions can be immediately inferred.

Consider the application of $\mathcal{O}_{SM}$ to a spin function in which the first n_α spins are in state α and the remaining n_β spins are in state β, where n_α and n_β are as given in Eq. (1). Any other values of n_α and n_β will lead to a zero result when $\mathcal{O}_{SM}$ is applied. We designate this product function $[\alpha^{n_\alpha}][\beta^{n_\beta}]$. The notation will be explained shortly.

Since all eigenfunctions of given S and M have equal numbers of α functions, $\mathcal{O}_{SM}$ converts $[\alpha^{n_\alpha}][\beta^{n_\beta}]$ into a linear combination of functions generated by permuting its α and β functions into

[4] F. E. Harris, Advanc. Quantum Chem. 3, 61 (1966).

[5] For a fuller discussion, see M. Kotani, A. Amemiya, E. Ishiguro, and T. Kimura, *Table of Molecular Integrals* (Maruzen Co. Ltd., Tokyo, 1955).

distinguishable orderings:

$$\mathcal{O}_{SM}[\alpha^{n_\alpha}][\beta^{n_\beta}] = \sum_P{}' C(P)P[\alpha^{n_\alpha}][\beta^{n_\beta}], \tag{9}$$

where the sum is over those P which give rise to distinguishable product functions.

We can relate $C(P)$ for various P by use of the fact that $\mathcal{O}_{SM}$ and every P commute. To do so, we apply to both sides of Eq. (9) a permutation Q which affects only the first n_α spins. Bringing Q through $\mathcal{O}_{SM}$ on the left side, we see that it only permutes the n_α α spins and therefore has no effect. This means that the right-hand side of Eq. (9) must also be unaffected. But there, the various P have produced functions not all of whose first n_α spins are identical. Since application of Q produces a nontrivial re-ordering of the first n_α spins in $P[\alpha^{n_\alpha}][\beta^{n_\beta}]$, it is necessary that all P which differ only by permutations among the first n_α functions have the same $C(P)$. Corresponding remarks apply to the application of permutations Q involving the last n_β spins. The conclusion is that $C(P)$ can depend only upon the numbers of α and β spins in the first n_α, and last n_β positions of $P[\alpha^{n_\alpha}][\beta^{n_\beta}]$.

As a result of the arguments of the preceding paragraph, we write

$$\mathcal{O}_{SM}[\alpha^{n_\alpha}][\beta^{n_\beta}] = \sum_j C_j(S, M, n)[\alpha^{n_\alpha-j}\beta^j][\alpha^j\beta^{n_\beta-j}], \tag{10}$$

where $[\alpha^p\beta^q]$ is the sum of all the $\binom{p+q}{p}$ possible orderings of p α-functions and q β-functions. The summation is over all j values for which the spin functions are defined, and the $C_j(S, M, n)$ are coefficients whose values are to be the subject of later sections of this paper. Note that we show explicitly the values of S, M, and $n = \frac{1}{2}N$ to which the coefficients refer.[6] We see also that j describes a set of permutations which when applied to $[\alpha^{n_\alpha}][\beta^{n_\beta}]$ cause j β-spins to be placed in the first n_α positions (with j α-spins replacing them in the last n_β positions).

We ordinarily use Eq. (10) in contexts involving the evaluation of a matrix element. For example,

$$\begin{aligned}\langle \mathcal{O}_{SM}[\alpha^{n_\alpha}][\beta^{n_\beta}] \mid \mathcal{O}_{SM}[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle &= \langle [\alpha^{n_\alpha}][\beta^{n_\beta}]|\, \mathcal{O}_{SM}^\dagger \mathcal{O}_{SM}\, |[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle \\ &= \sum_j C_j(S, M, n)\langle [\alpha^{n_\alpha}][\beta^{n_\beta}] \mid [\alpha^{n_\alpha-j}\beta^j][\alpha^j\beta^{n_\beta-j}]\rangle \\ &= C_0(S, M, n). \end{aligned} \tag{11}$$

To obtain Eq. (11) we used the self-adjointness and idempotency of $\mathcal{O}_{SM}$ and the orthonormality of different spin products. Another example, arising when the antisymmetrizer is used, is

$$\begin{aligned}\langle \mathcal{O}_{SM}[\alpha^{n_\alpha}][\beta^{n_\beta}]|\, P\, |\mathcal{O}_{SM}[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle &= \langle [\alpha^{n_\alpha}][\beta^{n_\beta}]|\, P\mathcal{O}_{SM}\, |[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle \\ &= \sum_j C_j(S, M, n)\langle P^{-1}[\alpha^{n_\alpha}][\beta^{n_\beta}] \mid [\alpha^{n_\alpha-j}\beta^j][\alpha^j\beta^{n_\beta-j}]\rangle \\ &= C_{j(P)}(S, M, n). \end{aligned} \tag{12}$$

Here, we require the fact that P and $\mathcal{O}_{SM}$ commute, and we define $j(P)$ to be the number of β spins in the first n_α positions of $P^{-1}[\alpha^{n_\alpha}][\beta^{n_\beta}]$. The scalar product multiplying $C_{j(P)}(S, M, n)$ is unity because there will be exactly one term in its right half which matches the single product $P^{-1}[\alpha^{n_\alpha}][\beta^{n_\beta}]$.

Finally, we consider the effect of applying $\sum_{S,M} \mathcal{O}_{SM}$ which according to Eq. (3) is the identity operator. We drop the M summation since only the M of Eq. (1) leads to nonzero contributions. Then, from Eq. (10),

$$\begin{aligned}\sum_S \langle [\alpha^{n_\alpha}][\beta^{n_\beta}]|\, \mathcal{O}_{SM}\, |[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle &= \sum_{S,j} C_j(S, M, n)\langle [\alpha^{n_\alpha}][\beta^{n_\beta}] \mid [\alpha^{n_\alpha-j}\beta^j][\alpha^j\beta^{n_\beta-j}]\rangle \\ &= \sum_S C_0(S, M, n). \end{aligned}$$

Thus we have a "normalization" condition

$$\sum_S C_0(S, M, n) = 1. \tag{13}$$

If we also apply any permutation, as in $P\sum_{S,M}\mathcal{O}_{SM}$, we have, from Eq. (12),

$$\sum_S \langle [\alpha^{n_\alpha}][\beta^{n_\beta}]|\, P\mathcal{O}_{SM}\, |[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle = \sum_S C_{j(P)}(S, M, n).$$

But, because $P\sum_{S,M}\mathcal{O}_{SM} = P$, this is equivalent to $\langle [\alpha^{n_\alpha}][\beta^{n_\beta}]|\, P\, |[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle$, and any P characterized by a nonzero $j(P)$ will make the matrix element vanish. Thus, we have the "orthogonality" condition

$$\sum_S C_j(S, M, n) = 0, \quad j \neq 0. \tag{14}$$

Since some of the methods for finding the $C_j(S, M, n)$ lead to homogeneous equations, Eq. (13) provides a means of determining the scale of these coefficients. Equation (14) can be used to check numerical work.

III. SPIN PROJECTION (SANIBEL) COEFFICIENTS

The coefficients $C_j(S, M, n)$ have now received extensive study.[2-4,7-15] It was first shown that they

[6] In the notation of Refs. 4, 21, 22, the coefficient here designated $C_j(S, M, n)$ is rendered $C_j(S, M, 2n)$. In Ref. 10, $C_{S,j}$ stands for $C_j(S, M, n)$; M and n are not explicitly shown.

[7] R. Pauncz, J. Chem. Phys. 37, 2739 (1962); $M = 0$.

[8] P.-O. Löwdin (private communication, 1962); $M = 0$.

[9] F. E. Harris (private communication, 1962); $M = 0$.

[10] F. Sasaki and K. Ohno, J. Math. Phys. 4, 1140 (1963); general S, M.

[11] V. H. Smith, Jr., J. Chem. Phys. 41, 277 (1964); general S, M. Note the misprints: in Eq. (1) replace $j = 1$ by $j = 0$, and in Eq. (6) replace k by j.

obey the following recurrence relation, originally derived for $M = 0$[2] and later for nonzero M[11]:

$$(n + M - j)(n - M - j)C_{j+1}(S, M, n) + [n + 2j(n - j) - S(S + 1) + M^2]C_j(S, M, n) + j^2C_{j-1}(S, M, n) = 0. \quad (15)$$

Subsequently, the $C_j(S, M, n)$ were explicitly related to the vector-coupling coefficients.[10] Alternatively, consideration of the behavior of the spin functions under the operations of the three-dimensional rotation group led to the integral formula[3]

$$C_j(S, M, n) = \frac{2S + 1}{2}(-1)^j \int_0^\pi {}_2F_1\left(M - S, M + S + 1; 1; \sin^2\frac{\theta}{2}\right) \times \left(\cos^2\frac{\theta}{2}\right)^{n-j+M}\left(\sin^2\frac{\theta}{2}\right)^j \sin\theta\, d\theta. \quad (16)$$

The vector-coupling method led to the following result, reported by Sasaki and Ohno[10]:

$$C_j(S, M, n) = \frac{(2S + 1)(n - M - j)!\,(S + M)!}{(S - M)!} \times \sum_k \frac{(-1)^k[(S-M+k)!]^2}{k!(S-M+k-j)!(n-S-k)!(2S+1+k)!}. \quad (17)$$

In Eq. (17) and all subsequent expressions involving factorials or binomial coefficients, it is to be understood that summations are to be over all values for which the summand is well defined.

Equation (16) was integrated by Smith,[11] using a representation of ${}_2F_1$ as a Jacobi polynomial, thereby obtaining

$$C_j(S, M, n) = \frac{2S + 1}{n + S + 1}\sum_k (-1)^{S-M+j-k}\binom{S - M}{k} \times \binom{S + M}{S - M - k}\binom{n + S}{n + M - j + k}^{-1}. \quad (18)$$

Replacing k by $S - M - k$, this expression is easily rearranged to the form

$$C_j(S, M, n) = \frac{2S + 1}{n + S + 1}\sum_k (-1)^{j+k}\binom{S - M}{k} \times \binom{S + M}{k}\binom{n + S}{j + k}^{-1}. \quad (19)$$

This is the most efficient formula for M near to $\pm S$.

Equation (19) also makes explicit the symmetry in M: $C_j(S, M, n) = C_j(S, -M, n)$.

Equations (17) and (18) or (19) are not easily shown to be identical. As a first step towards the demonstration, it is convenient to identify $C_j(S, M, n)$ in terms of generalized hypergeometric series. We use the notation[16]

$${}_pF_q\left[\begin{matrix}\alpha_1, \alpha_2, \cdots, \alpha_p; \\ \beta_1, \beta_2, \cdots, \beta_q\end{matrix} z\right] = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k(\alpha_2)_k \cdots (\alpha_p)_k z^k}{(\beta_1)_k(\beta_2)_k \cdots (\beta_q)_k k!}, \quad (20)$$

with $(u)_k = u(u + 1) \cdots (u + k - 1)$, $(u)_0 = 1$. If some α_i is a negative integer the series is deemed to terminate when $(\alpha_i)_k = 0$; no difficulty in interpretation arises unless a $(\beta_j)_k$ becomes zero for a smaller value of k. This case is not encountered here.

By a suitable grouping of the factors in Eq. (19), it may be brought to the form

$$C_j(S, M, n) = \frac{2S + 1}{n + S + 1}(-1)^j\binom{n + S}{j}^{-1} \times {}_3F_2\left[\begin{matrix}M - S, -M - S, j + 1; \\ j - n - S, 1\end{matrix} 1\right]. \quad (21)$$

Incidentally, Eq. (19) may itself be easily verified without the intermediate introduction of a Jacobi polynomial by use of Eq. (20) for ${}_2F_1$ in Eq. (16). We perform the integration directly in the variable $z = \sin^2(\theta/2)$. Each term involves a definite integral of the form

$$\int_0^1 z^p(1 - z)^q\, dz = p!\, q!/(p + q + 1)!.$$

Returning to Eq. (17), we now transform it first to some possibly useful explicit expressions for $C_j(S, M, n)$, and then to a hypergeometric series equivalent to Eq. (21). Rewriting Eq. (17) to cast it in terms of binomial coefficients, we obtain

$$C_j(S, M, n) = \frac{2S + 1}{n + S + 1}\binom{n + S}{S + M}^{-1}\binom{n - M}{j}^{-1} \times \sum_k (-1)^k\binom{n + S + 1}{n - S - k} \times \binom{S - M + k}{k}\binom{S - M + k}{j}. \quad (22)$$

Our next step is to apply the binomial-coefficient addition theorem[17]

$$\binom{a + c}{b + d} = \sum_l \binom{a}{b + l}\binom{c}{d - l} \quad (23)$$

[12] P.-O. Löwdin (private communication, 1964); general S, M.
[13] N. Karayianis and C. A. Morrison, J. Math. Phys. **6**, 876 (1965); general S, M.
[14] J. Shapiro, J. Math. Phys. **6**, 1680 (1965).
[15] J. E. Harriman, J. Chem. Phys. **40**, 2827 (1964).
[16] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill Book Co., Inc., New York, 1953), Vol. 1, Chap. 4.
[17] See, for example, H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry* (D. Van Nostrand Company, Inc., Princeton, N.J., 1956), 2nd ed., pp. 433ff.

in the form

$$\binom{S-M+k}{j} = \sum_l \binom{S-M}{j-l}\binom{k}{l}. \quad (24)$$

Inserting Eq. (24) into Eq. (22), interchanging the k and l summations, and cancelling factorials where possible, we obtain for the k summation of Eq. (22)

$$\sum_l \binom{S-M}{j-l}\binom{S-M+l}{l} \times \sum_k (-1)^k \binom{n+S+1}{n-S-k}\binom{S-M+k}{k-l}.$$

Using the relationship[17]

$$\binom{a}{r} = (-1)^r \binom{r-a-1}{r} \quad (25)$$

to convert $(-1)^k\binom{S-M+k}{k-l}$ to $(-1)^l\binom{M-S-l-1}{k-l}$, the k summation may be evaluated using the addition theorem to reach

$$C_j(S,M,n) = \frac{2S+1}{n+S+1}\binom{n+S}{S+M}^{-1}\binom{n-M}{j}^{-1} \times \sum_l (-1)^l \binom{S-M}{j-l}\binom{S-M+l}{l}\binom{n+M-l}{n-S-l}. \quad (26)$$

Similar steps starting from the replacement in Eq. (26) of $\binom{S-M+l}{l}$ by $\sum_k \binom{S-M+l-j}{l-k}\binom{j}{k}$ lead to

$$C_j(S,M,n) = \frac{(2S+1)(n+M)!\,(n-M)!}{(n+S+1)(n+S)!\,(n-S)!} \times \binom{n+M}{j}^{-1}\binom{n-M}{j}^{-1} \times \sum_k (-1)^{j-k}\binom{S+M}{k}\binom{S-M}{k}\binom{n-S}{j-k} \quad (27)$$

$$= C_0(S,M,n)\binom{n+M}{j}^{-1}\binom{n-M}{j}^{-1} \times \sum_k (-1)^{j-k}\binom{S+M}{k}\binom{S-M}{k}\binom{n-S}{j-k}. \quad (28)$$

Once again we have reached a form in which M and $-M$ appear symmetrically. Equation (28) is the most convenient form for small j values, as the k summation runs at most from 0 to j. This equation is also an appropriate point from which to obtain another expression in hypergeometric series. By straightforward comparison with the definition,

$$C_j(S,M,n) = \frac{2S+1}{n+S+1}(-1)^j\binom{n+S}{j}^{-1} \times \frac{(n-M-j)!\,(n+M-j)!}{(n-S-j)!\,(n+S-j)!} \times {}_3F_2\left[\begin{matrix}M-S, -M-S, -j; 1\\ n-S-j+1, 1\end{matrix}\right]. \quad (29)$$

MacRobert[18] gives an identity involving ${}_3F_2$ series which for the case at hand is

$${}_3F_2\left[\begin{matrix}M-S, -M-S, j+1; 1\\ j-n-S, 1\end{matrix}\right] = \frac{(j-n+M)_{S-M}}{(j-n-S)_{S-M}}\,{}_3F_2\left[\begin{matrix}M-S, -M-S, -j; 1\\ n-S-j+1, 1\end{matrix}\right]. \quad (30)$$

Since

$$\frac{(j-n+M)_{S-M}}{(j-n-S)_{S-M}} = \frac{(n-M-j)!\,(n+M-j)!}{(n-S-j)!\,(n+S-j)!},$$

the Sasaki–Ohno and Smith formulations are proven equivalent.

For completeness, we cite various special cases of the $C_j(S,M,n)$. The first three of these were reported by Sasaki and Ohno. For $j=0$, from Eq. (27),

$$C_0(S,M,n) = \frac{(2S+1)(n-M)!\,(n+M)!}{(n+S+1)(n-S)!\,(n+S)!}. \quad (31)$$

For $M=S$, from Eq. (19), noting that k has only the value zero:

$$C_j(S,S,n) = (-1)^j\left(\frac{2S+1}{n+S+1}\right)\binom{n+S}{j}^{-1}. \quad (32)$$

For $S=n$, using Eq. (27), noting that j and k must be equal,

$$C_j(n,M,n) = \frac{(n+M)!\,(n-M)!}{(2n)!}. \quad (33)$$

We have obtained one new result. For $j=n-M$, using Eq. (28), noting that k has only the value $S-M$,

$$C_{n-M}(S,M,n) = (-1)^{n-S}\left(\frac{2S+1}{n+S+1}\right) \times \binom{n-M}{S-M}\binom{n+S}{S+M}^{-1}. \quad (34)$$

IV. RECURRENCE FORMULAS

Since the $C_j(S,M,n)$ are proportional to generalized hypergeometric series, we may use the results available for these series to obtain recurrence formulas

[18] T. M. MacRobert, Phil. Mag. 28, 488 (1939).

among contiguous $C_j(S, M, n)$. Alternatively, we have found that the recurrence relations can be derived from a further analysis of the properties of the eigenfunctions θ_{SMk} which appear in the definition of $\mathcal{O}_{SM}$.

Taking first the approach based on hypergeometric functions, we start from the relations connecting contiguous functions. The functions contiguous to

$$ {}_pF_q\left[\begin{matrix}\alpha_1, \alpha_2, \cdots, \alpha_p; \\ \beta_1, \beta_2, \cdots, \beta_q\end{matrix} z\right] $$

are those which are obtainable by increasing or decreasing a single α_i or β_j by unity. We denote these contiguous functions by $F(\alpha_i+)$, $F(\alpha_i-)$, $F(\beta_j+)$, $F(\beta_j-)$. Among a given ${}_3F_2$ and its ten contiguous functions there are eight linearly independent linear recurrence relations. However, relations in which β_2 changes are not of use here because they involve hypergeometric functions which cannot be identified with any $C_j(S, M, n)$.

Rainville[19] has derived the recurrence relations needed for use here. Although his derivation is restricted to positive integral β_j, the results are also valid for the zero or negative β_j values needed here. The six contiguous relations not involving $\beta_2\pm$ are, for ${}_3F_2\left[\begin{matrix}\alpha_1,\alpha_2,\alpha_3;\\ \beta_1,\beta_2\end{matrix}1\right]$:

$$ (\alpha_1 - \alpha_2)F = \alpha_1 F(\alpha_1+) - \alpha_2 F(\alpha_2+), \tag{35} $$

$$ (\alpha_1 - \alpha_3)F = \alpha_1 F(\alpha_1+) - \alpha_3 F(\alpha_3+), \tag{36} $$

$$ (\alpha_1 - \beta_1 + 1)F = \alpha_1 F(\alpha_1+) - (\beta_1 - 1)F(\beta_1-), \tag{37} $$

$$ \rho F = \beta_1(\alpha_1 - \beta_2)F(\alpha_1-) - (\alpha_2 - \beta_1)(\alpha_3 - \beta_1)F(\beta_1+), \tag{38} $$

$$ \rho F = \beta_1(\alpha_2 - \beta_2)F(\alpha_2-) - (\alpha_1 - \beta_1)(\alpha_3 - \beta_1)F(\beta_1+), \tag{39} $$

$$ \rho F = \beta_1(\alpha_3 - \beta_2)F(\alpha_3-) - (\alpha_1 - \beta_1)(\alpha_2 - \beta_1)F(\beta_1+), \tag{40} $$

where

$$ \rho = \beta_1(\alpha_1 + \alpha_2 + \alpha_3 - \beta_1 - \beta_2). $$

From Eqs. (21) and (35)–(40) we obtain the following six recurrence relations among the $C_j(S, M, n)$:

$$ (S - M + j + 1)C_j(S, M, n) + (n + S + 2)C_{j+1}(S, M, n + 1) = (S - M)[(2S + 1)/2S] \times C_j(S - \tfrac{1}{2}, M + \tfrac{1}{2}, n + \tfrac{1}{2}), \tag{41} $$

$$ (S + M + j + 1)C_j(S, M, n) + (n + S + 2)C_{j+1}(S, M, n + 1) = (S + M)[(2S + 1)/2S] \times C_j(S - \tfrac{1}{2}, M - \tfrac{1}{2}, n + \tfrac{1}{2}), \tag{42} $$

$$ C_j(S, M, n) + C_{j+1}(S, M, n + 1) = C_j(S, M, n + 1), \tag{43} $$

$$ (n - S)(n + S + 1)C_j(S, M, n) + j^2 C_{j-1}(S, M, n - 1) = (n + M - j)(n - M - j)C_j(S, M, n - 1), \tag{44} $$

$$ (n - S)C_j(S, M, n) + (S - M + 1) \times [(2S + 1)/2S + 2]C_j(S + \tfrac{1}{2}, M - \tfrac{1}{2}, n - \tfrac{1}{2}) = (n - M - j)C_j(S, M, n - 1), \tag{45} $$

$$ (n - S)C_j(S, M, n) + (S + M + 1) \times [(2S + 1)/2S + 2]C_j(S + \tfrac{1}{2}, M + \tfrac{1}{2}, n - \tfrac{1}{2}) = (n + M - j)C_j(S, M, n - 1). \tag{46} $$

We note that Eq. (43) was previously given by Harriman.[15]

The alternative procedure for obtaining these recurrence relations involves an interesting synthesis of projection-operator and vector-coupling techniques. Rather than derive all six formulas, we illustrate with a typical example. Our starting point is the combination of Eqs. (2) and (12) to obtain

$$ \begin{aligned} C_j(S, M, n) &= \langle [\alpha^{n_\alpha}][\beta^{n_\beta}] | \, \mathcal{O}_{SM} P_j \, | [\alpha^{n_\alpha}][\beta^{n_\beta}] \rangle \\ &= \sum_k \langle [\alpha^{n_\alpha}][\beta^{n_\beta}] \mid \theta_{SMk} \rangle \langle \theta_{SMk} | \, P_j \, | [\alpha^{n_\alpha}][\beta^{n_\beta}] \rangle, \end{aligned} \tag{47} $$

where P_j is any permutation which interchanges j α, β-pairs in $[\alpha^{n_\alpha}][\beta^{n_\beta}]$. For the relation to be derived now, we choose P_j such that it leaves the first function unaltered.

The θ_{SMk} are now related to the eigenfunctions for smaller numbers of spins using the genealogical construction scheme. Writing θ^N_{SMk} to indicate explicitly that the functions refer to an N-spin system, the total set of θ^N_{SMk} consist of all those obtainable by either of the formulas

$$ \theta^N_{SMk} = -\left(\frac{S - M + 1}{2S + 2}\right)^{\frac{1}{2}} \alpha \theta^{N-1}_{S+\frac{1}{2}, M-\frac{1}{2}, k} + \left(\frac{S + M + 1}{2S + 2}\right)^{\frac{1}{2}} \beta \theta^{N-1}_{S+\frac{1}{2}, M+\frac{1}{2}, k}, \tag{48} $$

$$ \theta^N_{SMk} = \left(\frac{S + M}{2S}\right)^{\frac{1}{2}} \alpha \theta^{N-1}_{S-\frac{1}{2}, M-\frac{1}{2}, k} + \left(\frac{S - M}{2S}\right)^{\frac{1}{2}} \beta \theta^{N-1}_{S-\frac{1}{2}, M+\frac{1}{2}, k}. \tag{49} $$

[19] E. D. Rainville, Bull. Am. Math. Soc. 51, 266 (1945).

The α and β functions refer to spin 1; the $(N-1)$-spin functions describe spins 2 through N. The k values in Eq. (48) range over the $(N-1)$-spin eigenfunctions for $S+\frac{1}{2}$ while those in Eq. (49) refer to the $(N-1)$-spin eigenfunctions for $S-\frac{1}{2}$. If $S=0$, Eq. (49) does not apply. The summation in Eq. (47) is to be construed as including both these sets of θ^N_{SMk}.

Inserting Eq. (48) and (49) into Eq. (47), and dropping all terms which vanish due to orthogonality in spin 1, we have

$$\begin{aligned} C_j(S,M,n) &= \left(\frac{S-M+1}{2S+2}\right)\sum_k \langle[\alpha^n\alpha][\beta^n\beta] \mid \alpha\theta^{N-1}_{S+\frac{1}{2},M-\frac{1}{2},k}\rangle \\ &\times \langle\alpha\theta^{N-1}_{S+\frac{1}{2},M-\frac{1}{2},k}\mid P_j\mid[\alpha^n\alpha][\beta^n\beta]\rangle \\ &+ \left(\frac{S+M}{2S}\right)\sum_k \langle[\alpha^n\alpha][\beta^n\beta] \mid \alpha\theta^{N-1}_{S-\frac{1}{2},M-\frac{1}{2},k}\rangle \\ &\times \langle\alpha\theta^{N-1}_{S-\frac{1}{2},M-\frac{1}{2},k}\mid P_j\mid[\alpha^n\alpha][\beta^n\beta]\rangle. \end{aligned} \tag{50}$$

Integrating spin 1, we see that the remainder of each term on the right side of Eq. (50) is exactly of the form given in Eq. (47) for an $(N-1)$-spin system; in particular,

$$\begin{aligned} C_j(S,M,n) &= \left(\frac{S-M+1}{2S+2}\right)C_j(S+\tfrac{1}{2},M-\tfrac{1}{2},n-\tfrac{1}{2}) \\ &+ \left(\frac{S+M}{2S}\right)C_j(S-\tfrac{1}{2},M-\tfrac{1}{2},n-\tfrac{1}{2}). \end{aligned} \tag{51}$$

By replacing M by $-M$, and noting that

$$C_j(S,M,n) = C_j(S,-M,n),$$

we also have

$$\begin{aligned} C_j(S,M,n) &= \left(\frac{S+M+1}{2S+2}\right)C_j(S+\tfrac{1}{2},M+\tfrac{1}{2},n-\tfrac{1}{2}) \\ &+ \left(\frac{S-M}{2S}\right)C_j(S-\tfrac{1}{2},M+\tfrac{1}{2},n-\tfrac{1}{2}). \end{aligned} \tag{52}$$

Equations (51) and (52) are linearly dependent upon Eqs. (41)–(46). Similar techniques yield the remainder of the recurrence formulas.

Equations (51) and (52) are also illustrative of a fact which has proven useful in manipulations involving spin-dependent operators, namely that many of the relationships are entirely independent of j. This fact also facilitates the construction of j-independent recursive schemes for generating the $C_j(S,M,n)$.

Now that the recurrence relations are in hand, we can verify that the $C_j(S,M,n)$ satisfy the pure recurrence relation along j which we cited as Eq. (15). Writing Eq. (43) for j and $n-1$:

$$C_j(S,M,n-1) + C_{j+1}(S,M,n) = C_j(S,M,n) \tag{53}$$

and, for $j-1$ and $n-1$,

$$C_{j-1}(S,M,n-1) + C_j(S,M,n) = C_{j-1}(S,M,n), \tag{54}$$

we reach Eq. (15) by eliminating $C_{j-1}(S,M,n-1)$ and $C_j(S,M,n-1)$ from Eq. (44).

V. GENERATION OF THE $C_j(S,M,n)$

We have found that the most efficient way to generate the $C_j(S,M,n)$ is to start from

$$C_j(S,M,S) = (S+M)!\,(S-M)!/(2S)!. \tag{55}$$

Note that these coefficients are independent of j. We then make $C_0(S,M,n)$ for $n=S+1,\ S+2,\cdots$, up to the maximum n value needed. From Eq. (31) we easily obtain

$$C_0(S,M,n) = \left(\frac{n^2-M^2}{n(n+1)-S(S+1)}\right)C_0(S,M,n-1). \tag{56}$$

All remaining $C_j(S,M,n)$ are then built up by systematic use of Eq. (43). We make $C_1(S,M,n)$ for all n values starting from $n=S+1$, using the $C_0(S,M,n)$ values already available and the value of $C_1(S,M,S)$ from Eq. (55). Then from Eq. (55) and the $C_1(S,M,n)$ we make $C_2(S,M,n)$, etc.

Tables of the $C_j(S,M,n)$ for n values through 10 have been prepared by Manne.[20]

VI. PHYSICAL APPLICATIONS

The most frequent applications of the spin projectors discussed here have been to energy calculations. The usual situation is that a wavefunction Ψ is generated from an N-electron spatial function $\Xi(\mathbf{r})$ and a spin function $[\alpha^n\alpha][\beta^n\beta]$ by antisymmetrization and spin projection:

$$\Psi = \mathcal{O}_{SM}\mathcal{A}\Xi(\mathbf{r})[\alpha^n\alpha][\beta^n\beta]. \tag{57}$$

Here $\mathcal{A}$ is the antisymmetrizer, defined as

$$(N!)^{-1}\sum_P \epsilon_P P,$$

where the sum is over all permutations P and ϵ_P is the parity of P.

[20] R. Manne, Theor. Chim. Acta 6, 116 (1966); a more extensive table is available as Preprint No. 153, Quantum Chemistry Group, Uppsala University, Uppsala, Sweden, 1965.

We next form the matrix elements of an operator B, simplifying our results by recognizing that $\mathcal{O}_{SM}$ and $\mathcal{A}$ commute and are idempotent, and that $\mathcal{A}$ commutes with B.

$$\begin{aligned}&\langle\Psi|\, B\,|\Psi\rangle\\ &= \langle \mathcal{O}_{SM}\mathcal{A}\Xi(\mathbf{r})[\alpha^{n_\alpha}][\beta^{n_\beta}]|\, B\,|\mathcal{O}_{SM}\mathcal{A}\Xi(\mathbf{r})[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle\\ &= \langle \Xi(\mathbf{r})[\alpha^{n_\alpha}][\beta^{n_\beta}]|\, \mathcal{O}_{SM}B\mathcal{A}\mathcal{O}_{SM}\,|\Xi(\mathbf{r})[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle. \quad (58)\end{aligned}$$

If B is spin free, $\mathcal{O}_{SM}$ and B commute, and one of the $\mathcal{O}_{SM}$ factors can be dropped from Eq. (58). If we then introduce the definition of $\mathcal{A}$ and separate the space and spin integrations, we reach, with the aid of Eq. (12),

$$\langle\Psi|\, B\,|\Psi\rangle = (N!)^{-1}\sum_P \epsilon_P C_{j(P)}(S, M, n)\,\langle\Xi|\, BP\,|\Xi\rangle. \quad (59)$$

Equation (59) shows that the $C_j(S, M, n)$ entirely characterize the effect of spin upon the matrix element of a spin-free operator B.

Spin-dependent operators can be handled by methods similar to those just described. However, B no longer necessarily commutes with $\mathcal{O}_{SM}$ and it by hypothesis has spin dependence. The procedure to be followed depends upon the nature of the spin dependence of B. If we assume B to be a one-electron operator of the form

$$B = \sum_{i=1}^{N} b_r(\mathbf{r}_i)b_\sigma(\boldsymbol{\sigma}_i), \quad (60)$$

where $\mathbf{r}_i$ and $\boldsymbol{\sigma}_i$ refer to the spatial and spin coordinates of the ith electron, the equation corresponding to Eq. (59) is

$$\begin{aligned}\langle\Psi|\, B\,|\Psi\rangle = (N!)^{-1}\sum_i\sum_P \epsilon_P\,\langle[\alpha^{n_\alpha}][\beta^{n_\beta}]|\\ \times\, \mathcal{O}_{SM}b_\sigma(\boldsymbol{\sigma}_i)P\mathcal{O}_{SM}\,|[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle\\ \times\,\langle\Xi|\, b_r(\mathbf{r}_i)P\,|\Xi\rangle. \quad (61)\end{aligned}$$

Because $\mathcal{O}_{SM}$ and b_σ do not commute, no further direct simplification is possible. We see that the effect of the spin is thus carried in the factor

$$\langle[\alpha^{n_\alpha}][\beta^{n_\beta}]|\, \mathcal{O}_{SM}b_\sigma(\boldsymbol{\sigma}_i)P\mathcal{O}_{SM}\,|[\alpha^{n_\alpha}][\beta^{n_\beta}]\rangle.$$

This factor can be reduced to an expression involving several $C_j(S, M, n)$ differing in their values of j, S, M, and n. The recurrence relations of Sec. IV can be used to simplify the results. For further details the original literature can be consulted.[4,21,22]

ACKNOWLEDGMENTS

The authors gratefully acknowledge the encouragement and support of Professor Per-Olov Löwdin. Large parts of the contributions of both authors to the work reported here were made in the Quantum Chemistry Group of Uppsala University.

[21] F. E. Harris, Mol. Phys. 11, 243 (1966).
[22] F. E. Harris, J. Chem. Phys. 47, 1047 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

On the Isospin Crossing Matrix*

HUAN LEE
Lawrence Radiation Laboratory, University of California, Berkeley, California

(Received 4 May 1967)

New properties and relations to Clebsch–Gordan coefficients have been found and proved for a certain subset of isospin crossing matrix elements. These properties, coupled with the well-known fact that the crossing matrix is simply related to a real orthogonal matrix, provide a quick method to calculate crossing matrices of rank ≤ 4. Procedures of construction and tables are given for all crossing matrices involving $I \leq \frac{3}{2}$. For the sake of completeness as well as clarification of the confusing point about phases, we include a systematic discussion of crossing relations and a general expression for crossing matrices applicable to arbitrary phase conventions. For the crossing matrix in which both direct and crossed reactions are elastic, some interesting inequalities among elements in the first and last columns are noticed and their physical implication is briefly discussed.

I. INTRODUCTION

The relation between the isospin crossing matrix (denoted by C.M. hereafter) for a four-line connected part and the $6j$ symbol[1] has been known[2] for some time. General discussions of the crossing relations and formulas for the C.M., with convention-dependent phase factors, have been given by several authors.[3-5] Although the expression of the C.M. in terms of $6j$ symbol can be easily derived, the over-all phases, which are very important in practical calculations in order to be consistent within a given convention, often cause confusion. To settle this question of crossing phases once and for all, we give a detailed discussion of the crossing relation[6] in the Appendix A. From this result we derived a general formula for the C.M., with its over-all phase written in a compact and physically transparent way, which can suitably accommodate all the different conventions.

Our main interest, however, is to investigate whether we can calculate the low-rank C.M. by just making use of their general properties. One method of this kind, which has been used for[7] πN and for[8] $\pi\pi$ scattering is to obtain the C.M. by explicitly constructing projection operators of definite isospin in both direct and crossed reactions and then searching for their relations. Unfortunately, the approach along this line is quite complicated except for a few simple elastic cases. In the following, we want to present an algebraic method for calculating the C.M. We have found certain simple phase properties of the outermost layer of C.M. elements and their relations with Clebsch–Gordan coefficients. In addition we have proved some useful sum rules for the trace and for the column and row elements in elastic C.M.'s. These properties, coupled with the well-known fact that the C.M. is simply related to a real orthogonal matrix, provide us enough constraints to calculate all C.M.'s of rank equal to or lower than four. So far the largest isospin for established particles, with $B = 0, 1$, is $\frac{3}{2}$; hence, our method covers all C.M.'s involving known baryons and mesons.

As a by-product of our investigation, we have noticed that there exist interesting inequalities for the matrix elements of an arbitrary elastic C.M. The physical implication of this fact is discussed in Sec. VI.

In Sec. II, we discuss the phase convention used in this note and derive a general expression of the C.M. which covers all different conventions. The general properties of the C.M., together with simple sum rules among matrix elements, are given in Sec. III. In Sec. IV, we give procedures for constructing the C.M. and also tabulate all C.M.'s involving $I \leq \frac{3}{2}$. Formulas for all elastic C.M.'s up to rank four are also given; these might be of use for spin crossing matrices in a static model.[7] Finally, in Sec. V, we give the proof of those properties stated in Sec. III.

II. PHASE CONVENTION

For a general reaction involving two-particle channels, if we let $a + b \to c + d$ be the s reaction, as usual, we call $a + \bar{c} \to \bar{b} + d$ and $a + \bar{d} \to c + \bar{b}$ the t and u reactions, respectively. The letters a, b, c, and

* This work was done under the auspices of the United States Atomic Energy Commission.

[1] See, for example, A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1960). Also, M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooten, Jr., *3-j and 6-j Symbols* (Technology Press, Cambridge, Mass., 1959).

[2] F. J. Dyson, Phys. Rev. **100**, 344 (1955).

[3] A. O. Barut and B. C. Unal, Nuovo Cimento **28**, 112 (1963).

[4] P. A. Carruthers and J. P. Krisch, Ann. Phys. (N.Y.) **33**, 1 (1965). These authors divided particles into classes in their discussion of the crossing phase, a procedure which seems to us confusing and unnecessary.

[5] D. E. Neville, "Isospin Crossing Matrices," Lawrence Radiation Laboratory Report UCRL-70059, 1966.

[6] J. R. Taylor, J. Math. Phys. **7**, 181 (1966). Our discussion of isospin crossing relations follows a similar line.

[7] G. F. Chew and F. E. Low, Phys. Rev. **101**, 1570 (1956).

[8] G. F. Chew and S. Mandelstam, Phys. Rev. **119**, 467 (1960).

d stand not only for particle type but also for the isospin value of each particle; the same letters with and without bar are antiparticle conjugate to each other. Throughout this note, the Condon–Shortley convention[1] is used. We use a_α to denote a charge state of multiplet a with I_z component α, and $\langle c_\gamma d_\delta | A | a_\alpha b_\beta \rangle$ to denote the amplitude for a typical reaction $a_\alpha + b_\beta \to c_\gamma + d_\delta$. The phases of all amplitudes or, equivalently, the phases of all particles, are adjusted in such a way that isospin invariance is expressed by

$$\langle c_\gamma d_\delta | A | a_\alpha b_\beta \rangle = \sum_{\alpha' \cdots \delta'} D_{\alpha'\alpha}(u) D_{\beta'\beta}(u) D^*_{\gamma'\gamma}(u) D^*_{\delta'\delta}(u) \times \langle c_{\gamma'} d_{\delta'} | A | a_{\alpha'} b_{\beta'} \rangle, \quad (1)$$

where u is any group element of $SU(2, C)$[9] and $D(u)$ is the usual irreducible representation of u with dimension $(2I + 1)$ for each particle; star means complex conjugation.

As is shown in the Appendix A, the crossing relation compatible with (1) takes the general form

$$\langle c_\gamma d_\delta | A | a_\alpha b_\beta \rangle = \eta(b, \beta)\eta^*(d, \delta) \langle c_\gamma \bar{b}_{-\beta} | A | a_\alpha \bar{d}_{-\delta} \rangle, \quad (2)$$

where the η's are phase factors which can be written as $\eta(j, m) = \eta_j(-1)^m$ with η_j arbitrary and independent of m. Thus, we can take all crossing phases of the type $\eta(j, m)$ to be real for convenience, and rewrite (2) as

$$\langle c_\gamma d_\delta | A | a_\alpha b_\beta \rangle = \eta(b, \beta)\eta(d, \delta) \langle c_\gamma \bar{b}_{-\beta} | A | a_\alpha \bar{d}_{-\delta} \rangle. \quad (3)$$

In real phase convention, which we use hereafter, there is two-fold freedom in selecting η_a for a given isomultiplet a, i.e.,

$$\eta_a = -\eta_{\bar{a}} = \pm i, \quad \text{if } 2a \text{ is odd},$$
$$\eta_a = \eta_{\bar{a}} = \pm 1, \quad \text{if } 2a \text{ is even}.$$

We are now in a position to define the C.M. unambiguously: Let A_s, A_t, and A_u denote the amplitudes of definite isospin in s, t, and u channels respectively; they are defined through[10]

$$\langle c_\gamma d_\delta | A | a_\alpha b_\beta \rangle = \sum_s \langle a, \alpha; b, \beta \mid s \rangle \langle c, \gamma; d, \delta \mid s \rangle A_s, \quad (4)$$

$$\langle \bar{b}_{-\beta} d_\delta | A | a_\alpha \bar{c}_{-\gamma} \rangle = \sum_t \langle a, \alpha; c, -\gamma \mid t \rangle \langle b, -\beta; d, \delta \mid t \rangle A_t, \quad (5)$$

$$\langle c_\gamma \bar{b}_{-\beta} | A | a_\alpha \bar{d}_{-\delta} \rangle = \sum_u \langle a, \alpha; d, -\delta \mid u \rangle \langle c, \gamma; b, -\beta \mid u \rangle A_u. \quad (6)$$

[9] This is the group of complex 2 × 2 unitary unimodular matrix.

[10] Our notation for C-G coefficients is the same as in Ref. 1; sometimes we just write $\langle j_1, m_1; j_2, m_2 \mid j \rangle = \langle j_1, m_1; j_2, m_2 \mid j, m_1 + m_2 \rangle$.

The C.M.'s C_{st} and C_{su} are such that

$$A_s = \sum_t C_{st} A_t, \quad (7)$$

$$A_s = \sum_u C_{su} A_u. \quad (8)$$

Notice that the phases of the A's depend on the particle order of writing the Clebsch–Gordan coefficients, since[1]

$$\langle j_1, m_1; j_2, m_2 \mid j \rangle = (-1)^{j_1 + j_2 - j} \langle j_2, m_2; j_1, m_1 \mid j \rangle.$$

In our convention, Eqs. (4)–(6), we simply let the particle written in the "first" ("second") position in each channel also be the one located at the first (second) position in each C–G coefficient. Thus defined, it can be easily shown that

$$C_{st}(a + b \to c + d) = (-1)^{2d+b+c-s-u} C_{su}(a + b \to d + c), \quad (9)$$

where the reaction in the bracket specifies the s reaction for each C.M. Because of (9) we confine our attention to C_{su} from now on.

With the crossing relation (3) and the definition of the $6j$ symbol, together with (4) and (6), we finally get

$$C_{su} = \eta_b \eta_d (-1)^{b+d+2s} (2u + 1) \begin{Bmatrix} a & b & s \\ c & d & u \end{Bmatrix}. \quad (10)$$

The phase factor in (10) exhibits a symmetrical dependence on the two crossed particles, and in the general phase convention $\eta_d(-1)^d$ is simply replaced by $[\eta_d(-1)^d]^*$. A very important fact which follows from our analysis is that *there are only two cases in which the phase of* (10) *is independent of convention:* one case is when b and d belong to the same particle multiplet, then $\eta_b(-1)^b[\eta_d(-1)^d]^* = 1$; the other case is when b and d belong to two multiplets which are particle–antiparticle conjugate to each other, by (A9) we have

$$\eta_b(-1)^b[\eta_d(-1)^d]^* = \eta_b \eta_{\bar{b}}^* = (-1)^{2b}.$$

A special case of particular interest is the C.M. for elastic scattering, where the η factors drop out and we simply have[11]

$$C_{su}(a + b \to a + b) = (-1)^{2a+2b} (2u + 1) \begin{Bmatrix} a & b & s \\ a & b & u \end{Bmatrix}. \quad (10')$$

It is worthwhile to mention that the phase in (10′) is such as to make each row sum to unity,[12] or equivalently, the last column elements being all positive. (The proof is given in Sec. V.)

[11] For an interesting connection of this formula to the generalized Pomeranchuk theorem, see C. N. Yang, J. Math. Phys. **4**, 52 (1963).

[12] This is always the case if we define all elastic amplitudes in the same way; for example, a natural choice would be that all such amplitudes have positive imaginary parts. Note the C.M. for NN^* and N^*N^* given by Ref. 3 have opposite phases compared to our result.

As one last remark about the general formula (10), we noted that the ratios[13]

$$C_{su}(a+b\to c+d)/C_{su}(a'+b\to c'+\bar{d}) = (-1)^{2d}$$

and

$$C_{su}(a+b\to c+d)/C_{su}(a'+\bar{b}\to c'+d) = (-1)^{2b}$$

(with $I_a = I_{a'}$ and $I_c = I_{c'}$) *are independent of any particle-phase convention.* This can be seen from the group-theoretic result[14] (A9). Taking nucleon-antinucleon scattering as an example, we must have

$$C_{su}(N+\bar{N}\to N+\bar{N})/C_{su}(N+\bar{N}\to \bar{N}+N) = -1.$$

In the following sections, it is convenient to consider a matrix $\mathcal{C}_{su}$ whose phase differs from that of C_{su} by the rule

$$\begin{aligned} C_{su} &= \eta_b\eta_d(-1)^{d+b+2s}(-1)^{a+b+c+d}\mathcal{C}_{su} \\ &= \eta_b\eta_d(-1)^{a+2b+3c+4d}\mathcal{C}_{su}. \end{aligned} \tag{11}$$

The last form is written for mnemonic reasons.

III. GENERAL PROPERTIES OF THE CROSSING MATRIX

From the last section, $\mathcal{C}$ is given by

$$\mathcal{C}_{su} = (-1)^{a+b+c+d}(2u+1)\begin{Bmatrix} a & b & s \\ c & d & u \end{Bmatrix}. \tag{12}$$

With this relation and the properties of $6j$ symbols, we then infer the following:

(A)

$$\mathcal{C}_{su} = \left(\frac{2u+1}{2s+1}\right)^{\frac{1}{2}} O_{su}, \tag{13}$$

where O is a real orthogonal matrix; thus,

$$\mathcal{C}_{su} = \left(\frac{2u+1}{2s+1}\right)\mathcal{C}^{-1}_{us}. \tag{14}$$

(B) The last row and last column of $\mathcal{C}$ are all positive whereas the first row and first column are alternating in sign.

(C) The last row can be expressed in terms of Clebsch-Gordan coefficients as follows: If $c+d \geq a+b$,

$$\mathcal{C}_{a+b,u} = \frac{\langle a, a; d, c-a-b \,|\, u\rangle\langle c, c; b, -b \,|\, u\rangle}{\langle c, c; d, a+b-c \,|\, a+b, a+b\rangle}. \tag{15}$$

If $a+b \geq c+d$,

$$\mathcal{C}_{c+d,u} = \frac{\langle c, c; b, a-c-d \,|\, u\rangle\langle a, a; d, -d \,|\, u\rangle}{\langle a, a; b, c+d-a \,|\, c+d, c+d\rangle}. \tag{16}$$

It is clear (15) and (16) are related by the exchange $a \leftrightarrow c$ and $b \leftrightarrow d$.

(D) Coupling (13) and (C), we can get the last column of $\mathcal{C}$ in terms of last row of $\mathcal{C}^{-1}$:

$$\mathcal{C}_{s,u_M} = \left(\frac{2u_M+1}{2s+1}\right)\mathcal{C}^{-1}_{u_M,s}, \tag{17}$$

where u_M is the maximum of u.

(E) If $I_b = I_{\bar{d}}$, i.e., the two crossed particles have the same isospin, then $\mathcal{C}^2 = 1$ and $\mathbf{O}$ is symmetrical.

(F) For the case $I_a = I_c$ and $I_b = I_d$, we have sum rules for the matrix elements

$$\sum_u \mathcal{C}_{su} = 1, \tag{18}$$

$$\sum_s (2s+1)\mathcal{C}_{su} = (2u+1), \tag{19}$$

$$\operatorname{Tr}\mathcal{C} = \sum_\alpha \mathcal{C}_{\alpha\alpha} = \begin{cases} 0, & \text{if } n \text{ is even,} \\ 1, & \text{if } n \text{ is odd,} \end{cases} \tag{20}$$

where n is the rank of $\mathcal{C}$.

(G) Again for the elastic case only, the elements in the last row are simply

$$\mathcal{C}_{a+b,u} = \langle a, a; b, -b \,|\, u, a-b\rangle^2, \tag{21}$$

while those in the first column are given by (without loss of generality we take $a \geq b$):

$$\mathcal{C}_{s,a-b} = (-1)^l\langle a, a-l; b, -b+l \,|\, a-b, a-b\rangle^2, \tag{22}$$

where $l \equiv (a+b)-s$, taking values from $0, 1, \cdots,$ to $2b$.

(H) Another special case worth mentioning is that when $I_a = I_{\bar{d}}$ and $I_b = I_c$, then

$$\mathcal{C}_{s0} = (-1)^{a+b+s}[(2a+1)(2b+1)]^{-\frac{1}{2}}. \tag{23}$$

The properties stated above are useful for constructing the C.M. Following are properties with interesting dynamical implications.

(I) For $\mathcal{C}_{su}(a+b\to a+b)$ and $a \geq b$:

(i) $\mathcal{C}_{a-b,a+b}$ is greater than any other element.

(ii) The elements in the last column decrease with increasing s.

(iii) The elements in the first column increase their magnitude with increasing s, for $a \neq b$, and become all equal in magnitude when $a = b$.

(iv) The elements in the first row increase in magnitude with increasing u.

[13] The formula given by Ref. 4 fails to give distinct crossing phases between crossed particles and antiparticles of half-integral isospin.

[14] For an interesting implication of this result applied to self-conjugate particles with half-integral isospin, see H. Lee, Phys. Rev. Letters **18**, 1098 (1967).

(J) For $C(a+b\to a+b)$ with $a/b \gg 1$, the anti-diagonal elements are close to unity to the order of b/a, while others are all vanishingly small in the order

$$C_{su} \approx O((b/a)^{|s+u-2a|});$$

and the signs of C_{su} are given by

$$\text{sgn } C_{su} = (-1)^{s+u-2a}, \quad \text{for} \quad s+u<2a,$$
$$\text{sgn } C_{su} = 1, \quad \text{for} \quad s+u \geq 2a.$$

(K) For $C_{su}(a+a\to b+b)$, the element with largest magnitude in each column is the one with $s=0$.

Properties (A), (E), and (H) are well known,[1] (K) has been proved by Masuda,[15] and the remaining ones are proved in Sec. V.

IV. METHOD OF CONSTRUCTION AND TABLES FOR CROSSING MATRICES

Since by (11) we can obtain the most general crossing matrix for different conventions and particle sets, it is sufficient for us to consider the C matrix for various set of isospins. From the relation (12) and symmetry properties of the $6j$ symbol, it is clear that $C(a+b\to c+d)$ and $C(b+a\to d+c)$ are identical, while $C(a+b\to c+d)$ and $C(a+d\to c+b)$ are inverse to each other. Hence for a given set of isospin a, b, c, and d there are at most two independent C matrices to be considered, namely, $C(a+b\to c+d)$ and $C(a+b\to d+c)$.

The procedures for constructing C matrices are as follows:

Case of rank one: By (13) we trivially obtain

$$C_{su} = \left(\frac{2u+1}{2s+1}\right)^{\frac{1}{2}}.$$

Case of rank two: By (13) and (B) we may put

$$C_{su} = \left(\frac{2u+1}{2s+1}\right)^{\frac{1}{2}} O_{su}$$

with

$$O = \begin{vmatrix} -\alpha & (1-\alpha^2)^{\frac{1}{2}} \\ (1-\alpha^2)^{\frac{1}{2}} & \alpha \end{vmatrix},$$

where α can be determined from any element in the last row by using (C).

Case of rank three or four: Although our method can cover all C.M. up to rank four, we consider only those in which either (or both) the s and u reaction is elastic (in the sense of isospin), since others necessarily involve particles with $I > \frac{3}{2}$, in which we are not interested. First consider $C(a+b\to a+b)$; the last row and first column are immediately obtained by using (G). [It is amusing to notice that the numbers given by (21) and (22) also appear in a column and row in the table of Clebsch–Gordan coefficients.] With (14) relating symmetrical elements about diagonal for the case $C = C^{-1}$, we easily get the first row and last column. The remaining elements are then determined by sum rules in (F).

As for the case $a+b\to b+a$, we use (C) and (D) to get the last row and column, and (H) to get the first row; then the rest can be determined by orthogonality constraints imposed on O_{su}.

The main scheme of the above procedures is quite clear: first we calculate the outermost layer of elements; then we use the orthogonality constraints together with sum rules to get the remaining elements. Since the number of parameters needed to describe an arbitrary real orthogonal matrix of rank 2, 3, and 4 is 1, 3, and 6, respectively, we see that the available information is more than necessary to obtain a unique result.

In fact it is interesting to note that for $C(a+\frac{1}{2}\to a+\frac{1}{2})$, the sum rules from (F) are alone sufficient to fix all elements. By (18) and (20), C can be parameterized as

$$C = \begin{vmatrix} -\alpha & 1+\alpha \\ 1-\alpha & \alpha \end{vmatrix}.$$

Now using (19) we have

$$2a(-\alpha) + (2a+2)(1-\alpha) = 2a,$$

so

$$\alpha = (2a+1)^{-1}.$$

The following tables for C.M.'s are listed in the order of increasing rank. The reaction written in each case is the s reaction for C and the u reaction for C^{-1}:

$\frac{1}{2}+\frac{1}{2}\to\frac{1}{2}+\frac{1}{2}$

$$C = C^{-1} = \begin{vmatrix} -\frac{1}{2} & \frac{3}{2} \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix};$$

$1+\frac{1}{2}\to 1+\frac{1}{2}$

$$C = C^{-1} = \begin{vmatrix} -\frac{1}{3} & \frac{4}{3} \\ \frac{2}{3} & \frac{1}{3} \end{vmatrix};$$

$\frac{3}{2}+\frac{1}{2}\to\frac{3}{2}+\frac{1}{2}$

$$C = C^{-1} = \begin{vmatrix} -\frac{1}{4} & \frac{5}{4} \\ \frac{3}{4} & \frac{1}{4} \end{vmatrix};$$

[15] N. Masuda, Progr. Theoret. Phys. (Kyoto) 33, 864 (1965).

$a+\frac{1}{2}\to a+\frac{1}{2}$

$$\mathcal{C}=\mathcal{C}^{-1}=\begin{vmatrix} -\dfrac{1}{2a+1} & \dfrac{2(a+1)}{2a+1} \\ \dfrac{2a}{2a+1} & \dfrac{1}{2a+1} \end{vmatrix};$$

$1+\frac{1}{2}\to\frac{1}{2}+1$

$$\mathcal{C}=\begin{vmatrix} -\dfrac{1}{\sqrt{6}} & 1 \\ \dfrac{1}{\sqrt{6}} & \dfrac{1}{2} \end{vmatrix},\quad \mathcal{C}^{-1}=\begin{vmatrix} -\dfrac{\sqrt{6}}{3} & \dfrac{2\sqrt{6}}{3} \\ \dfrac{2}{3} & \dfrac{2}{3} \end{vmatrix};$$

$\frac{3}{2}+\frac{1}{2}\to\frac{1}{2}+\frac{3}{2}$

$$\mathcal{C}=\begin{vmatrix} -\dfrac{\sqrt{2}}{4} & \dfrac{\sqrt{10}}{4} \\ \dfrac{\sqrt{2}}{4} & \dfrac{3}{2\sqrt{10}} \end{vmatrix},\quad \mathcal{C}^{-1}=\begin{vmatrix} -\dfrac{3\sqrt{2}}{4} & \dfrac{5\sqrt{2}}{4} \\ \dfrac{\sqrt{10}}{4} & \dfrac{\sqrt{10}}{4} \end{vmatrix};$$

$a+\frac{1}{2}\to\frac{1}{2}+a$

$$\mathcal{C}=\frac{1}{[2(2a+1)]^{\frac{1}{2}}}\begin{vmatrix} -1 & \left[\dfrac{3(a+1)}{a}\right]^{\frac{1}{2}} \\ 1 & \left[\dfrac{3a}{(a+1)}\right]^{\frac{1}{2}} \end{vmatrix},$$

$$\mathcal{C}^{-1}=\frac{1}{[2(2a+1)]^{\frac{1}{2}}}\begin{vmatrix} -2a & 2(a+1) \\ 2\left[\dfrac{a(a+1)}{3}\right]^{\frac{1}{2}} & 2\left[\dfrac{a(a+1)}{3}\right]^{\frac{1}{2}} \end{vmatrix};$$

$\frac{1}{2}+1\to1+\frac{3}{2}$

$$\mathcal{C}=\begin{vmatrix} -\dfrac{1}{2} & \dfrac{5(3)^{\frac{1}{2}}}{6} \\ \dfrac{(5)^{\frac{1}{2}}}{2(2)^{\frac{1}{2}}} & \dfrac{(5)^{\frac{1}{2}}}{2(6)^{\frac{1}{2}}} \end{vmatrix},\quad \mathcal{C}^{-1}=\begin{vmatrix} -\dfrac{1}{3} & \dfrac{(10)^{\frac{1}{2}}}{3} \\ \left(\dfrac{1}{3}\right)^{\frac{1}{2}} & \left(\dfrac{2}{15}\right)^{\frac{1}{2}} \end{vmatrix};$$

$\frac{1}{2}+1\to\frac{3}{2}+1$

$$\mathcal{C}=\mathcal{C}^{-1}=\begin{vmatrix} -\frac{2}{3} & \frac{1}{3}(10)^{\frac{1}{2}} \\ \frac{1}{6}(10)^{\frac{1}{2}} & \frac{2}{3} \end{vmatrix};$$

$\frac{1}{2}+\frac{3}{2}\to\frac{3}{2}+\frac{3}{2}$

$$\mathcal{C}=\mathcal{C}^{-1}=\begin{vmatrix} -\frac{1}{2} & \frac{1}{2}(5)^{\frac{1}{2}} \\ 3(20)^{-\frac{1}{2}} & \frac{1}{2} \end{vmatrix};$$

$1+1\to1+1$

$$\mathcal{C}=\mathcal{C}^{-1}=\begin{vmatrix} \frac{1}{3} & -1 & \frac{5}{3} \\ -\frac{1}{3} & \frac{1}{2} & \frac{5}{6} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{vmatrix};$$

$\frac{3}{2}+1\to\frac{3}{2}+1$

$$\mathcal{C}=\mathcal{C}^{-1}=\begin{vmatrix} \frac{1}{6} & -\frac{2}{3} & \frac{3}{2} \\ -\frac{1}{3} & \frac{11}{15} & \frac{3}{5} \\ \frac{1}{2} & \frac{2}{5} & \frac{1}{10} \end{vmatrix};$$

$a+1\to a+1$

$$\mathcal{C}=\mathcal{C}^{-1}=\begin{vmatrix} \dfrac{1}{a(2a+1)} & -\dfrac{1}{a} & \dfrac{2a+3}{2a+1} \\ -\dfrac{(2a-1)}{a(2a+1)} & \dfrac{a^2+a-1}{a(a+1)} & \dfrac{2a+3}{(a+1)(2a+1)} \\ \dfrac{2a-1}{2a+1} & \dfrac{1}{a+1} & \dfrac{1}{(a+1)(2a+1)} \end{vmatrix};$$

$\frac{3}{2}+1\to1+\frac{3}{2}$

$$\mathcal{C}=\begin{vmatrix} \frac{1}{6}(3)^{\frac{1}{2}} & -\frac{1}{4}(10)^{\frac{1}{2}} & \frac{5}{2}(6)^{-\frac{1}{2}} \\ -\frac{1}{6}(3)^{\frac{1}{2}} & (10)^{-\frac{1}{2}} & \frac{1}{3}(6)^{\frac{1}{2}} \\ \frac{1}{6}(3)^{\frac{1}{2}} & \frac{3}{2}(10)^{-\frac{1}{2}} & \frac{1}{12}(6)^{\frac{1}{2}} \end{vmatrix},$$

$$\mathcal{C}=\begin{vmatrix} 3^{-\frac{1}{2}} & -2(3)^{-\frac{1}{2}} & 3^{\frac{1}{2}} \\ -\frac{1}{6}(10)^{\frac{1}{2}} & \frac{2}{15}(10)^{\frac{1}{2}} & 3(10)^{-\frac{1}{2}} \\ 6^{-\frac{1}{2}} & \frac{4}{15}(6)^{\frac{1}{2}} & \frac{1}{10}(6)^{\frac{1}{2}} \end{vmatrix};$$

$a+1\to1+a$

$$\mathcal{C}=\frac{1}{[3(2a+1)]^{\frac{1}{2}}}\begin{vmatrix} 1 & -3\left[\dfrac{(a+1)}{2a}\right]^{\frac{1}{2}} & \left[\dfrac{5(a+1)(2a+3)}{2a(2a-1)}\right]^{\frac{1}{2}} \\ -1 & \dfrac{3}{[2a(a+1)]^{\frac{1}{2}}} & \left[\dfrac{5(2a-1)(2a+3)}{2a(a+1)}\right]^{\frac{1}{2}} \\ 1 & 3\left[\dfrac{a}{2(a+1)}\right]^{\frac{1}{2}} & \left[\dfrac{5a(2a-1)}{2(a+1)(2a+3)}\right]^{\frac{1}{2}} \end{vmatrix},$$

$$\mathcal{C}^{-1}=\frac{1}{[3(2a+1)]^{\frac{1}{2}}}\begin{vmatrix} (2a-1) & -(2a+1) & (2a+3) \\ -(2a-1)\left[\dfrac{(a+1)}{2a}\right]^{\frac{1}{2}} & \dfrac{(2a+1)}{[2a(a+1)]^{\frac{1}{2}}} & (2a+3)\left[\dfrac{a}{2(a+1)}\right]^{\frac{1}{2}} \\ \left[\dfrac{(2a-1)(a+1)(2a+3)}{10a}\right]^{\frac{1}{2}} & (2a+1)\left[\dfrac{(2a-1)(2a+3)}{10a(a+1)}\right]^{\frac{1}{2}} & \left[\dfrac{(2a-1)a(2a+3)}{10(a+1)}\right]^{\frac{1}{2}} \end{vmatrix};$$

$\frac{3}{2}+\frac{3}{2}\to\frac{3}{2}+\frac{3}{2}$

$$\mathcal{C}=\mathcal{C}^{-1}=\begin{vmatrix} -\frac{1}{4} & \frac{3}{4} & -\frac{5}{4} & \frac{7}{4} \\ \frac{1}{4} & -\frac{11}{20} & \frac{1}{4} & \frac{21}{20} \\ -\frac{1}{4} & \frac{3}{20} & \frac{3}{4} & \frac{7}{20} \\ \frac{1}{4} & \frac{9}{20} & \frac{1}{4} & \frac{1}{20} \end{vmatrix};$$

$a+\frac{3}{2}\to a+\frac{3}{2}$

$$\mathcal{C}=\mathcal{C}^{-1}=\frac{1}{(2a+1)}\begin{vmatrix} -\frac{3}{a(2a-1)} & \frac{1}{2a-1} & -\frac{3(a+1)}{a} & 2(a+2) \\ \frac{6(a-1)}{a(2a-1)} & -\frac{8a^2+4a-13}{(a+1)(2a-1)} & \frac{2(a^2+a-3)}{a} & \frac{3(a+2)}{(a+1)} \\ -\frac{3(a-1)}{a} & \frac{2(a^2+a-3)}{(a+1)} & \frac{8a^2+8a-9}{a(2a+3)} & \frac{6(a+2)}{(a+1)(2a+3)} \\ 2(a-1) & \frac{3a}{a+1} & \frac{6}{2a+3} & \frac{3}{(a+1)(2a+3)} \end{vmatrix}.$$

V. PROOFS OF THE GENERAL PROPERTIES

To prove (B) we use a formula[1]

$$\begin{Bmatrix} j_1 & j_2 & l_1+l_2 \\ l_1 & l_2 & l_3 \end{Bmatrix}=(-1)^{j_1+j_2+l_1+l_2}\left[\frac{(2l_1)!\,(2l_2)!\,(j_1+j_2+l_1+l_2+1)!}{(2l_1+2l_2+1)!\,(j_1+j_2-l_1-l_2)!}\right.$$
$$\left.\times\frac{(j_1+l_1+l_2-j_2)!\,(j_2+l_1+l_2-j_1)!\,(j_1+l_3-l_2)!\,(j_2+l_3-l_1)!}{(l_1+j_2+l_3+1)!\,(j_1+l_2-l_3)!\,(l_2+l_3-j_1)!\,(j_1+l_2+l_3+1)!\,(l_1+j_2-l_3)!\,(l_1+l_3-j_2)!}\right]^{\frac{1}{2}}. \quad (24)$$

Let s_M and u_M represent the maximum of the s and u values that appeared in the C.M., while s_m and u_m represent the corresponding minimum values. Since s_M is equal to the smaller value of $(a+b)$ and $(c+d)$, while u_M equals the smaller of $(a+d)$ and $(b+c)$, the $6j$ symbols

$$\begin{Bmatrix} a & b & s_M \\ c & d & u \end{Bmatrix} \text{ and } \begin{Bmatrix} a & b & s \\ c & d & u_M \end{Bmatrix}$$

in all cases can be put into the form (24) by use of the following symmetries:

$$\begin{Bmatrix} a & b & s \\ c & d & u \end{Bmatrix}=\begin{Bmatrix} c & d & s \\ a & b & u \end{Bmatrix}=\begin{Bmatrix} a & d & u \\ c & b & s \end{Bmatrix}=\begin{Bmatrix} c & b & u \\ a & d & s \end{Bmatrix}.$$

By (24) and (12) we immediately get

$$\operatorname{sgn}\mathcal{C}_{s_M u}=\operatorname{sgn}\mathcal{C}_{s u_M}=(-1)^{2a+2b+2c+2d}=1,$$

where $\operatorname{sgn}x\equiv x/|x|$.

On the other hand, s_m is the larger value of $|a-b|$ and $|c-d|$, u_m the larger of $|a-d|$ and $|b-c|$. For example, let us assume for definiteness that $s_m=a-b$; this implies that

$$a-b\geq|c-d|,$$

so

$$a+d\geq c+b=u_M.$$

Now,

$$\begin{Bmatrix} a & b & s_m \\ c & d & u \end{Bmatrix}=\begin{Bmatrix} a & b & a-b \\ c & d & u \end{Bmatrix}=\begin{Bmatrix} a-b & b & a \\ u & d & c \end{Bmatrix}=\begin{Bmatrix} u & d & a \\ a-b & b & c \end{Bmatrix}.$$

Applying (24) to the last form, we get the phase $(-1)^{a+d+u}$, hence

$$\operatorname{sgn}\mathcal{C}_{(a-b),u}=(-1)^{2a+b+c+2d+u}=(-1)^{u-b-c}=(-1)^{u-u_M}.$$

Similar considerations for all other cases lead to the same result, namely,

$$\operatorname{sgn}\mathcal{C}_{s_M,u}=(-1)^{u-u_M}$$

and

$$\operatorname{sgn}\mathcal{C}_{s,u_m}=(-1)^{s-s_M}.$$

This completes the proof of assertion (B).

One can verify (C) straightforwardly by direct substitution of the explicit expression for each entry in (15) and (16), with the use of (24) and[1]

$$\langle j_1, j_1; j_2, m-j_1 \mid j, m\rangle=\left[\frac{(2j+1)(2j_1)!\,(-j_1+j_2+j)!\,(j_1+j_2-m)!\,(j+m)!}{(j_1+j_2-j)!\,(j_1-j_2+j)!\,(j_1+j_2+j+1)!\,(-j_1+j_2+m)!\,(j-m)!}\right]^{\frac{1}{2}} \quad (25)$$

But we prefer the following proof in view of its physical transparency. Take the case $c + d \geq a + b$; we consider the following set of charge states,

$$|a, a\rangle + |b, b\rangle \to |c, c\rangle + |d, a + b - c\rangle. \quad (26)$$

Now the lhs is a pure $I_s = a + b$ state while the rhs has only a fraction

$$\langle c, c; d, a + b - c \,|\, a + b, a + b\rangle$$

of an $I_s = a + b$ state, so the whole expression represents

$$\langle c, c; d, a + b - c \,|\, a + b, a + b\rangle A_{s=a+b}.$$

On the other hand, the crossed reaction

$$|a, a\rangle + |\bar{d}, -a - b + c\rangle \to |c, c\rangle + |\bar{b}, -b\rangle \quad (27)$$

represents a linear combination of different I_u amplitudes, namely,

$$\sum_u \langle a, a; d, c - b - a \,|\, u\rangle\langle c, c; b, -b \,|\, u\rangle A_u.$$

Equating this to the previous expression we arrive at (15). Notice that if we restore the crossing phase $\eta(d, a + b - c)\eta(b, b)$ to the amplitude represented by (27) according to (3), then instead of (15) we automatically have an expression for $C_{a+b,u}$.

The sum rule (18) can be proved most easily as follows: Consider $a = c$ and $b = d$, then $C_{su} = \mathcal{C}_{su}$ by (11); now let A be an identity operator, from (4) and (6) we have $A_s = A_u = 1$; since C_{su} is independent of A, (8) is just reduced to $\sum_u \mathcal{C}_{su} = 1$.

To prove the sum rule (19) we use (13) and (18):

$$\begin{aligned}\sum_s (2s + 1)\mathcal{C}_{su} &= \sum_s [(2s + 1)(2u + 1)]^{\frac{1}{2}} O_{su} \\ &= \sum_s [(2u + 1)(2s + 1)]^{\frac{1}{2}} O_{us} \\ &= \sum_s (2u + 1)\mathcal{C}_{us} = (2u + 1).\end{aligned}$$

The proof of the sum rules (20) is more elaborate. First we note that $\mathrm{Tr}\, \mathcal{C} = \mathrm{Tr}\, O$ from (13), and $O^2 = I$ from (E). Since now the eigenvalues of O can only be ± 1 and its trace is invariant under diagonalization, we conclude that $\mathrm{Tr}\, \mathcal{C} =$ integer. Now using (J), which we shall prove later, we can put

$$\mathcal{C}_{su} = \delta_{s+u,2a} + B_{su}, \quad (28)$$

where $B_{su} \to 0$ as $a/b \to \infty$. So for sufficiently large a we see that (20) indeed follows. To extend the proof to all values of a, we fix b and consider $\mathcal{C}_{su}(a)$ as a function of the real variable a; it is actually a rational function of a without poles for $a \geq b$, as can be seen from the analytic expression[1] for the $6j$ symbol. So, what we have proved about the trace for large a can be analytically continued down to all values of a as low as b; thus we have (20).

To prove the statements in (I), we use (25) in (21) and (22) to obtain the analytic expression for $\mathcal{C}_{a+b,u}$ and $\mathcal{C}_{s,a-b}$, and then use (14) to get the expressions for $\mathcal{C}_{s,a+b}$ and $\mathcal{C}_{a-b,u}$. Taking ratios we have

$$\frac{\mathcal{C}_{s,a+b}}{\mathcal{C}_{s+1,a+b}} = \frac{(a + b + 2 + s)}{(a + b - s)} > 1, \quad (29)$$

$$\frac{|\mathcal{C}_{s,a-b}|}{|\mathcal{C}_{s+1,a-b}|} = \frac{(s + 1 + b - a)}{(s + 1 + a - b)} < 1, \quad \text{for } a > b,$$

$$= 1, \quad \text{for } a = b, \quad (30)$$

$$\frac{|\mathcal{C}_{a-b,u}|}{|\mathcal{C}_{a-b,u+1}|} = \left(\frac{2u + 1}{2u + 3}\right)\left(\frac{u + 1 + b - a}{u + 1 + a - b}\right) < 1. \quad (31)$$

These relations confirm (i) through (iii). To show that $\mathcal{C}_{a-b,a+b}$ is the largest element, we employ

$$\left|(2a + 1)\begin{Bmatrix} a & b & s \\ a & b & u \end{Bmatrix}\right| = |\mathcal{C}_{aa}(s + b \to u + b)|$$

$$= |O_{aa}(s + b \to u + b)| < 1,$$

so

$$\frac{|\mathcal{C}_{su}|}{\mathcal{C}_{a-b,a+b}} = (2u + 1)\left|\begin{Bmatrix} a & b & s \\ a & b & u \end{Bmatrix}\right| \left(\frac{2a + 2b + 1}{2a + 1}\right)^{-1}$$

$$< \left(\frac{2u + 1}{2a + 2b + 1}\right) \leq 1. \quad (32)$$

Finally, for property (J), we have assumed $a \gg b$, so $(a - b)/(a + b) \approx 1$ and $s \approx u \approx a$. (We use $\approx$ to mean approximation to the order of b/a.) Using the explicit formula for the $6j$ symbol, we can write

---

$$\mathcal{C}_{su} = (-1)^{2a+2b}(2u + 1)P(a\ b\ s)P(a\ b\ u)$$

$$\times \sum_z \frac{(-1)^z(z + 1)!}{[(z - a - b - u)!]^2} \frac{1}{[(z - a - b - s)!]^2(2a + 2b - z)!\,(2b + s + u - z)!\,(2a + s + u - z)!},$$

where

$$P(a\ b\ c) \equiv \frac{(-a + b + c)!\,(a - b + c)!\,(a + b - c)!}{(a + b + c + 1)!}$$

and z takes all integer values such that the arguments in the factorials are nonnegative. It is easy to see that

the term with highest z in the sum is the one with the leading power in a; thus (let $s \equiv a + m$, $u \equiv a + n$):

$$\mathcal{C}_{su} \approx (-1)^{2a-s-u} P(a\ b\ s)P(a\ b\ u)\frac{(2b+s+u+1)!}{[(s+b-a)!]^2[(u+b-a)!]^2(2a-s-u)!\,(2a-2b-1)!}$$

$$\approx (-1)^{2a-s-u}\frac{(b-m)!\,(b-n)!}{(b+m)!\,(b+n)!\,(-m-n)!}(2a)^{s+u-2a},$$

for $s + u \leq 2a$,

$$\mathcal{C}_{su} \approx P(a\ b\ s)P(a\ b\ u)\frac{(2a+2b+1)!}{[(a+b-s)]^2[(a+b-u)!]^2(s+u-2a)!\,(s+u-2b)!}$$

$$\approx \frac{(b+m)!\,(b+n)!}{(b-m)!\,(b-n)!\,(m+n)!}(2a)^{2a-s-u},$$

for $s + u \geq 2a$. The proof is completed.

Equation (21) is just a special case of (15), while Eq. (22) can be verified by direct substitution:

$$\frac{\mathcal{C}_{s,a-b}}{(2a-2b+1)} = (-1)^{2a+2b}\begin{Bmatrix} a & b & s \\ a & b & a-b \end{Bmatrix}$$

$$= (-1)^{2a+2b}\begin{Bmatrix} s & b & a \\ a-b & b & a \end{Bmatrix}$$

$$= (-1)^{s-a-b}\frac{(2b)!\,(s+a-b)!}{(2a+1)!\,(s+b-a)!}.$$

In the last step we make use of (24). On the other hand,

$$\langle a, a-l; b, -b+l \,|\, a-b, a-b\rangle^2$$

$$= \left(\frac{2a-2b+1}{2a+1}\right)\langle a-b, a-b; b, b-l \,|\, a, a-l\rangle^2$$

$$= (2a-2b+1)\left[\frac{(2b)!\,(2a-l)!}{(2a+1)!\,(2b-l)!}\right],$$

so this is just $(-1)^l\mathcal{C}_{s,a-b}$ if we put $l = (a + b) - s$.

VI. DISCUSSION IN CONNECTION WITH THE RECIPROCAL BOOTSTRAP

Several years ago Chew proposed a very attractive mechanism,[16] the so-called reciprocal bootstrap, for the self-consistent calculation of the N, Δ system. One of the main features in the theory is that the crossing matrix plays an important role in determining the forces. Roughly speaking, the force due to the exchange of N is strongest in the (3, 3) channel, while the force from the exchange of Δ is strongest in the (1, 1) channel. This results from the fact that the crossing matrix in question is such that the bottom (top) element in the first (last) column is positive and largest in magnitude. From (29) and (30) we see, in fact, all elastic matrices with $a \neq b$ exhibit this general character. Furthermore, whenever property (J) is applicable [actually the condition $a \gg b$ for its validity can be relaxed to $a/b > b$ for small b, as can be verified by the general formulas for $\mathcal{C}(a + b \to a + b)$ given in Sec. IV], we expect that the strongest force in each column (or row) is given by the antidiagonal elements, and the reciprocal bootstrap mechanism is likely to operate. The sign and rough magnitude of the forces produced by other elements can also be estimated without the explicit C.M.

However, if in a system the spin C.M. or the isospin C.M. or both are of the type with $a = b$, the situation is quite different due to property (K). In this case, the exchange of a particle with zero spin or zero isospin or both would be important in all the direct channels, and we do not expect this to be a reciprocal bootstrap system.

[16] G. F. Chew, Phys. Rev. Letters 9, 233 (1962).

ACKNOWLEDGMENT

I am deeply grateful to Professor G. F. Chew for his encouraging guidance and valuable comments.

APPENDIX

In the following, we work with the spinless case for simplicity; however, our result is not affected if we have spin. The most general crossing relation without constraint from (1) can be expressed as

$$\langle K_1|\, A\, |K_2; a_\alpha, p\rangle = \eta\,\langle K_1; \bar{a}_{-\alpha} - p|\, A\, |K_2\rangle \quad \text{(A1)}$$

and

$$\langle K_1'|\, A\, |K_2'; \bar{a}_{-\alpha}, -p\rangle = \eta\,\langle K_1'; a_\alpha, p|\, A\, |K_2'\rangle, \quad \text{(A2)}$$

where the K's represent particle sets which are not crossed, and $p(-p)$ is the 4-momentum of $a_\alpha(\bar{a}_{-\alpha})$. That the crossing phase η is independent of p but depends only on the crossed particle can be understood from Lorentz invariance. We shall not attempt to give the proof but refer to Ref. 6 to prove that the same η appears in both (A1) and (A2). We take the convention of writing η as $\eta(a, \alpha)$ when a particle a_α in the ket vector (initial state) is crossed; then $\eta^*(a, \alpha)$ is the crossing phase when a_α is crossed from a bra

vector (final state). From (A2) we have

$$\eta(\bar{a}, -\alpha) = \eta(a, \alpha). \tag{A3}$$

In particular, for a four-line connected part the crossing relation is

$$\langle c_\gamma d_\delta | A | a_\alpha b_\beta \rangle = \eta(b, \beta)\eta^*(d, \delta) \langle c_\gamma \bar{b}_{-\beta} | A | a_\alpha \bar{d}_{-\delta} \rangle. \tag{A4}$$

We define the matrix Λ^j corresponding to each set of $\eta(j, m)$ so that

$$\Lambda^j_{mm'} = \eta(j, m')\delta_{m,-m'}; \tag{A5}$$

clearly

$$\Lambda^\dagger = \Lambda^{-1}.$$

Now (A4) can be rewritten as

$$\langle c_\gamma d_\delta | A | a_\alpha b_\beta \rangle = \sum_{\beta'\delta'} \langle c_\gamma \bar{b}_{\beta'} | A | a_\alpha \bar{d}_{\delta'} \rangle \Lambda^b_{\beta'\beta}\Lambda^{d*}_{\delta'\delta}. \tag{A6}$$

We apply the transformation (1) on both sides of (A6) to obtain

$$\begin{aligned}
&\sum \langle c_{\gamma'} d_{\delta'} | A | a_{\alpha'} b_{\beta'} \rangle D_{\alpha'\alpha} D_{\beta'\beta} D^*_{\gamma'\gamma} D^*_{\delta'\delta} \\
&\quad = \sum \langle c_{\gamma'} \bar{b}_{\beta''} | A | a_{\alpha'} \bar{d}_{\delta''} \rangle \\
&\qquad \times D_{\alpha'\alpha} D^*_{\beta''\beta'} D^*_{\gamma'\gamma} D_{\delta''\delta'} \Lambda^b_{\beta'\beta}\Lambda^{d*}_{\delta'\delta} \\
&\quad = \sum \langle c_{\gamma'} d_{\delta'} | A | a_{\alpha'} b_{\beta'} \rangle \\
&\qquad \times D_{\alpha'\alpha}(\Lambda^{b-1} D^* \Lambda^b)_{\beta'\beta} D^*_{\gamma'\gamma}(\Lambda^{d-1} D^* \Lambda^d)^*_{\delta'\delta},
\end{aligned}$$

where primed indices are to be summed over. Comparing both ends of the last equality we conclude that

$$\Lambda^{b-1} D^{b*}(u) \Lambda^b = D^b(u), \tag{A7}$$

$$\Lambda^{d-1} D^{d*}(u) \Lambda^d = D^d(u). \tag{A8}$$

First we note that any two unitary matrices, say Λ_1 and Λ_2, both of which satisfy either (A6) or (A7), can only differ by a phase. Since the relation $D^j(u) = \Delta_1^{-1} D^{j*}(u)\Lambda_1 = \Lambda_1^{-1} D^{j*}(u)\Lambda_2$ implies

$$[\Lambda_2^{-1}\Lambda_1, D^j(u)] = 0$$

for all u, by Schur's lemma[17] we get $\Lambda_2^{-1}\Lambda_1 = aI$. Taking the determinant of this equation on both sides yields $|a| = 1$. It is well known[1] that

$$d^j(\pi) D^{j*}(u)\, d^{j-1}(\pi) = D^j(u),$$

where $d^j_{mm'}(\pi) = (-1)^{j+m'}\delta_{m,-m'}$, so in general $\Lambda^j_{mm'} = \eta_j(-1)^{m'}\delta_{m,-m'}$, with η_j an arbitrary phase independent of m'. From (A4) we immediately get $\eta(b, \beta) = \eta_b(-1)^\beta$ and $\eta(d, \delta) = \eta_{\bar{d}}(-1)^\delta$. It is clear that (A3) implies

$$\eta_{\bar{a}} = (-1)^{2a}\eta_a, \tag{A9}$$

so we see it is an intrinsic characteristic that η_a and $\eta_{\bar{a}}$ have opposite signs when $2a$ is an odd integer.[14]

[17] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Partial Separation of the Schrödinger Equation for Two Charged Particles in a Magnetic Field*

BENJAMIN P. CARTER
Lawrence Radiation Laboratory, University of California, Livermore, California

(Received 2 August 1968)

An effective Hamiltonian is derived which describes the relative motion of two oppositely charged particles in a homogeneous magnetostatic field.

I. INTRODUCTION

In what follows, we consider the motion of a system of two charged particles in a constant, homogeneous magnetic field **B**. The central problem in the theory of the Zeeman effect for such a system is to compute the stationary states and energy levels as a function of $|\mathbf{B}|$. This is usually done relativistically to various orders in $\alpha = \frac{1}{137}$ and m_1/m_2, where m_1 and m_2 are the masses of the two particles. For definiteness, we always take particle 1 to be an electron, with charge $e_1 = -e$. The charge e_2 of the second particle will at first be left arbitrary, but certain of our results hold only for the case that the system as a whole is neutral, i.e., that $e_2 = -e_1$.

For the case that particle 2 is a proton ($m_2 = 1836\, m_1$), Lamb[1] has computed the Zeeman levels to first order in m_1/m_2 using the relativistic Breit equation. The result is what one would expect from simple physical arguments[2]: namely, that the term $\mu_0\mathbf{B}\cdot\mathbf{L}$ must be replaced by $(1 - m_1/m_2)\mu_0\mathbf{B}\cdot\mathbf{L}$, where μ_0 is the Bohr magneton and **L** the orbital angular momentum. The nonrelativistic result which we obtain below agrees with Lamb's result to lowest order in the mass ratio.

For systems with mass ratios closer to unity, such as muonium ($m_2 \approx 200\, m_1$) or $p\mu^-$ ($m_2 \approx 9\, m_1$), terms of higher order in the mass ratio are at least as important as terms of higher order in α. In the extreme case of positronium ($m_1 = m_2$), the expansion in powers of the mass ratio is not useful. As a first step toward a more satisfactory treatment of positronium and other systems with mass ratios close to unity, we treat the nonrelativistic Schrödinger equation in such a way that the dependence of the solutions on m_1/m_2 is made explicit to all orders, i.e., exactly. The main purpose of this nonrelativistic treatment is to provide insight regarding the mixing of relative and center-of-mass (CM) variables which occurs in the nonrelativistic Hamiltonian through the vector potential.

For the case of a neutral system ($e_1 = -e_2$), we find an exact solution to a somewhat tricky problem in elementary quantum mechanics. The problem is to separate out the CM variables and obtain an effective Hamiltonian which operates on functions of the relative coordinates. The effective Hamiltonian is formally identical to that for a particle of a different (reduced) mass moving in a different magnetic field and in a harmonic-oscillator potential. The latter has terms which represent the kinetic energy of the CM motion, and terms which represent the Stark effect due to the electric field which would be seen by a comoving observer.

II. DERIVATION OF THE HAMILTONIAN IN RELATIVE AND CM COORDINATES

The Hamiltonian for our problem is

$$H = \frac{\boldsymbol{\Pi}_1^2}{2m_1} + \frac{\boldsymbol{\Pi}_2^2}{2m_2} + \frac{e_1e_2}{|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{1}$$

where $\boldsymbol{\Pi}_i$ is the kinetic momentum of the ith particle:

$$\boldsymbol{\Pi}_i = -i\hbar\nabla_i - (e_i/c)\mathbf{A}(\mathbf{r}_i), \quad i = 1, 2. \tag{2}$$

We choose the z axis of our coordinate system in the direction of the magnetic field **B**. It is convenient at this point not to specify the gauge completely, but we assume that each of the three components of $\mathbf{A}(\mathbf{r})$ is a linear combination of x, y, and z. The magnetic field does not depend on position or time, and satisfies

$$\mathbf{B} = \nabla \times \mathbf{A}(\mathbf{r}). \tag{3}$$

The CM coordinates $\mathbf{R} = (X, Y, Z)$ and the relative coordinates $\mathbf{r} = (x, y, z)$ are related to the positions $\mathbf{r}_i$ by

$$\begin{aligned} \mathbf{r} &= \mathbf{r}_2 - \mathbf{r}_1, \\ \mathbf{R} &= \alpha\mathbf{r}_1 + \beta\mathbf{r}_2, \end{aligned} \tag{4}$$

where $\alpha = m_1/(m_1 + m_2)$ and $\beta = m_2/(m_1 + m_2)$. The inverse relations are

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{R} - \beta\mathbf{r}; \\ \mathbf{r}_2 &= \mathbf{R} + \alpha\mathbf{r}. \end{aligned} \tag{5}$$

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] W. E. Lamb, Phys. Rev. **85**, 259 (1952).

[2] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic Press Inc., New York, 1957), p. 214.

By virtue of the requirement that **A** be linear in **r**, we find that **A** transforms like **r** under the coordinate transformation (5):

$$\mathbf{A}(\mathbf{r}_1) = \mathbf{A}(\mathbf{R}) - \beta\mathbf{A}(\mathbf{r}),$$
$$\mathbf{A}(\mathbf{r}_2) = \mathbf{A}(\mathbf{R}) + \alpha\mathbf{A}(\mathbf{r}). \tag{6}$$

The gradients

$$\nabla_R = \left(\frac{\partial}{\partial X}, \frac{\partial}{\partial Y}, \frac{\partial}{\partial Z}\right)$$

and

$$\nabla_r = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$$

are related to those which appear in Eq. (2) by

$$\nabla_1 = -\nabla_r + \alpha\nabla_R,$$
$$\nabla_2 = \nabla_r + \beta\nabla_R. \tag{7}$$

From Eqs. (1), (2), (5), (6), and (7) we obtain the desired expression for H:

$$H = -\frac{\hbar^2}{2\mu}\nabla_r^2 - \frac{\hbar^2}{2M}\nabla_R^2 + \frac{e_1e_2}{r} + \frac{1}{2c^2} \times [DA^2(\mathbf{R}) + 2EA(\mathbf{R})\cdot\mathbf{A}(\mathbf{r}) + FA^2(\mathbf{r})] + \frac{ih}{c}\left\{G[\mathbf{A}(\mathbf{R})\cdot\nabla_r + \frac{\mu}{M}\mathbf{A}(\mathbf{r})\cdot\nabla_R] + IA(\mathbf{R})\cdot\nabla_R + J\mathbf{A}(\mathbf{r})\cdot\nabla_r\right\}. \tag{8}$$

In obtaining Eq. (8) we have assumed that the order of factors is immaterial in dot products between **A**'s and **∇**'s. To insure that this is the case, we restrict ourselves to gauges in which

$$\frac{\partial}{\partial x}A_x(\mathbf{r}) = \frac{\partial}{\partial y}A_y(\mathbf{r}) = \frac{\partial}{\partial z}A_z(\mathbf{r}) = 0. \tag{9}$$

Besides the usual total mass $M = m_1 + m_2$ and reduced mass $\mu = m_1m_2/M$, the other new constants in (8) are

$$\begin{aligned} D &= e_1^2/m_1 + e_2^2/m_2, \\ E &= \alpha e_2^2/m_2 - \beta e_1^2/m_1 \\ &= (m_1/M)e_2^2/m_2 - (m_2/M)e_1^2/m_1, \\ F &= \beta^2 e_1^2/m_1 + \alpha^2 e_2^2/m_2 \\ &= (m_2/M)^2 e_1^2/m_1 + (m_1/M)^2 e_2^2/m_2, \\ G &= e_2/m_2 - e_1/m_1, \\ I &= (e_1 + e_2)/M, \\ J &= (1/M)(e_1m_2/m_1 + e_2m_1/m_2). \end{aligned} \tag{10}$$

III. USE OF A GAUGE TRANSFORMATION TO FIND CONSTANTS OF MOTION

Up to now we have left the gauge arbitrary, subject to the restrictions mentioned above. We now consider two candidates for the vector potential, both of which are linear in **r** and satisfy Eqs. (3) and (9). For our first candidate we choose $A_y(\mathbf{r}) = Bx$, $A_x = A_z = 0$. The advantage of this choice is that the spatial dependence of **A** is on only one coordinate, so that it is easy to find coordinates which do not appear in the Hamiltonian. Landau[3] used such a gauge to solve the problem of a single charge moving in a magnetic field.

Our second candidate is the one which preserves the cylindrical symmetry of the problem:

$$\mathbf{A}'(\mathbf{r}) = -\tfrac{1}{2}\mathbf{r}\times\mathbf{B}. \tag{11}$$

Its components are $A_y' = \frac{1}{2}Bx$, $A_x' = -\frac{1}{2}By$, $A_z' = 0$. All quantities expressed in this gauge are given primes. The primed and unprimed vector potentials are related by

$$\mathbf{A}'(\mathbf{r}) = \mathbf{A}(\mathbf{r}) - \nabla F(\mathbf{r}), \tag{12}$$

where

$$F(\mathbf{r}) = \tfrac{1}{2}Bxy. \tag{13}$$

If $\Psi(\mathbf{r}_1, \mathbf{r}_2)$ represents a state in the unprimed gauge, the same state is represented by

$$\Psi'(\mathbf{r}_1, \mathbf{r}_2) = \exp\left[-(i/\hbar c)\sum e_iF(\mathbf{r}_i)\right]\Psi(\mathbf{r}_1, \mathbf{r}_2) \tag{14}$$

in the primed gauge. This transformation of wavefunctions insures that

$$\mathbf{\Pi}_i'(\mathbf{r}_i)\Psi' = [\mathbf{\Pi}_i(\mathbf{r}_i)\Psi]' \tag{15}$$

and, hence, that $H'\Psi' = (H\Psi)'$, where H' and $\mathbf{\Pi}_i'$ are obtained from H and $\mathbf{\Pi}_i$ by replacing **A** by **A**′. More generally, the transformation of an arbitrary operator from the unprimed to the primed gauge is as follows:

$$(\text{Operator})' = \exp\left[-(i/\hbar c)\sum e_iF(\mathbf{r}_i)\right] \times (\text{Operator})\exp\left[(i/\hbar c)\sum e_iF(\mathbf{r}_i)\right]. \tag{16}$$

Our notation is slightly inconsistent in that **A** and **A**′ are different operators according to Eq. (12), rather than representations of the same operator in different gauges. But, except for **A** and **A**′, all primed and unprimed operators are related by Eq. (16).

Inserting the (unprimed) vector potential **A** into Eq. (8) and setting the momentum conjugate to Z equal to zero, we obtain

$$H = -\frac{\hbar^2}{2\mu}\nabla_r^2 - \frac{\hbar^2}{2M}\left(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2}\right) + \frac{e_1e_2}{r} + \frac{B^2}{2c^2}(DX^2 + 2EXx + Fx^2) + \frac{i\hbar B}{c} \times \left[G\left(X\frac{\partial}{\partial y} + \frac{\mu}{M}x\frac{\partial}{\partial Y}\right) + IX\frac{\partial}{\partial Y} + Jx\frac{\partial}{\partial y}\right]. \tag{17}$$

[3] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1958), p. 474, and the reference cited therein.

Since Y does not appear explicitly in Eq. (17), the operator

$$\Pi_Y = \frac{\hbar}{i}\frac{\partial}{\partial Y} \tag{18}$$

commutes with H, and can be taken as a constant of motion. Applying the gauge transformation (16) to the right-hand side of (18), we obtain

$$\Pi'_Y = \Pi_Y + \frac{1}{c}\frac{\partial \Lambda}{\partial Y}, \tag{19}$$

where

$$\Lambda = e_1 F(\mathbf{r}_1) + e_2 F(\mathbf{r}_2). \tag{20}$$

From Eqs. (5), (13), and (20), we obtain

$$\Lambda = \tfrac{1}{2}B[e_1(X - \beta x)(Y - \beta y) + e_2(X + \alpha x)(Y + \alpha y)]. \tag{21}$$

From Eqs. (18), (19), and (21), and the definitions of α and β, we obtain

$$\Pi'_Y = \frac{\hbar}{i}\frac{\partial}{\partial Y} + \frac{B}{2c}\left[(e_1 + e_2)X + \left(\frac{m_1 e_2 - m_2 e_1}{m_1 + m_2}\right)x\right]. \tag{22}$$

By the symmetry of the primed gauge, we may perform a rotation through 90° about the z axis. Replacing Y by X, y by x, X by $-Y$, and x by $-y$ in Eqs. (19) and (22), we obtain another operator which commutes with H:

$$\Pi'_X = \Pi_X + \frac{1}{c}\frac{\partial \Lambda}{\partial X} = \frac{\hbar}{i}\frac{\partial}{\partial X} - \frac{B}{2c}\left[(e_1 + e_2)Y + \frac{m_1 e_2 - m_2 e_1}{m_1 + m_2}\,y\right]. \tag{23}$$

From Eq. (21) and the second of Eqs. (23), we obtain

$$\Pi_X = \frac{\hbar}{i}\frac{\partial}{\partial X} - \frac{B}{c}\left[(e_1 + e_2)Y + \frac{m_1 e_2 - m_2 e_1}{m_1 + m_2}\,y\right]. \tag{24}$$

It is worthwhile to verify directly from Eqs. (17) and (24) that Π_X commutes with H. But Π_X and Π_Y cannot, in general, both be constants of motion. Their commutator

$$[\Pi_X, \Pi_Y] = (B\hbar/ci)(e_1 + e_2) \tag{25}$$

is zero only if the over-all charge of the system is zero. This reflects the fact that the system as a whole moves in a straight line only if its charge is zero.[4,5]

IV. SEPARATION OF VARIABLES FOR A NEUTRAL SYSTEM

We now specialize to the case of a neutral system. Substituting $e_1 = -e$ and $e_2 = +e$ into Eqs. (10), we obtain

$$\begin{aligned} D &= e^2/\mu, \\ E &= (e^2/M)(m_1/m_2 - m_2/m_1), \\ F &= e^2(1/\mu - 3/M), \\ G &= e/\mu, \\ I &= 0, \\ J &= (e/M)(m_1/m_2 - m_2/m_1). \end{aligned} \tag{26}$$

The most general simultaneous eigenstate of H, Π_X, and Π_Y is of the form

$$\Psi(X, Y, \mathbf{r}) = \exp\left[\frac{i}{\hbar}\left(P_X + \frac{eB}{c}y\right)X\right] \times \exp\left(\frac{i}{\hbar}P_Y Y\right)\psi(\mathbf{r}), \tag{27}$$

where P_X and P_Y are the eigenvalues of Π_X and Π_Y. The Schrödinger equation $H\Psi = E\Psi$ will be satisfied if and only if $h\psi(\mathbf{r}) = E\psi(\mathbf{r})$, where h is defined by

$$H\Psi(X, Y, \mathbf{r}) = \exp\left[\frac{i}{\hbar}\left(P_X + \frac{eB}{c}y\right)X\right] \times \exp\left(\frac{i}{\hbar}P_Y Y\right)h\psi(\mathbf{r}). \tag{28}$$

It is clear that Eq. (28) is satisfied if h is obtained from H by making the following replacements:

$$\begin{aligned} \frac{\hbar}{i}\frac{\partial}{\partial Y} &\to P_Y, \\ \frac{\hbar}{i}\frac{\partial}{\partial X} &\to P_X + \frac{eB}{c}y, \\ \frac{\hbar}{i}\frac{\partial}{\partial y} &\to \frac{eB}{c}X + \frac{\hbar}{i}\frac{\partial}{\partial y}. \end{aligned} \tag{29}$$

Making these replacements in the right-hand side of Eq. (17) and using Eqs. (26), we obtain[6]

$$h = -\frac{\hbar^2}{2\mu}\nabla^2 - \frac{e^2}{r} + \frac{1}{2M} \times \left[\left(P_X + \frac{eB}{c}y\right)^2 + \left(P_Y - \frac{eB}{c}x\right)^2\right] + \left(\frac{1}{\mu} - \frac{4}{M}\right)\frac{e^2B^2}{2c^2}x^2 - \frac{eB}{Mc}\left(\frac{m_1}{m_2} - \frac{m_2}{m_1}\right)x\frac{\hbar}{i}\frac{\partial}{\partial y}. \tag{30}$$

[4] B. P. Carter, Report #UCID-15180, Lawrence Radiation Laboratory, University of California, Livermore, California, 1967.

[5] In Ref. 4, constants of motion are found for a neutral system containing an arbitrary number of particles in a magnetic field.

[6] From Eq. (30) on, we dispense with subscripts R and r, it being understood that all gradients operate on the relative coordinates.

We call h the effective Hamiltonian for the relative motion of the two particles, since h operates on functions of the relative coordinates and does not involve CM coordinates or their derivatives. We emphasize that P_X and P_Y in (30) are just numbers. Equation (30) may be compared to the radial Schrödinger equation for a particle in a central potential. In this comparison, P_X and P_Y are analogous to the orbital angular momentum L which appears in the centrifugal potential.

We may rewrite Eq. (30) in the suggestive form

$$h = \frac{1}{2\mu}\left(\frac{\hbar}{i}\nabla + \frac{e}{c}\mathbf{a}\right)^2 + \frac{\mathbf{P}^2}{2M} - \frac{e^2}{r} - \frac{1}{c}e\mathbf{r}\cdot(\mathbf{v}\times\mathbf{B}) + \frac{e^2B^2}{2Mc^2}(x^2+y^2), \quad (31)$$

where $\mathbf{a}$ is a vector potential for the effective magnetic field

$$\mathbf{b} = \frac{m_2 - m_1}{m_2 + m_1}\mathbf{B} = \nabla \times \mathbf{a} \quad (32)$$

and $\mathbf{v} = \mathbf{P}/M$ is the velocity of the CM of the system. If we had not set the momentum P_Z conjugate to Z equal to zero, Eq. (31) would still be valid for $\mathbf{P} = (P_X, P_Y, P_Z)$. The term in Eq. (31) involving $\mathbf{v}$ is just the potential for the electric field

$$\mathbf{E} = \mathbf{v}\times\mathbf{B}/c \quad (33)$$

which is present for a comoving observer, i.e., one for whom the CM is at rest and who is moving with a velocity $\mathbf{v}$ relative to the lab frame. We note in passing that we may easily generalize our treatment to the case of crossed electric and magnetic fields in the lab frame by transforming to a frame in which the electric field vanishes.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Classical Fields on Spacelike Mass Shells

W. RÜHL*
The Rockefeller University, New York, New York

(Received 27 February 1967)

We define classical fields, corresponding to unitary representations of the inhomogeneous Lorentz group with $M^2 < 0$, which belong to the discrete series. These fields satisfy Bargmann–Wigner equations which are given in explicit matrix form.

1. INTRODUCTION

The aim of this article is to define and describe classical fields which satisfy the equation

$$(\Box_x - 1)\psi(x) = 0,$$

where

$$\Box_x = \frac{\partial^2}{\partial x_0^2} - \sum_{k=1}^{3}\frac{\partial^2}{\partial x_k^2}.$$

These fields possess a spin degree of freedom which is described by spin indices (α, u). α is discrete and ranges over all nonnegative integers. u is continuous and it describes certain rest classes in $SL(2, C)$. The indices (α, u) transform as a nonunitary representation of the homogeneous Lorentz group. The fields can, moreover, be constructed in pairs $\psi_\alpha(x, u)$, $\psi^\alpha(x, u)$ which satisfy Bargmann–Wigner equations:

$$\sum_{\beta=0}^{\infty} P^{\alpha\beta}(p, u)\psi_\beta(p, u) = \psi^\alpha(p, u),$$

$$\sum_{\beta=0}^{\infty} \hat{P}_{\alpha\beta}(p, u)\psi^\beta(p, u) = \psi_\alpha(p, u)$$

(in momentum space). The matrices P and $\hat{P}$ are infinite matrices of triangular shape with elements which are constructed later in this article.

The problem whether objects belonging to masses $M^2 < 0$ might be physically relevant and possibly lead to a new kind of field theory has been studied recently.[1] We intend to contribute to this discussion only by enlarging the number of classical objects which can perhaps be submitted to a later quantization and by deriving field equations, which in the case of spin-$\frac{1}{2}$ particles proved so convenient for the description of electromagnetic interactions.

Fields on timelike mass shells, the familiar objects of field theory, can be defined in several fashions. The best known method is due to Bargmann and Wigner[2] who base their construction on the theory of unitary representations of the inhomogeneous Lorentz group. For spin S and $p_0 > 0$, an irreducible representation

* Permanent address: CERN, Geneva.

[1] G. Feinberg, Phys. Rev. **159**, 1089 (1967).

[2] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. US **34**, 211 (1946).

We call h the effective Hamiltonian for the relative motion of the two particles, since h operates on functions of the relative coordinates and does not involve CM coordinates or their derivatives. We emphasize that P_X and P_Y in (30) are just numbers. Equation (30) may be compared to the radial Schrödinger equation for a particle in a central potential. In this comparison, P_X and P_Y are analogous to the orbital angular momentum L which appears in the centrifugal potential.

We may rewrite Eq. (30) in the suggestive form

$$h = \frac{1}{2\mu}\left(\frac{\hbar}{i}\nabla + \frac{e}{c}\mathbf{a}\right)^2 + \frac{\mathbf{P}^2}{2M} - \frac{e^2}{r} - \frac{1}{c}e\mathbf{r}\cdot(\mathbf{v}\times\mathbf{B}) + \frac{e^2B^2}{2Mc^2}(x^2+y^2), \quad (31)$$

where $\mathbf{a}$ is a vector potential for the effective magnetic field

$$\mathbf{b} = \frac{m_2 - m_1}{m_2 + m_1}\mathbf{B} = \nabla \times \mathbf{a} \quad (32)$$

and $\mathbf{v} = \mathbf{P}/M$ is the velocity of the CM of the system. If we had not set the momentum P_Z conjugate to Z equal to zero, Eq. (31) would still be valid for $\mathbf{P} = (P_X, P_Y, P_Z)$. The term in Eq. (31) involving $\mathbf{v}$ is just the potential for the electric field

$$\mathbf{E} = \mathbf{v}\times\mathbf{B}/c \quad (33)$$

which is present for a comoving observer, i.e., one for whom the CM is at rest and who is moving with a velocity $\mathbf{v}$ relative to the lab frame. We note in passing that we may easily generalize our treatment to the case of crossed electric and magnetic fields in the lab frame by transforming to a frame in which the electric field vanishes.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 4 APRIL 1969

Classical Fields on Spacelike Mass Shells

W. RÜHL*
The Rockefeller University, New York, New York

(Received 27 February 1967)

We define classical fields, corresponding to unitary representations of the inhomogeneous Lorentz group with $M^2 < 0$, which belong to the discrete series. These fields satisfy Bargmann–Wigner equations which are given in explicit matrix form.

1. INTRODUCTION

The aim of this article is to define and describe classical fields which satisfy the equation

$$(\Box_x - 1)\psi(x) = 0,$$

where

$$\Box_x = \frac{\partial^2}{\partial x_0^2} - \sum_{k=1}^{3}\frac{\partial^2}{\partial x_k^2}.$$

These fields possess a spin degree of freedom which is described by spin indices (α, u). α is discrete and ranges over all nonnegative integers. u is continuous and it describes certain rest classes in $SL(2, C)$. The indices (α, u) transform as a nonunitary representation of the homogeneous Lorentz group. The fields can, moreover, be constructed in pairs $\psi_\alpha(x, u)$, $\psi^\alpha(x, u)$ which satisfy Bargmann–Wigner equations:

$$\sum_{\beta=0}^{\infty} P^{\alpha\beta}(p, u)\psi_\beta(p, u) = \psi^\alpha(p, u),$$

$$\sum_{\beta=0}^{\infty} \hat{P}_{\alpha\beta}(p, u)\psi^\beta(p, u) = \psi_\alpha(p, u)$$

(in momentum space). The matrices P and $\hat{P}$ are infinite matrices of triangular shape with elements which are constructed later in this article.

The problem whether objects belonging to masses $M^2 < 0$ might be physically relevant and possibly lead to a new kind of field theory has been studied recently.[1] We intend to contribute to this discussion only by enlarging the number of classical objects which can perhaps be submitted to a later quantization and by deriving field equations, which in the case of spin-$\frac{1}{2}$ particles proved so convenient for the description of electromagnetic interactions.

Fields on timelike mass shells, the familiar objects of field theory, can be defined in several fashions. The best known method is due to Bargmann and Wigner[2] who base their construction on the theory of unitary representations of the inhomogeneous Lorentz group. For spin S and $p_0 > 0$, an irreducible representation

* Permanent address: CERN, Geneva.

[1] G. Feinberg, Phys. Rev. **159**, 1089 (1967).

[2] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. US **34**, 211 (1946).

can be defined on the Hilbert space

$$\mathcal{H} = \Big\{\varphi_m(p): \sum_{m=-s}^{+s} \int |\varphi_m(p)|^2 \times \delta(p^2 - M^2)\theta(p_0)\, d^4p < \infty\Big\}$$

by

$$U(a, x)\varphi_m(p) = \sum_{m'=-s}^{+s} e^{ipx} D_{mm'}[R(a, p)]\varphi_{m'}(p'),$$

where

$$p' = a^{-1}p(a^{-1})^+,$$
$$R(a, p) = a(p)aa(p')^{-1} \in SU(2),$$

and

$$a(p)pa(p)^+ = M.$$

$a(p)$ is as usual a rotation-free boost. The essential step is the continuation of the representation $D_{mm'}(u)$, $u \in SU(2)$, to a representation of the whole homogeneous group $SL(2, C)$. This can be achieved in two ways:

$u_{\alpha\beta}$ substituted by $a_{\alpha\beta}$

or

$u_{\alpha\beta}$ substituted by $(a^{-1})^+_{\alpha\beta}$.

Both substitutions lead to representations of $SL(2, C)$ because $D_{mm'}(u)$ is a polynomial in $u_{\alpha\beta}$. Because the coefficients of this polynomial are real, we have moreover

$$D((a^{-1})^+)_{mm'} = \overline{D(a^{-1})_{m'm}} = D^{-1,+}(a)_{mm'}.$$

We define

$$\psi_\alpha(p) = \sum_m D_{\alpha m}(a(p)^{-1})\varphi_m(p),$$
$$\psi^\alpha(p) = \sum_m D_{\alpha m}(a(p)^+)\varphi_m(p),$$

thus obtaining quantities with simple transformation properties:

$$U_a\psi_\alpha(p) = \sum_\beta D_{\alpha\beta}(a)\psi_\beta(p'),$$
$$U_a\psi^\alpha(p) = \sum_\beta D_{\alpha\beta}((a^{-1})^+)\psi^\beta(p').$$

With

$$P^{\alpha\beta}(p) = \sum_m D_{\alpha m}(a(p)^+)D_{m\beta}(a(p)) = D_{\alpha\beta}((p/M)^{-1}),$$
$$\hat{P}_{\alpha\beta}(p) = \sum_m D_{\alpha m}(a(p)^{-1})D_{m\beta}((a(p)^{-1})^+) = D_{\alpha\beta}(p/M),$$

we get the Bargmann–Wigner equations

$$\sum_\beta P^{\alpha\beta}(p)\psi_\beta(p) = \psi^\alpha(p),$$
$$\sum_\beta \hat{P}_{\alpha\beta}(p)\psi^\beta(p) = \psi_\alpha(p),$$

where

$$\sum_\beta P^{\alpha\beta}\hat{P}_{\beta\gamma} = \delta^\alpha_\gamma.$$

After a Fourier transformation this yields fields $\psi_\alpha(x)$, $\psi^\alpha(x)$ and differential operators

$$P^{\alpha\beta}\Big(-i\frac{\partial}{\partial x}\Big), \quad \hat{P}_{\alpha\beta}\Big(-i\frac{\partial}{\partial x}\Big)$$

of order $2S + 1$. Local bilinear invariant densities can easily be found, e.g.,

$$\sum_\alpha \bar{\psi}^\alpha(x)\psi_\alpha(x).$$

Requiring, on the other hand, only the Bargmann–Wigner equations to hold for $\psi_\alpha(x)$ and $\psi^\alpha(x)$ together with the Klein–Gordon equation

$$(\Box_x + M^2)\psi(x) = 0, \quad \psi = \psi_\alpha \text{ or } \psi^\alpha,$$

yields a field which consists of two irreducible representations, one positive timelike, the other negative timelike. The equations

$$\sum_\gamma \Big(\sum_\beta P^{\alpha\beta}\Big(-i\frac{\partial}{\partial x}\Big)\hat{P}_{\beta\dot{\gamma}}\Big(-i\frac{\partial}{\partial x}\Big) - \delta^\alpha_{\dot{\gamma}}\Big)\psi^{\dot{\gamma}} = 0, \quad \text{etc.},$$

are less restrictive. This gives rise to the particle–antiparticle symmetry.

Gel'fand and Yaglom[3] based their approach on the theory of representations of the homogeneous Lorentz group. They require that $\psi(x)$ transforms under the homogeneous Lorentz group as a certain representation ("index transformations") and that it satisfies the linear field equation

$$\Big(-i\gamma_\mu\frac{\partial}{\partial x_\mu} + M\Big)\psi(x) = 0.$$

The finite or infinite matrices γ_μ are to be determined from the requirement that the field equation is covariant.

This approach is very general. Bargmann–Wigner fields can be obtained as a special case and after some superfluous field components have been eliminated. If $\psi(x)$ transforms as a finite direct sum of unitary irreducible representations of the homogeneous Lorentz group, it can in general be decomposed into irreducible unitary representations of the inhomogeneous Lorentz group such that

(1) representations with $M^2 > 0$ appear as a discrete direct sum, the discrete spectrum of M^2 is bounded from above and accumulates at zero;

(2) representations with $M^2 < 0$ appear as a direct integral and a direct sum, the spectrum of M^2 is continuous, contains a discrete set of eigenvalues, and is unbounded from below;

[3] M. A. Naimark, *Linear Representations of the Lorentz Group* (English transl.: Pergamon Press, Oxford, 1964), and the original publications quoted in this book.

(3) there exists a finite set of functions

$$M_i^2(S)$$

(mass relations) which relate the mass and the spin of the representations occurring.

These fields enable us to describe simultaneously an infinite spectrum of M^2. The Bargmann–Wigner fields to the mass $M^2 < 0$ which we are going to study now, are, however, much simpler objects.

2. UNITARY REPRESENTATIONS OF THE DISCRETE SERIES FOR THE GROUP $SU(1, 1)$ AND THEIR CONTINUATION

A. The Discrete Series D_k^+ and D_k^-

Let $\mathfrak{A}$ and $\overline{\mathfrak{A}}$ be two linear spaces defined by

$$\mathfrak{A} = \{f(z): f(z) \text{ holomorphic in } |z| < 1 \text{ and continuous in } |z| \leq 1\},$$
$$\overline{\mathfrak{A}} = \{f(z): \overline{f(z)} \in \mathfrak{A}\}.$$

Defining a norm by

$$\|f\|_k^2 = \frac{2k-1}{\pi} \int_{|z|\leq 1} (1 - |z|^2)^{2k-2} |f(z)|^2 \, dz$$

for $k = 1, \frac{3}{2}, 2, \cdots$, and as the limit $k \to \frac{1}{2}+$ for $k = \frac{1}{2}$, enables us to complete $\mathfrak{A}$ and $\overline{\mathfrak{A}}$. We call the Hilbert spaces $\mathcal{H}_k$ and $\overline{\mathcal{H}}_k$. Because of

$$|f(z)| \leq (1 - |z|^2)^{-k} \|f\|_k,$$

all elements of $\mathcal{H}_k(\overline{\mathcal{H}}_k)$ are holomorphic (antiholomorphic) for $|z| < 1$. We can give a basis in $\mathcal{H}_k(\overline{\mathcal{H}}_k)$ by

$$\{f_m\}, \quad f_m(z) = z^m, \quad \{f_m\}, f_m(\bar{z}) = \bar{z}^m, \quad m = 0, 1, 2, \cdots.$$

The basis elements have the norm

$$\|f_m\|_k = \binom{2k-1+m}{m}^{-\frac{1}{2}}.$$

For an element $v \in SU(1, 1)$,

$$v = \begin{pmatrix} v_{11} v_{12} \\ v_{21} v_{22} \end{pmatrix}, \quad v_{11} = \bar{v}_{22}, v_{12} = \bar{v}_{21},$$

we define an irreducible unitary representation in $\mathcal{H}_k$ by

$$T_v f(z) = (v_{12} z + v_{22})^{-2k} f(z'),$$

where

$$z' = \frac{v_{11} z + v_{21}}{v_{12} z + v_{22}}.$$

In this fashion we obtain the series D_k^+.[4] Similarly we define the series D_k^- in $\overline{\mathcal{H}}_k$ by

$$T_v f(\bar{z}) = (v_{21} \bar{z} + v_{11})^{-2k} f(\bar{z}'),$$
$$\bar{z}' = \frac{v_{22} \bar{z} + v_{12}}{v_{21} \bar{z} + v_{11}}.$$

Let us consider the special element $v_\phi \in SU(1, 1)$:

$$v_\phi = \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix}.$$

It turns out that the basis vectors f_m are eigenvectors of T_{v_ϕ}:

$$D_k^+: \quad T_{v_\phi} f_m = e^{i\phi(k+m)} f_m,$$
$$D_k^-: \quad T_{v_\phi} f_m = e^{-i\phi(k+m)} f_m.$$

B. The Continuation of T_v on Polynomial Spaces

The definition of T_v can be extended to all elements of $SL(2, C)$ by simply replacing

$$v_{\alpha\beta} \to a_{\alpha\beta}.$$

However, this operation leads in general to elements lying outside $\mathcal{H}_k$ or $\overline{\mathcal{H}}_k$.

In the sequel the polynomial spaces $\mathfrak{F}(\overline{\mathfrak{F}})$ which consist of polynomials in $z(\bar{z})$ and are dense in $\mathcal{H}_k(\overline{\mathcal{H}}_k)$ play a particular role. On these subspaces, T_v can be extended to an operator T_a, provided a satisfies the following conditions:

$$\mathcal{H}_k: a_{12} z + a_{22} \neq 0 \quad \text{for} \quad |z| \leq 1,$$

or

$$|z_0| > 1, \quad \text{when} \quad z_0 = -a_{22}/a_{12},$$
$$\overline{\mathcal{H}}_k: a_{21} z + a_{11} \neq 0 \quad \text{for} \quad |z| \leq 1,$$

or

$$|z_0| > 1, \quad \text{when} \quad z_0 = -a_{11}/a_{21}.$$

In that case,

$$\|T_a f_m\|_k < \infty$$

and T_a is an unbounded operator with domain $\mathfrak{O} \supset \mathfrak{F}$, or $\overline{\mathfrak{O}} \supset \overline{\mathfrak{F}}$. The conditions on a found above are referred to as conditions I.

C. The Continuation of T_v on the Whole Hilbert Space

Though not important for our purposes, we give a characterization of all those $a \in SL(2, C)$ for which T_a can be defined on the whole Hilbert space. This domain of elements a is called Graev's domain.[5] For the sake of simplicity we consider only the representation D_k^+. Conditions on the element a

[4] V. Bargmann, Ann. Math. (2) **48**, 568 (1947).

[5] M. I. Graev, Am. Math. Soc. Transl. (2) **16**, 393 (1960).

for D_k^- can be obtained by the substitution

$$a_{11} \leftrightarrow a_{22}, \quad a_{12} \leftrightarrow a_{21}.$$

The necessary and sufficient condition (Graev's condition) is that

$$z' = \frac{a_{11}z + a_{21}}{a_{12}z + a_{22}}$$

defines a contraction, i.e.,

$$|z'| < 1 \quad \text{for} \quad |z| \leq 1.$$

In this case $f(z')$ is well defined and analytic in $|z| \leq 1$, whatever the singularity of $f(z)$ on the unit circle is. Moreover, we find

$$\mathrm{Tr}\,(T_a^+ T_a) = \sum_{m=0}^{\infty} \binom{2k-1+m}{m} \|T_a f_m\|_k^2 \leq \|T_a f_0\|_k^2 (1-\rho^2)^{-2k} < \infty$$

when

$$\rho = \max_{|z|\leq 1} |z'| < 1.$$

Use has been made of

$$\|T_a f_m\|_k < \rho^m \|T_a f_0\|_k.$$

The existence of the traces of $T_a^+ T_a$ proves that T_a is bounded and completely continuous. When we introduce a real four-vector p_μ:

$$p = p_0 + \sum_{k=1}^{3} p_k \sigma_k = a \sigma_3 a^+,$$

$$p_0^2 - \sum_{k=1}^{3} p_k^2 = -1,$$

the condition that

$$z \xrightarrow[a]{} z'$$

is a contraction can be expressed by Graev's domain:

$$p_0 < 0, \quad |p_3| > 1.$$

D. Triangular Matrices

The simplest manner to satisfy conditions I is

$$D_k^+: \quad a_{12} = 0, \quad a_{22} \neq 0,$$
$$D_k^-: \quad a_{21} = 0, \quad a_{11} \neq 0.$$

Then a has the form, respectively,

$$a = k = \begin{pmatrix} \lambda & 0 \\ \mu & \lambda^{-1} \end{pmatrix}, \quad \text{in } D_k^+,$$

$$a = \tilde{k} = \begin{pmatrix} \lambda^{-1} & \mu \\ 0 & \lambda \end{pmatrix}, \quad \text{in } D_k^-,$$

with λ, μ complex. Such triangular matrices constitute a group K (respectively, $\tilde{K}$).

As we did in the case of Graev's domain, we can characterize conditions I by a four-vector (note the difference!).

$$p = p_0 + \sum_{k=1}^{3} p_k \sigma_k = a^{-1} \sigma_3 (a^{-1})^+,$$

$$p_0^2 - \sum_{k=1}^{3} p_k^2 = -1.$$

This yields for

$$D_k^+ \text{ domain } I: \quad p_0 + p_3 > 0,$$
$$D_k^- \text{ domain } \tilde{I}: \quad p_3 - p_0 > 0.$$

We continue the discussion first for the case D_k^+. Inserting

$$a = vk, \quad v \in SU(1,1), \quad v^+ \sigma_3 = \sigma_3 v^{-1},$$

into the definition of p, yields the elements λ, μ of $k(p)$:

$$\lambda = (p_0 + p_3)^{-\frac{1}{2}}, \quad \mu = (p_1 + ip_2)(p_0 + p_3)^{-\frac{1}{2}}.$$

These matrices $k(p)$ can be used as boosts. It is remarkable how natural the choice of boosts as triangular matrices is: The domain to which the representation of $SU(1,1)$ can be extended coincides with the domain in which a triangular boost can be defined. Triangular boosts are not rotation free; we should, therefore, better call them "twists."

The complementary domain

$$p_0 + p_3 < 0$$

can be reached by boosts $i\sigma_1 \cdot k(-p)$ (see Ref. 6). Triangular matrices leave the space $\mathfrak{F}$ invariant, this space carries therefore a representation of K by means of unbounded operators. Normalizing the basis f_m, we find the matrix elements

$$(T_k)_{m_1 m_2} = c_{m_1 m_2} \lambda^{2k+m_1+m_2} \mu^{m_2 - m_1},$$

where

$$c_{m_1 m_2} = \frac{1}{(m_2 - m_1)!} \left[\frac{m_2!\,(2k-1+m_2)!}{m_1!\,(2k-1+m_1)!} \right]^{\frac{1}{2}}.$$

In the case D_k^- we get the same result

$$(T_{\tilde{k}})_{m_1 m_2} = (T_k)_{m_1 m_2}.$$

In particular, both matrices are triangular matrices of the upper-right type. The space $\bar{\mathfrak{F}}$ is invariant against $T_{\tilde{k}}$, where $T_{\tilde{k}}$ is a representation of the element $\tilde{k} \in \tilde{K}$.

3. BARGMANN–WIGNER FIELDS

A. The Field Element to the Angle Ω_0

Let a Hilbert space $\mathcal{H}$ be defined by

$$\mathcal{H} = \left\{ \phi_m(p): \sum_{m=0}^{\infty} \int |\phi_m(p)|^2 \, d\mu(p) < \infty \right\},$$

[6] G. Rideau, Commun. Math. Phys. 3, 218 (1966).

where

$$d\mu(p) = d^4p\delta(p^2 + 1).$$

In this Hilbert space we define a unitary representation of the inhomogeneous Lorentz group by

$$U_{(a,x)}\phi_m(p) = \sum_{m=0}^{\infty} e^{ipx}D_{mm'}(R(a, p))\phi_{m'}(p'),$$

with $D \in D_k^+$, and

$$p' = a^{-1} \cdot p \cdot (a^{-1})^+,$$
$$R(a, p) = a(p) \cdot a \cdot a(p')^{-1} \in SU(1, 1),$$
$$a(p)pa(p)^+ = \sigma_3.$$

We choose $a(p)$ as

$$a(p) = \begin{cases} k(p), & \text{for } p_0 + p_3 > 0, \\ i\sigma_1 \cdot k(-p), & \text{for } p_0 + p_3 < 0. \end{cases}$$

$D_{mm'}(v)$ may be any representation of the series D_k^+. The series D_k^- can be treated similarly (see Sec. 4). For $p_0 + p_3 > 0$, we represent $a(p)$ by $A(p)$:

$$A(p) = T_{k(p)}.$$

If ϕ_m is in $\mathfrak{F}$, $A(p)$ and its inverse can be applied to it, and we are allowed to define

$$\psi_\alpha(p) = \sum_{m=0}^{\infty} A^{-1}(p)_{\alpha m} \cdot \phi_m(p) \cdot \theta(p_0 + p_3).$$

If p and p' are both in I, then we may further write

$$U_a\psi_\alpha(p) = \sum_{\alpha'=0}^{\infty} D_{\alpha\alpha'}(a)\psi_{\alpha'}(p'),$$

where by definition

$$D_{\alpha\alpha'}(a) = [A^{-1}(p)D(R(a, p))A(p')]_{\alpha\alpha'}.$$

For momenta p on a given orbit, we use the notation Ω for the direction of the momentum three-vector in the system $p_0 = 0$. In particular, we denote the direction of the positive third axis by Ω_0. Remembering that our definition of boosts involved a normal form $p_n = \sigma_3$, we write the "field element" $\psi_\alpha(p)$ just constructed as

$$\psi_\alpha(p, \Omega_0).$$

The extension of this definition to all Ω and the final substitution of Ω by rest classes of the group K are the next steps to take.

B. The Field Element for Arbitrary Angle Ω

The condition that p and p' are both in I is a condition on the element $a \in SL(2, C)$ which cannot be satisfied by all a's. To make an application of all a's possible, we proceed as follows. Let us define

$$\sigma_\Omega = u_\Omega^{-1}\sigma_3 u_\Omega = \sum_{k=1}^{3} e_{\Omega,k}\sigma_k,$$
$$\sum_{k=1}^{3} e_{\Omega,k}^2 = 1,$$

where u_Ω is any matrix of $SU(2)$. To make the relation between u_Ω and e_Ω unique, we may define

$$u_{\Omega,11} \geq 0, \quad u_\Omega \in SU(2)/U(1).$$

The equation

$$a_\Omega(p)pa_\Omega(p)^+ = \sigma_\Omega,$$

where

$$p_\Omega = u_\Omega p u_\Omega^{-1} \in I,$$

can be solved by

$$a_\Omega(p) = u_\Omega^{-1}k(p_\Omega)u_\Omega.$$

The corresponding little group $SU(1, 1)_\Omega$ consists of elements

$$R(a, p, \Omega) = a_\Omega(p)aa_\Omega(p')^{-1}$$
$$= u_\Omega^{-1}R(a_\Omega, p_\Omega)u_\Omega,$$

where

$$a_\Omega = u_\Omega a u_\Omega^{-1}.$$

We introduce a field element to the angle Ω by

$$\psi_\alpha(p, \Omega) = \sum_{m=0}^{\infty} [A^{-1}(p_\Omega)D(R(u_\Omega, p_\Omega))]_{\alpha m} \times \phi_m(p)\theta(\mathbf{p}\mathbf{e}_\Omega + p_0).$$

For this definition to make sense we must, moreover, take

$$\phi'(p) = D(R(u_\Omega, p_\Omega))\phi(p) \in \mathfrak{F}.$$

If $p \in I$ and $u \in SU(2)$ are such that also

$$p' = u^{-1}pu \in I,$$

we obtain

$$U_u\psi_\alpha(p, \Omega_0) = \sum_{\alpha'=0}^{\infty} D_{\alpha\alpha'}(u)\psi_{\alpha'}(p', \Omega_0)$$
$$= \sum_{m=0}^{\infty} [D(u)A^{-1}(p')]_{\alpha m}\phi_m(p').$$

Let us choose

$$u = u_\Omega \quad \text{and} \quad p'_\Omega = u_\Omega p' u_\Omega^{-1} = p.$$

Then we may insert our definition of the field element to the angle Ω,

$$\phi_m(p') = \sum_{\alpha=0}^{\infty} [D(R^{-1}(u_\Omega, p))A(p)]_{m\alpha}\psi_\alpha(p', \Omega),$$

and get

$$U_{u_\Omega}\psi_\alpha(p, \Omega_0) = \psi_\alpha(p', \Omega), \quad p' = u_\Omega^{-1}pu_\Omega,$$

because by definition

$$D(u_\Omega)A^{-1}(p')D(R^{-1}(u_\Omega, p))A(p) = 1.$$

For elements p, p' which are not simultaneously of I, we take

$$U_{u_\Omega}\psi_\alpha(p, \Omega_0) = \psi_\alpha(p', \Omega)$$

as a definition of U_{u_Ω}.

C. Application of $SL(2, C)$ to $\psi_\alpha(p, \Omega)$

We find in Sec. 3D that the field $\psi_\alpha(p, \Omega)$ contains many superfluous variables. This is the price we have to pay for a p-independent transformation property under application of the homogeneous Lorentz group. To find this transformation property, we first write

$$\psi_\alpha(p, u_\Omega) \quad \text{instead of} \quad \psi_\alpha(p, \Omega).$$

Second, we note that an arbitrary element $a \in SL(2, C)$ can be decomposed as

$$a = k \cdot u, \quad u_{11} \geq 0, \quad k \in K.$$

This decomposition tells us that the quotient

$$SL(2, C)/K$$

can be characterized by such elements u:

$$u \in SU(2) \mid U(1), \quad \text{i.e.,} \quad u_{11} \geq 0.$$

From the relations

$$k \in K: \qquad U_k\psi_\alpha(p, 1) = \sum_{\beta=0}^{\infty}(T_k)_{\alpha\beta}\psi_\beta(p', 1), \qquad p' = k^{-1}p(k^{-1})^+,$$

$$u \in SU(2)/U(1): \quad U_u\psi_\alpha(p, 1) = \psi_\alpha(p', u), \qquad p' = u^{-1}pu,$$

we find, for general $a \in SL(2, C)$,

$$\begin{aligned} U_a\psi_\alpha(p, u) &= U_uU_a\psi_\alpha(upu^{-1}, 1) \\ &= U_kU_{u'}\psi_\alpha(upu^{-1}, 1) \\ &= \sum_{\beta=0}^{\infty}(T_k)_{\alpha\beta}U_{u'}\psi_\beta(k^{-1}up(k^{-1}u)^+, 1) \\ &= \sum_{\beta=0}^{\infty}(T_k)_{\alpha\beta}\psi_\beta(a^{-1}p(a^{-1})^+, u'). \end{aligned}$$

Here we made use of

$$ua = ku'$$

which describes the transformation of points in the space $SL(2, C)/K$. The matrix elements of T_k were given in Section 2.4.

D. The Elimination of Superfluous Variables

The fact that a field $\psi_\alpha(p, u)$ belongs to an irreducible representation of the inhomogeneous Lorentz group leads to an identity which helps us to eliminate superfluous variables. If $u_1, u_2 \in SU(2)/U(1)$ are such that

$$p_{u_1}, \quad p_{u_2} \in I, \qquad p_u = upu^{-1},$$

$$\psi_\alpha(p, u_1) = \sum_{\beta=0}^{\infty}[A^{-1}(p_{u_1})D(R(u_1, p_{u_1})) \times D(R(u_2, p_{u_2}))^{-1}A(p_{u_2})]_{\alpha\beta}\psi_\beta(p, u_2)$$

or, as a definition of $\Delta\alpha\beta$,

$$\psi_\alpha(p, u_1) = \sum_{\beta=0}^{\infty}\Delta_{\alpha\beta}(p \mid u_1, u_2)\psi_\beta(p, u_2).$$

For p_u, $u = u_1$, and $u = u_2$, not simultaneously in I the $\Delta_{\alpha\beta}$ can, in general, not be defined.

4. THE PARITY TRANSFORMED FIELD

A. Definition of a Field $\psi^\alpha(p, u)$

If a representation of $SU(1, 1)$ belongs to the series D_k^+, the conjugate representation belongs to D_k^- with the same k. The conjugation can best be defined in the basis $\{f_m\}$. The idea of defining a field $\psi^\alpha(p, u)$ consists in continuing first the conjugate representation and to conjugate it back after the continuation. However, we emphasize that the field $\psi^\alpha(p, u)$ thus obtained is different from the conjugate contragredient representation of $\psi_\alpha(p, u)$. The reason is that

$$\tilde{k} = k^{-1,\mathrm{T}}$$

does not imply

$$T_{\tilde{k}} = (T_k)^{-1,\mathrm{T}}.$$

Indeed the latter equation is impossible, since $T_{\tilde{k}}$ is upper-right but the transpose of T_k^{-1} is lower-left triangular.

We proceed similarly as before. We define

$$a(p) = \begin{cases} \tilde{k}(p), & \text{for } p_3 - p_0 > 0, \\ i\sigma_1\tilde{k}(-p), & \text{for } p_3 - p_0 < 0, \end{cases}$$

where the elements of $\tilde{k}(p)$ are

$$\lambda = (p_3 - p_0)^{-\frac{1}{2}}, \quad \mu = -(-p_1 + ip_2)(p_3 - p_0)^{-\frac{1}{2}}.$$

For $p \in \bar{I}$ we represent $a(p)$ by $B(p)$:

$$B(p) = T_{\tilde{k}(p)}.$$

If P_t is a time inversion

$$P_tp_0 = -p_0, \quad P_tp_k = +p_k,$$

we may write

$$\overline{\tilde{k}(P_tp)} = k^{-1,\mathrm{T}}(p).$$

By means of the matrix $B(p)$ we define

$$\psi^\alpha(p, \Omega_0) = \sum_{m=0}^{\infty}B^{-1}(p)_{\alpha m}\phi_m(p)\theta(p_3 - p_0)$$

as before. For $\tilde{k} \in \tilde{K}$ we have

$$U_{\tilde{k}}\psi^{\alpha}(p, \Omega_0) = \sum_{\beta=0}^{\infty} \overline{(T_{\tilde{k}})}^{\alpha\beta}\psi^{\beta}(p', \Omega_0),$$

$$p' = \tilde{k}^{-1}p(\tilde{k}^{-1})^{+}.$$

In addition we introduce

$$\psi^{\alpha}(p, \Omega) = \sum_{m=0}^{\infty} [B^{-1}(p_\Omega)\, D(\tilde{R}(u_\Omega p_\Omega))]_{\alpha m}\phi_m(p) \cdot \theta(\mathbf{e}_\Omega \mathbf{p} - p_0).$$

Here we marked the difference in the definition of $R(a, p)$, which is due to the different boosts used.

The quantities $\psi^{\alpha}(p, u)$ transform, finally, as

$$U_a\psi^{\alpha}(p, u) = \sum_{\beta=0}^{\infty} \overline{(T_{\tilde{k}})}_{\alpha\beta}\psi^{\beta}(p', u''),$$

with

$$ua = \tilde{k}u'',$$

$$p' = a^{-1}p(a^{-1})^{+}.$$

We point out that the argument u'' is not equal to u' obtained from

$$ua = ku'.$$

B. Bargmann–Wigner Equations

After a Fourier transformation we obtain

$$\psi_{\alpha}(x, u) = (2\pi)^{-\frac{3}{2}} \int e^{ipx}\psi_{\alpha}(p, u)\, d\mu(p),$$

with the property

$$U_{(a,y)}\psi_{\alpha}(x, u) = \sum_{\beta=0}^{\infty} (T_k)_{\alpha\beta}\psi_{\beta}(a^{+}xa + y, u').$$

Our aim is to construct a dual object $\psi^{\alpha}(x, u)$ which is connected with $\psi_{\alpha}(x, u)$ by a covariant field equation. The form of $\psi^{\alpha}(x, u)$ is not yet suited for such a purpose since

(a) its support is different from the support of $\psi_{\alpha}(p, u)$;

(b) u transforms differently and so does the support.

To cure these defects we perform first a time reflection P_t:

$$\tilde{\psi}^{\alpha}(p, u) = \psi^{\alpha}(P_t p, u).$$

To adjust the transformation property of u we make an additional total inversion in $\phi_m(p)$:

$$\phi_m(P_t p) \to \phi'_m(P_t p) = \phi_m(P_s p).$$

Since $\hat{p} = P_s p$ transforms as

$$\hat{p} \to \hat{p}' = a^{+}pa,$$

i.e., contravariant as compared with p, we introduce a new representation by

$$U_a \to \tilde{U}_a = U_{(a^{-1})^{+}}, \quad U_x \to \tilde{U}_x = U_{P_s x} = U_{\hat{x}}.$$

The relation defining the change in u is now

$$u(a^{+})^{-1} = \tilde{k}u''$$

which implies

$$\tilde{k} = (k^{-1})^{+}, \quad u'' = u'.$$

This is in accord with the relation

$$\tilde{k}(P_t p) = [k(p)^{-1}]^{+}$$

already found above. In this fashion we obtain a quantity $\psi^{\alpha}(p, u)$ which transforms as

$$\tilde{U}_a\psi^{\alpha}(p, u) = \sum_{\beta=0}^{\infty} (T_{k^{-1,T}})_{\alpha\beta}\psi^{\beta}(p', u').$$

In the derivation of this equation we have to make use of the fact that u_Ω commutes with P_t and that, in the basis $\{f_m\}$,

$$T_{\bar{k}} = T_k^{*} \quad (\bar{k} \text{ is the conjugate of } k),$$

because the matrix elements depend polynomially on the matrix elements of k with real coefficients. Now we are in a position to derive the Bargmann–Wigner equations:

$$\psi^{\alpha}(\sigma_3, 1) = \psi_{\alpha}(\sigma_3, 1)$$

implies

$$\tilde{U}_a\psi^{\alpha}(\sigma_3, 1) = U_a\psi_{\alpha}(\sigma_3, 1)$$

or

$$\psi^{\alpha}(p, u) = \sum_{\beta=0}^{\infty} P^{\alpha\beta}(p, u)\psi_{\beta}(p, u),$$

$$\psi_{\alpha}(p, u) = \sum_{\beta=0}^{\infty} \hat{P}_{\alpha\beta}(p, u)\psi^{\beta}(p, u),$$

where

$$P^{\alpha\beta}(p, u) = \sum_{\gamma=0}^{\infty} (T_{k^T})_{\alpha\gamma}(T_k)_{\gamma\beta}.$$

Here p, k, and u are defined by:

$$p = a^{-1}\sigma_3(a^{-1})^{+},$$

$$a = ku.$$

Of course $P^{\alpha\beta}$ depends only on k. It is a triangular matrix with the elements

$$P^{\alpha\beta} = (2\lambda\mu)^{\beta-\alpha}c_{\alpha\beta}.$$

λ and μ are the elements of k. It is remarkable that all diagonal elements of P and $\hat{P}$ are equal to one.

A Fourier transformation of P and $\hat{P}$ makes not much sense because of the singularities involved in the elements of P and $\hat{P}$. Indeed, for $u = 1$, we have $k = k(p)$ and, therefore,

$$\lambda\mu = -(p_1 + ip_2)/(p_0 + p_3).$$

On the other hand, the fields ψ_{α} and ψ^{α} must satisfy

the boundary condition that

$$\psi_\alpha(\psi^\alpha) \sim (\mathbf{pe}_\Omega + p_0)^\alpha$$

or of higher order for $\mathbf{pe}_\Omega + p_0 \to 0+$.

From the fact that the elements of P and $\hat{P}$ are of homogeneity zero in the momentum follows immediately that the Klein–Gordon equation must be imposed on the fields as an independent condition.

5. DISCUSSION

We have obtained fields

$$\psi_\alpha(x, u)$$

which satisfy

(a) the Bargmann–Wigner equations;
(b) the Klein–Gordon equation;
(c) certain identities.

We recall that the identities have been introduced to eliminate superfluous variables and to connect our fields with an irreducible representation of the inhomogeneous Lorentz group. We emphasize, however, that it makes good sense to abandon these identities. Indeed, a timelike Bargmann–Wigner field is also reducible into two components satisfying

$$|p_0| = m.$$

Abandoning the identities in our case would lead to a representation which is reducible into components which can all be ascribed to the solutions of

$$|\mathbf{p}| = \lambda.$$

Instead of a particle–antiparticle duality this would entail a continuous internal symmetry group $SU(2)$.

Let us now make a comment on the transformation properties of our fields. A Gel'fand–Yaglom field $\psi(x, u)$ transforms as

$$U_a\psi(x, u) = \lambda^{t_1}\bar{\lambda}^{t_2}\psi(a^+xa, u'),$$

where $ua = ku'$, λ is the diagonal element k_{11} of k, t_1, t_2 are complex numbers. Let us decompose k as

$$k = \begin{pmatrix} \lambda & 0 \\ \mu & \lambda^{-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mu\lambda^{-1} & 1 \end{pmatrix}\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}.$$

The matrices

$$\begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}$$

constitute the nilpotent group N and the diagonal matrices constitute the Abelian group A, i.e.,

$$K = NA.$$

If in a completely irreducible representation of K, N is mapped on the identity, the representation is one-dimensional and a representation of A. This yields the multiplier of the Gel'fand–Yaglom field. In our case the nilpotent group N is represented by infinite triangular matrices.

Gel'fand-Yaglom fields have an important advantage over our fields, since they allow us to construct local invariant densities in a simple fashion, e.g.,

$$[\psi, \psi](x) = \int \bar{\psi}(x, u)\psi(x, u)\, d\mu(u)$$

in the case of the principal series. However, if we allow for a reducibility with respect to the inhomogeneous Lorentz group of an arbitrary degree, this statement is nearly trivial. Indeed, a field belonging to the regular representation of $SL(2, C)$,

$$\psi(x, a), \quad a \in SL(2, C),$$

gives the invariant density

$$[\psi, \psi](x) = \int d\mu(a)\bar{\psi}(x, a)\psi(x, a).$$

Such fields can even be made to satisfy the equation

$$(\Box_x + M^2)\psi(x) = 0.$$

Let us finish with a remark concerning the connection of spin and statistics. We know that Pauli's proof of the spin-statistics theorem is based on the degree $2S$ of the homogeneous polynomial $P^{\alpha\beta}(p)$. In this context, it is remarkable that our triangular matrices $P^{\alpha\beta}(p)$ are homogeneous of degree zero independently of what the spin k is.

ACKNOWLEDGMENT

We thank Professor A. Pais for his kind hospitality at the Rockefeller University.