JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Physical-Region Discontinuity Equations for Many-Particle Scattering Amplitudes. I*

JOSEPH COSTER AND HENRY P. STAPP
Lawrence Radiation Laboratory, University of California, Berkeley, California

(Received 15 April 1967)

The discontinuity equations are derived for all singularities of multiparticle scattering functions that enter the portion of the physical region lying below the four-particle threshold. These equations, which are the precise statements of the Cutkosky formulas, are calculated directly from the unitarity equations. The only analyticity properties used are those obtained from the S-matrix macroscopic-causality condition. That is, the scattering functions are taken to be analytic in the physical region except on positive-α Landau surfaces, around which they continue in accordance with the well-defined plus-$i\epsilon$ rule.

1. INTRODUCTION

The purpose of this paper, and of several that will follow, is to derive formulas for discontinuities across singularities of many-particle scattering functions. These formulas are needed for the extension of general dispersion methods to many-particle reactions.

Discontinuity equations have been derived by Cutkosky in the framework of perturbation theory.[1] His result states that the discontinuity around a surface $\mathcal{M}[D]$ corresponding to a Landau diagram D is obtained by associating the vertices of D with scattering functions and performing an integration over the momentum vectors associated with the internal lines of D. These formulas are, however, essentially incomplete, for they include no general rule specifying upon which sheets one evaluates the various functions that occur. In the simplest case of a normal threshold in a two-particle scattering function, the two functions that occur in the discontinuity formula are the physical scattering function and its Hermitian conjugate, whereas in the case of the triangle singularity of the $3 \to 3$ (three-particle to three-particle) amplitude, all the occurring functions are physical scattering amplitudes. In other cases, certain functions in the discontinuity formula are neither the physical function nor its Hermitian conjugate, but are functions on other sheets. In fact, it has not actually been proved that the functions in the discontinuity formulas are in all cases merely the continuation of a scattering function to some sheet.

The problem of finding discontinuity formulas has been examined earlier in the S-matrix framework by Gunson,[2] Stapp,[3] and Olive.[4] Their approach has been essentially to verify, within certain approximations, the consistency of certain conjectured discontinuity formulas. Certain normal-threshold discontinuity formulas have been derived by Hwa using crossing, and working to lowest order.[5]

More recently, Landshoff and Olive[6] have derived the discontinuity across the singularity of the triangle diagram of the $3 \to 3$ amplitude in the physical region, and their method has been applied by others[7–10] to singularities associated with various other diagrams. The method of Landshoff and Olive is, however, quite complicated. It requires a detailed investigation of specific features of Landau curves, an examination of the properties of certain integrals, and a tracing of paths of continuation, and it depends on delicate cancellations of various terms. Also, it requires a complete enumeration of the "generation" and "regeneration" mechanisms of the singularity, and this is not easily obtained except in the simplest cases. Finally, each singularity is a separate problem.

In the present paper we develop an alternative method for calculating the discontinuities of the (connected part) physical-region scattering amplitude M^+. This function has singularities only on positive-α Landau surfaces, and it can be continued past these in accordance with a well-defined plus-$i\epsilon$ rule.[11,12] The discontinuities around these singularities are obtained in this paper directly from the unitarity equations through manipulations that bring these equations into a form that displays explicitly the appropriate discontinuity function. Specifically, a unitarity equation

* This work was done under the auspices of the U.S. Atomic Energy Commission.

[1] R. E. Cutkosky, J. Math. Phys. **1**, 429 (1960); Phys. Rev. Letters **4**, 624 (1960).

[2] J. Gunson, University of Birmingham, Birmingham, England, report of work prior to publication, 1962; published in J. Math. Phys. **6**, 827, 845, and 852 (1965).

[3] H. P. Stapp, *Lectures on Analytic S-Matrix Theory* (Matscience, Madras, India, 1963); "Discontinuity Equations," Trieste Report IC/65/17.

[4] D. I. Olive, Nuovo Cimento **37**, 1422 (1965).

[5] R. C. Hwa, Phys. Rev. **134**, B1086 (1964).

[6] P. V. Landshoff and D. I. Olive, J. Math. Phys. **7**, 1464 (1966).

[7] P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, J. Math. Phys. **7**, 1593 (1966).

[8] P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, J. Math. Phys. **7**, 1600 (1966).

[9] M. J. W. Bloxham, Nuovo Cimento **44A**, 794 (1966).

[10] J. K. Storrow, Nuovo Cimento (to be published).

[11] C. Chandler and H. P. Stapp, "Macroscopic Causality Conditions and Properties of Scattering Amplitudes," (to be published in J. Math. Phys.); and D. Iagolnitzer and H. P. Stapp (in preparation).

[12] F. Pham, CERN preprint, 1966; Ann. Inst. Henri Poincaré **6**, 89 (1967).

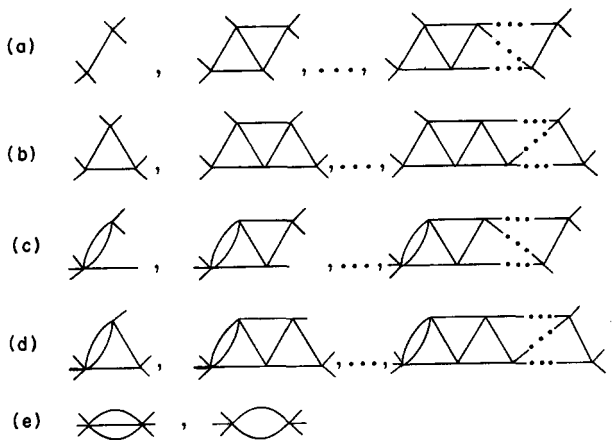


FIG. 1. The elementary diagrams of the $3 \to 3$ amplitude below the four-particle threshold. All positive-α Landau surfaces for the $3 \to 3$ amplitude below the four-particle threshold are contained in the union of the positive-α Landau surfaces corresponding to the set of diagrams consisting of (1) the set of elementary diagrams, (2) the set of reflections of elementary diagrams, and (3) the set of diagrams obtained from diagrams of these two sets by replacing any set of simple inner vertices by single two-particle closed loops. Simple vertices are vertices directly connected to no other vertex by more than one line. Inner vertices are vertices not standing on the extreme right or left of the diagram. The second figure in (e) contains a single two-particle closed loop. The analogous part of the first figure in (e) is called a three-particle closed loop.

is converted, using unitarity, into the form

$$M^+ = T(D) + R(D), \tag{1.1}$$

where $T(D)$ vanishes on one side (called the unphysical side) of the positive-α Landau surface $\mathcal{M}^+[D]$ associated with the Landau diagram D, and $R(D)$ is known to have a minus-$i\epsilon$ continuation around $\mathcal{M}^+[D]$. That $R(D)$ has a minus-$i\epsilon$ continuation around $\mathcal{M}^+[D]$ means that the function $R(D)$ is carried into itself by a continuation around $\mathcal{M}^+[D]$ in the sense opposite to the sense that carries M^+ into itself. Since $T(D)$ vanishes on the unphysical side of $\mathcal{M}^+[D]$, the function $R(D)$ constitutes an explicit expression for the continuation of M^+ around $\mathcal{M}^+[D]$ in the minus-$i\epsilon$ sense, and (1.1) displays $T(D)$ as the discontinuity of M^+ around $\mathcal{M}^+[D]$.

In this first article, a bubble-diagram notation is set up that facilitates manipulations of the unitarity equations. Various results needed from earlier work[13] concerning the analytic structure of bubble-diagram functions are summarized, and a general theorem fundamental to our approach is proved. The method is then exhibited for the special case of $3 \to 3$ reactions below the four-particle threshold, and we obtain the discontinuity formulas for every physical-region singularity. The results for the $3 \to 2$ and $2 \to 3$ reactions can be obtained from these formulas by regarding an appropriate pair of external lines as a single line.

The results in the $3 \to 3$ case are easily summarized. All the positive-α Landau surfaces below the four-particle threshold are contained in the surfaces corresponding to the set of diagrams shown in Fig. 1 together with those obtained from these diagrams by the procedure explained in the caption. (The other possible Landau diagrams, some of which are shown in Fig. 2, give no additional surfaces and hence can be ignored.) The discontinuity of M^+ around the Landau surface $\mathcal{M}^+[D]$ for any one of these diagrams D is expressed as an integral over a product of a set of functions consisting of one physical scattering amplitude (the connected part of the S matrix, which is diagrammatically represented by a plus bubble) for each vertex V_n of D and one function F_{nm} for each pair of vertices (V_n, V_m). If there is no elementary line segment L_j directly connecting V_n to V_m, then F_{nm} is unity. If exactly one L_j directly connects V_n to V_m, then F_{nm} is $2\pi\delta(p_j^2 - \mu_j^2)\theta(p_j^0)$. If two or three lines L_j directly connect V_n to V_m, then F_{nm} is a function defined in terms of physical scattering amplitudes by a Fredholm integral equation. It is formally the inverse S_α^{-1} of the restriction S_α of S to the space associated with the set of lines α connecting V_n and V_m.

The rules giving the discontinuity can be expressed

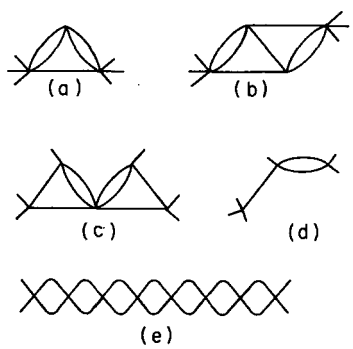
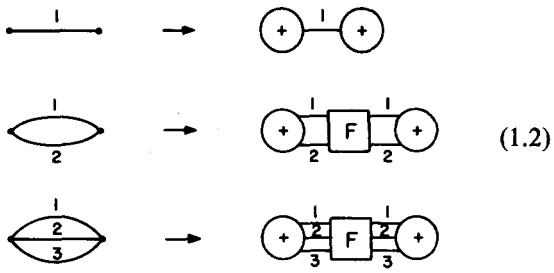


FIG. 2. Various Landau diagrams corresponding to Landau surfaces contained in the Landau surfaces of the diagrams described in Fig. 1. A diagram containing more than one $2 \to 3$, $3 \to 2$, or $3 \to 3$ vertex, such as one of the above diagrams (a) and (b), corresponds to a Landau surface that is confined to the surface corresponding to one of the diagrams (e) of Fig. 1. The diagrams (c) and (d) contain two "independent parts" that are diagrams described in Fig. 1. The Landau surfaces of the full diagrams are confined to the intersections of the Landau surfaces corresponding to their independent parts. The Landau surface corresponding to the chain of two-particle closed loops (e) is confined to the Landau surface corresponding to the single two-particle closed loop in (e) of Fig. 1.

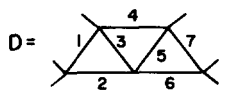
[13] H. P. Stapp, J. Math. Phys. 9, 1548 (1968).

in the diagrammatic form



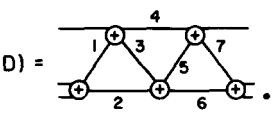
(1.2)

where the left-hand diagrams denote the part of a Landau diagram D that connects V_n to V_m, with all the lines not directly connecting V_n to V_m suppressed. The right-hand diagrams denote the corresponding part of the bubble-diagram function that gives the discontinuity across $\mathcal{M}^+[D]$, with the F box representing F_{nm}. Thus, for example, the discontinuity of M^+ across $\mathcal{M}^+[D]$, where



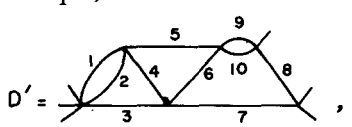
(1.3)

is represented by



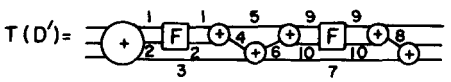
(1.4)

But if the diagram has two-particle closed loops, as does, for example,



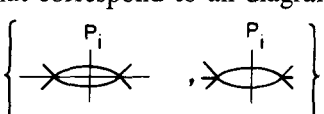
(1.5)

then, according to (1.2), an F box is added for each of these. Thus the discontinuity of M^+ around the Landau surface corresponding to (1.5) is given by



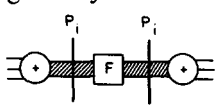
(1.6)

We also obtain formulas for the discontinuities across various classes of singularities. For example, the discontinuity of M^+ across the class of Landau surfaces that correspond to all diagrams of the form



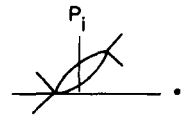
(1.7)

plus all diagrams that can be contracted to a diagram of this class is given by



(1.8)

Here the P_i bar restricts the sets of particles represented by the lines it intersects to sets having a sum of rest masses greater than a given mass M_i. A similar result is valid for all diagrams that can be contracted to a diagram of the class



(1.9)

The results described above are derived strictly from the physical-region unitarity equations. They provide a complete solution of the problem of physical-region discontinuities below the four-particle threshold. If one also admits the so-called extended unitarity equations, then it can be shown that the functions F_{nm} convert the scattering functions upon which they operate to their values on other sheets.

The present work deals only with the singularities lying in the physical region. One ultimately wants to have discontinuity formulas also for singularities lying outside the physical region, but the evident initial step is to establish the results first in the physical region, where the unitarity equations apply. On the basis of earlier work,[11] it is assumed that the physical-region singularities are confined to the union of the positive-α Landau surfaces, and that the physical continuation around these singularities follows the so-called plus-$i\epsilon$ rules. The existence of the unphysical continuations via minus-$i\epsilon$ rules is proved by using Fredholm theory, apart from possible zeros of the Fredholm denominator.

It is possible that the positive-α Landau surfaces corresponding to two different Landau diagrams are identical. Indeed, if two Landau diagrams are "equivalent," then their Landau surfaces are certainly identical. (Equivalent diagrams are diagrams having the same set of vertices and the same set of masses μ_{nm}. The mass μ_{nm} is defined, in the positive-α case, to be the sum of the rest masses of the set of particles γ_{nm} associated with the set of lines Γ_{nm} that directly connect vertex V_n to vertex V_m.) It is assumed in the present work that the positive-α Landau surfaces corresponding to basic inequivalent diagrams are nonidentical. The diagrams of Fig. 1 and those obtained from them by the procedure of the caption are all the $3 \to 3$ basic diagrams. (Generally, a basic diagram is one such that the α's are uniquely defined at some point on the positive-α Landau surfaces. The diagrams shown in Fig. 2 are not basic.) This assumption allows us to consider separately the discontinuities associated with different sets of equivalent basic diagrams.

It is possible for a positive-α surface to coincide with a Landau surface associated with α's of mixed

sign.[14] A second assumption used in the present work is that there is no cancellation between mixed-α singularities and positive-α singularities. That is, it is assumed that in the unitarity equations, and equations arising from them, the singularities associated with positive-α Landau surfaces cancel among themselves, as do the singularities associated with the mixed-α diagrams, even though these singularities may happen to occur at the same point. This seems reasonable. Since singularities in the physical region can occur only on positive-α surfaces, it would be unnatural for singularities associated with mixed-α surfaces to contribute to the physical-region discontinuities, even though a mixed-α surface might happen to coincide with a positive-α surface. This assumption allows us effectively to ignore the singularities associated with the mixed-α surfaces.

The validity of the assumptions described in the preceding two paragraphs really should be proved, but this is not attempted here.

Note added in proof: After this work was submitted we received reprints of three papers by M. J. Bloxham, D. I. Olive, and J. C. Polkinghorne containing some similar results. They obtain the formula for the discontinuity around an arbitrary physical-region singularity corresponding to a "simple diagram," which is a Landau diagram having at most one line connecting any pair of vertices. Our results for singularities corresponding to simple diagrams are special cases of their general formula.

Our main focus is on singularities associated with nonsimple diagrams. We have recently extended our results to cover all diagrams, both simple and nonsimple. The general result is that the discontinuity is obtained by replacing each vertex of the basic diagram corresponding to the singularity by the corresponding physical sheet scattering function, and replacing each set of lines α running between a pair of vertices by the inverse S_α^{-1} of the restriction S_α of S to the space corresponding to the set α. [S_α^{-1} is a generalization of the F encountered above.]

Our procedure differs significantly from that of Bloxham, Olive, and Polkinghorne. In the first place, we deal mainly with strict identities that follow from the cluster decompositions of S and S^{-1} above, and introduce analytic continuations only at the final step. Secondly, we do not separate the singularity into parts corresponding to various "mechanisms," and thus avoid the complicated question of the interference between different mechanisms, which makes the arguments of Bloxham, Olive, and Polkinghorne difficult even for the case of simple diagrams. Finally, we accept, on the basis of macrocausality, that the physical-region singularities are confined to positive-α surfaces, whereas Bloxham, Olive, and Polkinghorne derive this result. However, they accept the $i\epsilon$ rules for positive-α surfaces and assume certain similar rules for nompositive-α surfaces. Whether these rules can be justified without appealing to a causality condition is not yet known.

[14] D. Branson, Nuovo Cimento **44A**, 1081 (1966).

2. BUBBLE DIAGRAMS AND LANDAU DIAGRAMS

The basic quantities in this discussion are bubble-diagram functions. These are functions of scattering functions that can be represented by bubble diagrams.

A *bubble diagram* B is a collection of leftward-directed line segments L_j and signed circles called bubbles. Each bubble has at least one line issuing from it and at least one line terminating on it. The bubbles are partially ordered by the requirement that a bubble on which a line terminates stands left of the bubble from which this line issues. A line of B that issues from some bubble of B and also terminates on some bubble of B is called an internal line of B. The other lines of B are called external.

A circle with a plus sign inside represents the connected part of the S matrix for the process obtained by associating initial particles with lines that terminate on the bubble and final particles with lines that issue from the bubble. A circle with a minus sign inside represents the complex conjugate of the connected part of the S matrix for the transpose (initial $\leftrightarrow$ final) of that process. That is, the leftward-directed lines that terminate on a minus bubble are associated with final particles, and the leftward-directed lines issuing from the bubble are associated with initial particles.

Spin variables will be ignored.[15] Then each line L_j is associated with a variable (p_j, t_j), where t_j is an index specifying a type of particle (electron, proton, positron, etc.), the mass of which is $\mu_j \equiv \mu(t_j)$, and p_j is a physical momentum–energy vector satisfying $p_j^2 = \mu_j^2$ and $p_j^0 > 0$.

The bubble diagram B represents a corresponding function $M^B(K'; K'')$, which is just the product of the functions represented by the bubbles of B, with the understanding that there is a sum over repeated indices (variables) of these functions. In particular, for each internal line L_j of B there is a sum over all particle types t_j, and for each value of t_j, there is an integration over all physical values of the corresponding momentum vectors p_j. Thus, each internal

[15] The inclusion of spin is a trivial complication in the M-function formalism. See H. P. Stapp, Phys. Rev. **125**, 2139 (1962); "The Analytic S-Matrix Framework" in "The Trieste Lectures," *High Energy Physics and Elementary Particles* (IAEA, Vienna, 1965).

line L_j of a bubble diagram corresponds to a factor

$$\sum_{t_j} \int \frac{d^4p_j}{(2\pi)^4} 2\pi\theta(p_j^0)\delta[p_j^2 - \mu^2(t_j)], \qquad (2.1)$$

where a covariant volume element has been chosen. The integration is restricted so that topologically equivalent contributions are counted only once, as is discussed in Appendix A.

Occasionally we shall wish to restrict the sum over t_j associated with internal lines to a sum over particles having a given rest mass. Then the line L_j will be labeled by an integer, which is regarded as specifying a particular value of the rest mass.

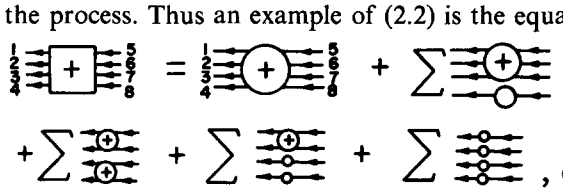
The argument $(K'; K'')$ of $M^B(K'; K'')$ is the set of variables associated with the external lines of B. The lines that issue from bubbles of B are associated in a one-to-one fashion with the variables of the set $K' = (p_1', t_1', \cdots, p_n', t_n')$, and the lines that terminate on bubbles of B are associated in a one-to-one fashion with the variables of the set $K'' = (p_1'', t_1'', \cdots, p_m'', t_m'')$. These two sets of lines are also called the outgoing and incoming lines of B, respectively.

Bose statistics is assumed throughout. Then the cluster decomposition of the S matrix is expressed by the equation[16]

$$S(K'; K'') = \sum_{B \in B_0^+(K'; K'')} M^B(K'; K''). \qquad (2.2)$$

Here $B_0^+(K'; K'')$ is the set consisting of all bubble diagrams that have only plus bubbles, have no internal lines, and have external lines specified by $(K'; K'')$. That is, the sum is over all different ways of connecting the specified set of external lines to columns of plus bubbles. Only topologically different diagrams are regarded as different; reorderings of lines on a given bubble, or reorderings of the bubbles do not give additional terms (see Appendix A).

The function $S(K'; K'')$ is represented by a box with a plus sign inside. A box with a minus sign inside represents the complex conjugate for the transpose of the process. Thus an example of (2.2) is the equation



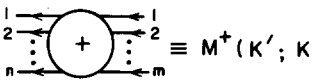
(2.3)

where the sums are over all topologically different ways of connecting the specified set of external lines to bubbles having the indicated numbers of incoming and outgoing lines. Some bubble diagrams always vanish by virtue of conservation laws and mass constraints, and these have been omitted. For example, each nontrivial bubble must connect to at least two incoming and two outgoing lines by virtue of the stability conditions on the physical masses. Trivial bubbles are bubbles from which just one line issues and upon which just one line terminates. The function corresponding to a trivial bubble is defined to be the inverse of (2.1),

$$\frac{(2\pi)^4\delta^4(p_r' - p_s'')\delta(t_r' - t_s'')}{2\pi\delta(p_r'^2 - \mu_r^2)\theta(p_r'^0)} \equiv \delta(p_r', t_r'; p_s'', t_s''), \qquad (2.4)$$

where $\delta(a - b) = \delta_{ab}$ for discrete indices. Since (2.4) holds for all trivial bubbles, the signs in these bubbles can be omitted. Often the trivial bubbles themselves are omitted.

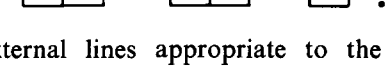
The connected part of the S matrix is denoted by



$$\equiv M^+(K'; K'') \equiv M^+_{nm}. \qquad (2.5)$$

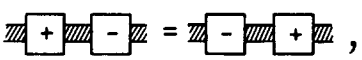
More generally, the function represented by a bubble with the symbol σ inside is denoted by $M^\sigma(K'; K'') = M^\sigma_{nm}$.

Unitarity is written in box notation as



(2.6)

The external lines appropriate to the process in question are represented by shaded strips. There is a sum over all possible numbers of lines crossing the interface of the rectangles. Sometimes these lines are indicated by writing the left side of Eq. (2.6) as



(2.7)

where the shaded strip between the boxes represents the sum over all possible numbers of lines. The right side of (2.6) is zero unless m is equal to n, in which case it is given by

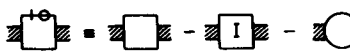
$$\delta(K'; K'') = \delta(K''; K') = \sum_{\alpha} \prod_{i=1}^{n} \delta(p_i', t_i'; p_{\alpha i}'', t_{\alpha i}''), \qquad (2.8)$$

where the α are the $n!$ permutations on n objects. The last term in (2.3) is the bubble-diagram representation of I for the case $n = 4$.

We often need to subtract from the S matrix the identity, the connected part of the S matrix, or both. The remainders are denoted by special symbols:

(2.9)

(2.10)



(2.11)

[16] See Ref. 12 for a discussion of the necessary phase factors in the Fermi case.

where a common label inside the boxes and bubbles of each equation has been suppressed.

It was shown in Ref. 13 that all singularities of bubble-diagram functions are associated with Landau diagrams.[17] A Landau diagram D is a set of directed line segments L_j and a set of vertices V_n. Each vertex contains end points of three or more lines L_j, and no vertex contains both end points of any one line. Denoting the leading and trailing end points of the line segment L_j by the symbols L_j^+ and L_j^-, respectively, one can characterize the diagram D by the set of numbers ϵ_{jn} defined by

$$\epsilon_{jn} = 1 \quad \text{if} \quad L_j^+ \subset V_n,$$
$$\epsilon_{jn} = -1 \quad \text{if} \quad L_j^- \subset V_n,$$

or

$$\epsilon_{jn} = 0 \quad \text{otherwise.}$$

Each line L_j of D is associated with a particle of type t_j and mass μ_j. If particles of type t_j carry a_j units of an additively conserved quantum number "a," then the conditions

$$\sum_j a_j \epsilon_{jn} = 0 \quad (\text{all } n) \tag{2.12}$$

are required of D.

The lines L_j of D are characterized as being incoming, outgoing, or internal according to the following rules:

L_j is incoming if $\epsilon_{jn} \geq 0$ for all n,

L_j is outgoing if $\epsilon_{jn} \leq 0$ for all n,

L_j is internal otherwise.

The incoming and outgoing lines of D are called external lines of D. A line that is both incoming and outgoing is called an unscattered line.

A *connected* Landau diagram is an arcwise-connected Landau diagram. A *trivial* Landau diagram is a connected Landau diagram with no internal lines.[18]

For each Landau diagram D there is corresponding Landau surface $\mathcal{M}[D]$. The surface $\mathcal{M}[D]$ is the set of variables (p_j, t_j) associated with the external lines of D via associations

$$L_j \leftrightarrow (p_j, t_j, \alpha_j) \tag{2.13}$$

that satisfy the (loop) equations

$$\sum_j \alpha_j p_j n_{jf} = 0 \quad (\text{all } f), \tag{2.14}$$

the mass constraints

$$p_j^2 = \mu_j^2 \quad (\text{all } j), \tag{2.15}$$

and the conservation laws (2.12). A particular case of (2.12) is the momentum–energy conservation law

$$\sum_j p_j \epsilon_{jn} = 0 \quad (\text{all } n). \tag{2.16}$$

In Eq. (2.14), n_{jf} is the algebraic number of times the Feynman loop f passes along line L_j in the positive sense, and the α_j associated with each internal line L_j is a *nonzero* number. The α_j associated with the external L_j play no role, and can be set equal to zero. Each p_j is a real energy–momentum vector with $p_j^0 > 0$. The part of $\mathcal{M}[D]$ that can be realized with all α's positive is denoted by $\mathcal{M}^+[D]$.

A connection between Landau diagrams and bubble diagrams is set up using the following terminology. A Landau diagram D *corresponding* to a bubble b is a D with its external lines in one-to-one correspondence with the lines of b. The incoming lines of D are to correspond to the lines terminating on b, and the outgoing lines of D are to correspond to the lines issuing from b. The internal lines of a D corresponding to a bubble b will be said to lie *inside* b.

A $D' \subset B$ is a Landau diagram D' that can be constructed by replacing each nontrivial bubble b of B by a corresponding connected diagram D_c^b, which might be simply a trivial point vertex V^b. This D_c^b is required to be such that $\mathcal{M}^+[D_c^b]$ is nonempty. Lines containing only trivial bubbles are replaced by lines containing no vertices.

A contraction $D \supset D'$ of a Landau diagram D' is a Landau diagram D that can be obtained by shrinking to points certain of the internal lines L_i of D', and then removing all those lines L_j of the resulting diagram for which L_j^+ and L_j^- coincide. The diagram D' is considered a trivial contraction of itself.

A $D \supset \subset B$ is a D such that for some $D' \subset B$, D is a $D \supset D'$. The phrase B *supports* D means that D is a $D \supset \subset B$. Thus, for example, the bubble diagram B of (1.4) supports the Landau diagram D of (1.3). This terminology is used in the next section to describe the locations of the singularities of bubble-diagram functions.

3. LANDAU SINGULARITIES, STRUCTURE THEOREMS FOR BUBBLE DIAGRAMS, AND THE $i\epsilon$ RULE

In this section, we summarize some pertinent results obtained earlier regarding the location and nature of singularities of bubble-diagram functions.

According to the First Structure Theorem of Ref. 13, the singularities of $M^B(K'; K'')$ [divided by $(2\pi)^4\delta^4(\sum p_j' - \sum p_j'')$] for any connected B are confined to the closure of the union over nontrivial $D \supset \subset B$ of the Landau surfaces $\mathcal{M}[D]$.[17]

[17] A similar result was obtained by J. C. Polkinghorne, Nuovo Cimento **25**, 901 (1962).

[18] In Ref. 13, trivial Landau diagrams were not included among the set of Landau diagrams.

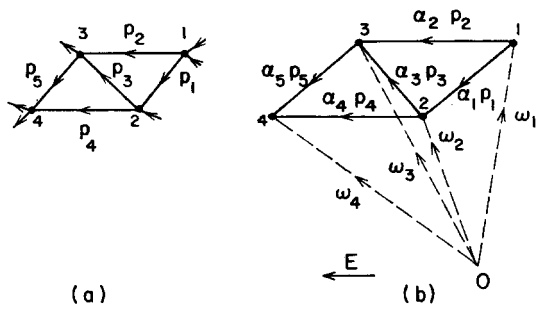


FIG. 3. Figure (a) shows a diagram D, and Fig. (b) shows a corresponding $\bar{D}$ with all α's positive. In diagram (b), positive-energy vectors point left. The condition that all α's be positive, together with the energy-conservation law, ensures that the diagram $\bar{D}$ has the "physical" ordering, with energy flowing into the right-most vertex V_1 and out of the left-most vertex V_n. If the signs of the α's are all reversed, then the relative positions of the vertices of the new $\bar{D}$ are obtained by reflecting diagram (b) through the origin 0.

A concrete representation for the surface $\mathcal{M}[D]$ is now described.[19]

The geometric significance of the loop equations is that the set of momentum–energy vectors

$$\Delta_j \equiv \alpha_j p_j \tag{3.1}$$

fit together to form a momentum–energy space diagram $\bar{D}$ that is topologically equivalent to the diagram D. That is, the directed internal line segment L_j of D leading from a vertex m to a vertex n (i.e., $\epsilon_{jm} = -1$, $\epsilon_{jn} = 1$) is mapped to the four-vector

$$\Delta_j = \omega_n - \omega_m = \sum_r \epsilon_{jr}\omega_r \tag{3.2}$$

of $\bar{D}$, where ω_r is the four-vector from some arbitrary origin to the vertex V_r of $\bar{D}$. The allowed values of the ω_n are those values such that each vector

$$(\omega_n - \omega_m)\epsilon_{jn}|\epsilon_{jm}| = \alpha_j p_j \epsilon_{jn}|\epsilon_{jm}|$$

is positive timelike, negative timelike, or zero, according to whether $\alpha_j\epsilon_{jn}|\epsilon_{jm}|$ is positive, negative, or zero. This is just the requirement that p_j be positive timelike. A typical diagram $\bar{D}$ is shown in Fig. 3.

Each diagram $\bar{D}$ corresponds to exactly one point on $\mathcal{M}[D]$, and each point of $\mathcal{M}[D]$ corresponds to at least one diagram $\bar{D}$. The correspondence is given by the mapping function

$$q_n(\omega) = \sum_{m\neq n} \frac{\omega_n - \omega_m}{|\omega_n - \omega_m|}\mu_{nm}, \tag{3.3}$$

where

$$\mu_{nm} = \sum_j \mu_j |\epsilon_{jn}\epsilon_{jm}\alpha_j|/\alpha_j \tag{3.4}$$

and where the denominator is the Lorentz length

$$|\omega_n - \omega_m| \equiv [(\omega_n - \omega_m)(\omega_n - \omega_m)]^{\frac{1}{2}}. \tag{3.5}$$

[19] An independent derivation of this representation was given by A. A. Logunov, I. T. Todorov, and N. A. Chernikov in the *Proceedings of the 1962 International Conference on High Energy Physics at* CERN (CERN, Geneva, 1962), p. 695; and Nucl. Phys. **50**, 273 (1964).

This length is necessarily positive for $\mu_{nm} \neq 0$, since $\omega_n - \omega_m$ is timelike in this case. The terms where the $\mu_{nm} = 0$ do not contribute to (3.3).

The vector q_n occurring in (3.3) is the total outgoing momentum at vertex V_n,

$$q_n = -\sum_{\text{ex } j} p_j\epsilon_{jn}, \tag{3.6}$$

where the sum over j runs over the j corresponding to external lines of D. The Landau surface $\mathcal{M}[D]$ depends on the external p_j only through these combinations q_n. If no external lines are incident on vertex V_n, then q_n is required to vanish. If exactly one external line is incident on V_n, then q_n is required to satisfy the corresponding mass constraint.

Equation (3.3) is obtained by first using the conservation law (2.16) to convert (3.6) to a sum over internal lines of D and then using (3.1) and (3.2):

$$\begin{aligned} q_n &= \sum_{\text{int } i} p_i\epsilon_{in} = \sum_{\text{int } i} \frac{\Delta_i}{\alpha_i}\epsilon_{in} \\ &= \sum_{\text{int } i} \epsilon_{in} \sum_m \epsilon_{im}\omega_m/\alpha_i \\ &= -\sum_{\text{int } i} \sum_{m\neq n} \epsilon_{in}\epsilon_{im}(\omega_n - \omega_m)/\alpha_i \\ &= \sum_{\text{int } i} \sum_{m\neq n} |\epsilon_{in}\epsilon_{im}|\,(\omega_n - \omega_m)/\alpha_i, \end{aligned} \tag{3.7}$$

where we have observed that for a given internal i only two values of m give a nonzero ϵ_{im}, and that these two ϵ_{im} have opposite signs. From (3.1) and (3.2), one obtains

$$\frac{|\alpha_i|\,\mu_i}{|\omega_n - \omega_m|} = 1, \tag{3.8}$$

which combines with (3.7) to give (3.3):

$$\begin{aligned} q_n &= \sum_{m\neq n} \sum_{\text{int } i} |\epsilon_{im}\epsilon_{in}\alpha_i| \frac{(\omega_n - \omega_m)\mu_i}{|\omega_n - \omega_m|\,\alpha_i} \\ &= \sum_{m\neq n} \sum_i |\epsilon_{im}\epsilon_{in}\alpha_i| \frac{(\omega_n - \omega_m)\mu_i}{|\omega_n - \omega_m|\,\alpha_i}. \end{aligned} \tag{3.9}$$

A point $q(\omega)$ of $\mathcal{M}[D]$ such that the first-order variations $\delta q = (\partial q/\partial\omega)\delta\omega$ generate the tangent space to $\mathcal{M}[D]$ at $q(\omega)$ is called a *simple point* $q(\omega)$ of $\mathcal{M}[D]$. If $\mathcal{M}[D]$ is considered as a surface in $q \equiv \{q_n\}$ space, then the tangent space to $\mathcal{M}[D]$ at a simple point $q(\omega)$ of $\mathcal{M}[D]$ lies in the linear manifold defined by

$$q \cdot \omega = q(\omega) \cdot \omega, \tag{3.10}$$

where

$$q \cdot \omega \equiv \sum_n q_n \cdot \omega_n \equiv \sum_{n,\mu} q_n^\mu \omega_{n\mu} \equiv \sigma(q, \omega). \tag{3.11}$$

This follows from the fact that, at $\omega' = \omega$,

$$\partial\sigma[q(\omega'), \omega]/\partial\omega' = 0. \tag{3.12}$$

Equation (3.12) is readily verified by substituting the

right side of (3.3) into (3.11) and taking the partial derivative with respect to a component $\omega'_{n\mu}$ of ω'. The vector ω is, in this sense, a normal vector to $\mathcal{M}[D]$ at a simple point $q(\omega)$ of $\mathcal{M}[D]$.

The mapping $q(\omega)$ is not one-to-one. All values of ω that are related by changes of the origin or by rescalings of the α_i give the same $q(\omega)$ and are called *equivalent*. It is convenient to fix the origin by requiring $\sum \omega_n = 0$. Then ω lies in the same manifold as q, which is restricted by $\sum q_n = 0$ due to momentum–energy conservation. The scaling can be fixed up to a single sign by requiring that

$$\sum_{n>m} \sum_i |\omega_m - \omega_n| \, |\epsilon_{in}\epsilon_{im}| \equiv \sum_{\text{int } i} |\alpha_i p_i| = 1.$$

As already mentioned, the first structure theorem asserts that the singularities of $M^B(K'; K'') \equiv M^B(K)$ for any connected B are confined to the closure of the union over nontrivial $D \supset \subset B$ of the Landau surfaces $\mathcal{M}[D]$. Let this union be denoted by $\mathcal{M}^B$. A *simple point* of $\mathcal{M}^B$ is a point such that a complete neighborhood of K in $\mathcal{M}^B$ is generated by an arbitrarily small neighborhood of some unique (up to equivalence) point ω, for some unique nontrivial $D \supset \subset B$. According to the second structure theorem of Ref. 13 such a point cannot actually be a singularity of $M^B(K)$ unless $\bar{D}$ can be realized by taking all $\alpha_i \eta_i \geq 0$, where η_i is the sign of the bubble inside which L_i lies, or is zero in case L_i does not lie inside any bubble of B. That is, we can require the α_i of lines L_i of $D \supset \subset B$ that lie in plus (minus) bubbles to be positive (negative), but the α_i of lines of $D \supset \subset B$ that are also lines of the original bubble diagram B are allowed to be either positive or negative.[20]

According to the third structure theorem of Ref. 13 the functions $M^B(K)$ lying on the two sides of the singularity surface at a simple point $\bar{K}$ of $\mathcal{M}^B$ are boundary values of a single function analytic near $\bar{K}$ in the upper-half $\sigma(K; \bar{K}) \equiv \sigma[q(K), \omega(\bar{K})]$ plane, provided at least one line of the corresponding $D \supset \subset B$ lies inside some bubble b of B. The sign of σ is fixed through (3.1) and (3.2) by the requirement $\alpha_i \eta_i \geq 0$, which has force only if at least one line L_i of D lies inside some bubble b of B.

In accordance with the second structure theorem, we may consider the signs of the α_i to be restricted by the condition $\alpha_i \eta_i \geq 0$. Then the singularity at a simple point of $\mathcal{M}^B$ can be classified as either a positive-α, negative-α, or mixed-α singularity, according to whether the various α_i corresponding [via (3.1) and (3.2)] to the singularity are all positive, all negative, or neither all positive nor all negative. Thus, if B consists of a single plus bubble, then all its singularities are positive-α singularities, but if B consists of a single minus bubble, then all of its singularities are negative-α singularities. The positive-α and negative-α singularity surfaces corresponding to a diagram D both occupy the same position, $\mathcal{M}^+[D]$, but the sign of $\sigma(q, \omega)$ is opposite, and hence the two continuations pass into opposite half-planes. A continuation in accordance with the rules for going around a positive-α singularity is called a continuation according to the plus-$i\epsilon$ rule, whereas a continuation in accordance with the rules for going around a negative-α singularity is called a continuation according to the minus-$i\epsilon$ rule.

The function $q(K)$ appearing in $\sigma(K; \bar{K}) \equiv \sigma[q(K); \omega(\bar{K})]$ is the expression for the q_n in terms of external vectors given in (3.6). One can also write q as a function of the internal vectors p_i as in (3.7). This gives, again via (3.2) and then (3.1),

$$\begin{aligned} \sigma[q(p_i), \omega] &= \sum_n q_n(p_i) \cdot \omega_n \\ &= \sum_n \sum_{\text{int } i} \epsilon_{in} p_i \cdot \omega_n \\ &= \sum_{\text{int } i} p_i \cdot \Delta_i(\omega) \\ &= \sum_{\text{int } i} \alpha_i p_i \cdot p_i(\omega). \end{aligned} \tag{3.13}$$

The vector $p_i(\omega)$ is a real positive-energy vector satisfying $p_i(\omega)^2 = \mu_i^2$. Hence the minimum value of $p_i \cdot p_i(\omega)$, when the real vector p_i is restricted by $p_i^2 = \mu_i^2$ and $p_i^0 > 0$, is precisely $\mu_i^2 = p_i(\omega) \cdot p_i(\omega)$. Thus if all the α_j are positive, then we have

$$\sigma[q(p_i), \omega] \geq \sigma[q(p_i(\omega)), \omega]. \tag{3.14}$$

That is, the function $\sigma[q(p_i), \omega]$ takes its minimum value at $p_i = p_i(\omega)$.

Equation (3.14) has several important consequences. Consider any Landau diagram D. The "physical region" of D is the set of all points in the space of external variables such that there is a set of real p_i associated with the internal lines of D for which the mass constraints (2.15), the energy condition $p_i^0 > 0$, and the conservation laws (2.16) are satisfied. Equation (3.14) says that for any ω corresponding to a point on $\mathcal{M}^+[D]$, the entire physical region of D, considered as a region in q space, lies on the positive side of the hyperplane $\sigma(q, \omega) = \sigma[q(\omega), \omega]$, except for the point of contact $q = q(\omega)$. This fact was first noticed by Pham.[12] It tells us in particular that the points of $\mathcal{M}^+[D]$ are necessarily on the boundary of the physical region of D (a result that is not always true on $\mathcal{M}[D] - \mathcal{M}^+[D]$) and moreover that $\sigma(q, \omega)$ increases (rather than decreases) as one moves from outside the physical region to inside the physical

[20] A similar result was obtained by P. V. Landshoff and D. I. Olive in the Appendix of Ref. 6.

region at the point $q(\omega)$. It follows from this that the continuation according to the plus-$i\epsilon$ rule around any positive-α singularity is such that it passes into the upper half-plane in any variable z for which $\delta q = (\partial q/\partial z)\delta z$ moves the point q from $\mathcal{M}^+[D]$ into the physical region of D when δz is real and positive. That is, the requirement that $\delta\sigma = \delta q \cdot \omega$ be positive for positive δz implies that $(\partial q/\partial z) \cdot \omega > 0$, which means that for $\operatorname{Im} \delta z > 0$ one has

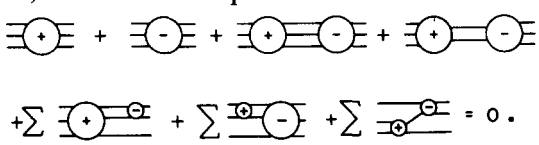
$$\delta \operatorname{Im} \sigma = \omega \cdot \operatorname{Im} \delta q = \omega \cdot (\partial q/\partial z) \operatorname{Im} \delta z > 0.$$

Thus, the continuation according to the plus-$i\epsilon$ rule always goes into the *upper half plane* of a variable that is increasing as one moves *into* the physical region.

The Landau diagram $D \subset B$ obtained by replacing each bubble b of B by a point vertex V^b is denoted by D^B. The function M^B is not generally continuable past the singularity at $\mathcal{M}[D^B]$. Indeed, the above result shows that the function M^B vanishes on one side of $\mathcal{M}^+[D^B]$ but not on the other side. These two sides are called the unphysical and physical sides, respectively. The function M^B evidently has both positive-α and negative-α singularities at $\mathcal{M}^+[D^B]$, since the constraints $\alpha_i\eta_i \geq 0$ allow for $\bar{D}$ with either all α_i positive or all negative, since all the η_i are zero.

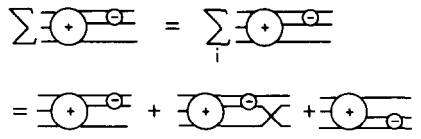
4. DEVELOPMENT OF THE UNITARITY EQUATION FOR THE $3 \to 3$ SCATTERING AMPLITUDE

Suppose the center-of-mass energy E of the initial (or final) particles is below the four-particle threshold. Then the connected part of the unitarity equation (2.6) for the $3 \to 3$ amplitude can be written in the form



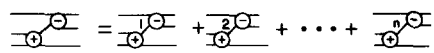
(4.1)

The unmarked summation signs in (4.1) and in the subsequent equations refer to the external lines (incoming and outgoing). Those that are marked "i" or "f" refer only to incoming or outgoing lines, respectively. These sums are over all topologically different ways of connecting the specified set of external lines to the rest of the bubble diagram. For example, the last term in (4.1) has nine contributions, whereas the fifth term has three:



(4.2)

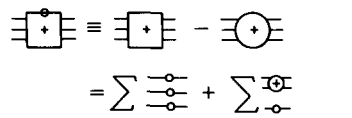
The external lines, reading from top to bottom, represent fixed variables (p_j, t_j), but the internal lines are subject to the summation convention of Sec. 2 unless it is restricted by explicitly labeling an internal line by an integer. Thus, for instance, we have



(4.3)

where the number r on an internal line restricts the summation to particles of a fixed mass M_r.

Postmultiplying (4.1) by

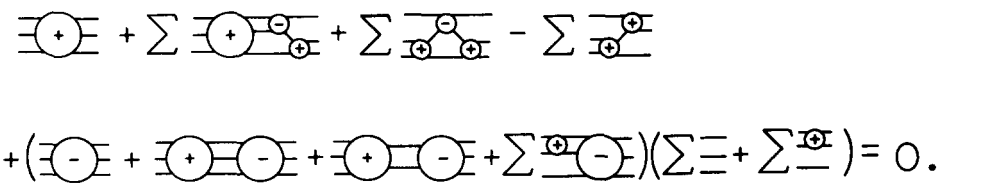


(4.4)

and simplifying by means of two-particle unitarity,

(4.5)

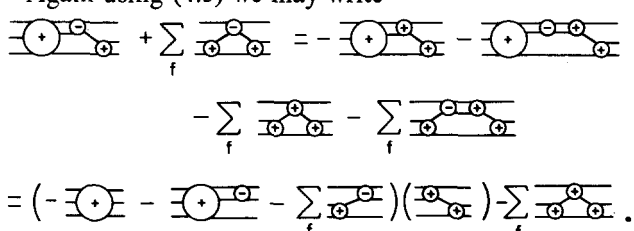
we obtain



(4.6)

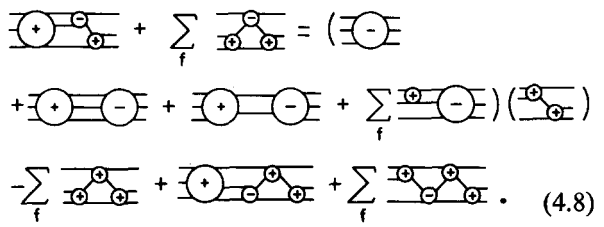
[A detailed discussion of combinatorial questions is given in Appendix A. Our rules are such that the products of functions represented by diagrams such as (4.1) and (4.4) are always represented simply as the sum of the topologically distinct diagrams that can be obtained by combining the diagrams in the natural fashion. By product we mean, of course, matrix product; there is an integration (2.1) over the internal lines.]

Again using (4.5) we may write



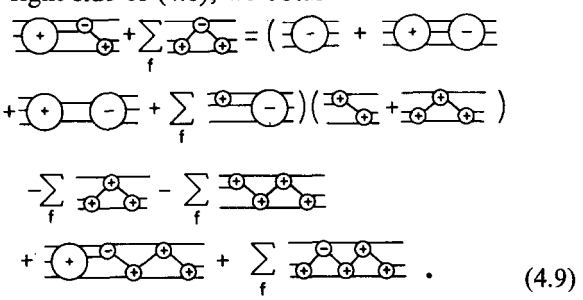
(4.7)

Application of (4.1) to the expression in the left parentheses gives



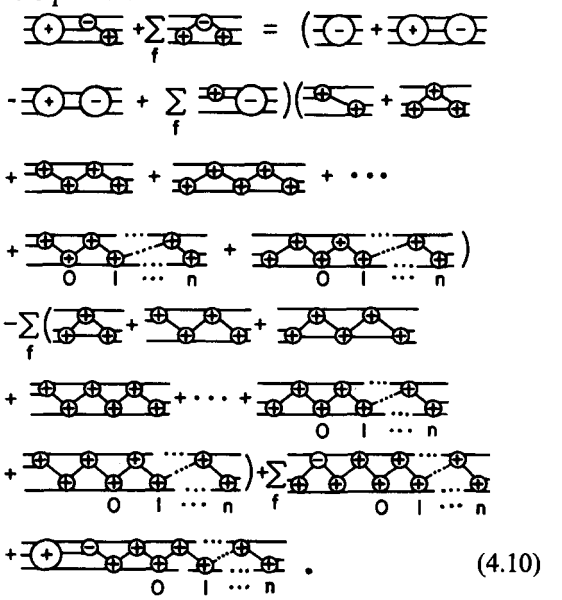
(4.8)

There is an equation similar to (4.8) but with the right plus bubble connected to the upper two lines. Substituting that equation into the last two terms of the right side of (4.8), we obtain



(4.9)

Equation (4.9) can be iterated by substituting the right-hand side of (4.9) into the last two terms of the right-hand side of (4.9). Iterating n times, we end up with the equation

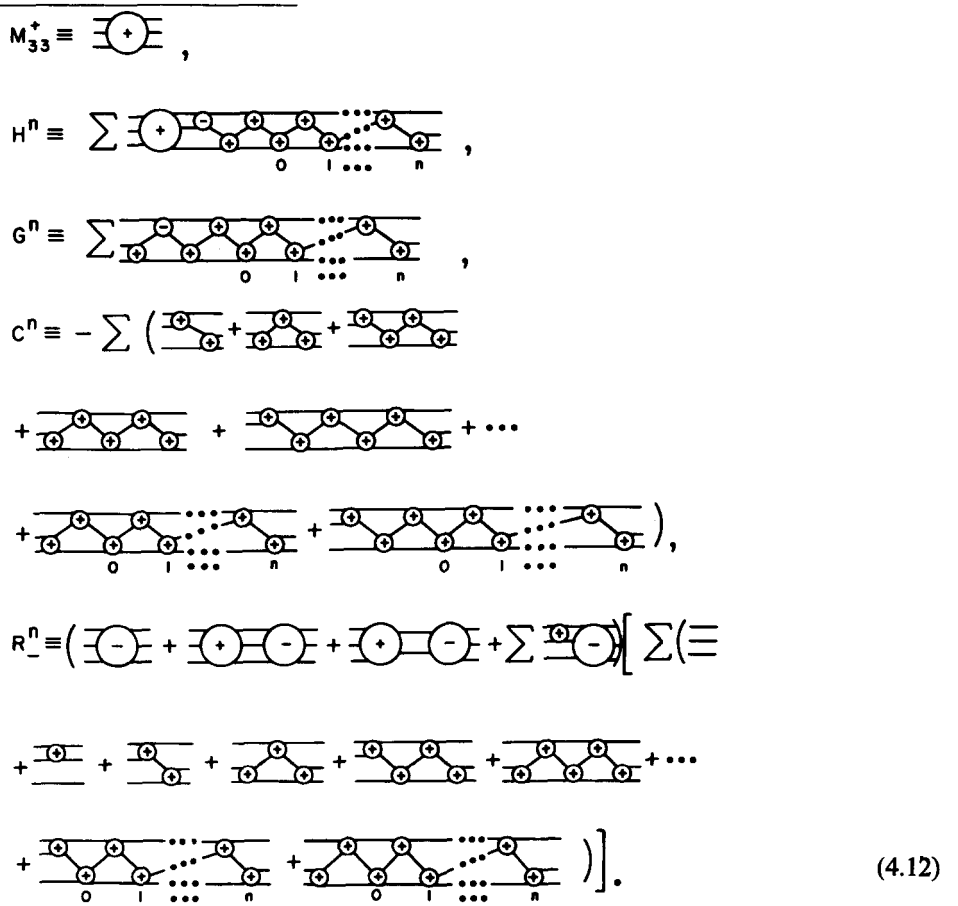


(4.10)

Substituting (4.10) into (4.6), we obtain for any positive integer n

$$M_{33}^{+} + H^{n} + G^{n} + C^{n} + R_{-}^{n} = 0, \tag{4.11}$$

where



(4.12)

5. DERIVATION OF DISCONTINUITY EQUATIONS

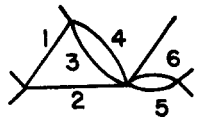
A. Discontinuities across Singularities Depending on a Cross-Energy

Certain Landau diagrams D are such that every continuous curve within D connecting any incoming line to any outgoing line must pass through one and the same vertex V (see Fig. 4). The Landau surfaces corresponding to such diagrams can depend on the total energy and on the various subenergies, but with a suitable choice of variables they are independent of all cross-energy invariants.[21] Landau surfaces $\mathcal{M}[D]$ corresponding to diagrams D of this type are called surfaces of Type A. Singularities not lying on surfaces of Type A are called cross-energy singularities. The discontinuity equations for all cross-energy singularities of M_{33}^{+} can be obtained directly from (4.11) and the structure theorems.

The singularities of the various terms in (4.11) must cancel. Since M_{33}^{+} is regular at points not lying on $\mathcal{M}^{+}$, the singularities of $H^n + G^n + C^n + R_{-}^n$ corresponding to α's of mixed sign (mixed-α singularities) must cancel among themselves at these points. As discussed in the introduction, we shall assume that this cancellation among mixed-α singularities also holds true at points of $\mathcal{M}^{+}$. We also assume that the surfaces $\mathcal{M}^{+}[D]$ corresponding to inequivalent basic diagrams D are nonidentical and consider points lying on a single one of these surfaces.

By virtue of the first assumption, we can, in deriving the discontinuities of M_{33}^{+}, ignore the mixed-α singularities of the various terms in (4.11). According to the first structure theorem, the singularities of M^B are confined to the union over $\mathcal{M}[D]$ of the nontrivial $D \supset \subset B$. Consider first G^n. (We sometimes use the same symbol to denote both the bubble diagram and its function.) If any one of the diagonal lines of G^n is contracted to give D, then $\mathcal{M}[D]$ is of Type A. This is because the only allowed diagrams in the $2 \to 2$ bubbles must have, as a consequence of the stability and the positive- (or negative-) α requirements, both incoming lines terminating on a common vertex and both outgoing lines issuing from a common vertex. On the other hand, if any of the horizontal lines of G^n is contracted, then the two diagonal lines

FIG. 4. An example of a Landau diagram containing a vertex V through which must pass every path connecting any initial line to any final line. A path connecting lines is required to connect internal points of these lines.

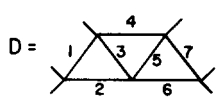


lying above or below this horizontal line must also be contracted, as a consequence of the Landau loop equations and the positive- (or negative-) α requirement. Thus, all the positive- (or negative-) α cross-energy singularities of G^n correspond to diagrams D in which none of the (explicitly appearing) lines of G^n are contracted. But for sufficiently large n, the $\mathcal{M}^{+}[D]$ for such a D does not enter the region below the four-particle threshold. (This was proved in Ref. 22 by proving that for sufficiently large n, the classical point-particle multiscattering process pictured by D is dynamically impossible.)

A similar argument shows that H^n can have no positive- (or negative-) α cross-energy singularities, provided n is taken sufficiently large.

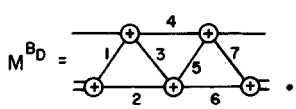
The diagrams $D \supset \subset B$ for $M^B \in R_{-}^n$ have, for the Landau diagrams D_c^b corresponding to the $3 \to 3$ or the $3 \to 2$ minus bubbles, either a point vertex V^b or a nontrivial diagram with internal lines. In the former case, $\mathcal{M}[D]$ is of type A. In the latter case, the fact that D contains some line that lies inside a minus bubble means that the α_i of at least one line of $\bar{D}$ must be negative. Since all α_i must have the same sign, this universal sign must be negative. Thus, the σ of the third structure theorem that defines the continuation past this singularity of R_{-}^n is the negative of the σ associated with the same singularity of M_{33}^{+}. That is, the continuation around this singularity of R_{-}^n follows the *minus-i*ϵ rule.

We may now derive the discontinuity of M_{33}^{+} across the singularity corresponding to the Landau surface $\mathcal{M}^{+}[D]$ associated with any diagram D that may be formed by contracting to points the bubbles of a bubble diagram B_D having labeled lines that represents a contribution to C^n. For example, the diagram



(5.1a)

corresponds to the contribution to C_n



(5.1b)

According to the remarks at the end of Sec. 3, the term $T(D) \equiv M^{B_D}$ is a threshold term present in

[21] A subenergy is the energy of a proper subset of initial or final particles in its own center-of-mass frame. A cross-energy invariant is the square of the sum of the momentum–energy vectors of a set of particles that contains both initial and final particles. The eight invariants needed to describe a $3 \to 3$ reaction are chosen to include two initial subenergies, two final subenergies, and the total center-of-mass energy E of the reaction. Then all six subenergies are functions only of these five invariants and are therefore independent of the remaining three variables, which are cross-energy invariants. The choice of the invariants is discussed by V. E. Asribekov, Zh. Eksp. Teor. Fiz. **42**, 565 (1962) [Soviet Phys.—JETP **15**, 394 (1962)].

[22] H. P. Stapp, J. Math. Phys. **8**, 1606 (1967).

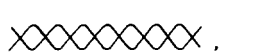
(4.11) for points on the physical side of $\mathcal{M}^+[D]$, but absent on the unphysical side. If one takes Eq. (4.11) on the unphysical side and continues around the singularity in accordance with the minus-$i\epsilon$ rule, then R^n_- is continued into the same function R^n_- that occurs on the physical side of the $\mathcal{M}^+[D]$, by virtue of the third structure theorem. However, the function M^+_{33} is continued to its value underneath the cut (or underneath the pole). This continuation is denoted by $M^+_{33}(D^-)$. The remaining terms are not singular at the point in question and are therefore continued into the same functions that occur on the physical side of $\mathcal{M}^+[D]$. Thus, the difference between the continuation of (4.11) from the unphysical side and (4.11) evaluated on the physical side of $\mathcal{M}^+[D]$ gives simply

$$\Delta_D M^+_{33} \equiv M^+_{33} - M^+_{33}(D^-) = M^{B_D} \equiv T(D). \quad (5.2)$$

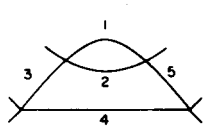
That is, the discontinuity around the singularity surface $\mathcal{M}^+[D]$ is just M^{B_D}.

The essential point in this method is that the expansion of M^+_{33} be such that all positive-α contributions to the singularity in question that are not in M^+_{33} itself are contained in a "threshold term," which is a term that contributes on only one side of the singularity surface in question. This is the key point of this paper.

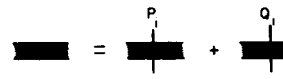
The surfaces $\mathcal{M}^+[D]$ considered above are not the only positive-α Landau surfaces that are not of Type A. Other D's can be obtained by inserting some internal structure in some of the plus bubbles of a bubble diagram corresponding to a term of C^n. The stability and positive-α requirements, together with the total-energy limitation, allow the insertions only of chains of the form

$D^b_c =$  , (5.3)

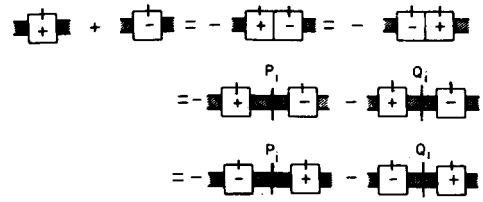
where the Landau equations require that the sum of the masses of each link of the chain be the same. Since the Landau surfaces corresponding to such a chain do not depend on the number $n \geq 1$ of closed loops, we can, without loss of generality, limit the chains to a single closed loop. Then, we obtain, for instance, the Landau diagram

$D_1 =$  . (5.4)

In order to obtain the discontinuity equation for the cut starting at $\mathcal{M}^+[D_1]$, and also for later use, we now cast the unitarity equation into a form where only the scattering amplitude itself and a threshold term have a positive-α singularity corresponding to a given *normal* threshold. Let us introduce the decomposition

 , (5.5)

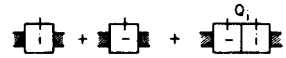
where the P_i (or Q_i) bar restricts each of the sets of particles corresponding to the lines intersected by the bar to a set having a sum of rest masses greater than or equal to (or smaller than) a given mass M_i. Using (2.9) and (5.5), we can write (2.6) in the form

 . (5.6)

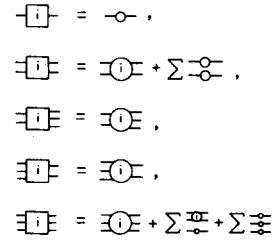
Instead of the bars, we often use the more compact notation exhibited in

 . (5.7)

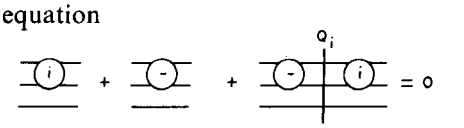
Denoting the center-of-mass energy of the incoming particles by E, we see that only the second term on the right-hand side of (5.6) contributes below $E = M_i$. We define the *i box* by the equation[23]

 (5.8)

for both $E < M_i$ and $E > M_i$. (In this equation, and those that follow, E is assumed to be below the four-particle threshold.) An i box is related to i bubbles in the same way as a plus box is related to plus bubbles. Thus, we have

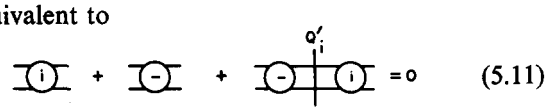
 (5.9a)–(5.9e)

The functions represented by the i bubbles are denoted by $M^i(K'; K'')$. These functions depend on the mass M_i associated with Q_i. The $2 \to 2$ i bubble in (5.9e) is defined by the required vanishing of the various disconnected parts of (5.8). The disconnected part equation

 (5.10)

[23] Cf. J. Gunson, Ref. 2, and D. I. Olive, Nuovo Cimento **29**, 326 (1963).

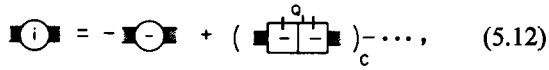
is equivalent to



(5.11)

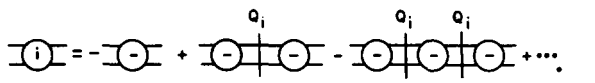
if the mass M_i' associated with the Q_i' of (5.11) is related to the mass M_i associated with the Q_i of (5.10) by $M_i' = M_i - \mu_i$, where μ_i is the rest mass of the spectator particle in (5.10).

It is shown in Appendix B that (5.8) can be solved for $M^i(K'; K'')$ by using Fredholm theory and that $M^i(K'; K'')$ has a minus-$i\epsilon$ continuation past the normal threshold singularity at $E = M_i$. This result follows formally, directly from the iterative solution,



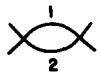
(5.12)

where the subscript C indicates the connected part of the expression in parentheses. For example, in the $2 \to 2$ case (5.12) becomes



(5.13)

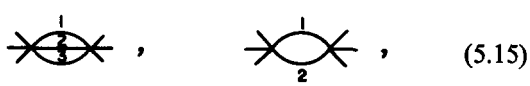
Let B be any term on the right-hand side of (5.13) and consider the normal threshold corresponding to the Landau diagram D



(5.14)

where $M_i = \mu_1 + \mu_2$. Because of the Q_i projections, any $D \supset \subset B$ must have negative α's associated with its lines. Then the third structure theorem prescribes a minus-$i\epsilon$ continuation past the singularity of M^B at $\mathcal{M}^+[D]$. Thus each term on the right-hand side of (5.13) must follow a minus-$i\epsilon$ rule past this normal threshold.

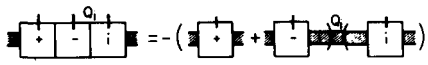
Similarly, it follows from (5.12) that M_{33}^i has a minus-$i\epsilon$ continuation past the normal thresholds corresponding to the diagrams



(5.15)

where we put M_i equal to $\mu_1 + \mu_2 + \mu_3$ or $\mu_1 + \mu_2$, respectively. These results can be made rigorous by using Fredholm theory in place of (5.12), as is shown in Appendix B.

From (5.6) and (5.8) we obtain[24]



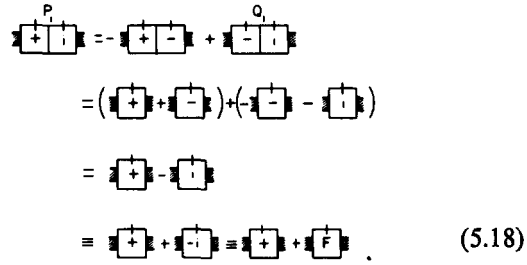
(5.16)

[24] This argument is taken from Ref. 3.

and also

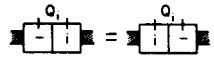
(5.17)

Comparing (5.16) and (5.17), we obtain



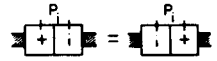
(5.18)

In Appendix C, we show that



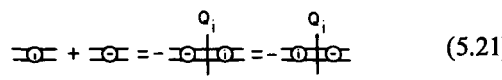
(5.19)

Using (5.19), one also obtains (5.18) with the i box and plus box interchanged on the left-hand side, and hence also



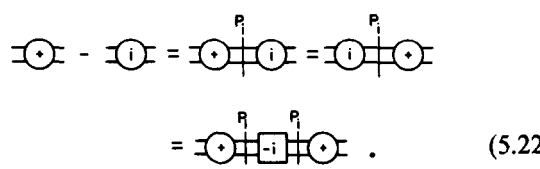
(5.20)

It follows from (5.11) and (5.19), and from (5.18), (5.20), and (2.9), that



(5.21)

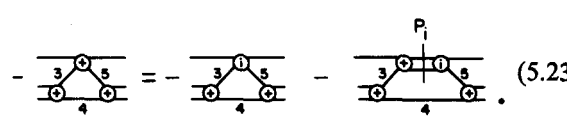
and that



(5.22)

In (5.22) only the first term on the left-hand side and the threshold term on the right-hand side have a positive-α singularity corresponding to the diagram (5.14) where $M_i = \mu_1 + \mu_2$. This shows that M_{22}^i is the continuation of M_{22}^+ below this singularity and that the discontinuity is the right side of (5.22).

Returning now to the problem of the discontinuity around $\mathcal{M}^+[D_1]$, where D_1 is defined in (5.4), we use (5.22) to expand the relevant term T of C^n as follows:



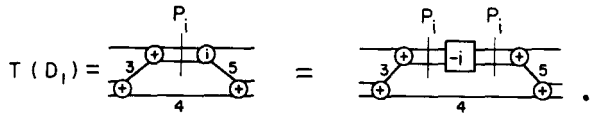
(5.23)

Setting M_i equal to $\mu_1 + \mu_2$, we see that the first term of this expansion has a minus-$i\epsilon$ continuation past the surface $\mathcal{M}^+[D_1]$ for just the same reason that R_-^n does.

The second term has its threshold at $\mathcal{M}^+[D_1]$. Therefore, we obtain in the same way as before the discontinuity equation

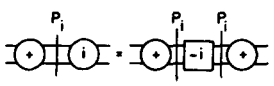
$$\Delta_{D_1} M_{33}^+ = T(D_1), \tag{5.24}$$

where



(5.25)

The method given above can be generalized. Any set of inner plus bubbles (that is, plus bubbles not standing at the extreme right or the extreme left of the bubble diagram) of any bubble diagram T corresponding to a term T of C^n can be replaced by closed loops, while the remaining plus bubbles are contracted to points. The corresponding discontinuity of M_{33}^+ is then given by T with every plus bubble corresponding to a two-particle closed loop replaced by

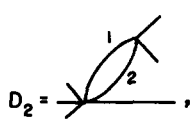


(5.26)

with the index i referring to the sum of the rest masses of the two particles in the two-particle closed loop.

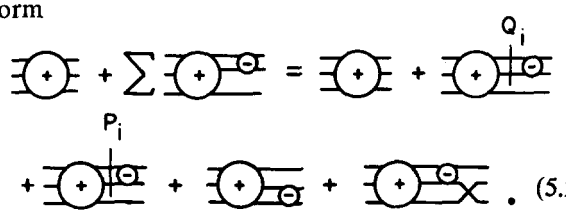
B. Discontinuity Equation for Normal Singularities in a Subenergy

To derive the discontinuity of M_{33}^+ across $\mathcal{M}^+[D_2]$, where



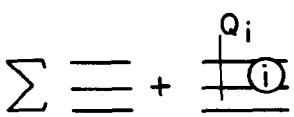
(5.27)

we first write the first and fifth term of (4.1) in the form



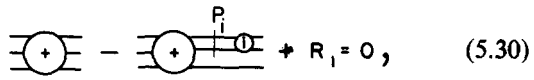
(5.28)

Substituting (5.28) into (4.1), postmultiplying the result by



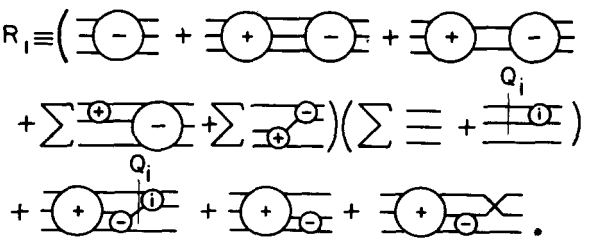
(5.29)

and using (5.21), we obtain



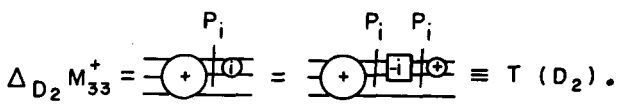
(5.30)

where



(5.31)

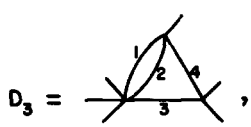
Again setting M_i equal to $\mu_1 + \mu_2$, we see[25] that the only terms of (5.30) that can support D_2 with all α's positive are the first and the second. Thus we obtain[26]



(5.32)

C. Discontinuity Equation for the Ice-Cream-Cone Diagram

To find the discontinuity of M_{33}^+ across $\mathcal{M}^+[D_3]$ where

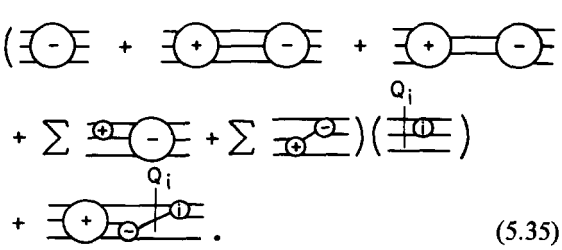


(5.33)

we first subtract (4.1) from (5.30), and obtain

(5.34)

where

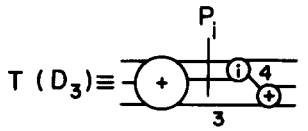


(5.35)

[25] Details are given in Appendix E.

[26] This is essentially the result of Ref. 7.

The diagram R_2 cannot support D_2 with all α's positive.[25] Substituting (5.34) into the second term of (4.6) we see[25] that the only terms of (4.6) that can support D_3 with all α's positive are M_{33}^{+} and $-T(D_3)$, where



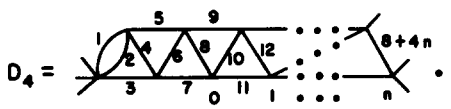
(5.36)

is a threshold term. Thus we obtain[27]

$$\Delta_{D_3} M_{33}^{+} = T(D_3). \tag{5.37}$$

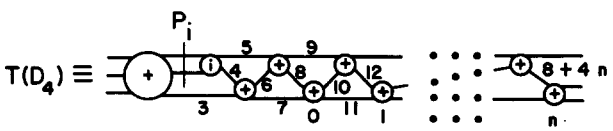
D. Discontinuity Equation for the Extended Ice-Cream-Cone Diagram

Consider the Landau diagram



(5.38)

Substituting (5.34) into the second term of (4.11) we see[25] that the only terms of (4.11) that can support D_4 with all α's positive are M_{33}^{+} and $-T(D_4)$, where

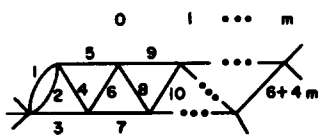


(5.39)

is a threshold term. It therefore follows that

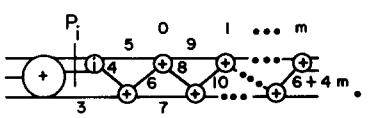
$$\Delta_{D_4} M_{33}^{+} = T(D_4). \tag{5.40}$$

Similarly, by combining (4.8) and (4.6), or (4.8) and (4.11), and then substituting (5.34), we find that the discontinuity of M_{33}^{+} corresponding to the diagram



(5.41)

is given by

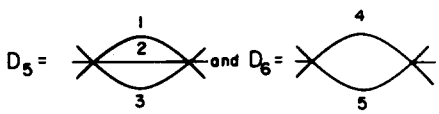


(5.42)

Any set of simple inner vertices (defined in the caption of Fig. 1) of any of the diagrams (5.38) or (5.41) can be replaced by a corresponding set of two-particle closed loops. Then the corresponding discontinuity of M_{33}^{+} is given by (5.39) or (5.42) with every plus bubble corresponding to a two-particle closed loop replaced by the expression (5.26).

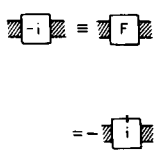
E. Discontinuity Equation for Normal Singularities in the Total Energy

The problem is to find the discontinuity of M_{33}^{+} around $\mathcal{M}^{+}[D_5]$ and $\mathcal{M}^{+}[D_6]$, where



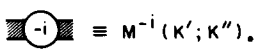
(5.43)

The $-i$ (or F) box was defined in (5.18). By virtue of (2.9) this definition is equivalent to



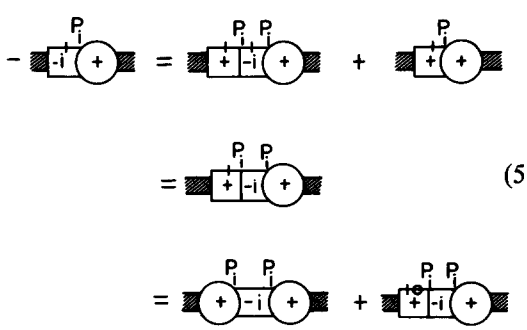
(5.44)

The connected part of the left side of (5.44) is denoted by



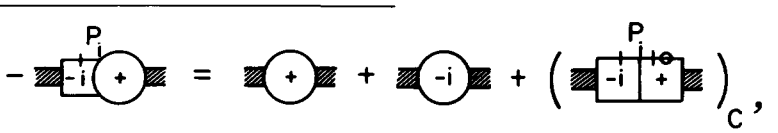
(5.45)

It then follows from (5.18) and (2.11) that



(5.46)

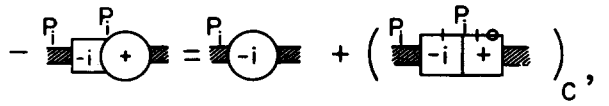
Using (5.20) we can write the connected part of (5.18) in the form



(5.47)

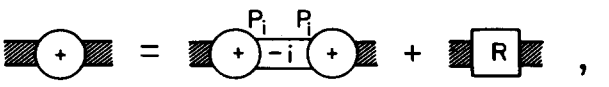
[27] This is essentially the result of Ref. 8.

which gives



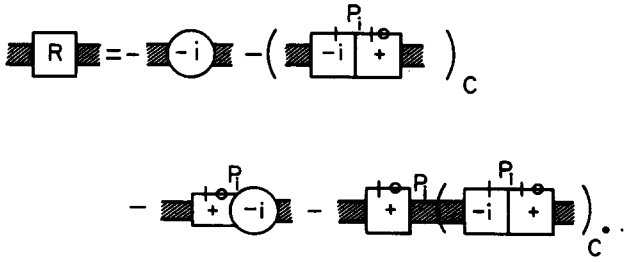
(5.48)

where the subscript C indicates the connected part. Substituting (5.47) and (5.48) into (5.46) we obtain



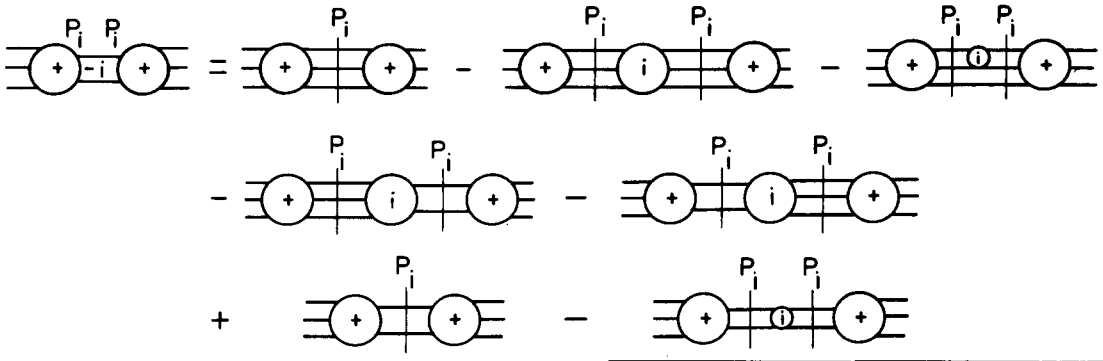
(5.49)

where



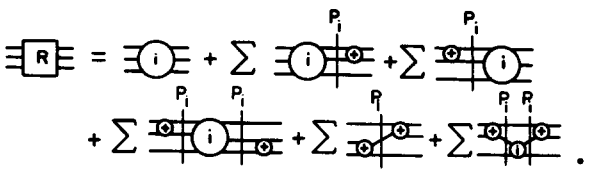
(5.50)

For the special case of three ingoing and three outgoing particles, the terms on the right-hand side of (5.49) are given by



(5.51)

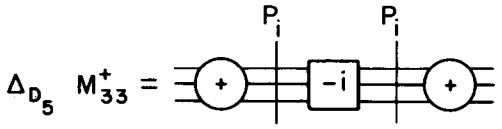
and



(5.52)

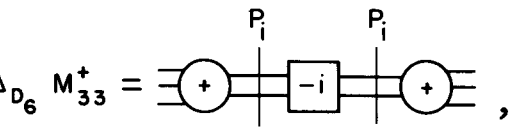
Equation (5.22) is used to obtain the fifth term of (5.52).

The only terms in (5.49) that can support D_5 or D_6 with all α's positive are M_{33}^{+} and the first term on the right-hand side.[25] Thus, we obtain



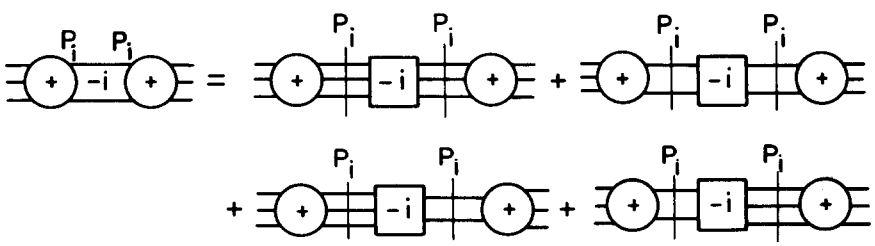
(5.53)

and



(5.54)

where we set M_i equal to $\mu_1 + \mu_2 + \mu_3$ in (5.53) and equal to $\mu_4 + \mu_5$ in (5.54) and we assume that $\mu_1 + \mu_2 + \mu_3 \neq \mu_4 + \mu_5$. For $\mu_1 + \mu_2 + \mu_3 = \mu_4 + \mu_5$, the discontinuity across $\mathcal{M}^{+}[D_5] = \mathcal{M}^{+}[D_6]$ is given by



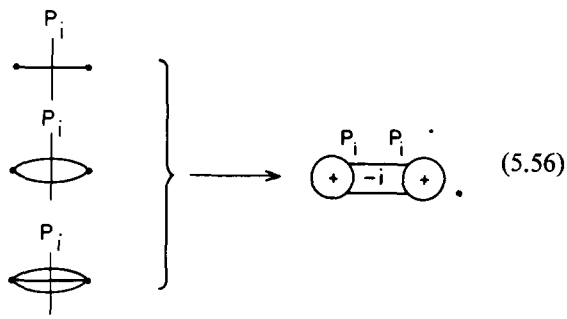
(5.55)

F. Discontinuity Equation for the $3 \to 2$ and $2 \to 3$ Amplitudes

The $3 \to 2$ and $2 \to 3$ amplitudes have no physical-region singularities depending on a cross-energy invariant. The discontinuities of these amplitudes across their singularities are given by formulas completely analogous to those obtained for the $3 \to 3$ amplitude except for the change in the external lines.

G. Equivalent Diagrams

The Landau surfaces corresponding to equivalent (but not identical) diagrams coincide. The discontinuity formulas given by the rules (1.2) of the Introduction have partially taken this into account, since a summation over all particles of the same mass M_r is implied for a line labeled with the number r. However, the sum of the rest masses $M_1 + M_2$ of the particles in a two-particle closed loop may be equal to the sum of the rest masses $M_1' + M_2' + M_3'$ of the particles in a three-particle closed loop; or it may be equal to the sum of the rest masses $M_1'' + M_2''$ with $M_1'' \neq M_1$ of another set of particles corresponding to the two-particle loop. The rules (1.2) do not cover such cases. In the derivations given in this section, the possibility of equivalent diagrams is not excluded, and the results cover these cases also. They give, instead of (1.2), the more general rule



(5.56)

These rules are equivalent to those of (1.2), if there are no equivalent diagrams of the type just mentioned and provided the M_i corresponding to the bar in (5.56) is set equal to the appropriate sum of the masses occurring in (1.2).

It is understood here, as it is in the discontinuity formulas derived in this section, that the bubble-diagram functions on the right-hand side are evaluated just above their threshold. We shall see in the next section that if this restriction is relaxed, then the rule (5.56) can, in some cases, give the discontinuity across a whole set of cuts.

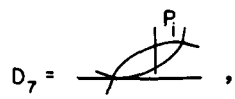
H. Surfaces of Lower Dimension

The third structure theorem applies only to simple points. Thus it would not apply, for example, to D_1 of (5.4) if the masses satisfy $\mu_1 + \mu_2 = \mu_3 + \mu_6$, where 6 is the upper left-hand line; for then $\mathcal{M}^+(D_1)$ would lie in $\mathcal{M}^+(D_3)$ [see (5.33)]. Thus it would contain no simple points and would in fact be of lower dimension. It is the discontinuity around the imbedding codimension-one surface $\mathcal{M}^+(D_3)$ that would be calculated. Other threshold singularities for internal bubbles are treated analogously.

6. DISCONTINUITY OF THE SCATTERING AMPLITUDE AROUND SEVERAL CUTS

The method used in Sec. 5 to derive the discontinuity of M_{33}^+ across a single Landau surface can be generalized to give the discontinuity across several Landau surfaces.[28]

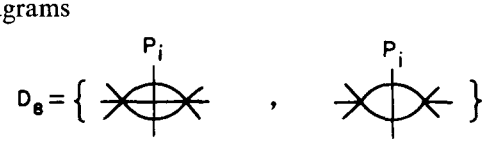
Consider the set of Landau diagrams



(6.1)

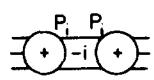
where the unlabeled internal lines correspond to all possible types of particles that are compatible with the conservation and mass constraints and which satisfy the requirement that the sum of the masses of the particles corresponding to the lines cut by the P_i bar be greater than a given mass M_i. The only terms of (5.30) that can support a Landau diagram of the set D_7 with all α_i positive are the first and the second.[25] On the unphysical side of all the Landau surfaces $\mathcal{M}^+[D_7]$, Eq. (5.30) is $M_{33}^+ + R_1 = 0$. If this equation is continued to a point that lies on the physical side of all the Landau surfaces $\mathcal{M}^+[D_7]$ in such a way that R_1 is continued into itself, then M_{33}^+ is continued around $\mathcal{M}^+[D_7]$ according to a minus-$i\epsilon$ rule and into a function denoted by $M_{33}^+(D_7^-)$. Subtracting the continued equation from (5.30), we find that the discontinuity $M_{33}^+ - M_{33}^+(D_7^-)$ is again given by the right-hand side of (5.32).

A similar argument, based on the formulas of subsection 5E shows that the discontinuity of M_{33}^+ across the set of Landau surfaces corresponding to the diagrams



(6.2)

is given by

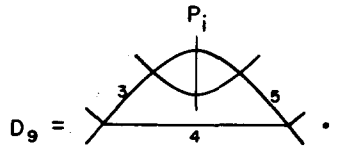


(6.3)

The same procedure also works for nonnormal singularities. Consider, for instance, the set of Landau

[28] This topic will be discussed in more detail in later work.

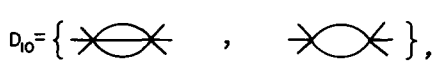
diagrams



(6.4)

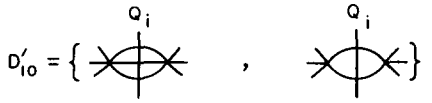
The discontinuity $M_{33}^{+} - M_{33}^{+}(D_9^{-})$ across the set of all cuts that begin at the Landau surfaces $\mathcal{M}^{+}[D_9]$ is given by the function $T(D_1)$ defined in (5.25), provided we remain in the region where $M_{33}^{+}(D_9^{-})$ has no mixed-α singularities corresponding to D_9 or to the contractions of D_9. This restriction is imposed because the third structure theorem cannot be used to continue the right-hand side of $M_{33}^{+}(D_9^{-}) = M_{33}^{+} - T(D_1)$ into itself around such mixed-α singularities.

Finally, let us examine the total discontinuity of M_{33}^{+} across all normal thresholds in the total energy or in a subenergy. Consider the set of diagrams



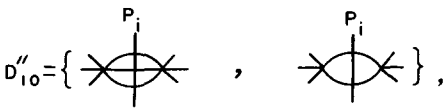
(6.5)

where the unlabeled internal lines correspond to all possible types of particles that are compatible with the conservation and mass constraints. This set is composed of the subsets



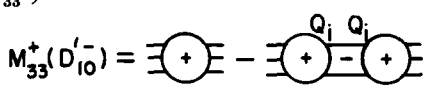
(6.6)

and



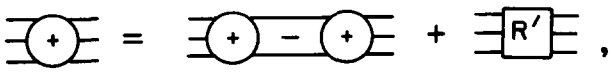
(6.7)

where we choose M_i in such a way that $E = M_i$ is the energy corresponding to the lowest normal threshold lying within the physical region of M_{33}^{+}. (Thus, the normal thresholds corresponding to the diagrams D'_{10} lie outside or on the boundary of the physical region of M_{33}^{+}.) We define



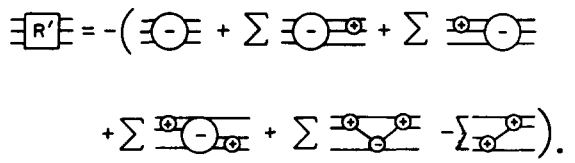
(6.8)

The argument which led from (5.46) to (5.49) also works if we replace $-i$ by a minus sign and omit the P_i bar in these formulas. Then, Eq. (5.49) becomes



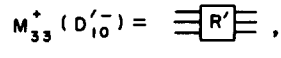
(6.9)

where



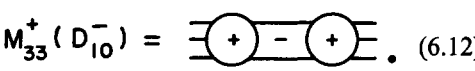
(6.10)

On the unphysical side of the set of singularities corresponding to D''_{10} we see that



(6.11)

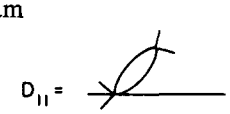
where we have used (6.9). If this equation is continued to a point which lies on the physical side of the surfaces $\mathcal{M}^{+}[D_{10}]$ in such a way that the right-hand side of the equation is continued into itself, then $M_{33}^{+}(D_{10}^{\prime -})$ is continued below the cuts corresponding to $\mathcal{M}^{+}[D''_{10}]$ and into a function denoted by $M_{33}^{+}(D_{10}^{-})$.[25] Subtracting the continued equation from (6.9) we obtain



(6.12)

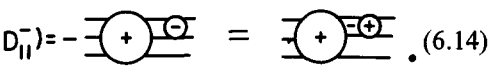
Equation (6.12) gives discontinuity of M_{33}^{+} across all the Landau surfaces $\mathcal{M}^{+}[D_{10}]$ provided $M_{33}^{+}(D_{10}^{\prime -})$ can be regarded as a continuation of M_{33}^{+} to some unphysical sheet.

With a similar proviso, we find that the total discontinuity across the singularities corresponding to the diagram



(6.13)

is given by



(6.14)

Equations (6.12) and (6.14) can be regarded as just definitions of $M_{33}^{+}(D_{10}^{-})$ and $M_{33}^{+}(D_{11}^{-})$, respectively. The nontrivial aspect is the result that these functions have minus-$i\epsilon$ continuations around the normal threshold $\mathcal{M}^{+}[D''_{10}]$ and $\mathcal{M}^{+}[D''_{11}] \equiv \mathcal{M}^{+}[D_7]$, respectively.

APPENDIX A: COMBINATORICS

A. Cluster Decomposition of the S Matrix

For bosons, the cluster-decomposition property of the S matrix is expressed by

$$M(K'; K'') = \sum_{p} M_p(K'; K'') \tag{A1a}$$

and

$$M_p(K'; K'') = \prod_{s=1}^{N_p} M_1(K'_{ps}; K''_{ps}). \tag{A1b}$$

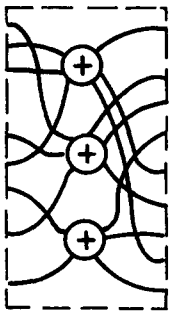


FIG. 5. Diagram representing a typical term M_p of the M function in (A1a).

The sum over p is a sum over all partitions of the set of variables $(K'; K'')$ into disjoint subsets. The set of variables $(K'_{ps}; K''_{ps})$ is the sth subset of the pth partition. The pth partition has altogether N_p subsets, and the first partition, $p = 1$, is the unique partition with $N_p = 1$ and $K'_{11} = K'$, $K''_{11} = K''$.

The cluster decomposition is graphically represented in terms of bubble diagrams. A bubble with a plus sign inside represents the connected part, $M_1(K'_{ps}; K''_{ps})$, of the S matrix $M(K'_{ps}; K''_{ps})$. Then $M(K'; K'')$ is represented by a sum of terms each of which is a column of plus bubbles. (The set of bubbles includes trivial bubbles, which are bubbles connected to just one initial and one final line.) Counting is important. There is precisely one term for each topologically different way of connecting a column of plus bubbles to the given set of external lines specified by $(K'; K'')$. The topological structure is determined completely by specifying the grouping of the external lines into subsets. The lines of each individual subset $(K'_{ps}; K''_{ps})$ of external lines are drawn as emerging from a single bubble. Two diagrams that differ only in the ordering of the bubbles in the column are not topologically different. Similarly, two diagrams that differ only in the ordering of the lines emerging from any given bubble are not topologically different. This latter fact allows us to always draw the diagrams so that the lines emerging from any given bubble never cross. However, lines emerging from different bubbles may cross. A typical term thus has a structure like that shown in Fig. 5.

B. Counting the Intermediate States

The unitarity equation is

$$\sum_K M(K'; K)M^\dagger(K; K'') = \delta(K'; K''). \qquad \text{(A2)}$$

Here $K = (p_1, t_1, p_2, t_2, \cdots, p_n, t_n)$ is a set of variables $V_j \equiv (p_j, t_j)$. The sum over K includes a sum over all n. For each n, each of the n indices t_j is summed over all possible values, and for each value of t_j there is an integration over all values of p_j satisfying $p_j^2 = \mu_j^2 \equiv \mu^2(t_j)$. The states are labeled by *unordered* sets K. That is, states labeled by sets K that differ only in the order of the variables of K are not counted as different. Thus, one must either restrict the range of integration by some normal ordering convention, as in Ref. 15, or divide by $n!$. Let us temporarily adopt the latter method, so that there is no restriction on the range of integration. Then the summation on the left-hand side (A2) can be written in the explicit form

$$\begin{aligned}\sum_K = \sum_{j=1}^{n} \frac{1}{n!} \sum_{t_1=1}^{\tau} \sum_{t_2=1}^{\tau} \cdots \sum_{t_n=1}^{\tau} \\ \times \int \frac{d^4p_1}{(2\pi)^4} 2\pi\theta(p_1^0)\delta(p_1^2 - \mu_1^2) \\ \times \frac{d^4p_2}{(2\pi)^4} 2\pi\theta(p_2^0)\delta(p_2^2 - \mu_2^2) \cdots \\ \times \frac{d^4p_n}{(2\pi)^4} 2\pi\theta(p_n^0)\delta(p_n^2 - \mu_n^2). \qquad \text{(A3)}\end{aligned}$$

Here n is the number of lines in the intermediate state, and τ is the number of types of particles. The momentum p_j is associated with line j and also with the type variable t_j and can therefore be written as $p_j(t_j)$.

When we transcribe unitarity (A2) into bubble notation, we find that topologically indistinguishable diagrams occur. That is, even though the individual M functions are expressed as a sum of topologically different diagrams, the topological product of these diagrams contains diagrams that are not topologically different. The topological structure of a contribution to the product is specified by specifying first which subsets of the set of outgoing variables K' are grouped together (i.e., are attached to a common bubble) and which subsets of the set of incoming variables K'' are grouped together. (The various variables of K' and K'' are always considered as distinct and identifiable. One can, for instance, take all the p_j in K' and K'' to have different fixed values.) The various groups of incoming and outgoing variables can be labeled by indices i and f, respectively. These indices i and f then label the bubbles of the right and left columns of the product. It is important to note that this labeling does not refer to the position of the bubbles in the column but rather to the sets of external lines connected to these bubbles.

The number of lines connecting bubble i to bubble f is called N_{fi}. The topological structure is specified by these numbers N_{fi}, together with the specifications of the subsets of incoming and outgoing lines labeled by i and f.

Bubble diagrams of the same topological structure give exactly the same bubble-diagram functions. Thus, the product on the left of (A2) can be expressed in the

form

$$F(K'; K'') = \sum_B c_B \bar{M}^B(K'; K''), \qquad \text{(A4)}$$

where the sum is over all the topologically different bubble diagrams B contained in the topological product of the two boxes. The coefficient c_B for a diagram with n internal lines is $N_B/n!$, where N_B is the number of diagrams topologically equivalent to B in the topological product of the two boxes, and the factor $1/n!$ comes from (A3). The bar on $\bar{M}^B$ indicates that, contrary to the convention adopted in the main text (see below), the regions of integration are not restricted by any ordering convention, but are as given in (A3), without the $1/n!$. We show immediately that

$$N_B = n!/\prod (N_{fi})!. \qquad \text{(A5)}$$

This result gives

$$c_B = 1/\prod (N_{fi})!, \qquad \text{(A6)}$$

the product being over all pairs (f, i). As usual, one takes $0! = 1$.

To derive (A5) one first labels the intermediate lines in accordance with their topological character: Each line is labeled by a unique triple (f, i, m), where f and i label the final and initial bubbles that the line joins, and for any particular values of f and i the m in (f, i, m) is an index that runs from 1 to N_{fi} and specifies the particular one of these N_{fi} lines. There is also the index j that runs from 1 to n, and identifies the n variables of $K = (p_1, t_1, p_2, t_2, \cdots, p_n, t_n)$.

For definiteness, one may specify that the ordering of the lines of any box reading from top to bottom is the same as the ordering of the associated variables of the corresponding set K. Thus, j specifies the geometric location of the intermediate line L_j, reading from top to bottom of the box. The index m of (f, i, m) may also be considered to specify the position, reading from top to bottom, of line (f, i, m) relative to the other lines of the set of lines Γ_{fi} that joins bubbles i and f. The condition imposed earlier that lines attached to a given bubble do not cross within the box insures that the ordering of the lines of Γ_{fi} is well defined; the relative ordering within one box of any set of lines of Γ_{fi} is the same as the ordering of this set of lines in the other box. This condition that the ordering of the lines Γ_{fi} be given by m is, however, the only restriction on the ordering of the intermediate lines; one readily confirms that the various intermediate lines, as identified by their topological indices (f, i, m), can occur in any possible order (reading from top to bottom), subject only to this condition that the relative ordering of lines in the various sets Γ_{fi} be in accordance with the index m. The term coming from each of these allowed orderings is a different contribution to the product (A2). Thus, in this product, the number of different contributions that are topologically equivalent to a diagram B is just the number of different allowed orderings of these intermediate lines. This is just the total number of orderings $n!$ divided by the product of the number of orderings within each set Γ_{fi}. Thus we obtain (A5).

In the text it was specified that the region of integration in the definition of bubble-diagram functions M^B be restricted so that contributions from topologically equivalent diagrams are counted only once. In the derivation of (A4) no such restriction on the range of integration was imposed, and the corresponding functions were written as $\bar{M}^B$. These two functions are related by the factor c_B. The point here is that the various lines of a set of lines connecting a given pair of bubbles are regarded as topologically equivalent. Thus, in computing M^B, the integration region is restricted so as to include only one of the set of contributions obtained by interchanging the lines of such a set. This restriction on the domain of integration in the definition of the functions M^B means that the $\bar{M}^B$ in (A4) divided by $\prod N_{fi}!$ is just M^B. Thus, in place of (A4) one obtains

$$F(K'; K'') = \sum_B M^B(K'; K''). \qquad \text{(A7)}$$

The notion of topological distinctness has been applied on two different levels: When in (A4) or (A7) we say the sum over diagrams B is over topologically different diagrams B, we are considering B to be simply a collection of lines and bubbles joined to give a geometric figure; the lines are not yet assigned particular variables. But when for a fixed B we say that the integration region defining M^B is restricted so that topologically equivalent diagrams are counted only once, then we are considering variables (p_j, t_j) to be assigned to the lines. This separation into two levels is evidently arbitrary.

The proof given above can be shortened and extended to products of arbitrary numbers of boxes by arguing as follows. In the integration corresponding to the sum over the set of intermediate states, one is supposed to count only one of the set of possible contributions obtained from the various possible reorderings of the variables. A reordering of variables corresponds to a reordering of the lines associated with the intermediate particles. Thus, the ordering of the intermediate lines can be considered to be completely irrelevant; the intermediate lines can be identified by the value of the associated variables alone. For every way of connecting the various bubbles of the various adjacent columns by lines, and assigning a fixed variable to each line, there is at least one

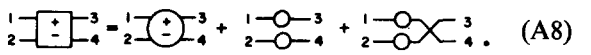
contribution to the over-all function. We now define diagrams with fixed variables attached to each line to be topologically equivalent if and only if they can be made identical by some reordering of the various bubbles within the various columns. It then follows that the contributions from two topologically equivalent diagrams should not both be counted. For they must both arise from contributions to the individual boxes that are topologically equivalent, and hence identical. On the other hand, no two contributions that are not topologically equivalent in this sense can come from a single set of contributions from the various boxes. Hence the restriction to topologically different diagrams leaves one with precisely one complete set of independent contributions.

The consequence of this argument is that the function corresponding to a product of any number of functions S and $S^\dagger$, with the intermediate sums defined as in the unitarity equations, is represented by the function M^B, where B is the natural topological product of the boxes representing the individual functions S and $S^\dagger$. One can decompose the various boxes into sums of terms represented by different columns of bubbles. The natural topological product does not include diagrams that are topologically equivalent in the sense that they differ only in the ordering of bubbles within a column or by the path followed by intermediate lines. (Only the end points of the lines are significant.) For each diagram B of the natural topological product there is one term M^B. In evaluating this term the integration is restricted so that topologically equivalent contributions are counted precisely once, where now each line is identified by a variable $(t_j p_j)$.

C. Example

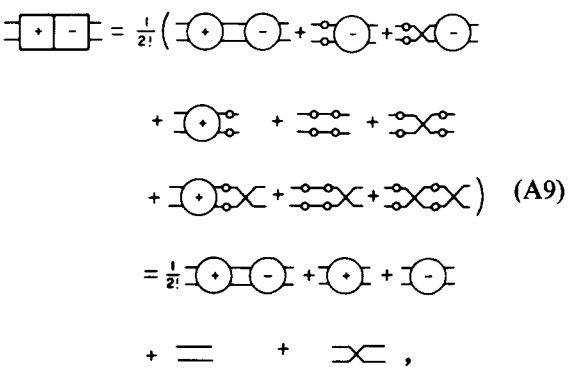
In the main text the combinatoric questions are automatically taken care of by the use of functions M^B; the restriction on the ranges of integrations of these functions makes everything correct. To exhibit the combinatoric questions resolved by this notation, and to confirm our basic formulas, we rederive the formulas of Sec. 4 starting directly from Eqs. (A1)–(A3), and using instead of the functions M^B rather the functions $\bar{M}^B$, which have no restriction on the domain of integration.

First consider two-particle unitarity. The two-particle box is given by



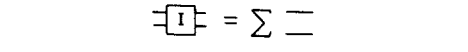
(A8)

In (A8) the incoming and outgoing lines are identified redundantly both by an integer and also by the vertical position of the external end points. In the remainder of Appendix A we suppress the integer and use only the latter method of labeling. Below the three-particle threshold, we obtain



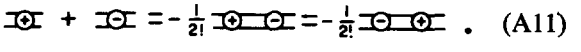
(A9)

where the factor 1/2! comes from (A3). (This factor 1/2! does not appear in the equations of the text because there the diagrams represent always the function M^B whereas in this section they represent the functions $\bar{M}^B$.) Equation (A9) is evidently in agreement with (A4). The last two terms on the right-hand side of (A9) are equal to the identity



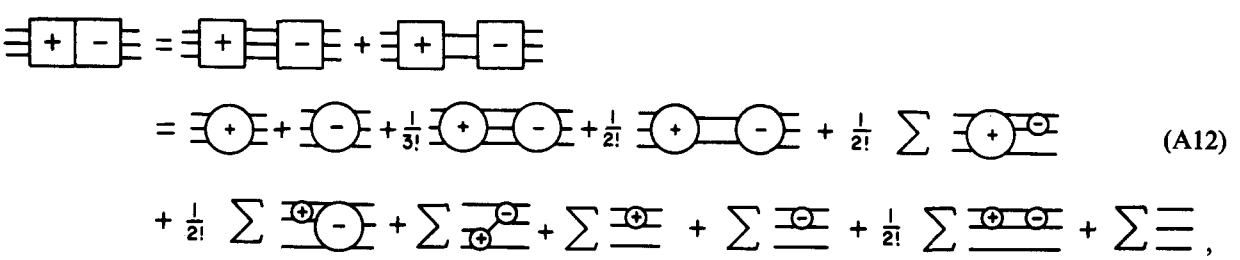
(A10)

so that the connected part of the unitarity equation is



(A11)

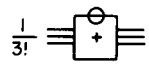
By a completely analogous procedure we obtain, after combining various terms,



(A12)

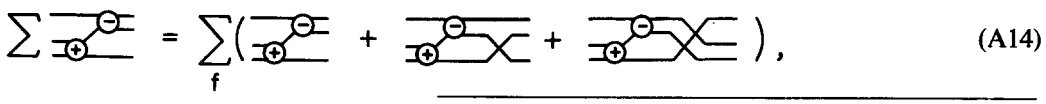
where the summation signs are interpreted as in Sec. 4. The disconnected parts of (A12) are equal to the identity by virtue of (A11), so that the connected part of (A12) vanishes, by unitary.

To obtain (4.6) when there is no restriction on the range of integration, we postmultiply the connected part of (A12) by



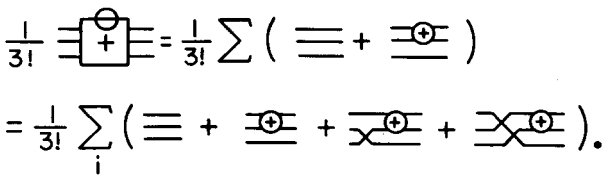
(A13)

Consider, for instance, the postmultiplication of the nine dumbbell terms of (A12),



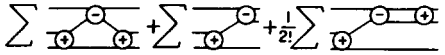
(A14)

by the 6 + 9 terms of



(A15)

The result of the multiplication is

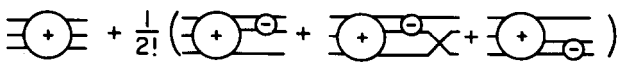


(A16)

The last two terms combine to give

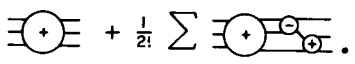
(A17)

Similarly, the postmultiplication of the terms



(A18)

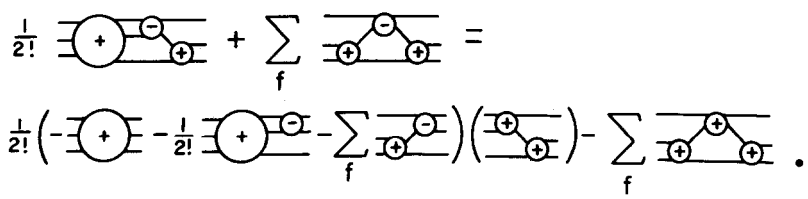
of (A12) by (A13) gives



(A19)

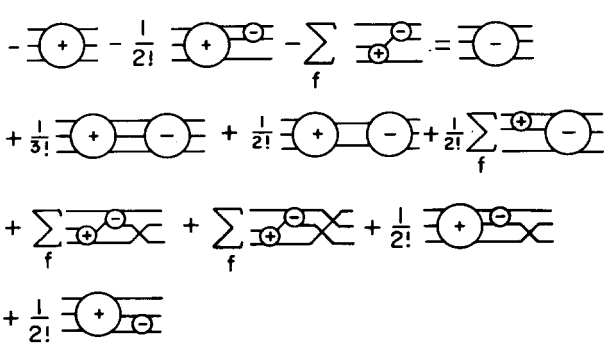
These results check with (4.6) if we take account of the factors $N_{fi}!$ that relate $\bar{M}^B$ to M^B.

Using (A11), we obtain



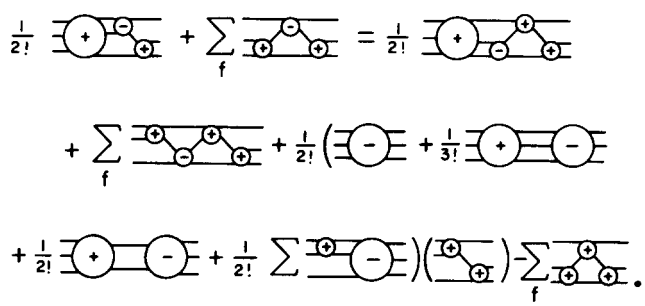
(A20)

Substituting the unitarity equation



(A21)

into (A20), we obtain



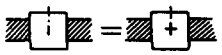
(A22)

Equations (A20) and (A22) check with Eqs. (4.7) and (4.8) of the text if the factors $N_{fi}!$ relating $\bar{M}^B$ to M^B are considered.

The remaining equations of Sec. 4 now follow from the equations already derived and pose no combinatorial problems.

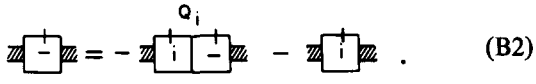
APPENDIX B: DETERMINATION OF $M^i(K'; K'')$ USING FREDHOLM THEORY[29]

The $M^i(K'; K'')$ are defined by (5.8). Let us first assume that $E < M_i$. Then a comparison of (5.8) and (5.6) shows that



(B1)

and that (5.19) must hold. Thus, (5.8) can be written in the form



(B2)

Let us now postmultiply (B2) by

(B3)

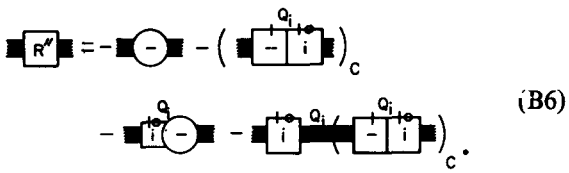
and go through exactly the steps that led from (5.46) to (5.49), but making the replacements

$$+ \to i, \quad -i \to -, \quad P_i \to Q_i. \tag{B4}$$

Then in place of (5.49) we obtain

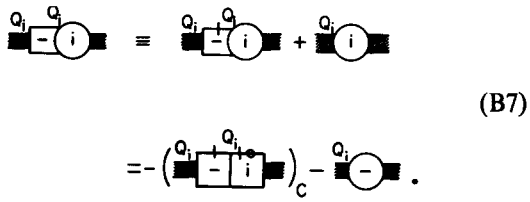
(B5)

where



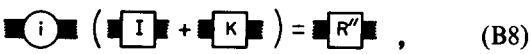
(B6)

The connected part of (5.8), premultiplied by Q_i, can be written in the form



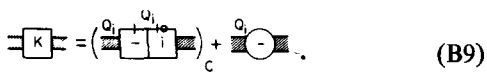
(B7)

Substituting (B7) into (B5) we obtain



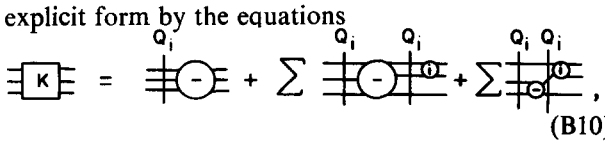
(B8)

where

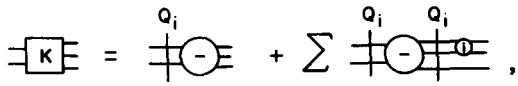


(B9)

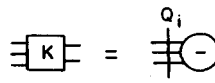
The kernel of the integral equation (B8) is given in explicit form by the equations



(B10)

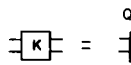


(B11)



(B12)

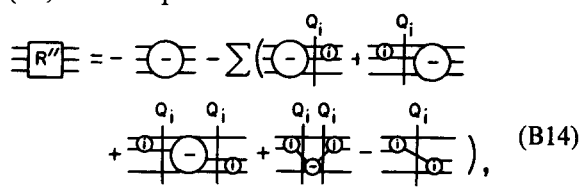
and



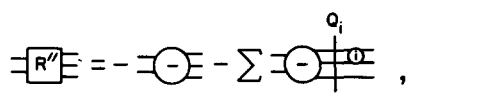
(B13)

The delta function appearing in (B10) combines with the remaining terms of this equation to give a pole with a plus-$i\epsilon$ rule. This follows from an argument similar to that used by Olive, except that one uses (5.21) rather than (4.5) to combine the residues of the poles.[30] The contour of integration can be distorted away from the remaining singularities of the kernel.[31] Thus Eq. (B8) can be solved for $M^i(K'; K'')$ through the Fredholm formula.

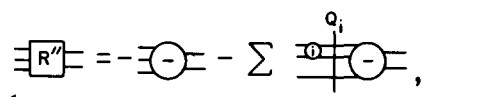
Using Eq. (5.21) we can express the right-hand side of (B8) in the explicit form



(B14)

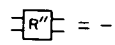


(B15)



(B16)

and



(B17)

We suppose that (5.21) has been solved already for M^i_{22} and that this solution has been substituted in (B14), (B15), and (B16), as well as in (B10) and (B11). It is then seen that the Fredholm solution of (B8) expresses M^i_{33}, M^i_{23}, and M^i_{32} in terms of bubble-diagram functions all of which follow a minus-$i\epsilon$ prescription at the normal two- or three-particle threshold at $E = M_i$. It then follows[31] that the solution $M^i(K'; K'')$ of (B8), and moreover Eq. (5.8)

[29] The application of Fredholm theory to the problem of continuation of many-particle scattering amplitudes to unphysical sheets has been discussed earlier by H. P. Stapp in Lawrence Radiation Laboratory Report UCRL-10261, 1962, and Nuovo Cimento **32**, 103, 1964; J. Gunson, Ref. 2; D. I. Olive, Nuovo Cimento **28**, 1318 (1962).

[30] D. I. Olive, Phys. Rev. **135**, B745 (1964); see also R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, Cambridge, England, 1966), p. 212.

[31] The argument is essentially the same as in the proof of the third structure theorem of Ref. 13.

itself, can be analytically continued from $E < M_i$ to $E > M_i$ by following a minus-$i\epsilon$ rule.

The original restriction to $E < M_i$, which we used to justify (B2), entails that the set of incoming particles and the set of outgoing particles each have a sum of rest masses less than M_i. Thus these sets are, in effect, cut by a Q_i bar. Hence the Fredholm solution of (B8) is an explicit expression for

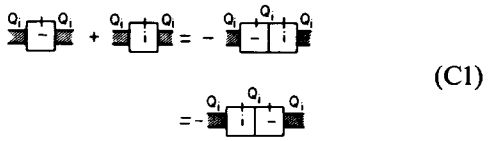
(B18)

for $E < M_i$ that has a minus-$i\epsilon$ rule for continuation past the normal threshold at $E = M_i$. In Appendix C we will enlarge upon this result and show that Eq. (B8) also determines $M^i(K'; K'')$. (That is, the Q_i bars can be omitted.)

The fundamental result established above is that the physical scattering function has a continuation past the normal-threshold singularity in the minus-$i\epsilon$ sense, apart from possible poles coming from the vanishing of the Fredholm denominator. Discounting the possibility that these poles become dense,[32] we obtain, using the same arguments that led to the third structure theorem, the result that all terms in (5.8) can be continued in the minus-$i\epsilon$ sense around the normal threshold at $E = M_i$. Thus, this equation can be regarded as valid on both sides of $E = M_i$, with $M^i(K'; K'')$ a function that has a minus-$i\epsilon$ continuation around the singularity at $E = M_i$.

APPENDIX C: PROOF OF EQUATION (5.19)

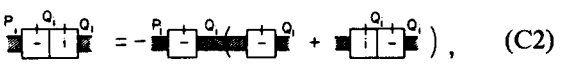
In Appendix B we showed that the equation



(C1)

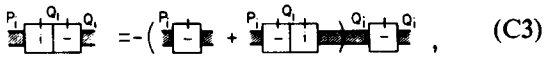
valid for $E < M_i$ can be analytically continued to the region $E > M_i$. We now show that Eq. (5.19) is a consequence of (C1) and the definition (5.8).

It follows from (C1) and (5.8) that



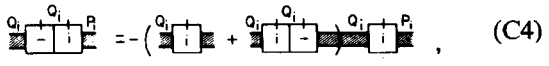
(C2)

and



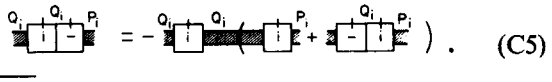
(C3)

and also that



(C4)

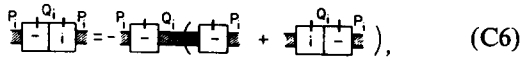
and



(C5)

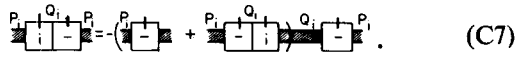
[32] A. Martin, CERN preprint.

Comparison of the right-hand sides shows that the left-hand side of (C2) is equal to the left-hand side of (C3), and that the left-hand side of (C4) is equal to the left-hand side of (C5). Using this result and Eq. (5.8), we obtain



(C6)

and



(C7)

Since the right-hand sides are equal, so are the left-hand sides, and hence the proof of (5.19) is complete.

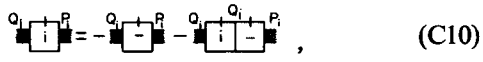
Since Eq. (B8) was based on (5.8), (5.19), and nothing else, we see that (B8) can be used to determine

(C8)

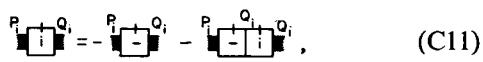
as well as

(C9)

Alternatively, we can determine the quantities (C8) in terms of (C9) by directly using (5.8) and (5.19). Thus, we may write

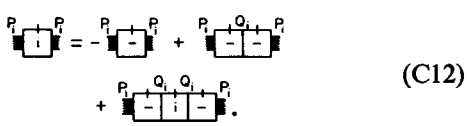


(C10)



(C11)

and



(C12)

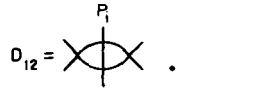
APPENDIX D: THE F_{nm} AS SHEET CONVERTERS

The right-hand sides of the discontinuity equations given in the previous sections are expressed in terms of plus bubbles and F boxes. These boxes are defined in terms of physical scattering amplitudes by Fredholm integral equations. It is shown in this appendix that the effect of applying a (nontrivial) F box to the physical scattering amplitude is to convert the latter to its value on the unphysical side of a certain cut. This result allows one to express the discontinuity formulas in all cases considered in this paper in terms of the scattering amplitudes evaluated on various sheets, instead of in terms of physical amplitudes and F boxes. The proof depends, however, on the so-called extended unitarity equations.

In the main text we have been careful to use only physical unitarity equations. In particular, the various momentum vectors are always real, apart from infinitesimal variations needed in the continuation around

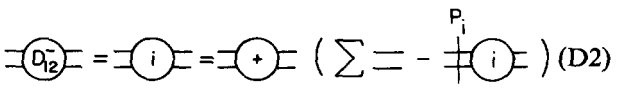
Landau singularities. This restriction means that the unitarity equations below a certain threshold in the total energy at $E = M_i$ hold only if the sums of the rest masses of the initial and final particles are both less than M_i. There are similar restrictions on the masses of subsets of initial and final particles associated with subenergies. The equations obtained if one relaxes these conditions are called extended unitarity equations. Their justification within the S-matrix framework is discussed in Refs. 13, 30, and 33. In this appendix these extended unitarity equations are assumed without further comment.

Consider first the set of Landau diagrams

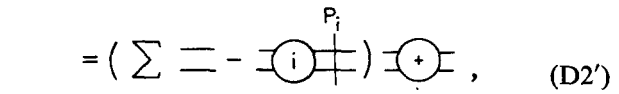


(D1)

From (5.22) and the fact that M^i_{22} has a minus-$i\epsilon$ continuation past $\mathcal{M}^+[D_{12}]$, we obtain

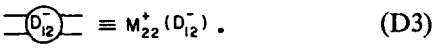


(D2)



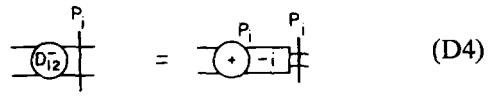
(D2′)

where



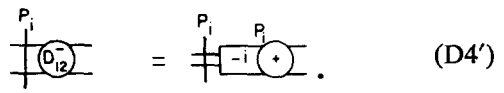
(D3)

This derivation depends on the assumption that the region $E < M_i$ contains physical points. That is, within the framework of the physical unitarity equations, the derivation of (D2) and (D2′) is only justified if the set of incoming particles and the set of outgoing particles each has a sum of rest masses smaller than M_i. This restriction can be represented by placing Q_i bars on the external lines. By virtue of the extended unitarity assumption, this restriction can be dropped, and one obtains as special cases of (D2) and (D2′) the results



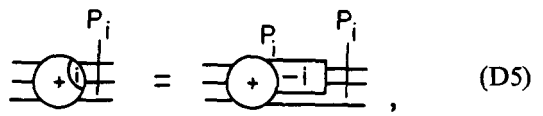
(D4)

and



(D4′)

Similarly, using the results of Sec. 4, we find that

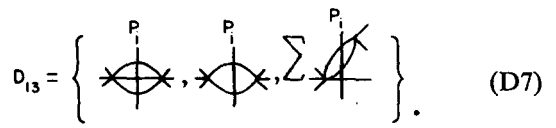


(D5)

where

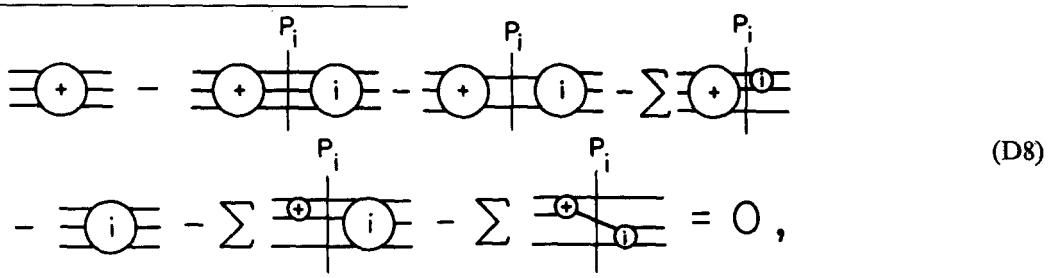
(D6)

Next consider the set of Landau diagrams



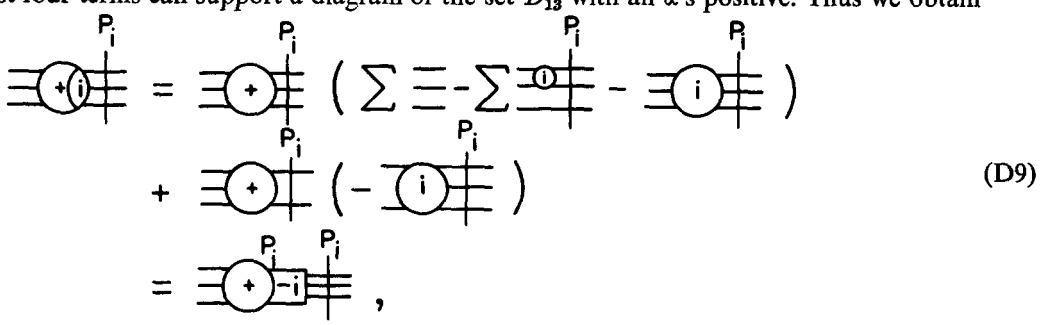
(D7)

The discontinuity around the set of Landau surfaces $\mathcal{M}^+[D_{13}]$ is derived, according to the method of Sec. 6, by noting that the connected part of (5.18) can be written in the form



(D8)

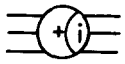
where only the first four terms can support a diagram of the set D_{13} with all α's positive. Thus we obtain



(D9)

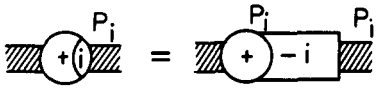
[38] J. B. Boyling, Nuovo Cimento 33, 1356 (1964).

where



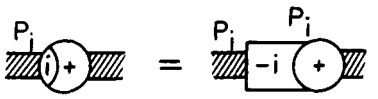
$$\equiv M^{+}_{33}(D^{-}_{13}). \qquad \text{(D10)}$$

Using similar methods we can establish the more general results



(D11)

and



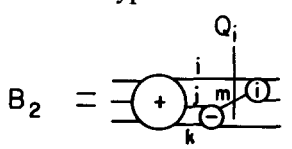
(D12)

Formulas (D4), (D4′), (D5), (D11), and (D12) can be substituted on the right-hand side of the various discontinuity equations derived earlier. The F boxes are thereby eliminated, but the scattering functions are evaluated on unphysical sheets.

APPENDIX E: PROOF OF THE ABSENCE OF CERTAIN POSITIVE-α LANDAU SINGULARITIES IN CERTAIN BUBBLE-DIAGRAM FUNCTIONS

In this appendix we prove that certain sets of bubble diagrams occurring in the equations of the main text and in Appendix D cannot support certain Landau diagrams with all α's positive, provided the signs of the α's are restricted in accordance with the second structure theorem.

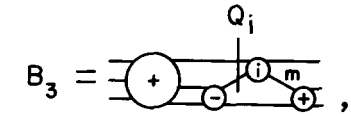
Consider first the Landau diagram D_2 defined in (5.27) and the set of bubble diagrams R_1 defined in (5.31). Let each M^i_{22} occurring in R_1 be replaced by the right-hand side of (5.13). Since we are interested only in Landau diagrams with all α's positive, all minus bubbles in R_1 can be replaced by point vertices. Then the function M^i_{22} can also be replaced by a point vertex since none of the internal lines shown in the right-hand side of (5.13) can be a line of D_2. It is seen by inspection that no term in R_1, except possibly contributions of the type



(E1)

can support D_2 with all α's positive. [The letters labeling the internal lines in (E1) stand for a specific set of integers.] Clearly, the bubble diagram B_2 cannot support D_2 with all α's positive if line m is contracted. Thus, this line must be one of the two internal lines of D_2. The other line of D_2 cannot be i, because of the projection Q_i, nor can the remaining line of D_2 be line j (or k) for the resulting Landau loop equation; $\alpha_j p_j + \alpha_m p_m = 0$ cannot be satisfied with all α's and all p^0_r positive. Finally, the second line of D_2 cannot be an internal line of the $3 \to 3$ plus bubble because the stability requirement would then force one final external line to emerge from the right-hand vertex of D_2, contrary to its definition. Thus all the possibilities are eliminated, and B_2 cannot support D_2 with all α's positive. Hence, neither can R_1.

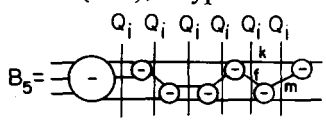
Next consider the diagram D_3 defined in (5.33). After carrying out the substitutions specified in Sec. 5C, we see by inspection that no term of (4.6), except M^{+}_{33}, $T(D_3)$, and possibly contributions of the type



(E2)

can support D_3 with all α's positive. Line m of B_3 cannot be contracted and must be line 4 of D_3. Then applying the same arguments as above, we conclude that B_3 cannot support D_3 with all α's positive. Similar arguments can be made for the extended ice-cream cone diagram.

Consider next the diagram D_5 defined in (5.43). The $3 \to 3$ i bubble occurring in (5.52) cannot support D_5 with all α's positive. This follows by considering the right side of (5.12), a typical term of which is



(E3)

All the bubbles of B_5 can be contracted to points. Line m must evidently also be contracted if one is to obtain D_5. One then sees that any way of picking out three internal lines of B_5 such that the contraction of all others leads to a diagram with the structure of D_5 is such that these three lines are cut by the same Q_i bar of B_5. But then these three lines cannot be lines 1, 2, and 3 of D_5, and hence B_5 cannot support D_5 with all α's positive. Similar arguments show that none of the terms on the right-hand side of (5.52) can support D_5 or D_6 [or in fact any of the diagrams of D_8 defined in (6.2)] with all α's positive.

The $3 \to 3$ i bubble cannot support the last diagram, D'_{13}, on the right side of Eq. (D7) with all α's positive. To see this, consider again the typical term B_5. Evidently, D'_{13} cannot be obtained if m is contracted. Thus m must become one of the two internal lines of D'_{13}. The second one cannot be f, because the Landau equations are not then solvable with positive α_j's and p^0_j's. Thus the second line of D'_{13} must be line k of B_5. Then B_5 cannot support D'_{13} with all α's positive because of the restrictions imposed by the Q_i bar. Similarly arguments show that the last two terms on the left of (D8) cannot support D'_{13} with all α's positive.

Because minus bubbles can be contracted to points, we see that no term on the right side of (6.10) can support a diagram of the set D''_{10} with all α's positive.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Normal and Abnormal Diffusion in Ehrenfests' Wind-Tree Model

E. H. HAUGE
Institutt for Teoretisk Fysikk, Norges Tekniske Høgskole, Trondheim, Norway
AND
E. G. D. COHEN
Rockefeller University, New York

(Received 25 August 1968)

The Ehrenfest wind-tree model—a special case of the Lorentz model—where noninteracting point particles move in the plane through a random array of square scatterers, is used to study the divergences previously discovered in the density expansions of the transport coefficients. Two cases for which the results are qualitatively different are discussed. When the scatterers are *not* allowed to overlap, the diffusion of particles through the array of scatterers is normal, characterized by a diffusion constant D. The calculation of D^{-1} is carried to the second order in the density of the scatterers, and involves a discussion of the above-mentioned divergences and a resummation of all most-divergent terms in the straightforward expansion of D^{-1}. If, however, the trees *are* allowed to overlap, the growth of the mean-square displacement with time is *slower* than linear, so that no diffusion coefficient can be defined. The origin and possible relevance of this new phenomenon to other problems in kinetic theory is discussed. The above results have stimulated molecular dynamics calculations by Wood and Lado on the same model. Their preliminary results seem to confirm the theoretical predictions.

1. INTRODUCTION

It is well established by now that a straightforward generalization of the Boltzmann equation to higher densities is not possible.[1] As a consequence, a density expansion of the transport coefficients does not, in general, exist. In fact, a natural generalization to nonequilibrium of the methods used to obtain virial expansions in equilibrium leads to divergent collision integrals, and thereby to infinite virial coefficients for the transport coefficients.[1,2]

The same infinities are found in the density expansions[3] of the transport coefficients, as expressed in terms of time-correlation functions.[1,4,5] Kawasaki and Oppenheim,[4] in addition, proposed a formal resummation of the expansions which removes a certain class of most-divergent diagrams in each formal order of the density.

Although no rigorous proof of the divergences exists,[6] their origin is in principle well understood. An explicit discussion of the difficulties, as well as of the proposed remedies, however, is complicated by the intricacies of the dynamics of groups of particles interacting through a general short-range potential. For such a discussion it seems therefore natural to consider models which are complicated enough to contain the essential difficulties, but sufficiently simple to permit explicit calculations. The Lorentz models, where one studies the motion of particles without mutual interaction through a random array of stationary scatterers, form one such class of models.

The purpose of the present paper is to investigate the difficulties outlined above as they manifest themselves in a special case of the 2-dimensional Lorentz gas. Before this work was completed, the results of similar studies became available.[7–9] In particular, van Leeuwen and Weijland[7] considered the 2- and 3-dimensional Lorentz models with circular (spherical) scatterers and demonstrated explicitly the existence of divergences in the density expansion of the diffusion coefficient. They were also able to elucidate the resummation of Kawasaki and Oppenheim. Their model was still too complicated, however, to make feasible a computation of the full first order (in the density) for which resummations become necessary.[10]

For a special case of a 2-dimensional Lorentz model, the wind-tree model introduced by Ehrenfest and Ehrenfest[11]—where the randomly distributed scatterers

[1] See for example, J. R. Dorfman and E. G. D. Cohen, J. Math. Phys. **8**, 282 (1967).
[2] J. V. Sengers, Phys. Rev. Letters **15**, 515 (1965).
[3] R. Zwanzig, Phys. Rev. **129**, 486 (1963); K. Kawasaki and I. Oppenheim, Phys. Rev. **136A**, 1519 (1964); M. H. Ernst, J. R. Dorfman and E. G. D. Cohen, Physica **31**, 493 (1965).
[4] K. Kawasaki and L. Oppenheim, Phys. Rev. **139A**, 1763 (1965). See also their paper in *Proceedings of the IUPAP Meeting, Copenhagen, 1966*, T. A. Bak, Ed. (W. A. Benjamin, Inc., New York, 1967), p. 313.
[5] L. K. Haines, J. R. Dorfman, and M. H. Ernst, Phys. Rev. **144**, 207 (1966).
[6] Except in the case of special models.
[7] J. M. J. van Leeuwen and J. Weijland, Phys. Letters **19**, 562 (1966); Physica **36**, 457 (1967) and **38**, 35 (1968).
[8] J. L. Lebowitz and J. K. Percus, Phys. Rev. **155**, 122 (1967).
[9] W. Hoegy, thesis, University of Michigan, Ann Arbor, Mich., 1967.
[10] For example in the case of circular scatterers, where the $O(n^2)$ contribution is the first order in the density n of the scatterers that diverges, van Leeuwen and Weyland computed the coefficient of the term of $O(n^2 \ln n)$, but not of $O(n^2)$.
[11] P. Ehrenfest, *Collected Scientific Papers* (North-Holland Publishing Co., Amsterdam, 1959), p. 229.

are squares with parallel diagonals—we show[12] that in the case where the squares are *not* allowed to overlap each other, it is indeed possible not only to demonstrate the existence of the divergences and discuss their elimination, but also to obtain essentially the complete first correction to the Boltzmann result for the diffusion constant. This again has stimulated a direct evaluation of the diffusion coefficient for this case by molecular dynamics by Wood and Lado, and thus made possible a confrontation between theory and "experiment."

We also consider the case that the squares *are* allowed to overlap each other. In this case, however, the very simplicity of the model introduces new difficulties which lead to the discovery of a novel type of divergence, which does *not* exist in the case of nonoverlapping scatterers within the order to which the calculations are carried out. This *qualitative* difference between the two cases seems to be confirmed by the machine calculations.[13]

In Sec. 2, the precise definition of the model treated in the present paper is given. Section 3 contains the basic formulas on which the subsequent discussion is based. The Boltzmann result for the inverse diffusion coefficient and that part of the first correction to it for which resummations are unnecessary are given in Sec. 4. The existence of the divergences is demonstrated in Sec. 5, where a general prescription for their removal is also given. In Secs. 6, 7, and 8, the contributions to the first correction to the Boltzmann result stemming from the various classes of most divergent diagrams are calculated. The significance of the results, particularly of the new type of divergence discovered in Sec. 8, are discussed in Sec. 9.

2. THE MODEL

In the Ehrenfests' wind-tree model, classical point particles without mutual interaction (the "wind" particles[14]) move in a plane through a random array of immovable square scatterers (the "trees"[14]) with parallel diagonals of length $2a$. The particles move only in the four directions parallel to the diagonals of the squares.

In this paper we are interested in the asymptotic time dependence of the mean-square displacement,

$$\Delta(t) \equiv \langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle, \tag{2.1}$$

of the moving particles, where $\mathbf{r}(t)$ is the position of a particle at time t. The average $\langle\ \rangle$ is to be understood as follows:

First, the scatterers are distributed randomly over the plane according to one of the following two prescriptions:

(A) All configurations are equally probable (i.e., overlapping squares are allowed).

(B) Configurations with overlapping squares are excluded. All other configurations are equally probable. (These are referred to as Cases A and B, respectively.)

Second, the moving particles are inserted between the trees at $t = 0$.

Third, one averages over the square displacement of the moving particles at t for all allowed configurations of the scatterers.

By this prescription[16] for the average, we have excluded (as we are free to do in our model) any influence of the moving particles on the distribution of the trees. Furthermore, since the moving particles have no mutual interaction, we have at the same time reduced the problem to that of a *single* particle moving through a random array of scatterers.

For $\Delta(t)$ we appeal to the well-known equation

$$\frac{d}{dt}\Delta(t) = 2\int_0^t d\tau' \langle \mathbf{v}(t') \cdot \mathbf{v}(t) \rangle = 2\int_0^t d\tau \langle \mathbf{v}(0) \cdot \mathbf{v}(\tau) \rangle, \tag{2.2}$$

where all we have used is the property of time translation invariance of the average, i.e., $\langle \mathbf{v}(t') \cdot \mathbf{v}(t) \rangle = \langle \mathbf{v}(0) \cdot \mathbf{v}(t - t') \rangle$. When the diffusion coefficient D defined by

$$D \doteq \lim_{t\to\infty} \frac{1}{2} \int_0^t d\tau \langle \mathbf{v}(0) \cdot \mathbf{v}(\tau) \rangle \tag{2.3}$$

exists and is nonzero, the asymptotic time-dependence of the mean-square displacement is clearly

$$\Delta(t) \sim 4Dt, \tag{2.4}$$

and the problem is thus reduced to the computation of D.

Strictly speaking, with the system enclosed in a box of finite "volume" $V = L^2$, Δ can at most become of $\mathcal{O}(L^2)$ and it follows from (2.2) that the resulting diffusion coefficient (2.3) vanishes identically. For (2.3) to be a meaningful definition of D, it must be

[12] The main results have been published previously; see E. H. Hauge and E. G. D. Cohen, Phys. Letters, **25A**, 78 (1967). The following detailed report on the present problem is available: E. H. Hauge and E. G. D. Cohen, "Divergences in Non-Equilibrium Statistical Mechanics and Ehrenfest's Wind-Tree Model," *Det Fysiske Seminar i Trondheim, No. 7* (Institute for Theoretical Physics, N.T.H., Trondheim, Norway, 1968).

[13] W. W. Wood and F. Lado (private communication).

[14] This picturesque vocabulary is not found in Ref. 11. However, P. Ehrenfest used it in his lectures, and it has also appeared previously in the literature.[15]

[15] A. J. F. Siegert, Phys. Rev. **75**, 1322 (1949).

[16] Which is precisely the one adopted by Wood and Lado in their molecular dynamics calculations.

understood that the thermodynamic limit $N \to \infty$, $V \to \infty$, $N/V = n$, where N is the number of scatterers in V, must be taken *before* the limit $t \to \infty$.

Even with the limits taken in this order it will be shown in Sec. 10 that with overlapping trees the diffusion coefficient vanishes. This should be interpreted as a result of an abnormal diffusion process in which $\Delta(t)$ grows *slower* than linearly with time for long times. The detailed discussion of this phenomenon will be postponed to Sec. 9.

3. BASIC FORMULAS

In this section we give the formulas on which the discussions in later sections are based. For details of their derivation the reader is referred to Appendix A.

For the inversion introduced later in this section, but in particular for the binary collision expansion used to discuss the divergent terms and their resummation in Secs. 5–8, it is convenient to introduce Laplace transforms. Accordingly we adopt a definition of the diffusion coefficient differing slightly from (2.3):

$$D' = \lim_{\epsilon \to 0} \frac{1}{2} \int_0^\infty dt e^{-\epsilon t} \langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle. \tag{3.1}$$

For the purposes of the present paper the two definitions are completely equivalent, and the distinction will not be made from now on.

For simplicity we assume that the moving particles all have the same absolute value of the velocity v, and one can then write

$$D = \lim_{\epsilon \to 0} \tfrac{1}{2} \mathbf{v} \cdot \hat{\mathbf{\Phi}}(\mathbf{v}, \epsilon), \tag{3.2}$$

with

$$\hat{\mathbf{\Phi}} = \lim_{\substack{N,V \to \infty \\ N/V = n}} \int \cdots \int_V d\mathbf{r}\, dQ^N \rho(\mathbf{r}, Q^N) G(Q^N, x, \epsilon) \mathbf{v}, \tag{3.3}$$

where $Q^N = \{\mathbf{Q}_1, \cdots, \mathbf{Q}_N\}$ are the positions of the N scatterers. $x = (\mathbf{r}, \mathbf{v})$ is the phase of the moving particle and $G(Q^N, x, \epsilon) \equiv G(1 \cdots N)$ is the Laplace transform of the dynamical operator $\exp(t\mathcal{H})$:

$$\begin{aligned} G(1 \cdots N) &= \int_0^\infty dt \exp(-\epsilon t) \exp[t\mathcal{H}(x, Q^N)] \\ &= [\epsilon - \mathcal{H}(x, Q^N)]^{-1}, \end{aligned} \tag{3.4}$$

where $\exp[t\mathcal{H}(x, Q^N)]$ transforms the phase of the moving particle at $t = 0$ into that at t, and is defined by

$$F(x(t)) = \exp[t\mathcal{H}(x, Q^N)] F(x(0)), \tag{3.5}$$

for any $F(x)$. The probability distribution $\rho(\mathbf{r}, Q^N)$ for one moving particle and N scatterers is, according to the prescription in Sec. 2 for the average, given as

$$\rho(\mathbf{r}, Q^N) = Z_N^{-1} \exp[-H(\mathbf{r}, Q^N)], \tag{3.6}$$

with

$$Z_N = \int \cdots \int_V d\mathbf{r}\, dQ^N \exp[-H(\mathbf{r}, Q^N)]. \tag{3.7}$$

One can write

$$H(\mathbf{r}, Q^N) = \sum_{k=1}^N V(\mathbf{r} - \mathbf{Q}_k) + \sum_{k<l} W(\mathbf{Q}_k - \mathbf{Q}_l) \tag{3.8}$$

and in the two cases treated in this paper one has

A and B: $V(\mathbf{r} - \mathbf{Q}_k)$

$$= \begin{cases} \infty, & \text{when } \mathbf{r} \text{ is inside scatterer number } k, \\ 0, & \text{otherwise.} \end{cases} \tag{3.9}$$

(A) $W_A(\mathbf{Q}_k - \mathbf{Q}_l) = 0;$ (3.10)

(B) $W_B(\mathbf{Q}_k - \mathbf{Q}_l)$

$$= \begin{cases} \infty, & \text{when scatterers numbers} \\ & k \text{ and } l \text{ overlap}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.11}$$

A direct density expansion of $\hat{\mathbf{\Phi}}$, and thus of the diffusion coefficient, can be shown not to be meaningful in the limit $\epsilon \to 0$.[3] However, when (3.3) is inverted to give $\mathbf{v}$ in terms of $\hat{\mathbf{\Phi}}$, one can derive the following expression for the inverse diffusion coefficient (see Appendix A):

$$D^{-1} = 2a^{-1}v^{-1}\gamma = \lim_{\epsilon \to 0} 2a^{-1}v^{-1}\gamma(\epsilon), \tag{3.12}$$

with

$$\gamma(\epsilon) = av^{-3}\mathbf{v} \cdot K(\mathbf{v}, \epsilon, n)\mathbf{v}. \tag{3.13}$$

The operator K is given as a formal power series in the density

$$K = \sum_{l=1}^\infty n^l K_l, \tag{3.14}$$

where K_l is expressed in terms of Laplace transforms of dynamical operators involving l scatterers only. The explicit expressions for $l = 1$, 2 are given in (A18) and (A19), and will be used in the next section to compute the corresponding contributions γ_1, γ_2 to γ, which are well behaved in the limit $\epsilon \to 0$.

For $l \geq 3$, the expansion (3.14) is shown in Sec. 5 to contain terms leading to divergent contributions to γ in the limit $\epsilon \to 0$. In Sec. 5 we express the K_l operators in terms of binary collision operators, and it is shown that the expansion (3.14) then serves as a useful starting point for resummations by which these divergences are removed. Consequently, Eqs. (3.12)–(3.14) are adopted as the basis for the subsequent discussion.

4. FINITE RESULTS FROM THE INVERTED EXPANSION TO $O(n^2)$

The two first terms in the formal density expansion of the operator K [see (A18)–(A21)] give rise to contributions to γ that are finite in the limit $\epsilon \to 0$.

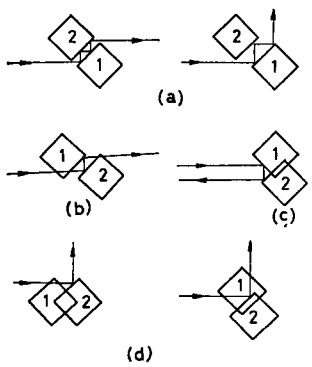


FIG. 1. Events contributing to γ_2.

Except for their numerical values, they are consequently not of central concern in this paper, and no detailed derivation of the corresponding contributions to γ is given here.[17] However, it is interesting that they can all be stated in the form of phase integrals associated with certain collision events, as is exhibited below.

The term of $\mathcal{O}(n)$ (the "Boltzmann term") follows from (A18) and (A20) in the limit $\epsilon \to 0$ and is proportional to the total cross section of the scatterers (cf. Fig. 2.):

$$\gamma_1 = \lim_{\epsilon \to 0} \gamma_1(\epsilon) = an \int_{-a}^{a} db_1 = 2\rho, \qquad (4.1)$$

where b_1 is the impact parameter in the collision between the moving particle and tree 1, and where $\rho = a^2 n$ is the dimensionless density of the scatterers.

The operator K_2 of (A19) gives rise to contributions to γ of $\mathcal{O}(\rho^2)$ [see (A21)]. Some of these terms correspond to actual collision events with two trees, while others should be viewed as corrections to the $\mathcal{O}(\rho)$ result due to the nature of the expansion formalism. In the limit $\epsilon \to 0$, however, they can all be expressed in terms of phase integrals of the type

$$\gamma_2^{(i)} = \lim_{\epsilon \to 0} \gamma_2^{(i)}(\epsilon) = an^2 \int_{-a}^{a} db_1 \int^{(i)} d\mathbf{Q}_2, \qquad (4.2)$$

where $\mathbf{Q}_1$ is kept fixed, and with various restrictions, here indicated by (i), on the integration over $\mathbf{Q}_2$. We define three types of contributions:

(A) The "real" contribution $\gamma_2^{(r)}$ stems from events of the type shown in Fig. 1(a) with real collisions only. The corresponding restrictions on the integration over $\mathbf{Q}_2$ are such that:

(i) the moving particle collides with tree 1 first;

(ii) the total number of collisions in the event is an odd integer larger than, or equal, to 3.

[17] Details are given in Ref. 12.

With these restrictions integration yields

$$\gamma_2^{(r)} = (\pi^2/9 - 1)\rho^2. \qquad (4.3)$$

(B) The "virtual" contribution $\gamma_2^{(v)}$ corresponds to the event shown in Fig. 1(b) and the integration is over configurations such that

(i) the particle first has a "virtual" collision (i.e., it passes through tree 1);

(ii) it subsequently suffers a real collision with tree 2 and a second, real collision with tree 1. Possible additional collisions should not be taken into account.

The result of the integration is

$$\gamma_2^{(v)} = \rho^2. \qquad (4.4)$$

The contributions (4.3) and (4.4) are independent of whether the trees are allowed to overlap or not. In the derivation of $\gamma_2^{(r)}$, however, the factor $g(\mathbf{r}, 1, 2)$ in (A21) was put equal to unity. This is strictly correct in the limit $\epsilon \to 0$ for Case A, but with nonoverlapping trees it leads to an error which is corrected for by the following.

(C) The last type is the "overlap" contribution, which vanishes by definition in Case A:

$$\gamma_2^{(O)A} \equiv 0. \qquad (4.5)$$

In Case B it can be written as a sum of two terms, both of the form (4.2):

$$\gamma_2^{(O)B} = -I^{(O)} + J^{(O)}. \qquad (4.6)$$

$I^{(O)}$ stems from events of the type shown in Fig. 1(c) and the restrictions on the integration are accordingly:

(i) tree 1 overlaps with tree 2;

(ii) the particle has a real collision, first with tree 1 and subsequently with tree 2.

The result is

$$I^{(O)} = \tfrac{4}{3}\rho^2. \qquad (4.7)$$

Figure 1(d) shows the type of events leading to $J^{(O)}$, where

(iii) trees 1 and 2 overlap;

(iv) the particle first has a virtual collision with tree 1 and subsequently a real collision with tree 2.

One finds

$$J^{(O)} = \tfrac{16}{3}\rho^2, \qquad (4.8)$$

so that

$$\gamma_2^{(O)B} = 4\rho^2. \qquad (4.9)$$

Summing (4.3), (4.4), and (4.5) or (4.9), for the complete $\mathcal{O}(\rho^2)$ contribution stemming from K_2 one finds

$$\gamma_2^A = (\pi^2/9)\rho^2, \qquad (4.10)$$

$$\gamma_2^B = (\pi^2/9 + 4)\rho^2. \qquad (4.11)$$

Two remarks are here in order:

(a) According to the rule given by Kawasaki and Oppenheim,[4] γ_2 should diverge for 2-dimensional models like $\ln \epsilon$ in the limit $\epsilon \to 0$. This is in fact borne out by the Lorentz model with *circular* scatterers,[7] where the divergence can be traced to the event corresponding to the first one shown in our Fig.1 (a). In the present case, however, the divergence is suppressed because the discreteness of the velocity space poses a severe restriction on the allowed phase of tree 2 relative to tree 1. The example shows that the details of the interaction cannot be completely neglected in arguments like the one given in Ref. 4.

(b) It is also worth noting that in a Lorentz model with oriented polygons of $4N$ sides as scatterers, in the limit $\epsilon \to 0$, one finds

$$\gamma_2(0) \sim \rho^2 \ln N + \text{terms finite when } N \to \infty,$$

so that in the circular limit $N \to \infty$ (at constant diameter), we retrieve the logarithmic divergence mentioned under point (a).

If the operator K of (3.14) had been well behaved in the limit $\epsilon \to 0$, the results (4.10) and (4.11) would have been the complete contribution of $\mathcal{O}(\rho^2)$. We show in the following section, however, that the K expansion still contains divergences as $\epsilon \to 0$, and as a result (4.10) and (4.11) do *not* represent the complete contribution to γ of $\mathcal{O}(\rho^2)$.

5. DIVERGENT TERMS

A. Resummations

For $l \geq 3$ the expansion (3.14) gives rise to divergent contributions to $\gamma(\epsilon)$ in the limit $\epsilon \to 0$, as is shown explicitly below. The great majority of the most divergent terms to each order in n, which are of central interest in the following, are independent of the statistical factor $\rho(\mathbf{r}, Q^N)$ in (3.3). Accordingly it is convenient to split the operators K_l ($l \geq 3$) of (3.14) into two parts,

$$K_l = K_l^{\mathrm{d}} + K_l^{\mathrm{st}}. \tag{5.1}$$

The "dynamical" part K_l^{d} is defined as the resulting K_l if $\rho(\mathbf{r}, Q^N)$ in (3.3) is put equal to unity. Equation (5.1) then defines the "statistical" part K_l^{st} as the corresponding correction term. By definition, K_l^{d} is the same for both Cases A and B, while K_l^{st} is different in the two cases.

A simple representation of K_l^{d} in terms of the *paths* of collision events with l trees (essential for the classification of events) is derived by the following two steps:

First, we express K_l^{d} as an expansion in terms of the binary collision operator $C(1, x, \epsilon) = C(1)$, given by

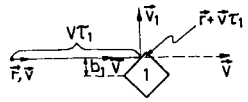


FIG. 2. Effect of the binary-collision operator.

$$C(1)F(\mathbf{v}, \mathbf{r}) = e^{-\epsilon\tau_1}\{F(\mathbf{v}_1, \mathbf{r} + \mathbf{v}\tau_1) - F(\mathbf{v}, \mathbf{r} + \mathbf{v}\tau_1)\}, \tag{5.2}$$

for any function $F(\mathbf{v}, \mathbf{r})$ (see Appendix B). Here τ_1 is the time from the initial configuration $(\mathbf{r}, \mathbf{v}, \mathbf{Q}_1)$ until the collision of the moving particle with tree 1. If there is no such collision, $C(1) = 0$ and we conveniently define τ_1 to be infinite. The velocities before and after the collision are $\mathbf{v}$ and $\mathbf{v}_1$, respectively (see Fig. 2).

The binary-collision expansion[3,18] of K_l^{d} reads as follows (see Appendix B):

$$K_l^{\mathrm{d}}\mathbf{v} = -\epsilon \sum_{p=l+1}^{\infty} \sum_{\substack{\{i\} \\ (1,\cdots,5)}}{}' \int \cdots \int dQ^l C(i_1) \cdots C(i_p)\mathbf{v}, \tag{5.3}$$

where conditions on the sum over the sets of labels $\{i\}$ are the following:

(1) All labels $i_1 \cdots i_p$ belong to the set $\{1 \cdots l\}$.

(2) $i_{n+1} \neq i_n$, $n = 1, 2, \cdots, p-1$.

(3) All labels in the set $\{1 \cdots l\}$ occur at least once in the set $\{i_1 \cdots i_p\}$.

(4) The labels are ordered with respect to their first appearance in the sequence $i_1 \cdots i_p$.

(5) The sum is over irreducible products only.

The term "irreducible" is defined in Appendix B. A simpler definition in terms of the path of collision events is given below.

The second step is to split the binary-collision operator C into two parts, each of which has a well-defined effect on the velocity:

$$C(1) = C^r(1) + C^v(1), \tag{5.4}$$

where the real[19] collision operator $C^r(1)$ is defined by [cf. (5.2)]:

$$C^r(1)F(\mathbf{v}, \mathbf{r}) = e^{-\epsilon\tau_1}F(\mathbf{v}_1, \mathbf{r} + \mathbf{v}\tau_1), \tag{5.5}$$

and the virtual[19] collision operator $C^v(1)$ is correspondingly given by

$$C^v(1)F(\mathbf{v}, \mathbf{r}) = -e^{-\epsilon\tau_1}F(\mathbf{v}, \mathbf{r} + \mathbf{v}\tau_1). \tag{5.6}$$

[18] T. D. Lee and C. N. Yang, Phys. Rev. **113**, 1165 (1959); A. J. F. Siegert and E. Teramoto, Phys. Rev. **110**, 1238 (1958).

[19] The terms "real" and "virtual" are here used in a similar sense as in Sec. 4.

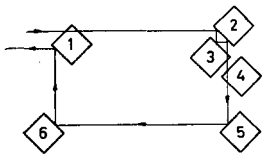


FIG. 3. Graph representing the product $C^v(1)C^r(2)C^r(3)C^r(2)C^v(3) \times C^v(4)C^r(5)C^r(6)C^r(1)$, *or* the same product with $C^v(3)$ missing.

Introduction of this splitting into (5.3) gives

$$K_l^d \mathbf{v} = -\epsilon \sum_{p=l+1}^{\infty} \sum_{\substack{\{i\} \\ (1,\cdots,5)}}' \sum_{\substack{s_1=r,v;\cdots; \\ s_p=r,v}} \int \cdots \int dQ^l \times C^{s_1}(i_1) \cdots C^{s_p}(i_p)\mathbf{v}. \quad (5.7)$$

The point distinguishing (5.7) from (5.3) is that now every operator in the integrand has a well-defined effect on the velocity (with C^r it still depends on the impact parameter, of course). A given term in (5.7) can therefore be given a simple graphical representation in terms of a path.[20] An example is shown in Fig. 3. The notion of an irreducible product is readily interpreted on the basis of such diagrams. A product of binary-collision operators (real or virtual) is called irreducible if one cannot make the corresponding graph disconnected by cutting the path at any single point.

Using (3.4), (5.5), and (5.6), and introducing

$$\int dQ_1 \to v \int_0^{\infty} d\tau_1 \int_{-a}^{a} db_1, \quad (5.8)$$

we can immediately perform the τ_1 integration in (5.7) to get

$$\begin{aligned} n^l K_l^d \mathbf{v} &= n^l \sum_{p=l+1} \sum_{\{i\}}' \sum_{\{s\}} (-1)^{V+1} v \\ &\quad \times \int_{-a}^{a} db_1 \int \cdots \int' dQ_2 \cdots dQ_l \\ &\quad \times e^{-\epsilon(\tau_2+\cdots+\tau_p)} \mathbf{v}_p(\{i\},\{s\}) \\ &= n^l \sum_{\text{irr},r,v} (-1)^{V+1} v \\ &\quad \times \int_{-a}^{a} db_1 \int \cdots \int' dQ_2 \cdots dQ_l e^{-\epsilon\tau} \mathbf{v}_f, \quad (5.9) \end{aligned}$$

where V is the number of virtual collisions in the sequence $\{s\}$, τ_k is the time between the $(k-1)$th and the kth collision, and $\mathbf{v}_p(\{i\},\{s\})$ the velocity after the sequence of p collisions has been completed. [That is, $\mathbf{v}_p(\{i\},\{s\})$ is the result of a product of C-operators, characterized by the sets $\{i\}$ and $\{s\}$, acting on the initial velocity $\mathbf{v}$.] The prime indicates that the integrations over $\mathbf{Q}_2 \cdots \mathbf{Q}_l$ are restricted to configurations such that the given sequence of collisions can take place. In the second line we have simply indicated that the sum should go over all irreducible (irr) sequences of real (r) and virtual (v) collisions involving l trees. Furthermore, $\tau = \sum_i \tau_i$ is the time between the first and the last collision in the sequence, and $\mathbf{v}_f$ the final velocity after the event has been completed.

[20] Note that there is not a one-to-one correspondence between a product of operators and a graph. A product of $C^{s_i}(i)$-operators determines uniquely (apart from the trivial dependence on the impact parameter of C^r) the corresponding graph. A given graph, however, can correspond to several products of operators. The point is illustrated by Fig. 3 where to the second collision with tree 3, one may or may not associate a virtual collision operator. The respective contributions to γ will be of opposite sign, but will not cancel since the additional $C^v(3)$ represents a restriction on the integration over $\mathbf{Q}_3$. (See, however, Ref. 12, Appendix F, where a class of events is considered for which the corresponding cancellation is complete.)

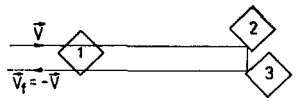


FIG. 4. Event leading to a divergence.

B. Divergences

It is now easy to see from (5.9) that all divergences in the limit $\epsilon \to 0$ have *not* been removed by passing from the B expansion to the K expansion as was done in Appendix A. Consider, for example, the sequence of operators $C^v(1)C^r(2)C^r(3)C^v(1)$. The corresponding event is shown in Fig. 4. The leading contribution to $\mathbf{v} \cdot n^3 K_3^d \mathbf{v}$ from this event for small ϵ is, by (5.8) and (5.9),

$$\begin{aligned} &\mathbf{v} \cdot n^3(-1)^3 v \int_{-a}^{a} db_1 \int_{-a}^{0} db_2 \int_{-a}^{0} db_3 \int_0^{\infty} d(v\tau_2) \\ &\quad \times \int_0^{\mathcal{O}(a)} d(v\tau_3) e^{-\epsilon(\tau_2+\tau_3+\tau_4)}(-\mathbf{v}) \sim n^3 v^4 a^4 \epsilon^{-1}, \quad (5.10) \end{aligned}$$

which clearly does not exist in the limit $\epsilon \to 0$. The divergence stems from the $v\tau_2$ integration, i.e., from the integration over infinitely long paths between tree 1 and trees 2 and 3. We remark that

(i) the divergence difficulty in the example is *not* spurious and one can convince oneself that the divergent contributions to any single K_l^d do not cancel in general[21];

(ii) for a classification of the various contributions the splitting (5.4) of C leading to a well-defined path is essential; and

(iii) the order of the divergences increases with increasing order of n and, in general, one finds that the most divergent contributions in a given $\mathcal{O}(n)$

[21] The order of the limiting processes is unimportant [the literature contains statements to the contrary; see J. Stecki, Phys. Letters **19**, 123 (1965)]. If we put $\epsilon = 0$ in the integrand of (5.10), we get a long path divergence. If we integrate first, we get a divergence for small ϵ. And finally, if we had also performed an inverse Laplace transform on (5.10) the result would have been a long *time* divergence.

behave like

$$\gamma_l(\epsilon) \equiv n^l a v^{-3}\mathbf{v}\cdot K_l\mathbf{v} \sim n^l\epsilon^{-(l-2)}, \quad l \geq 3.^{22} \quad (5.11)$$

C. Heuristic Discussion

The divergence difficulties encountered in Sec. 5B are of precisely the same origin as those occurring with more realistic models. The essence of cluster expansions, such as the one introduced in Appendix A, is to express the properties of the *total* N-body system as a series expansion, the terms of which are successively determined by the properties of the 2, 3, 4, $\cdots$ body systems. This is precisely the program that leads to the equilibrium virial series for systems with forces of finite range. In nonequilibrium, however, a new type of infinite-range difficulty arises in such expansions from correlations due to particles traversing very long paths.

Physically speaking, what is wrong with (5.9) is that, in the limit $\epsilon \to 0$, events that need very long times for their completion are overemphasized. From a purely heuristic point of view, one is tempted to correct for this by introducing a probability $\exp(-l/\lambda) = \exp(-\alpha\tau)$ that a straight path of length $l = v\tau$ be left undisturbed. (Here λ is the mean free path, and $\alpha^{-1} = \lambda/v$ is the mean free time between collisions.) If this is done, the effect on (5.9) is that even in the limit $\epsilon \to 0$ the damping $\exp(-\alpha\tau)$ on the integrand remains.

All the integrals thereby become finite. However, the density dependence of every term will be different, since to lowest order $\alpha \sim \lambda^{-1} \sim n$. Effectively then, all ϵ's can be replaced by (const $\times$ n) and the estimate (5.11) for the most divergent contributions in any formal order of n is changed to

$$\gamma_l = \lim_{\epsilon\to 0}\gamma_l(\epsilon) \sim n^l n^{-(l-2)} = n^2. \quad (5.12)$$

Thus, that part within every formal order of n (in the K expansion) which is most divergent as $\epsilon \to 0$, in the final analysis contributes to the first correction to the Boltzmann result, and will thus be of the same order as the γ_2 computed in Sec. 4. Less divergent pieces only contribute in higher orders of the density.

D. Formal Argument

The beauty of the Lorentz models is that the heuristic arguments of the preceding paragraphs can in a simple way be paralleled by explicit calculations based on (5.7) and (5.9).

Take an arbitrary term in the sum $n^l K_l^{\mathrm{d}}\mathbf{v}$ [Eq. (5.7)]

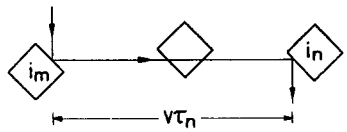


FIG. 5. Uncorrelated virtual collision.

and consider the integrand with a fixed configuration $\mathbf{Q}_1, \cdots, \mathbf{Q}_l$, so that the path is well defined. Let $C^{s_m}(i_m)C^{s_n}(i_n)$ be two consecutive operators in the product. Form the sum

$$\begin{aligned} &C^{s_1}(i_1)\cdots C^{s_m}(i_m)\\ &\quad\times\Big[1 + n\int' d\mathbf{Q}_{j_1}C^v(j_1)\\ &\quad + n^2\int' d\mathbf{Q}_{j_1}\int'' d\mathbf{Q}_{j_2}C^v(j_1)C^v(j_2) + \cdots\Big]\\ &\quad\times C^{s_n}(i_n)\cdots C^{s_p}(i_p)G\mathbf{v}. \end{aligned} \quad (5.13)$$

The first term reproduces the integrand considered. In the second one, tree j_1 is *not* a member of the basic set of l trees. Since the integration is restricted so that the virtual collision with j_1 occurs between the collisions with trees i_m and i_n, the second term is part of the sum $n^{l+1}K_{l+1}^{\mathrm{d}}\mathbf{v}$, although the *path* is exactly the same as in the first term. In the third term none of the trees j_1, j_2 belongs to the basic set, and the integrations are such that the virtual collision with j_1 occurs after the collision with i_m; the virtual collision with j_2 occurs after the collision with j_1, but before the collision with i_n [see condition 4 on the sum in (5.3) and (5.7)]. The third term is thus a part of $n^{l+2}K_{l+2}^{\mathrm{d}}\mathbf{v}$, and in the same way one goes on.

All of the inserted virtual collisions are with trees occurring only once in the complete integrand. We call them *uncorrelated* virtual collisions. They can clearly be integrated over with the basic set $\mathbf{Q}_1\cdots\mathbf{Q}_l$ (or equivalently the *path*) fixed, and since in these dynamical contributions all g's are replaced by unity even in Case B, the inserted trees can freely overlap. Thus the sum in the bracket of (5.13) reduces to

$$\sum_{k=0}^{\infty}\frac{n^k}{k!}\Big[-\int' d\mathbf{Q}\Big]^k = \exp\Big[-n\int' d\mathbf{Q}\Big]. \quad (5.14)$$

The restriction on the integration is the same as for $\int' d\mathbf{Q}_{j_1}$ in (5.13). From Fig. 5 it follows that

$$\int' d\mathbf{Q} = 2av\tau_n + \mathcal{O}(a^2). \quad (5.15)$$

Since, to $\mathcal{O}(\rho^2)$ in γ, we are only interested in the most divergent parts of every formal order in n, the term of $\mathcal{O}(a^2)$ in (5.15) can be neglected.

Clearly the sum in (5.13) can be inserted between *any* two collision operators in the basic sequence

[22] To *prove* this result a general classification scheme of all irreducible graphs, which is not available at present, would be needed. See Sec. 5E.

without affecting the path. The result is a damping

$$\exp\left[-2nav(\tau_1+\tau_2+\cdots+\tau_p)\right]=\exp\left(-2nav\tau\right), \tag{5.16}$$

on a path of length $v\tau$ between the first and the last collision in the sequence, precisely as anticipated from the heuristic argument above.

Through this damping, one implicitly takes into account all uncorrelated virtual collisions, so that only real (r) and correlated virtual (cv) ones (i.e., with the label "tree" occurring more than once in the product of operators) should be taken explicitly into account. With the approximation (5.15), (5.9) is thus changed to

$$n^l K_l^{\mathrm{d}}\mathbf{v}=n^l\sum_{\substack{\text{irr}\\ r,cv}}(-1)^{V+1}v \times\int_{-a}^{a}db_1\int\cdots\int' d\mathbf{Q}_2\cdots d\mathbf{Q}_l e^{-2nav\tau}\mathbf{v}_f, \tag{5.17}$$

where V here is the number of *correlated* virtual collisions, and we have gone to the limit $\epsilon\to 0$, since all integrals now clearly exist in this limit. Note that a single K_l^{d} contains parts from every K_m^{d} with $m\geq l$.

E. Classification of the Most-Divergent Diagrams

As we remarked in Sec. 5C, all most-divergent terms in a given formal order of n, after a resummation contribute to $\mathcal{O}(\rho^2)$ in γ. Since the divergences in (5.9) stem from the fact that an integration over an infinitely long path leads to a factor $1/\epsilon$, the topology of the path corresponding to a given product of operators in (5.7) is clearly essential. A classification based on (5.7) and (5.9), of all paths according to their orders in n and ϵ, would therefore be desirable. Such a general scheme is not available, and consequently no criterion exists by which one can decide whether all diagrams contributing to $\mathcal{O}(\rho^2)$ have been found. The following classification of diagrams with asymptotic behavior $n^l\epsilon^{-(l-2)}$ must be considered with this qualification in mind.

Since intermediate virtual collisions do not influence the path, our classification is based on sequences of *real* collisions, except for Class I and a corresponding part of Class III. The motivation for this exception becomes clear later. The three contributing classes of most-divergent diagrams are then given as:

Class I: Ring events, characterized by sequences of collision operators (real or virtual) of the type $1, 2, 3, \cdots, r, 1$, rather than by a prescribed path. The numbers label the trees in the collision sequence. An example with $r=3$ was shown in Fig. 4.

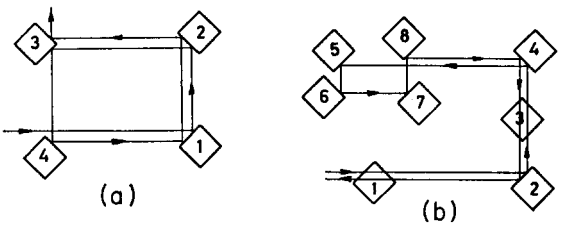


FIG. 6. (a) Orbiting event with $r=4$, $p=1$, $k=3$. (b) Retracing event with $m=4$, $r=4$.

Class II: Orbiting events, where the particle stays in the "orbit" for an arbitrary number of additional collisions, characterized by the following sequence of real collisions: $(1, 2, 3, \cdots, r)^p 1, 2, \cdots, k$. Here it is understood that: (1) $r\geq 4$, $p\geq 1$, $1\leq k<r$, and the case $p=1$, $k=1$ is excluded (it corresponds to the ring R_r); (2) the very first and the very last tree in the sequence may suffer real *or* virtual collisions; and (3) the basic sequence above may be decorated by any combination of intermediate virtual collisions. An example with $r=4$, $p=1$, and $k=3$ is shown in Fig. 6a.

Class III: Retracing events, with sequences of the type $1, 2, 3, \cdots, m, m+1, m+2, \cdots, m+r, m, m-1, \cdots, 2, 1$. It is understood that: (1) $m\geq 2$, $r\geq 2$; (2) the two collisions with trees $2, 3, \cdots, m-1$ must either be both real or both virtual; (3) any combination of real and virtual collisions with trees 1 and m is allowed (if *all* collisions with $1\cdots m$ are virtual and $r=2$, we get the exceptional path identical with that of the ring shown in Fig. 4); and (4) the basic sequence may be decorated by uncorrelated virtual collisions.

An example is shown in Fig.6(b),and the characteristic feature is the double path between trees 1 and 4 along which several trees are encountered twice, closed by a "reflecting chain."

All three classes give rise to contributions of $\mathcal{O}(\rho^2)$ to γ and will be considered in the three subsequent sections. There is also a fourth class which is a combination of Class II and Class III, but the corresponding contributions cancel exactly.[12]

6. CLASS I: THE RINGS

In this section we calculate the contribution of $\mathcal{O}(\rho^2)$ to γ from the first class of most divergent diagrams. We start by considering Case A, overlapping trees, and the modifications necessary in Case B are treated at the end of this section.

A. Overlapping Trees

The simplest ring R_2 with a collision sequence 121 [Figs. 1(a), 1(b)] is not a divergent diagram. It was

already calculated in Sec. 4 as a part of γ_2. What we are interested in then are the contributions γ_{R_l} with $l = 3, 4, \cdots$, from the rings R_l calculated to $\mathcal{O}(\rho^2)$.

The straightforward way to do this is to specialize the general formula (5.17) to the case of rings; for the contribution to γ from the rings with l trees, (3.12) and (3.13) then give

$$\gamma^A_{R_l} = n^l a v^{-2} \sum_{\text{rings}}{}' (-1)^{V+1} \cdot \mathbf{v} \int_{-a}^{a} db_1 \int \cdots \int' d\mathbf{Q}_2 \cdots d\mathbf{Q}_l e^{-2nav\tau} \mathbf{v}_f, \quad (6.1)$$

where it follows from the conditions on the sum (5.17) that here all intermediate collisions have to be real, while the first and the final collision with tree 1 can be real or virtual. The restriction on the integrations is simply that the configurations must be such that a ring event takes place.

Capitalizing on the fact that we are only interested in γ_{R_l} to $\mathcal{O}(\rho^2)$, we can carry out the integrals. We do not give details here, since the final result is derived by a more elegant procedure in Appendix C. The results to $\mathcal{O}(\rho^2)$ are

$$\gamma^A_{R_3} = \rho^2,$$
$$\gamma^A_{R_{2p}} = 0,$$
$$\gamma^A_{R_{2p+1}} = \rho^2 2^{5-4p} \left\{ \binom{2p-4}{p-2} + \binom{2p-4}{p-1} \right\} \times \left\{ \binom{2p-2}{p-1} - \binom{2p-2}{p} \right\}, \quad p \geq 2, \quad (6.2)$$

and the sum is

$$\gamma^A_I \equiv \sum_{l=3}^{\infty} \gamma^A_{R_l} = \frac{4}{\pi} \rho^2. \quad (6.3)$$

B. Nonoverlapping Trees

Does it make a difference to $\mathcal{O}(\rho^2)$, whether the trees are allowed to overlap or not? For contributions from events where all paths are integrated to infinity the answer is *no*, since the difference between our cases A and B becomes manifest only when the distance between two trees is of $\mathcal{O}(a)$. For all rings with more than two real intermediate collisions, the difference between $\gamma^A_{R_l}$ and $\gamma^B_{R_l}$ thus becomes of $\mathcal{O}(\rho^3)$, and therefore is immaterial in the present context.

The preceding reveals the motivation for the splitting of the K_l operators into a "dynamical" and a "statistical" part [Eq. (5.1)]. The hope was that the dynamical terms with fewer restrictions on the integrations would contain all of the most-divergent diagrams, and therefore be the only important part to $\mathcal{O}(\rho^2)$.

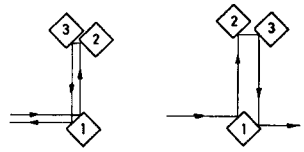


FIG. 7. Events contributing to γ_{R_3}.

The resummed ring R_3 (one example is shown in Fig. 4) is an exception to this rule, however. Here trees 2 and 3 are bound to be close together throughout the integration, and whether the trees are allowed to overlap or not *does* have an effect on the resulting contribution. This distinction becomes the crucial point in the calculation of the Class III events in Sec. 8.

To calculate $\gamma^B_{R_3}$, one needs simply to insert a factor

$$g(2, 3) = \exp\{-W^B(2, 3)\} + \mathcal{O}(\rho) \quad (6.4)$$

into (6.1) with $l = 3$, rather than split into a dynamical and a statistical part. The computation is straightforward; we only mention that now one has to include the contributions from the events in Fig. 7 [their sum is of $\mathcal{O}(\rho^3)$ in Case A], besides the one from the event of Fig. 4.

The result of the calculation to $\mathcal{O}(\rho^2)$ is

$$\gamma^B_{R_3} = \tfrac{7}{24}\rho^2, \quad (6.5)$$

and since

$$\gamma^B_{R_l} = \gamma^A_{R_l} + \mathcal{O}(\rho^3), \quad l = 4, 5, \cdots, \quad (6.6)$$

it follows from (6.2) and (6.3) that to $\mathcal{O}(\rho^2)$

$$\gamma^B_I = \sum_{l=3}^{\infty} \gamma^B_{R_l} = \left(\frac{4}{\pi} - \frac{17}{24}\right)\rho^2. \quad (6.7)$$

We finally remark that the difference between the two cases A and B, to $\mathcal{O}(\rho^2)$ for the rings, is a result of the discreteness of velocity space in our model. With circular scatterers replacing the square trees, there is no such distinction.

7. CLASS II: ORBITING EVENTS

The next class of most-divergent diagrams listed at the end of Sec. 5 consists of the orbiting events. The orbit is determined by an even number $2r$ of trees. In the example of Fig. 8, the trees 2, 3, 4, 5, 6,

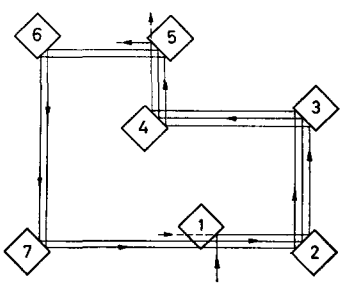


FIG. 8. Orbiting event with $r = 3$, $s = 11$.

and 7 define the corners in the orbit. All intermediate collisions with the corners must clearly be real. The very first and the very last collision in the sequence may be real or virtual, however, and may occur at one of the corners *or* along one of the sides (see Fig. 8[23]). If the initial or the final collision is with a tree along one of the sides in the orbits, all additional collisions with that tree must be virtual. In Ref. 12 it is shown that all other intermediate virtual collisions cooperate to give a damping $\exp\{-n[2a + (k-1)x]L\}$ on k parallel paths on length L, and with a distance x between the subsequent paths. These collisions should not,therefore, be taken explicitly into account.

Defining s as the number of collisions in an orbiting event in addition to those necessary to complete the corresponding ring, we can on the basis of the above discussion write the contribution of $\mathcal{O}(\rho^2)$ from Class II to γ as

$$\gamma_{II} = \sum_{r=2}^{\infty} \sum_{s=1}^{\infty} T(2r, s). \tag{7.1}$$

It is easily seen on the basis of the discussion in Sec. 6B that γ_{II} is the *same* in cases A and B to $\mathcal{O}(\rho^2)$.

Any term $T(2r, s)$ can in principle be calculated. Unfortunately, neither the general term $T(2r, s)$ nor the sum γ_{II} is known explicitly. However, the following partial results are available.

The first "row" of the double sum (7.1) is found to give[12]

$$\gamma_{4,\infty} \equiv \sum_{s=1}^{\infty} T(4, s) = (0.0627 - 0.0238 + \cdots)\rho^2 = 0.0295\rho^2. \tag{7.2}$$

Using the operator Ω (see Appendix C) to sum over all uncorrelated intermediate collisions, one can also find the two first columns[12]

$$\gamma_{\infty,1} = \sum_{r=2}^{\infty} T(2r, 1) = +0.1118\rho^2, \tag{7.3}$$

$$\gamma_{\infty,2} = \sum_{r=2}^{\infty} T(2r, 2) = -0.0353\rho^2. \tag{7.4}$$

To arrive at an order-of-magnitude estimate on the double sum (7.1), we conjecture that the terms $T(2r, s)$ have roughly the same s-dependence for all r,[24] in particular that asymptotically for $s \gg 1$, $T(2r, s) \sim s^{-2}$ [which is the asymptotic behavior of $T(4, s)$]. On the basis of this conjecture we find

$$\gamma_{II} \sim \gamma_{\infty,1}\gamma_{4,\infty}/T(4, 1) \approx 0.05\rho^2 \tag{7.5}$$

or

$$\gamma_{II} \lesssim 0.1\rho^2. \tag{7.6}$$

Clearly, the material available is not sufficient to allow us to take the estimate (7.6) very seriously.[25]

Our discussion in Sec. 9 of the final result, however, will be based on the conjecture (7.6). At this point we only remark that one can look upon γ_{II} as a correction to the ring contribution; and comparing it with (6.3) and (6.7), which numerically read

$$\gamma_I^A = 1.2732 \cdots \rho^2, \tag{7.7}$$

$$\gamma_I^B = 0.5649 \cdots \rho^2, \tag{7.8}$$

one finds that the correction from the orbiting events seems to be an order of magnitude smaller.

8. CLASS III: RETRACING EVENTS

The retracing events form the final contributing class of most divergent diagrams listed in Sec. 5 and are considered in the present section. Again it is convenient from a computational point of view to start with the calculations for overlapping trees and then treat the necessary modifications for Case B.

A. Overlapping Trees

The typical Class III event $1, 2, \cdots, m, m+1, m+2, \cdots, m+r, m, m-1, \cdots, 2, 1$ naturally splits into two distinct parts. First, a double path from tree 1 to tree m along which the moving particle retraces its previous steps. Second, a chain of uncorrelated collisions from tree m and back to the same tree [cf. Fig. 6(b)]. Accordingly, the computation of γ_{III} is done in two steps. First, for any fixed double path, the ring operator Ω introduced in Appendix C is used to sum over all chains of uncorrelated collisions between the two encounters with the final tree along the double path. This sum can be represented by two reflectors with certain weights. Then, using these reflectors, one sums over all possible double paths.

1. The Reflectors

The sum over all chains of real and virtual uncorrelated collisions between the two collisions with tree m is given by

$$nC(m)\Big[\sum_{r=2}^{\infty} n^r \int \cdots \int d\mathbf{Q}_{m+1} \cdots d\mathbf{Q}_{m+r} \times C(m+1) \cdots C(m+r)\Big] C(m) \equiv \bar{B}(\mathbf{Q}_m, \mathbf{Q}_m). \tag{8.1}$$

[23] Note that at least one encounter with tree 1, in addition to the initial real or virtual one, must be counted as a virtual collision; otherwise the event would not be irreducible.

[24] Compare the two first terms in the sum (7.2) with (7.3) and (7.4).

[25] Not even the *convergence* of the double sum (7.1) has been proved; and it seems quite difficult to do so, since the binary-collision expansion splits the Class II contributions into terms that have to be partially recombined before a divergence-free summation can be performed.

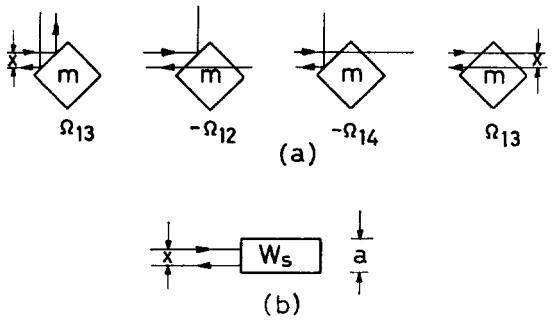


FIG. 9. The reflector of type s.

That the sum starts at $r = 2$ follows from the obvious fact that at least two collisions are needed to reflect the moving particle back onto the double path. Chain integrals like those in (8.1) are conveniently handled by the Fourier technique used in Appendix C, and to $\mathcal{O}(n^2)$ one has[12]

$$\bar{B}(\mathbf{Q}_m, \mathbf{Q}_m) = e^{-\epsilon\tau_m}\bar{C}(m)an^2 \times \int \frac{d\mathbf{K}}{(2\pi)^2}\,\Omega(K_1, K_2)\int_{-a}^{a} db'_m \bar{C}'(m). \tag{8.2}$$

Here $\bar{C}(m) = e^{\epsilon\tau_m}C(m)$, and the matrix elements of the dimensionless operator $\Omega(K_1K_2)$—where K_1 and K_2 are the components of the dimensionless vector $\mathbf{K}$ along the $\mathbf{v}_1$ and $\mathbf{v}_2$ axes, respectively—are given in Appendix C. The prime is only used to distinguish the variables characterizing the *second* collision with tree m. In (8.2) the exponential $\exp\{ian\mathbf{K}\cdot(\boldsymbol{\sigma}_m - \boldsymbol{\sigma}'_m)\}$ has been put equal to unity to $\mathcal{O}(n^2)$ by the same argument as used in Appendix C.

For actual computations we have to pick that part of the operator $\bar{B}$ of (8.2) which reflects the particle to make it retrace its previous steps from 1 to m. [Or more formally, we are only interested in the matrix element $(\bar{C}(m)\Omega\bar{C}'(m))_{13}$ of the operator product.] In Fig. 9(a) are shown the four possible situations at tree m when we insist that the particle hits the upper half of the tree both in the first *and* in the second collision (same-side reflection: type s).

Below each combination of collisions with tree m, the corresponding matrix element Ω_{1j} is given (when $\mathbf{v}_1$ in each case is defined as the direction of the moving particle immediately *after* the first collision with tree m). The sign depends on whether the real or the virtual part of $C(m)$ and $C'(m)$ has been used.

The same combination of matrix elements occurs when both paths touch the lower half of tree m, leading to a factor 2. Since the integration over $\mathbf{Q}_m$ leads to an integration over a collision parameter b_m, we can write the integrations over b_m and over b'_m in (8.2) as follows:

$$\int_{-a}^{a} db_m \int^{(s)} db'_m \cdots \to \sum_P \int_0^a dx 2(a-x)\cdots \equiv \sum_P \int_0^{2a} dx R_s(x)\cdots. \tag{8.3}$$

The restriction on the b'_m integration is such that a reflection of type s occurs. On the right-hand side in (8.3), we integrate over the width of the double path and sum over the two "permutations" $\rightrightarrows$ and $\rightleftarrows$. These two operations can only be carried out after the weight of the double path is known, and thus are postponed to Sec. 8A.2. The remaining factor $R_s(x)$ [defined by (8.3)] can then be combined with the 1, 3 matrix element of (8.2), without the b'_m integration, to give the weight $W_s^A(x)$ of a reflector as shown in Fig. 9(b), representing the sum of all chains with overlapping trees leading to a type s reflection. To $\mathcal{O}(n^2)$ and in the limit $\epsilon \to 0$, one finds

$$W_s^A(x) = R_s(x)an^2\int \frac{d\mathbf{K}}{(2\pi)^2}(2\Omega_{13} - \Omega_{12} - \Omega_{14}), \tag{8.4}$$

or by (C4)

$$W_s^A(x) = \tfrac{1}{2}R_s(x)an^2. \tag{8.5}$$

A reflector of the second type arises from the four possible arrangements when the first collision is with the upper half of tree m, and the second collision is with the lower half of tree m (or vice versa; see Fig. 10). The type d reflector has the property that the two paths are connected with *different* halves, each of width a [see Fig. 10(b)]. In this case, (8.3) is replaced by

$$\int_{-a}^{a} db_m \int^{(d)} db'_m \cdots \to \sum_P \left\{\int_0^a dx x\cdots + \int_a^{2a} dx(2a-x)\cdots\right\} \equiv \sum_P \int_0^{2a} dx R_d(x)\cdots. \tag{8.6}$$

Combining the factor $R_d(x)$ as defined by (8.6) with

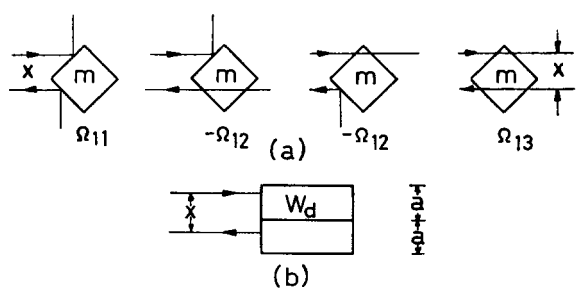


FIG. 10. The reflector of type d.

the 1, 3 matrix element of (8.2), we find

$$W_d^A(x) = R_d(x)an^2\int\frac{d\mathbf{K}}{(2\pi)^2}(\Omega_{11} + \Omega_{13} - 2\Omega_{12}), \quad (8.7)$$

in analogy with (8.4). By (C4)

$$W_d^A(x) = (\tfrac{1}{2} - \pi^{-1})R_d(x)an^2, \quad (8.8)$$

so that by (8.5) and (8.8), all types of collisions with the final tree on the double path, plus all reflecting chains, have been taken into account.

2. *The Double Path*

What remains to be considered are all possible sequences of collisions along the double path. In this case it is convenient to use (5.9) as a starting point, rather than to perform the partial resummation of Sec. 5 separately. On the basis of (5.9) and the discussion in Sec. 8A.1, one can write the total contribution to γ from Class III events as

$$\gamma_{\mathrm{III}} = \frac{a}{v^2}\lim_{\epsilon\to 0}\mathbf{v}\cdot\int_{-a}^{a}db_1\sum_P\int_0^{2a}dx\int_0^{\infty}dL \times\sum_{k_p=0}^{\infty}n^{k_p+1}\sum_{\mathrm{r,v}}'(-1)^{V+1}\int\cdots\int'd\mathbf{Q}_{1_p}\cdots d\mathbf{Q}_{k_p} \times\exp(-\epsilon 2L/v)\{W_s(x) + W_d(x)\}\mathbf{v}_f. \quad (8.9)$$

Here $\sum_{\mathrm{r,v}}'$ sums over all types of collisions with tree 1 and all allowed combinations of the k_p collisions along the double path of total length L, such that the double path is not destroyed. The integrations over $\{\mathbf{Q}_{1_p}\cdots\mathbf{Q}_{k_p}\}$ are restricted accordingly.

Consider first the collisions with tree 1. The two possible configurations are shown in Fig. 11. (The initial velocity can be chosen arbitrarily.) The real collisions consistent with a given double path are drawn with full lines, the virtual ones with broken lines. Since only the dot product contributes to γ, we are left with the combinations real-real and virtual-virtual. These two terms give rise to a factor 2 in Fig. 11(a), whereas they cancel in Fig. 11(b). Consequently, we need only consider double paths of width $0 \leq x \leq a$. Furthermore, with the configuration in Fig. 11(a), $\sum_P$ simply gives rise to a factor 2, and at fixed

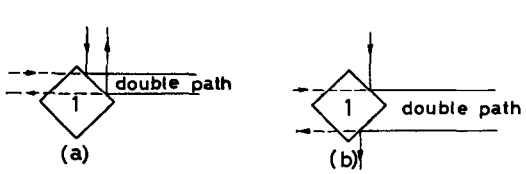


FIG. 11. Collisions with tree 1.

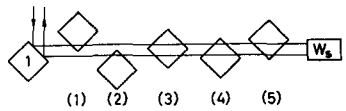


FIG. 12. The double path.

x, the b_1 integration yields the factor $2(a - x)$, so that (8.9) reduces to

$$\gamma_{\mathrm{III}} = 8a\lim_{\epsilon\to 0}\int_0^a dx(a - x)[W_s(x) + W_d(x)] \times\int_0^{\infty}dL\exp(-\epsilon 2L/v) \times\sum_{k=0}^{\infty}n^{k+1}\sum_{\mathrm{r,v}}'(-1)^V\int\cdots\int'd\mathbf{Q}_1\cdots d\mathbf{Q}_k, \quad (8.10)$$

where we have dropped the subscript p on the labels, since the collisions with tree 1 have now been taken care of, and all that remain to be considered are those along the double path itself.

To simplify the picture one can clearly deform the double path into a straight one of total length L, without changing the lengths of the paths. On the basis of (5.7), then, five types of collisions along the double path are possible (see Fig. 12):

(1) Uncorrelated virtual collision on the upper path.
(2) The same on the lower path.
(3) (Correlated) virtual collisions on both paths with the same tree.
(4) Real collisions on both paths with the upper half of a given tree.
(5) The same with the lower half.

(Note that with a type 1 or type 2 collision the tree can be penetrated by both paths, although only one of the encounters is taken explicitly into account.[20])

The sum over all possible collisions (at positions $l_1, l_2\cdots$) along a double path of length L and width x ($\leq a$), is to lowest order [cf. (5.15)] given as

$$\sum_{k=0}^{\infty}\sum_{\left(\sum_{i=1}^{5}k_i=k\right)}'\frac{k!}{k_1!\cdots k_5!} \times\int_0^L dl_1\int_{l_1}^L dl_2\cdots\int_{l_{k-1}}^L dl_k\,[-2an]^{k_1}[-2an]^{k_2} \times[(2a - x)n]^{k_3}[(a - x)n]^{k_4}[(a - x)n]^{k_5} = \exp(-3nxL). \quad (8.11)$$

This "superdamping" can be interpreted as the

probability that a double path of width x is not split up over a total length L.[26]

Inserting (8.11) in (8.10), one can perform the L-integration and go to the limit $\epsilon \to 0$. With $x = a\xi$, the final integrals yielding the two Class III contributions with overlapping trees read [use (8.3)–(8.8)]:

$$\gamma_s^A = \rho^2 \frac{8}{3} \int_0^1 d\xi \xi^{-1}(1 - \xi)^2, \tag{8.12}$$

$$\gamma_d^A = \rho^2 \left(\frac{1}{2} - \frac{1}{\pi}\right) \frac{8}{3} \int_0^1 d\xi (1 - \xi). \tag{8.13}$$

The first integral diverges logarithmically as $\xi \to 0$; however, the second one is finite:

$$\gamma_d^A = \tfrac{2}{3}(1 - 2\pi^{-1})\rho^2. \tag{8.14}$$

We postpone a discussion of this divergence to Sec. 9, and first compute the corresponding quantities with nonoverlapping trees.

B. Nonoverlapping Trees

It follows from the discussion in Sec. 6B that the only terms in the sum (8.1), which depend on whether or not the trees are allowed to overlap, are those with two real collisions between the encounters with tree m. From the same discussion it is also clear that the double-path calculation to (8.11) remains valid to $\mathcal{O}(\rho^2)$ in Case B. One therefore calculates the Class III contribution with nonoverlapping trees in two steps. First, the terms representing the exceptional events are subtracted from the operator Ω, and the corresponding reduced weights W^B and contributions to γ are computed. Second, the case of exactly two real collisions in the reflecting chain is treated separately.

(1) The reduced weights are easily found to be [compare (8.5) and (8.8)]:

$$W_s^B(x) = 0, \tag{8.15}$$

$$W_d^B(x) = -(\pi^{-1} - \tfrac{1}{4})R_d(x)an^2. \tag{8.16}$$

Two remarks are in order. First, the crucial fact that the entire weight of the s reflector in Case A (8.5) stems from that part of the Ω operator which corresponds to two intermediate real collisions, is a consequence of (8.15); it can also be checked directly. Next, from their definitions it follows that the weights W can be of either sign, depending on the prevalent type of collisions. W_d^A is positive, whereas W_d^B is negative.

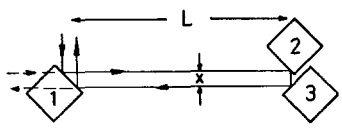


FIG. 13. The exceptional event in Case B.

The contributions to γ_{III}^B corresponding to (8.15) and (8.16) are

$$\gamma_s^B = 0, \tag{8.17}$$

$$\gamma_d^B = -\tfrac{1}{3}(4\pi^{-1} - 1)\rho^2. \tag{8.18}$$

(2) Finally, one has to consider the situation shown in Fig. 13. The weight associated with the two nonoverlapping trees 2 and 3, serving as a reflector in this case, is found by integrating over $\mathbf{Q}_2$, $\mathbf{Q}_3$ at constant x, L. Thus,

$$W_{r2}^B(x) = (ax - x^2/4)n^2. \tag{8.19}$$

With the weight (8.19) we repeat the argument of Sec. 8A.2 and arrive at the following contribution to γ:

$$\gamma_{r2}^B = \rho^2 \frac{8}{3} \int_0^1 d\xi \xi^{-1}(1 - \xi)(\xi - \xi^2/4)$$
$$= \tfrac{11}{9}\rho^2. \tag{8.20}$$

This result, however, contains the contribution from the ring R_3 in addition to those from the Class III events, since the sum (8.11) also includes the case of all collisions along the double path being uncorrelated virtual ones. The ring part of (8.20) [which is found by replacing $\frac{1}{3}\xi^{-1}$ by $\frac{1}{4}$ in (8.20)] is given by (6.5), and must be subtracted when we are interested in the Class III contribution only. Altogether, one finds to $\mathcal{O}(\rho^2)$ that

$$\gamma_{\text{III}}^B = \gamma_s^B + \gamma_d^B + \gamma_{r2}^B - \gamma_{R_3}^B,$$
$$\gamma_{\text{III}}^B = (\tfrac{91}{72} - \tfrac{4}{3}\pi^{-1})\rho^2. \tag{8.21}$$

Thus the retracing events give rise to a *finite* contribution to γ with nonoverlapping trees, whereas with overlapping ones the corresponding integral [(8.12) and (8.13)] diverges logarithmically. The physical meaning of this qualitative difference is discussed in Sec. 9.

9. DISCUSSION

The results of the preceding sections are summarized in Table I. Below we first discuss the "physical" reason for the divergence of γ_{III}^A. Subsequently the long time behavior of the diffusion process in the two cases A and B is considered on the basis of the results of this paper. We close with some final remarks.

[26] An immediate "physical" argument (*not* based on the binary-collision expansion) supports this interpretation. There are three ways of splitting up a double path: (1) the first path is broken by a collision, but not the second; (2) vice versa; (3) both paths are broken, but on different sides of the tree. All three possibilities contribute xn to an "absorption coefficient" and hence (8.11) follows.

TABLE I. Contributions to $\gamma = av/2D$.

	Name	Derived in Sec.	Case A: overlapping trees	Case B: nonoverlapping trees
$\mathcal{O}(\rho)$	γ_1	4	2ρ	2ρ
$\mathcal{O}(\rho^2)$	γ_2	4	$(\pi^2/9)\rho^2 = 1.097\rho^2$	$(\pi^2/9 + 4)\rho^2 = 5.097\rho^2$
	γ_{I}	6	$(4/\pi)\rho^2 = 1.273\rho^2$	$(4/\pi - 17/24)\rho^2 = 0.565\rho^2$
	γ_{II}	7	$\lesssim 0.1\rho^2$	$\lesssim 0.1\rho^2$
	γ_{III}	8	$\ln\infty \cdot \rho^2$	$(91/72 - 4/3\pi)\rho^2 = 0.839\rho^2$
Sum $\mathcal{O}(\rho^2)$			$\ln\infty \cdot \rho^2$	$(\pi^2/9 + 8/3\pi + 41/9)\rho^2 + \gamma_{\mathrm{II}}$ $= 6.501\rho^2 + (\lesssim 0.1\rho^2)$

A. The Divergence of γ^A_{III}

To get a better "physical" understanding of the divergence in the resummed theory discovered in Sec. 8, consider a particle moving around in the "forest" of trees. During a time interval of order $v^{-1}a^{-3}n^{-2}$, the particle will probably encounter two trees very close together. Given that they are close, the probability is of $\mathcal{O}(1)$ that they are arranged in such a way that the particle is reflected more or less precisely onto its previous path (see Fig. 14).

The probability that the two paths, distance x apart, stay together over a total length L was, in Sec. 8, found to be $\exp(-3nxL)$. Whether this type of event will *qualitatively* alter the nature of the diffusion process depends on the outcome of the competition between the small phase-space associated with the reflector, and the long "memory" associated with the narrow double path. The memory associated with this excessive back scattering extends over a time of the order $(3nxv)^{-1}$ which tends to infinity as $x \to 0$. With nonoverlapping trees, the phase-space associated with the reflector [$\sim$ the weight W^B_{r2} of (8.19)] goes down linearly with x when $x \to 0$. With overlapping trees, however, this phase space is independent of x. The result of the competition in the two cases can be seen from (8.12) and (8.20). The corresponding contribution to the inverse diffusion coefficient is well behaved in Case B, while it diverges in Case A.

B. The Long-Time Behavior

To understand the qualitative significance of this divergence on the long-time behavior of the diffusion process described by the fundamental equation (2.2), we go back to (8.10) and keep ϵ small but finite. With finite ϵ, Eq. (8.11) has to be replaced by

$$\exp(-3nxL) \to \exp(-3nxL - \epsilon \cdot 2L/v), \quad (9.1)$$

so that it follows asymptotically from (8.10), using (8.5) and (8.8), that

$$\gamma^A_{\mathrm{III}} \sim -(8\rho^2/3)\ln(2\epsilon/3nav) + \text{finite terms of } \mathcal{O}(\rho^2). \quad (9.2)$$

With (2.2), (2.3), (3.1), and (3.12), it follows from (9.2) that with overlapping trees the mean-square displacement asymptotically obeys the relation

$$d\Delta^A(t)/dt \sim 2av[2\rho + (\tfrac{8}{3}\rho^2)\ln(\tfrac{3}{2}navt) + \text{terms of } \mathcal{O}(\rho^2) \text{ finite in the limit } t \to \infty]^{-1}. \quad (9.3)$$

The dominant behavior for very long times follows immediately as

$$\Delta^A(t) \sim (3av/4\rho^2)t/\ln t. \quad (9.4)$$

The asymptotic growth of the mean-square displacement is therefore *slower* in case A than in a normal diffusion process. However, the anomalies become substantial only after times T of the order [see (9.3)]:

$$T \sim (\tfrac{3}{2}nav)\exp(\tfrac{3}{4}\rho) = \tfrac{4}{3}\tau\exp(\tfrac{3}{4}\rho), \quad (9.5)$$

where τ is the mean free time which follows from the Boltzmann equation.

The above discussion was based on the divergent contribution of $\mathcal{O}(\rho^2)$ to γ. The precise form of the asymptotic behavior (9.4) and the corresponding time scale (9.5) would probably have to be modified if terms of higher orders in ρ were taken into account. The *qualitative* point, however, that the diffusion process is slower in the anomalous Case A than in a normal diffusion process, is seen to depend on the *existence* of the divergence only, not on its precise nature.

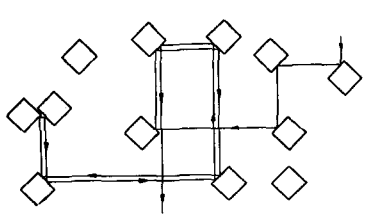


FIG. 14. Retracing event with real collisions only.

In contrast to Case A, Case B represents perfect normality up to $\mathcal{O}(\rho^2)$. The growth of the mean-square displacement is therefore

$$\Delta^B(t) \sim 4D^B t,$$

where the contributions to $av/2D^B$ of orders ρ and ρ^2 are given in Table I.

C. Concluding Remarks

(A) Within the framework of our present calculation to $\mathcal{O}(\rho^2)$, it makes a qualitative difference as far as the diffusion process is concerned, whether the trees are allowed to overlap or not. It is an open question, however, whether this difference persists to higher orders, since:

(i) We have no *proof* that in a complete theory, γ^A does not exist in the limit $\epsilon \to 0$, and that the divergence in γ^A of $\mathcal{O}(\rho^2)$ is not an artifact of the expansion in powers of ρ.

(ii) We do not know whether γ^B remains finite in higher orders than $\mathcal{O}(\rho^2)$.

From the computer study of the wind-tree model by Wood and Lado,[13] however, there are strong indications that our results are indeed qualitatively correct not only for very small ρ.

(B) The questions above are intimately connected with the following: What is the precise asymptotic behavior of $\Delta^A(t)$ for long times when higher orders in ρ are taken into account?

(C) It is worth noting that at very high densities a qualitative difference of a second kind between the two cases becomes apparent. With overlapping trees there is always a nonvanishing probability that a moving particle started in a random position between the trees is trapped in a finite volume. In such a case the displacement can never exceed finite bounds and nothing resembling a diffusive process is ever realized.[27]

(D) We would like to stress that both types of divergences discussed in this paper point to a deeper physical insight into the problem of kinetic theory.

(i) The divergences discussed in Sec. 5 seem to make a non-Markovian description of the *kinetic* stage (in the Bogoliubov scheme, $t \lesssim$ mean free time) necessary. This point has not been discussed in the present paper.

(ii) It follows from (9.2) and (9.4) that the divergences in the resummed theory in Case A qualitatively affect the hydrodynamical stage ($t \gg$ mean free time). The diffusion equation is clearly not valid in Case A.

The divergences of the first type are essentially model-independent; therefore, the conclusions drawn from them have a bearing on the general case. The divergences of the second type seem to be closely associated with special features of our model; their importance to kinetic theory in general is therefore less clear.

(E) In this connection, however, we point out that it should be interesting to see if a microscopic interpretation in some way analogous to that of (9.4) can be given to the vanishing of the mutual diffusion coefficient in a binary mixture at the critical point.

[27] For low densities the probability of starting in a trapped position is clearly of $\mathcal{O}(\rho^4)$ in our model; and in analogy with the percolation problem, it is tempting to conjecture that this probability will increase with ρ and be unity above a certain critical density ρ_c, which then defines the end of the diffusive regime.

ACKNOWLEDGMENTS

We are very much indebted to T. S. Nilsen, lic. techn., for having carried out the computations leading to (7.4). It is also a pleasure to thank Professor Z. W. Salsburg and Dr. W. W. Wood for stimulating discussions. Finally, one of the authors (E. H. H.) would like to express his sincere gratitude to Professor M. Kac and Rockefeller University for the hospitality extended to him during the academic years 1965–67, when the major part of the present work was done.

APPENDIX A

In this appendix we first give some of the details in the derivation of (3.13), then go on to derive a formal density expansion of the operator K.

Inversion of Equation (3.3)

One first rewrites (3.3) as follows:

$$\hat{\mathbf{\Phi}}_1 = G_0[1 + B(\mathbf{v}, \epsilon, n)]\mathbf{v}, \tag{A1}$$

where

$$G_0\mathbf{v} = [\epsilon - \mathcal{K}(x)]^{-1}\mathbf{v} = \epsilon^{-1}\mathbf{v}, \tag{A2}$$

by the definition (3.4). It is easily seen that a direct expansion of the operator $B(\mathbf{v}, \epsilon, n)$ in powers of the density gives rise to divergent terms in the limit $\epsilon \to 0$. Following Zwanzig,[3] we therefore invert (A1) to get

$$\begin{aligned}\mathbf{v} &= [1 + B(\mathbf{v}, \epsilon, n)]^{-1}G_0^{-1}\hat{\mathbf{\Phi}}_1 \\ &\equiv [G_0^{-1} + K(\mathbf{v}, \epsilon, n)]\hat{\mathbf{\Phi}}_1,\end{aligned} \tag{A3}$$

and subsequently study the power series expansion of the operator K. By this procedure the most divergent terms in the expansion of B are removed. Although it is shown in Sec. 5 that the density expansion of K still contains divergences, we adopt it as a basis for resummations.

To express the diffusion coefficient in terms of K we first note that for reasons of symmetry we can

write[7]:

$$\hat{\Phi}(\mathbf{v}, \epsilon) = \mathbf{v}\hat{\phi}(\epsilon), \tag{A4}$$

and since the operators in Lorentz models act only on the *direction* of $\mathbf{v}$, as the absolute value v of $\mathbf{v}$ remains constant, (A3) gives

$$\mathbf{v} = \hat{\phi}(\epsilon)[G_0^{-1} + K]\mathbf{v}. \tag{A5}$$

Multiplying (A5) from the left by $\mathbf{v}$ and using (3.2) and (A4), one finds

$$D^{-1} = \lim_{\epsilon \to 0} 2v^{-4}(\epsilon v^2 + \mathbf{v} \cdot K\mathbf{v}), \tag{A6}$$

so that introduction of the dimensionless quantity $\gamma(\epsilon)$ defined by

$$\gamma(\epsilon) = av^{-3}\mathbf{v} \cdot K(\mathbf{v}, \epsilon, n)\mathbf{v} \tag{A7}$$

gives, finally,

$$D^{-1} = 2a^{-1}v^{-1} \lim_{\epsilon \to 0} \gamma(\epsilon). \tag{A8}$$

Formal Expansion of K

We are interested only in the first correction to the Boltzmann result for the diffusion coefficient. However, as a consequence of the above-mentioned divergences, one can only hope to get systematic results to this order in the density after a classification of the general term in K, and a resummation over certain classes of terms. The remainder of this appendix is consequently devoted to combinatorial aspects of the expansion problem. The goal is to derive a formal density expansion of the operator K in (A7), and thus of the inverse diffusion coefficient.

The starting point is the cluster expansion of the operator $G(1 \cdots N)$ in (3.3) (the arguments x and ϵ are henceforth suppressed, and the N position vectors of the scatterers are replaced by the set of numbers labeling them):

$$G(1 \cdots N) = \sum_{\{A\} \subset \{1, \cdots, N\}} U(\{A\}), \tag{A9}$$

where the sum goes over all subsets $\{A\}$ of the set $\{1, \cdots, N\}$, and the empty set is included [$G(\{0\}) \equiv G_0 = U_0$]. The inversion of (A9) reads

$$U(1 \cdots m) = \sum_{\{A\} \subset \{1, \cdots, m\}} (-1)^{m-\nu(A)} G(\{A\}), \tag{A10}$$

where $\nu(A)$ is the number of elements in the set $\{A\}$.

For $m = 1, 2$ this means that

$$\begin{aligned} U(1) &= G(1) - G_0, \\ U(12) &= G(12) - G(1) - G(2) + G_0. \end{aligned} \tag{A11}$$

Defining the reduced distribution functions

$$n^l g(\mathbf{r}, 1, \cdots, l) = \lim_{\substack{N,V \to \infty \\ N/V = n}} \frac{VN!}{(N-l)!} \int_V \cdots \int dQ^{N-l} \rho(\mathbf{r}, Q^N), \tag{A12}$$

one can write (3.3) as the formal power series

$$\hat{\Phi}_1 = \left[G_0 + \sum_{l=1}^{\infty} \frac{n^l}{l!} \int \cdots \int dQ^l \times g(\mathbf{r}, 1, \cdots, l) U(1, \cdots, l) \right] \mathbf{v}. \tag{A13}$$

This is not yet a *bona fide* series expansion, since:

(i) the terms blow up in the limit $\epsilon \to 0$;

(ii) the reduced distribution functions are, in general, complicated functions of the density.

However, one can still regard (A13) as a formal density expansion of the operator B in (A1) with coefficients B_l:

$$\begin{aligned} B &= \sum_{l=1}^{\infty} n^l B_l, \\ B_l &= \frac{G_0^{-1}}{l!} \int \cdots \int dQ^l g(\mathbf{r}, 1, \cdots, l) U(1, \cdots, l). \end{aligned} \tag{A14}$$

An expansion of the operator K in (A3),

$$K = \sum_{l=1}^{\infty} n^l K_l, \tag{A15}$$

which is formal in the same sense as (A14), is obtained from (A1) and (A3) by the requirement that

$$G_0^{-1} + \sum_{l=1}^{\infty} n^l K_l = \left[1 + \sum_{l=1}^{\infty} n^l B_l \right]^{-1} G_0^{-1} \tag{A16}$$

should be satisfied to every formal order in n. From (A16) one can derive[12] the following explicit expression for K_l in terms of the B_k's:

$$K_l = \sum_{n=1}^{l} (-1)^n \sum_{\{\alpha\}}{}' B_{\alpha_1} \cdots B_{\alpha_n} G_0^{-1}, \tag{A17}$$

$$\sum_{i=1}^{n} \alpha_i = l$$

where the primed sum is over all products of B_k's such that the sum of their subscripts (positive integers) equals l. For $l = 1, 2$, Eq. (A17) reads

$$K_1 = -B_1 G_0^{-1}, \tag{A18}$$

$$K_2 = -B_2 G_0^{-1} + B_1 B_1 G_0^{-1}, \tag{A19}$$

and using (A7) and (A14), one finds that the corresponding contributions, $\gamma_1(\epsilon)$, $\gamma_2(\epsilon)$, to $\gamma(\epsilon)$, can be written

$$\gamma_1(\epsilon) = -nav^{-3}\epsilon \mathbf{v} \cdot \int d\mathbf{Q}_1 g(\mathbf{r}, 1) C(1) \mathbf{v}, \tag{A20}$$

$$\begin{aligned} \gamma_2(\epsilon) = -\tfrac{1}{2} n^2 av^{-3} \epsilon \mathbf{v} \cdot \iint d\mathbf{Q}_1 \, d\mathbf{Q}_2 [& g(\mathbf{r}, 1, 2) U G_0^{-1}(12) \\ & - g(\mathbf{r}, 1) C(1) g(\mathbf{r}, 2) C(2) \\ & - g(\mathbf{r}, 2) C(2) g(\mathbf{r}, 1) C(1)] \mathbf{v}, \end{aligned} \tag{A21}$$

where we have introduced $C(i) \equiv U(i) G_0^{-1}$.

APPENDIX B: THE BINARY COLLISION EXPANSION

In this appendix, some of the steps leading to (5.2) and (5.3) are sketched.

The C Operator

The binary-collision operator $C(1)$ is defined by[28]

$$C(1) = U(1)G_0^{-1} = [G(1) - G_0]G_0^{-1}. \quad \text{(B1)}$$

By (3.5), for any function $F(\mathbf{v}, \mathbf{r})$ one has

$$e^{t\mathcal{H}(x)}F(\mathbf{v}, \mathbf{r}) = F(\mathbf{v}, \mathbf{r} + \mathbf{v}t)$$

so that, by (3.4),

$$G_0^{-1}F(\mathbf{v}, \mathbf{r}) = [\epsilon - \mathcal{H}(x)]F(\mathbf{v}, \mathbf{r}) = \left(\epsilon - v \cdot \frac{\partial}{\partial \mathbf{r}}\right)F(\mathbf{v}, \mathbf{r}). \quad \text{(B2)}$$

The result of $[G(1) - G_0]$ acting on $G_0^{-1}F(\mathbf{v}, \mathbf{r})$ is again given by (3.4) as

$$C(1)F(\mathbf{v}, \mathbf{r}) = \int_{\tau_1}^{\infty} dt e^{-\epsilon t}\left\{\left[\epsilon - \mathbf{v}_1 \cdot \frac{\partial}{\partial \mathbf{R}_1}\right]F(\mathbf{v}_1, \mathbf{R}_1) - \left[\epsilon - \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{R}}\right]F(\mathbf{v}, \mathbf{R})\right\}, \quad \text{(B3)}$$

where $\mathbf{v}_1$, τ_1 are defined as in (5.2), and where

$$\mathbf{R}_1 = \mathbf{r} + \mathbf{v}\tau_1 + \mathbf{v}_1(t - \tau_1); \quad \mathbf{R} = \mathbf{r} + \mathbf{v}t.$$

The integrand in (B3) can be rewritten to give

$$\begin{aligned} C(1)F(\mathbf{v}, \mathbf{r}) &= \int_{\tau_1}^{\infty} dt e^{-\epsilon t}\left(\epsilon - \frac{\partial}{\partial t}\right) \\ &\quad \times \{F[\mathbf{v}_1, \mathbf{r} + \mathbf{v}\tau_1 + \mathbf{v}_1(t - \tau_1)] - F(\mathbf{v}, \mathbf{r} + \mathbf{v}t)\} \\ &= \int_{\tau_1}^{\infty} dt\left(-\frac{\partial}{\partial t}\right)e^{-\epsilon t}\{\ \} \\ &= e^{-\epsilon\tau_1}\{F(\mathbf{v}_1, \mathbf{r} + \mathbf{v}\tau_1) - F(\mathbf{v}, \mathbf{r} + \mathbf{v}\tau_1)\}, \end{aligned} \quad \text{(B4)}$$

which is precisely (5.2).

The Expansion

The steps leading to (5.3) can be summarized as follows (for details, see Ref. 12):

First, the G operator of (3.4) is expanded in terms of the C operators:

$$G(1 \cdots m) = \left[1 + \sum_{p=1}^{\infty} \sum_{\substack{\{i\} \\ (1.,2.)}}' C(i_1) \cdots C(i_p)\right]G_0, \quad \text{(B5)}$$

where the primed sum is over all products of p binary-collision operators subject to restrictions (1., 2.):

(1) All labels $i_1 \cdots i_p$ belong to the set $\{1 \cdots m\}$;

(2) $i_{n+1} \neq i_n$, $n = 1, \cdots, p - 1$.

Second, the expansion (B5) is inserted into (A10) to yield

$$U(1 \cdots m) = G_0 \sum_{p=m}^{\infty} \sum_{\substack{\{i\} \\ (1.2.3.)}}' C(i_1) \cdots C(i_p), \quad \text{(B6)}$$

where the additional restriction on the primed sum reads:

(3) All labels in the set $\{1, \cdots, m\}$ occur at least once in the set $\{i_1, \cdots, i_p\}$.

Equation (B6) constitutes an expansion of the "dynamical" part of the B_l in (A14), since

$$B_l^{\mathrm{d}} = \frac{G_0^{-1}}{l!} \int \cdots \int dQ^l G U(1, \cdots, l). \quad \text{(B7)}$$

The third and final step is to go from the B_l^{d}'s to the K_l^{d}'s by (A15). Using (B6) one obtains (5.3), where "irreducible" is defined as follows:

A sequence is called irreducible if it is impossible to insert a partition at any point in the sequence in such a way that there is an element that occurs on one side of the partition only. Thus the sequence 12123 is reducible since it can be partitioned (1212/3), while the sequence 1212 is irreducible.

APPENDIX C. FOURIER TRANSFORMS

In this appendix we shall rederive the result (6.3) by a technique which is useful whenever intermediate chains of uncorrelated collisions should be summed over. The case of rings will be treated as an example, but only slight modifications are necessary to apply the same method to similar problems encountered in Secs. 7 and 8.

In the ring case, the starting point for this method is (5.3), and denoting the ring part of K_l^{d} by R_l one can write

$$n^l R_l \mathbf{v} = -\epsilon n^l \int \cdots \int dQ^l C(1) \cdots C(l)C(1)\mathbf{v}. \quad \text{(C1)}$$

Since the integral is of the convolution type, it is natural to introduce Fourier transforms. It is shown in Ref. 12 that by performing a scaled transform on (C1), and subsequently summing over l from 3 to ∞, one arrives at (in the limit $\epsilon \to 0$):

$$\sum_{l=3}^{\infty} n^l R_l \mathbf{v} = -avn^2 \int_{-a}^{a}\int db_1\, db_1' \iint_{-\infty}^{\infty} \frac{dK_1\, dK_2}{(2\pi)^2} e^{i\mathbf{K}\cdot\sigma_1 an}\bar{C}(1) \times \Omega(K_1, K_2)e^{-i\mathbf{K}\cdot\sigma_1' an}\bar{C}'(1)\mathbf{v}, \quad \text{(C2)}$$

[28] *Note Added in Proof:* In Ref. 12, $C(1)$ as given by (B4) is erroneously identified with $G_0^{-1}U(1)$ instead of $U(1)G_0^{-1}$. No results are affected by this error, but in a number of equations G_0 and G_0^{-1} should be commuted with the other operators present. We are indebted to Dr. M. H. Ernst (private communication) for having pointed out this error to us.

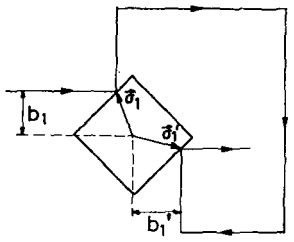


FIG. 15. The initial and final collisions in a ring event.

where σ_1 is the vector from the center of tree 1 to the point on the edge where the first collision takes place, and b_1 is the corresponding impact parameter. The primes are used throughout to distinguish the variables associated with the second collision with tree 1 (see Fig. 15). The $\bar{C}$ operator is defined by [see (5.2)]:

$$C(1) \equiv e^{-\epsilon r_1}\bar{C}(1). \tag{C3}$$

The dimensionless operator $\Omega(K_1, K_2)$ can be given the following 4×4 matrix representation with the four allowed velocity directions $\mathbf{v}_1, \cdots, \mathbf{v}_4$ ($\mathbf{v}_3 = -\mathbf{v}_1$, $\mathbf{v}_4 = -\mathbf{v}_2$) labeled in counterclockwise manner as a basis:

$$\begin{aligned} \Omega_{11} &= A^{-1}(2 + iK_1)(4 + K_2^2), \\ \Omega_{12} &= A^{-1}4(2 + iK_2), \\ \Omega_{13} &= A^{-1}(2 - iK_1)(4 + K_2^2), \\ \Omega_{14} &= A^{-1}4(2 - iK_1), \end{aligned} \tag{C4}$$

with

$$A = [K_1^2K_2^2 + 4(K_1^2 + K_2^2)][2 - iK_1][4 + K_2^2],$$

and with

$$\mathbf{K} = K_1v^{-1}\mathbf{v}_1 + K_2v^{-1}\mathbf{v}_2.$$

The remaining matrix elements are obtained by cyclic permutations on the four indices (i.e., $K_1 \to K_2$, $K_2 \to K_3 = -K_1$, etc.).

From (3.13) and (6.3) it follows that

$$\gamma_I^A = av^{-3}\mathbf{v} \cdot \sum_{l=3}^{\infty} n^l R_l \mathbf{v}, \tag{C5}$$

and to extract the part of $\mathcal{O}(\rho^2)$ from (C5), (C2), we first observe that γ_I^A depends on n in two ways only, through the factor n^2 in front, and through the factor n in the exponential. To $\mathcal{O}(\rho^2)$ we can therefore put the exponential equal to unity, provided that the resulting integrals are finite.

The next step is to observe that whatever the results of the preceding operators,

$$\int_{-a}^{a} db_1'\bar{C}'(1)\mathbf{v} = \int_{-a}^{a} db_1'\bar{C}'^{v}(1)\mathbf{v} = -2a\mathbf{v}. \tag{C6}$$

Since the exponentials in (C2) have been put equal to unity, one can use (C6) and an equivalent argument on $\bar{C}(1)$ to obtain

$$\gamma_I^A = -\pi^{-2}a^4n^2v^{-2}\mathbf{v} \cdot \iint_{-\infty}^{\infty} dK_1\, dK_2 \Omega\mathbf{v}. \tag{C7}$$

By (C4), Eq. (C7) reduces to

$$\begin{aligned} \gamma_I^A &= \pi^{-2}a^4n^2 \iint_{-\infty}^{\infty} dK_1\, dK_2(\Omega_{13} - \Omega_{11}) \\ &= (4/\pi)\rho^2, \end{aligned} \tag{C8}$$

in agreement with (6.3).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Functional Integration Theory for Incompressible Fluid Turbulence. II

GERALD ROSEN, J. A. OKOLOWSKI, AND GENE ECKSTUT
Drexel Institute of Technology, Philadelphia, Pennsylvania

(Received 25 June 1968)

Two iterative solutional procedures are reported for the nonlinear integro-differential dynamical equation obtained previously by the stationary functionality method and supplemented here by a subsidiary dynamical condition. Rapidly convergent for grid Reynolds numbers between about 100 and 300, both methods yield general expressions for the two-point equal-time velocity-correlation tensor in approximate agreement with experiment for the initial period of decay.

1. INTRODUCTION

A purely deductive approximation theory for incompressible fluid turbulence has been reported recently,[1,2] a theory based exclusively on the Hopf characteristic functional space–time formulation and free of any additive statistical postulate. With the use of explicit functional integration techniques, a nonlinear integro-differential dynamical equation for the two-point velocity-correlation tensor was obtained by evaluating $\Lambda_{\mu\nu}(x', x'')$, the real symmetric solenoidal two-point tensorial stationary functionality, with a zero-mean velocity-field Gaussian approximation for the characteristic functional, and by equating the resulting expression for $\Lambda_{\mu\nu}(x', x'')$ to zero. The nonlinear integro-differential dynamical equation so obtained was noted to be identical to a specific two-point Navier–Stokes expectation-value equation with a zero-mean velocity-field probability distribution such that the fourth-order velocity-field product expectation values are related to lower-order product expectation values in the same way as for a zero-mean Gaussian probability distribution. Specialized for isotropic homogeneous turbulence, the nonlinear integro-differential dynamical equation was shown to take the form

$$\left[\frac{\partial^2}{\partial t'\partial t''} - \nu\left(\frac{\partial}{\partial t'} + \frac{\partial}{\partial t''}\right)\frac{\partial^2}{\partial r^2} + \nu^2\frac{\partial^4}{\partial r^4}\right]\frac{\partial^2(r\xi)}{\partial r^2} + 2\eta - 6r\int_r^\infty \eta(s, t', t'')s^{-2}\,ds = 0, \quad (1.1)$$

with

$$\eta = \eta(r, t', t'') \equiv \frac{\partial\xi}{\partial r}\left[\frac{\partial^2(\nabla^2\xi)}{\partial r^2} - r^{-1}\frac{\partial(\nabla^2\xi)}{\partial r}\right], \quad (1.2)$$

the two-point velocity-correlation tensor being expressed in terms of the scalar function

$$\xi = \xi(r, t', t'') = \xi(r, t'', t')$$

(regarded as an even function of r) by

$$\langle u_\mu(x')u_\nu(x'')\rangle \equiv f_{\mu\nu}(x', x'') = \delta_{\mu\nu}\nabla^2\xi - \xi_{,\mu\nu}, \quad (1.3)$$

where

$$r \equiv (r_\alpha r_\alpha)^{\frac{1}{2}}, \quad r_\mu \equiv x'_\mu - x''_\mu,$$

and

$$\xi_{,\mu} \equiv \frac{\partial\xi}{\partial r_\mu} = \frac{r_\mu}{r}\frac{\partial\xi}{\partial r}, \quad \nabla^2 \equiv \frac{\partial^2}{\partial r_\alpha \partial r_\alpha}. \quad (1.4)$$

In the present paper we supplement the dynamical equation (1.1) with a subsidiary dynamical condition of the form

$$\left(\frac{\partial^2\xi}{\partial t'\partial t'} - 2\frac{\partial^2\xi}{\partial t'\partial t''} + \frac{\partial^2\xi}{\partial t''\partial t''}\right)_{t'=t''} = 0, \quad (1.5)$$

which implies that

$$\xi = \xi(r, t', t'') = \bar{\xi}(r, t) + (t' - t'')^2\zeta(r, t', t''), \quad (1.6)$$

where $t \equiv \frac{1}{2}(t' + t'')$, $\bar{\xi}(r, t)$ is an arbitrary function, and $\zeta(r, t', t'')$ vanishes for $t' = t''$, but is otherwise an arbitrary function. Observe that if (1.5) is satisfied as an initial condition at $t' = t'' = 0$, then Eq. (1.5) can be expected to hold for all $t' = t'' > 0$ because this subsidiary dynamical condition is compatible with Eq. (1.1) for all $t', t'' > 0$. By substituting (1.6) into (1.1), putting $t' = t''$, and using the condition $\zeta(r, t, t) = 0$, we find the associated dynamical equation for $\bar{\xi} = \bar{\xi}(r, t)$:

$$\left[\frac{1}{4}\frac{\partial^2}{\partial t^2} - \nu\frac{\partial}{\partial t}\frac{\partial^2}{\partial r^2} + \nu^2\frac{\partial^4}{\partial r^4}\right]\frac{\partial^2(r\bar{\xi})}{\partial r^2} + 2\bar{\eta} - 6r\int_r^\infty \bar{\eta}(s, t)s^{-2}\,ds = 0, \quad (1.7)$$

with

$$\bar{\eta} = \bar{\eta}(r, t) \equiv \frac{\partial\bar{\xi}}{\partial r}\left(\frac{\partial^2(\nabla^2\bar{\xi})}{\partial r^2} - r^{-1}\frac{\partial(\nabla^2\bar{\xi})}{\partial r}\right). \quad (1.8)$$

Note that the two-point equal-time velocity-correlation tensor

$$f_{\mu\nu}(x', x'')\big|_{t'=t''=t} = \delta_{\mu\nu}\nabla^2\bar{\xi} - \bar{\xi}_{,\mu\nu} \quad (1.9)$$

follows from a solution to Eq. (1.7). The purpose of the present paper is to develop systematic approximation methods for obtaining solutions to the nonlinear integro-differential dynamical equation (1.7).

[1] G. Rosen, Phys. Letters **25A**, 644 (1967).
[2] G. Rosen, Phys. Fluids **10**, 2614 (1967).

Two iterative solutional procedures are given for Eq. (1.7) in the following sections, the first procedure being based on the similitude invariance of Eq. (1.7) and the second on the quasilinear differential structure of Eq. (1.7).

2. APPROXIMATE SIMILITUDE SOLUTION OF EQ. (1.7)

Let us assume that $\bar{\xi}(r, t)$ has the form

$$\bar{\xi}(r, t) = \nu^2\omega(\alpha), \tag{2.1}$$

where $\alpha \equiv r/(\nu t)^{\frac{1}{2}}$. Then Eq. (1.7) becomes

$$\left(\frac{3}{4} + \frac{\alpha}{4}\frac{d}{d\alpha} + \frac{d^2}{d\alpha^2}\right)\left(\frac{1}{4} + \frac{\alpha}{4}\frac{d}{d\alpha} + \frac{d^2}{d\alpha^2}\right)\frac{d^2}{d\alpha^2}\alpha\omega(\alpha)$$
$$= \frac{2}{\alpha}\frac{d}{d\alpha}\alpha^3\int_\alpha^\infty \beta^{-5}\omega'(\beta)[6\omega'(\beta) - 6\beta\omega''(\beta)$$
$$+ \beta^2\omega'''(\beta) + \beta^3\omega''''(\beta)]\,d\beta \equiv \Gamma(\alpha). \tag{2.2}$$

Introducing the differential operator

$$\Lambda \equiv \frac{\alpha}{4}\frac{d}{d\alpha} + \frac{d^2}{d\alpha^2}, \tag{2.3}$$

it is readily verified that the solutions of the eigenvalue equation $\Lambda f_n(\alpha) = \lambda_n f_n(\alpha)$ for eigenfunctions $f_n(\alpha)$, such that $f_n(\pm\infty) = 0$, are given by

$$f_n(\alpha) = H_n\left(\frac{\alpha}{\sqrt{8}}\right)e^{-\alpha^2/8}, \quad n = 0, 1, 2, \cdots, \tag{2.4}$$

where $H_n(\alpha/\sqrt{8})$ is the nth Hermite polynomial expressed as a function of $\alpha/\sqrt{8}$, and the eigenvalues of Λ are

$$\lambda_n = -\tfrac{1}{4}(n + 1). \tag{2.5}$$

Notice that $\omega(\alpha)$ must be an even function of α since $\bar{\xi}(r, t)$ is regarded as an even function of r; from this it follows that $\Gamma(\alpha)$ must be an odd function of α.

The right side of (2.2) is now expanded in terms of the odd eigenfunctions of Λ as

$$\Gamma(\alpha) = \sum_{n=0}^{\infty} A_n f_{2n+1}(\alpha) = \sum_{n=0}^{\infty} A_n H_{2n+1}\left(\frac{\alpha}{\sqrt{8}}\right)e^{-\alpha^2/8}. \tag{2.6}$$

The A_n's are computed by using the orthogonality properties of the Hermite polynomials, and we obtain

$$A_n = \frac{1}{2^{2n+1}(2n + 1)!\,(2\pi)^{\frac{1}{2}}}\int_0^\infty \Gamma(\alpha)H_{2n+1}\left(\frac{\alpha}{\sqrt{8}}\right)d\alpha. \tag{2.7}$$

By putting (2.6) into (2.2) and integrating, we find

$$\frac{d^2}{d\alpha^2}\alpha\omega(\alpha) = \sum_{n=0}^{\infty}\frac{A_n H_{2n+1}(\alpha/\sqrt{8})e^{-\alpha^2/8}}{(\lambda_{2n+1} + \frac{3}{4})(\lambda_{2n+1} + \frac{1}{4})}. \tag{2.8}$$

An arbitrary linear combination of $f_0(\alpha)$ and $f_2(\alpha)$, eigenfunctions of $(\frac{3}{4} + \Lambda)(\frac{1}{4} + \Lambda)$ with eigenvalues zero, does not appear on the right side of (2.8) because $\omega(\alpha)$ is an even function of α. Inserting (2.5) into (2.8) gives

$$\frac{d^2}{d\alpha^2}\alpha\omega(\alpha) = \sum_{n=0}^{\infty}\frac{16A_n H_{2n+1}(\alpha/\sqrt{8})e^{-\alpha^2/8}}{(2n + 1)(2n - 1)}. \tag{2.9}$$

From the Hermite polynomial definition

$$H_n(x) \equiv (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2}, \tag{2.10}$$

we have

$$H_{2n+1}\left(\frac{\alpha}{\sqrt{8}}\right)e^{-\alpha^2/8} = 8\frac{d^2}{d\alpha^2}H_{2n-1}\left(\frac{\alpha}{\sqrt{8}}\right)e^{-\alpha^2/8}, \quad n \geq 1. \tag{2.11}$$

Therefore, substituting (2.11) into (2.9) leads to

$$\frac{d^2}{d\alpha^2}\alpha\omega(\alpha) = \frac{d^2}{d\alpha^2}\sum_{n=1}^{\infty}\frac{128A_n H_{2n-1}(\alpha/\sqrt{8})e^{-\alpha^2/8}}{(2n + 1)(2n - 1)}$$
$$+ 32\sqrt{2}\,A_0\frac{d}{d\alpha}e^{-\alpha^2/8}. \tag{2.12}$$

Equation (2.12) is integrated subject to the condition

$$\lim_{\alpha\to\pm\infty}\alpha\omega(\alpha) = 0, \tag{2.13}$$

with the result that

$$\omega(\alpha) = \frac{128}{\alpha}\sum_{n=1}^{\infty}\frac{A_n H_{2n-1}(\alpha/\sqrt{8})e^{-\alpha^2/8}}{(2n + 1)(2n - 1)}$$
$$- \frac{32\sqrt{2}\,A_0}{\alpha}\int_\alpha^\infty e^{-\alpha^2/8}\,d\alpha. \tag{2.14}$$

To ensure that $\omega(\alpha)$ remains finite at $\alpha = 0$, we must have $A_0 = 0$. But by (2.7), this condition is equivalent to the requirement

$$\int_0^\infty \alpha\Gamma(\alpha)\,d\alpha = 0, \tag{2.15}$$

which follows from the second and third members in (2.2). Hence, $\omega(\alpha)$ may be expressed as

$$\omega(\alpha) = \frac{128}{\alpha}\sum_{n=1}^{\infty}\frac{A_n H_{2n-1}(\alpha/\sqrt{8})e^{-\alpha^2/8}}{(2n + 1)(2n - 1)}. \tag{2.16}$$

An iterative solutional procedure for Eq. (2.2) goes as follows. Suppose that $\omega^{(N)}(\alpha)$ is the Nth approximation to $\omega(\alpha)$; then substituting $\omega^{(N)}(\alpha)$ into the right side of (2.2) gives the corresponding approximation for $\Gamma(\alpha)$, namely $\Gamma^{(N)}(\alpha)$. Putting $\Gamma^{(N)}(\alpha)$ into (2.7) enables one to calculate the $A_n^{(N)}$'s which are substituted into (2.16) to yield $\omega^{(N+1)}(\alpha)$, the

$(N+1)$th approximation to $\omega(\alpha)$; that is,

$$\omega^{(N+1)}(\alpha) = \frac{128}{\alpha}\sum_{n=1}^{\infty}\frac{A_n^{(N)}H_{2n-1}(\alpha/\sqrt{8})e^{-\alpha^2/8}}{(2n+1)(2n-1)} \tag{2.17a}$$

with

$$A_n^{(N)} = \frac{1}{2^{2n+1}(2n+1)!\,(2\pi)^{\frac{1}{2}}}\int_0^{\infty}\Gamma^{(N)}(\alpha)H_{2n+1}\left(\frac{\alpha}{\sqrt{8}}\right)d\alpha. \tag{2.17b}$$

If $\omega^{(N)}(\alpha)$ is expanded in terms of the even eigenfunctions of Λ, then the coefficients of the expansion are related to the $A_n^{(N)}$'s. To develop equations which give the $A_n^{(N)}$'s explicitly, first note that the definition of $\Gamma(\alpha)$ produces

$$\alpha^5\Gamma''(\alpha) = -6R(\alpha) + 6\alpha R'(\alpha) - 2\alpha^2R''(\alpha), \tag{2.18}$$

where

$$R(\alpha) = \omega'(\alpha)[6\omega'(\alpha) - 6\alpha\omega''(\alpha) + \alpha^2\omega'''(\alpha) + \alpha^3\omega''''(\alpha)]. \tag{2.19}$$

Next, substitute Eq. (2.6) into the left side of (2.18), multiply both sides of (2.18) by $H_{2l}(\alpha/\sqrt{8})$, and finally integrate over α to obtain

$$-\frac{1}{16}\sum_{n=1}^{\infty}A_n\int_{-\infty}^{\infty}\alpha^5H_{2n+3}\left(\frac{\alpha}{\sqrt{8}}\right)H_{2l}\left(\frac{\alpha}{\sqrt{8}}\right)e^{-\alpha^2/8}\,d\alpha = \int_{-\infty}^{\infty}\left[3R(\alpha) - 3\alpha\frac{d}{d\alpha}R(\alpha) + \alpha^2\frac{d^2}{d\alpha^2}R(\alpha)\right]H_{2l}\left(\frac{\alpha}{\sqrt{8}}\right)d\alpha. \tag{2.20}$$

Apply the identity $xH_k(x) = kH_{k-1}(x) + \frac{1}{2}H_{k+1}(x)$ to the right side of (2.20), integrate by parts, and use $H_k'(x) = 2kH_{k-1}(x)$ to get

$$-32\sum_{n=1}^{\infty}A_nJ_{2n+3,2l} = \int_{-\infty}^{\infty}\left[(2l+4)(2l+2)H_{2l}\left(\frac{\alpha}{\sqrt{8}}\right) + 8l(2l-1)(2l+2)H_{2l-2}\left(\frac{\alpha}{\sqrt{8}}\right) + 8l(2l-1)(2l-2)(2l-3)H_{2l-4}\left(\frac{\alpha}{\sqrt{8}}\right)\right]R(\alpha)\,d\alpha, \tag{2.21}$$

with

$$J_{k,s} \equiv \int_{-\infty}^{\infty}x^5H_k(x)H_s(x)e^{-x^2}\,dx = 2_k(\pi)^{\frac{1}{2}}[\tfrac{1}{32}k!\,\delta_{k,s+5} + \tfrac{5}{16}k!\,(k-1)\delta_{k,s+3} + \tfrac{5}{8}k!\,(2k^2+1)\delta_{k,s+1} + \tfrac{5}{4}(k+1)!\,(2k^2+4k+3)\delta_{k,s-1} + \tfrac{5}{2}(k+3)!\,(k+2)\delta_{k,s-3} + (k+5)!\,\delta_{k,s-5}], \tag{2.22}$$

where

$$\delta_{ij} = \begin{cases}1 & \text{for } i=j,\\ 0 & \text{for } i\neq j.\end{cases}$$

From (2.22) it follows that

$$\sum_{n=1}^{\infty}A_nJ_{2n+3,2l} = \sum_{n=0}^{\infty}A_{(n+l-4)}J_{2(n+l)-5,2l}. \tag{2.23}$$

If $\omega(\alpha)$ is expanded in terms of the even eigenfunctions of Λ, then

$$\omega(\alpha) = \sum_{n=0}^{\infty}b_{2n}f_{2n}(\alpha) = \sum_{n=0}^{\infty}b_{2n}H_{2n}\left(\frac{\alpha}{\sqrt{8}}\right)e^{-\alpha^2/8}. \tag{2.24}$$

Therefore

$$R(\alpha) = -\frac{e^{-\alpha^2/4}}{8}\sum_{m,n=0}^{\infty}b_{2m}B_{2n}H_{2m+1}\left(\frac{\alpha}{\sqrt{8}}\right)H_{2n+1}\left(\frac{\alpha}{\sqrt{8}}\right), \tag{2.25}$$

where

$$B_{2n} = \tfrac{1}{8}b_{2n-6} + \tfrac{1}{2}(3n+1)b_{2n-4} + \tfrac{1}{2}(6n+1)(2n+3)b_{2n-2} + 2n(2n+3)(2n+5)b_{2n}.$$

Substitute (2.23) and (2.25) into (2.21); the result is

$$32\sqrt{8}\sum_{n=0}^{5}A_{(n+l-4)}J_{2(n+l)-5,2l} = \sum_{m,n=0}^{\infty}b_{2m}B_{2n}[4(l+1)(l+2)I_{2m+1,2n+1,2l} + 16l(l+1)(2l-1)I_{2m+1,2n+1,2l-2} + 16l(l-1)(2l-1)(2l-3)I_{2m+1,2n+1,2l-4}] \equiv Q_l(b), \tag{2.26}$$

where

$$I_{2m+1,2n+1,2k} \equiv \int_{-\infty}^{\infty}H_{2m+1}(x)H_{2n+1}(x)H_{2k}(x)e^{-2x^2}\,dx. \tag{2.27}$$

With the help of

$$e^{x^2-(x-s)^2} = \sum_{n=0}^{\infty}\frac{H_n(x)s^n}{n!},$$

Eq. (2.27) becomes

$$I_{2m+1,2n+1,2k} = \left(\frac{\pi}{2}\right)^{\frac{1}{2}}(2m+1)!\,(2n+1)!\,2^{n+m+k+1} \times \sum_{r=m-n+k}^{m+n+k+1}\frac{(-1)^{3r}(2r)!\,2^{-2r}}{[m+n+k+1-r]!\,[r+(m-n+k)]!\,[r-(m-n+k)]!\,r!}, \quad m\geq n.$$

Note that the condition $m \geq n$ entails no loss of generality. In the Nth approximation, (2.24) becomes

$$\omega^{(N)}(\alpha) = \sum_{n=0}^{\infty} b_{2n}^{(N)} f_{2n}(\alpha); \tag{2.28}$$

then, from (2.26),

$$32\sqrt{8} \sum_{n=0}^{5} A_{(n+l-4)} J_{2(n+l)-5,2l} = Q_l(b^{(N)}). \tag{2.29}$$

As an illustration of this iterative solutional procedure, we consider the first approximation to $\omega(\alpha)$:

$$\omega^{(1)}(\alpha) = -ce^{-\alpha^2/8}, \quad c \equiv \text{constant}. \tag{2.30}$$

Then from (2.30) and (2.2) we find

$$\Gamma^{(1)}(\alpha) = \frac{c^2}{256}(72\alpha - 17\alpha^3 + \tfrac{1}{2}\alpha^5)e^{-\alpha^2/4}. \tag{2.31}$$

Hence it follows from (2.17b) that

$$A_n^{(1)} = \frac{c^2(-1)^n 2^{-3n}(5 + 2n)}{128(n - 1)!}, \quad n \geq 1. \tag{2.32}$$

We now combine (2.17a), (2.32), and (2.1) to obtain

$$\bar{\xi}_2 = \frac{\nu^2 c^2}{\alpha} \sum_{n=1}^{\infty} \frac{(-1)^n 2^{-3n}(5 + 2n)}{(2n + 1)(2n - 1)(n - 1)!} \times H_{2n-1}\left(\frac{\alpha}{\sqrt{8}}\right) e^{-\alpha^2/8}, \tag{2.33}$$

where $\bar{\xi}_2$ is the second approximation to $\bar{\xi}$. Equation (2.33) leads to

$$f_{\mu\mu}^{(2)}(r, t) = 2\nabla^2 \bar{\xi}_2 = \frac{\nu c^2}{4\alpha t} \sum_{n=1}^{\infty} \frac{2^{-3n}(-1)^n(5 + 2n)}{(2n + 1)(2n - 1)(n - 1)!} \times H_{2n+1}\left(\frac{\alpha}{\sqrt{8}}\right) e^{-\alpha^2/8}, \tag{2.34}$$

and therefore

$$f_{\mu\mu}^{(2)}(0, t) = \frac{\nu c^2}{4\sqrt{2}\, t} \sum_{n=1}^{\infty} \frac{2^{-3n}(5 + 2n)(2n)!}{(2n - 1)(n - 1)!\, n!}, \tag{2.35}$$

where we have used

$$\left.\frac{H_{2n+1}(x)}{x}\right|_{x=0} = 2(-1)^n(2n + 1)\frac{(2n)!}{n!}. \tag{2.36}$$

Observe that the t^{-1} dependence of the turbulent kinetic energy, proportional to the quantity (2.35), is characteristic of the initial period of decay observed experimentally.[3]

Since the infinite series in (2.35) has the approximate value 2.83, Eq. (2.35) may be written

$$f_{\mu\mu}^{(2)}(0, t) \cong c^2\nu/2t. \tag{2.37}$$

We compute $f_{\mu\mu}^{(1)}(0, t)$ from (2.30) and find

$$f_{\mu\mu}^{(1)}(0, t) = \tfrac{3}{2}(c\nu/t). \tag{2.38}$$

Comparing (2.37) and (2.38) we see that if $c \cong 3$, we have rapid convergence of the iteration procedure for

$$f_{\mu\mu}^{(2)}(0, t) \cong f_{\mu\mu}^{(1)}(0, t) \cong 9\nu/2t, \tag{2.39}$$

a relation which would hold experimentally for a grid Reynolds number[4] $UM/\nu \cong \frac{3}{2}a$. An approximation for the two-point velocity correlation tensor is given by

$$f_{\mu\gamma}^{(2)}(r, t) = \frac{c^2 e^{-\alpha^2/8}}{\alpha t} \sum_{n=1}^{\infty} \frac{(-1)^n 2^{-3n}(5 + 2n)}{(2n + 1)(2n - 1)(n - 1)!} \times \left\{ \nu\delta_{\mu\gamma}\left[\frac{1}{\alpha^2} H_{2n-1}\left(\frac{\alpha}{\sqrt{8}}\right) + \frac{1}{\sqrt{8}\,\alpha} H_{2n}\left(\frac{\alpha}{\sqrt{8}}\right) + \tfrac{1}{8} H_{2n+1}\left(\frac{\alpha}{\sqrt{8}}\right)\right] - \frac{x_\mu x_\gamma}{\alpha^2 t}\left[\frac{3}{\alpha^2} H_{2n-1}\left(\frac{\alpha}{\sqrt{8}}\right) + \frac{3}{\sqrt{8}\,\alpha} H_{2n}\left(\frac{\alpha}{\sqrt{8}}\right) + \tfrac{1}{8} H_{2n+1}\left(\frac{\alpha}{\sqrt{8}}\right)\right]\right\},$$

where[4]

$$c^2 = (6/a)(UM/\nu).$$

3. ITERATIVE SOLUTION BASED ON THE QUASILINEAR CHARACTER OF EQ. (1.7)

In this section we apply an iteration procedure for the determination of solutions to Eq. (1.7), expressed as

$$\left(\frac{1}{2}\frac{\partial}{\partial t} - \nu\frac{\partial^2}{\partial r^2}\right)^2 \frac{\partial^2(r\bar{\xi})}{\partial r^2} = -\zeta[\bar{\xi}], \tag{3.1}$$

where

$$\zeta[\bar{\xi}] = 2\bar{\eta} - 6r\int_r^\infty \bar{\eta} s^{-2}\, ds = -2r\int_r^\infty s^{-3}\frac{d}{ds}(s^2\bar{\eta})\, ds \tag{3.2}$$

and

$$\bar{\eta} = r\frac{\partial\bar{\xi}}{\partial r}\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\nabla^2\bar{\xi}}{\partial r}\right). \tag{3.3}$$

The Green's function of the operator

$$\left(\frac{1}{2}\frac{\partial}{\partial t} - \nu\frac{\partial^2}{\partial r^2}\right)^2$$

is defined by the equations

$$\left(\frac{1}{2}\frac{\partial}{\partial t} - \nu\frac{\partial^2}{\partial r^2}\right)^2 G(r, r_0, t, t_0) = \delta(r - r_0)\delta(t - t_0) \quad \text{for} \quad t \geq t_0, \tag{3.4}$$

[3] G. K. Batchelor, *Theory of Homogeneous Turbulence* (Cambridge University Press, London, 1960), p. 134.

[4] The dimensionless constant a, depending principally on the grid shape and having a value near 134 for typical square-mesh grids, was introduced empirically by G. K. Batchelor, *Theory of Homogeneous Turbulence* (Cambridge University Press, London, 1960), p. 135.

and

$$G(r, r_0, t, t_0) = 0 \quad \text{for} \quad t < t_0. \tag{3.5}$$

With Fourier representations for the delta functions, the inverse of the differential operator on the left side of (3.4) can be applied to the equation to give an expression that must be integrated twice, first by a contour integration and then by a standard integration. The result is

$$G(r, r_0, t, t_0) = \begin{cases} \left[\frac{2(t-t_0)}{\pi\nu}\right]^{\frac{1}{2}} \exp\left(-\frac{(r-r_0)^2}{8\nu(t-t_0)}\right), & t \geq t_0, \\ 0, & t \leq t_0. \end{cases} \tag{3.6}$$

Equation (3.1) then becomes

$$\frac{\partial^2(r\bar{\xi})}{\partial r^2} = -\int_{-\infty}^{\infty}\int_0^t G(r, r_0, t, t_0)\zeta[\bar{\xi}(r_0, t_0)]\, dt_0\, dr_0, \tag{3.7}$$

which leads to the iterative series of approximate expressions for $\bar{\xi}$:

$$\frac{\partial^2(r\bar{\xi}_{n+1})}{\partial r^2} = -\int_{-\infty}^{\infty}\int_0^t G(r, r_0, t, t_0)\zeta[\bar{\xi}_n(r_0, t_0)]\, dt_0\, dr_0. \tag{3.8}$$

Hence, if an approximate form is given for $\bar{\xi}$, say $\bar{\xi}_1$, we can calculate the improved approximation $\bar{\xi}_2$ by integrating Eq. (3.8).

To illustrate this iterative approximation procedure, let

$$\bar{\xi}_1 = F(t)e^{-\gamma(t)r^2}, \tag{3.9}$$

which produces

$$\nabla^2\bar{\xi}_1 = 2Fe^{-\gamma r^2}(2\gamma^2r^2 - 3\gamma), \tag{3.10}$$

$$\bar{\eta}_1 = 16F^2\gamma^4e^{-2\gamma r^2}(7r^3 - 2\gamma r^5), \tag{3.11}$$

and

$$\zeta_1 = -32F^2\gamma^4e^{-2\gamma r^2}\left(2\gamma r^5 - \tfrac{17}{2}r^3 + \frac{9r}{2\gamma}\right). \tag{3.12}$$

Thus

$$\frac{\partial^2(r\bar{\xi}_2)}{\partial r^2} = \int_{-\infty}^{\infty}\int_0^t\left[\frac{2(t-t_0)}{\pi\nu}\right]^{\frac{1}{2}} 32F^2\gamma^4\left(2\gamma r_0^5 - \tfrac{17}{2}r_0^3 + \frac{9r_0}{2\gamma}\right) \times \exp\left(-2\gamma r_0^2 - \frac{(r-r_0)^2}{8\nu(t-t_0)}\right) dt_0\, dr_0, \tag{3.13}$$

or finally

$$\frac{\partial^2(r\bar{\xi}_2)}{\partial r^2} = \int_0^t\left[\frac{2(t-t_0)}{\pi\nu}\right]^{\frac{1}{2}} 32F^2\gamma^4 \exp\left(-\frac{2\gamma r^2}{1+16\gamma\nu(t-t_0)}\right) \times \int_{-\infty}^{\infty}\left(2\gamma r_0^5 - \tfrac{17}{2}r_0^3 + \frac{9r_0}{2\gamma}\right) \times \exp\left[-\left(\frac{1+16\gamma\nu(t-t_0)}{8\nu(t-t_0)}\right) \times \left(r_0 - \frac{r}{1+16\gamma\nu(t-t_0)}\right)^2\right] dr_0\, dt_0. \tag{3.14}$$

Concentrating on the inner integral, set

$$\frac{r}{1+16\gamma\nu(t-t_0)} \equiv a \equiv a'r$$

and

$$\frac{1+16\gamma\nu(t-t_0)}{8\nu(t-t_0)} \equiv b.$$

Then, by a translation of the integration variable $r_0 \to (r_0 + a)$, we get

$$\int_{-\infty}^{\infty}\left(2\gamma r_0^5 - \tfrac{17}{2}r_0^3 + \frac{9r_0}{2\gamma}\right)e^{-b(r_0-a)^2} dr_0 = 2\gamma a'^5(\pi/b)^{\frac{1}{2}}r^5 + \left(\frac{10\gamma a'^3}{b} - \frac{17a'^3}{2}\right)(\pi/b)^{\frac{1}{2}}r^3 + \left(\frac{15\gamma a'}{2b^2} - \frac{51a'}{4b} + \frac{9a'}{2\gamma}\right)(\pi/b)^{\frac{1}{2}}r. \tag{3.15}$$

Equation (3.14) thus becomes

$$\frac{\partial^2(r\bar{\xi}_2)}{\partial r^2} = \frac{8\sqrt{2}}{\nu^{\frac{1}{2}}}\int_0^t F^2\gamma^4(t-t_0)^{\frac{1}{2}}b^{-\frac{5}{2}}(lr^5 + mr^3 + nr)e^{-Dr^2} dt_0, \tag{3.16}$$

where

$$l \equiv 8\gamma a'^5b^2, \quad m \equiv 40\gamma a'^3b - 34a'^3b^2, \quad n \equiv 30\gamma a' - 51a'b + 18a'b^2\gamma^{-1}, \quad D \equiv \{2\gamma/[1+16\gamma\nu(t-t_0)]\} \equiv 2\gamma a'. \tag{3.17}$$

Both sides of (3.16) can be integrated with respect to r to give

$$\frac{\partial(r\bar{\xi}_2)}{\partial r} = -\frac{8\sqrt{2}}{\nu^{\frac{1}{2}}}\int_0^t F^2\gamma^4(t-t_0)^{\frac{1}{2}}b^{-\frac{5}{2}} \times e^{-Dr^2}\left[\frac{l}{2D}r^4 + \left(\frac{m}{2D} + \frac{l}{D^2}\right)r^2 + \left(\frac{n}{2D} + \frac{m}{2D^2} + \frac{l}{D^3}\right)\right] dt_0, \tag{3.18}$$

where $\bar{\xi}_2 \to 0$ as $r \to \infty$ has been used. We now substitute for D, l, m, n, b, and a' to obtain

$$\frac{\partial(r\bar{\xi}_2)}{\partial r} = -\frac{\sqrt{2}}{32\nu^{\frac{5}{2}}}\int_0^t F^2\gamma^4(t-t_0)^{-\frac{3}{2}}b^{-\frac{5}{2}} \times \exp\left(-\frac{2\gamma r^2}{1+16\gamma\nu(t-t_0)}\right) \times \left\{\frac{8r^4}{[1+16\gamma\nu(t-t_0)]^2} - \frac{26+14(16)\gamma\nu(t-t_0)}{\gamma[1+16\gamma\nu(t-t_0)]}r^2 + \frac{5+56\gamma\nu(t-t_0)}{\gamma^2}\right\} dt_0. \quad (3.19)$$

Equation (3.19) can be integrated again with respect to r by using $r\bar{\xi}_2 \to 0$ as $r \to \infty$ to give

$$\bar{\xi}_2 = 8\int_0^t \frac{F^2\gamma^4(t-t_0)}{[1+16\gamma\nu(t-t_0)]^{\frac{5}{2}}} \times \left(\frac{2r^2}{\gamma[1+16\gamma\nu(t-t_0)]} - \frac{5+56\gamma\nu(t-t_0)}{\gamma^2}\right) \times \exp\left(-\frac{2\gamma r^2}{1+16\gamma\nu(t-t_0)}\right) dt_0, \quad (3.20)$$

where the expression for b has been substituted.

Note that as yet $\gamma(t)$ and $F(t)$ have not been specified. A form with some approximate validity is $F(t) =$ constant $\equiv -k$ and $\gamma(t) = 1/(8\nu t)$. If we set

$$\bar{\xi}_1 = -k\exp\left(-\frac{r^2}{8\nu t}\right), \quad (3.21)$$

it follows that

$$f_{\alpha\alpha}^{(1)}(0,t) = 2\nabla^2\bar{\xi}_1 = \frac{3k}{2\nu t}, \quad (3.22)$$

a time dependence observed experimentally during the initial period of decay. Using these values for F and γ in Eq. (3.20) yields

$$\bar{\xi}_2 = \frac{k^2}{32\nu^3}(r^2tM_{21} - r^2M_{31} - 28\nu t^2M_{10} + 36\nu tM_{20} - 8\nu M_{30}), \quad (3.23)$$

where M_{mn} is defined by

$$M_{mn} = \int_0^t \frac{t_0^{m-\frac{3}{2}}}{(2t-t_0)^{n+\frac{5}{2}}}\exp\left(-\frac{r^2}{4\nu(2t-t_0)}\right) dt_0. \quad (3.24)$$

By making the substitution

$$V = \frac{r^2}{4\nu(2t-t_0)} - \theta^2$$

where

$$\theta = r/(8\nu t)^{\frac{1}{2}}$$

and simplifying the resulting expression, we obtain

$$M_{mn} = (2t)^{m-n-3}\theta^{-2n-3}e^{-\theta^2} \times \int_0^{\theta^2} V^{m-\frac{3}{2}}(V+\theta^2)^{n-m+2}e^{-V}\,dV. \quad (3.25)$$

For the combinations of m and n which occur in (3.23), it is found that only four integrals appear in (3.25). These integrals are

$$\int_0^{\theta^2} V^{\frac{1}{2}}e^{-V}\,dV \equiv \Lambda(\theta), \quad (3.26)$$

$$\int_0^{\theta^2} V^{-\frac{1}{2}}e^{-V}\,dV = 2\theta e^{-\theta^2} + 2\Lambda(\theta), \quad (3.27)$$

$$\int_0^{\theta^2} V^{\frac{3}{2}}e^{-V}\,dV = -\theta^3e^{-\theta^2} + \tfrac{3}{2}\Lambda(\theta), \quad (3.28)$$

and

$$\int_0^{\theta^2} \frac{V^{\frac{3}{2}}e^{-V}}{V+\theta^2}\,dV = \Lambda(\theta) - \theta^3N(\theta), \quad (3.29)$$

where

$$N(\theta) \equiv \int_0^1 \frac{W^{\frac{1}{2}}}{W+1}e^{-\theta^2W}\,dW. \quad (3.30)$$

Thus, $\Lambda(\theta)$ and $N(\theta)$ are the only two integrals that remain to be determined. Now

$$\Lambda(\theta) = -\theta e^{-\theta^2} + I(\theta), \quad (3.31)$$

where

$$I(\theta) \equiv \int_0^\theta e^{-W^2}\,dW = \frac{\sqrt{\pi}}{2}\phi(\theta) \quad (3.32)$$

and ϕ is the well-known error function normalized to $\phi(\infty) = 1$. $N(\theta)$ can of course be evaluated numerically for any given θ, but such an evaluation is not required for our final result.

We now use (3.26) through (3.32) in (3.25) to get the M_{mn}'s and then evaluate (3.23). Thus

$$\bar{\xi}_2 = \frac{k^2e^{-\theta^2}}{8\nu^2}[2N(\theta) - 3\theta^{-1}I(\theta)]. \quad (3.33)$$

In order to evaluate $f_{\alpha\alpha}^{(2)}(r,t)$ we need the relations

$$N'(\theta) = 2\theta N(\theta) + 2\theta^{-1}e^{-\theta^2} - 2\theta^{-2}I(\theta) \quad (3.34)$$

and

$$I'(\theta) = e^{-\theta^2}. \quad (3.35)$$

Then

$$f_{\alpha\alpha}^{(2)}(r,t) = 2\nabla^2\bar{\xi}_2 = \frac{k^2e^{-\theta^2}}{16\nu^3t}[7\theta^{-1}I(\theta) - 6\theta I(\theta) + e^{-\theta^2}]. \quad (3.36)$$

This result is in contrast to

$$f_{\alpha\alpha}^{(1)}(r,t) = 2\nabla^2\bar{\xi}_1 = \frac{k}{\nu t}e^{-\theta^2}(\tfrac{3}{2} - \theta^2). \quad (3.37)$$

For the case when $r = 0$ and $t \neq 0$, we have $\theta = 0$ and

$$f_{\alpha\alpha}^{(2)}(0, t) = \frac{k^2}{2\nu^3 t}, \tag{3.38}$$

where

$$I(\theta) = \theta - \tfrac{1}{3}\theta^3 + \tfrac{1}{10}\theta^5 - \cdots \tag{3.39}$$

has been used. Comparing (3.22) and (3.38), we see that $f_{\alpha\alpha}^{(2)}$ is in close correspondence with $f_{\alpha\alpha}^{(1)}$, indicating rapid convergence of our iteration procedure, if k is of the order $3\nu^2$; for $k \simeq 3\nu^2$ we have

$$f_{\alpha\alpha}^{(2)}(0, t) \equiv \langle u_\alpha(\mathbf{x}, t)u_\alpha(\mathbf{x}, t)\rangle \simeq \frac{9\nu}{2t}. \tag{3.40}$$

The relation (3.40) obtains in experimental situations for a grid Reynolds number[4] $UM/\nu \simeq (\frac{3}{2})a$. We would expect

$$\begin{aligned} f_{\mu\nu}^{(2)}(r, t) = {} & \frac{k^2 x_\mu x_\nu}{512\nu^4 t^2} e^{-\theta^2}[12\theta^{-1}I(\theta) + 4\theta^{-3}I(\theta) \\ & - 3\theta^{-5}I(\theta) - 2\theta^{-2}e^{-\theta^2} + 3\theta^{-4}e^{-\theta^2}] \\ & + \frac{k^2\delta_{\mu\nu}}{64\nu^3 t} e^{-\theta^2}[8\theta^{-1}I(\theta) - 12\theta I(\theta) \\ & + \theta^{-3}I(\theta) + 2e^{-\theta^2} - \theta^{-2}e^{-\theta^2}] \end{aligned} \tag{3.41}$$

to be a satisfactory approximation for the two-point velocity-correlation tensor for grid Reynolds numbers between 100 and 300 with the constant[4]

$$k^2 = \frac{6UM\nu^3}{a}.$$

ACKNOWLEDGMENT

This work was supported by a National Science Foundation grant.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Spectral Theory of the Difference Equation $f(n + 1) + f(n - 1) = [E - \varphi(n)]f(n)$

PLATON C. DELIYANNIS AND EVANGELOS K. IFANTIS
"Democritus" Nuclear Research Center, Aghia Paraskevi, Attikis, Athens, Greece

(Received 17 April 1968)

In this work, the spectrum of the second-order difference equation

$$f(n + 1) + f(n - 1) = [E - \varphi(n)]f(n)$$

in the l^2 Hilbert space is studied for the case in which the limit of the sequence $\{\varphi(n)\}_{n=1}^{\infty}$ exists. By means of a simple representation the problem is transferred to one about the spectrum of an abstract operator in a separable Hilbert space. This operator T has a form analogous to the Schrödinger operator, namely $T = T_0 + A$, where T_0 is self-adjoint with a purely continuous spectrum but bounded, while A depends on the sequence $\{\varphi(n)\}$. In fact, A is of Hilbert–Schmidt type for any $\{\varphi(n)\}$ in l^2, and of trace class if the series $\sum_{n=1}^{\infty} |\varphi(n)|$ converges. Sufficient conditions for the existence of a discrete spectrum and more generally, of proper values, are found. Using the theory of the wave operators $\Omega_\pm = s - \lim_{t\to\pm\infty} \exp(iTt)$ range $\exp(-iT_0 t)$, results on the existence of a mixed spectrum are obtained.

I. INTRODUCTION

Problems leading to difference and, more generally, to functional equations have appeared in mathematics and physics[1] (some even with a boundary-value character[2]) before classical analysis was developed. These problems were, however, rather simple and could be solved by direct algebraic methods.

Later on, more complex forms appeared, while the methods of treatment varied according to the known methods of analysis.[3]

Most of the time, due to the analogies between difference and differential equations, methods used with success in the case of differential equations were applied, and results of differential equations were translated into analogous properties of difference equations.

[1] H. Meschkowski, *Differenzengleichungen* (Vandenhöch und Rupprecht, Göttingen, 1959); E. Bohmer, *Differenzengleichungen und bestimmte Integrale* (K. F. Köhler Verlag, Leipzig, 1939).

[2] Z. Aczel, *Vorlesungen über Functionalgleichungen* (Birkhäuser Verlag, Basel, 1961), p. 106.

[3] G. Doetsch, *Handbuch der Laplace-Transformation* (Birkhäuser Verlag, Basel, 1956), Vol. 3, p. 91.

For the case when $r = 0$ and $t \neq 0$, we have $\theta = 0$ and

$$f_{\alpha\alpha}^{(2)}(0, t) = \frac{k^2}{2\nu^3 t}, \tag{3.38}$$

where

$$I(\theta) = \theta - \tfrac{1}{3}\theta^3 + \tfrac{1}{10}\theta^5 - \cdots \tag{3.39}$$

has been used. Comparing (3.22) and (3.38), we see that $f_{\alpha\alpha}^{(2)}$ is in close correspondence with $f_{\alpha\alpha}^{(1)}$, indicating rapid convergence of our iteration procedure, if k is of the order $3\nu^2$; for $k \simeq 3\nu^2$ we have

$$f_{\alpha\alpha}^{(2)}(0, t) \equiv \langle u_\alpha(\mathbf{x}, t)u_\alpha(\mathbf{x}, t)\rangle \simeq \frac{9\nu}{2t}. \tag{3.40}$$

The relation (3.40) obtains in experimental situations for a grid Reynolds number[4] $UM/\nu \simeq (\frac{3}{2})a$. We would expect

$$\begin{aligned} f_{\mu\nu}^{(2)}(r, t) = {} & \frac{k^2 x_\mu x_\nu}{512\nu^4 t^2} e^{-\theta^2}[12\theta^{-1}I(\theta) + 4\theta^{-3}I(\theta) \\ & - 3\theta^{-5}I(\theta) - 2\theta^{-2}e^{-\theta^2} + 3\theta^{-4}e^{-\theta^2}] \\ & + \frac{k^2\delta_{\mu\nu}}{64\nu^3 t} e^{-\theta^2}[8\theta^{-1}I(\theta) - 12\theta I(\theta) \\ & + \theta^{-3}I(\theta) + 2e^{-\theta^2} - \theta^{-2}e^{-\theta^2}] \end{aligned} \tag{3.41}$$

to be a satisfactory approximation for the two-point velocity-correlation tensor for grid Reynolds numbers between 100 and 300 with the constant[4]

$$k^2 = \frac{6UM\nu^3}{a}.$$

ACKNOWLEDGMENT

This work was supported by a National Science Foundation grant.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Spectral Theory of the Difference Equation $f(n + 1) + f(n - 1) = [E - \varphi(n)]f(n)$

PLATON C. DELIYANNIS AND EVANGELOS K. IFANTIS
"Democritus" Nuclear Research Center, Aghia Paraskevi, Attikis, Athens, Greece

(Received 17 April 1968)

In this work, the spectrum of the second-order difference equation

$$f(n + 1) + f(n - 1) = [E - \varphi(n)]f(n)$$

in the l^2 Hilbert space is studied for the case in which the limit of the sequence $\{\varphi(n)\}_{n=1}^{\infty}$ exists. By means of a simple representation the problem is transferred to one about the spectrum of an abstract operator in a separable Hilbert space. This operator T has a form analogous to the Schrödinger operator, namely $T = T_0 + A$, where T_0 is self-adjoint with a purely continuous spectrum but bounded, while A depends on the sequence $\{\varphi(n)\}$. In fact, A is of Hilbert–Schmidt type for any $\{\varphi(n)\}$ in l^2, and of trace class if the series $\sum_{n=1}^{\infty} |\varphi(n)|$ converges. Sufficient conditions for the existence of a discrete spectrum and more generally, of proper values, are found. Using the theory of the wave operators $\Omega_\pm = s - \lim_{t\to\pm\infty} \exp(iTt)$ range $\exp(-iT_0 t)$, results on the existence of a mixed spectrum are obtained.

I. INTRODUCTION

Problems leading to difference and, more generally, to functional equations have appeared in mathematics and physics[1] (some even with a boundary-value character[2]) before classical analysis was developed. These problems were, however, rather simple and could be solved by direct algebraic methods.

Later on, more complex forms appeared, while the methods of treatment varied according to the known methods of analysis.[3]

Most of the time, due to the analogies between difference and differential equations, methods used with success in the case of differential equations were applied, and results of differential equations were translated into analogous properties of difference equations.

[1] H. Meschkowski, *Differenzengleichungen* (Vandenhöch und Rupprecht, Göttingen, 1959); E. Bohmer, *Differenzengleichungen und bestimmte Integrale* (K. F. Köhler Verlag, Leipzig, 1939).

[2] Z. Aczel, *Vorlesungen über Functionalgleichungen* (Birkhäuser Verlag, Basel, 1961), p. 106.

[3] G. Doetsch, *Handbuch der Laplace-Transformation* (Birkhäuser Verlag, Basel, 1956), Vol. 3, p. 91.

The second-order recurrence equation

$$f(n+1)+f(n-1)=[E-\varphi(n)]f(n) \quad (1)$$

was studied very closely, because of the analogy with the second-order differential equation of Liouville or Schrödinger type, and all that was established for the Schrödinger equation was carried over to Eq. (1) (e.g., linear independence of solutions, asymptotic forms,[4] constancy of the Wronskian, approximation methods,[5] etc.).

In this work we study Eq. (1) in the l^2 Hilbert space, using methods of functional analysis. By translating the problem into a problem on abstract operators in a separable Hilbert space, we find a large class of sequences $\{\varphi(n)\}$ for which the solutions of (1) form a complete system in l^2 (discrete spectrum). We also exhibit another large class of sequences $\{\varphi(n)\}$ for which proper values exist, but the corresponding proper vectors do not form a complete system (mixed spectrum). As a mixed spectrum in quantum mechanics characterizes the scattering systems, we find something analogous to the admissible interaction operators in Schrödinger's equation for which the scattering operator can be defined.[6-8]

In general, we study qualitatively the spectrum of Eq. (1) for real-valued sequences $\{\varphi(n)\}$, such that either

$$\lim_{n\to\infty}\varphi(n)=\infty \quad (2)$$

or

$$\lim_{n\to\infty}\varphi(n)=a\neq\infty. \quad (3)$$

II. REDUCTION TO OPERATOR FORM

Consider an arbitrary but fixed basis $\{e_n\}$, $n=1, 2, \cdots$, in a separable Hilbert space H and define the operators V and A as follows:

$$Ve_n=e_{n+1},$$
$$Ae_n=\varphi(n)e_n.$$

The operator V is an isometry and its adjoint V^* a partial isometry:

$$V^*e_n=\begin{cases}0, & \text{for } n=1,\\ e_{n-1}, & \text{for } n>1,\end{cases}$$

so that $V^*V=I$ (identity) while $VV^*=P$ (projection on $H-\{e_1\}$).

Because of the isomorphism between H and l^2, the proper-value problem for Eq. (1) in the space l^2 is equivalent to the proper-value problem for the operator

$$T=T_0+A,$$

where

$$T_0=V+V^*$$

in the space H.

The operator T_0 plays the role of a free Hamiltonian while A is a perturbation. As T_0 is bounded, there is no problem whether T is self-adjoint or not.

III. THE OPERATORS T_0, A, AND T

We establish below the properties of the operators T_0, A, and T, which we use later.

Proposition 1: The spectrum of T_0 is continuous extending from -2 to 2.

Proof: First observe that

$$\|T_0\|=2.$$

As $\|V\|=\|V^*\|=1$, we have $\|T_0\|\leq 2$. Now let $f=n^{-\frac{1}{2}}\sum_{m=1}^{n}e_m$ so that $\|f\|=1$ and $\|T_0f\|^2=4-5n^{-1}$; since $\|T_0\|=\sup\|T_0f\|$ over all f with $\|f\|=1$, we obtain $\|T_0\|=2$. Now let λ be a proper value of T_0 with proper vector $f=\sum_{m=1}^{\infty}a_me_m$; then we have $a_{m+1}+a_{m-1}=\lambda a_m$ (with $a_0=0$); Hence $a_m=c(l_1^m-l_2^m)$ where l_1, l_2 are the roots of $l^2-\lambda l+1=0$. But then $\sum_{m=1}^{\infty}|a_m|^2$ does not converge and thus f is not a vector in H. Therefore T_0 has no proper values.

Proposition 2: In case

$$\lim_{n\to\infty}\varphi(n)=\infty,$$

the operator A has a self-adjoint extension with discrete spectrum.

Proof: We may assume $\varphi(n)\neq 0$ for all n without loss of generality, since addition of a constant multiple of I does not change the nature of the spectrum. First observe that the operator B defined by

$$Be_n=\varphi(n)^{-1}e_n$$

is self-adjoint [since the $\varphi(n)$ are real] and completely continuous[9] because $\varphi(n)^{-1}$ tends to zero. Since $AB=BA=I$ on the dense linear manifold spanned by the basis $\{e_n\}$, the inverse of B, which is self-adjoint, is an extension of A. To show that the spectrum is discrete, let $\lambda\neq\varphi(n)$ for all n. Then $[\lambda-\varphi(n)]^{-1}$

[4] J. Meixner and F. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen* (Springer-Verlag, Berlin, 1954).
[5] H. Schmidt, Math. Nachr. **1**, 377 (1948); **2**, 35 (1949); P. Harper, Proc. Phys. Soc. (London) **A68**, 874 (1955).
[6] J. M. Jauch and I. I. Zinnes, Nuovo Cimento **11**, 553 (1959).
[7] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966).
[8] J. M. Jauch, Helv. Phys. Acta **31**, 127, 661 (1958).
[9] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Frederick Ungar Publishing Co., New York, 1955), p. 235.

tends to zero and, since

$$(A - \lambda I)^{-1}e_n = [\varphi(n) - \lambda]^{-1}e_n,$$

the operator $(A - \lambda I)^{-1}$ is bounded. Thus λ is not in the spectrum of A, which therefore consists of the points $\varphi(n)$.

Proposition 3: The operator T is self-adjoint on a suitable domain; its spectrum may extend beyond $[-2, 2]$ by a distance $\|A\| = \sup|\varphi(n)|$.

The first part follows from Proposition 2 and the fact that T_0 is bounded. The second is obvious.

Proposition 4: The operator $T - aI$, where a real, cannot be completely continuous.

Proof: If the operator is completely continuous, then the sequence $\{e_n\}$ which converges weakly to zero would be mapped onto a sequence strongly convergent to zero; i.e., we should have

$$\lim_{n\to\infty}\|(T - aI)e_n\|^2 = \lim_{n\to\infty}[2 + (\varphi(n) - a)^2] = 0$$

which is impossible for real a.

Proposition 5: The one-parameter unitary group

$$U_t = \exp(-iT_0t)$$

acting on e_1 produces the vector

$$U_te_1 = \sum_{m=1}^{\infty} m(-i)^{m-1}t^{-1}J_m(2t)e_m, \tag{4}$$

where J_m is the ordinary Bessel function of order m.

Proof: By induction we have

$$T_0^n e_1 = \sum_{k=0}^{[n/2]}\left[\binom{n-1}{k} - \binom{n-1}{n+1-k}\right]e_{n+1-2k},$$

where we set $\binom{\mu}{\lambda} = 0$ if $\lambda > \mu$, and

$$U_te_1 = \sum_{n=0}^{\infty}\frac{(-it)^n}{n!} \times \sum_{k=0}^{[n/2]}\left[\binom{n-1}{k} - \binom{n-1}{n+1-k}\right]e_{n+1-2k};$$

as

$$\left(\frac{1}{n!}\right)\left[\binom{n-1}{k} - \binom{n-1}{n+1-k}\right] = \frac{(n+1-2k)}{k!\,(n+1-k)!},$$

we obtain

$$U_te_1 = \sum_{n=0}^{\infty}(-it)^n\sum_{k=0}^{[n/2]}\frac{n+1-2k}{k!\,(n+1-k)!}e_{n+1-2k}.$$

Setting $n + 1 - 2k = m$, $m = 1, 2, 3, \cdots$, and rearranging, we obtain

$$U_te_1 = \sum_{m=1}^{\infty} m(-i)^{m-1}t^{-1}\sum_{k=0}^{\infty}\frac{(-i)^k t^{2k+m}}{k!\,(m+k)!}.$$

Q.E.D.

IV. THE SPECTRUM OF T: FIRST CASE

We consider the case where the perturbation A satisfies condition (2). Then T has a discrete spectrum, because of the following theorem.

Theorem 1: If T_0 is bounded self-adjoint and A self-adjoint with A^{-1} completely continuous, then $T_0 + A$ has a discrete spectrum.

This theorem is also valid for T_0 not self-adjoint, as shown by Osborn[10] for the special case where A^{-1} is of Hilbert–Schmidt type. For the case that concerns us, i.e., T_0 self-adjoint, the converse is also true if and only if T is not bounded.

Given the above theorem, we observe that without loss of generality we may assume that $\varphi(n) \neq 0$. Then A^{-1} exists and $A^{-1}e_n = [\varphi(n)]^{-1}e_n$ so that, since

$$\lim_{n\to\infty}(\varphi(n))^{-1} = 0,$$

A^{-1} is completely continuous. Hence $T_0 + A$ has a discrete spectrum.

Proof of Theorem 1: First assume that $\|T_0\| \cdot \|A^{-1}\| < 1$; then, as $T_0 + A = (I + T_0A^{-1})A$ and $\|T_0A^{-1}\| < 1$, we see that

$$(T_0 + A)^{-1} = A^{-1}(I + T_0A^{-1})^{-1}$$

because $I + T_0A^{-1}$ has a bounded inverse. Since A^{-1} is completely continuous and $(I + T_0A^{-1})^{-1}$ is bounded, $(T_0 + A)^{-1}$ is completely continuous and therefore the self-adjoint operator $T_0 + A$ has a discrete spectrum as is seen by the argument of Proposition 2. To get rid of the extra hypothesis

$$\|T_0\| \cdot \|A^{-1}\| < 1,$$

we replace A by $A_1 = A + \lambda I$, where λ is to be chosen later, so that

$$\|T_0\| \cdot \|A_1^{-1}\| < 1;$$

then the argument applies to $T_0 + A_1$ and the result holds for $T_0 + A = T_0 + A_1 - \lambda I$.

Suppose that for some basis $\{e_n\}$ we have

$$A^{-1}e_n = \lambda_n e_n,$$

where $|\lambda_1| \geq |\lambda_2| \geq \cdots \to 0$, so that

$$A_1^{-1}e_n = (\lambda + \lambda_n^{-1})^{-1}e_n.$$

Since $(\lambda + \lambda_n^{-1})^{-1} \to 0$ as $n \to \infty$, A_1^{-1} is completely

[10] J. E. Osborn, J. Math. & Phys. **45**, 391 (1966).

continuous. Further, we have

$$\|A_1^{-1}\| = |\lambda + \lambda_i^{-1}|^{-1}$$

for some λ_i, and it suffices to choose

$$\lambda > \|T_0\| + |\lambda_i|^{-1}.$$

V. THE SPECTRUM OF T: SECOND CASE

We now consider the case where A satisfies condition (3); then $A - aI$ is completely continuous, and as the addition of $-aI$ simply translates the spectrum of T, we may assume A itself to be completely continuous, i.e., $\lim_{n\to\infty} \varphi(n) = 0$. Then the limit points of T_0 are the same as those of T[7,9] and, therefore, the limit points of T cover the interval $[-2, 2]$.

Using the results of Sec. III we have the following theorem.

Theorem 2: The operator T cannot have a discrete spectrum $\{E_n\}$ such that $\lim_{n\to\infty} E_n$ exists, i.e., it cannot have a pure point spectrum with a single limit point. In the case

$$\|T\| > 2, \tag{5}$$

the point spectrum of T is not empty.

The first follows from the fact that, if

$$\lim_{n\to\infty} E_n = a,$$

then $T - aI$ would be completely continuous. The second follows from the above remark that the limit points of T cover the interval $[-2, 2]$.

Remarks: (1) As $\|Te_n\|^2 = 2 + [\varphi(n)]^2$, we see that, if some $|\varphi(n)| > \sqrt{2}$, then $\|T\| > 2$, and a point spectrum exists. (2) Another case in which $\|T\| > 2$ is obtained as follows:

Let f be a normalized proper vector of V^* with a real proper value λ, i.e.,

$$f = (1 - \lambda)^{\frac{1}{2}} \sum_{n=1}^{\infty} \lambda^{n-1} e_n.$$

Then (5) holds, provided

$$(1 - \lambda^2)[\lambda + \varphi(1)]^2 + \lambda^4 + 2\lambda^2 + \lambda^{-2}(1 - \lambda^2) \times \sum_{n=2}^{\infty} \lambda^{2n}[\varphi(n)]^2 + 2\lambda^{-3}(1 - \lambda^4) \sum_{n=2}^{\infty} \lambda^{2n}\varphi(n) > 3. \tag{6}$$

For example, in the case of $\varphi(n) = \lambda^{\beta n}$, $\lambda < 1$, we obtain from (6), with some manipulation, that (5) holds if $\lambda > (\frac{2}{3})^{1/\beta}$.

Now it may be the case that T has a pure-point spectrum dense in $[-2, 2]$, as the Weyl–von Neumann theorem predicts.[7,11] This, however, cannot happen whenever A is of trace class, according to a theorem of Kato (valid for arbitrary self-adjoint T_0).[7] Using the theory of wave operators, we give below a simple argument which covers a special case of Kato's theorem.

Theorem 3: In case there exists an $\alpha > 1$ such that

$$\sum_{m=1}^{\infty} [\varphi(m)]^2 m^{\alpha} < \infty, \tag{7}$$

at least part of the spectrum of T is continuous.

In particular, we see that in case $\varphi(m) \sim 1/m^{1+\epsilon}$ for large m, the spectrum of T contains a continuous part. This is a result analogous to that for ordinary differential operators of Schrödinger type.[7] Our argument can be used to give a proof for the case where A is of trace class analogous to, but naturally simpler than, Kato's argument.

We first summarize the properties of the wave operators that we need; proofs are to be found in the literature.[6–8] Let $V_t = \exp(-iTt)$, $U_t = \exp(-iT_0t)$, for $-\infty < t < \infty$. The wave operators are defined by

$$\Omega_{\pm} = s - \lim_{t\to\pm\infty} V_t^* U_t.$$

(a) A necessary and sufficient condition for the existence of $\Omega_{\pm} f$ is the existence of the strong integral

$$\int^{\pm\infty} V_t^* A U_t f \, dt.$$

From this the sufficient condition

$$\int^{\pm\infty} \|AU_t f\| \, dt < +\infty \quad \text{(Cook's criterion)},$$

is easily obtained.

(b) Since $V_t^* U_t$ is unitary, if $\Omega_{\pm} f$ exists, we have

$$\|\Omega_{\pm} f\| = \|f\|,$$

so that $\Omega_{\pm}$ are isometries on their nontrivial respective domains, and therefore their ranges are not trivial.

(c) If $\Omega_{\pm} f$ exists for some vector f, then it exists for an infinite-dimensional subspace of H, because T_0 has a purely continuous spectrum.

(d) The ranges of $\Omega_{\pm}$ are contained in the continuous subspace of H. [A short proof of (d) is given in the appendix.]

[11] J. von Neumann, *Collected Works* (Pergamon Press, N.Y., 1961), Vol. 4, p. 38.

Proof of Theorem 3: We apply Cook's criterion; from Eq. (4) we obtain

$$\|AU_t e_1\| = \left[\sum_{m=1}^{\infty} m^2 t^{-2}[J_m(2t)]^2[\varphi(m)]^2\right]^{\frac{1}{2}}$$

and hence, using Schwartz's inequality,

$$\left[\int^{\infty} \|AU_t e_1\|\, dt\right]^2 \leq \int^{\infty} t^{-1-\epsilon}\, dt \int^{\infty} \sum_{m=1}^{\infty} m^2[\varphi(m)]^2 t^{-1+\epsilon}[J_m(2t)]^2\, dt, \tag{8}$$

where $\epsilon > 0$ is to be chosen later. The first factor is finite, while the second is bounded by

$$\sum_{m=1}^{\infty} m^2[\varphi(m)]^2 \int^{\infty} t^{-1+\epsilon}[J_m(2t)]^2\, dt.$$

Since[12]

$$\int_0^{+\infty} [J_m(t)]^2 t^{-s}\, dt = \frac{\Gamma(s)\Gamma(m + \frac{1}{2}(1-s))}{2^s\Gamma(\frac{1}{2}(s+1))^2\Gamma(m + \frac{1}{2}(1+s))},$$

we have for large m that

$$\int_0^{+\infty} [J_m(2t)]^2 t^{-1+\epsilon}\, dt \sim m^{-1+\epsilon}.$$

Thus the second factor in (8) is finite, provided the series $\sum_{m=1}^{\infty} [\varphi(m)]^2 m^{1+\epsilon}$ converges. This is insured by our hypothesis (7), provided we choose $0 < \epsilon < 1 - \alpha$.

Remarks: (1) We can see easily that the condition

$$\sum_{m=1}^{\infty} \varphi(m)^2 m^{1+\epsilon} < \infty$$

(for some $\epsilon > 0$) implies that A is of trace class. For if we write $|\varphi(m)| = m^{-[1+a(m)]}$, then

$$\sum_{m=1}^{\infty} |\varphi(m)| = X + Y,$$

where X contains all terms with $a(m) \geq \epsilon$, and Y all those with $a(m) < \epsilon$. Then X is majorized by $\sum_{m=1}^{\infty} m^{-1-\epsilon}$ and Y by $\sum_{m=1}^{\infty} \varphi(m)^2 m^{1+\epsilon}$, and hence both converge.

(2) The full theorem of Kato for our case can be obtained in an easier way if we observe that the really hard part in the argument of Ref. 7 concerns the case of a perturbation of finite rank. Here this is trivial because the integral

$$\int^{\infty} V_t^* A U_t e_1\, dt = \int^{\infty} \sum_{m=1}^{n} m(-i)^{m-1} t^{-1} J_m(2t) V_t^* e_m\, dt$$

obviously exists.

VI. A TYPICAL CASE: COULOMB PERTURBATION

For $\varphi(n) = 2b/n$, the operator A is called a Coulomb perturbation. The proper values of T are[13]

$$E_k = \pm 2[1 + (b/k)^2]^{\frac{1}{2}}, \quad k = 1, 2, 3, \cdots, \tag{9}$$

corresponding to $b > 0$ or $b < 0$. The proper vectors have the form

$$f_k(n) = n(E_k + b/k)^{-n} P_k(n), \tag{10}$$

where the $P_k(n)$ are polynomials in n of degree k, and they do not form a complete system in l^2 as follows from Theorem 2. This is very difficult to prove by classical methods.

Remark: The proper values (9) and proper vectors (10) have been obtained by classical methods. It is clear that an abstract approach to this special problem may lead to a general method of obtaining the proper values of Eq. (1).

APPENDIX

We give a short proof of statement (d) in Sec. V for a general scattering system. To begin with, we can see immediately from the definition that any wave operator Ω intertwines U_t and V_t:

$$\Omega U_t = V_t \Omega.$$

Let R be the range of Ω (a subspace $\neq 0$), and P the projection on R: $P = \Omega\Omega^*$. If E_λ, F_λ are the spectral families of T_0, we have

$$\Omega E_\lambda = F_\lambda \Omega, \quad \Omega^* F_\lambda = E_\lambda \Omega^*,$$

so that $F_\lambda P = F_\lambda \Omega\Omega^* = \Omega E_\lambda \Omega^* = \Omega\Omega^* F_\lambda = PF_\lambda$ and thus R reduces T. Suppose $Tf = \lambda f$ for some f in R: then $f = \Omega g$ for some g in H, and hence $T\Omega g = \lambda\Omega g$. As $\Omega E_\lambda = F_\lambda \Omega$ implies $\Omega T_0 = T\Omega$, we have $\Omega(T_0 g - \lambda g) = 0$.

As Ω is an isometry, we get $T_0 g = \lambda g$ which is impossible, because T_0 has continuous spectrum. Thus $g = 0$, and hence, $f = 0$. Therefore no vector of R is a proper vector of T, i.e., R is contained within the continuous subspace of T.

[12] I. M. Ryshik and I. S. Gradstein, *Tables* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1957), p. 231.

[13] E. K. Ifantis, Z. Angew. Math. Mech. **48**, 66 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Invariants to All Orders in Classical Perturbation Theory

TIMOTHY P. COFFEY, JR.
Edgerton, Germeshausen and Grier, Inc., Arlington, Virginia

(Received 24 May 1968)

The method of averaging is applied to systems having Hamiltonians of the form

$$H = H_0(\mathbf{J}) + \epsilon H_1(\mathbf{J}, \psi, \mathbf{p}, \mathbf{q}),$$

where H_1 is periodic in each of the components of ψ. When the system is nondegenerate it is shown that corresponding to each component of ψ there is a quantity K_i which is invariant to all orders in ϵ. When the system has an m-fold degeneracy, somewhat weaker results are obtained. In this case it is shown that the Hamiltonian, when expressed in terms of the average variables, depends on the angle variables only through their m degenerate combinations. This is true to all orders in ϵ. Thus, if ψ has s components there are $s - m$ invariants provided that the average variables can be made canonical. However, the conditions under which degenerate perturbation theory can be made canonical are not known. The invariants which arise when the Hamiltonian has an adiabatic or a harmonic time dependence are also discussed. The techniques are applied to the simple case of a harmonic oscillator whose frequency varies slowly with time.

1. INTRODUCTION

In recent years there have appeared several publications which discuss the existence of quantities which are invariant to all orders in some small parameter. Apparently the first such discussion was given by Kulsrud who considered a harmonic oscillator whose frequency varied slowly from an initial constant value to a final constant value.[1] He showed that the ratio of the energy to the frequency was the same in the final state as it was in the initial state to all orders in the slowness parameter. Shortly thereafter, Kruskal considered a charged particle moving in a magnetic field which varied slowly from an initial constant value to a final constant value.[2] Using perturbation theory he was able to demonstrate the existence of a quantity (the magnetic moment) whose final value was the same as its initial value to all orders in perturbation theory. Chandrasekhar then considered a harmonic oscillator whose frequency varied in a slow but arbitrary fashion.[3] He showed how to construct a quantity which is constant to any desired order in the slowness parameter. This was accomplished by a sequence of transformations of the dependent and independent variables. Next Lenard considered a one-dimensional oscillator whose energy varied slowly with time.[4] He was able to show that the action integral, extended over a period of the instantaneous time-independent problem, was invariant to all orders in the slowness parameter. Gardner also considered a one-dimensional oscillatory system whose Hamiltonian varied slowly with time.[5] He showed that, by performing a certain sequence of canonical transformations, one could construct a quantity which is constant to any desired order in the slowness parameter. Finally Kruskal discussed a general class of Hamiltonian systems whose motions were nearly periodic.[6] He was able to show that the action integral when defined in an appropriate way is constant to all orders in perturbation theory.

The papers which we have discussed above have one thing in common: they deal with systems having a single rapid phase. The invariants are in some sense related to an averaging over this rapid phase. The purpose of the present paper is to extend the results obtained above to systems having several rapid phases. As our starting point, we employ a perturbation theory which we developed in a forthcoming publication (henceforth cited as I).[7] This perturbation theory endeavors to separate the average or secular motion from the rapidly fluctuating motion. The introduction of average variables at the very outset is quite advantageous.[8] As we shall see, the invariants are, in a sense, built into the method of averaging.

Our program is as follows: In Sec. 2 we apply the

[1] R. Kulsrud, Phys. Rev. **106**, 205 (1957).
[2] M. Kruskal, *Rendiconti del Terzo Congresso Internazionale sui Fenomeni D'Ionizzazione nei Gas tenuto a Venezia* (Societa Italiana di Fisica, Milan, 1957).
[3] S. Chandrasekhar, *The Plasma in a Magnetic Field*, R. Landshoff, Ed. (Stanford University Press, Stanford, Calif., 1958).
[4] A. Lenard, Ann. Phys. (N.Y.) **6**, 261 (1959).
[5] C. S. Gardner, Phys. Rev. **115**, 791 (1959).
[6] M. Kruskal, J. Math. Phys. **3**, 806 (1962).
[7] T. P. Coffey and G. W. Ford, J. Math. Phys. (to be published).
[8] It should be pointed out here that the perturbation theory developed by Kruskal (Ref. 6) attempts to introduce average or "nice" variables. Indeed, Kruskal's perturbation theory is intimately related to the method of rapidly rotating phase developed by Bogoliubov and Mitropolsky [N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic Methods in the Theory of Non-Linear Oscillations* (Hindustan Publishing Company, Delhi, India, 1961)].

method of averaging to nondegenerate Hamiltonian systems. We show that for each angular degree of freedom there is a quantity which is invariant to all orders. These invariants are a direct consequence of the fact that nondegenerate perturbation theory can be made canonical to all orders. In Sec. 3 we apply the method of averaging to a Hamiltonian system which has an m-fold degeneracy.[9] We show that the Hamiltonian when expressed in terms of the average variables depends on the angles only through their m degenerate combinations. This result holds to all orders. If the system has s angular degrees of freedom then there will be $s - m$ invariants provided that degenerate perturbation theory can be made canonical to all orders. However, the conditions under which degenerate perturbation theory can be made canonical are not known. Thus, the results which we obtain for degenerate systems are weaker than those which we obtain for nondegenerate systems. In Sec. 4 we discuss the simple problem of a harmonic oscillator whose frequency varies slowly with time.

A few words are now in order concerning the statement that a quantity is invariant to all orders. This does not mean that this quantity is a rigorous constant of the motion. For such a quantity to be a rigorous constant of the motion, perturbation theory would have to converge to the exact solution. Perturbation theory, however, makes no claims of convergence. To date only some rather weak theorems concerning the asymptotic convergence of perturbation theory have been given.[10] By invariance to all orders we simply mean that a quantity is constant to all orders in the formal structure of perturbation theory. The practical usefulness of these formal constants depends on the extent to which perturbation theory provides an approximate description of the true motion. We should also mention here that our classifications of systems as nondegenerate or m-fold degenerate are idealizations. A system is truly nondegenerate or m-fold degenerate only to some finite order of perturbation theory due to the appearance of small divisors in the higher orders of perturbation theory. The appearance of these small divisors influences the extent to which our idealized perturbation theories represent the true motion. However, if these small divisors occur only at high orders of perturbation theory, we can expect that the formal constants which we obtain will be quite useful from a practical viewpoint.

2. NONDEGENERATE SYSTEMS

Let us first consider systems which are described by a Hamiltonian of the form[11]

$$H = H_0(\mathbf{J}) + \epsilon H_1(\mathbf{J}, \boldsymbol{\psi}, \mathbf{p}, \mathbf{q}). \tag{2.1}$$

Here ϵ is a small parameter, J_i is the momentum conjugate to the angle variable ψ_i $(i = 1, 2, \cdots, s)$, and p_k is the momentum conjugate to the variable q_k $(k = 1, 2, \cdots, r)$. We require that H_1 is a periodic function of each of the ψ_i's and that both H_0 and H_1 are infinitely differentiable functions of their arguments.

Hamilton's equations of motion are found from Eq. (2.1) to be

$$\dot{J}_i = -\epsilon \frac{\partial H_1}{\partial \psi_i}, \tag{2.2a}$$

$$\dot{p}_k = -\epsilon \frac{\partial H_1}{\partial q_k}, \tag{2.2b}$$

$$\dot{q}_k = \epsilon \frac{\partial H_1}{\partial p_k}, \tag{2.2c}$$

$$\dot{\psi}_i = \omega_i^0(\mathbf{J}) + \epsilon \frac{\partial H_1}{\partial J_i}, \tag{2.2d}$$

where

$$\omega_i^0(J) = \frac{\partial H_0}{\partial J_i}. \tag{2.3}$$

These equations are in the standard form given in I. The object of the perturbation theory presented in I is to achieve a separation of the rapidly fluctuating motion from the slow secular motion. In the case of nondegenerate systems this separation is achieved by the following change of variables:

$$p_k = P_k + \sum_{n=1}^{\infty} \epsilon^n D_k^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}), \tag{2.4a}$$

$$q_k = Q_k + \sum_{n=1}^{\infty} \epsilon^n E_k^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}), \tag{2.4b}$$

$$J_i = K_i + \sum_{n=1}^{\infty} \epsilon^n F_i^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}), \tag{2.4c}$$

$$\psi_i = \phi_i + \sum_{n=1}^{\infty} \epsilon^n G_i^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}), \tag{2.4d}$$

where the $D_k^{(n)}$'s, $E_k^{(n)}$'s, $F_i^{(n)}$'s, and $G_i^{(n)}$'s are all required to be periodic functions of each of the ϕ_i's. In order that the P_k's, Q_k's, K_i's, and ϕ_i's represent

[9] In this paper, we interpret the term degeneracy to mean a near commensurability of the unperturbed frequencies as well as a pure commensurability of these frequencies.

[10] For a discussion of the asymptotic convergence of perturbation theory, see J. Berkowitz and C. S. Gardner, Commun. Pure Appl. Math. 12, 501 (1959). See also Ref. 6 and T. P. Coffey, "Analytical Methods in the Theory of Non-Linear Oscillations" (Ph.D. thesis, The University of Michigan, Ann Arbor, 1966).

[11] The extension to cases where the Hamiltonian contains second- and higher-order terms is straightforward. Hamiltonians of the form (2.1) describe systems which are nearly multiple-periodic.

only the secular motion, we require that

$$\dot{P}_k = \sum_{n=1}^{\infty} \epsilon^n a_k^{(n)}(\mathbf{P}, \mathbf{Q}, \mathbf{K}), \tag{2.5a}$$

$$\dot{Q}_k = \sum_{n=1}^{\infty} \epsilon^n b_k^{(n)}(\mathbf{P}, \mathbf{Q}, \mathbf{K}), \tag{2.5b}$$

$$\dot{K}_i = \sum_{n=1}^{\infty} \epsilon^n A_i^{(n)}(\mathbf{P}, \mathbf{Q}, \mathbf{K}), \tag{2.5c}$$

$$\dot{\phi}_i = \omega_i^0(\mathbf{K}) + \sum_{n=1}^{\infty} \epsilon^n B_i^{(n)}(\mathbf{P}, \mathbf{Q}, \mathbf{K}), \tag{2.5d}$$

where the right-hand sides of (2.5) are independent of the ϕ_i's.

Equations (2.4) and (2.5) are simply the formulation of nondegenerate perturbation theory as presented in I. In I we pointed out that the functions $D_k^{(n)}$, $E_k^{(n)}$, $F_i^{(n)}$, and $G_i^{(n)}$ are determined only to within an arbitrary additive function of $\mathbf{K}$, $\mathbf{P}$, and $\mathbf{Q}$. We shall show that these arbitrary functions may be chosen so that the transformation (2.4) is canonical. For the moment let us assume that this is so. It is then straightforward to show that the canonical K_i's are constant to all orders in nondegenerate perturbation theory. First we write the Hamiltonian (2.1) in terms of the new variables and denote it by $h(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q})$. Since K_i and ϕ_i are assumed to be canonically conjugate variables to all orders, we must have that

$$\dot{K}_i = -\frac{\partial h(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q})}{\partial \phi_i}. \tag{2.6}$$

Upon comparing Eqs. (2.5c) and (2.6) we see that

$$\frac{\partial h}{\partial \phi_i} = -\sum_{n=1}^{\infty} \epsilon^n A_i^{(n)}(\mathbf{P}, \mathbf{Q}, \mathbf{K}). \tag{2.7}$$

Now, by construction, the Hamiltonian $h(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q})$ must be a periodic function of each of the ϕ_i's. Therefore, the left-hand side of Eq. (2.7) can contain no zero harmonic in ϕ_i. However, by construction the right-hand side of Eq. (2.7) is independent of ϕ_i. We conclude then that the $A_i^{(n)}$'s must vanish to all orders. Thus the canonical K_i's are constant to all orders in nondegenerate perturbation theory.

We must now demonstrate that the transformation (2.4) can be made canonical to all orders. In order to do this in a succinct fashion we introduce the following change in notation:

$$(K_1, K_2, \cdots, K_s) = (V_{-N}, V_{-N+1}, \cdots, V_{-N+s-1}), \tag{2.8a}$$

$$(\phi_1, \phi_2, \cdots, \phi_s) = (V_N, V_{N-1}, \cdots, V_{N-s+1}), \tag{2.8b}$$

$$(P_1, P_2, \cdots, P_r) = (V_{-N+s}, \cdots, V_{-1}), \tag{2.8c}$$

$$(Q_1, Q_2, \cdots, Q_r) = (V_{N-s}, \cdots, V_1), \tag{2.8d}$$

where $N = r + s$. The equations of motion (2.5) now have the form

$$\dot{V}_k = \epsilon a_k(V_{-N}, \cdots, V_r), \quad k = -N, \cdots, r, \tag{2.9a}$$

$$\dot{V}_j = \omega_j(V_{-N}, \cdots, V_r), \quad j = r+1, \cdots, N, \tag{2.9b}$$

where the a_k's and ω_j's are formal infinite series in powers of ϵ.

If the transformation is to be canonical, we must have

$$\{V_i, V_j\} = \operatorname{sgn} i\delta_{i,-j}, \quad i, j = -N, \cdots, N, \tag{2.10}$$

where the Poisson brackets are formed with respect to any complete set of canonical variables. We now show that the left-hand sides of Eq. (2.10) are independent of the angle variables V_j $(j = r+1, \cdots, N)$ no matter how we choose the arbitrary zero harmonics in the transformation (2.4). In order to do this we construct the $(N+r) \times (N+r)$-dimensional matrix L whose (i, k)th component L_{ik} is

$$L_{ik} = \{V_i, V_k\}, \quad i, k \neq r+1, \cdots, N. \tag{2.11}$$

We also construct the $[(N+r) \times (N+r)]$-dimensional matrix M whose (i, k)th component M_{ik} is

$$M_{ik} = \frac{\partial a_k}{\partial V_i}, \quad i, k \neq r+1, \cdots, N. \tag{2.12}$$

The time derivative of L_{ik} is

$$\begin{aligned} \dot{L}_{ik} &= \{\dot{V}_i, V_k\} + \{V_i, \dot{V}_k\} \\ &= \{\epsilon a_i, V_k\} + \{V_i, \epsilon a_k\}. \end{aligned} \tag{2.13}$$

Equation (2.13) may be rewritten as[12]

$$\dot{L}_{ik} = \epsilon \sum_{m=-N}^{r} \{V_m, V_k\} \frac{\partial a_i}{\partial V_m} + \epsilon \sum_{m=-N}^{r} \{V_i, V_m\} \frac{\partial a_k}{\partial V_m}. \tag{2.14}$$

It follows from Eq. (2.14) that

$$\dot{L} = \epsilon \tilde{M} L + \epsilon L M, \tag{2.15}$$

where $\tilde{M}$ is the transpose of the matrix M. It must also be true that

$$\dot{L} = \sum_{j=r+1}^{N} \omega_j \frac{\partial L}{\partial V_j} + \epsilon \sum_{k=-N}^{r} a_k \frac{\partial L}{\partial V_k}. \tag{2.16}$$

We find from Eqs. (2.15) and (2.16) that

$$\sum_{j=r+1}^{N} \omega_j \frac{\partial L}{\partial V_j} = \epsilon \tilde{M} L + \epsilon L M - \epsilon \sum_{k=-N}^{r} a_k \frac{\partial L}{\partial V_k}. \tag{2.17}$$

[12] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1959), p. 254, Eq. (8.50).

The matrix L must be a periodic function of the angle variables V_j $(j = r+1, \cdots, N)$. Let us suppose that L is known to be independent of the V_j's $(j = r+1, \cdots, N)$ through order ϵ^n. Thus,

$$L = L^{(n)} + \sum_{m=n+1}^{\infty} \epsilon^m L^{(m)} \tag{2.18}$$

where $L^{(n)}$ is independent of the V_j's $(j = r+1, \cdots, N)$. Upon substituting Eq. (2.18) into (2.17) and equating the $(n+1)$th-order terms on each side we find that

$$\sum_{j=r+1}^{N} \omega_j^0 \frac{\partial L^{(n+1)}}{\partial V_j} = \left(\epsilon \tilde{M} L^{(n)} + \epsilon L^{(n)} M - \epsilon \sum_{k=-N}^{r} a_k \frac{\partial L^{(n)}}{\partial V_k}\right)_{n+1}, \tag{2.19}$$

where the $\omega_j^0(V_{-N}, \cdots, V_{-r-1})$ are the unperturbed frequencies and where the right-hand side includes only those terms of order $n+1$. The point to observe is that the right-hand side of (2.19) is independent of the V_j's $(j = r+1, \cdots, N)$. Since $L^{(n+1)}$ must be a periodic function of the V_j's $(j = r+1, \cdots, N)$ the right-hand side of (2.19) can contain no zero harmonic and must therefore vanish. This leaves

$$\sum_{j=r+1}^{N} \omega_j^0 \frac{\partial L^{(n+1)}}{\partial V_j} = 0. \tag{2.20}$$

The homogeneous equation (2.20) has two possible types of solutions: either $L^{(n+1)}$ is independent of the V_j's $(j = r+1, \cdots, N)$ or there is a solution

$$L^{(n+1)} = \exp\left(i \sum_{j=r+1}^{N} p_j V_j\right), \tag{2.21}$$

where the p_k's are integers satisfying the identity

$$\sum_{j=r+1}^{N} p_j \omega_j^0 = 0. \tag{2.22}$$

In nondegenerate perturbation theory we assume that there are no sets of integers satisfying this identity except for the trivial set in which all the p_k's are zero. We conclude then that $L^{(n+1)}$ is independent of the V_j's $(j = r+1, \cdots, N)$. Thus if L is independent of the V_j's $(j = r+1, \cdots, N)$ through order ϵ^n then it is independent of the V_j's through order ϵ^{n+1}. Since the transformation (2.4) is the identity transformation through order ϵ^0, L is independent of the V_j's $(j = r+1, +\cdots, N)$ through order ϵ^0. It follows, therefore, by mathematical induction that L is independent of the V_j's $(j = r+1, \cdots, N)$ to all orders in nondegenerate perturbation theory.

We must now investigate those brackets which involve combinations of the angle and nonangle variables. To do this we construct the $(N+r) \times 1$ dimensional matrix $R^{(j)}$ whose ith component $R_{i1}^{(j)}$ is

$$R_i^{(j)} = \{V_i, V_j\}, \quad i \neq r+1, \cdots, N, \quad j = r+1, \cdots, N. \tag{2.23}$$

We also construct the $(N+r) \times 1$ dimensional matrix $S^{(j)}$ whose ith component $S_{i1}^{(j)}$ is

$$S_{i1}^{(j)} = \frac{\partial \omega_j}{\partial V_i}, \quad i \neq r+1, \cdots, N, \quad j = r+1, \cdots, N. \tag{2.24}$$

Proceeding as we did for L, we find that

$$\sum_{j=r+1}^{N} \omega_j \frac{\partial R^{(m)}}{\partial V_j} = LS + \epsilon \tilde{M} R^{(m)} - \epsilon \sum_{k=-N}^{r} a_k \frac{\partial R^{(m)}}{\partial V_k}. \tag{2.25}$$

The matrix $R^{(m)}$ must be periodic in the angle variables and is known to be independent of the angle variables through order ϵ^0. It follows, therefore, by the same arguments which were used on Eq. (2.17) that the $R^{(m)}$'s are independent of the V_j's $(j = r+1, \cdots, N)$ to all orders in nondegenerate perturbation theory.

We must now consider the remaining case of the brackets $\{V_i, V_j\}$ $(i, j = r+1, \cdots, N)$. Proceeding as before, we find that

$$\sum_{m=r+1}^{N} \omega_m \frac{\partial \{V_i, V_j\}}{\partial V_m} = \tilde{R}^{(j)} S^{(i)} - \tilde{S}^{(j)} R^{(i)} - \epsilon \sum_{k=-N}^{r} a_k \frac{\partial \{V_i, V_j\}}{\partial V_k}. \tag{2.26}$$

Since $\{V_i, V_j\}$ must be a periodic function of the angle variables and is known to be independent of the angle variables through order ϵ^0, it follows that $\{V_i, V_j\}$ $(i, j = r+1, \cdots, N)$ is independent of the angle variables V_k $(k = r+1, \cdots, N)$ to all orders in nondegenerate perturbation theory.

We have succeeded in showing that the Poisson brackets $\{V_i, V_j\}$ $(i, j = -N, \cdots, N)$ are independent of the angle variables V_k $(k = r+1, \cdots, N)$. This proof is independent of how we choose the arbitrary zero harmonic in the transformation (2.4). Thus we are at liberty to choose the arbitrary zero harmonics so that Eq. (2.10) is satisfied. In other words, we can choose the arbitrary zero harmonics so that the transformation (2.4) is canonical to all orders in nondegenerate perturbation theory.

Consider now the special case where the system has only angular degrees of freedom; in other words,

when the Hamiltonian (2.1) depends only on the J_i's and the ψ_i's. In this case, the K_i's are constant to all orders whether the transformation (2.4) is canonical or not. This is because the noncanonical K_i's are equal to the canonical K_i's plus some function of the canonical K_i's alone. Since the canonical K_i's are known to be constant it must be that all choices of the K_i's turn out to be constant to all orders. We also see that, in the case we are discussing here, the canonical K_i's and ϕ_i's are just the classical action and angle variables of the perturbed system. The solutions of the perturbed problem are all multiple-periodic to all orders in nondegenerate perturbation theory. Thus, in this special case, nondegenerate perturbation theory is equivalent to the nondegenerate classical perturbation theory of Poincaré. It is interesting to note that the perturbation theory of Poincaré seeks only the multiple-periodic solutions while our nondegenerate perturbation theory imposes no such restrictions. It simply turns out that, in this special case, nondegenerate perturbation theory yields only multiple-periodic solutions.

We conclude this discussion of nondegenerate invariants by briefly considering two other types of systems which commonly occur in practice. First we consider oscillatory systems whose Hamiltonians are slowly varying functions of some coordinates q and of the time t. We take the Hamiltonian to be of the form

$$H = H_0(\mathbf{J}, \epsilon\mathbf{q}, \epsilon t) + \epsilon H_1(\mathbf{J}, \boldsymbol{\psi}, \mathbf{p}, \mathbf{q}, \epsilon t), \quad (2.27)$$

where ϵ is a small parameter and H_1 is a periodic function of each of the ψ_i's ($i = 1, \cdots, s$). Here, as usual, J_i is the momentum conjugate to ψ_i and p_j is the momentum conjugate to q_j ($j = 1, \cdots, r$). Hamilton's equations of motion are found from (2.27) to be

$$\dot{J}_i = -\epsilon \frac{\partial H_1}{\partial \psi_i}, \quad i = 1, \cdots, s, \quad (2.28a)$$

$$\dot{p}_k = -\epsilon \frac{\partial H_0}{\partial \epsilon q_k} - \epsilon \frac{\partial H_1}{\partial q_k}, \quad k = 1, \cdots, r, \quad (2.28b)$$

$$\dot{q}_k = \epsilon \frac{\partial H_1}{\partial p_k}, \quad k = 1, \cdots, r, \quad (2.28c)$$

$$\dot{\psi}_i = \omega_i^0(\mathbf{J}, \epsilon\mathbf{q}, \epsilon t) + \epsilon \frac{\partial H_1}{\partial J_i}, \quad i = 1, \cdots, s, \quad (2.28d)$$

where

$$\omega_i^0 = \frac{\partial H_0}{\partial J_i}. \quad (2.29)$$

We perform nondegenerate perturbation theory on Eqs. (2.28) by introducing new variables $\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}$, and $\mathbf{Q}$ such that

$$p_k = P_k + \sum_{n=1}^{\infty} \epsilon^n D_k^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}, \epsilon t), \quad (2.30a)$$

$$q_k = Q_k + \sum_{n=1}^{\infty} \epsilon^n E_k^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}, \epsilon t), \quad (2.30b)$$

$$J_i = K_i + \sum_{n=1}^{\infty} \epsilon^n F_i^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}, \epsilon t), \quad (2.30c)$$

$$\psi_i = \phi_i + \sum_{n=1}^{\infty} \epsilon^n G_i^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}, \epsilon t). \quad (2.30d)$$

Here the $D_k^{(n)}$'s, $E_k^{(n)}$'s, $F_i^{(n)}$'s, and $G_i^{(n)}$'s are required to be periodic functions of each of the ϕ_i's. One should note that the transformation (2.30) is time-dependent. The P_k's, Q_k's, K_i's, and ϕ_i's are to satisfy the equations

$$\dot{P}_k = \sum_{n=1}^{\infty} \epsilon^n a_k^{(n)}(\mathbf{K}, \mathbf{P}, \mathbf{Q}, \epsilon t), \quad (2.31a)$$

$$\dot{Q}_k = \sum_{n=1}^{\infty} \epsilon^n b_k^{(n)}(\mathbf{K}, \mathbf{P}, \mathbf{Q}, \epsilon t), \quad (2.31b)$$

$$\dot{K}_i = \sum_{n=1}^{\infty} \epsilon^n A_i^{(n)}(\mathbf{K}, \mathbf{P}, \mathbf{Q}, \epsilon t), \quad (2.31c)$$

$$\dot{\phi}_i = \omega_i^0(\mathbf{K}, \epsilon\mathbf{Q}, \epsilon t) + \sum_{n=1}^{\infty} \epsilon^n B_i^{(n)}(\mathbf{K}, \mathbf{P}, \mathbf{Q}, \epsilon t). \quad (2.31d)$$

Equations (2.30) and (2.31) are simply an attempt to separate the slow secular motion from the rapidly fluctuating motion. We now require that the transformation (2.30) be canonical to all orders.[13] The Hamiltonian $h(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}, \epsilon t)$ which is appropriate to the new canonical variables $\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}$, and $\mathbf{Q}$ must be a periodic function of each of the ϕ_i's. This fact allows us to prove that the canonical K_i's ($i = 1, \cdots, s$) are constant to all orders in nondegenerate perturbation theory. Since K_i is canonically conjugate to ϕ_i, we must have that

$$\dot{K}_i = -\frac{\partial h}{\partial \phi_i}. \quad (2.32)$$

Upon comparing Eq. (2.32) with Eq. (2.31) we find that

$$\frac{\partial h}{\partial \phi_i} = -\sum_{n=1}^{\infty} \epsilon^n A_i^{(n)}(\mathbf{K}, \mathbf{P}, \mathbf{Q}, \epsilon t). \quad (2.33)$$

Since $h(\mathbf{K}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{Q}, \epsilon t)$ must be a periodic function of the ϕ_i's, we conclude from Eq. (2.33) that the $A_i^{(n)}$'s must vanish to all orders. We have, therefore, proved that the canonical K_i's are constant to all orders in nondegenerate perturbation theory. These invariants are usually called adiabatic invariants because of the adiabatic time dependence of the Hamiltonian (2.27).

[13] The proof that this can be done is a logical extension of the proof that (2.4) can be made canonical.

We conclude this section by discussing the invariants which arise when a nondegenerate oscillatory system is driven by a weak harmonic force. The Hamiltonian for such a system has the form

$$H = H_0(\mathbf{J}) + \epsilon H_1(\mathbf{J}, \boldsymbol{\psi}, \omega t), \tag{2.34}$$

where ϵ is a small parameter and H_1 is a periodic function of each of the ψ_i's $(i = 1, \cdots, s)$ and of ωt. Hamilton's equations of motion are found from Eq. (2.34) to be

$$\dot{J}_i = -\epsilon \frac{\partial H_1}{\partial \psi_i}, \tag{2.35a}$$

$$\dot{\psi}_i = \omega_i^0(\mathbf{J}) + \epsilon \frac{\partial H_1}{\partial J_i}, \tag{2.35b}$$

where

$$\omega_i^0 = \frac{\partial H_0}{\partial J_i}. \tag{2.36}$$

We can apply nondegenerate perturbation theory to Equations (2.35) by introducing new variables $\mathbf{K}$ and $\boldsymbol{\phi}$ such that

$$J_i = K_i + \sum_{n=1}^{\infty} \epsilon^n F_i^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \omega t), \tag{2.37a}$$

$$\psi_i = \phi_i + \sum_{n=1}^{\infty} \epsilon^n G_i^{(n)}(\mathbf{K}, \boldsymbol{\phi}, \omega t), \tag{2.37b}$$

where the $F_i^{(n)}$'s and $G_i^{(n)}$'s are required to be periodic functions of each of the ϕ_i's and of ωt. The time development of the K_i's and ϕ_i's is governed by the equations

$$\dot{K}_i = \sum_{n=1}^{\infty} \epsilon^n A_i^{(n)}(\mathbf{K}), \tag{2.38a}$$

$$\dot{\phi}_i = \omega_i^0(\mathbf{K}) + \sum_{n=1}^{\infty} \epsilon^n B_i^{(n)}(\mathbf{K}). \tag{2.38b}$$

We now require that the transformation (2.37) be canonical to all orders. Since the Hamiltonian $h(\mathbf{K}, \boldsymbol{\phi}, \omega t)$ which is appropriate to the new variables must be a periodic function of each of the ϕ_i's, it is straightforward to show that the canonical K_i's are invariant to all orders. Since K_i is canonically conjugate to ϕ_i, we must have that

$$\dot{K}_i = -\frac{\partial h}{\partial \phi_i}. \tag{2.39}$$

Upon comparing Eqs. (2.39) and (2.38a) we find that

$$\frac{\partial h}{\partial \phi_i} = -\sum_{n=1}^{\infty} \epsilon^n A_i^{(n)}(\mathbf{K}). \tag{2.40}$$

Since $h(\mathbf{K}, \boldsymbol{\phi}, \omega t)$ is a periodic function of the ϕ_i's, we conclude from Eq. (2.40) that the $A_i^{(n)}$'s are zero to all orders. Thus the K_i's are constant to all orders in nondegenerate perturbation theory.

3. DEGENERATE SYSTEMS

In I we pointed out that nondegenerate perturbation theory will often fail because certain linear combinations of the angle variables give rise to small divisors in the transformation (2.4). We showed that this problem of small divisions can be avoided by allowing the differential equations which describe the average motion to depend upon these degenerate combinations of the angle variables. We have called this modification degenerate perturbation theory. In degenerate perturbation theory, as in nondegenerate perturbation theory, the transformation to the average variables is determined only to within an arbitrary zero harmonic in the average angle variables. In nondegenerate perturbation theory we found that we could exploit this arbitrariness to prove that nondegenerate perturbation theory can be performed in a canonical fashion. Once we proved the existence of canonical average variables, it was then rather simple to demonstrate the existence of nondegenerate invariants. In degenerate perturbation theory, however, the situation is not so simple. Because of the appearance of the angle variables in the equations for the average motion, it is no longer clear that the existence of an arbitrary zero harmonic in the transformation to the average variables is sufficient to guarantee that this transformation can be made in a canonical fashion. We, therefore, have to content ourselves with somewhat weaker results in this section than those which were obtained in the previous section. Here we prove certain results about the Hamiltonian when it is written in terms of the average variables. These results imply the existence of degenerate invariants provided that the transformation to the average variables can be made canonical.

We again consider systems which are described by Hamiltonians of the form (2.1). Hamilton's equations of motion are given by Eqs. (2.2), and the transformation to the average variables P_k, Q_k $(k = 1, \cdots, r)$, and K_i, ϕ_i $(i = 1, \cdots, s)$ can be written as in Eqs. (2.4) where the $D_k^{(n)}$'s, $E_k^{(n)}$'s, $F_i^{(n)}$'s, and $G_i^{(n)}$'s are required to be periodic functions of each of the ϕ_i's.

We now specify that our system is m-fold degenerate $(m < s)$ through the m linearly-independent angular combinations

$$\theta_i = p_{i1}\phi_1 + p_{i2}\phi_2 + \cdots + p_{is}\phi_s, \quad i = 1, \cdots, m, \tag{3.1}$$

where the p_{ij}'s are integers not all of which are zero. In particular we assume that the set of integers a_{ij},

not all zero, for which

$$a_{i1}\omega_1^0 + a_{i2}\omega_2^0 + \cdots + a_{is}\omega_s^0 = 0, \tag{3.2}$$

is either the null set or is a subset of the p_{ij}'s. The equations of motion for the average variables take the form

$$\dot{P}_k = \epsilon a_k(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}), \tag{3.3a}$$

$$\dot{Q}_k = \epsilon b_k(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}), \tag{3.3b}$$

$$\dot{K}_i = \epsilon A_i(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}), \tag{3.3c}$$

$$\dot{\phi}_i = \omega_i(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}), \tag{3.3d}$$

where a_k, b_k, A_i, and ω_i, are formal infinite series in powers of ϵ with the lowest-order contribution to ω_i being the unperturbed frequency $\omega_i^0(\mathbf{K})$. Again let us point out that the $\boldsymbol{\theta}$ dependence has been included in Eqs. (3.3) simply to avoid the small divisors which would occur in nondegenerate perturbation theory because of the angular combinations given in (3.1).

We now prove that the Hamiltonian (2.1), when expressed in terms of the average variables $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{K}$, and $\boldsymbol{\phi}$, will depend upon the ϕ_i's only through the θ_j's defined by (3.1). To do this we replace the ϕ_j's ($j = 1, \cdots, s$) by the θ_i's ($i = 1, \cdots, m$) and the μ_k's ($k = m + 1, \cdots, s$) defined by

$$\mu_k = \phi_k, \quad k = m + 1, \cdots, s. \tag{3.4}$$

The θ_i's ($i = 1, \cdots, m$) together with the μ_k's ($k = m + 1, \cdots, s$) are equivalent to the set of ϕ_j's. The Hamiltonian (2.1), when expressed in terms of the new variables $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{K}$, $\boldsymbol{\theta}$, $\boldsymbol{\mu}$, is denoted by $h(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}, \boldsymbol{\mu})$. Since the Hamiltonian is autonomous, its time derivative must vanish. Thus we find that

$$\sum_{i=1}^{m}(p_{i1}\omega_1 + p_{i2}\omega_2 + \cdots + p_{is}\omega_s)\frac{\partial h}{\partial \theta_i} + \sum_{j=m+1}^{s}\omega_j\frac{\partial h}{\partial \mu_j} = -\sum_{i=1}^{s}\epsilon A_i\frac{\partial h}{\partial K_i} - \sum_{j=1}^{r}\epsilon\left\{a_j\frac{\partial h}{\partial P_j} + b_j\frac{\partial h}{\partial Q_j}\right\}. \tag{3.5}$$

Now let us suppose that h is known to be independent of the μ_j's ($j = m + 1, \cdots, s$) through order ϵ^n. In other words,

$$h = h^{(n)}(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}) + \sum_{m=n+1}^{\infty}\epsilon^m h^{(m)}(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}, \boldsymbol{\mu}). \tag{3.6}$$

When we substitute Eq. (3.6) into Eq. (3.5) we observe that the rhs of (3.5) is independent of $\boldsymbol{\mu}$ through order ϵ^{n+1}. However, the lhs of (3.5) can have an $(n + 1)$th order $\boldsymbol{\mu}$ dependence coming from the term

$$\sum_{i=1}^{m}[p_{i1}\omega_i^0(K) + p_{i2}\omega_2^0(K) + \cdots + p_{is}\omega_s^0(K)]\frac{\partial h^{(n+1)}}{\partial \theta_i} + \sum_{j=m+1}^{s}\omega_j^0(\mathbf{K})\frac{\partial h^{(n+1)}}{\partial \mu_j}. \tag{3.7}$$

The expression given by (3.7) must be independent of $\boldsymbol{\mu}$. Since $h^{(n+1)}$ must be a periodic function of the θ_i's and the μ_i's, expression (3.7) is independent of $\boldsymbol{\mu}$ only if $h^{(n+1)}$ is independent of $\boldsymbol{\mu}$ or if there exist some integers b_{kj}, not all of which are zero, such that

$$\sum_{i=1}^{m}b_{ki}[p_{i1}\omega_1^0 + p_{i2}\omega_2^0 + \cdots + p_{is}\omega_s^0] + \sum_{j=m+1}^{s}b_{kj}\omega_j^0 = 0. \tag{3.8}$$

However, Eq. (3.8) implies the existence of commensurabilities other than those given by Eq. (3.2). We have excluded this possibility. We, therefore, conclude that if $h^{(n+1)}$ is independent of $\boldsymbol{\mu}$ through order ϵ^n then it is independent of $\boldsymbol{\mu}$ through order ϵ^{n+1}. Now $h(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}, \boldsymbol{\mu})$ is certainly independent of $\boldsymbol{\mu}$ through order ϵ^0. It follows by mathematical induction that $h(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}, \boldsymbol{\mu})$ is independent of $\boldsymbol{\mu}$ through all orders of degenerate perturbation theory.

It is now quite simple to demonstrate the possible existence of degenerate invariants. If the transformation from the original variables $\mathbf{p}$, $\mathbf{q}$, $\mathbf{J}$, $\boldsymbol{\psi}$ to the average variables $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{K}$, $\boldsymbol{\phi}$ can be made canonical to all orders, then there are $s - m$ linearly independent combinations of the K_i's ($i = 1, \cdots, s$) which are constant to all orders in degenerate perturbation theory. These invariants are the momenta conjugate to the μ_j's defined earlier. They are constant because the Hamiltonian $h(\mathbf{P}, \mathbf{Q}, \mathbf{K}, \boldsymbol{\theta}, \boldsymbol{\mu})$ is cyclic in the μ_i's to all orders.

The discussions given above can be extended in a straightforward fashion to degenerate systems whose Hamiltonians contain an adiabatic time dependence or whose Hamiltonians contain a weak harmonic time dependence. In the latter case one must allow for forced resonance. In both these cases one is forced to prove rather weak results, namely that degenerate invariants exist provided that the transformation to the average variables can be made canonical.

4. THE ADIABATIC OSCILLATOR

As a simple illustration of the techniques which we have been considering we now discuss an oscillator whose frequency is a slowly varying function of time. In the $q - p$ representation, the Hamiltonian which describes the oscillator has the form

$$H_0 = \tfrac{1}{2}p^2 + \tfrac{1}{2}\omega^2(T)q^2, \tag{4.1}$$

where

$$T = \epsilon t, \tag{4.2}$$

is the slow time. The Hamiltonian (4.1) is not in the

standard form required for the application of perturbation theory. To put the Hamiltonian in standard form we introduce the new momentum

$$J_1 = (2\pi)^{-1} \oint p\, dq, \tag{4.3}$$

where the integral is performed over one period of the time-independent problem. In order to perform this integral we set

$$H_0 = W_0(J_1, T). \tag{4.4}$$

We find from (4.1) that

$$p = \pm(2W_0 - \omega^2 q^2)^{\frac{1}{2}}. \tag{4.5}$$

Upon setting

$$q = (2W_0/\omega^2)^{\frac{1}{2}} \sin\theta, \tag{4.6}$$

we find that

$$p = (2W_0)^{\frac{1}{2}} \cos\theta. \tag{4.7}$$

Thus

$$\begin{aligned} J_1 &= (2W_0/2\pi\omega)\int_0^{2\pi} \cos^2\theta\, d\theta \\ &= W_0/\omega. \end{aligned} \tag{4.8}$$

The function S_0, which generates the new canonical momentum J_1 and its conjugate coordinate ψ_1, is found from the differential equation

$$p = \frac{\partial S_0}{\partial q} = (2J_1\omega - \omega^2 q^2)^{\frac{1}{2}}. \tag{4.9}$$

This equation has the solution

$$S_0 = \tfrac{1}{2}[q(2J_1\omega - \omega^2 q^2)^{\frac{1}{2}} + 2J_1 \sin^{-1}(q^2\omega/2J_1)^{\frac{1}{2}}]. \tag{4.10}$$

The angle variable ψ_1 conjugate to J_1 is found from Eq. (4.10) to be

$$\psi_1 = \frac{\partial S_0}{\partial J_1} = \sin^{-1}\left(\frac{q^2\omega}{2J_1}\right)^{\frac{1}{2}}. \tag{4.11}$$

The Hamiltonian in the $J_1 - \psi_1$ representation is

$$\begin{aligned} H_1 &= W_0(J_1, T) + \epsilon \frac{\partial S_0}{\partial T} \\ &= J_1\omega(T) + \epsilon\left(\frac{\omega'}{2\omega}\right) J_1 \sin 2\psi_1, \end{aligned} \tag{4.12}$$

where

$$\omega' = \frac{d\omega}{dT}. \tag{4.13}$$

The Hamiltonian (4.12) is in the standard form required for application of the method of averaging. The equations of motion are found from (4.12) to be

$$\dot{J}_1 = -\epsilon(\omega'/\omega) J_1 \cos 2\psi_1, \tag{4.14a}$$

$$\dot{\psi}_1 = \omega(T) + \epsilon(\omega'/2\omega) \sin 2\psi_1. \tag{4.14b}$$

According to the remarks made in Sec. 2, we should seek solutions of (4.14) in the form

$$J_1 = K + \sum_{n=1}^{\infty} \epsilon^n F^{(n)}(K, \phi, T), \tag{4.15a}$$

$$\psi_1 = \phi + \sum_{n=1}^{\infty} \epsilon^n G^{(n)}(K, \phi, T), \tag{4.15b}$$

where the $F^{(n)}$'s and $G^{(n)}$'s are required to be periodic functions of ϕ. We further require that

$$\dot{K} = \sum_{n=1}^{\infty} \epsilon^n A^{(n)}(K, T), \tag{4.16a}$$

$$\dot{\phi} = \omega(T) + \sum_{n=1}^{\infty} \epsilon^n B^{(n)}(K, T). \tag{4.16b}$$

These equations when substituted into (4.14) produce an infinite set of coupled differential equations. The first-order members of this set are

$$A^{(1)} + \omega \frac{\partial F^{(1)}}{\partial \phi} = -\left(\frac{\omega'}{\omega}\right) K \cos 2\phi, \tag{4.17a}$$

$$B^{(1)} + \omega \frac{\partial G^{(1)}}{\partial \phi} = \left(\frac{\omega'}{2\omega}\right) \sin 2\phi. \tag{4.17b}$$

The solutions of Eqs. (4.17) are

$$A^{(1)} = B^{(1)} = 0, \tag{4.18}$$

$$F^{(1)} = -(\omega'/2\omega^2) K \sin 2\phi, \tag{4.19a}$$

$$G^{(1)} = -(\omega'/4\omega^2) \cos 2\phi. \tag{4.19b}$$

The transformation determined by (4.19) is canonical through first order. According to the results proved in Sec. 2, K must be constant through order ϵ. This is easily seen to be true since $A^{(1)} = 0$. However, we can show that K is constant through first order in a more illuminating way. Let us first invert the transformation (4.15). We find that

$$K = J_1 + \epsilon(\omega'/2\omega^2) J_1 \sin 2\psi_1 + O(\epsilon^2). \tag{4.20}$$

Now let us introduce the new momentum

$$J_2 = (2\pi)^{-1} \int_0^{2\pi} J_1\, d\psi_1. \tag{4.21}$$

If we set

$$H_1 = W_1(J_2, T), \tag{4.22}$$

we find that

$$J_2 = \frac{W_1}{\omega[1 - (\epsilon\omega'/2\omega^2)^2]^{\frac{1}{2}}}. \tag{4.23}$$

The generating function S_1 which produces the momentum J_2 and its canonically conjugate coordinate ψ_2 from J_1 and ψ_1 is found from the differential

equation

$$J_1 = \frac{\partial S_1}{\partial \psi_1} = \frac{W_1}{\omega[1 + (\epsilon\omega'/2\omega^2)\sin 2\psi_1]}. \quad (4.24)$$

This equation has the solution

$$S_1 = J_2 \tan^{-1}\left\{\frac{\tan\psi_1 + \epsilon\omega'/2\omega^2}{[1 - (\epsilon\omega'/2\omega^2)^2]^{\frac{1}{2}}}\right\}. \quad (4.25)$$

The variable ψ_2 conjugate to J_2 is

$$\psi_2 = \frac{\partial S_1}{\partial J_2} = \tan^{-1}\left\{\frac{\tan\psi_1 + \epsilon\omega'/2\omega^2}{[1 - (\epsilon\omega'/2\omega^2)^2]^{\frac{1}{2}}}\right\}. \quad (4.26)$$

The Hamiltonian in the $J_2 - \psi_2$ representation is

$$H_2 = W_1(J_2, T) + \epsilon\frac{\partial S_1}{\partial T} = J_2\left\{\omega\left[1 - \left(\frac{\epsilon\omega'}{2\omega^2}\right)^2\right]^{\frac{1}{2}} + \epsilon^2\frac{d(\omega'/2\omega^2)/dT}{1 + \tan^2\psi_2} \times \left(\frac{1}{[1 - (\epsilon\omega'/2\omega^2)^2]^{\frac{1}{2}}} + \epsilon\frac{(\omega'/2\omega^2)\tan\psi_2}{1 - (\epsilon\omega'/2\omega^2)^2}\right)\right\}. \quad (4.27)$$

The point to observe here is that $\dot{J}_2 = -\partial H_2/\partial\psi_2$ is rigorously of order ϵ^2. Now from (4.23) and (4.12) we observe that

$$J_2 = J_1 + \epsilon(\omega'/2\omega^2)J_1 \sin 2\psi_1 + O(\epsilon^2). \quad (4.28)$$

Upon comparing (4.20) with (4.28) we see that

$$K = J_2 + O(\epsilon^2). \quad (4.29)$$

In other words, K defined by (4.20) is the first-order approximant to J_2 which is rigorously constant through order ϵ.

Now let us proceed to second-order perturbation theory. The second-order perturbation equations are

$$A^{(2)} + \omega\frac{\partial F^{(2)}}{\partial\phi} - \frac{\partial}{\partial T}\left(\frac{\omega'}{2\omega^2}\right)K\sin 2\phi = 0, \quad (4.30a)$$

$$B^{(2)} + \omega\frac{\partial G^{(2)}}{\partial\phi} - \frac{\partial}{\partial T}\left(\frac{\omega'}{4\omega^2}\right)\cos 2\phi = -(\omega'/2\omega)(\omega'/2\omega^2)\cos^2 2\phi. \quad (4.30b)$$

These equations have the solutions

$$A^{(2)} = 0, \quad (4.31a)$$

$$B^{(2)} = -\left(\frac{\omega'}{4\omega}\right)\left(\frac{\omega'}{2\omega^2}\right), \quad (4.31b)$$

$$F^{(2)} = -\frac{1}{2\omega}\frac{d}{dT}\left(\frac{\omega'}{2\omega^2}\right)K\cos 2\phi + \frac{1}{4}\left(\frac{\omega'}{\omega^2}\right)^2 K, \quad (4.32a)$$

$$G^{(2)} = \frac{1}{4\omega}\frac{d}{dT}\left(\frac{\omega'}{2\omega^2}\right)\sin 2\phi - \frac{1}{16}\left(\frac{\omega'}{\omega^2}\right)^2\sin 4\phi. \quad (4.32b)$$

The zero harmonic of $F^{(2)}$ has been chosen so that the transformation is canonical through second order. The transformation (4.15a) through second order is

$$J_1 = K - \epsilon\left(\frac{\omega'}{2\omega^2}\right)K\sin 2\phi - \epsilon^2\left\{\frac{1}{2\omega}\frac{d}{dT}\left(\frac{\omega'}{2\omega^2}\right)K\cos 2\phi - \frac{1}{4}\left(\frac{\omega'}{\omega^2}\right)^2 K\right\} + O(\epsilon^3), \quad (4.33a)$$

$$\psi_1 = \phi - \epsilon\left(\frac{\omega'}{4\omega^2}\right)\cos 2\phi + \epsilon^2\left\{\frac{1}{4\omega}\frac{d}{dT}\left(\frac{\omega'}{2\omega^2}\right)\sin 2\phi - \frac{1}{16}\left(\frac{\omega'}{\omega^2}\right)^2\sin 4\phi\right\} + O(\epsilon^3). \quad (4.33b)$$

The quantity K defined by (4.33) is constant through order ϵ^2. This is seen by the fact that $A^{(1)} = A^{(2)} = 0$. We can gain some further insight into the nature of K by introducing the momentum

$$J_3 = (2\pi)^{-1}\int_0^{2\pi} J_2\, d\psi_2. \quad (4.34)$$

Upon setting $H_2 = W_2(J_3, T)$ we find that

$$J_3 = \frac{2W_2}{\omega[1 - (\epsilon\omega'/2\omega^2)^2]^{\frac{1}{2}}(4a - b^2)^{\frac{1}{2}}}, \quad (4.35)$$

where

$$a = 1 + \epsilon^2\frac{d(\omega'/2\omega^2)/dT}{\omega[1 - (\epsilon\omega'/2\omega^2)]}, \quad (4.36a)$$

$$b = \epsilon^3\frac{(\omega'/2\omega^3)\, d(\omega'/2\omega^2)/dT}{[1 - (\epsilon\omega'/2\omega^2)]^{\frac{3}{2}}}. \quad (4.36b)$$

The quantity J_3 is defined such that $\dot{J}_3$ is rigorously of order ϵ^3. If we expand (4.35) and retain terms through second order, then we find that

$$J_3 = J_1\left\{1 + \epsilon\left(\frac{\omega'}{2\omega^2}\right)\sin 2\psi_1 + \epsilon^2\left[\frac{1}{2\omega}\frac{d}{dT}\left(\frac{\omega'}{2\omega^2}\right)\cos 2\psi_1 + \frac{1}{2}\left(\frac{\omega'}{2\omega^2}\right)^2\right]\right\} + O(\epsilon^3). \quad (4.37)$$

Now let us invert the transformation (4.33). We find that

$$K = J_1\left\{1 + \epsilon\left(\frac{\omega'}{2\omega^2}\right)\sin 2\psi_1 + \epsilon^2\left[\frac{1}{2\omega}\frac{d}{dT}\left(\frac{\omega'}{2\omega^2}\right)\cos 2\psi_1 + \frac{1}{2}\left(\frac{\omega'}{2\omega^2}\right)^2\right]\right\} + O(\epsilon^3). \tag{4.38}$$

In other words, K is the second-order approximant to the quantity J_3 which is known to be rigorously constant through second order.

This process can be continued by defining a sequence of canonical transformations such that[14]

$$J_n = (2\pi)^{-1}\int_0^{2\pi} J_{n-1}\, d\psi_{n-1}. \tag{4.39}$$

This sequence is such that $\dot{J}_n = O(\epsilon^n)$. Furthermore, canonical perturbation theory carried through order ϵ^{n-1} produces the $(n-1)$th approximant to J_n. We see then that we have two ways of producing quantities which are invariant to any desired order. One way is to do perturbation theory, the other is to construct the sequence of J_n's. Both techniques become rather tedious at higher orders. Perturbation theory, however, involves much simpler integrations than does the construction of the J_n's. This is evident even in the simple problem which we have discussed here.

We conclude this section by choosing an explicit time dependence for $\omega(\epsilon t)$ and by discussing the sense in which the perturbation solution represents the exact solution. We choose ω such that

$$\omega^2 = \omega_0^2(1-u) + \omega_1^2 u - \tfrac{1}{4}(\eta^2 + \epsilon^2)u(1-u), \tag{4.40}$$

where

$$u = (1 + e^{-\epsilon t})^{-1}. \tag{4.41}$$

In order that $\omega^2 > 0$ for all times we require that

$$\eta^2 + \epsilon^2 < (\omega_0 + \omega_1)^2. \tag{4.42}$$

The oscillator described by (4.40) is such that its frequency changes in a continuous fashion from an initial ($t = -\infty$) constant value ω_0 to a final ($t = +\infty$) constant value ω_1.

The differential equation with which we are concerned is

$$\frac{d^2x}{dt^2} + [\omega_0^2(1-u) + \omega_1^2 u - \tfrac{1}{4}(\eta^2 + \epsilon^2)u(1-u)]x = 0. \tag{4.43}$$

This equation has been discussed by Eckart[15] in connection with the penetration of electrons through a potential barrier, by Epstein[16] in connection with wave propagation in inhomogeneous media, and most recently by Backus, Lenard, and Kulsrud[17] in connection with the adiabatic invariance of the magnetic moment. It is found that the solutions of (4.43) are expressible in terms of the hypergeometric function. We choose the particular solution[18]

$$x = A\,\mathrm{Im}\,[u^{i\omega_0/\epsilon}(1-u)^{-i\omega_1/\epsilon}F(a;b;c;u)], \tag{4.44}$$

where A is a real constant, $F(a;b;c;u)$ is the hypergeometric function, and

$$a = \frac{1}{2} + \frac{i}{2}\frac{2\omega_0 - 2\omega_1 + \eta}{\epsilon}, \tag{4.45}$$

$$b = \frac{1}{2} + \frac{i}{2}\frac{2\omega_0 - 2\omega_1 - \eta}{\epsilon}, \tag{4.46}$$

$$c = 1 + \frac{2i\omega_0}{\epsilon}. \tag{4.47}$$

This solution has the property that

$$\lim_{t\to-\infty} x = A\sin\omega_0 t. \tag{4.48}$$

We must now examine the perturbation solution. We have from (4.11) that

$$x = (2J/\omega)^{\frac{1}{2}}\sin\psi. \tag{4.49}$$

Upon making use of (4.15) and (4.19) we find that

$$J = K\left(1 - \epsilon\frac{\omega'}{2\omega^2}\sin 2\phi\right) + O(\epsilon^2), \tag{4.50a}$$

$$\psi = \phi - \epsilon\frac{\omega'}{4\omega^2}\cos 2\phi + O(\epsilon^2). \tag{4.50b}$$

If we substitute (4.50a) and (4.50b) into (4.49) and retain terms through order ϵ, we find that the perturbation solution x_p is given by

$$x_p = \left(\frac{2K}{\omega}\right)^{\frac{1}{2}}\left[\sin\phi - \epsilon u(1-u) \times \frac{4\omega_1^2 - 4\omega_0^2 + \eta^2(2u-1)}{32\omega^3}\cos\phi\right] + O(\epsilon^2), \tag{4.51}$$

where

$$\phi = \int \omega\, dt + O(\epsilon^2). \tag{4.52}$$

In order to compare (4.51) with (4.44) we must evaluate ϕ. This can best be done by introducing u

[14] The hierarchy of invariants J_n generated by this process is mathematically equivalent to the hierarchy of invariants produced by Gardner (Ref. 5). However, our technique for generating the hierarchy is different from that used by Gardner.

[15] C. Eckart, Phys. Rev. **35**, 1303 (1930).

[16] P. S. Epstein, Proc. Natl. Acad. Sci. **16**, 627 (1930).

[17] G. Backus, A. Lenard, and R. Kulsrud, Z. Naturforsch. **15a**, 1007 (1960).

[18] By Im we mean "imaginary part of."

as the independent variable. We find that

$$\phi = \frac{1}{\epsilon}\left(\int \frac{\omega}{u}\,du + \int \frac{\omega\,du}{1-u}\right). \tag{4.53}$$

The integrals involved in (4.53) are easily evaluated. The result is

$$\begin{aligned}\phi = {} & \frac{\omega_0}{\epsilon}\log u - \frac{\omega_1}{\epsilon}\log(1-u) \\ & - \frac{\eta}{2\epsilon}\log\left[\frac{2\omega_0^2 - 2\omega_1^2 + \frac{1}{2}\eta^2 - \eta^2 u - 2\eta\omega}{2(\omega_0 - \omega_1 - \frac{1}{2}\eta)(\omega_0 + \omega_1 - \frac{1}{2}\eta)}\right] \\ & - \frac{\omega_1}{\epsilon}\log\left[\frac{(\omega_0+\omega_1+\frac{1}{2}\eta)(\omega_0+\omega_1-\frac{1}{2}\eta)}{2\omega_1\omega + 2\omega_1^2 - (\omega_1^2 - \omega_0^2 + \frac{1}{4}\eta^2)(1-u)}\right] \\ & + \frac{\omega_0}{\epsilon}\log\left[\frac{4\omega_0^2}{(2-u)\omega_0^2 + \omega_1^2 u - \frac{1}{4}\eta^2 u + 2\omega_0\omega}\right],\end{aligned} \tag{4.54}$$

where the integration constant has been chosen such that

$$\lim_{t\to-\infty}\phi = \omega_0 t. \tag{4.55}$$

When we substitute (4.54) into (4.51) we find that

$$\begin{aligned}x_p = {} & \left(\frac{2K}{\omega}\right)^{\frac{1}{2}} \operatorname{Im}\left\{u^{i\omega_0/\epsilon}(1-u)^{-i\omega_1/\epsilon}e^{if(u)}\right. \\ & \left.\times\left[1 - \epsilon u(1-u)\frac{4\omega_1^2 - 4\omega_0^2 + \eta^2(2u-1)}{32\omega^3}\,e^{i\pi/2}\right]\right\},\end{aligned} \tag{4.56}$$

where

$$\begin{aligned}f(u) = {} & \frac{\omega_0}{\epsilon}\log\left[\frac{4\omega_0^2}{(2-u)\omega_0^2 + \omega_1^2 u - \frac{1}{4}\eta^2 u + 2\omega_0\omega}\right] \\ & - \frac{\omega_1}{\epsilon}\log\left[\frac{(\omega_0+\omega_1+\frac{1}{2}\eta)(\omega_0+\omega_1-\frac{1}{2}\eta)}{2\omega_1\omega + 2\omega_1^2 - (\omega_1^2 - \omega_0^2 + \frac{1}{4}\eta^2)(1-u)}\right] \\ & - \frac{\eta}{2\epsilon}\log\left[\frac{2\omega_0^2 - 2\omega_1^2 + \frac{1}{2}\eta^2 - \eta^2 u - 2\eta\omega}{2(\omega_0 - \omega_1 - \frac{1}{2}\eta)(\omega_0 + \omega_1 - \frac{1}{2}\eta)}\right].\end{aligned} \tag{4.57}$$

We see immediately that (4.56) agrees with the exact solution (4.44) at $t = -\infty$, provided that we choose K such that

$$(2K)^{\frac{1}{2}} = \omega_0^{\frac{1}{2}} A. \tag{4.58}$$

In order to investigate the connection between (4.56) and (4.44) for $-\infty < t \leq \infty$, we must examine the hypergeometric function. Clearly, as $\epsilon \to 0$, the constants a, b, and c develop large imaginary parts. We should, therefore, seek the asymptotic expansion of the hypergeometric function with large constants. In order to do this we employ the integral representation of the hypergeometric function.[19] In our case

$$F(a;b;c;u) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 g(t)e^{ih(t)/\epsilon}\,dt, \tag{4.59}$$

where

$$g(t) = [t(1-t)(1-tu)]^{-\frac{1}{2}}, \tag{4.60}$$

$$h(t) = \log[t^\beta(1-t)^{\gamma-\beta}(1-tu)^\alpha], \tag{4.61}$$

and

$$\alpha = \omega_0 - \omega_1 + \tfrac{1}{2}\eta, \tag{4.62a}$$

$$\beta = \omega_0 - \omega_1 - \tfrac{1}{2}\eta, \tag{4.62b}$$

$$\gamma = 2\omega_0. \tag{4.62c}$$

If $h(t)$ is such that

$$\frac{dh}{dt} = \frac{\beta}{t} - \frac{\gamma-\beta}{1-t} + \frac{\alpha u}{1-tu} \tag{4.63}$$

is zero at some point $0 < \tau < 1$, then we can employ the method of stationary phase to evaluate the integral in (4.59) as $\epsilon \to 0$.[20] Clearly when $\beta > 0$, $(0 \leq u \leq 1)$, there is a stationary point in the interval of integration. The case when $\beta < 0$ can be approached by the method of steepest descents. However, except for some rather special values of ω_0 and ω_1, the contours involved are rather complex. Therefore, we restrict our discussion to the case $\beta > 0$. This case encompasses an infinite number of oscillatory systems and is certainly adequate for a discussion of the connection between the perturbation solution and the exact solution.

The stationary point is readily found from (4.63) to be

$$\tau = \frac{2\omega_0 - \eta u - 2\omega}{2(\omega_0 + \omega_1 - \frac{1}{2}\eta)u}. \tag{4.64}$$

In general, the primary contribution to an integral like (4.59) comes from the end points and from the stationary point. In our case the end points do not contribute. We have, therefore, only to consider the contribution from the stationary point. If we introduce a new variable w such that

$$w^2 = h(\tau) - h(t), \tag{4.65}$$

then the integral in (4.59) becomes

$$\begin{aligned}\Phi &= \int_0^1 g(t)e^{ih(t)/\epsilon}\,dt \\ &= e^{ih(\tau)/\epsilon}\int_{w_1}^{w_2} g(t(w))e^{-iw^2/\epsilon}\frac{dt}{dw}\,dw,\end{aligned} \tag{4.66}$$

where

$$w_1 = -(h(\tau) - h(0))^{\frac{1}{2}}, \tag{4.67a}$$

$$w_2 = (h(\tau) - h(1))^{\frac{1}{2}}. \tag{4.67b}$$

[19] See P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., Inc., New York, 1953), p. 591, Eq. (5.3.16).

[20] For a discussion of the method of stationary phase, see E. T. Copson, *Asymptotic Expansions* (Cambridge University Press, London, England, 1965). See also A. Erdélyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956).

By the usual methods[21] we can invert (4.65) to express t as a power series in w. The result is

$$t = \tau + a_0 w + a_1 w^2 + a_2 w^3 + \cdots, \quad (4.68)$$

where

$$a_0 = [2/-h''(\tau)]^{\frac{1}{2}}, \quad (4.69a)$$

$$a_1 = 0, \quad (4.69b)$$

$$a_2 = a_0\left\{\frac{5}{36}\frac{[h^{(3)}(\tau)]^2}{[-h''(\tau)]^3} + \frac{1}{12}\frac{h^{(4)}(\tau)}{(h''(\tau))^2}\right\}. \quad (4.69c)$$

We now substitute (4.68) into (4.66) and extend the limits of integration from $-\infty$ to $+\infty$, since the contribution to the integral comes only from the neighborhood of the stationary point. The result is

$$\Phi = (\pi\epsilon)^{\frac{1}{2}} g(\tau) a_0 \exp i\left(\frac{h(\tau)}{\epsilon} - \frac{\pi}{4}\right) \times \left\{1 - \frac{\epsilon}{2}\left[\frac{3a_2}{a_0} + \frac{1}{2}\frac{g''(\tau)}{g(\tau)} a_0^2\right] \exp\frac{i\pi}{2} + O(\epsilon^2)\right\}. \quad (4.70)$$

We must now examine the coefficient of Φ in (4.59). Using the asymptotic expansion of $\Gamma(z)$ for large $|z|$, we find that[22]

$$\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} = \left(\frac{\gamma}{2\pi\epsilon}\right)^{\frac{1}{2}}\left[1 - \frac{i\epsilon}{12}\left(\frac{1}{\gamma} + \frac{1}{2\beta} + \frac{1}{2(\gamma-\beta)}\right) + O(\epsilon^2)\right] \times \exp\left[i\frac{\gamma}{\epsilon}\log\left(\frac{\gamma}{\gamma-\beta}\right) - i\frac{\beta}{\epsilon}\log\left(\frac{\beta}{\gamma-\beta}\right) + i\frac{\pi}{4}\right]. \quad (4.71)$$

If we now substitute (4.70) and (4.71) into (4.59) and retain terms through order ϵ, we find that

$$F(a; b; c; u) = \omega_0^{\frac{1}{2}} g(\tau) a_0 [1 - \epsilon B e^{i\pi/2} + O(\epsilon^2)] \exp\frac{i}{\epsilon} \times \left\{\omega_0 \log\left[\frac{4\omega_0^2}{(\omega_0+\omega_1+\frac{1}{2}\eta)(\omega_0-\omega_1-\frac{1}{2}\eta)}\frac{\tau(1-\tau)}{1-\tau u}\right] - \omega_1 \log\left[\frac{\omega_0+\omega_1+\frac{1}{2}\eta}{\omega_0-\omega_1-\frac{1}{2}\eta}\frac{\tau}{(1-\tau)(1-\tau u)}\right] - \frac{\eta}{2}\log\left[\frac{\omega_0+\omega_1+\frac{1}{2}\eta}{\omega_0-\omega_1-\frac{1}{2}\eta}\frac{\tau(1-\tau u)}{1-\tau}\right]\right\}, \quad (4.72)$$

where

$$B = -\frac{12\omega_0^2 - 4\omega_1^2 - 4\omega_1\eta - \eta^2}{24\omega_0(4\omega_0^2 - 4\omega_1^2 - 4\omega_1\eta - \eta^2)} + \frac{1}{2}\left[\frac{3a_2}{a_0} + \frac{1}{2}\frac{g''(\tau)}{g(\tau)} a_0^2\right]. \quad (4.73)$$

[21] See Ref. 19, pp. 411ff.
[22] See Ref. 19, p. 443.

Upon making use of (4.64) we are able, after some rather tedious algebra, to make the following identifications:

$$g(\tau)a_0(\tau) = \omega_0^{-\frac{1}{2}}, \quad (4.74)$$

$$\frac{\tau(1-\tau)}{1-\tau u} = \frac{(\omega_0+\omega_1+\frac{1}{2}\eta)(\omega_0-\omega_1-\frac{1}{2}\eta)}{(2-u)\omega_0^2 + \omega_1^2 u - \frac{1}{4}\eta^2 u + 2\omega_0\omega}, \quad (4.75)$$

$$\frac{\tau}{(1-\tau)(1-\tau u)} = \frac{(\omega_0-\omega_1-\frac{1}{2}\eta)(\omega_0+\omega_1-\frac{1}{2}\eta)}{2\omega_1\omega + 2\omega_1^2 - (\omega_1^2 - \omega_0^2 + \frac{1}{4}\eta^2)(1-u)}, \quad (4.76)$$

$$\frac{\tau(1-\tau u)}{1-\tau} = \frac{2\omega_0^2 - 2\omega_1^2 + \frac{1}{2}\eta^2 - \eta^2 u - 2\eta^2\omega}{2(\omega_0+\omega_1+\frac{1}{2}\eta)(\omega_0+\omega_1-\frac{1}{2}\eta)}, \quad (4.77)$$

$$B = u(1-u)\frac{4\omega_1^2 - 4\omega_0^2 + \eta(2u-1)}{32\omega^3}. \quad (4.78)$$

Upon substituting this information into (4.72) we find that the asymptotic expansion of the exact solution is

$$x = A(\omega_0/\omega)^{\frac{1}{2}} \operatorname{Im}\left\{u^{i\omega_0/\epsilon}(1-u)^{-i\omega_1/\epsilon} e^{if(u)} \times \left[1 - \epsilon u(1-u)\frac{4\omega_1^2 - 4\omega_0^2 + \eta^2(2u-1)}{32\omega^3} \times e^{i\pi/2} + O(\epsilon^2)\right]\right\}, \quad (4.79)$$

where $f(u)$ is given by (4.57). The constant A is related to K by (4.58). We see then that (4.79) is identical, through first order, with the perturbation solution. Perturbation theory is simply constructing the asymptotic expansion of the exact solution. Perhaps the most interesting aspect of this example is that the perturbation solution is asymptotic to the exact solution for all times. This is a much stronger result than is obtained from the usual theorems of asymptotic convergence. These theorems generally conclude that the perturbation solution is asymptotic to the exact solution for times of order $1/\epsilon$. Our discussion of the adiabatic oscillator indicates that there are classes of differential equations for which the general theorems of asymptotic convergence can be improved.

We conclude our discussion with a few remarks about the final state of the oscillator. The exact solution of the final state is

$$x = A \operatorname{Im}\{e^{i\omega_1 t} F(a; b; c; 1)\}. \quad (4.80)$$

By using the analytic continuation of $F(a; b; c; u)$ through the singular point at $u = 1$, we can write (4.80) in the form[23]

$$x = A \operatorname{Im} \left\{ \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} e^{i\omega_1 t} + \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} e^{-i\omega_1 t} \right\}. \quad (4.81)$$

If we now examine this expression, we find that perturbation theory is giving us only the contribution from the first term in (4.81). The reason why we do not get a contribution from the second term is easily found by examining the magnitude of this term. Using the relations

$$\Gamma(z)\Gamma(-z) = -\pi/z \sin(\pi z),$$
$$\Gamma(\tfrac{1}{2}+z)\Gamma(\tfrac{1}{2}-z) = \pi/\cos(\pi z),$$

it is straightforward to show that

$$\left| \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} \right|^2 = \frac{\omega_0}{\omega_1} \times \frac{\cosh\left(\frac{\pi}{\epsilon}\right)\left(\omega_0 - \omega_1 + \frac{\eta}{2}\right) \cosh\left(\frac{\pi}{\epsilon}\right)\left(\omega_0 - \omega_1 - \frac{\eta}{2}\right)}{\sinh(2\pi\omega_0/\epsilon) \sinh(2\pi\omega_1/\epsilon)}. \quad (4.82)$$

If we now consider the limit of (4.82) as $\epsilon \to 0$, we find that

$$\lim_{\epsilon \to 0} \left| \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} \right|^2 = \frac{\omega_0}{\omega_1} e^{-4\pi\omega_1/\epsilon}. \quad (4.83)$$

This expression vanishes faster than any power of ϵ. Perturbation theory, which orders things in powers of ϵ, would never construct such a function. We have here a typical example of how perturbation theory can construct quantities which are invariant to all orders in ϵ and yet are not rigorous constants of the motion.

[23] See Ref. 19, p. 546, Eq. (5.2.49).

5. CONCLUSION

We have discussed the application of the method of averaging as developed in I to nearly multiple-periodic Hamiltonian systems. In the case of nondegenerate systems we have demonstrated the existence of a quantity which is invariant to all orders in perturbation theory for each angular degree of freedom. These invariants resulted directly from the fact that nondegenerate perturbation theory can be made canonical to all orders. In the case where the system has an m-fold degeneracy we have shown that the Hamiltonian, when expressed in terms of the average variables, depends on the angle variables only through their m degenerate combinations. Thus if there are s angular degrees of freedom, then there will be $s - m$ invariants to all orders provided that degenerate perturbation theory can be made canonical to all orders. However, the question of when degenerate perturbation theory can be made canonical is left unanswered.

The techniques which were developed were applied to the simple problem of a harmonic oscillator whose frequency varies slowly with time. It was shown that the perturbation theory, when applied to a special time-dependent oscillator, produced a solution which was asymptotic to the exact solution for all time.

ACKNOWLEDGMENT

The author wishes to acknowledge many helpful discussions with Professor G. W. Ford.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Isoperimetric Solutions Related to the Thermodynamics of Plasmas

ROBERT L. W. CHEN AND THOMAS H. NEIGHBORS III*
Physics Department, Emory University, Atlanta, Georgia

(Received 22 July 1968)

More solutions of isoperimetric problems are obtained which lead to several extensions and refinements of previous results in the thermodynamics of plasmas: (1) Energy requirement for density fluctuations in a relativistic plasma is determined. At the typical relativistic temperature of $kT = mc^2$ it becomes $\sim nkT(|\overline{\Delta n}|/n)^2$ as compared with the previous nonrelativistic value of $\sim\frac{2}{5}nkT(|\overline{\Delta n}|/n)^2$. (2) An optimum expression for the free energy in a plasma is derived in terms of familiar thermodynamic variables. (3) For a collisional plasma, the energy requirement again assumes the same general form and is equal to $\frac{1}{2}nkT(|\overline{\Delta n}|/n)^2$.

INTRODUCTION

The thermodynamic approach to plasma physics[1–6] makes use of the constraints inherent in the equation of motion for a plasma. The existence of constraints imposes bounds on important quantities of physical interest. Information of this kind can be useful in plasma research, particularly in the area of thermonuclear work. We do not consider the physical aspects of the theory here as this has already been done in previous papers on the subject. This note only reports on some mathematical solutions which lead to several extensions or refinements of earlier results. We are dealing with the following problems. First, it was previously shown that a plasma obeying the nonrelativistic Vlasov equation requires a certain minimum amount of energy associated with its nonlinear density fluctuations. If Liouville's theorem on the conservation of phase space is used as the constraint, this lower bound on the energy is found to be $\sim\frac{2}{5}nkT(|\overline{\Delta n}|/n)^2$. The question then arises: For very-high-temperature plasmas ($kT \approx mc^2$) how would this expression modify in accordance with its relativistic nature? We surmised that the quadratic dependence on $|\overline{\Delta n}|$ would be retained, but the numerical factor may be changed. As it turns out, $\Delta E \cong nkT(|\overline{\Delta n}|/n)^2$ for $kT = mc^2$ with a numerical factor about *twice* that of the nonrelativistic case. Second, for a plasma which obeys the H-theorem, we seek the optimum bound on its free energy. The expression we found is in terms of the initial entropy, energy, volume, and the number of particles of the plasma:

$$E_0 - (3m/4\pi)N^{\frac{5}{3}}V^{-\frac{2}{3}} \exp[(2S_0/3kn) - 1].$$

It in fact corresponds to the minimum value of the plasma Lyapunov function of Fowler. Third, similar to the first problem, we consider the modification of the energy expression when particle collisions are taken into account. We again obtain the quadratic dependence on $|\overline{\Delta n}|$, which confirms our expectation that this is a general feature of the plasma, not at all a peculiarity connected with the nonrelativistic collisionless plasma treated previously.[3] The actual expression we found is $\frac{1}{2}nkT(|\overline{\Delta n}|/n)^2$.

1. RELATIVISTIC PLASMA[7]

For a plasma which obeys the relativistic Vlasov equation,[8] its motion in the $\mathbf{r}$, $\mathbf{u}$ space is measure-preserving, just as the motion of nonrelativistic plasma in $\mathbf{r}$, $\mathbf{v}$ space, $\mathbf{u}$ being the relativistic velocity $\mathbf{u} \equiv \gamma\mathbf{v} \equiv [1 - (v^2/c^2)]^{-\frac{1}{2}}\mathbf{v}$. The isoperimetric problem to be solved is as follows:

Let $f(\mathbf{r}, \mathbf{u})$ be a nonnegative function of $\mathbf{r}$ and $\mathbf{u}$ which physically stands for the particle distribution. Determine f which minimizes the kinetic energy $\mathcal{E}$,

$$\mathcal{E} \equiv \iint_{L^3} mc^2(\gamma - 1)f(\mathbf{r}, \mathbf{u})\, d^3r\, d^3u,$$

under the constraints

$$\Phi(\alpha) = \text{const}, \tag{1.1}$$

$$|\overline{\Delta n}| = \text{const}. \tag{1.2}$$

Where $\Phi(\alpha)$ is defined as the *measure of the point set*

* Present address: Physics Department, Georgetown University, Washington, D.C.

[1] T. K. Fowler, J. Math. Phys. **4**, 559 (1963).
[2] T. K. Fowler, Phys. Fluids **8**, 459 (1965).
[3] R. L. W. Chen, J. Math. Phys. **8**, 2410 (1967).
[4] R. L. W. Chen, Phys. Fluids **9**, 761 (1966).
[5] C. S. Gardner, Phys. Fluids **6**, 839 (1963).
[6] T. K. Fowler and G. E. Guest, *Plasma Physics and Controlled Nuclear Fusion Research, Vol. 1* (International Atomic Energy Agency, Vienna, 1966).

[7] This section is based on the Master's thesis of T. H. Neighbors at Emory University.
[8] P. C. Clemmow and A. J. Wilson, Proc. Cambridge Phil. Soc. **53**, 222 (1957).

in the $\mathbf{r}$, $\mathbf{u}$ phase space $\{\mathbf{r}, \mathbf{u}: f(\mathbf{r}, \mathbf{u}) > \alpha\}$. Δn is the deviation of the number density

$$n = \int f(\mathbf{r}, \mathbf{u})\, d^3u$$

from its mean. L^3 is the volume in $\mathbf{r}$ space, to which the $\mathbf{r}$ integration is confined. Similar to the nonrelativistic problem, $\Phi(\alpha)$ is set equal to that of a Maxwellian. The relativistic Maxwellian is[9]

$$f_0(u) = \frac{nm}{4\pi ckTK_2(mc^2/kT)} \exp\left(-\frac{mc^2\gamma}{kT}\right),$$

where m, k, T are, respectively, particle mass, Boltzmann constant, and temperature, and K_2 is the modified Bessel function. The corresponding Φ, to be designated as Φ_0, is found to be

$$\Phi_0(\alpha) = \tfrac{4}{3}\pi c^3\left\{\left[\frac{kT}{mc^2}\ln\left(\frac{nm}{4\pi ckTK_2\alpha}\right)\right]^2 - 1\right\}^{\frac{3}{2}}.$$

The expression (kT/mc^2) is henceforth denoted by β. If one divides L^3 into two regions I and II, wherein Δn is negative and nonnegative, respectively—the ratio of the volumes of I and II being fixed at the constant value w—then, it was shown in the nonrelativistic case, the minimizing $f(\mathbf{r}, \mathbf{v})$ must be of the form of two monotone functions of v. This remains true in the relativistic case. We thus restrict ourselves to considering $f(\mathbf{r}, \mathbf{u})$ of the form $f_1(u)$ and $f_2(u)$ in regions I and II, respectively. Under the circumstances, δ becomes dependent on a function of *only one* variable, $\phi(\alpha)$ where $\phi(\alpha)$ is defined as ϕ_0 $(\equiv \Phi_0 L^{-3})$ minus the measure in $\mathbf{u}$ space of the point set $\{\mathbf{u}: f_1(u) > \alpha\}$. With suitable choice of variables x and y,

$$x \equiv \beta \ln\left(\frac{nm}{4\pi ckTK_2(\beta)\alpha}\right),$$

$$y \equiv -3K_2(\beta)\int_0^x \frac{\phi'\alpha^2}{n} e^{Z/\beta}\, dZ,$$

the kinetic energy per unit volume $E \equiv \delta L^{-3}$ becomes

$$E = \frac{nkT\beta^2}{(1+w)K_2}\int F(x, y, y')\, dx, \qquad (1.3)$$

and the second constraint in (1.2) becomes

$$\frac{1}{\beta K_2}\int G(x, y')\, dx = b, \qquad (1.4)$$

where

$$F \equiv w(u_1 - 1)[e^{-x/\beta}x(x^2-1)^{\frac{1}{2}} - \tfrac{1}{3}y'e^{-x/\beta}] + (u_2 - 1)[e^{-x/\beta}x(x^2-1)^{\frac{1}{2}} + \tfrac{1}{3}wy'e^{-x/\beta}],$$

$$G \equiv \tfrac{1}{3}y'e^{-x/\beta}.$$

The mathematical problem is to minimize (1.3) under the constraint (1.4). $y(x)$ is the unknown function to be determined. The solution must satisfy certain conditions in order to be acceptable. $\phi_0(\alpha) - \phi(\alpha)$ must be *monotonically increasing* with respect to α in accordance with its definition. $\phi'(\alpha)$ is, therefore, bounded. The appropriate solution can be found, based on a theorem due to Weierstrass[10]:

Theorem: If the minimizing curve has a segment in common with the boundary R, then along this segment the following condition must be satisfied:

$$\mathcal{F}_y - \frac{d}{dx}\mathcal{F}_{y'} \geq 0, \quad \text{if } R \text{ lies above the segment,}$$

$$\mathcal{F}_y - \frac{d}{dx}\mathcal{F}_{y'} \leq 0, \quad \text{if } R \text{ lies below the segment,}$$

subject to the requirement that the *Weierstrass E function* must be equal to zero at the end points of the segment. The Weierstrass E function is defined as

$$\mathcal{F}(x, y, y') - \mathcal{F}(x, y, p) - (y' - p)\frac{\partial \mathcal{F}}{\partial p}(x, y, p).$$

$\mathcal{F}$ is the integrand of the functional being considered, which is $F = \lambda G$ in the present case. Our solution consists of two segments. The first one with x between 1 and x_0, $x_0 > 1$, lies in common with the boundary, i.e., $\phi(\alpha) = \phi_0(\alpha)$ or

$$y = (x^2 - 1)^{\frac{1}{2}}. \qquad (1.5a)$$

The second segment with x between x_0 and ∞ is found by solving the Euler equation:

$$\frac{\partial}{\partial y}(F + \lambda G) - \frac{d}{dx}\left[\frac{\partial}{\partial y'}(F + \lambda G)\right] = 0.$$

We obtain

$$w(u_2 - u_1) = -\lambda. \qquad (1.5b)$$

x_0 is related to λ as

$$x_0 = [2^{-\frac{2}{3}}(\lambda^2 - 2\lambda) + 1]^{\frac{1}{2}}.$$

It can be readily verified that (1.5a) satisfies the *Weierstrass conditions* and (1.5b) possesses *zero first variation*

[9] J. L. Synge, *The Relativistic Gas* (North-Holland Publishing Co., Amsterdam, 1957).

[10] O. Bolza, *Lectures on the Calculus of Variations* (G. E. Stechert and Co., New York, 1946), pp. 41–43.

TABLE I. Numerical values of b and the corresponding values of $\Delta E/nkT$ computed electronically.

b	$\Delta E/nkT$
0	0
0.099	0.012
0.239	0.062
0.3447	0.118
0.5070	0.257
0.6810	0.496
0.7976	0.741
0.9019	1.057

and *positive second variation.* When this solution is substituted in (1.3) and (1.4), we obtain two quadratures which relate E and b to the value of λ. For the case of $w = 1$ we have

$$E = \frac{nkT\beta^2}{K_2(\beta)}\int_1^{x_0}[1 + 2^{\frac{3}{2}}(x^2-1)^{\frac{1}{2}} - 1]e^{-x/\beta}x(x^2-1)^{\frac{1}{2}}\,dx + \frac{nkT\beta^2}{6K_2(\beta)}\int_{\eta_0}^{\infty}([(1+\eta^{\frac{2}{3}})^{\frac{1}{2}} - 1] + [(1+\eta^{\frac{2}{3}})^{\frac{1}{2}} + \lambda - 1]\{[(1+\eta^{\frac{2}{3}})^{\frac{1}{2}} + \lambda]^2 - 1\}^{\frac{1}{2}} \times [(1+\eta^{\frac{2}{3}})^{\frac{1}{2}} + \lambda](1+\eta^{\frac{2}{3}})^{-\frac{1}{2}}\eta^{-\frac{1}{3}})e^{-X/\beta}\,d\eta$$

and

$$X \equiv [2^{-\frac{2}{3}}(\{[(1+\eta^{\frac{2}{3}})^{\frac{1}{2}} + \lambda]^2 - 1\}^{\frac{2}{3}} + \eta)^{\frac{2}{3}} + 1]^{\frac{1}{2}},$$

where

$$\eta \equiv (x^2-1)^{\frac{3}{2}} + y,$$
$$\eta_0 \equiv 2(x_0^2-1)^{\frac{3}{2}},$$

$$b = \frac{1}{\beta K_2}\int_1^{x_0} e^{-X/\beta}(x^2-1)^{\frac{1}{2}}x\,dx + \frac{1}{6\beta K_2}\int_{\eta_0}^{\infty}(1 - \{[(1+\eta^{\frac{2}{3}})^{\frac{1}{2}} + \lambda]^2 - 1\}^{\frac{1}{2}} \times [(1+\eta^{\frac{2}{3}})^{\frac{1}{2}} + \lambda])e^{-X/\beta}\,d\eta. \quad (1.6)$$

We have chosen to evaluate E and b numerically at a typical *relativistic* temperature $kT = mc^2$ or $\beta = 1$. Samples of these values are given in Table I, which are taken from a more extensive set of data of electronic calculations. From this we obtain, to a good approximation, the following expression (in the same manner as in Ref. 3):

$$\Delta E \cong nkT(|\overline{\Delta n}|/n)^2, \quad \text{for} \quad kT = mc^2. \quad (1.7)$$

This may be compared to the nonrelativistic case of[2]

$$\Delta E \cong \tfrac{3}{5}nkT(|\overline{\Delta n}|/n)^2.$$

2. FREE ENERGY

We seek the minimum E under the constraints $N = \text{const}$ and $S = \text{const}$, where

$$E \equiv \int [\tfrac{1}{2}mv^2 + \pi(\mathbf{r})]f\,d^3r\,d^3v, \quad (2.1)$$

$$S \equiv -k\int f \ln f\,d^3r\,d^3v, \quad (2.2a)$$

$$N \equiv \int f\,d^3r\,d^3v, \quad (2.2b)$$

where $f \equiv f(\mathbf{r}, \mathbf{v})$ is the function to be varied and $\pi(\mathbf{r})$ denotes the potential of an external conservative force field (if present). By standard methods of Lagrange multipliers we obtain

$$f = \nu e^{-\mu\epsilon}, \quad (2.3)$$

where ν and μ are both positive constants, and $\epsilon \equiv \frac{1}{2}mv^2 + \pi(\mathbf{r})$. The values of ν and μ are related to S and N. For most situations of physical interest, π is negligible, and it is possible to manipulate the interdependence between S, N, ν, μ,and E so that E becomes expressible in terms of S and N. We obtain

$$E_{\min}(S, N) = \frac{3m}{4\pi}N^{\frac{5}{3}}V^{-\frac{2}{3}}\exp\left(\frac{2S}{3kN} - 1\right), \quad (2.4)$$

where V is the volume to which the plasma is confined. $E_{\min}$ in (2.4) increases with S. Hence, for a plasma which obeys the H-theorem we have

$$S_t > S_0,$$
$$E_t > E_{\min}(S_t, N) > E_{\min}(S_0, N),$$

where the subscript 0 and t indicate the initial time and any later time, respectively. The free energy available for driving instabilities is therefore

$$E_0 - \frac{3m}{4\pi}N^{\frac{5}{3}}V^{-\frac{2}{3}}\exp\left(\frac{2S_0}{3kN} - 1\right), \quad (2.5)$$

where E_0 and S_0 are completely determined by the initial distribution $f_0(\mathbf{r}, \mathbf{v})$.

With a nonzero $\pi(\mathbf{r})$ the monotonic-increasing character of $E_{\min}$ can also be established. Substituting (2.3) in (2.1) and (2.2), and taking the partial derivative $(\partial S/\partial\mu)_N$, keeping N fixed, we get

$$\left(\frac{\partial S}{\partial\mu}\right)_N = -k\mu\int\epsilon^2\nu e^{-\mu\epsilon}\,d^3r\,d^3v = k\mu\left(\frac{\partial E_{\min}}{\partial\mu}\right)_N.$$

It follows that

$$\left(\frac{\partial E_{\min}}{\partial S}\right)_N = \frac{1}{k\mu} > 0,$$

which proves the point. Thus, the free energy as $E_0 - E_{\min}(S_0, N)$ is generally valid, although the explicit expression for $E_{\min}$ may not be obtainable.

The bound is the closest one possible if the H-theorem and the conservation of particles are assumed to be the only constraints. It can be shown to be equivalent to the *minimized* value of the plasma Lyapunov function of Fowler,[1,2] attainable by the most judicious choice of parameters T and C.

3. ΔE OF COLLISIONAL PLASMA

In calculating the additional energy associated with density fluctuations in a collisional plasma, we have the following problem which is analogous to that in Section 1.

Determine $f(\mathbf{r}, \mathbf{v})$ which minimizes $\mathcal{E}$:

$$\mathcal{E} \equiv \iint_{L^3} \tfrac{1}{2} m v^2 f(\mathbf{r}, \mathbf{v})\, d^3r\, d^3v \tag{3.1}$$

under the constraints

$$\begin{aligned} N &= \text{const}, \\ S &= \text{const}, \\ \overline{|\Delta n|} &= \text{const}. \end{aligned} \tag{3.2}$$

We note that for a collisional plasma S actually increases rather than remaining constant. But this makes no difference insofar as finding the lower bound on ΔE is concerned.

Consider again the two regions as in Sec. 1. Let N_1, N_2, S_1, S_2 be the respective N and S in region I and region II, respectively. S_1, for example, is defined as

$$-k \iint_{\mathbf{r}\in I} f \ln f\, d^3r\, d^3v,$$

with $\mathbf{r}$ confined to I and $\mathbf{v}$ over all $\mathbf{v}$ space. The minimum $\mathcal{E}$ for given N_1, N_2, S_1, S_2, and w, in accordance with the result of (2.4), is

$$\mathcal{E} = \frac{3m}{4\pi}\left[N_\eta^{\frac{5}{3}} \left(\frac{w}{1+w} L^3\right)^{-\frac{2}{3}} \exp\left(\frac{2S_1}{3kN_1} - 1\right) + N_2^{\frac{5}{3}} \left(\frac{1}{1+w} L^3\right)^{-\frac{2}{3}} \exp\left(\frac{2S_2}{3AN_2} - 1\right)\right]. \tag{3.3}$$

We then minimize (3.3) further by varying N_1, N_2, S_1, S_2, and w subject to the conditions of (3.2). These conditions, in terms of N_1, N_2, S_1, S_2, and w, now become

$$S_1 + S_2 = S = \text{const},$$

$$N_1 + N_2 = N = \text{const},$$

$$L^{-3}\left[\left(N_1 - \frac{w}{1+w} N\right) - \left(N_2 - \frac{1}{1+w} N\right)\right] = \overline{|\Delta n|} = \text{const}.$$

The problem is now reduced to minimizing a *function of several variables* under some constraints and can be solved in a comparatively straightforward manner. After a fair amount of manipulation, we obtain a series expansion of the minimized E ($\equiv \mathcal{E}L^{-3}$):

$$\frac{3}{4}\frac{m}{\pi} n^{\frac{5}{3}} \exp\left(\frac{2S}{3kn} - 1\right) \cdot \left[1 + \frac{1}{3}\left(\frac{\overline{|\Delta n|}}{n}\right)^2 + \cdots\right],$$

where $s = SL^{-3}$ is the entropy density.[3] The linear term in the bracket is zero and, as shown by exact numerical work, the higher terms contribute no more than a few percent. The expression preceding the bracket is, in fact, the kinetic energy density of a uniform Maxwellian gas, and can be expressed in terms of its temperature as $\frac{3}{2}nkT$. The requirement for additional energy beyond that of the uniform state is, therefore,

$$\Delta E \cong \tfrac{1}{2} nkT(\overline{|\Delta n|}/n)^2.$$

We again note that the *collisionless* case has

$$\Delta E \cong \tfrac{3}{5} nkT(\overline{|\Delta n|}/n)^2.$$

We conclude that $nkT(\overline{|\Delta n|}/n)^2$ dependence is a ubiquitous feature of ΔE in all kinds of plasma. There are some differences, however, in the value of the numerical factor.

ACKNOWLEDGMENTS

The author wishes to thank the Emory Biomedical Data Processing and Analysis Center and its director Dr. Buell Evan. Special thanks are due Robert Harbort for his programming.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Negative Finding for the Three-Dimensional Dimer Problem*

J. M. HAMMERSLEY, A. FEUERVERGER, A. IZENMAN, AND K. MAKANI
Statistical Laboratory, University of California, Berkeley

(Received 30 August 1968)

The dimer problem can be solved if one can evaluate the permanent of $\mathbf{P} = (p_{ij})$, the incidence matrix of the lattice. All known methods of solving the two-dimensional case consist (explicitly or implicitly) in finding another matrix $\mathbf{Q} = (q_{ij})$, such that $p_{ij} = |q_{ij}|$ and per $\mathbf{P} = |\det \mathbf{Q}|$, and then computing the determinant of $\mathbf{Q}$. We show that in the three-dimensional case no such matrix $\mathbf{Q}$ exists for any choice of elements q_{ij}, whether real or complex numbers, or quaternions. A stronger negative result of an asymptotic character seems to be true, but this rests upon a plausible but unproved conjecture.

INTRODUCTION AND STATEMENT OF RESULTS

Let a, b, c be positive integers with $N = abc$ even, and define the lattice L to be the set of N points in three-dimensional Euclidean space with integer coordinates (x, y, z) such that $1 \leq x \leq a$, $1 \leq y \leq b$, $1 \leq z \leq c$. A dimer is a pair of points of L which are unit distance apart; and a dimer configuration is a partitioning of L into $\frac{1}{2}N$ disjoint dimers. Let f denote the number of dimer configurations on L. It can be proved[1] that $N^{-1} \log f$ tends to a limit (denoted by λ) as $a \to \infty$, $b \to \infty$, $c \to \infty$ independently. (For brevity, we hereafter write $N \to \infty$ to signify $a, b, c \to \infty$.) The dimer problem is to determine f as a function of a, b, c and hence (or otherwise) to calculate λ.

Number the points of L from 1 to N in a fixed arbitrary way, and write $p_{ij} = 1$ or 0, according as the ith and jth points of L are or are not unit distance apart. The $N \times N$ matrix $\mathbf{P} = (p_{ij})$ is called the incidence matrix of L; and it can be shown[2] that $f^2 = \text{per } \mathbf{P}$, the permanent of $\mathbf{P}$. Thus a solution of the dimer problem is equivalent to an evaluation of this permanent. Unlike determinants, to which they bear a superficial algebraic resemblance, permanents do not enjoy any practicable algorithms for their evaluation when N is large. However, most of the elements of $\mathbf{P}$ are zero, and this has suggested the possibility of finding another matrix $\mathbf{Q}$, such that

$$p_{ij} = |q_{ij}| \quad \text{and} \quad \text{per } \mathbf{P} = |\det \mathbf{Q}|, \tag{1}$$

and so calculating f via det $\mathbf{Q}$. Here the q_{ij} are real or complex numbers or quaternions; and, if q is a real or complex number, $|q|$ denotes its modulus in the ordinary sense; while, if q is a quaternion, its modulus $|q|$ is the positive square root of its norm. (The possibility of using quaternions in this context has not, so far as we know, been mentioned in the literature, but from private conversations we know that this idea has occurred independently to several colleagues; for, indeed, there are anticommuting features in the dimer problem which lend appeal to use of quaternions as a tool.) We discuss below two methods of defining the determinant of a matrix of quaternions.

In the two-dimensional case (i.e., when $a = 1$ and only $b \to \infty$ and $c \to \infty$) all known methods of solving the dimer problem depend, explicitly or implicitly, on finding a solution of (1), and a variety of such real and complex solutions are known. Here we prove that no solutions exist in the three-dimensional case: specifically, we show that, when $a \geq 2$, $b \geq 4$, and $c \geq 4$, then

$$p_{ij} = |q_{ij}| \Rightarrow \text{per } \mathbf{P} > |\det \mathbf{Q}|, \tag{2}$$

for any choice of real or complex or quaternion $\mathbf{Q}$. This does not, however, completely dispose of (1) as a device for computing λ; for, despite (2), it might still be true that

$$\limsup_{N \to \infty} \tfrac{1}{2}N^{-1} \log |\det \mathbf{Q}| = \lim_{N \to \infty} \tfrac{1}{2}N^{-1} \log \text{per } \mathbf{P} = \lambda, \tag{3}$$

when $p_{ij} = |q_{ij}|$. We believe that (3) is actually false; but the best we can do in this direction is to deduce the falsity of (3) from the following plausible but unproved conjecture. Define a *block* of L to be a set of 32 points of L whose coordinates (x, y, z) satisfy $\xi \leq x < \xi + 2$, $\eta \leq y < \eta + 4$, $\zeta \leq z < \zeta + 4$ for some integers ξ, η, ζ. Thus L contains $(a - 1)(b - 3)(c - 3)$ different blocks when $a \geq 2$, $b \geq 4$, $c \geq 4$. Given a dimer configuration on L, we say that a particular block is *smooth* if there is no dimer of the configuration with one of its points in this block and the other point not in the block. A dimer configuration is called *rough* if no block of L is smooth. Let g denote the number of rough configurations on L. We conjecture

* This investigation was partially supported by research grant No. GM 10525-05 from the National Institutes of Health, Public Health Service.

[1] J. M. Hammersley, "Existence Theorems and Monte Carlo Methods for the Monomer-Dimer Problem" in *Research Papers in Statistics: Festschrift for J. Neyman* (John Wiley & Sons, Inc., New York, 1966), pp. 125–146.

[2] J. M. Hammersley, Proc. Cambridge Phil. Soc. **64**, 455 (1968).

that

$$\liminf_{N\to\infty} N^{-1}\log g < \lambda. \tag{4}$$

However, we do not give here the proof that (3) is false if (4) is true.

DETERMINANTS OF QUATERNION MATRICES

The literature contains two slightly different definitions of the determinant of a matrix $\mathbf{Q}$ with quaternion elements. They are due to Moore[3] and Dieudonné,[4] and we denote them by $\det_M \mathbf{Q}$ and $\det_D \mathbf{Q}$, respectively. We recall that a quaternion can be written $q = \alpha + \beta i + \gamma j + \delta k$, where α, β, γ, δ are real numbers and i, j, k are indeterminates satisfying $i^2 = j^2 = k^2 = ijk = -1$; that the conjugate of q is $\bar{q} = \alpha - \beta i - \gamma j - \delta k$; that the norm of q is $N(q) = q\bar{q} = \bar{q}q = \alpha^2 + \beta^2 + \gamma^2 + \delta^2$; and we write $|q|$ for the modulus of q, i.e., the positive square root of $N(q)$. Then $|q_1q_2| = |q_2q_1| = |q_1|\,|q_2|$ for any quaternions q_1, q_2. We write $\mathbf{Q}^*$ for the transposed conjugate of a quaternion matrix $\mathbf{Q}$; we call $\mathbf{Q}$ Hermitian if $\mathbf{Q} = \mathbf{Q}^*$, i.e., if $q_{ij} = \bar{q}_{ji}$. All matrices mentioned below are square matrices with quaternion elements, unless the contrary is explicitly stated.

Dieudonné's paper deals with a slightly more general case than we need. Reduced to the quaternion case in hand, and stripped of its abstract terminology, it boils down to the following. If $\mathbf{Q}$ is a diagonal matrix, $\det_D \mathbf{Q}$ is defined to be $|q|$, where q is the product of the diagonal elements of $\mathbf{Q}$. For general $\mathbf{Q}$, the value of $\det_D \mathbf{Q}$ is (by definition) unchanged if, to any row of $\mathbf{Q}$ we add a constant multiple of any other row, it being understood that the constant multiplier (which is a quaternion) acts as a left-hand multiplier of the row. Similarly the value is unchanged for similar operations on columns, the constant multiplier now being a right-hand multiplier of the column. The value of $\det_D \mathbf{Q}$ also is unchanged by any permutation of rows or of columns of $\mathbf{Q}$. As with ordinary determinants, these row and column operations let us reduce a general $\mathbf{Q}$ to diagonal form, and so to determine the value of $\det_D \mathbf{Q}$. Dieudonné shows that the foregoing requirements are self-consistent and uniquely determine $\det_D \mathbf{Q}$, and that $\det_D (\mathbf{Q}_1\mathbf{Q}_2) = \det_D \mathbf{Q}_1 \det_D \mathbf{Q}_2$ for any two matrices $\mathbf{Q}_1$, $\mathbf{Q}_2$.

Moore's definition applies only to the case when $\mathbf{Q}$ is Hermitian. Suppose $\mathbf{Q}$ has N rows and N columns, and write Z for the set $\{1, 2, \cdots, N\}$. Let z be some given nonempty subset of Z, and suppose that z has s elements. Define

$$q(z, i_1) = \sum (-1)^{s-1} q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{s-1} i_s} q_{i_s i_1}, \tag{5}$$

where i_1 is a selected element of z, and the sum in (5) is taken over all permutations of the remaining unselected elements $i_2, i_3, \cdots, i_s$ of z. Thus there are $(s-1)!$ summands in (5). The Hermitian character of $\mathbf{Q}$ ensures, as Moore proves, that $q(z, i_1)$ is a real number (i.e., a quaternion with $\beta = \gamma = \delta = 0$) and that $q(z, i_1)$ is independent of the choice of i_1 in z. We may thus write $q(z)$ in place of $q(z, i_1)$ and regard the sum in (5) as being taken over the $(s-1)!$ different cycles which can be formed from the elements of z. Next, let T be a partition of Z into disjoint nonempty subsets $z_1^T, z_2^T, \cdots, z_t^T$ (whose union is Z, of course); and define

$$\det_M \mathbf{Q} = \sum_T q(z_1^T) q(z_2^T) \cdots q(z_t^T), \tag{6}$$

where the sum in (6) is over all possible distinct partitions of Z. [As is usual in a partition, the order of the parts and the order of the elements in each part is immaterial; but the order of the parts does not affect the definition (6), because the $q(z^T)$ are all real and therefore commute; and the order of the elements in each part does not affect the definition (5), because $q(z, i_1)$ is independent of the selection i_1 and the summation in (5) is over all $(s-1)!$ permutations of the remaining unselected elements.]

Moore also proves that, if $\mathbf{Q}_0$ is an arbitrary quaternion matrix and if $\mathbf{Q}$ is Hermitian, then $\mathbf{Q}_0\mathbf{Q}_0^*$ and $\mathbf{Q}_0\mathbf{Q}\mathbf{Q}_0^*$ are both Hermitian and

$$\det_M (\mathbf{Q}_0\mathbf{Q}\mathbf{Q}_0^*) = \det_M (\mathbf{Q}_0\mathbf{Q}_0^*) \det_M \mathbf{Q}.$$

In particular, if we take the diagonal elements of $\mathbf{Q}_0$ to be all 1, and all the nondiagonal elements, except just one of them, to be zero, then it is easy to verify from (5) that $\det_M (\mathbf{Q}_0\mathbf{Q}_0^*) = 1$. However, by choosing a succession of such $\mathbf{Q}_0$'s to premultiply and postmultiply $\mathbf{Q}$ in the fashion of $\mathbf{Q}_0\mathbf{Q}\mathbf{Q}_0^*$ we can reduce $\mathbf{Q}$ to diagonal form, just as with ordinary Hermitian transformations.[5] Since clearly $\det_D \mathbf{Q}_0 = \det_D \mathbf{Q}_0^* = 1$, we can prove in this way that

$$\det_D \mathbf{Q} = |\det_M \mathbf{Q}| \tag{7}$$

for any Hermitian matrix $\mathbf{Q}$. It follows that the notation $|\det \mathbf{Q}|$ may be used without ambiguity for either $|\det_D \mathbf{Q}|$ or $|\det_M \mathbf{Q}|$ when $\mathbf{Q}$ is Hermitian. (It is easy to see from simple examples that $\det_D \mathbf{Q} = \det_M \mathbf{Q}$ is not always true for Hermitian matrices $\mathbf{Q}$.)

[3] E. H. Moore, *General Analysis, Part I* (Memoirs series, Vol. 1, The American Philosophical Society, Philadelphia, Pa., 1935).

[4] J. Dieudonné, Bull. Soc. Math. de France 71, 27 (1943).

[5] H. W. Turnbull and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices* (Blackie & Son Ltd., Glasgow, 1932), p. 85.

PROOF OF EQUATION (2)

We assume throughout that $a \geq 2$, $b \geq 4$, $c \geq 4$, and show that

$$p_{ij} = |q_{ij}| \quad \text{and} \quad \text{per } \mathbf{P} \leq |\det \mathbf{Q}| \tag{8}$$

leads to a contradiction. The case when the q_{ij} are real or complex numbers is a particular case of quaternions q_{ij} with $\gamma = \delta = 0$ for all elements, since then the modulus of the ordinary determinant of $\mathbf{Q}$ coincides with $\det_D \mathbf{Q}$, as the above definition of the latter shows. Hence we may suppose that $\mathbf{Q}$ is a quaternion matrix satisfying (8). Color the points of L black and white after the fashion of a chessboard, i.e., all points of L at unit distance from a white point shall be black and vice versa. We say that the ith row of $\mathbf{Q}$ is black or white according as the ith point of L is black or white. Let T be the permutation of rows of $\mathbf{Q}$ which places all the black rows before all the white rows, while leaving the relative order of the black rows among themselves unchanged and similarly preserving the relative order of the white rows. Apply this permutation to the rows of $\mathbf{Q}$ and the *same* permutation to the columns of $\mathbf{Q}$. Since each dimer contains one black and one white point wherever it may be on L, $\mathbf{Q}$ is transformed to the form

$$\mathbf{Q}' = \begin{pmatrix} \mathbf{0} & \mathbf{Q}_1 \\ \mathbf{Q}_2 & \mathbf{0} \end{pmatrix}$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are $\frac{1}{2}N \times \frac{1}{2}N$ quaternion matrices. Then from (8) we have

$$f^2 = \text{per } \mathbf{P} \leq |\det \mathbf{Q}| = \det_D \mathbf{Q} = \det_D \mathbf{Q}' = \det_D \mathbf{Q}_1 \det_D \mathbf{Q}_2 . \tag{9}$$

Hence there exists $\mathbf{Q}_0$, equal to one or another of $\mathbf{Q}_1$ or $\mathbf{Q}_2$, such that $f \leq \det_D \mathbf{Q}_0$. But $\det_D \mathbf{Q}_0 = \det_D \mathbf{Q}_0^*$ from the definitions. Hence,

$$\text{per } \mathbf{P} = f^2 \leq \det_D \mathbf{Q}_0 \det_D \mathbf{Q}_0^* = \det_D \begin{pmatrix} \mathbf{0} & \mathbf{Q}_0 \\ \mathbf{Q}_0^* & \mathbf{0} \end{pmatrix} = \det_D \mathbf{Q}'' = |\det_D \mathbf{Q}''|, \tag{10}$$

where $\mathbf{Q}''$ is the matrix obtained by applying the inverse permutation T^{-1} to both the rows and columns of $\begin{pmatrix} \mathbf{0} & \mathbf{Q}_0 \\ \mathbf{Q}_0^* & \mathbf{0} \end{pmatrix}$. We have $p_{ij} = |q''_{ij}|$; and $\mathbf{Q}''$ is Hermitian. Thus if any solution $\mathbf{Q}$ of (8) exists, there is at least one Hermitian solution of (8). Hereafter we suppose that $\mathbf{Q}$ is such a Hermitian solution of (8); and accordingly we may now interpret $\det \mathbf{Q}$ as $\det_M \mathbf{Q}$.

Let $P_1, P_2, \cdots, P_N$ denote the points of L in the fixed enumeration used for specifying the incidence matrix $\mathbf{P}$. We define an (oriented) polygon on L as a cyclic sequence of distinct points of L, say $(P_{j_1} P_{j_2} \cdots P_{j_s})$. We further say that a polygon is a *nonzero polygon* if its sides $P_{j_1}P_{j_2}, P_{j_2}P_{j_3}, \cdots, P_{j_{s-1}}P_{j_s}, P_{j_s}P_{j_1}$ are all of unit length. The cubic character of L guarantees that a nonzero polygon must have an even number of sides. We include two-sided polygons (i.e., ones with only a pair of sides $P_{j_1}P_{j_2}$, $P_{j_2}P_{j_1}$) in our discussion; indeed, nonzero two-sided polygons play an important role, and we call them *degenerate polygons*.

Let π be any permutation of $Z = \{1, 2, \cdots, N\}$. This permutation can be written, in the usual way, as a product of disjoint cycles $\sigma_1^\pi \sigma_2^\pi \cdots \sigma_t^\pi$ (including 1-cycles if they occur). This product is unique apart from the order of its terms. A cycle in the product, say $\sigma = (j_1 j_2 \cdots j_s)$, corresponds naturally to a polygon $(P_{j_1} P_{j_2} \cdots P_{j_s})$; hence, there is a one-to-one correspondence between a permutation π and a partition of L into disjoint oriented polygons. However, each permutation π is in one-to-one correspondence with a product in the expansion of $\text{per } \mathbf{P} = \Sigma_\pi p_{1\pi(1)} p_{2\pi(2)} \cdots p_{N\pi(N)}$. Moreover, the nonzero products in this expansion correspond to the partitions of L into polygons which are all nonzero polygons.

Again, in any cycle $\sigma = (j_1 j_2 \cdots j_s)$ we can select a particular element i_1, say the numerically smallest element in σ, and then write $\sigma = (i_1 i_2 \cdots i_s)$, where $i_1 i_2 \cdots i_s$ is obtained from $j_1 j_2 \cdots j_s$ by cyclic permutation. Thus a cycle corresponds to a product in the sum (5); and a permutation $\pi = \sigma_1^\pi \sigma_2^\pi \cdots \sigma_t^\pi$ corresponds to a term (a product of N quaternions) in the sum obtained by substituting (5) into (6). This correspondence is one to one and again maps the nonzero terms in the expansion of $\det_M \mathbf{Q}$ onto the partitions of L into nonzero polygons.

Thus the number of nonzero products in the expansions of $\text{per } \mathbf{P}$ and $\det_M \mathbf{Q}$ is f^2 in both cases. Since each nonzero product in the expansion of $\det_M \mathbf{Q}$ is a product of N unit quaternions, such a product is a unit quaternion. It now follows from (8) that the modulus of a sum of f^2 unit quaternions can only be not less than $f^2 = \text{per } \mathbf{P}$ if all these unit quaternions are equal. Hence every nonzero term in the expansion of $\det_M \mathbf{Q}$ must equal $(-1)^{N/2}$, because this is the value of one such particular product obtained when all the polygons are degenerate, whereupon each quantity in (5) takes the form

$$-q_{i_1 i_2} q_{i_2 i_1} = -q_{i_1 i_2} \bar{q}_{i_1 i_2} = -1.$$

We say that a polygon on L is *admissible* if it is a nonzero polygon, and if there exists a polygon-partition of L, containing this polygon and having all its other polygons degenerate. Consider any admissible polygon with $2r$ sides, and suppose that

$q_1, q_2, \cdots, q_{2r}$ are the values of $q_{i_1 i_2}, \cdots, q_{i_{2r} i_1}$ encountered in following the cycle σ around this polygon. There is a polygon partition with $\frac{1}{2}N - r$ degenerate polygons besides the given admissible polygon. Hence the corresponding summand in (6) yields

$$(-q_1 q_2 \cdots q_{2r})(-1)^{\frac{1}{2}N-r} = (-1)^{\frac{1}{2}N}, \tag{11}$$

i.e.,

$$q_1 q_2 \cdots q_{2r} = (-1)^{r-1} \tag{12}$$

for any admissible polygon. In particular,

$$q_1 q_2 q_3 q_4 = -1 \tag{13}$$

for an admissible square; and

$$q_1 q_2 q_3 q_4 q_5 q_6 = +1 \tag{14}$$

for an admissible hexagon (not necessarily a planar hexagon).

Suppose temporarily that $a = 2$, $b = 4$, $c = 4$. We show that an admissible hexagon, whose opposite sides are opposite sides of a cube, lies near the center of L. This follows from the diagram in Fig. 1. Here the points of L are denoted by crosses or circles according to their x coordinate. It is easy to see from similar diagrams that any square, forming a face of the cube in the above diagram, is also admissible. Moreover,

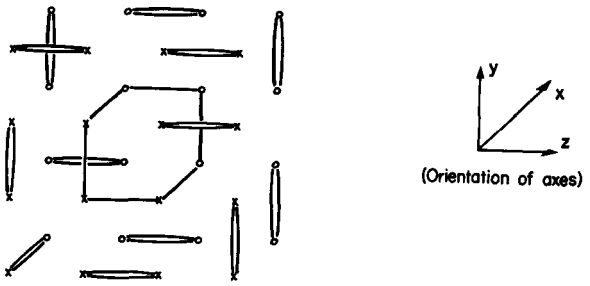


FIG. 1. Admissible hexagon in a smooth block.

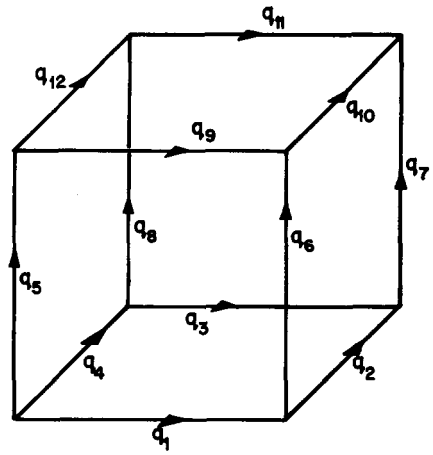


FIG. 2. Numbering and orientation of quaternions on a cube.

this diagram can be embedded in a larger polygon partition with $a \geq 2$, $b \geq 4$, $c \geq 4$ by pairing the points outside this $2 \times 4 \times 4$ configuration in an obvious fashion. So the existence of this admissible hexagon also follows for $a \geq 2$, $b \geq 4$, $c \geq 4$.

Now consider the cube carrying this admissible hexagon, and let the q_{ij} on its sides, with respect to the marked orientations, be $q_1, q_2, \cdots, q_{12}$ as shown in Fig. 2. From (14) we have

$$q_1 q_2 q_7 \bar{q}_{11} \bar{q}_{12} \bar{q}_5 = +1 \tag{15}$$

and from (13) we have

$$q_1 q_2 \bar{q}_3 \bar{q}_4 = q_3 q_7 \bar{q}_{11} \bar{q}_8 = q_8 \bar{q}_{12} \bar{q}_5 q_4 = -1. \tag{16}$$

Hence,

$$\begin{aligned} +1 &= q_1 q_2 q_7 \bar{q}_{11} \bar{q}_{12} \bar{q}_5 \\ &= q_1 q_2 (\bar{q}_3 q_3) q_7 \bar{q}_{11} (\bar{q}_8 q_8) \bar{q}_{12} \bar{q}_5 (q_4 \bar{q}_4) \\ &= q_1 q_2 \bar{q}_3 (q_3 q_7 \bar{q}_{11} \bar{q}_8)(q_8 \bar{q}_{12} \bar{q}_5 q_4) \bar{q}_4 \\ &= q_1 q_2 \bar{q}_3 \bar{q}_4 = -1. \end{aligned} \tag{17}$$

This contradiction denies (8) and completes the proof of (2).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Approach to the Mode Conversion Problem in Nonuniform Acoustic Waveguides

JAMES R. WAIT
CIRES, University of Colorado, Boulder, Colorado

(Received 23 July 1968)

A method is described for treating wave propagation in a waveguide structure whose cross section varies in the direction of propagation. Special attention is given to the conversion of energy between waveguide modes of different order. For simplicity, the waveguide is bounded by impedance-type walls and the lateral height variation is assumed to have circular symmetry. This is considered to be an idealized model of an atmospheric waveguide for acoustic-wave propagation in the case when there is a localized depression.

1. INTRODUCTION

There has been a great deal of attention paid to acoustic-wave propagation in uniform guiding structures. While the terrestrial environment does tend to be uniformly stratified in its gross characteristics, there are many important instances where the effective cross section of the guide varies significantly in the direction of propagation. A concrete example is the propagation of sound in shallow water where there is a sudden change in depth. An equally striking illustration is when guided acoustic waves in the atmosphere pass over a mountain range.

Previous work in nonuniform waveguides[1–4] has been devoted to situations where the coupling between the modes is sufficiently small to be ignored. In this paper, we present a method, albeit approximate, for handling the more general case where mode coupling is retained. However, to simplify the discussion, we consider a waveguide whose lower boundary is flat but the height of the upper boundary is variable. In order to simulate earth curvature and to allow for the sound speed to vary with altitude, we take the wavenumber of the medium to be a specified function of height. The lower boundary may be regarded as an idealized smooth earth, while the variable upper boundary is some height where there is strong coupling between upgoing and downgoing wave types. As such, both walls of the waveguide are assumed to be characterized by impedance-type boundary conditions. The validity of such a description is not investigated here, but it has met with considerable success in electromagnetic-wave propagation.[5,6]

2. FORMULATION

The situation is illustrated in Fig. 1. Essentially, we are dealing with a parallel-plate waveguide of thickness h_1 except for a depressed region of radius a where the thickness is h_2. The respective regions are designated (1) and (2). The field quantity ψ is assumed to satisfy a scalar wave equation

$$(\nabla^2 + k^2)\psi = 0, \tag{1}$$

where $k(z)$, the wavenumber, is a function of z only.

Boundary conditions on the horizontal surfaces are of the type

$$\frac{\partial \psi}{\partial z} + K\psi = 0, \tag{2}$$

where K is independent of the radial coordinate in either region (1) or (2). For example, in dealing with acoustic waves, where ψ is the velocity potential, K is zero for a rigid surface or infinity for a free surface. In general, K may be regarded as a boundary-impedance parameter.

Solutions of (1) may be written in the form

$$\psi = Z_p(k_0 S\rho)e^{+ip\phi}G(z), \tag{3}$$

where Z_p is a cylindrical Bessel function of order p,

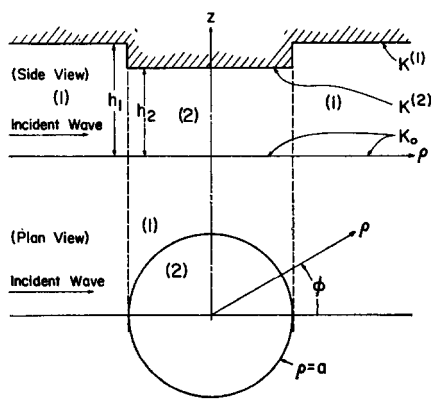


FIG. 1. The idealized height variation considered in this analysis. (The appropriate boundary-impedance parameters K_0, $K^{(1)}$ and $K^{(2)}$ are indicated.)

[1] D. E. Weston, Proc. Phys. Soc. (London) 73, 365 (1959).
[2] M. Redwood, *Mechanical Waveguides* (Pergamon Press Ltd., London, 1960), pp. 89–91.
[3] A. D. Pierce, J. Acoust. Soc. Am. 37, 19 (1965).
[4] I. Tolstoy and C. S. Clay, *Ocean Acoustics* (McGraw-Hill Book Co., Inc., New York, 1966).
[5] J. R. Wait, *Electromagnetic Waves in Stratified Media* (Pergamon Press, Ltd.,London, 1962).
[6] J. R. Wait, *Advances in Electronics and Electron Physics, Vol. 25*, L. Marton, Ed. (Academic Press Inc., New York, 1968).

while $G(z)$ satisfies

$$\frac{\partial^2 G}{\partial z^2} + [k^2(z) - k_0^2 S^2]G = 0, \tag{4}$$

where $k_0 = k(0)$ and S is a separation constant.

Some further consideration shows that the general solution is of the form

$$\psi = \sum_{p=-\infty}^{+\infty} e^{ip\phi} \sum_m (A_m)_p Z_p(k_0 S_m \rho) G_m(z), \tag{5}$$

where $(A_m)_p$ is a coefficient to be determined from the boundary conditions. The summation over p involves integers only, and the summation index m indicates that we superimpose all "modes." The latter are exemplified by a discrete set of S_m, obtained as a result of the boundary condition (2) which states that

$$\frac{\partial G_m}{\partial z} + KG_m|_{\text{at horizontal surfaces}} = 0. \tag{6}$$

Thus, $k_0 S_m^{(1)}$ is the transverse wavenumber for a waveguide mode of order m in region (1), and $k_0 S_m^{(2)}$ is the corresponding wavenumber in region (2). For convenience, we describe $G_m(z)$ as a "height-gain function" for a mode of order m and normalize it such that $G_m(0) = 1$. Again, a superscript (1) or (2) is added as appropriate.

We now imagine that a mode of order m is incident from a distant source in region (1). The problem is to calculate the total field everywhere. The incident field ψ_{inc}, suitably normalized, is written

$$\psi_{\text{inc}} = \frac{\Lambda_m^{(1)}}{h_1} \exp(-ik_0 S_m^{(1)} \rho \cos\phi) G_m^{(1)}(z_0) G_m^{(1)}(z), \tag{7}$$

where the "excitation factor" $\Lambda_m^{(1)}$ is a measure of the strength of the incident mode of order m for a distant point source located at a height z_0. Summation over m is implied.

Without difficulty,[5] we may demonstrate that

$$\int_0^{h_1} G_m^{(1)}(z) G_n^{(1)}(z)\, dz = 0, \quad \text{for} \quad m \neq n, \tag{8}$$

while

$$\int_0^{h_1} [G_m^{(1)}(z)]^2\, dz = \frac{h_1}{2\Lambda_m^{(1)}} \tag{9}$$

may be regarded as a definition of the "excitation factor" $\Lambda_m^{(1)}$ which appears in (7).

3. CONSTRUCTION OF THE SOLUTION

To deal with the specific problem posed above, we employ an addition theorem and rewrite (7) in an equivalent form.[7] Thus,

$$\psi_{\text{inc}} = \frac{\Lambda_m^{(1)}}{h_1} G_m^{(1)}(z_0) \sum_{p=-\infty}^{+\infty} e^{-i\pi p/2} J_p(k_0 S_m^{(1)} \rho) e^{ip\phi} G_m^{(1)}(z), \tag{10}$$

where J_p is the Bessel function of the first type. The total field $\psi^{(1)}$ in region (1) now consists of ψ_{inc} plus a scattered field ψ_{sc} which must have the form

$$\psi_{\text{sc}} = \frac{\Lambda_m^{(1)}}{h_1} G_m^{(1)}(z_0) \sum_{p=-\infty}^{+\infty} e^{-i\pi p/2} e^{ip\phi} \times \sum_s (B_{s,m})_p H_p(k_0 S_s^{(1)} \rho) G_s^{(1)}(z), \tag{11}$$

where $H_p = J_p - iY_p$ is the Hankel function of the second kind. Here, $(B_{s,m})_p$ may be regarded as a coefficient which describes the scattering from mode of order m into a mode of order s for a wave whose azimuthal variation is $\exp(ip\phi)$. It may be verified that (11) satisfies the wave equation (1), the appropriate boundary conditions, and it gives rise to outgoing waves as $\rho \to \infty$. The latter statement is a consequence of the asymptotic relation[7]

$$\lim_{Z \to \infty} H_p(Z) \simeq [2i/(\pi Z)]^{\frac{1}{2}} e^{i\pi p/2} e^{-iZ}. \tag{12}$$

Furthermore, when there is any loss in the system, the eigenvalue solution for the wavenumbers $k_0 S_s^{(1)}$ shows that $\text{Im}\, k_0 S_s^{(1)} < 0$, which leads to radial damping of the scattered modes.

The total field $\psi^{(1)}$ in region (1), i.e., where $\rho > a$ and $0 < z < h_1$, is thus given by $\psi^{(1)} = \psi_{\text{inc}} + \psi_{\text{sc}}$.

We now construct an expression for the field $\psi^{(2)}$ in region (2), i.e., where $\rho < a$ and $0 < z < h_2$. Without difficulty, we find that

$$\psi^{(2)} = \frac{\Lambda_m^{(1)}}{h_1} G_m^{(1)}(z_0) \sum_{p=-\infty}^{+\infty} e^{-i\pi p/2} e^{ip\phi} \times \sum_n (A_{n,m})_p J_p(k_0 S_n^{(2)} \rho) G_n^{(2)}(z), \tag{13}$$

where the Bessel function J_p is chosen to be finite at $\rho = 0$ and $(A_{n,m})_p$ is a coefficient which describes the modes of order n transmitted into region (2) for an incident mode of order m from region (1). We note that the height-gain function $G_n^{(2)}(z)$, for region (2), must be employed if $\psi^{(2)}$ is to satisfy the appropriate boundary conditions.

4. FIELD MATCHING

In order to obtain relationships between the unknown coefficients, we impose continuity conditions at the cylindrical interface between regions (1) and (2). Specifically, we require that ψ and $\partial\psi/\partial\rho$ be continuous at $\rho = a$, for $0 < z < h_2$, where $h_2 < h_1$.

[7] G. N. Watson, *Theory of Bessel Functions*, 2nd ed. (Cambridge University Press, Cambridge, England, 1944).

Then, on using (10), (11), and (13), we find that

$$J(k_0S_m^{(1)}a)G_m^{(1)}(z) + \sum_s B_{s,m}H(k_0S_s^{(1)}a)G_s^{(1)}(z) = \sum_n A_{n,m}J(k_0S_n^{(2)}a)G_n^{(2)}(z) \quad (14)$$

and

$$S_m^{(1)}J'(k_0S_m^{(1)}a)G_m^{(1)}(z) + \sum_s B_{s,m}S_s^{(1)}H'(k_0S_s^{(1)}a)G_s^{(1)}(z) = \sum_n A_{n,m}S_n^{(2)}J'(k_0S_n^{(2)}a)G_n^{(2)}(z), \quad (15)$$

where the primes indicate differentiation with respect to the argument of the Bessel function. In writing (14) and (15) the subscript p has been dropped, since term-wise matching of the p series is permitted.

We are now faced with the question of what boundary condition to impose on the vertical face $\rho = a$ when $h_1 > z > h_2$. For the moment, we assume that both $\psi^{(1)}$ and $\partial\psi^{(1)}/\partial\rho$ are zero over this surface. An equivalent statement is to say that (14) and (15) apply over the interval $0 < z < \infty$ but $G_n^{(1)}(z) = 0$ for $z > h_1$ and $G_n^{(2)}(z) = 0$ for $z > h_2$. A method to estimate the error of this assumption and a first-order correction has been discussed elsewhere[8] in connection with electromagnetic waves in the earth–ionosphere waveguide.

5. REDUCTION OF THE SOLUTION

To reduce (14) and (15), for the idealizations indicated, we multiply both sides by $G_q^{(2)}(z)$ and integrate from 0 to h_2. Thus, we obtain

$$J(k_0S_m^{(1)}a)\hat{\Gamma}_{q,m} + \sum_s B_{s,m}H(k_0S_s^{(1)}a)\hat{\Gamma}_{q,s} = A_{q,m}J(k_0S_q^{(2)}a)\Gamma_{q,q} \quad (16)$$

and

$$S_m^{(1)}J'(k_0S_m^{(1)}a)\hat{\Gamma}_{q,m} + \sum_s B_{s,m}S_s^{(1)}H'(k_0S_a^{(1)}a)\hat{\Gamma}_{q,s} = A_{q,m}S_q^{(2)}J'(k_0S_q^{(2)}a)\Gamma_{q,q}, \quad (17)$$

where

$$\Gamma_{q,q} = \frac{h_2}{2\Lambda_q^{(2)}} = \int_0^{h_2}[G_q^{(2)}(z)]^2\,dz, \quad (18)$$

$$\hat{\Gamma}_{q,m} = \int_0^{h_2} G_m^{(1)}(z)G_q^{(2)}(z)\,dz, \quad (19)$$

and

$$\hat{\Gamma}_{q,s} = \int_0^{h_2} G_s^{(1)}(z)G_q^{(2)}(z)\,dz. \quad (20)$$

In getting from (14) and (15) to (16) and (17), we exploit the orthogonality property

$$\int_0^{h_2} G_n^{(2)}(z)G_q^{(2)}(z)\,dz = 0 \quad \text{for} \quad n \neq q. \quad (21)$$

On eliminating $A_{q,m}$ from (16) and (17), we obtain the single infinite system

$$[S_q^{(2)}J'(k_0S_q^{(2)}a)J(k_0S_m^{(1)}a) - S_m^{(1)}J'(k_0S_m^{(1)}a)J(k_0S_q^{(2)}a)]\hat{\Gamma}_{q,m} + \sum_s B_{s,m}[S_q^{(2)}J'(k_0S_q^{(2)}a)H(k_0S_s^{(1)}a) - S_s^{(1)}H'(k_0S_s^{(1)}a)J(k_0S_q^{(2)}a)]\hat{\Gamma}_{q,s} = 0. \quad (22)$$

In principle, we may solve for the coefficients $B_{s,m}$ for each p by inverting an infinite square matrix. In practice, we truncate the system. In effect, this means that we replace the summation over s by a finite number of terms, say N. Then, by letting $q = 1, 2, 3, \cdots, N$, we have a sufficient number of equations to solve for $B_{s,m}$. The convergence of solutions of this type as N becomes large has been discussed previously.[9,10] For the present situation, we focus our attention on the situation where $h_1 - h_2 \ll h_1$. In this instance, we can expect that

$$|S_q^{(1)} - S_q^{(2)}| \ll 1.$$

Furthermore, in such cases, it is found that $|\hat{\Gamma}_{q,m}| \ll |\Gamma_{q,q}|$ when $q \neq m$. Thus, in order to get a zero-order solution for a typical coefficient $B_{s,m}$ in (22), we retain only the terms for $s = q$ and $s = m$. First of all, when $q = m$, we easily get

$$B_{m,m} \simeq B_{m,m}^0 = -\left[\frac{S_m^{(2)}J'(k_0S_m^{(2)}a)J(k_0S_m^{(1)}a) - S_m^{(1)}J'(k_0S_m^{(1)}a)J(k_0S_m^{(2)}a)}{S_m^{(2)}J'(k_0S_m^{(2)}a)H(k_0S_m^{(1)}a) - S_m^{(1)}H'(k_0S_m^{(1)}a)J(k_0S_m^{(2)}a)}\right], \quad (23)$$

while, if $q \neq m$, we must deal with more terms in (22). In most cases it appears that

$$B_{q,m} \simeq B_{q,m}^0 = -B_{m,m}^0 P, \quad (24)$$

where P is of the order of 1. The zero-order solutions given by (23) and (24) should be applicable for the dominant modes such that $S_q^{(1)}$ and $S_q^{(2)}$ are not significantly different from unity.

The final result for the scattered field is given by (11) with the coefficient $(B_{s,m})_p$ given approximately

[8] J. R. Wait, Can. J. Phys. **46**, 1979 (1968).

[9] E. Bahar and J. R. Wait, *Quasi Optics*, J. Fox, Ed. (Polytechnic Institute of Brooklyn Press, Brooklyn, New York, 1964), pp. 447–464.

[10] A. Wexler, IEEE Trans. Microwave Theory Tech. **15**, 508 (1967).

by (23) and (24) above, where we set $q = s$ and remember that the Bessel functions are of order p. Of course, if the accuracy is insufficient we should return to (22) and retain additional terms.

6. A SIMPLIFICATION

To provide insight into the behavior of the solution, we now consider a simplification which is valid when the perturbation is small. First of all, we note that, in the zeroth approximation, the scattered field is expressible in the form

$$\psi_{sc} \simeq \text{const} \times \sum_{q} \sum_{p=-\infty}^{+\infty} e^{-i\pi/2}(B^0_{q,m})_p H_p(k_0 S_q^{(1)}\rho) e^{ip\phi}, \tag{25}$$

which amounts to changing the order of the summations indicated in (11). First of all, we focus our attention on the terms where $q = m$. These are the contribution to the scattered field in mode m for an incident mode also of order m. Thus, according to (24), we have

$$(B^0_{m,m})_p = -\left[\frac{\dfrac{uJ_p'(u)}{J_p(u)} - \dfrac{vJ_p'(v)}{J_p(v)}}{\dfrac{uJ_p'(u)}{J_p(u)} - \dfrac{vH_p'(v)}{H_p(v)}}\right]\frac{J_p(v)}{H_p(v)}, \tag{26}$$

where $u = k_0 S_m^{(2)} a$ and $v = k_0 S_m^{(1)} a$. We now set $\Delta = u - v$ and expand (26) in a series in powers of Δ. To within a first order of smallness in Δ, we find that

$$(B^0_{m,m})_p \simeq [(\pi\Delta)/2i]v[J_p^2(v) - J_{p+1}(v)J_{p-1}(v)]. \tag{27}$$

In arriving at (27), we have used the following well-known identities for Bessel functions of argument v:

$$J_p H_p' - J_p' H_p = -2i/(\pi v),$$
$$J_p' = -(p/v)J_p + J_{p-1} = (p/v)J_p - J_{p+1}.$$

Another simplification in (25) is to replace the Hankel function of argument $k_0 S_s^{(1)}\rho$ by the first term of its asymptotic expansion [e.g., see (12) above]. This is justified at sufficiently large distances from the scatterer. In other words, we require that both $k_0 a \gg 1$ and $\rho \gg a$. Then, on using (27), we find from (25) that the scattered mth mode field is

$$\psi_{sc}]_m \simeq \left(\frac{2i}{\pi k_0 S_m^{(1)}\rho}\right) e^{-ik_0 S_m{}^{(1)}\rho}\,\frac{\pi v\Delta}{2i} \times \sum_{p=-\infty}^{+\infty} [J_p^2(v) - J_{p-1}(v)J_{p+1}(v)]e^{-ip\phi}, \tag{28}$$

omitting the constant term $(\Lambda_m^{(1)}/h_1)G_m^{(1)}(z_0)$. In order to sum these series, we make use of the following identities:

$$J_2\left(2v\sin\frac{\phi}{2}\right) = -\sum_{p=-\infty}^{+\infty} J_{p-1}(v)J_{p+1}(v)e^{-ip\phi} \tag{29}$$

and

$$J_0\left(2v\sin\frac{\phi}{2}\right) = \sum_{p=-\infty}^{+\infty} J_p^2(v)e^{-ip\phi}, \tag{30}$$

which follow immediately from formulas given by Erdelyi *et al.*[11] Then, on using the well-known identity

$$J_2(Z) = (2/Z)J_1(Z) - J_0(Z),$$

we see that (28) is expressible in the closed form

$$\psi_{sc}]_m \simeq \left(\frac{\pi}{2k_0 S_m^{(1)}\rho}\right)^{\frac{1}{2}} e^{-i\pi/4} e^{-ik_0 S_m{}^{(1)}\rho}\,\frac{\Delta}{\sin(\phi/2)} \times J_1\left(2k_0 S_m^{(1)} a\sin\frac{\phi}{2}\right), \tag{31}$$

where $\Delta = k_0 a(S_m^{(2)} - S_m^{(1)})$.

We see that the scattered field has a strong maximum in the forward direction. In fact, as $\phi \to 0$,

$$\frac{J_1(2k_0 S_m^{(1)} a\sin\phi/2)}{\sin\phi/2} \to k_0 S_m^{(1)} a.$$

Thus, the forward scattered field is proportional to a^2 (i.e., the area of the perturbation) and the contrast between the wavenumbers

$$k_0 S_m^{(2)} \quad \text{and} \quad k_0 S_m^{(1)}.$$

The development of the coefficients $(B^0_{q,m})$ for $q \neq n$ into a series development in Δ follows in a similar fashion to that outlined above. If we restrict attention to modes of order q such that both

$$|k_0 a(S_m^{(2)} - S_m^{(1)})| \ll 1 \quad \text{and} \quad |k_0 a(S_q^{(1)} - S_m^{(1)})| \ll 1,$$

then (24) reduces to

$$(B^0_{q,m})_p \sim (B^0_{m,m})_p P. \tag{32}$$

Thus, for the approximation indicated, it follows that the cross-coupled scattered field in mode q for an incident mode of order m is given by

$$\psi_{sc}]_q \sim \psi_{sc}]_m P, \tag{33}$$

where P is roughly proportional to

$$S_{q,m} = \frac{\hat{\Gamma}_{q,m}}{\Gamma_{q,q}} = \frac{\displaystyle\int_0^{h_2} G_m^{(1)}(z)G_q^{(2)}(z)\,dz}{\displaystyle\int_0^{h_2} [G_q^{(2)}(z)]^2\,dz}. \tag{34}$$

[11] A. Erdelyi, Ed., *Higher Transcendental Functions, Vol. 2* (McGraw-Hill Book Co., Inc., New York, 1953), p. 101.

7. MODE COUPLING

The coupling parameter given by (34) may be expressed in closed form. We proceed by using a variant of the standard analysis for demonstrating orthogonality of modes in a closed system.[12]

In region (1), for a mode of order m, we have

$$\frac{\partial^2 G_m^{(1)}}{\partial z^2} + [k^2(z) - (k_0 S_m^{(1)})^2] G_m^{(1)} = 0, \quad (35a)$$

while in region (2), for a mode of order n,

$$\frac{\partial^2 G_n^{(2)}}{\partial z^2} + [k^2(z) - (k_0 S_n^{(2)})^2] G_n^{(2)} = 0. \quad (35b)$$

Multiplying (35a) by $G_n^{(2)}$ and (35b) by $G_m^{(1)}$ and subtracting the resulting equations, we find that

$$\frac{\partial}{\partial z}\left[G_n^{(2)} \frac{\partial G_m^{(1)}}{\partial z} - G_m^{(1)} \frac{\partial G_m^{(2)}}{\partial z}\right] = [(k_0 S_n^{(2)})^2 - (k_0 S_n^{(2)})^2] G_m^{(1)} G_n^{(2)}. \quad (36)$$

Both sides of (36) are now integrated over the interval from 0 to h_2. Thus, we find that

$$\hat{\Gamma}_{n,m} = \int_0^{h_2} G_m^{(1)}(z) G_n^{(2)}(z)\, dz = \frac{[G_m^{(1)'}(h_2) + K^{(2)} G_m^{(1)}(h_2)] G_n^{(2)}(h_2)}{(k_0 S_m^{(1)})^2 - (k_0 S_n^{(2)})^2}, \quad (37)$$

where $G^{(1)'}(h_2) = \partial G^{(1)}/\partial z$ evaluated at $z = h_2$. Also, in obtaining (37), we have used the boundary condition

$$\left[\frac{\partial G^{(2)}}{\partial z} + K^{(2)} G^{(2)}\right]_{z=h_2} = 0. \quad (38)$$

In a similar manner, we find that, for $n \neq m$,

$$\Gamma_{n,m} = \int_0^{h_2} G_m^{(2)}(z) G_n^{(2)}(z)\, dz = 0,$$

which is a demonstration of the orthogonality property already used. For the case $m = n$, we consider the height-gain function $G^{(2)}(z)$, in region (2), which satisfies

$$\frac{\partial^2 G^{(2)}}{\partial z^2} + [k^2(z) - (k_0 S)^2] G^{(2)} = 0. \quad (39)$$

Combining (35b) and (39), we obtain

$$\frac{\partial}{\partial z}\left[G^{(2)} \frac{\partial}{\partial z} G_n^{(2)} - G_n^{(2)} \frac{\partial}{\partial z} G^{(2)}\right] = [(k_0 S_n^{(2)}) - (k_0 S)^2] G_n^{(2)} G^{(2)}.$$

We again integrate both sides with respect to z over the range 0 to h_2 and assume that *both* $G_n^{(2)}$ and $G^{(2)}$ satisfy the impedance boundary conditions at $z = h_2$, but only $G_n^{(2)}$ satisfies the impedance boundary condition $G_n^{(2)} = K_0 G_n^{(2)}$ at $z = 0$. Thus, we find that

$$\int_0^{h_2} G_n^{(2)}(z) G^{(2)}(z)\, dz = -\frac{G_n^{(2)}(0)}{k_0^2[S^2 - (S_n^{(2)})^2]}\left[\frac{\partial}{\partial z} G^{(2)}(z) - K_0 G^{(2)}(z)\right]_{z=0}. \quad (40)$$

We *now* require that $G^{(2)}(z)$ satisfy the same impedance boundary condition as $G_n^{(2)}(z)$ at $z = 0$. Thus, $S \to (S_n^{(2)})$. Then, on applying Cauchy's theorem to the right-hand side of (40), we get

$$\Gamma_{n,n} = \int_0^{h_2} [G_n^{(2)}(z)]^2\, dz = -\frac{G_n^{(2)}(0)}{2k_0^2 S_n^{(2)}}\left\{\frac{\partial}{\partial S}[G^{(2)'}(z) - K_0 G(z)]\right\}_{\substack{S=S_n^{(2)} \\ z=0}}. \quad (41)$$

An explicit evaluation of this normalizing factor requires that we are able to solve (39) subject to the boundary condition at $z = h_2$. For example, if $k^2(z)$ is a linear function of z, the solutions of (39) are Airy functions, and a relatively simple closed form for $\Gamma_{n,n}$ is available.[5] The important point is that $\Gamma_{n,n}$ is of the order of h_2 for the propagating modes of low attenuation. On the other hand, $\hat{\Gamma}_{n,m}$ (for $n \neq m$) is much less than h_2 when $h_1 - h_2 \ll h_1$. Thus, we are justified in our exploiting the condition $|\hat{\Gamma}_{n,m}/\Gamma_{n,n}| \ll 1$ in considering first-order perturbation effects.

More work on this subject is obviously needed if we are to obtain useful information about acoustic-wave propagation (with or without gravity) in the nonuniform atmospheric ducts.

[12] A. N. Sommerfeld, *Partial Differential Equations* (Academic Press Inc., New York, 1960), pp. 81–91.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Ray Velocity and Exceptional Waves: A Covariant Formulation

GUY BOILLAT
Département de Mathématiques, Université de Clermont, Clermont-Ferrand, France

(Received 1 October 1968)

In a nonlinear field an accelerated wave sooner or later turns into a shock. When this is not the case, the wave is exceptional (e.g., Alfvén waves of magnetohydrodynamics). Then the normal speed of the wave remains undisturbed. This criterion is given a convenient covariant form in terms of the ray velocity. As an example a thermodynamical relativistic fluid is studied.

1. INTRODUCTION

The field is represented by a column vector $\mathbf{v}(x^\alpha)$ of N components, functions of the coordinates x^α ($\alpha = 0, 1, 2, \cdots, n$); $x^0 = t$ is a time coordinate and x^i ($i = 1, 2, \cdots, n$) are space coordinates. We assume that the field satisfies a set of N partial differential equations. By differentiating (if necessary) some of these equations with respect to t, we obtain a system of quasilinear differential equations which through the introduction of new dependent variables[1–3] can always be reduced to the form

$$A^\alpha(\mathbf{u}, x^\beta)\mathbf{u}_\alpha = \mathbf{f}(\mathbf{u}, x^\beta). \tag{1}$$

(The subscript α denotes partial differentiation with respect to x^α.)

We assume that across some characteristic (wave) surface,

$$\varphi(x^\alpha) = 0, \tag{2}$$

the first-order derivatives of the field $\mathbf{u}$ have a finite jump for which we use the following notations:

$$\left[\frac{\partial \mathbf{u}}{\partial \varphi}\right] = \boldsymbol{\pi} = \delta\mathbf{u}, \tag{3}$$

while the field itself is continuous

$$[\mathbf{u}] = 0.$$

Now it is well known that (2) must be a solution of the characteristic equation

$$|A^\alpha \varphi_\alpha| = 0, \quad \varphi_\alpha = \partial_\alpha \varphi,$$

or

$$|A_n - \lambda I| = 0, \quad A_n = A^i n_i,$$

if we assume $A^0 = I$ and if we put

$$\lambda = -\frac{\varphi_t}{|\nabla\varphi|}, \quad \mathbf{n} = \frac{\nabla\varphi}{|\nabla\varphi|}. \tag{4}$$

We suppose that the system (1) is hyperbolic; i.e., that the matrix

$$A_n - \lambda I \tag{5}$$

has a full set of linearly independent eigenvectors.

It is also well known that the discontinuities—or (weak) disturbances—(3) propagate along rays. Their laws of propagation can be given a very simple form when the waves invade a constant state and the matrices A^α do not depend explicitly on the x^β.[4] When a wave is accelerated the disturbances grow until they cease to be finite, thus tending to a shock. (Nonhyperbolic fields show a different behavior.[3])

However, waves might exist for which this phenomenon does not arise. This remarkable fact, first noted by Lax,[5] requires that the gradient of the normal velocity λ with respect to the field components be orthogonal to all the corresponding right eigenvectors of (5); this is equivalent to

$$\delta\lambda(\mathbf{u}, \mathbf{n}) \equiv \nabla\lambda \cdot \delta\mathbf{u} \equiv 0. \tag{6}$$

A wave whose normal speed is not disturbed (according to this equation) is called, after Lax, an exceptional wave. If all the waves of a given field are exceptional, the field is said to be completely exceptional. For instance, Alfvén waves of magnetohydrodynamics are exceptional.[2,4,6]

2. RAY VELOCITY

If the field equations are covariant the characteristic equations also have a covariant form. Thus,

$$\psi = 0, \quad \psi = p^{-1}G^{\alpha_1\alpha_2\cdots\alpha_p}\varphi_{\alpha_1}\varphi_{\alpha_2}\cdots\varphi_{\alpha_p},$$
$$\alpha_1, \alpha_2, \cdots, \alpha_p = 0, 1, \cdots, n, \tag{7}$$

where $G(\mathbf{u})$ is a completely symmetric tensor which is not a product of tensors of lower orders.

[1] R. Courant and D. Hilbert, *Methods of Mathematical Physics II* (Interscience Publishers, Inc., New York, 1962).
[2] A. Jeffrey and T. Taniuti, *Non-Linear Wave Propagation* (Academic Press Inc., New York, 1964).
[3] G. Boillat, "Définition et propagation des ondes" (to be published).
[4] G. Boillat, *La propagation des ondes* (Gauthier-Villars, Paris, 1965).
[5] P. D. Lax, Ann. Math. Studies **33**, 211 (1954); Commun. Pure Appl. Math. **10**, 537 (1957).
[6] K. O. Friedrichs and H. Kranzer, Courant Institute of Mathematical Sciences, New York University, Report NYO-6486, 1958.

The ray velocity has the components

$$\partial\psi/\partial\varphi_\alpha. \tag{8}$$

It is the speed of propagation of the disturbances and therefore it must not be, in general relativity, a spacelike vector, i.e., we must have

$$\mathcal{N}^2 = g^{\alpha\beta}\frac{\partial\psi}{\partial\varphi_\alpha}\frac{\partial\psi}{\partial\varphi_\beta} \geq 0; \tag{9}$$

$g_{\alpha\beta}$ is the metric tensor of signature -2.

If $\mathcal{N} \neq 0$, we can normalize (8); thus,

$$u^\alpha = \frac{1}{\mathcal{N}}\frac{\partial\psi}{\partial\varphi_\alpha}, \quad u_\alpha u^\alpha = 1. \tag{10}$$

From (7) we have

$$u^\alpha\varphi_\alpha = 0.$$

Of special interest are the cases $p = 1$ and $p = 2$. When $p = 1$, Eq. (7) is simply

$$u^\alpha\varphi_\alpha = 0,$$

and the u^α are field variables. The nonrelativistic equivalent is

$$\lambda = \mathbf{u}\cdot\mathbf{n}.$$

This kind of wave appears in fluid moving with velocity $\mathbf{u}$.

When $p = 2$, then

$$\psi = \tfrac{1}{2}G^{\alpha\beta}\varphi_\alpha\varphi_\beta.$$

We assume that G is a normal tensor[7]: its eigenvalues $s_{(\alpha)}$ are real, different from zero; one eigenvector (corresponding to $s_{(0)}$) is timelike and the three others are spacelike. In the eigenframe

$$G^{\alpha\beta} = \text{diag}\,(s_{(0)}, -s_{(1)}, -s_{(2)}, -s_{(3)}),$$
$$g^{\alpha\beta} = \text{diag}\,(1, -1, -1, -1),$$

so that $\psi = 0$ becomes

$$\lambda^2 = \sum_{i=1}^{3}\frac{s_{(i)}}{s_{(0)}}n_i^2.$$

As $\lambda^2 \leq 1$ is to be satisfied for any direction $\mathbf{n}$, we must have

$$0 < s_{(i)}/s_{(0)} \leq 1, \quad i = 1, 2, 3. \tag{11}$$

(The restriction $s_{(i)} \neq 0$ allows one to express the φ_α in terms of the u^β.) Then the condition (9) is automatically satisfied.

We note that, if one equality holds in (11), light velocity is reached in one direction:

$$s_{(1)} = s_{(0)} \to \lambda^2 = 1, \quad \mathbf{n} = (\pm 1, 0, 0).$$

[7] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme* (Masson, Paris, 1955).

3. EXCEPTIONAL WAVES

Inserting (4) into (7) we have

$$|\nabla\varphi|\frac{\partial\psi}{\partial\varphi_0}\nabla\lambda = \nabla\psi = \frac{1}{p}\varphi_{\alpha_1}\varphi_{\alpha_2}\cdots\varphi_{\alpha_p}\nabla G^{\alpha_1\alpha_2\cdots\alpha_p}.$$

Multiplying by $\delta\mathbf{u}$ and taking account of (6), we see that the waves (7) are exceptional provided[8] that

$$\delta\psi \equiv 0, \quad \delta\psi = p^{-1}\varphi_{\alpha_1}\varphi_{\alpha_2}\cdots\varphi_{\alpha_p}\delta G^{\alpha_1\alpha_2\cdots\alpha_p} \tag{12}$$

when (7) is satisfied.

From (10) we get

$$\nabla u^\alpha = \frac{1}{\mathcal{N}}(\delta^\alpha_\beta - u^\alpha u_\beta)\frac{\partial}{\partial\varphi_\beta}(\nabla\psi).$$

Hence,

$$\delta u^\alpha = \mathcal{N}^{-1}(\delta^\alpha_\beta - u^\alpha u_\beta)\varphi_{\alpha_2}\varphi_{\alpha_3}\cdots\varphi_{\alpha_p}\delta G^{\beta\alpha_2\alpha_3\cdots\alpha_p},$$

and the condition (12) can also be written

$$\varphi_\alpha\delta u^\alpha \equiv 0.$$

(The metric tensor $g_{\alpha\beta}$ only admits discontinuities of the second order across a null surface.[7])

By applying the criterion (12) to nonlinear electrodynamics we selected an exceptional Lagrangian which generalizes the Born–Infeld Lagrangian and gives birth to a completely exceptional system of field equations.[9]

4. THERMODYNAMICAL FLUID

The field equations of a perfect thermodynamical fluid derived from a nonsymmetric energy tensor can be written[10]

$$\nabla_\alpha(ru^\alpha) = 0,$$
$$u^\alpha(r\partial_\alpha f - \partial_\alpha p) - \nabla_\alpha q^\alpha = 0,$$
$$(rfu^\alpha - q^\alpha)\nabla_\alpha u^\beta - \gamma^{\alpha\beta}\partial_\alpha p = 0,$$
$$\gamma^{\alpha\beta} = g^{\alpha\beta} - u^\alpha u^\beta, \quad u_\alpha u^\alpha = 1.$$

The metric tensor is given; r, f, p, and u^α, are respectively the density, index, pressure, and velocity of the fluid. The heat-current vector is assumed to be a given function of the field variables

$$q^\alpha = q^\alpha(u^\beta, p, \phi, \phi_\gamma), \quad \phi_\gamma = \partial_\gamma\phi.$$

Furthermore it is supposed that

$$r = r(p, \phi), \quad f = f(p, \phi),$$

where ϕ is some thermodynamical quantity (e.g., the temperature) which can be discontinuous in the second

[8] G. Boillat, Compt. Rend. **262A**, 1285 (1966).

[9] G. Boillat, "Non-Linear Electrodynamics; Lagrangians and Equations of Motion" (to be published).

[10] Y. Bruhat, Commun. Math. Phys. **3**, 334 (1966).

order ($\delta\phi = 0$). Since our purpose is merely to illustrate the preceding sections we do not go on to discuss these hypotheses.

The wave surfaces have already been determined in a study of the Cauchy problem,[10] but as we also need the disturbances we make the calculations. Making the replacement

$$\nabla_\alpha \to \varphi_\alpha \delta$$

in the field equations, we obtain

$$u^\alpha \varphi_\alpha r' \delta p + r\varphi_\alpha \delta u^\alpha = 0, \tag{13}$$

$$u^\alpha \varphi_\alpha (rf' - 1)\delta p - \varphi_\alpha \delta q^\alpha = 0, \tag{14}$$

$$(rfu^\alpha - q^\alpha)\varphi_\alpha \delta u^\beta - \gamma^{\alpha\beta}\varphi_\alpha \delta p = 0, \tag{15}$$

where the prime denotes partial differentiation with respect to p. The solutions are as follows:

(a) $(rfu^\alpha - q^\alpha)\varphi_\alpha = 0$, together with

$$\varphi_\alpha \delta u^\alpha = \varphi_\alpha \delta q^\alpha = \delta p = 0. \tag{16}$$

Equation (9) yields

$$(rf - q_\alpha u^\alpha)^2 \geq -\gamma^{\alpha\beta} q_\alpha q_\beta, \tag{17}$$

an inequality already given.[10] But we shall find stronger conditions.

It appears at once, by virtue of (16), that this wave is exceptional:

$$\varphi_\alpha \delta(rfu^\alpha - q^\alpha) \equiv 0.$$

(b) Multiplying (15) by φ_β and taking account of (13) we get

$$G^{\alpha\beta}\varphi_\alpha \varphi_\beta = 0, \quad \delta p \neq 0,$$

$$G^{\alpha\beta} = g^{\alpha\beta} + u^\alpha v^\beta + u^\beta v^\alpha,$$

$$v^\alpha = \tfrac{1}{2}(r'f - 1)u^\alpha - \tfrac{1}{2}r'r^{-1}q^\alpha.$$

Let V_α be an eigenvector corresponding to the eigenvalue s

$$(G^{\alpha\beta} - sg^{\alpha\beta})V_\alpha = 0. \tag{18}$$

To the eigenvalue $s = 1$ correspond two eigenvectors satisfying

$$u^\alpha V_\alpha = v^\alpha V_\alpha = 0.$$

These eigenvectors are necessarily spacelike for they are orthogonal to the timelike vector u^α. Thus we can put

$$s_{(2)} = s_{(3)} = 1.$$

Now we look for a and b such that

$$V_\alpha = au_\alpha + bv_\alpha.$$

Inserting this expression into (18) results in

$$a(1 - s + u_\alpha v^\alpha) + bv_\alpha v^\alpha = 0,$$

$$a + b(1 - s + u_\alpha v^\alpha) = 0,$$

whence we get

$$(1 - s + u_\alpha v^\alpha)^2 - v_\alpha v^\alpha = 0$$

and the necessary condition

$$v_\alpha v^\alpha > 0,$$

i.e.,

$$[rf - q_\alpha u^\alpha - (r/r')]^2 > -\gamma^{\alpha\beta} q_\alpha q_\beta. \tag{19}$$

Then,

$$s = 1 + u_\alpha v^\alpha (1 \pm w),$$

$$w^2 = v_\alpha v^\alpha / (u_\beta v^\beta)^2, \quad 0 < w < 1,$$

and

$$V_\alpha V^\alpha = \pm 2b^2 (u_\alpha v^\alpha)^2 w(1 \pm w).$$

$V_\alpha V^\alpha > 0$ corresponds to the upper sign, and $V_\alpha V^\alpha < 0$ to the lower one. Therefore,

$$s_{(0)} = 1 + u_\alpha v^\alpha (1 + w), \quad s_{(1)} = 1 + u_\alpha v^\alpha (1 - w)$$

and according to (11) we must have

$$u_\alpha v^\alpha > 0,$$

that is,

$$(r'/r)(rf - q_\alpha u^\alpha) > 1. \tag{20}$$

It is easy to see that (19) and (20) are more restrictive than (17) by noting that

$$(rf - q_\alpha u^\alpha)^2 > [rf - q_\alpha u^\alpha - (r/r')]^2$$

if (20) is true.

It can be checked that these waves are not exceptional.

(c) The remaining solution is

$$\delta u^\alpha = \delta p = 0, \quad \varphi_\alpha \delta q^\alpha = 0.$$

Since

$$\delta\phi_\beta = \varphi_\beta \delta^2 \phi,$$

these equations give

$$\frac{\partial q^\alpha}{\partial \phi_\beta} \varphi_\alpha \varphi_\beta = 0$$

or

$$a^{\alpha\beta}\varphi_\alpha \varphi_\beta = 0, \quad a^{\alpha\beta} = \frac{1}{2}\left(\frac{\partial q^\alpha}{\partial \phi_\beta} + \frac{\partial q^\beta}{\partial \phi_\alpha}\right).$$

The wave is exceptional if

$$\frac{\partial a^{\alpha\beta}}{\partial \phi_\gamma} \varphi_\alpha \varphi_\beta \varphi_\gamma = 0.$$

It is the case if the heat current is a linear function of the gradient of ϕ and, more generally, if there exists a vector a^α, which is a function of the field variables, such that

$$\mathop{S}_{\alpha,\beta,\gamma}\left(\frac{\partial a^{\alpha\beta}}{\partial \phi_\gamma} - a^\gamma a^{\alpha\beta}\right) \equiv 0.$$

ACKNOWLEDGMENT

We thank Professor P. D. Lax for his kind interest in this work.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

General Approach to Fractional Parentage Coefficients

M. MOSHINSKY*
Instituto de Física, Universidad de México
AND
V. SYAMALA DEVI
Andhra University, Waltair, India

(Received 5 November 1967)

The purpose of this paper is to achieve a clearer understanding of the problems involved in the determination of a closed formula for fractional parentage coefficients (fpc). The connection between the fpc and one-block Wigner coefficients of a unitary group of dimension equal to that of the number of states is explicitly derived. Furthermore, these Wigner coefficients are decomposed into ones characterized by a canonical chain of subgroups (for which an explicit formula is given) and transformation brackets from the canonical to the physical chain. It is in the explicit and systematic determination of the states in the latter chain where the main difficulty appears. We fully analyze the case of the p shell to show that a complete nonorthonormal set of states in the physical chain $\mathfrak{U}(3) \supset R(3)$ can be derived easily using Littlewood's procedure for the reduction of irreducible representations (IR) of $SU(3)$ with respect to the subgroup $R(3)$. This procedure gives a deeper understanding of the free exponent appearing in the polynomials in the creation operators defining the states in the $\mathfrak{U}(3) \supset R(3)$ chain. As Littlewood's procedure applies to the $\mathfrak{U}(n) \supset R(n)$ chain, and probably can be generalized to other noncanonical chains of groups, it opens the possibility of obtaining general closed formulas for the fpc in a nonorthonormal basis.

1. INTRODUCTION

The importance of the role that fractional parentage coefficients (fpc) play in problems of atomic and molecular structure has been appreciated for a long time. In particular, Racah,[1] Jahn,[2] and Flowers[3] have stressed their usefulness and have given tables obtained by systematic application of projection techniques.

Yet we do not have for the fpc the type of closed formula we are familiar with in the case of Wigner or Racah coefficients of the rotation group. The purpose of this article is to try to understand fully the reasons that have prevented the derivation of these closed formulas in the past, and to indicate the recent developments that could make possible a systematic derivation of such formulas in the future.

For the sake of clearness, we shall restrict ourselves in this article to the discussion of fpc in the configuration space of a single orbital, and in some aspects of the analysis, we shall further limit this orbital to the angular momentum $l = 1$, i.e., the p shell. Yet the approach will be such that everything we say can be extended, (and in some cases will already be applicable) to problems of several orbitals, to states in spin–isospin space, to spin-orbital states (i.e., j–j coupling) and even to some aspects of the fpc for n-particle harmonic-oscillator states developed recently.[4]

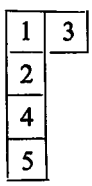
While the reader is undoubtedly familiar with the concept of fpc, we would like to review it briefly so as to formulate it in a way useful to the type of approach we shall outline in this article.

In a fixed orbital, the single-particle states are characterized by the projection m of the orbital angular momentum, as the total quantum number v and the angular momentum l are fixed. Thus these single particle states could be denoted by

$$\psi_{vlm}(\mathbf{r}^s) \equiv \psi_m(\mathbf{r}^s), \quad m = l, \cdots, -l, \tag{1.1}$$

where $s = 1, \cdots, n$ is the index characterizing the particle.

If we have a system of n particles in the vl orbital, we could form an n-particle state characterized by the total angular momentum L and projection M, i.e., irreducible representations (IR) of the chain of rotation groups $R(3) \supset R(2)$, as well as by a partition $f \equiv [f_1 f_2, \cdots, f_n]$, $f_1 \geq \cdots \geq f_n \geq 0$ of n and a Yamanouchi symbol $r \equiv (1 r_2, \cdots, r_{n-1} r_n)$, i.e., an IR of the chain of symmetric groups $S(n) \supset S(n-1) \supset \cdots \supset S(1)$. An example of f and r is given below:



$$f = [2111], \quad r = (12134). \tag{1.2}$$

* Asesor Director de la Comisión Nacional de Energía Nuclear (México).

[1] G. Racah, Phys. Rev. **76**, 1352 (1949).

[2] H. A. Jahn, Proc. Roy. Soc. (London) **A201**, 516 (1950); **A205**, 192 (1951), and following articles of the series.

[3] B. H. Flowers, Proc. Roy. Soc. (London) **A212**, 248 (1952); A. R. Edmonds and B. H. Flowers, Proc. Roy. Soc. (London) **A214**, 515 (1952); **A215**, 120 (1952).

[4] P. Kramer and M. Moshinsky, Nuclear Physics **82**, 241 (1966); also in *Group Theory and Applications*, E. M. Loebl, Ed. (Academic Press Inc., New York, 1968).

The n-particle states are then represented by the kets

$$|l^n fr\Omega LM\rangle = \sum_{m_i} A^{m_1\cdots m_n}_{fr\Omega LM} \psi_{m_1}(\mathbf{r}^1)\cdots\psi_{m_n}(\mathbf{r}^n), \quad (1.3)$$

where we indicate by Ω the additional quantum numbers that complete the characterization of the states. The A's in (1.3) are the coefficients in a development of our ket in terms of the single-particle states (1.1). The numbers f, r, L, M would, of course, remain unchanged if we apply to (1.3) any operator invariant under the groups $R(3)$ and $S(n)$.

The fpc is then the coefficient required for the expansion of the state (1.3) in terms of a similar state of $n-1$ particles and the single state of particle n, coupled to a total orbital angular momentum L and projection M (denoted by the bracket notation $[\ ]_{LM}$), i.e., it is the last term in the following expansion:

$$|l^n fr\Omega LM\rangle = \sum_{\bar{\Omega}\bar{L}} \{[|l^{n-1}\bar{f}\bar{r}\bar{\Omega}\bar{L}\rangle\psi(\mathbf{r}^n)]_{LM} \times \langle l^{n-1}\bar{f}\bar{\Omega}\bar{L}\ ,\ l\,|\}\, l^n f\Omega L\rangle\}. \quad (1.4)$$

The partition $\bar{f}$ is obtained from f by suppressing the block containing n in the Young tableau, and the Yamanouchi symbols $\bar{r}$, r are related in the same way.

The usefulness of the development (1.4) lies in the fact that the matrix elements of a symmetric, one-body, irreducible tensor operator

$$\sum_{s=1}^{n} T_{kq}(\mathbf{r}^s, \mathbf{p}^s) \quad (1.5)$$

with respect to states (1.3) that are bases for the IR of $S(n)$ is the same as the matrix elements of the operator[1–3]

$$nT_{kq}(\mathbf{r}^n, \mathbf{p}^n), \quad (1.6)$$

which could be evaluated immediately using the expansion (1.4). The extension of this type of analysis to two-body operators by using double expansions of the type (1.4), i.e., two-body fpc is of course very familiar.[1–3]

It is known[1–3] that the fpc in (1.4) are connected with particular Wigner coefficients of $\mathfrak{U}(2l+1)$. We proceed to give a systematic derivation of this connection by a procedure developed recently[5a] for the group-theoretical characterization of an alternative description of the states (1.3). We later discuss the systematic evaluation of these Wigner coefficients.

2. THE fpc AS WIGNER COEFFICIENTS OF A UNITARY GROUP[5b]

We shall introduce the creation operators a^s_m with two indices $m = l, \cdots, -l$, $s = 1, \cdots, n$, and establish the following correlation between them and the single-particle states (1.1):

$$\psi_m(\mathbf{r}^s) \leftrightarrow a^s_m\,|0\rangle. \quad (2.1)$$

In (2.1), $|0\rangle$ stands for the vacuum state characterized by the property that

$$\bar{a}^s_m\,|0\rangle = 0 \quad \text{for any} \quad m, s, \quad (2.2)$$

with $\bar{a}^s_m$ being the annihilation operator associated with a^s_m, so it satisfies with the latter the commutation relations

$$[\bar{a}^{s'}_{m'}, a^s_m] = \delta^{s's} g_{m'm}, \quad (2.3)$$

$$g_{m'm} = (-1)^m \delta_{-mm'}. \quad (2.4)$$

The appearance of the metric tensor $g_{m'm}$ is due to the fact that, from the correlation (2.1), a^s_m, $m = l, \cdots, -l$ transforms under rotations as a basis for an IR l of $R(3)$, i.e., in the same way as $Y_{lm}(\theta, \varphi)$. The annihilation operator $\bar{a}^s_m$, which is the transposed conjugate of a^s_m, will transform then as

$$(-1)^m Y_{l-m}(\theta, \varphi),$$

hence the appearance of $g_{m'm}$ rather than Kronecker delta.

From the correlation (2.1) an alternative expression for the n-particle state (1.3) will be given by

$$|l^n fr\Omega LM\rangle = P_n(fr\Omega LM)\,|0\rangle \quad (2.5a)$$

where P_n is the homogeneous polynomial

$$P_n(fr\Omega LM) = \sum_{m_i} A^{m_1\cdots m_n}_{fr\Omega LM} a^1_{m_1}\cdots a^n_{m_n}. \quad (2.5b)$$

We note now that the operators a^s_m, $m = l, \cdots, -l$; $s = 1, \cdots, n$ can be thought of as the components of a vector of $[(2l+1)n]$ dimensions. The vector $\{a^s_m\}$ could then be characterized by the IR [1] of a unitary group $U[(2l+1)n]$ acting in a space of the same number of dimensions. The polynomials (2.5b), being of degree n in the components of the vector $\{a^s_m\}$, are characterized by the IR $[n]$, i.e., the completely symmetric one of the same group.

We proceed now to show that the states (2.5) are further characterized by the IR of the following chain of subgroups associated with the $(2l+1)$-dimensional space of the indices m and the n-dimensional space of

[5] (a) M. Moshinsky, J. Math. Phys. 7, 691 (1966). (b) The discussion in this section is based on the paper cited in (a). To have a more systematic notation, the creation operators $a^{\dagger}_{\mu s}$ and annihilation operators $a^{\mu s}$ in that paper will here be replaced by a^s_m and $\bar{a}^s_m$, respectively, and the notation for the generators of the unitary groups will be changed accordingly.

the indices s:

$$\mathbf{U}[(2l+1)n] \supset \mathfrak{U}(2l+1) \times U(n), \quad (2.6a)$$

$$\mathfrak{U}(2l+1) \supset \cdots \supset \mathfrak{D}^l(R(3)) \supset \mathfrak{D}^l\left(\begin{bmatrix} R(2) & 0 \\ 0 & 1 \end{bmatrix}\right), \quad (2.6b)$$

$$U(n) \supset \begin{bmatrix} U(n-1) & 0 \\ 0 & 1 \end{bmatrix} \supset \begin{bmatrix} U(n-2) & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\supset \cdots \supset \begin{bmatrix} U(1) & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}; \quad (2.6c)$$

where $\mathfrak{D}^l$ are the $(2l+1) \times (2l+1)$ unitary matrices associated with the IR l of $R(3)$.

The IR of a unitary group of j dimensions $U(j)$ are characterized by a partition $[f_{1j} \cdots f_{jj}]$ and those of $R(3)$ and $R(2)$, as before, by the angular momentum L and its projection M. A polynomial associated with the IR of the chains (2.6b), (2.6c) could then be characterized

$$P\begin{pmatrix} [h_1 \cdots h_{2l+1}] & f_{1n} f_{2n} & \cdots & f_{nn} \\ \Omega LM & ; & f_{1,n-1} \cdots f_{n-1,n-1} & \\ & & \cdots & \\ & & f_{11} & \end{pmatrix}, \quad (2.7)$$

where $[h_1 \cdots h_{2l+1}]$ is the partition characterizing the IR of $\mathfrak{U}(2l+1)$ and Ω stands for the extra quantum numbers necessary to characterize the states in the chain (2.6b), which could include numbers associated with the IR of subgroups between $\mathfrak{U}(2l+1)$ and $\mathfrak{D}^l(R(3))$.

If the polynomial (2.7) belongs to the IR $[n]$ of $\mathbf{U}[(2l+1)n]$, then the partitions

$$[h_1 \cdots h_{2l+1}], \quad [f_{1n} \cdots f_{nn}]$$

must be identical[5,6] partitions of n, i.e.,

$$h_1 = f_{1n}, \quad h_2 = f_{2n}, \cdots, \quad (2.8a)$$

$$f_{1n} + f_{2n} + \cdots + f_{nn} = n, \quad (2.8b)$$

which implies incidentally that the number of terms different from zero is $\leq \min(2l+1, n)$. Furthermore, if we want to correlate the polynomials (2.5b) and (2.7), we first note that the former are of degree one in each value $s = 1, \cdots, n$ of the upper index. As was shown in Ref. 5, this implies that the partition in (2.7) must be restricted by the conditions

$$(f_{1j} + \cdots + f_{jj}) - (f_{1,j-1} + \cdots + f_{j-1,j-1}) = 1, \quad \text{for } j = 1, \cdots, n. \quad (2.9)$$

When taking into account the restrictions (2.8) and (2.9) on the partitions, one realizes that they could be replaced by the set of numbers $(1r_2 \cdots r_n)$, where r_j indicates in which position on the jth row we must add a unit to get the partition $[f_{1j} \cdots f_{jj}]$ from $[f_{1,j-1} \cdots f_{j-1,j-1}]$. As an example, for $n = 4$, $f = [31]$, we indicate the set of partitions $[f_{ij}]$, $1 \leq i \leq j \leq 4$, and the corresponding Yamanouchi symbols:

$$\left|\begin{matrix} 3\ 1\ 0\ 0 \\ 2\ 1\ 0 \\ 2\ 0 \\ 1 \end{matrix}\right\rangle (1121); \quad \left|\begin{matrix} 3\ 1\ 0\ 0 \\ 2\ 1\ 0 \\ 1\ 1 \\ 1 \end{matrix}\right\rangle (1211);$$

$$\left|\begin{matrix} 3\ 1\ 0\ 0 \\ 3\ 0\ 0 \\ 2\ 0 \\ 1 \end{matrix}\right\rangle (1112). \quad (2.10)$$

A special class of polynomials in the components of $\{a_m^s\}$ characterized by IR of the chains of groups (2.6), i.e., those whose partitions are restricted by (2.8), (2.9), are then homogeneous polynomials of degree n characterized by the same quantum numbers as (2.5b). In Ref. 5 we actually show that these polynomials are identical if $f_{in} = f_i$, $i = 1, \cdots, n$, so in our further discussion of the polynomials (2.5b), we shall use for them the notation (2.7) with restrictions (2.8) and (2.9).

To determine now the fpc, let us consider a polynomial (2.7) in which the index s is restricted to $s = 1, \cdots, n-1$, i.e.,

$$P\begin{pmatrix} [\bar{h}_1 \cdots \bar{h}_{2l+1}] & f_{1,n-1} f_{2,n-1} \cdots f_{n-1,n-1} \\ \bar{\Omega} \bar{L} \bar{M} & ; & f_{1,n-2} \cdots f_{n-2,n-2} \\ & & \cdots \\ & & f_{11} \end{pmatrix}. \quad (2.11)$$

To represent a polynomial of $n-1$ particles of the type (2.5b), we must have again

$$\bar{h}_1 = f_{1n-1}, \quad \bar{h}_2 = f_{2,n-1}, \cdots; \quad (2.12a)$$

$$f_{1,n-1} + \cdots + f_{n-1,n-1} = n-1 \quad (2.12b)$$

and the restrictions (2.9). We could then couple the P of (2.11) and a_m^n to a given IR $[h_1 \cdots h_{2l+1}]$ of

[6] T. A. Brody, M. Moshinsky, and I. Renero, J. Math. Phys. 6, 1540 (1965).

$\mathfrak{U}(2l+1)$, and angular momentum L and projection M, by means of the Wigner coefficients of $\mathfrak{U}(2l+1)$ in the chain (2.6b), i.e.,

$$\sum_{\bar{\Omega}\bar{L}\bar{M}}\sum_{m}\left[P\begin{pmatrix}[h_1\cdots h_{2l+1}] & f_{1,n-1}f_{2,n-1}\cdots f_{n-1,n-1}\\ \bar{\Omega}\bar{L}\bar{M} \quad ; & f_{1,n-2}\cdots f_{n-2,n-2}\\ & \cdots\\ & f_{11}\end{pmatrix}a^n_m\right.$$
$$\left.\times\left\langle\begin{matrix}[h_1\cdots h_{2l+1}];\,[1]\\ \bar{\Omega}\bar{L}\bar{M}\quad lm\end{matrix}\middle|\begin{matrix}[h_1\cdots h_{2l+1}]\\ \Omega LM\end{matrix}\right\rangle\right]. \quad (2.13)$$

The resulting polynomial will be homogeneous and of degree n and so correspond to the IR $[n]$ of $\mathbf{U}[(2l+1)n]$. Furthermore, by definition it will be characterized by the IR $[h_1\cdots h_{2l+1}]$ of $\mathfrak{U}(2l+1)$ and so from (2.8a) by the IR $[f_{1n}\cdots f_{nn}]$ of $U(n)$. As a^n_m is invariant under the subgroup $U(n-1)$ of $U(n)$, it is clear that (2.13) leads to the polynomial (2.7). Comparing this with (1.4), and using the orthonormality property of the Wigner coefficients of $R(3)$ and the orthogonality of bases of IR of groups,[7] we conclude that if

$$\bar{h}_1=\bar{f}_1,\quad \bar{h}_2=\bar{f}_2,\cdots;$$
$$h_1=f_1,\quad h_2=f_2,\cdots, \quad (2.14)$$

then

$$\langle l^{n-1}\bar{f}\bar{\Omega}\bar{L};\,l\,|\}\,l^n f\Omega L\rangle^{\bar{\gamma}}$$
$$=\sum_{\bar{M}m}\left\{\langle L l\bar{M}m\mid LM\rangle\left\langle\begin{matrix}[\bar{f}_1\bar{f}_2\cdots];\,[1]\\ \bar{\Omega}\bar{L}\bar{M}\quad lm\end{matrix}\middle|\begin{matrix}[f_1f_2\cdots]\\ \Omega LM\end{matrix}\right\rangle\right\} \quad (2.15)$$

with the f and $\bar{f}$ related as in Sec. 1.

The problem of the fpc reduces then to the one of determining the Wigner coefficients of $\mathfrak{U}(2l+1)$ in (2.15). This can be accomplished in two steps. First we consider for the $\mathfrak{U}(2l+1)$ group the chain of subgroups

$$\mathfrak{U}(2l+1)\supset\begin{bmatrix}\mathfrak{U}(2l) & 0\\ 0 & 1\end{bmatrix}$$
$$\supset\cdots\supset\begin{bmatrix}\mathfrak{U}(1) & & & \\ & 1 & & \\ & & \cdot & \\ & & & \cdot \\ & & & & 1\end{bmatrix}. \quad (2.16)$$

This would give us kets, now called Gel'fand states,[8] characterized by the partitions

$$\left|\begin{matrix}f_{1k}f_{2k}\cdots & f_{kk}\\ f_{1,k-1}\cdots f_{k-1,k-1}\\ \cdot\\ \cdot\\ \cdot\\ f_{11}\end{matrix}\right\rangle, \quad k\equiv 2l+1. \quad (2.17)$$

We could then determine the transformation brackets between the states in the chain (2.6b) and those in the chain (2.17), i.e.,

$$\left\langle\begin{matrix}f_{1k}f_{2k}\cdots & f_{kk}\\ f_{1,k-1}\cdots f_{k-1,k-1}\\ \cdot\\ \cdot\\ \cdot\\ f_{11}\end{matrix}\middle|\begin{matrix}[f_{1k}\cdots f_{kk}]\\ \Omega LM\end{matrix}\right\rangle. \quad (2.18)$$

Once these are available, we can use them to reduce the Wigner coefficients in (2.13) to the corresponding ones of $\mathfrak{U}(k)$, $k\equiv 2l+1$ in the canonical chain, i.e.,

$$\left\langle[\bar{f}_{ij}],\begin{bmatrix}1\\ \mu\end{bmatrix}\middle|[f_{ij}]\right\rangle\equiv\left\langle\begin{matrix}\bar{f}_{1k}\bar{f}_{2k} & \cdots & \bar{f}_{kk} & 1\,0 & & \cdots & 0\\ \bar{f}_{1,k-1} & \cdots & \bar{f}_{k-1,k-1} & & 1\,0 & \cdots & 0\\ & \cdots & & & & \cdots & \\ & \bar{f}_{11} & & ; & & 1\,0\cdots 0 & \\ & & & & & 0\cdots 0 & \\ & & & & & \cdots & \\ & & & & & 0 & \end{matrix}\middle|\begin{matrix}f_{1k}f_{2k} & \cdots & f_{kk}\\ f_{1,k-1} & \cdots & f_{k-1,k-1}\\ & \cdots & \\ & f_{11} & \end{matrix}\right\rangle, \quad (2.19)$$

[7] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959), p. 115.
[8] I. M. Gel'fand and M. I. Zetlin, Dokl. Akad. Nauk SSSR 71, 825 (1950).

where we have

$$\left|\begin{matrix}[1]\\ lm\end{matrix}\right\rangle \equiv \left|\begin{bmatrix}1\\ \mu\end{bmatrix}\right\rangle = \left|\begin{matrix} 1\ 0 & \cdots & 0 \\ 1\ 0 & \cdots & 0 \\ & \cdots & \\ 1\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ & \cdots & \\ & 0 & \end{matrix}\right\rangle, \quad (2.20a)$$

$$\mu = l - m + 1 \quad (2.20b)$$

where there are $k - \mu$ rows "1 0 $\cdots$ 0" and μ rows "0 $\cdots$ 0," and

$$f_{ij} = \bar{f}_{ij} + \delta_{il_j} \sum_{j'=\mu}^{k} \delta_{jj'}. \quad (2.21)$$

In the next section we proceed to give a closed general expression for (2.19), while in Sec. 4 we discuss the transformation brackets (2.18).

3. ONE-BLOCK WIGNER COEFFICIENTS IN THE CANONICAL CHAIN

As the most general unitary matrix could be written as

$$U = e^{i\delta}U', \quad \det U' = 1, \quad (3.1)$$

where U' is unitary unimodular, the Wigner coefficient (2.19) of $\mathfrak{U}(k)$ is identical to the corresponding one[5] of $S\mathfrak{U}(k)$

$$\left\langle [\bar{f}'_{ij}], \begin{bmatrix}1\\ \mu\end{bmatrix} \middle| [f'_{ij}] \right\rangle, \quad (3.2a)$$

where

$$\bar{f}'_{ij} = \bar{f}_{ij} - \bar{f}_{kk}, \quad f'_{ij} = f_{ij} - \bar{f}_{kk}. \quad (3.2b)$$

As $\bar{f}'_{kk} = 0$, the state associated with the partitions $[\bar{f}'_{ij}]$ will be a polynomial function only of the vectors a^s_m, $s = 1, \cdots, k-1$, acting on $|0\rangle$, so the independent state $\begin{bmatrix}1\\ \mu\end{bmatrix}$ could be characterized by the vector a^k_m acting on $|0\rangle$. It is convenient to denote the components of the vector by the index μ defined in (2.20b) that takes the more natural sequence of values $\mu = 1, 2, \cdots, k$, in which case we could write

$$\left|\begin{bmatrix}1\\ \mu\end{bmatrix}\right\rangle \equiv a^k_\mu |0\rangle, \quad \mu = 1, \cdots, k. \quad (3.3)$$

From (3.3) the Wigner coefficient (3.2), or equivalently (2.19), reduces to the matrix element

$$\left\langle \begin{matrix} f'_{1k} f'_{2k} \cdots & f'_{kk} \\ f'_{1,k-1} \cdots f'_{k-1,k-1} & \\ \cdots & \\ f'_{11} & \end{matrix} \middle| a^k_\mu \middle| \begin{matrix} \bar{f}'_{1k} & \cdots \bar{f}'_{k-1,k} & 0 \\ \bar{f}'_{1,k-1} & \cdots \bar{f}'_{k-1,k-1} & \\ & \cdots & \\ & \bar{f}'_{11} & \end{matrix} \right\rangle. \quad (3.4)$$

To evaluate the latter, let us consider the generators $\mathcal{C}_{\mu\mu'}$ of $\mathfrak{U}(k+1)$ in a Cartesian metric, i.e.,

$$\mathcal{C}_{\mu\mu'} = \sum_s a^s_\mu \bar{a}^s_{\mu'}, \quad [\bar{a}^{s'}_{\mu'}, a^s_\mu] = \delta^{ss'}\delta_{\mu\mu'},$$
$$\mu, \mu' = 1, \cdots, k, k+1. \quad (3.5)$$

Clearly $\mathcal{C}_{\mu k+1}$, $\mu = 1, 2, \cdots, k$ has the same transformation properties as a^s_μ, $\mu = 1, 2, \cdots, k$ under the group $\mathfrak{U}(k)$, so the Wigner–Eckart theorem leads to

$$\left\langle \begin{matrix} f'_{1k} & \cdots & f'_{k-k1,k} f'_{kk} \\ f'_{1,k-1} \cdots f'_{k-1,k-1} & & \\ \cdots & & \\ f'_{11} & & \end{matrix} \middle| a^k_\mu \middle| \begin{matrix} \bar{f}'_{1k} & \cdots \bar{f}'_{k-1,k} & 0 \\ \bar{f}'_{1,k-1} & \cdots \bar{f}'_{k-1,k-1} & \\ & \cdots & \\ & \bar{f}'_{11} & \end{matrix} \right\rangle = \langle f'_{1k} \cdots f'_{k-1,k} f'_{kk} || a^k || \bar{f}'_{1k} \cdots \bar{f}'_{k-1,k} \ 0 \rangle$$
$$\times \left\langle \begin{matrix} f'_{1k} & \cdots & f'_{kk} \ 0 \\ f'_{1k} & \cdots f'_{k-1,k} f'_{kk} & \\ f'_{1,k-1} \cdots f'_{k-1,k-1} & & \\ \cdots & & \\ f'_{11} & & \end{matrix} \middle| \mathcal{C}_{\mu,k+1} \middle| \begin{matrix} f'_{1k} & \cdots & f'_{kk} \ 0 \\ \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} \ 0 \\ \bar{f}'_{1,k-1} & \cdots \bar{f}'_{k-1,k-1} & \\ & \cdots & \\ & \bar{f}'_{11} & \end{matrix} \right\rangle \quad (3.6)$$

with

$$f'_{ij} = \bar{f}'_{ij} + \delta_{il_j} \sum_{j'=\mu}^{k} \delta_{jj'}. \quad (3.7)$$

Actually we could have taken for the IR of the $\mathfrak{U}(k+1)$ group any partition $[f'_{i,k+1}]$ satisfying the branching rule

$$f'_{1,k+1} \geq f'_{1k} \geq f'_{2,k+1} \geq f'_{2k} \geq \cdots,$$

and a similar one for $\bar{f}'_{ik}$. In that case the reduced matrix element would depend also on $f'_{i,k+1}$ in such a way as to cancel the dependence on $f'_{i,k+1}$ in the

matrix element of $\mathcal{C}_{\mu,k+1}$. We prefer to take the simplest partition satisfying the branching rule, i.e., $f'_{i,k+1} = f'_{ik}$, $f'_{k+1,k+1} = 0$, as the reduced matrix element in (3.6) has been calculated by Brody, Moshinsky, and Renero.[9]

The matrix elements of $\mathcal{C}_{\mu,k+1}$ were given in a fundamental work by Gel'fand and Zetlin,[8] while the reduced matrix element of a a^k_μ was given in Ref. 9. Combining these results and simplifying some terms, as indicated explicitly in the Appendix, and making use of the relations (3.2b), we obtain finally for the one-block Wigner coefficient in the canonical chain the expression

$$\left\langle [f'_{ij}], \begin{bmatrix} 1 \\ \mu \end{bmatrix} \middle| [f_{ij}] \right\rangle = \prod_{\lambda=\mu+1}^{k} S(l_{\lambda-1} - l_\lambda)[(f'_{l_\lambda\lambda} - f'_{l_{\lambda-1}\lambda-1} + l_{\lambda-1} - l_\lambda)(f'_{l_\lambda\lambda} - f'_{l_{\lambda-1}\lambda-1} + l_{\lambda-1} - l_\lambda + 1)]^{-\frac{1}{2}}$$
$$\times \prod_{\lambda=\mu}^{k} \left[\frac{\prod_{j=1}^{\lambda} (f'_{l_\lambda\lambda} - f'_{j,\lambda-1} + j - l_\lambda + 1)}{\prod_{\substack{j=1 \\ j\neq l_\lambda}}^{\lambda} (f'_{l_\lambda\lambda} - f'_{j\lambda} + j - l_\lambda)} \right]^{\frac{1}{2}} \prod_{\lambda=\mu}^{k-1} \left[\frac{\prod_{j=1}^{\lambda+1} (f'_{l_\lambda\lambda} - f'_{j,\lambda+1} + j - l_\lambda)}{\prod_{\substack{j=1 \\ j\neq l_\lambda}}^{\lambda} (f'_{l_\lambda\lambda} - f'_{j\lambda} + j - l_\lambda + 1)} \right]^{\frac{1}{2}},$$

$$S(x) = \begin{cases} +1 & \text{if } x \geq 0; \\ -1 & \text{if } x < 0 \end{cases} \quad 1 \leq l_\lambda \leq \lambda. \quad (3.8)$$

An alternative approach to the evaluation of the one-block Wigner coefficients, also using the matrix elements of the generator $\mathcal{C}_{\mu,k+1}$, was discussed previously by Baird and Biedenharn.[10]

4. THE TRANSFORMATION BRACKETS FROM THE CANONICAL TO THE PHYSICAL CHAIN

Having obtained in the previous section the one-block Wigner coefficient in the canonical chain, the remaining problem for the determination of the fpc is the evaluation of the transformation bracket (2.18) between the states in the physical and the canonical chains. A purely computational procedure for evaluating these brackets is easily available. We note that the operator of angular momentum $\mathfrak{L}_\tau$, $\tau = 1, 0, -1$, is given in terms of

$$\mathcal{C}^{m'}_m = \sum_s (-1)^{m'} a^s_m \bar{a}^s_{-m'}$$

by[11]

$$\mathfrak{L}_\tau = [l(l+1)]^{\frac{1}{2}} \sum_{m,m'} \langle l1m'\tau \mid lm \rangle \mathcal{C}^{m'}_m, \quad (4.1)$$

so that the Casimir operator of the $R(3)$ group

$$\mathfrak{L}^2 = \sum_\tau (-1)^\tau \mathfrak{L}_\tau \mathfrak{L}_{-\tau} \quad (4.2)$$

can also be expressed in terms of generators of $\mathfrak{U}(2l+1)$. Using then the result of Gel'fand and Zetlin,[8] we could find the matrix of $\mathfrak{L}^2$ with respect to the states (2.17) and from the diagonalization of the same, the transformation brackets we are looking for. This program has actually been implemented numerically, i.e., a computer program is available[11,12] for $l = 1$.

While this approach to the transformation brackets is quite general (it can be easily extended to configuration space states in several orbitals, spin–isospin states, etc.), it is, in the opinion of the authors, profoundly unsatisfactory, as it requires the diagonalization of matrices [whose dimension incidentally increases with that of the IR of $\mathfrak{U}(2l+1)$] and so does not provide us with closed formulas. Clearly then, what is required is a procedure by which we could calculate explicitly the states in the physical chain, so as to obtain a closed formula for the transformation brackets when we evaluate the scalar product of these states, with those of (2.17) associated with the canonical chain.

For the sake of simplicity we shall illustrate this procedure for $l = 1$, where what is required are the polynomials in a^s_m, $m = 1, 0, -1$; $s = 1, 2, 3$ that are bases for IR in the $\mathfrak{U}(3) \supset R(3)$ chain.

These polynomials were obtained long ago by Bargmann and Moshinsky,[13] but in the present paper we shall follow an analysis similar to the one given recently for the $\mathfrak{U}(4) \supset \mathfrak{U}(2) + \mathfrak{U}(2)$ chain by Syamala Devi and Venkatarayudu.[14] This goes much deeper into the group-theoretical nature of the problem, is immediately generalizable to the $\mathfrak{U}(n) \supset R(n)$ chain, and, we hope, eventually to the chain of groups (2.6b).

A. The IR of $R(3)$ Contained in an IR of $S\mathfrak{U}(3)$

Our starting point is a theorem given in Littlewood's book[15] on the procedure for getting the IR of $R(n)$ contained in an IR of $S\mathfrak{U}(n)$. For $n = 3$ this result

[9] T. A. Brody, M. Moshinsky, and I. Renero, Rev. Mexicana de Física **15**, 145 (1966).

[10] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).

[11] *Many Body Problems and Other Selected Topics in Theoretical Physics*, M. Moshinsky, T. A. Brody, and G. Jacob, Eds. (Gordon and Breach Science Publishers, New York, 1967), pp. 291–377).

[12] M. Moshinsky, M. Berrondo, and J. Pineda, *Structure of Low-Medium Mass Nuclei*, P. Goldhammer, Ed. (University of Kansas Press, Lawrence, Kansas, 1966), pp. 129–194.

[13] V. Bargmann and M. Moshinsky, Nuclear Physics **23**, 177 (1961).

[14] V. Syamala Devi and T. Venkatarayudu, J. Math. Phys, **9**, 1057 (1968).

[15] D. E. Littlewood, *The Theory of Group Characters* (Clarendon Press, Oxford, 1950), 2nd Ed., p. 240.

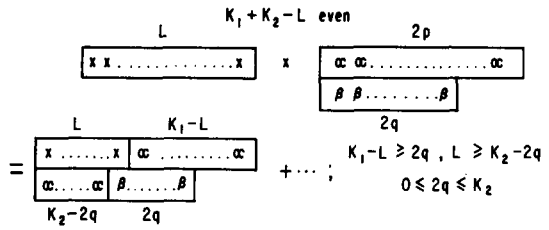


FIG. 1. Diagram of the IR $(\kappa_1\kappa_2)$ of $\mathfrak{U}(3)$ contained in the direct product $(L) \times (2p, 2q)$ and inequalities imposed by Littlewood rules on the components of the partitions.

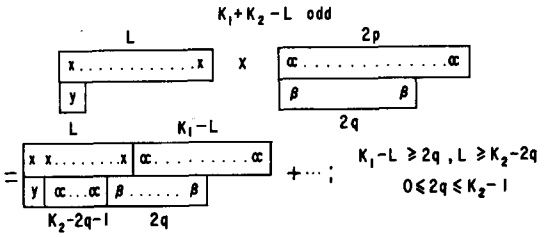


FIG. 2. Diagrams of the IR $(\kappa_1\kappa_2)$ of $\mathfrak{U}(3)$ contained in the direct product $(L1) \times (2p, 2q)$ and inequalities imposed by Littlewood rules on the components of the partition.

can be stated as follows: The IR of $\mathfrak{U}(3)$ is characterized by the partition $[f_1f_2f_3]$ and that of $S\mathfrak{U}(3)$ by the partition

$$(\kappa_1\kappa_2) \equiv (f_1 - f_3, f_2 - f_3). \tag{4.3}$$

On the other hand, the IR of $R(3)$ are characterized by the single number L giving the orbital angular momentum. The IR of the orthogonal group $O(3)$ [which contains the group $R(3)$ and the inversion] are characterized by the partitions (L) or $(L1)$. Both of these IR of $O(3)$ contain the same IR of $R(3)$ characterized by L, but in the first case the matrix in the IR corresponding to the inversion is the unit matrix I multiplied by $(-1)^L$, while in the second case, I is multiplied by $(-1)^{L+1}$.

Littlewood's theorem[15] states that if in the reduction to IR of $U(3)$, the product representation $(\lambda) \times (\delta)$ contains $(\kappa) \equiv (\kappa_1\kappa_2)$ a certain number of times, which we denote by $g_{\lambda\delta\kappa}$, then the IR $(\kappa_1\kappa_2)$ of $U(3)$ breaks into IR$[\lambda]$ of $O(3)$ according to[15]

$$(\kappa) = \sum_{\{\delta\}} g_{\lambda\delta\kappa}[\lambda], \tag{4.4}$$

the summation being over all partitions that characterize the IR of $U(3)$ of the type $(\delta) = (2p, 2q)$ with p, q integers and where (λ) is an IR of $U(3)$ of form (L) or $(L1)$. For example, the IR (32) of $U(3)$ contains the IR (3), (21), and (1) of $O(3)$, and so the possible values of L will be $L = 1, 2, 3$.

Littlewood's theorem (Ref. 15, p. 240) is actually available only for the IR (κ_1) of $U(3)$ [as is pointed out by Littlewood himself (Ref. 15, p. 294)], but it can be extended to the partitions $(\kappa_1\kappa_2)$ of $U(3)$ in the way elucidated in the previous paragraph, when the modification rules of Murnaghan[16] are taken into account. The coefficients $g_{\lambda\delta\kappa}$ are determined from the usual Littlewood rules (Ref. 15, p. 94) for the reduction of the direct product $(\lambda) \times (\delta)$. We therefore have to consider only two cases:

$$\kappa_1 + \kappa_2 = L + 2p + 2q, \tag{4.5a}$$

$$\kappa_1 + \kappa_2 = L + 1 + 2p + 2q. \tag{4.5b}$$

The reduction of $(\lambda) \times (\delta)$ is illustrated for the two cases in Figs. 1 and 2, respectively, in which we mark the blocks in the single row (L) by x and those in the rows $2p$ and $2q$ of (δ) by α and β, respectively. The single block in the second row of $(L1)$ is marked by y. The Littlewood rules restrict the distribution of the α's and β's in the way shown in Figs. 1 and 2, so that $2p$, $2q$ are restricted by the conditions also indicated there.

B. Decomposition of the Diagrams into Elementary Permissible Diagrams

Our next step will be to break the diagrams of $(\kappa_1\kappa_2)$ containing x, y, α, β of Figs. 1 and 2 into elementary permissible diagrams (epd). A permissible diagram would be any one- or two-rowed marked diagrams that could be formed by a direct product of the type $(\lambda) \times (\delta)$ for particular L's and $2p$, $2q$'s. An epd would be a permissible diagram that cannot be decomposed further into permissible diagrams.

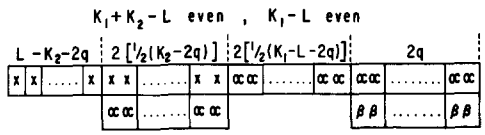


FIG. 3(a). Decomposition of the diagram of the IR $(\kappa_1\kappa_2)$ with $\kappa_1 + \kappa_2 - L$ even, $\kappa_1 - L$ even, into elementary permissible diagrams.

To find out the epd we need and to carry out the decomposition in a systematic way, we redraw the diagram of partitions $(\kappa_1\kappa_2)$ in Figs. 1, 2 by translating, as far as possible, the blocks of α's and β's in the second row. The expression "as far as possible" takes into account that no column should contain two α's or that a β block should not extend farther than the end of the α blocks above it. Under this restriction we carry out the translation in the following four cases: Fig. 3(a): $\kappa_1 + \kappa_2 - L$ even and $\kappa_1 - L$ (and therefore κ_2) even; Fig. 3(b): $\kappa_1 + \kappa_2 - L$ even, $\kappa_1 - L$ (and therefore κ_2) odd; Fig. 4(a): $\kappa_1 + \kappa_2 - L$ odd, $\kappa_1 - L$ even (κ_2 odd); Fig. 4(b): $\kappa_1 + \kappa_2 - L$ odd, $\kappa_1 - L$ odd (κ_2 even). We indicate the length of each segment of

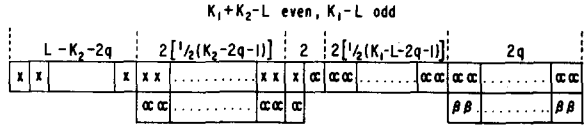


FIG. 3(b). Decomposition of the diagram of the IR $(\kappa_1\kappa_2)$ with $\kappa_1 + \kappa_2 - L$ even, $\kappa_1 - L$ odd.

[16] F. D. Murnaghan, Proc. Natl. Acad. Sci. 24, 184 (1938).

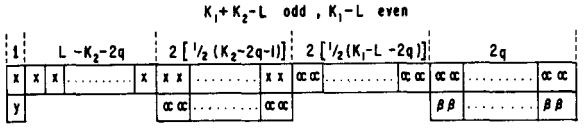


FIG. 4(a). Decomposition of the diagram of the IR $(\kappa_1\kappa_2)$ with $\kappa_1 + \kappa_2 - L$ odd, $\kappa_1 - L$ even.

the diagrams in a notation convenient for our later analysis.

Now we start by determining the epd in Fig. 3(a), reading from right to left. Clearly the first permissible diagram we see is

$$\begin{array}{|c|c|}\hline \alpha & \alpha \\ \hline \beta & \beta \\ \hline \end{array} \quad \text{as} \quad (00) \times (22) = (22), \qquad (4.6a)$$

which is an epd, as its decomposition into the possible subdiagrams $\boxed{\begin{array}{c}\alpha\\ \beta\end{array}}$ is forbidden by the rules (4.4). There are q epd of this type as indicated by the number in the square bracket above the corresponding part of Fig. 3(a). Continuing the reading from left to right, we find the permissible diagram

$$\boxed{\alpha \mid \alpha} \quad \text{as} \quad (00) \times (20) = (20), \qquad (4.6b)$$

which is an epd, as the subdiagram $\boxed{\alpha}$ is again forbidden by (4.4). There are $\frac{1}{2}(\kappa_1 - L - 2q)$ epd of this type as indicated in Fig. 3(a). The next permissible diagram is

$$\begin{array}{|c|c|}\hline x & x \\ \hline \alpha & \alpha \\ \hline \end{array} \quad \text{as} \quad (20) \times (20) = (22) + \cdots, \qquad (4.6c)$$

which is an epd, as its decomposition into allowed subdiagrams $\boxed{\begin{array}{c}x\\ \alpha\end{array}}$ is forbidden by (4.4). There are $\frac{1}{2}(\kappa_2 - 2q)$ epd of this type as indicated in Fig. 3(a). Finally, we get the epd

$$\boxed{x} \quad \text{as} \quad (10) \times (00) = (10), \qquad (4.6d)$$

of which there are $L - \kappa_2 - 2q$ also indicated in Fig. 3(a).

A similar analysis can be carried out in Figs. 3(b), 4(a), and 4(b), noting only that in, Figs. 3(b) and 4(b), there appears, only once, the permissible diagram

$$\begin{array}{|c|c|}\hline x & \alpha \\ \hline \alpha & \\ \cline{1-1} \end{array} \quad \text{as} \quad (10) \times (20) = (21) + \cdots, \qquad (4.6e)$$

which is an epd, as the possible subdiagrams $\boxed{\begin{array}{c}x\\ \alpha\end{array}}$, $\boxed{\alpha}$ are not permitted by the rules (4.4). Also in Figs. 4(a) and 4(b), there appears, again only once, the epd

$$\boxed{\begin{array}{c}x\\ y\end{array}} \quad \text{as} \quad (11) \times (00) = (11). \qquad (4.6f)$$

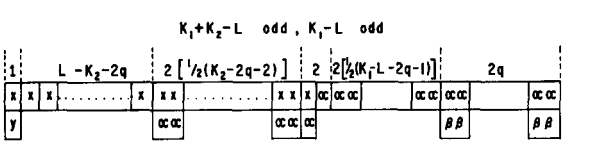


FIG. 4(b). Decomposition of the diagram of the IR $(\kappa_1\kappa_2)$ with $\kappa_1 + \kappa_2 - L$ odd, $\kappa_1 - L$ odd.

Having achieved the decomposition into epd of the Young diagrams formed from following the analysis of Littlewood for the determination of the IR of $R(3)$ contained in a given IR of $S\mathfrak{U}(3)$, we shall proceed to associate definite polynomials with these epd, and later with the full states in the chain $S\mathfrak{U}(3) \supset R(3)$.

C. Polynomials Associated with the Diagrams

The bases for the IR $(\kappa_1\kappa_2)$ of $S\mathfrak{U}(3)$ are identical[13] to the bases for the IR $[\kappa_1\kappa_2 0]$ of $\mathfrak{U}(3)$, and the latter would only depend[13] on two vectors a_m^s, $s = 1, 2$. From these creation operators and the corresponding annihilation operators $\bar{a}_{m'}^{s'}$, we can construct the generators of the following unitary groups:

$$U(2);\ C^{ss'} = \sum_{mm'} g^{mm'} a_m^s \bar{a}_{m'}^{s'} = \sum_m (-1)^m a_m^s \bar{a}_{-m}^{s'}, \quad s, s' = 1, 2, \qquad (4.7)$$

$$\mathfrak{U}(3);\ \mathcal{C}_m^{m'} = \sum_{m''} g^{m'm''} \sum_s a_m^s \bar{a}_{m''}^s = (-1)^{m'} \sum_s a_m^s \bar{a}_{-m'}^s, \quad m, m' = 1, 0, -1. \qquad (4.8)$$

From (4.1) the generators of the subgroup $R(3)$ of $\mathfrak{U}(3)$ are given by $\mathfrak{L}_m$ defined by[11]

$$\mathfrak{L}_1 = -(\mathcal{C}_1^0 + \mathcal{C}_0^{-1}), \quad \mathfrak{L}_0 = (\mathcal{C}_1^1 - \mathcal{C}_{-1}^{-1}), \quad \mathfrak{L}_{-1} = (\mathcal{C}_0^1 + \mathcal{C}_{-1}^0). \qquad (4.9)$$

As shown by Bargmann and Moshinsky,[13] the polynomials $P(a_m^s)$ corresponding to the IR $(\kappa_1\kappa_2)$ of $S\mathfrak{U}(3)$ and to IR L of $R(3)$ but of highest weight in the latter group, satisfy the equations

$$C^{11}P = \kappa_1 P, \qquad (4.10a)$$

$$C^{12}P = 0, \qquad (4.10b)$$

$$C^{22}P = \kappa_2 P, \qquad (4.10c)$$

$$\mathfrak{L}_1 P = 0, \qquad (4.11a)$$

$$\mathfrak{L}_0 P = LP, \qquad (4.11b)$$

where the $C^{ss'}$, $\mathfrak{L}_m$ can be interpreted as first-order differential operators, as from the commutation relations (2.3), the $\bar{a}_m^s$, when acting on polynomials $P(a_m^s)$, are equivalent to the operators

$$\sum_{m'} g^{mm'} \bar{a}_{m'}^s = (-1)^m \bar{a}_{-m}^s = \partial/\partial a_m^s. \qquad (4.12)$$

Because of this character of the operators $C^{ss'}$, $\mathfrak{L}_m$, it is clear that if we have a set of polynomial solutions of the Eqs. (4.10) and (4.11), a product of powers of this set will again be a solution, but with different κ_1, κ_2, L.

TABLE I. Elementary permissible diagrams (epd) and the IR $(\kappa_1\kappa_2)$ of $S\mathfrak{U}(3)$ and L of $R(3)$ that characterize them. The polynomials corresponding to these epd are given in terms of the variables $\Delta^s_m \equiv a^s_m$, $\Delta^{ss'}_{mm'} = a^s_m a^{s'}_{m'} - a^s_{m'} a^{s'}_m$.

epd	IR $S\mathfrak{U}(3)$ $(\kappa_1\kappa_2)$	IR $R(3)$ L	Polynomials
	(10)	1	Δ^1_1
	(11)	1	Δ^{12}_{10}
	(21)	1	$w_+ \equiv \sum_m (-1)^m \Delta^{12}_{1m} \Delta^1_{-m}$
	(20)	0	$s \equiv \sum_m (-1)^m \Delta^1_m \Delta^1_{-m}$
	(22)	0	$t \equiv \sum_{mm'} (-1)^{m+m'} \Delta^{12}_{mm'} \Delta^{12}_{-m-m'}$
	(22)	2	$(\Delta^{12}_{10})^2$

The analysis of the previous paragraph suggests that we first consider the epd (4.6) and see what are the polynomials associated with them. In Table I we give each epd with its corresponding IR of $S\mathfrak{U}(3)$ and $R(3)$, and then the polynomial solution of Eqs. (4.10), (4.11) for those values of $(\kappa_1\kappa_2)$, L. As the diagrams of Figs. 3 and 4 can be broken into epd repeated a certain number of times, the polynomial solution of (4.10), (4.11) corresponding to a definite $(\kappa_1\kappa_2)$, L, can be constructed in terms of powers of the epd polynomials of Table I. These polynomials would be further characterized by the index q associated with the specific Littlewood procedure[15] for determining each possible way in whicn the IR L of $R(3)$ appears in the IR $(\kappa_1\kappa_2)$ of $S\mathfrak{U}(3)$. In fact, from Figs. 1 and 2 and Table I we find that there are only two independent cases for the polynomials:

$$\kappa_1 - L \text{ even}: \quad (\Delta^1_1)^{L-\kappa_2+2q} (\Delta^{12}_{10})^{\kappa_2-2q} s^{\frac{1}{2}(\kappa_1-L-2q)} t^q, \tag{4.13a}$$

$$\kappa_1 - L \text{ odd}: \quad w_+ (\Delta^1_1)^{L-\kappa_2+2q} (\Delta^{12}_{10})^{\kappa_2-2q-1} s^{\frac{1}{2}(\kappa_1-L-2q-1)} t^q. \tag{4.13b}$$

The procedure we followed associates a polynomial with each diagram that we construct from Littlewood's analysis,[15] so we will have found all the polynomials of our problem, and therefore all the states, if we can prove that they are linearly independent.

D. Linear Independence of the Polynomials

The states associated with the polynomials (4.13), will be orthogonal if they differ in any of the eigenvalues κ_1, κ_2, L, since the operators C^{11}, C^{22}, $\mathfrak{L}_0$ are Hermitian. We need therefore only worry about the linear independence of the polynomials (4.13) belonging to different q. We note, from the definition of Δ^1_1, s, Δ^{12}_{10}, w_+ in Table I, that they depend only on the five variables

$$\Delta^1_1, \quad \Delta^1_0, \quad \Delta^1_{-1}, \quad \Delta^{12}_{10}, \quad \Delta^{12}_{1-1}, \tag{4.14a}$$

where, as indicated in Table I,

$$\Delta^s_m \equiv a^s_m, \quad \Delta^{ss'}_{mm'} \equiv a^s_m a^{s'}_{m'} - a^s_{m'} a^{s'}_m. \tag{4.14b}$$

Furthermore, from the relation

$$\Delta^1_1 \Delta^{12}_{0-1} = -\Delta^1_{-1} \Delta^{12}_{10} + \Delta^1_0 \Delta^{12}_{1-1}, \tag{4.15}$$

we see that t of Table I is also a function of the five variables of (4.14), i.e.,

$$t = (\Delta^1_1)^{-1} [4\Delta^1_0 \Delta^{12}_{10} \Delta^{12}_{1-1} - 4(\Delta^{12}_{10})^2 \Delta^1_{-1} - 2\Delta^1_1 (\Delta^{12}_{1-1})^2]. \tag{4.16}$$

The five variables (4.14) are functionally independent because for example if we add to them the variable Δ^2_1, the Jacobian of this set, with respect to the set Δ^s_m, $m = 1, 0, -1$, $s = 1, 2$, is $(\Delta^1_1)^2 \neq 0$.

To see that polynomials of different q are linearly independent, let us write (e.g., for $\kappa_1 - L$ even) the polynomial in terms of the five independent variables (4.14):

$$\begin{aligned} P^{\kappa_1\kappa_2 L}_q = {} & (\Delta^1_1)^{L-\kappa_2+q} (\Delta^{12}_{10})^{\kappa_2-2q} \\ & \times [-2\Delta^1_1 \Delta^1_{-1} + (\Delta^1_0)^2]^{\frac{1}{2}(\kappa_1-L)-q} \\ & \times [4\Delta^1_0 \Delta^{12}_{10} \Delta^{12}_{1-1} - 4(\Delta^{12}_{10})^2 \Delta^1_{-1} - 2\Delta^1_1 (\Delta^{12}_{1-1})^2]^q. \end{aligned} \tag{4.17}$$

Now from (4.17) it is clear that the highest power of Δ^1_1 appearing in P_q is

$$\begin{aligned} & [L - \kappa_2 + q] + [\tfrac{1}{2}(\kappa_1 - L) - q] + q \\ & \qquad = [(L - \kappa_2) + \tfrac{1}{2}(\kappa_1 - L)] + q, \end{aligned} \tag{4.18}$$

so it depends on q, and therefore the set of polynomials corresponding to different values of q cannot be linearly dependent. A similar reasoning applies to $\kappa_1 - L$ odd.

The polynomials (4.13) are, of course, the same as those obtained by Bargmann and Moshinsky[13] who directly solved the set of partial differential equations (4.10), (4.11). The difference in the procedure of obtaining them rests on the fact that the polynomials in the present analysis can be obtained by inspection from the diagrams associated with Littlewood's analysis[15] in terms of the polynomials of the epd, which are very easy to determine. Furthermore, the exponent q acquires a definite meaning associated with the Littlewood analysis[15]; finally, the procedure can be extended to the chain $\mathfrak{U}(n) \supset R(n)$, as Littlewood's analysis was, in fact, carried out for arbitrary n. This last point is very important, as the analytical procedure of Bargmann and Moshinsky[13] is based on divisibility arguments very difficult to extend to cases where we have more than one free exponent of the type q, as happens, for example, already in the case of

$\mathfrak{U}(4) \supset R(4)$, of great importance for the fpc of spin–isospin states. The present technique will be applied to the $\mathfrak{U}(4) \supset R(4)$ chain in future publications.

E. Determination of the Transformation Brackets from the $\mathfrak{U}(3) \supset R(3)$ to the Canonical Chain

In the case $l = 1$, the transformation brackets we need to determine are

$$\left\langle \begin{matrix} f_{13}f_{23}f_{33} \\ f_{12}f_{22} \\ f_{11} \end{matrix} \middle| \begin{matrix} f_{13}f_{23}f_{33} \\ qLM \end{matrix} \right\rangle = \left\langle \begin{matrix} \kappa_1\kappa_2 0 \\ f_1'f_2' \\ f_1'' \end{matrix} \middle| \begin{matrix} \kappa_1\kappa_2 0 \\ qLM \end{matrix} \right\rangle \tag{4.19a}$$

where

$$\kappa_s = f_{s3} - f_{33}, \quad f_s' = f_{s2} - f_{33}, \quad s = 1, 2,$$
$$f_1'' = f_{11} - f_{33}. \tag{4.19b}$$

Equation (4.19) stems from the fact[13,17] that both in the canonical chain and the $\mathfrak{U}(3) \supset R(3)$ chain, there is a term $(\Delta^{123}_{10-1})^{f_{33}}$, where

$$\Delta^{123}_{10-1} = \begin{vmatrix} a_1^1 & a_0^1 & a_{-1}^1 \\ a_1^2 & a_0^2 & a_{-1}^2 \\ a_1^3 & a_0^3 & a_{-1}^3 \end{vmatrix}, \tag{4.20}$$

with a corresponding normalization factor that disappears in the scalar product,[17] so we need only concern ourselves with the $S\mathfrak{U}(3)$ transformation bracket in the right-hand side of (4.19a). Furthermore, as

$$\left| \begin{matrix} \kappa_1\kappa_2 0 \\ qLM \end{matrix} \right\rangle = \left[\frac{(L+M)!\, 2^{L-M}}{(L-M)!\,(2L)!} \right]^{\frac{1}{2}} (\mathfrak{L}_{-1})^{L-M} \left| \begin{matrix} \kappa_1\kappa_2 0 \\ qLL \end{matrix} \right\rangle, \tag{4.21}$$

and $\mathfrak{L}_{-1}$ is given by (4.9), the transformation bracket (4.19a) for arbitrary M could be obtained from the one with $M = L$ and the well-known matrix elements[8,11] of $\mathfrak{C}^{m'}_{m}$ with respect to the canonical chain. For the case $M = L$, the ket (4.21) is given by the polynomials (4.13) acting on $|0\rangle$. As the polynomials for the canonical chain are also available,[17] the scalar product (4.19a) for $M = L$ could, and in fact was, evaluated in Ref. 17. We note from the way the indices $m = 1$, 0, -1 were enumerated in Ref. 17 that, in (4.19a),

$$M = 2f_1'' - (f_1' + f_2') \tag{4.22}$$

and so the transformation bracket (4.19a) for $M = L$ depends only on κ_1, κ_2, q, L, f_1', f_2' and is given by

$$\left\langle \begin{matrix} \kappa_1\kappa_2 0 \\ f_1'f_2' \\ f_1'' \end{matrix} \middle| \begin{matrix} \kappa_1\kappa_2 0 \\ qLL \end{matrix} \right\rangle = \left[\frac{[\tfrac{1}{2}(f_1' - f_2' + L)!\, f_2'!\, (f_1' - f_2' + 1)!\, (f_1' + 1)!\, (\kappa_1 - f_2' + 1)!\, (f_1' - \kappa_2)!\, (\kappa_1 - f_1')!]^{\frac{1}{2}}}{[\tfrac{1}{2}(f_1' - f_2' - L)]!\, (\kappa_1 - \kappa_2 + 1)!\, (\kappa_2 - f_2')!} \right]$$
$$\times (-2)^{\frac{1}{2}(f_1' - f_2' - L)} \sum_{\rho\beta} E^q_\rho \frac{[\tfrac{1}{2}(\kappa_1 - L - \rho)]!\, \rho!\, (\kappa_2 - \beta)!}{2^\beta [\tfrac{1}{2}(\kappa_1 - f_1' - f_2' - \rho) + \beta]!\, (f_2' - \beta)!\, (\rho - \beta)!\, \beta!\, (f_1' + 1 - \beta)!}, \tag{4.23a}$$

with ρ even (odd) if $\kappa_1 - L$ is even (odd) and

$$E^q_\rho = \binom{q}{\frac{1}{2}\rho} (-1)^{\frac{1}{2}\rho} \quad \text{if} \quad \rho \text{ even}, \quad E^q_\rho = \binom{q}{\frac{1}{2}(\rho - 1)} (-1)^{\frac{1}{2}(\rho - 1)} \quad \text{if} \quad \rho \text{ odd}, \tag{4.23b}$$

where the formula (4.14) in Ref. 17 was simplified using the following identities twice:

$$\sum_{\alpha=0}^{c} \frac{(a-\alpha)!\, (b+\alpha)!}{\alpha!\, (c-\alpha)!} = \frac{b!\, (a-c)!\, (a+b+1)!}{c!\, (a+b-c+1)!}. \tag{4.24}$$

We note that in Ref. 17 the enumeration of the indices $m = 1$, 0, -1 is $1 \to 1$, $-1 \to 2$, $0 \to 3$, while in the enumeration of (2.20b) it is $1 \to 1$, $0 \to 2$, $-1 \to 3$. To apply the transformation brackets (4.23) to the Wigner coefficient (3.8), we must then either apply first the transposition (2, 3) to the bra in (4.23), which gives the explicit result of Chacón and Moshinsky,[18] or what is much simpler, apply the transposition (2, 3) to the states $\left| \begin{bmatrix} 1 \\ \mu \end{bmatrix} \right\rangle$ which implies only the interchange

$$\left| \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\rangle \leftrightarrow \left| \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\rangle. \tag{4.25}$$

Combining (4.23) and (3.8) for $k = 3$ and taking into account the observation of the previous paragraph, we can get the fpc in the p shell

$$\langle (l=1)^{n-1} \bar{f}\bar{q}\bar{L},\ (l=1)\,|\}\,(l=1)^n f q L \rangle \tag{4.26}$$

for arbitrary n in a closed form. A very interesting point though is that these fpc are in a nonorthonormal basis, i.e., they are characterized by the numbers $\bar{q}$, q that are not eigenvalues of any operator, but appear through the Littlewood analysis[15] for finding the IR of $R(3)$ contained in an IR of $S\mathfrak{U}(3)$.

It is possible to transform these fpc to an orthonormal basis through the eigenvalues of the Hermitian

[17] M. Moshinsky, Rev. Mod. Phys. **34**, 813 (1962).
[18] E. Chacón and M. Moshinsky, Phys. Letters **23**, 567 (1966)

operator

$$\mathbf{\Omega} = \tfrac{1}{2}(\mathcal{C}^{mm'} + \mathcal{C}^{m'm})\mathcal{L}_m\mathcal{L}_{m'}, \quad \mathcal{C}^{mm'} = \sum_{m''} g^{mm''}\mathcal{C}^{m'}_{m''}, \tag{4.27}$$

introduced by Bargmann and Moshinsky,[13] but this would imply again a diagonalization of matrices, and so there would be no closed form. Clearly this brings in the view, first presented by Racah,[19] that a non-orthonormal basis may provide a much deeper and simpler description of the states characterized by noncanonical chains of groups than does orthonormal basis.

ACKNOWLEDGMENTS

The authors are indebted to E. Chacón for many discussions and for the checking and simplifying of several formulas. They would also like to express their appreciation to Prof. T. Venkatarayudu for the discussions that led to the concepts developed in Sec. 4.

APPENDIX: THE EVALUATION OF THE MATRIX ELEMENT OF a^k_μ

The set of vectors $a^s_m, \mu = 1, \cdots, k; s = 1, \cdots, k$ could be considered as components of a single k^2-dimensional vector. On the other hand, the states (2.17), as indicated in Ref. 6, are actually characterized by two Gel'fand patterns, one for index μ, and another, of highest weight, for the index s. The matrix element (3.4) should be rewritten to take this into account. Using the decomposition (3.6), both with respect to the μ and $s = k$ indices, we get

$$\left\langle \begin{matrix} f'_{1k} & \cdots & f'_{kk} \\ f'_{1,k-1} & \cdots & f'_{k-1,k-1} \\ \cdots & & \\ f'_{11} & & \end{matrix} ; \begin{matrix} f'_{1k} & \cdots & f'_{kk} \\ f'_{1k} & \cdots & f'_{k-1,k} \\ \cdots & & \\ f'_{1k} & & \end{matrix} \middle| a^k_\mu \middle| \begin{matrix} \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & 0 \\ \bar{f}'_{1,k-1} & \cdots & \bar{f}'_{k-1,k-1} & \\ \cdots & & & \\ \bar{f}'_{11} & & & \end{matrix} ; \begin{matrix} \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & 0 \\ \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & \\ \cdots & & & \\ \bar{f}'_{1k} & & & \end{matrix} \right\rangle$$

$$= \langle f'_{1k} \cdots f'_{kk} \| \, a \, \| \bar{f}'_{1k} \cdots \bar{f}'_{k-1,k} \; 0 \rangle \left\langle \begin{matrix} f'_{1k} & \cdots & f'_{kk} & 0 \\ f'_{1k} & \cdots & f'_{kk} & \\ f'_{1,k-1} & \cdots & f'_{k-1,k-1} & \\ \cdots & & & \\ f'_{11} & & & \end{matrix} \middle| \mathcal{C}_{\mu,k+1} \middle| \begin{matrix} f'_{1k} & \cdots & f'_{kk} & 0 \\ \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & 0 \\ \bar{f}'_{1,k-1} & \cdots & \bar{f}'_{k-1,k-1} & \\ \cdots & & & \\ \bar{f}'_{11} & & & \end{matrix} \right\rangle$$

$$\times \left\langle \begin{matrix} f'_{1k} & \cdots & f'_{kk} & 0 \\ f'_{1k} & \cdots & f'_{kk} & \\ f'_{1k} & \cdots & f'_{k-1,k} & \\ \cdots & & & \\ f'_{1k} & & & \end{matrix} \middle| C^{k,k+1} \middle| \begin{matrix} f'_{1k} & \cdots & f'_{kk} & 0 \\ \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & 0 \\ \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & \\ \cdots & & & \\ f'_{1k} & & & \end{matrix} \right\rangle \tag{A1}$$

To determine the last matrix element we note that

$$C^{k,k+1}C^{k+1,k} - C^{k+1,k}C^{k,k+1} = C^{kk} - C^{k+1,k+1}, \tag{A2}$$

so that evaluating the left-hand side between the states

$$\left\langle \begin{matrix} f'_{1k} & \cdots & f'_{kk} & 0 \\ \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & 0 \\ \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & \\ \cdots & & & \\ \bar{f}'_{1k} & & & \end{matrix} \middle| C^{k,k+1}C^{k+1,k} - C^{k+1,k}C^{k,k+1} \middle| \begin{matrix} f'_{1k} & \cdots & f'_{kk} & 0 \\ \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & 0 \\ \bar{f}'_{1k} & \cdots & \bar{f}'_{k-1,k} & \\ \cdots & & & \\ f'_{1k} & & & \end{matrix} \right\rangle = -1 \tag{A3}$$

where the value -1 on the right-hand side has to do with the eigenvalues[20] of C^{ij} and the relation (3.7). Furthermore, as $C^{k+1,k}$ is the lowering[20] operator L^k_{k+1}, which decreases the last term of the second row of the ket by 1, and as this term is zero already, the matrix element of the first product in (A3) is zero. For the second product, as $C^{k,k+1}$ is a raising generator, it can only take us from the ket in (A3) to the highest-weight state of $\mathfrak{U}(k+1)$, associated with the partition $[f'_{1k} \cdots f'_{kk} 0]$. We obtain therefore from

[19] G. Racah in *Istambul Lectures*, F. Gursey, Ed. (Gordon and Breach, Science Publishers, New York, 1962).

[20] J. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965); Rev. Mexicana Fis. **14**, 29 (1965).

(A3) that the square of the last matrix element in (A1) is equal to 1; because of the phase convention used,[20] it is actually 1.

The matrix element of $C_{\mu,k+1}$ was obtained by Gel'fand and Zetlin[8]; in our notation,[20] using the relation (3.7), it is given by

$$\left\langle \begin{vmatrix} f'_{1k} & \cdots & f'_{kk} & 0 \\ f'_{1k} & \cdots & f'_{kk} & \\ f'_{1,k-1} & \cdots & f'_{k-1,k-1} & \\ & \cdots & & \\ & f'_{11} & & \end{vmatrix} C_{\mu,k+1} \begin{vmatrix} f'_{1k} & \cdots & f'_{kk} & 0 \\ f'_{1k} & \cdots & f'_{k-1,k} & 0 \\ f'_{1,k-1} & \cdots & f'_{k-1,k-1} & \\ & \cdots & & \\ & f'_{11} & & \end{vmatrix} \right\rangle$$

$$= \prod_{\rho=\mu+1}^{k} S(l_{\rho-1}-l_\rho)[(f'_{l_\rho\rho}-f'_{l_{\rho-1}\rho-1}+l_{\rho-1}-l_\rho)(f'_{l_\rho\rho}-f'_{l_{\rho-1}\rho-1}+l_{\rho-1}-l_\rho+1)]^{-\frac{1}{2}}$$

$$\times \prod_{\lambda=\mu}^{k-1} \left[\frac{\prod_{\kappa=1}^{\lambda-1}(f'_{l_\lambda\lambda}-f'_{\kappa,\lambda-1}+\kappa-l_\lambda+1)\prod_{\kappa=1}^{\lambda+1}(f'_{l_\lambda\lambda}-f'_{\kappa,\lambda+1}+\kappa-l_\lambda)}{\prod_{\substack{\kappa=1\\ \kappa\neq l_\lambda}}^{\lambda}(f'_{l_\lambda\lambda}-f'_{\kappa\lambda}+\kappa-l_\lambda+1)\prod_{\kappa=1}^{\lambda}(f'_{l_\lambda\lambda}-f'_{\kappa\lambda}+\kappa-l_\lambda)} \right]^{\frac{1}{2}}$$

$$\times \left[-\frac{\prod_{\kappa=1}^{k-1}(f'_{l_kk}-f'_{\kappa,k-1}+\kappa-l_k+1)\prod_{\kappa=1}^{k}(f'_{l_kk}-f'_{\kappa k}+\kappa-l_k)}{\prod_{\substack{\kappa=1\\ \kappa\neq l_k}}^{k}(f'_{l_kk}-f'_{\kappa k}+\kappa-l_k+1)\prod_{\substack{\kappa=1\\ \kappa\neq l_k}}^{k}(f'_{l_kk}-f'_{\kappa k}+\kappa-l_k)}(f'_{l_kk}+k+1-l_k) \right]^{\frac{1}{2}},$$

$$S(x)=\begin{cases}+1 & \text{if } x\geq 0,\\ -1 & \text{if } x<0.\end{cases} \tag{A4}$$

The first matrix element in (A1) was obtained in a paper of Brody, Moshinsky, and Renero[9] and is given by

$$\langle f'_{1k}\cdots f'_{kk}\| a \|f'_{1k}\cdots f'_{k-1,k}\,0\rangle = \left[\frac{\prod_{1\leq i<j\leq k}(f'_{ik}-f'_{jk}+j-i+\delta_{il_k}-\delta_{jl_k})}{(f'_{l_kk}+k-l_k+1)\prod_{1\leq i<j\leq k}(f'_{ik}-f'_{jk}+j-i)} \right]^{\frac{1}{2}}. \tag{A5}$$

Combining (A4) and (A5), and taking into account the following product identities:

$$\begin{aligned} \prod_{1\leq i<j\leq k}(f'_{ik}-f'_{jk}+j-i+\delta_{il_k}-\delta_{jl_k}) &= \prod_{\kappa=1}^{k}(f'_{l_kk}-f'_{\kappa k}+\kappa-l_k+1),\\ \prod_{1\leq i<j\leq k}(f'_{ik}-f'_{jk}+j-i) &= \prod_{\substack{\kappa=1\\ \kappa\neq l_k}}^{k}(f'_{l_kk}-f'_{\kappa k}+\kappa-l_k),\\ \prod_{\kappa=1}^{k}(f'_{l_kk}-f'_{\kappa k}+\kappa-l_k) &= -\prod_{\substack{\kappa=1\\ \kappa\neq l_k}}^{k}(f'_{l_kk}-f'_{\kappa k}+\kappa-l_k), \end{aligned} \tag{A6}$$

we obtain finally that the matrix element of (A1) has the value

$$\prod_{\rho=\mu+1}^{k} S(l_{\rho-1}-l_\rho)[(f'_{l_\rho\rho}-f'_{l_{\rho-1}\rho-1}+l_{\rho-1}-l_\rho)(f'_{l_\rho\rho}-f'_{l_{\rho-1}\rho-1}+l_{\rho-1}-l_\rho+1)]^{-\frac{1}{2}}$$

$$\times \prod_{\lambda=\mu}^{k-1} \left[\frac{\prod_{\kappa=1}^{\lambda-1}(f'_{l_\lambda\lambda}-f'_{\kappa,\lambda-1}+\kappa-l_\lambda+1)\prod_{\kappa=1}^{\lambda+1}(f'_{l_\lambda\lambda}-f'_{\kappa,\lambda+1}+\kappa-l_\lambda)}{\prod_{\substack{\kappa=1\\ \kappa\neq l_\lambda}}^{\lambda}(f'_{l_\lambda\lambda}-f'_{\kappa\lambda}+\kappa-l_\lambda+1)\prod_{\substack{\kappa=1\\ \kappa\neq l_\lambda}}^{\lambda}(f'_{l_\lambda\lambda}-f'_{\kappa\lambda}+\kappa-l_\lambda)} \right]^{\frac{1}{2}} \left[\frac{\prod_{\kappa=1}^{k-1}(f'_{l_kk}-f'_{\kappa,k-1}+\kappa-l_k+1)}{\prod_{\substack{\kappa=1\\ \kappa\neq l_k}}^{k}(f'_{l_kk}-f'_{\kappa k}+\kappa-l_k)} \right]^{\frac{1}{2}}. \tag{A7}$$

Making use of the relations (3.2b), we can then obtain the one block Wigner coefficients given in (3.8).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Hermite's Reciprocity Law and the Angular-Momentum States of Equivalent Particle Configurations*

B. G. WYBOURNE
Physics Department, University of Canterbury, Christchurch, New Zealand

(Received 30 August 1968)

Hermite's reciprocity law is applied to the calculation of the angular-momentum states of equivalent particle configurations. Connections between boson and fermion states are considered and a method is given for determining the number of times a given term appears in l^3 without requiring a complete term analysis. A succinct expression for the number of S-states arising in the $[2^2]$ representation of R_{2l+1} under $R_{2l+1} \rightarrow R_3$ is developed.

I. INTRODUCTION

The group-theoretical classification of the angular-momentum states of equivalent electron or nucleon configurations is well known.[1-3] In LS coupling, the orbital angular-momentum states of the n-particle configuration l^n are classified using the chain of groups

$$U_{2l+1} \rightarrow R_{2l+1} \rightarrow R_3, \tag{1}$$

while in the jj-coupled configuration j^n, the chain of groups

$$U_{2j+1} \rightarrow Sp_{2j+1} \rightarrow R_3 \tag{2}$$

is appropriate.

The irreducible representations of the unitary group U_N in N dimensions are labeled by partitions $\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$ of the integer n into not more than N parts where

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0 \tag{3}$$

and

$$\lambda_1 + \lambda_2 + \cdots + \lambda_N = n. \tag{4}$$

The character of the irreducible representation $\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$ is just the S function $\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$ formed on the characteristic roots of the $N \times N$ unitary matrices that constitute elements of the group U_N. Littlewood[4] has exploited the properties of S-functions to express the characters of U_N as a linear combination of the characters of R_N or Sp_N and, conversely, he has developed prescriptions for expressing the characters of R_N or Sp_N in terms of S-functions defined on N variables. If the decomposition of the character of the irreducible representation $\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$ of U_N into the characters of the irreducible representations of R_3 is established, then the corresponding decompositions under the restrictions $R_N \rightarrow R_3$ or $Sp_N \rightarrow R_3$ may be found by the method of differences.[5] Thus, the principal problem in using the chains of groups, given in Eqs. (1) and (2), to classify angular-momentum states is the determination of the character decompositions under the restriction $U_N \rightarrow R_3$.

In the present paper we briefly discuss the calculation of the character decompositions under the restriction $U_N \rightarrow R_3$, emphasizing in particular the applications of Hermite's law of reciprocity[6,7] which lead to a number of illuminating correspondences between the angular-momentum states of many-boson systems with those of many-fermion systems. Finally, a simple method is given for determining the number of times a given representation of R_3 occurs in the decomposition of the irreducible representations $\{\lambda_1, \lambda_2, \lambda_3\}$ of U_N where $\lambda_1 + \lambda_2 + \lambda_3 = 3$. This leads to a simple prescription for determining the angular-momentum states in any l^3 or j^3 configuration and the distribution of S-states in l^4 or j^4.

II. BRANCHING RULES UNDER $U_n \rightarrow R_3$

Littlewood's algebra of plethysm[8,9] may be readily used to obtain the branching rules for $U_{2l+1} \rightarrow R_3$ or $U_{2j+1} \rightarrow R_3$ by evaluating the terms contained in the plethysms

$$[l] \otimes \{\lambda_1, \lambda_2, \cdots, \lambda_N\} \quad \text{and} \quad [j] \otimes \{\lambda_1, \lambda_2, \cdots, \lambda_N\}, \tag{5}$$

where, under $U_{2l+1} \rightarrow R_3$, $\{1\} \rightarrow [l]$, and under $U_{2j+1} \rightarrow R_3$, $\{1\} \rightarrow [j]$. Littlewood[10] has used the

* Research sponsored in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant No. 1275-67.

[1] H. A. Jahn, Proc. Roy. Soc. (London) **A201**, 516 (1950).
[2] B. H. Flowers, Proc. Roy. Soc. (London) **A212**, 248 (1952).
[3] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill Book Co., New York, 1962).
[4] D. E. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups* (Oxford University Press, London, 1950), 2nd ed.
[5] F. D. Murnaghan, *The Theory of Group Representations* (Johns Hopkins Press, Baltimore, 1938).
[6] F. D. Murnaghan, Proc. Natl. Acad. Sci. **37**, 439 (1951).
[7] F. D. Murnaghan, *The Unitary and Rotation Groups* (Spartan Books, Washington, D.C., 1962).
[8] D. E. Littlewood, Phil. Trans. Roy. Soc. (London) **A239**, 305 (1944).
[9] D. E. Littlewood, Phil. Trans. Roy. Soc. (London) **A239**, 387 (1944).
[10] D. E. Littlewood, Proc. London Math. Soc. **50**, 349 (1948).

known isomorphism between the representations $[\mu]$ of the ternary orthogonal group and the representations $\{2\mu\}$ of the binary full linear group to evaluate the plethysms of Eq. (5) in a direct manner. In this case we focus our attention on the evaluation of the plethysms

$$\{2l\} \otimes \{\lambda_1, \lambda_2, \cdots, \lambda_N\} \text{ and } \{2j\} \otimes \{\lambda_1, \lambda_2, \cdots, \lambda_N\}. \tag{6}$$

Each partition into two parts $\{\mu_1, \mu_2\}$, which arises in the binary analysis of the above plethysms, may be converted into a partition into just one part by making use of the equivalence

$$\{\mu_1, \mu_2\} = \{\mu_1 - \mu_2\}. \tag{7}$$

Thus, the character decompositions under $U_{2l+1} \rightarrow R_3$ and $U_{2j+1} \rightarrow R_3$ may be found by simply replacing the characters $\{2\mu\}$ that arise in the plethysms of Eq. (6) by the characters $[\mu]$ of R_3.

The operation of plethysm is distributive with respect to multiplication on the right and thus we may establish the equivalence

$$A \otimes (BC \cdots X) = (A \otimes B)(A \otimes C) \cdots (A \otimes X). \tag{8}$$

This simple result may be used to simplify the evaluation of the plethysms of Eq. (6) by noting that any S function having N parts may be written as a linear combination of products of S-functions having one part by use of the determinantal relation,[4]

$$\{\lambda_1, \lambda_2, \cdots, \lambda_N\} = \begin{vmatrix} \{\lambda_1\} & \{\lambda_1 + 1\} & & \cdots & \\ \{\lambda_2 - 1\} & \{\lambda_2\} & \{\lambda_2 + 1\} & & \\ & & \cdots & & \\ & & & \{\lambda_N - 1\} & \{\lambda_N\} \end{vmatrix}, \tag{9}$$

e.g.,

$$\{4, 3, 1\} = \begin{vmatrix} \{4\} & \{5\} & \{6\} \\ \{2\} & \{3\} & \{4\} \\ 0 & \{0\} & \{1\} \end{vmatrix}$$
$$= \{6\}\{2\} + \{4\}\{3\}\{1\} - \{5\}\{2\}\{1\} - \{4\}\{4\}.$$

Thus, noting Eqs. (8) and (9), we conclude that every plethysm of Eq. (6) may be evaluated in terms of the simpler plethysms $\{m\} \otimes \{n\}$ where m and n are positive integers.

III. HERMITE'S RECIPROCITY LAW

Hermite's reciprocity law states that[6] "the number of invariants and covariants of degree m of a binary form of degree n is the same as the number of invariants and covariants of degree n of a binary form of degree m." In terms of plethysm, this is equivalent to the statement that the binary analysis of the plethysms $\{m\} \otimes \{n\}$ and $\{n\} \otimes \{m\}$ for the linear group of any dimension coincide, i.e.,

$$\{m\} \otimes \{n\} = \{n\} \otimes \{m\}. \tag{10}$$

Murnaghan[7] has used this result to deduce the important recursive relation

$$\{m\} \otimes \{n\} = \{m\} \otimes \{n-2\} + \{m-2\} \otimes \{n\} + (\{m-1\} \otimes \{n-1\}) \times (\{m+n-1\} - \{m+n-3\}). \tag{11}$$

Furthermore, for the group GL_2 we have:

(a) if p is any positive integer $p \geq m + k - 1$, then

$$\{p\}(\{m+n-1\} - \{m+n-3\}) = \{p+m+n-1\} + \{p-m-n+1\}; \tag{12a}$$

(b) if $p = m + n - 2$, then

$$\{p\}(\{m+n-1\} - \{m+n-3\}) = \{p+m+n-1\}; \tag{12b}$$

and

(c) if $p \leq m + n - 3$, then

$$\{p\}(\{m+n-1\} - \{m+n-3\}) = \{p+m+n-1\} - \{m+n-3-p\}. \tag{12c}$$

Equation (11) gives a rapid method for determining plethysms of the type $\{m\} \otimes \{n\}$, and is particularly suited to machine calculation. Relevant tables of these plethysms are being published elsewhere.[11]

IV. BOSON AND FERMIONS

The plethysm $\{2l\} \otimes \{1^n\}$ plays a key role in the determination of the orbital angular-momentum (L) states of maximum multiplicity in l^N-type configurations of LS-coupled fermions, while the plethysm $\{2j\} \otimes \{1^n\}$ is important in describing the total angular-momentum (J) states in j^n-type configurations of jj-coupled fermions. Alternatively, we may identify the analysis of the plethysm $\{m\} \otimes \{1^n\}$ with the totally antisymmetric states of a system of n identical particles, each of angular momentum $m/2$.

Correspondingly, the analysis of the plethysm $\{m\} \otimes \{n\}$ may be identified with the totally symmetric states of a system of n identical particles, each of angular momentum $m/2$. Physical realizations of these

[11] B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (John Wiley & Sons, Inc., New York, 1969).

states would occur where the identical particles are bosons.

Murnaghan[7] has shown that in the case of the binary full linear group, Hermite's reciprocity principle leads naturally to the identity

$$\{m\} \otimes \{1^n\} = \{m + 1 - n\} \otimes \{n\}, \quad m + 1 \geq n. \tag{13}$$

This result gives a direct link between the totally antisymmetric states of a configuration of n identical particles each having angular momentum $m/2$, and the totally symmetric states of a configuration of n identical particles each having angular momentum $(m + 1 + n)/2$. For example, with $m = 8$ and $n = 3$ we see that the totally antisymmetric orbital angular states of g^3 are the same as for the totally symmetric orbital states of f^3. With $m = 8$ and $n = 4$ the correspondence is between g^4 and $(\frac{5}{2})^4$.

If Eq. (10) is used on the rhs of Eq. (13) we obtain the result

$$\{m\} \otimes \{1^n\} = \{n\} \otimes \{m + 1 - n\}, \quad m + 1 \geq n, \tag{14}$$

from which we conclude that the totally antisymmetric orbital states of n identical particles each of angular momentum $m/2$, are the same as the totally symmetric orbital $(m + 1 - n)$ particles each of angular momentum $n/2$. For example, with $m = 8$ and $n = 4$, the correspondence is between the antisymmetric states of g^4 and the symmetric states of d^5.

If Eq. (13) is used again in Eq. (14) we find

$$\{m\} \otimes \{1^n\} = \{m\} \otimes \{1^{m+1-n}\}. \tag{15}$$

In terms of jj-coupled states, the above relationship corresponds to the well-known particle–hole equivalence theorem[12] that states that the angular-momentum states occurring in j^n are the same as those in j^{2j+1-n}. In terms of LS-coupled states, Eq. (15) demonstrates the equivalence of the orbital angular-momentum states in l^n and l^{2l+1-n} for states of maximum multiplicity, the so-called quarter-shell symmetry.[13]

Equations (13) and (14) are particularly valuable in enumerating the angular-momentum states of jj-coupled configurations and of the orbital angular-momentum states associated with the states of maximum multiplicity in LS-coupled configurations, since they reduce the problem down to the evaluation of plethysms of the simple type $\{m\} \otimes \{n\}$. For example, the orbital states having maximum spin in g^4 are readily determined from the correspondence $\{8\} \otimes \{1^4\} = \{5\} \otimes \{4\}$.

V. THREE-PARTICLE CONFIGURATIONS

The preceding results may be readily applied to the angular-momentum analysis of configurations of three identical LS-coupled particles, i.e., the configuration l^3. In GL_2 we have

$$\{2l\} \otimes (\{1^2\}\{1\}) = \{2l\} \otimes \{1^3\} + \{2l\} \otimes \{21\}. \tag{16}$$

The terms in $\{2l\} \otimes \{1^3\}$ give the orbital angular-momentum states associated with maximum spin $(S = \frac{3}{2})$, and those in $\{2l\} \otimes \{21\}$, are those associated with spin $S = \frac{1}{2}$.

We first establish the orbital angular-momentum states without regard to their spin classification. The lhs of Eq. (16) may be written [using Eq. (8)] as

$$\{2l\} \otimes (\{1^2\}\{1\}) = (\{2l\} \otimes \{1^2\})\{2l\} = \sum_{\alpha=0}^{l-1} \{4l - 2 - 4\alpha\}\{2l\}, \tag{17}$$

where

$$l - 1 \geq \alpha \geq 0 \tag{18}$$

and α is a positive integer.

Recalling Eq. (7) and multiplying the S-functions in Eq. (17) for the two possibilities:

(a) $4l - 2 - 4\alpha \geq 2l$, i.e., $l - 1 \geq 2\alpha \geq 0$,

$$\{4l - 2 - 4\alpha\}\{2l\} = \sum_{\beta=0}^{2l} \{6l - 2 - 4\alpha - 2\beta\}, \tag{19}$$

where β is a positive integer and where

$$2l \geq \beta \geq 0; \tag{20}$$

(b) $4l - 2 - 4\alpha < 2l$, i.e., $2l - 2 \geq 2\alpha > l - 1$,

$$\{4l - 2 - 4\alpha\}\{2l\} = \sum_{\beta=0}^{4l-2-4\alpha} \{6l - 2 - 4\alpha - 2\beta\}, \tag{21}$$

where

$$4l - 2 - 4\alpha \geq \beta \geq 0; \tag{22}$$

we conclude that the given value of the orbital angular momentum L occurs if

$$2\alpha + \beta = 3l - L - 1, \tag{23}$$

where α and β are positive integers subjected to the restrictions of Eqs. (18)–(22).

As an example, consider the case of k^3, i.e., $l = 7$. Case (a) involves the restrictions $6 \geq 2\alpha \geq 0$ and $14 \geq \beta \geq 0$, while case (b) involves the restrictions $6 \geq \alpha \geq 4$ and $26 - 4\alpha \geq \beta \geq 0$. Furthermore, Eq. (23) is only satisfied if $2\alpha + \beta = 20 - L$. Using these results to enumerate the partitions (α, β) that

[12] A. de Shalit and I. Talmi, *Nuclear Shell Theory* (Academic Press Inc., New York, 1963).
[13] B. R. Judd, Phys. Rev. **125**, 613 (1962).

TABLE I. Enumeration of the orbital states of k^3.

L	$2\alpha+\beta$	(α,β)	nL
0	20	(3, 14)	1S
1	19		1P
2	18	(4, 10) (2, 14)	3D
3	17		3F
4	16	(5, 6) (1, 14)	5G
5	15		5H
6	14	(6, 2) (0, 14)	7I
7	13	↕	7K
8	12	(6, 0)	7L
9	11		6M
10	10	(5, 0)	6N
11	9		5O
12	8	(4, 0)	5Q
13	7		4R
14	6	(3, 0)	4T
15	5		3U
16	4	(2, 0)	3V
17	3		2W
18	2	(1, 0)	2X
19	1		Y
20	0	(0, 0)	Z

satisfy Eqs. (18)–(23) as in Table I rapidly gives the orbital states of k^3.

The orbital states associated with $S=\frac{3}{2}$ in l^3 may be readily determined by first analyzing the plethysm $\{2l\}\otimes\{3\}$ in terms of the recursive relation of Eq. (11), and the known result[8] for $\{2l\}\otimes\{2\}$, to give

$$\{2l\}\otimes\{3\} = \sum_{x=0}^{l}\{2(l-x)\} + \sum_{\alpha=0}^{l-1}\sum_{\beta=0}^{l-1-\alpha}\{4l-2-4\alpha-4\beta\} \times (\{2l+2-2\alpha\}-\{2l-2\alpha\}), \quad (24)$$

and then noting the restrictions on l, α, and β that arise in performing the S-function multiplication in terms of Eqs. (12a)–(12c). In this way, we readily establish that subject to the restrictions

$$l\geq x\geq 0, \quad l-1\geq\alpha\geq 0,$$

and

$$l-1-\alpha\geq\beta\geq 0, \quad (25)$$

the angular-momentum state characterized by L occurs once, whenever $L=l-x$ or whenever α and β satisfy the conditions

$$3\alpha+2\beta=3l-L \quad \text{or} \quad \alpha+2\beta=l-2-L. \quad (26)$$

TABLE II. Calculation of the states of maximum multiplicity in k^3.

L	(L) in x	(L) in $3\alpha+2\beta=18-L$	(L) in $\alpha+2\beta=4-L$	$(-L)$ in $\alpha+2\beta=L+5$	nL
0	1		(4, 0) (2, 1) (0, 2) (5, 0)	(3, 1) (1, 2)	1S
1	1		(3, 0) (1, 1)	(4, 1) (2, 2) (0, 3)	
2	1		(2, 0) (0, 1)	(3, 2) (1, 3)	1D
3	1	(5, 0)	(1, 0)	(2, 3) (0, 4)	1F
4	1	(4, 1)	(0, 0)	(1, 4)	2G
5	1	(3, 2)		(0, 5)	1H
6	1	(4, 0) (2, 3)			3I
7	1	(3, 1) (1, 4)			2K
8		(2, 2) (0, 5)			2L
9		(3, 0) (1, 3)			2M
10		(2, 1) (0, 4)			2N
11		(1, 2)			1O
12		(2, 0) (0, 3)			2Q
13		(1, 1)			1R
14		(0, 2)			1T
15		(1, 0)			1U
16		(0, 1)			1V
17					
18		(0, 0)			1X

TABLE III. Terms of k^3.

U_{15}	R_{15}	^{2S+1}L
$\{1^3\}$	$[1^3]$	$^4SDFG_2HI_3K_2L_2M_2N_2OQ_2RTUVX$
$\{21\}$	$[21]$	$^2PD_2F_2G_3H_4I_4K_4L_5M_4N_4O_4Q_3R_3T_3U_2V_2W_2XYZ$
	$[1]$	^2K

From the resultant, we subtract the number of times L satisfies the condition

$$\alpha + 2\beta = L + l - 1. \tag{27}$$

Thus, to establish the orbital states associated with maximum spin in k^3, we note from Eq. (13) that

$$\{14\} \otimes \{1^3\} = \{12\} \otimes \{3\}$$

and from Eq. (25) that

$$5 \geq x \geq 0, \quad 5 \geq \alpha \geq 0, \quad \text{and} \quad 5 - \alpha \geq \beta \geq 0.$$

Using these results in Eqs. (26) and (27) we readily deduce the results of Table II. The terms associated with $S = \frac{1}{2}$ may be found by subtracting the entries of nL in Table II from the corresponding entries in Table I, to yield the final classification of k^3 under $U_{30} \rightarrow SU_2 \times SU_{15} \rightarrow SU_2 \times R_{15} \rightarrow SU_2 \times R_3$ as given in Table III.

The procedure we have outlined has the advantage over the traditional methods of giving the number of times a given L value occurs, without requiring the complete enumeration of all the states of l^3, and is particularly useful in the systematic investigation of the properties of equivalent particle configurations having large values of angular momentum.

The classification of the states of j^3 may be made in a similar manner. A term with total angular momentum J is found for every solution of the equation

$$2\alpha + \beta = 3j - 1 - J, \tag{28}$$

where, for $j - 1 \geq 2\alpha \geq 0$, we have

$$2j \geq \beta \geq 0 \tag{29a}$$

and, for $2j - 2 \geq 2\alpha > j - 1$, we have

$$4j - 2 - 4\alpha \geq \beta \geq 0. \tag{29b}$$

The states associated with the $\{1^3\}$ representation of U_{2j+1} are found by replacing Eqs. (25)–(27) by the analogous equations

$$j - \tfrac{1}{2} \geq x \geq 0, \quad j - \tfrac{1}{2} \geq \alpha \geq 0,$$

and

$$j - \tfrac{1}{2} - \alpha \geq \beta \geq 0, \tag{25'}$$

$$3\alpha + 2\beta = 3j - J \quad \text{or} \quad \alpha + 2\beta = j - 2 - J, \tag{26'}$$

$$\alpha + 2\beta = J + j - 1, \tag{27'}$$

and noting that

$$\{2j\} \otimes \{1^n\} = \{2j + 1 - n\} \otimes \{n\}. \tag{30}$$

VI. NUMBER OF S STATES IN l^4

Judd[14] has shown that, while the Coulomb energies of a configuration of equivalent electrons l^n are expressible in terms of just $(l + 1)$ Slater radial integrals, the relative Coulomb energies of the terms of maximum spin may be expressed in terms of a number of parameters equal to the number of S states contained in the representation $[2^2]$ of R_{2l+1}. These parameters correspond to the particular linear combinations of the $(l + 1)$ Slater integrals. The results obtained for l^3 may be applied to give a remarkably succinct expression for the number of S states that occur in $[2^2]$ of R_{2l+1} for any l. We obtain this result as follows.

We first note that under $U_{2l+1} \rightarrow R_3$, the irreducible representations $\{2\}$ and $\{1^2\}$ yield no common terms, and hence the product $\{2\}\{1^2\}$ cannot yield an S state. But by ordinary S-function multiplication,

$$\{2\}\{1^2\} = \{21^2\} + \{31\}$$

and, hence under $U_{2l+1} \rightarrow R_3$, neither $\{21^2\}$ or $\{31\}$ can yield an S state. Furthermore,

$$\{1^3\}\{1\} = \{1^4\} + \{21^2\},$$

from which it follows that the number of S states in $[1^4]$ of R_{2l+1} is equal to the number of times $L = l$ appears in the states of maximum multiplicity of l^3, which is itself equal to the number of partitions (α, β), such that

$$3\alpha + 2\beta = 2l \quad \text{where} \quad l - 1 \geq \alpha \geq 2$$

and

$$l - 1 - \alpha \geq \beta \geq 0, \tag{31}$$

a result that follows as a consequence of Eqs. (25)–(27). In deriving this result we have made use of the fact that the number of S states in $[1^4]$ of R_{2l+1} is the same as that in [4], which follows upon noticing that

$$\{2\}\{2\} = \{4\} + \{2^2\} + \{31\} \supset (l + 1)S, \tag{32a}$$

$$\{1^2\}\{1^2\} = \{1^4\} + \{2^2\} + \{21^2\} \supset lS. \tag{32b}$$

Equation (31) can only be satisfied if α is even, from which we conclude that the number of S states in $[1^4]$ of R_{2l+1} is equal to the largest integer τ such that $3\tau + x = l$ where $l > x$. It follows from Eq. (32b) that the desired result is that the number of S states in $[2^2]$ of R_{2l+1} is $l - 1 - \tau$. Thus, we may readily determine the number of S states in $[2^2]$ without knowing the complete decomposition under

$$R_{2l+1} \rightarrow R_3.$$

ACKNOWLEDGMENT

I am grateful to Dr. Yves Bordarier for a penetrating question that led to the results of Sec. IV of this paper.

[14] B. R. Judd, Phys. Rev. 162, 28 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Complementary Bounds for Ground-State Energies

PETER D. ROBINSON*
Battelle Memorial Institute, Columbus, Ohio

(Received 7 June 1968)

Complementary upper and lower bounds are derived for the ground-state energies of Hamiltonians $H = h + V$, where V is positive-definite, by considering the Schrödinger equation in an integral form. Some simple applications are presented.

1. INTRODUCTION

In this paper, complementary upper and lower bounds are derived for the ground-state energy E of a Hamiltonian H which can be decomposed in the form

$$H = h + V. \tag{1}$$

It is assumed that the operator V is positive-definite and can be regarded as a perturbation to h, which we suppose has eigenvalues $\{\epsilon_n\}$ and normalized eigenfunctions $\{\phi_n\}$. Further, we assume that ψ, the ground state of H, is approximated by ϕ_0 and that E lies between ϵ_0 and ϵ_1, the two lowest eigenvalues of h, i.e.,

$$\epsilon_0 < E < \epsilon_1. \tag{2}$$

Adopting a customary starting point of perturbation theories,[1] we consider the Schrödinger equation

$$(h - E)\psi = -V\psi \tag{3}$$

as an equivalent integral equation

$$\psi(\mathbf{r}) = \phi_0(\mathbf{r}) - \int \mathcal{K}(\mathbf{r}, \mathbf{s})V(\mathbf{s})\psi(\mathbf{s})\, d\mathbf{s}, \tag{4}$$

which is subject to the condition

$$E - \epsilon_0 = \int \phi_0(\mathbf{r})V(\mathbf{r})\psi(\mathbf{r})\, d\mathbf{r}. \tag{5}$$

The operator

$$K = \sum_{n>0} \frac{|\phi_n\rangle\langle\phi_n|}{\epsilon_n - E} \tag{6}$$

is the integral operator with kernel

$$\mathcal{K}(r, s) = \sum_{n>0} \frac{\phi_n(\mathbf{r})\phi_n(\mathbf{s})}{\epsilon_n - E}. \tag{7}$$

For simplicity, let all functions be real. Then $\mathcal{K}(\mathbf{r}, \mathbf{s})$ is a symmetric kernel in the ordinary sense and by virtue of (2) it is also positive-definite.

Recently,[2] complementary variational principles have been developed for integral equations of the type

$$m\{\Phi(\mathbf{r})\} = \int \mathcal{K}(\mathbf{r}, \mathbf{s})\Phi(\mathbf{s})\, d\mathbf{s}, \tag{8}$$

which have symmetric positive-definite kernels, and here that theory is applied to Eq. (4). Implicit upper and lower bounds are obtained for E, which can be simplified to yield explicit bounds in certain instances. To illustrate the formulas, they are applied to the heliumlike atom and the perturbed linear oscillator.

2. COMPLEMENTARY BOUNDS

Equation (8) reduces to (4) with the substitutions

$$\Phi(\mathbf{r}) = V(\mathbf{r})\psi(\mathbf{r}) \tag{9}$$

and

$$m(\Phi) = \phi_0 - V^{-1}\Phi = \phi_0 - \psi. \tag{10}$$

The theory of Ref. 2 then shows that the functional

$$I(\psi) = \int \phi_0(\mathbf{r})V(\mathbf{r})\psi(\mathbf{r})\, d\mathbf{r} = \langle\phi_0|\, V\, |\psi\rangle \tag{11}$$

has *complementary* lower and upper bounds

$$G(\Psi) \leq I(\psi) \leq J(\Theta), \tag{12}$$

where

$$G(\Psi) = \int [-(V\Psi)K(V\Psi) - \Psi V\Psi + 2\phi_0 V\Psi]\, d\mathbf{r}, \tag{13}$$

$$J(\Theta) = \int [(V\Theta)K(V\Theta) + \{\phi_0 - K(V\Theta)\}V\{\phi_0 - K(V\Theta)\}]\, d\mathbf{r}, \tag{14}$$

and Ψ and Θ are trial functions. That G and J are indeed lower and upper bounds is easy to see directly from the relations

$$I - G = \int [\{V(\Psi - \psi)\}K\{V(\Psi - \psi)\} + (\Psi - \psi)V(\Psi - \psi)]\, d\mathbf{r} \geq 0, \tag{15}$$

$$J - I = \int \{\{V(\Theta - \psi)\}K\{V(\Theta - \psi)\} + [K\{V(\Theta - \psi)\}]V[K\{V(\Theta - \psi)\}]\}\, d\mathbf{r} \geq 0. \tag{16}$$

---

* On leave from Mathematics Department, York University, England. The author is most grateful to the Battelle Institute for a visiting Fellowship.

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., New York, 1953), Chap. 9.

[2] P. D. Robinson and A. M. Arthurs, J. Math. Phys. **9**, 1364 (1968).

(The first-order terms in $(\Psi - \psi)$ and $(\Theta - \psi)$ have disappeared from (15) and (16) by virtue of (4) and the self-adjointness of K.) Thus from (5), (11), and (12) we have the result

$$G(\Psi) \leq E - \epsilon_0 \leq J(\Theta), \tag{17}$$

which gives implicit bounds for E, since the functionals G and J depend on E via the operator K.

3. SIMPLE EXPLICIT BOUNDS

If we take as trial functions

$$\Psi = \Theta = \alpha V^{-1}\phi_0, \tag{18}$$

where α is an arbitrary factor, then, since

$$K\phi_0 = 0, \tag{19}$$

(13) and (14) simplify to give

$$G(\alpha V^{-1}\phi_0) = 2\alpha - \alpha^2\langle\phi_0|\, V^{-1}\,|\phi_0\rangle \tag{20}$$

and

$$J(\alpha V^{-1}\phi_0) = \langle\phi_0|\, V\,|\phi_0\rangle = E_1, \tag{21}$$

E_1 being the first-order perturbation correction to ϵ_0. If we choose

$$\alpha = \langle\phi_0|\, V^{-1}\,|\phi_0\rangle^{-1} \tag{22}$$

to maximize (20), then the result

$$\langle\phi_0|\, V^{-1}\,|\phi_0\rangle^{-1} \leq E - \epsilon_0 \leq E_1 \tag{23}$$

is obtained.

The upper bound for E in (23) is well known for nonnegative perturbations, and the lower bound has previously been found as an example of partitioning and projection techniques.[3-5] The derivation here is elementary, and exhibits the complementary nature of the bounds.

An illustration of (23) arises for the ground state of a heliumlike system with nuclear charge $Z \geq 2$, so that condition (2) holds. With

$$\phi_0 = (Z^3/\pi)e^{-Z(r_1+r_2)}, \quad V = 1/r_{12}, \tag{24}$$

(23) becomes

$$\tfrac{16}{35}Z \leq E + Z^2 \leq \tfrac{5}{8}Z, \tag{25}$$

which gives

$$-3.0857 \leq E \leq -2.75 \tag{26}$$

for helium when $Z = 2$.

4. BRILLOUIN-WIGNER BOUNDS

Another simplifying choice of trial function involves $\chi(E)$, the first-order Brillouin–Wigner correction to ϕ_0. This is defined by

$$\chi = -K(V\phi_0) \tag{27}$$

or, equivalently, as the acceptable solution of the nonhomogeneous equation

$$(h - E)\chi = (E_1 - V)\phi_0. \tag{28}$$

If we set

$$\Psi = \beta\phi_0, \quad \Theta = \gamma\phi_0, \tag{29}$$

then (13) and (14) reduce to

$$G(\beta\phi_0) = 2\beta E_1 - \beta^2(E_1 - \mathcal{E}_2) \tag{30}$$

and

$$J(\gamma\phi_0) = E_1 + 2\gamma\mathcal{E}_2 + \gamma^2(\mathcal{E}_3 - \mathcal{E}_2), \tag{31}$$

where

$$\mathcal{E}_2(E) = \langle\chi(E)|\, V\,|\phi_0\rangle < 0 \tag{32}$$

and

$$\mathcal{E}_3(E) = \langle\chi(E)|\, V\,|\chi(E)\rangle > 0. \tag{33}$$

The signs of $\mathcal{E}_2$ and $\mathcal{E}_3$ are consequences of K and V being positive-definite. Choosing

$$\beta = \frac{E_1}{E_1 - \mathcal{E}_2}, \quad \gamma = -\frac{\mathcal{E}_2}{\mathcal{E}_3 - \mathcal{E}_2} \tag{34}$$

to maximize G and minimize J, we obtain from (17) the result

$$G_{\max} = \frac{E_1^2}{E_1 - \mathcal{E}_2} \leq E - \epsilon_0 \leq E_1 - \frac{\mathcal{E}_2^2}{\mathcal{E}_3 - \mathcal{E}_2} = J_{\min}, \tag{35}$$

which may be written in the alternative form

$$\left\{\frac{1}{\mathcal{E}_2} - \frac{1}{E_1}\right\}^{-1} \leq E - \epsilon_0 - E_1 \leq \left\{\frac{1}{\mathcal{E}_2} - \frac{\mathcal{E}_3}{\mathcal{E}_2^2}\right\}^{-1}. \tag{36}$$

The bounds in (36) are still implicit, because $\mathcal{E}_2$ and $\mathcal{E}_3$ depend on E. However, progress is possible because as E increases, $\mathcal{E}_2(E)$ decreases, and therefore so does the lhs of (36) decrease. Thus by using an upper bound E_+ for E, e.g.,

$$E_+ = \epsilon_0 + E_1 \tag{37}$$

(or a better one if available) on the left of (36), we get an explicit lower bound

$$E_- = \epsilon_0 + E_1 + \left\{\frac{1}{\mathcal{E}_2(E_+)} - \frac{1}{E_1}\right\}^{-1} \leq E. \tag{38}$$

Whenever

$$\frac{d}{dE}(\mathcal{E}_2 + \mathcal{E}_3) < 0, \tag{39}$$

the rhs of (36) increases as E decreases. Then the lower bound E_- can be substituted for E to give a better upper bound and this iterative process continued to improve both bounds.

[3] P.-O. Löwdin, Phys. Rev. **139**, A357 (1965).
[4] P.-O. Löwdin, J. Chem. Phys. **43**, S175 (1965).
[5] J. H. Choi and D. W. Smith, J. Chem. Phys. **45**, 4425 (1966).

As an illustration, we consider the linear oscillator with

$$h = \frac{1}{2}\left(x^2 - \frac{d^2}{dx^2}\right), \quad \phi_0 = \pi^{-\frac{1}{4}} e^{-\frac{1}{2}x^2}, \quad \epsilon_0 = \tfrac{1}{2}, \quad (40)$$

in the two cases (i) $V = \lambda x^2$, and (ii) $V = \lambda x^4$, $\lambda > 0$. If we introduce a function $f(x)$ so that

$$\chi = f\phi_0, \quad (41)$$

then (28) gives

$$\frac{d^2f}{dx^2} - 2x\frac{df}{dx} + (2E - 1)f = 2(V - E_1), \quad (42)$$

which can be solved for f in these cases. The following results are obtained:

(*i*) *Perturbed oscillator;* $V = \lambda x^2$, $\lambda > 0$:

$$E_1 = \tfrac{1}{2}\lambda; \quad f = \frac{\lambda}{5 - 2E}(1 - 2x^2);$$

$$\mathcal{E}_2 = -\frac{\lambda^2}{5 - 2E}; \quad \mathcal{E}_3 = \frac{5\lambda^3}{(5 - 2E)^2};$$

$$\frac{-\lambda^2}{5 - 2E + 2\lambda} \leq E - \tfrac{1}{2}(1 + \lambda) \leq \frac{-\lambda^2}{(5 - 2E) + 5\lambda}.$$

Successive bounds with $\lambda = \frac{1}{4}$:

$$E_+ = 0.6250,\ 0.6126,\ 0.6126,$$
$$E_- = 0.6153,\ 0.6104,\ 0.6104$$
$$[\text{true } E = \tfrac{1}{2}(1 + 2\lambda)^{\frac{1}{2}} = 0.6124].$$

(*ii*) *Perturbed oscillator;* $V = \lambda x^4$, $\lambda > 0$:

$$E_1 = \tfrac{3}{4}\lambda;$$

$$f = \frac{\lambda}{9 - 2E}\left[\frac{3}{2}\cdot\frac{(13 - 2E)}{(5 - 2E)} - \frac{24x^2}{5 - 2E} - 2x^4\right];$$

$$\mathcal{E}_2 = -3\lambda^2\left[\frac{3}{5 - 2E} + \frac{1}{9 - 2E}\right];$$

$$\mathcal{E}_3 = \frac{9\lambda^3}{2}\left[\frac{39}{(5 - 2E)^2} + \frac{56}{(9 - 2E)(5 - 2E)} + \frac{41}{(9 - 2E)^2}\right].$$

Successive bounds with $\lambda = \frac{1}{4}$:

$$E_+ = 0.6875,\ 0.6329,\ 0.6329,$$
$$E_- = 0.5957,\ 0.5970,\ 0.5970$$
$$(\text{true } E = 0.6209).^6$$

5. DISCUSSION

More sophisticated bounds can be derived from (13) and (14) by using such trial functions as

$$V^{-1}(h - E)Y \quad (43)$$

or

$$(KV)^n\phi_0 \quad (44)$$

or optimized linear combinations of them. Löwdin[3] has discussed the use of the Brillouin–Wigner corrections (44) in obtaining bounds; the lower bound in (35) would seem to be a corrected version of Eq. (134) in his paper. However, it is not our aim to present highly accurate formulas. These are to be found in the *tours de force* of Löwdin,[3,4] and of Bazley and Fox.[6] We are primarily content with emphasizing the complementary nature of the bounds $G(\Psi)$ and $J(\Theta)$, and showing how easily they can be derived.

[6] N. W. Bazley and D. W. Fox, Phys. Rev. **124**, 483 (1961); see also further papers cited in Ref. 3.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Frequency Spectrum and Momentum Autocorrelation Function for a Simple Lattice

E. M. BAROODY
Battelle Memorial Institute, Columbus Laboratories, Columbus, Ohio

(Received 20 April 1968)

Calculations are made on the dynamics of a familiar lattice model: the simple cubic lattice with central and noncentral harmonic forces between nearest neighbors only. For the case of equal force constants, natural extensions are made of existing analytical approximations for the spectrum of squared frequency. The behavior of the spectrum near its singular points is described more accurately than before and expressions are derived which make it easy to obtain the density of modes at any frequency to about one part in a thousand. The new description of the spectrum is used to improve existing approximations for the classical momentum autocorrelation function for the infinite lattice $\rho(\tau)$, and for the function $X_3(\tau')$ used by Goodman in calculations on the response of surface atoms in a semi-infinite lattice. Good agreement with numerical results of Goodman for $\tau' = 20, 25$, and 30 is obtained. The results for the spectrum also apply to the density of states of electrons in a simple cubic lattice in the tight-binding approximation.

1. INTRODUCTION

Probably the most tractable crystal model which shows a reasonable three-dimensional vibrational behavior is the infinite simple cubic lattice with central and noncentral harmonic forces between nearest neighbors only. It has the special feature that the components of displacement along the three coordinate axes are uncoupled, but in spite of this its qualitative implications are similar to those of more realistic three-dimensional lattices. The model was used as early as 1927 by Waller,[1] but recent interest has followed mainly from work done in the 1950's, particularly that of Rosenstock[2,3] and his collaborators. It has frequently been taken as a starting point in discussions of the influence of defects on lattice dynamics and a number of workers[4–7] have used it in calculations of correlation (or response) functions. A substantial part of the work done with the model has been reviewed by Maradudin, Montroll, and Weiss.[8]

The chief merit of the model is that it has permitted many interesting questions to be studied analytically. A full plot of the frequency spectrum has apparently been obtained only by numerical integration, however. Moreover, in his recent work on response functions, Goodman[7] has used a direct summation over many normal modes to fill a gap between approximations which apply for early and for late times. The purpose of the present paper is to show that, for the case of equal force constants, rather straightforward extensions of existing approximations remove a good deal of the need for numerical work. Section 2 provides a description of the spectrum of squared frequency which makes it easy to calculate the density of normal modes at any frequency to about one part in a thousand. This has a double interest since the same function gives the density of states for electrons in a simple cubic lattice in the tight-binding approximation.[9] In Sec. 3 the available asymptotic approximation for the momentum autocorrelation function at long times is improved by calculating the terms varying as $t^{-\frac{5}{2}}$.

Until Sec. 4, all of our calculations are for an atom in an infinite lattice. There it is shown that the inclusion of the additional terms in the asymptotic approximation for the autocorrelation function leads to good agreement with numerical results of Goodman on the response of a surface atom in a semi-infinite lattice model.

2. SPECTRUM OF SQUARED FREQUENCY AND ITS MOMENTS

For this model, the frequency ω of a lattice wave with propagation vector $\mathbf{k}$ is independent of polarization, all three branches of the spectrum satisfying the relation

$$\omega^2 = (2/m) \sum \gamma_j (1 - \cos k_j a). \qquad (2.1)$$

Here, m is the mass of a lattice particle, a is the lattice constant, and the quantities γ_j (with $j = 1, 2, 3$)

[1] I. Waller, Ann. Physik **83**, 153 (1927).
[2] W. A. Bowers and H. B. Rosenstock, J. Chem. Phys. **18**, 1056 (1950).
[3] H. B. Rosenstock and G. F. Newell, J. Chem. Phys. **21**, 1607 (1953).
[4] P. Mazur and E. Montroll, J. Math. Phys. **1**, 70 (1960).
[5] R. J. Rubin, J. Math. Phys. **1**, 309 (1960); **2**, 373 (1961); Phys. Rev. **131**, 964 (1963).
[6] F. O. Goodman, J. Phys. Chem. Solids **23**, 1269 (1962).
[7] F. O. Goodman, Surface Sci. **3**, 386 (1965).
[8] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press Inc., New York, 1963).
[9] N. F. Mott and H. Jones, *Theory of the Properties of Metals and Alloys* (Oxford University Press, London, 1936). Curve (1) of Fig. 38a shows this density of states, but with the area under the curve equal to 2 (rather than to 1), corresponding to 2 electron states per atom of the lattice. The first evaluation of the function in connection with lattice dynamics appears to be that of Bowers and Rosenstock. (See Fig. 6 of Ref. 2.)

are central and noncentral force constants. We state basic equations for general values of the force constants, but our calculations deal only with the case of equal force constants.

Letting $x = (\omega/\omega_L)^2$, where ω_L is the maximum lattice frequency, the spectrum of squared frequency may be expressed as a single integral[8]

$$G(x) = (6/\pi)\int_0^\infty \cos 3\eta(1-2x) \times J_0(\lambda_1\eta)J_0(\lambda_2\eta)J_0(\lambda_3\eta)\,d\eta, \quad (2.2)$$

where $\lambda_j = 3\gamma_j/(\gamma_1+\gamma_2+\gamma_3)$ and J_0 is the zero-order Bessel function of the first kind. As this expression makes clear, the spectrum satisfies $G(x) = G(1-x)$. It is also very helpful to have the Laplace transform of the spectrum

$$\int_0^\infty \exp(-\eta x)G(x)\,dx = \exp(-\eta/2)I_0(\lambda_1\eta/6)I_0(\lambda_2\eta/6)I_0(\lambda_3\eta/6), \quad (2.3)$$

where I_0 is the zero-order modified Bessel function. A careful discussion of this equation has been given by Peretti.[10]

Spectrum for Equal Force Constants

For this case the most interesting features of $G(x)$ are infinities in slope at $x = \frac{1}{3}$ and $x = \frac{2}{3}$. To concentrate attention on these points we write (2.2) in the form

$$G(x) = G(\tfrac{1}{3}) - (6/\pi) \times \int_0^\infty [\cos\eta - \cos 3\eta(1-2x)]J_0^3(\eta)\,d\eta. \quad (2.4)$$

When x is near $\frac{1}{3}$ or $\frac{2}{3}$, heavy contributions to the integral are to be expected from the region of large η. This suggests that a good approximation can be obtained by introducing the asymptotic expansion

$$J_0(\eta) = \left(\frac{2}{\pi\eta}\right)^{\frac{1}{2}}\left[\cos\left(\eta-\frac{\pi}{4}\right) + \left(\frac{1}{8\eta}\right)\sin\left(\eta-\frac{\pi}{4}\right) + \cdots\right]. \quad (2.5)$$

Using two terms of $J_0^3(\eta)$ in the integral leads to no convergence problems at $\eta = 0$ and, without additional approximations, yields

$$G(x) = G(\tfrac{1}{3}) - (18/\pi^2)(1-3x)^{\frac{1}{2}} + (3/2\pi^2) \times (4\sqrt{2}-5)(1-3x) + (3/\pi^2)(1-3x)^{\frac{3}{2}}, \quad x \le \tfrac{1}{3}, \quad (2.6)$$

[10] J. Peretti, J. Phys. Chem. Solids 12, 216 (1960).

TABLE I. Coefficients in Eqs. (2.9) and (2.11).

j	c_j	d_j	j	c_j	d_j
1	$\frac{3}{2}$	$\frac{3}{4}$	7	102.13	0.90
2	$\frac{99}{40}$	$\frac{9}{40}$	8	248.87	2.20
3	$\frac{2599}{560}$	$\frac{155}{1120}$	9	622.4	5.5
4	$\frac{5751}{640}$	$\frac{81}{640}$	10	1588	13.7
5	19.189	0.1861	11	4119	34.3
6	43.324	0.382	12	10826	87

and

$$G(x) = G(\tfrac{1}{3}) - (3/\pi^2)(3+4\sqrt{2}) + (3\sqrt{3}/\pi^2) \times [x^{\frac{1}{2}}(2+3x) + (1-x)^{\frac{1}{2}}(5-3x)], \quad \tfrac{1}{3} \le x \le \tfrac{2}{3}. \quad (2.7)$$

Equation (2.6) should be good for the region just below $x = \frac{1}{3}$; an explicit expression for the region just beyond $x = \frac{2}{3}$ may be obtained by replacing x by $(1-x)$. The actual range of usefulness of these expressions is seen below. It turns out that (2.7) is a good approximation throughout the indicated range, not just very near the end points, as might have been imagined.

We now turn to the region near $x = 0$. An asymptotic expansion in ascending powers of x may be obtained from the Laplace transform

$$\int_0^\infty \exp(-\eta x)G(x)\,dx = \exp(-\eta/2)I_0^3(\eta/6). \quad (2.8)$$

Using the asymptotic series for the modified Bessel function, and recognizing that for large η the major contributions to the Laplace integral come from small x, one obtains

$$G(x) = (6/\pi^2)(3x)^{\frac{1}{2}}(1 + c_1x + c_2x^2 + c_3x^3 + \cdots), \quad (2.9)$$

where the coefficients c_j are given through $j = 12$ in Table I. For large j only a few terms in the cube of the Bessel function are important, and one may obtain the asymptotic approximation

$$c_j = 3(\tfrac{3}{4})^j \frac{(2j)!}{(2j+1)(j!)^2}\left(1 + \frac{1}{2j} + \frac{9}{8j^2} + \frac{57}{16j^3} + \cdots\right) = \frac{3^{j+1}}{2\pi^{\frac{1}{2}}j^{\frac{3}{2}}}\left(1 - \frac{1}{8j} + \frac{145}{128j^2} + \frac{2933}{1024j^3} + \cdots\right). \quad (2.10)$$

This representation of $G(x)$ is well known, but as it stands its range of usefulness is small. In seeking improvement we may take into account Eq. (2.6) which shows that as x approaches $\frac{1}{3}$, the derivative $G'(x)$ diverges as $27/\pi^2(1-3x)^{\frac{1}{2}}$. The expression

$$G(x) = (6/\pi^2)(3x)^{\frac{1}{2}}[4(1 - d_1x - d_2x^2 - \cdots) - 3(1-3x)^{\frac{1}{2}}] \quad (2.11)$$

TABLE II. Comparison of approximations for $G(x)$.

x	From Eq. (2.11) Terms through $j=5$	From Eq. (2.11) Terms through $j=12$	From Eq. (2.6) with $G(\frac{1}{3})=1.737$
0.20	0.6881	0.6880	0.70
0.25	0.8858	0.8853	0.888
0.30	1.182	1.180	1.180
0.31	1.270	1.267	1.267
0.32	1.383	1.380	1.379
$\frac{1}{3}$	1.745	1.740	1.737

incorporates this feature and may be made consistent with (2.9) by proper choice of the coefficients d_j. Values of these new coefficients through $j=12$ are included in Table I. For $j \geq 5$, the ratio c_j/d_j exceeds 100, promising a considerable increase in the range of x for which calculations can be made easily. For large j, one may obtain the approximation

$$d_j = \frac{3^{j+1}}{16\pi^{\frac{1}{2}} j^{\frac{5}{2}}}\left(1 - \frac{15}{8j} - \frac{707}{128j^2} - \cdots\right). \quad (2.12)$$

To investigate the accuracy of the above approximations we first observe that $G(\frac{1}{2})$ has the value[11]

$$G(\tfrac{1}{2}) = \left(\frac{6}{\pi}\right)\int_0^\infty J_0^3(\eta)\,d\eta = \left[\frac{2}{\Gamma(\frac{5}{6})\Gamma(\frac{2}{3})}\right]^2 = 1.712. \quad (2.13)$$

Secondly, Eq. (2.7) shows that as x decreases from $\frac{1}{2}$, $G(x)$ increases very slowly, reaching $1.712 + 0.025 = 1.737$ at $x = \frac{1}{3}$. With this value available, Eqs. (2.6) and (2.11) can be compared in the region below $x = \frac{1}{3}$. According to Table II, the expressions agree very well close to $x = 0.30$. We have, in fact, arrived at approximations which make it very easy to compute $G(x)$ at any point to about one part in a thousand. It appears that for $x \leq 0.25$, Eq. (2.11) should be used, while for x very near $\frac{1}{3}$, Eq. (2.6) is needed.

Moments of the Spectrum

Knowledge of the Laplace transform of $G(x)$ provides a highly convenient way of generating the moments

$$\mu_j = \int_0^1 x^j G(x)\,dx. \quad (2.14)$$

Expanding both sides of (2.8) in ascending powers of η and identifying corresponding coefficients yields the values through $j = 14$ which are listed in Table III.

[11] Y. L. Luke, *Integrals of Bessel Functions* (McGraw-Hill Book Co., Inc., New York, 1962). Formula (8) on p. 333 gives twice the quantity which we use. A suspected error in the statement of the formula was confirmed by evaluating the integral of the cube of $J_{\frac{1}{2}}(\eta) = (2/\pi\eta)^{\frac{1}{2}} \sin\eta$.

For large j, the major contribution to (2.14) comes from x near unity. Using this fact along with (2.9), and recalling $G(x) = G(1-x)$, one obtains the asymptotic relation

$$\mu_j = \left(\frac{3}{\pi j}\right)^{\frac{3}{2}}\left(1 + \frac{3}{8j} + \frac{313}{128j^2} + \frac{13{,}257}{1024j^3} + \cdots\right). \quad (2.15)$$

For $j = 15$, this approximation is correct to about 4 parts in a thousand. This is useful since, in some applications, the higher moments do not need to be known to great precision.

3. CLASSICAL MOMENTUM AUTOCORRELATION FUNCTION

The function of interest is

$$\rho(\tau) = (8\pi^3)^{-1}\int_0^{2\pi}\!\!\int\!\!\int \cos x^{\frac{1}{2}}\tau\; d\theta_1\, d\theta_2\, d\theta_3, \quad (3.1)$$

where $\tau = \omega_L t$ is the dimensionless time variable used by Mazur and Montroll,[4] and $\boldsymbol{\theta}$ is the product of the wave vector $\mathbf{k}$ and the lattice constant. The function may be interpreted as the momentum of a particle at time τ, given an initial state in which this particle had unit momentum in its equilibrium position and all other particles were at rest in equilibrium positions. Integration over a surface of constant frequency yields a single integral involving the frequency spectrum

$$\rho(\tau) = \int_0^1 G(x)\cos x^{\frac{1}{2}}\tau\,dx. \quad (3.2)$$

Expansion of the cosine gives

$$\rho = \sum \frac{(-1)^j \mu_j \tau^{2j}}{(2j)!}, \quad (3.3)$$

where the quantities μ_j are the moments defined in (2.14). This expression is very suitable for small τ and can be used without too much difficulty up to $\tau = 10$. Beyond this point one would like to depend upon the approximations suitable for large τ which are now discussed.

TABLE III. Moments of the spectrum of squared frequency.

j	μ_j	j	μ_j
1	1/2	8	0.04558853
2	7/24	9	0.03777040
3	3/16	10	0.03190726
4	149/1152	11	0.0273970
5	217/2304	12	0.0238498
6	17813/(12)^5	13	0.0210092
7	28031/2(12)^5	14	0.0186861

Introducing $y = x^{\frac{1}{2}} = \omega/\omega_L$ gives

$$\rho = \int_0^1 g(y) \cos y\tau \, dy, \tag{3.4}$$

where $g(y) = 2yG(y^2)$. According to the theory of Fourier transforms the behavior of ρ at large τ is determined by the form of $g(y)$ near its singularities (infinities in derivatives) at the three points $(\frac{1}{3})^{\frac{1}{2}}$, $(\frac{2}{3})^{\frac{1}{2}}$, and 1. Lighthill[12] discusses this theory and, in his Theorem 19 (p. 52) and Table I (p. 43), provides specific guidance in the calculation of asymptotic terms.

When the brief calculations are carried through for the singularity at $y = (\frac{1}{3})^{\frac{1}{2}}$, the corresponding contribution to ρ is

$$\rho_1 = 3^{\frac{1}{4}}\left(\frac{6}{\pi\tau}\right)^{\frac{3}{2}} \times \left[\cos\left(\frac{\tau}{\sqrt{3}} + \frac{\pi}{4}\right) - \frac{19\sqrt{3}}{8\tau}\sin\left(\frac{\tau}{\sqrt{3}} + \frac{\pi}{4}\right)\right]. \tag{3.5}$$

Equation (2.6) has been used, the important terms being those involving $(1 - 3x)^{\frac{1}{2}}$ and $(1 - 3x)^{\frac{3}{2}}$. Since $G(x) = G(1 - x)$, the relation

$$g(y) = 2yG(1 - y^2)$$

can be applied in working with the singularities at $(\frac{2}{3})^{\frac{1}{2}}$ and 1. Just beyond $(\frac{2}{3})^{\frac{1}{2}}$, the relevant terms in $g(y)$ are $-(36y/\pi^2)(3y^2 - 2)^{\frac{1}{2}}$ and $(6y/\pi^2)(3y^2 - 2)^{\frac{3}{2}}$. These lead to

$$\rho_2 = (24)^{\frac{1}{4}}\left(\frac{6}{\pi\tau}\right)^{\frac{3}{2}}\left[\sin\left(\left(\frac{2}{3}\right)^{\frac{1}{2}}\tau + \frac{\pi}{4}\right) + \frac{7}{8\tau}\left(\frac{3}{2}\right)^{\frac{1}{2}}\cos\left(\left(\frac{2}{3}\right)^{\frac{1}{2}}\tau + \frac{\pi}{4}\right)\right]. \tag{3.6}$$

A similar use of (2.9) for the singularity at $y = 1$ yields

$$\rho_3 = -\left(\frac{6}{\pi\tau}\right)^{\frac{3}{2}}\left[\cos\left(\tau + \frac{\pi}{4}\right) + \left(\frac{21}{8\tau}\right)\sin\left(\tau + \frac{\pi}{4}\right)\right]. \tag{3.7}$$

The sum $\rho = \rho_1 + \rho_2 + \rho_3$ gives an asymptotic expansion which is correct through terms of order $\tau^{-\frac{5}{2}}$. The main terms (those involving $\tau^{-\frac{3}{2}}$) were calculated earlier by Mazur and Montroll working with an integral in wave-vector space and similar calculations have been made by Goodman. The additional terms, however, are new.

[12] M. J. Lighthill, *Fourier Analysis and Generalised Functions* (Cambridge University Press, Cambridge, England, 1958).

TABLE IV. Values of $\rho(\tau)$ at some intermediate times.

τ	From Eq. (3.3)	From asymptotic approximation Main terms only	All terms
10	0.181	0.208	0.199
11	−0.0655	−0.043	−0.072
12	−0.199	−0.179	−0.215
13	−0.180	−0.165	−0.193

Table IV lists some values of $\rho(\tau)$ for the range $\tau = 10$ to $\tau = 13$. The first column follows from Eq. (3.3), enough terms being used to yield accurately the number of places stated. The later columns are based on the asymptotic approximation $\rho = \rho_1 + \rho_2 + \rho_3$. It is seen that the terms of order $\tau^{-\frac{5}{2}}$ improve the approximation a little, and that the final results are correct to about 10%. In this range of τ, the asymptotic approximation is of help, but it is not yet entirely satisfactory. In the next section, comparisons are made involving a related function, Goodman's $X_3(\tau')$, at a substantially later time. There the $\tau^{-\frac{5}{2}}$ terms improve the approximation greatly and agreement to 1% or better is obtained.

4. EVALUATION OF GOODMAN'S FUNCTION $X_3(\tau')$

Goodman's calculations on the dynamics of the lattice model under discussion were part of his study of the theory of thermal accommodation coefficients. For this reason he was particularly interested in a function X_3 which gives the displacement of a surface atom in a three-dimensional semi-infinite lattice at some time, the initial disturbance having been a unit velocity of this atom perpendicular to the lattice surface. He used a dimensionless time variable (which we denote by τ') which is related to the variable τ of Mazur and Montroll[4] by $\tau' = \tau/(3)^{\frac{1}{2}}$. In this notation the momentum autocorrelation function of a surface atom is $dX_3/d\tau'$.

Goodman[6] developed a relationship between the autocorrelation function of a surface atom in the semi-infinite model and that for an atom deep in the interior. In the notation of the present paper, this relationship becomes

$$\frac{dX_3}{d\tau'} = 2\int_0^1 (1 - x)G(x)\cos(3x)^{\frac{1}{2}}\tau' \, dx. \tag{4.1}$$

This quantity may be obtained from our $\rho(\tau)$ by calculating $2(\rho + d^2\rho/d\tau^2)$ and replacing τ by $(3)^{\frac{1}{2}}\tau'$. An integration then gives $X_3(\tau')$. At early times,

$$X_3(\tau') = \tau' - (5\tau'^3/24) + (\tau'^5/64) - \cdots, \tag{4.2}$$

in agreement with Goodman. For late times, use of

(3.5), (3.6), and (3.7) gives

$$X_3(\tau') = 2\left(\frac{2}{\pi\tau'}\right)^{\frac{3}{2}} \times \left\{\left[2\sin\left(\tau' + \frac{\pi}{4}\right) - 2^{\frac{1}{4}}\cos\left(\sqrt{2}\,\tau' + \frac{\pi}{4}\right)\right] - \frac{1}{\tau'}\left[\tfrac{5}{4}\cos\left(\tau' + \frac{\pi}{4}\right) + \frac{53(8^{\frac{1}{4}})}{16} \times \sin\left(\sqrt{2}\,\tau' + \frac{\pi}{4}\right) - (27)^{\frac{1}{4}}\cos\left(\sqrt{3}\,\tau' + \frac{\pi}{4}\right)\right]\right\}. \quad (4.3)$$

The terms decreasing as $\tau'^{-\frac{3}{2}}$ agree with those given by Goodman in Eq. (5.22) of Ref. 6; the terms involving $\tau'^{-\frac{5}{2}}$ are new.

For a range of intermediate times, Goodman felt that neither of the approximations available to him was adequate. Accordingly, he later (see Ref. 7) carried out calculations of $X_3(\tau')$ by a direct summation over a large number of normal modes. This is equivalent to a numerical evaluation of the integral

$$X_3(\tau') = \frac{1}{4\pi^3}\int_0^{2\pi}\!\!\int\!\!\int (3x)^{-\frac{1}{2}}(1-x) \times \sin(3x)^{\frac{1}{2}}\tau'\, d\theta_1\, d\theta_2\, d\theta_3, \quad (4.4)$$

which follows from our Eq. (3.1). As a check of accuracy, Goodman actually made two calculations in which the summations were carried over $M^3 = 1000$ and $M^3 = 1728$ normal modes. The accompanying Table V compares values from Table II of Ref. 7 with values computed from our Eq. (4.3). For $\tau' = 20$, 25, and 30, agreement to 1% or better is obtained. For $\tau' = 35$ and 40, discrepancies between results for $M = 10$ and $M = 12$ indicate a failure of Goodman's method, as he has pointed out. This is expected at sufficiently long times since calculations using a limited number of modes cannot reproduce accurately the effects of the singular points in the frequency spectrum. It is interesting to note that at $\tau' = 35$ the result for $M = 12$ is still good.

TABLE V. Values of Goodman's function $X_3(\tau')$.

	From Eq. (4.3)			
	Main terms		Goodman's values	
τ	only	All terms	$M = 10$	$M = 12$
20	0.0307	0.0324	0.032099	0.032099
25	0.00975	0.0120	0.011942	0.011943
30	−0.01257	−0.01235	−0.012409	−0.012293
35	−0.0151	−0.0150	−0.018594	−0.015004
40	−0.0029	−0.0030	−0.027946	−0.003576

ACKNOWLEDGMENTS

This study was begun in connection with research supported by the Aerospace Research Laboratories, U.S. Office of Aerospace Research, Wright-Patterson Air Force Base, Ohio. The author has appreciated this support as well as the interest of associates at Battelle which made possible a continuation of the work. He is also indebted to Dr. I. Adawi and Dr. E. J. Drauglis for helpful discussions and suggestions during the course of the work, and to Dr. A. C. Beer for comments on the manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Theoretical Study of Finite Dielectric-Coated Cylindrical Antenna*

CHUNG-YU TING
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts

(Received 23 March 1967)

A finite cylindrical antenna which is imbedded in a concentric dielectric rod has been investigated by employing a rigorous formulation. When the antenna is relatively short, a numerical method is used; when the antenna is long, the Wiener–Hopf technique is applied. In both cases the input admittance and the current distribution are obtained. It is found that the input conductances are larger than for the corresponding free-space antennas, the field patterns tend to be more broadside and, as the antenna gets longer and longer, the locus of the input admittance becomes a circle instead of converging to one point as it does for a bare cylindrical antenna. The first method is applicable regardless of the thickness of the antenna and the dielectric rod; the second method can be applied only to a sufficiently long antenna. The minimum length is determined by the thickness of the dielectric rod. This study is limited to thin antenna in rather thick dielectric cylinders. However, the dielectric rod is still not thick enough to support a transverse magnetic (T.M.) mode.

I. INTRODUCTION

In a cylindrical dielectric-coated antenna with infinite length,[1] the current can be separated into two parts: the radiation current which is associated with the radiation field, and the transmission current which is associated with the Goubau surface wave. The finite dielectric-coated antenna was first discussed by Wu[2] who showed that, when both the coating and the antenna itself are very thin, the current distribution differs very little from that of a thin bare dipole in free space and can be expressed in a form equivalent to that for a thin dipole with slightly modified radius and with a surface impedance. As the dielectric coating becomes thicker and thicker, changes are to be expected. For a very thick dielectric rod, the current in the antenna should behave more or less like that in a homogeneous infinite dielectric medium. However, due to the complexity of Green's function, an exact solution is very difficult to obtain.

In this study, an exact integral equation for the current in a finite dipole in an infinitely long dielectric rod was formulated and solved by a numerical method. The accuracy depends on the number of points taken and the accumulated round-off error. For a reasonable number of points, the results show excellent agreement with experiments; they are consistent with the prediction made from the infinite antenna. That is, when the dielectric layer is thick the current is dominated by the transmission current.[1] In principle, the method can be applied to an antenna of arbitrary length. However, due to the restricted number of storage locations available in a computer, it is limited to the relatively short dielectric-coated cylindrical antenna. To overcome this difficulty a new method is developed for a long dielectric-coated cylindrical antenna. This makes use of the fact[1] that the radiation current decays very quickly and can be neglected at the end of a long dielectric-coated antenna when compared with the transmission current. If the reflection coefficient of the transmission current is then found, the characteristics of the antenna can be determined.

II. FORMULATION OF THE INTEGRAL EQUATION

One way to formulate the integral equation for the current in a finite antenna is to derive Green's function first. Suppose there is a ring delta-function current source with radius a, oriented in the z direction inside a concentric dielectric rod with radius b as shown in Fig. 1(a). Rotational symmetry is assumed, and only

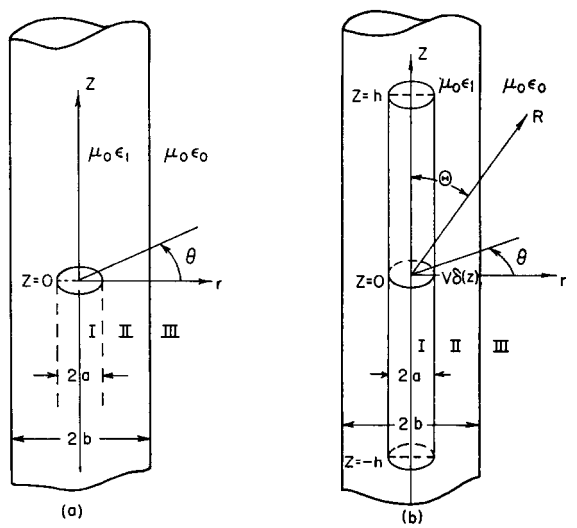


FIG. 1. Schematic diagrams of a ring delta source (a) and a finite dipole (b) in an infinite dielectric rod.

* This work was supported in part by the Office of Naval Research under Contract Nonr-1866 (32), and by the Division of Engineering and Applied Physics, Harvard University.

[1] C. Y. Ting, Radio Sci. 2, 325 (1967).

[2] T. T. Wu, J. Math. Phys. 2, 550 (1961).

the z component of the vector potential, $\vec{G} = G\hat{z}$, is excited. The vector potential G in the dielectric medium and in free space resulting from this delta-function source satisfies the following wave equations:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{\partial^2 G}{\partial z^2} + k_1^2 G = -\frac{\mu_0}{2\pi r}\delta(z)\delta(r-a), \quad 0 < r < b, \quad (1)$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{\partial^2 G}{\partial z^2} + k_0^2 G = 0, \quad b < r < \infty, \quad (2)$$

where $k_1 = \omega(\mu_0\epsilon_1)^{\frac{1}{2}}$, $k_0 = \omega(\mu_0\epsilon_0)^{\frac{1}{2}}$, μ_0 and ϵ_0 are the free-space permeability and dielectric constant, and $\epsilon_1 = \epsilon_r\epsilon_0$ is the dielectric constant in the dielectric medium. The time dependence is $e^{-i\omega t}$. Let the Fourier transforms of (1) and (2) be taken according to the relations

$$\bar{F}(k) = \int_{-\infty}^{\infty} F(z)e^{ikz}\,dz, \quad (3a)$$

$$F(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \bar{F}(k)e^{-ikz}\,dk. \quad (3b)$$

The Fourier-transformed solutions of (1) and (2) are given by

$$\bar{G}_1(k, r) = C_1 J_0(\xi r), \quad 0 < r \leq a, \quad (4)$$

$$\bar{G}_2(k, r) = C_2 J_0(\xi r) + C_3 Y_0(\xi r), \quad a \leq r < b, \quad (5)$$

$$\bar{G}_3(k, r) = C_4 H_0(\phi r), \quad b < r < \infty, \quad (6)$$

where $\xi = (k_1^2 - k^2)^{\frac{1}{2}}$, $\phi = (k_0^2 - k^2)^{\frac{1}{2}}$. By means of the boundary conditions: (a) tangential electric field continuous at $r = a$, (b) tangential magnetic field discontinuous at $r = a$ by the delta-function current source, (c) tangential electric field continuous at $r = b$, and (d) tangential magnetic field continuous at $r = b$, the constants C_1, C_2, C_3, and C_4 can be evaluated. Green's function in each region may also be found. In region I, where $0 < r \leq a$, it is

$$\bar{G}_1(k,r) = -\frac{\mu_0 J_0(\xi r)}{4D(k)} \times \{[\epsilon_r\phi Y_1(\xi b)H_0^{(1)}(\phi b) - \xi Y_0(\xi b)H_1^{(1)}(\phi b)]J_0(\xi a) + [\xi J_0(\xi b)H_1^{(1)}(\phi b) - \epsilon_r\phi J_1(\xi b)H_0^{(1)}(\phi b)]Y_0(\xi a)\}. \quad (7)$$

In region II, where $a \leq r < b$, it is

$$\bar{G}_2(k, r) = -\frac{\mu_0 J_0(\xi a)}{4D(k)} \times \{[\epsilon_r\phi Y_1(\xi b)H_0^{(1)}(\phi b) - \xi Y_0(\xi b)H_1^{(1)}(\phi b)]J_0(\xi r) + [\xi J_0(\xi b)H_1^{(1)}(\phi b) - \epsilon_r\phi J_1(\xi b)H_0^{(1)}(\phi b)]Y_0(\xi r)\}. \quad (8)$$

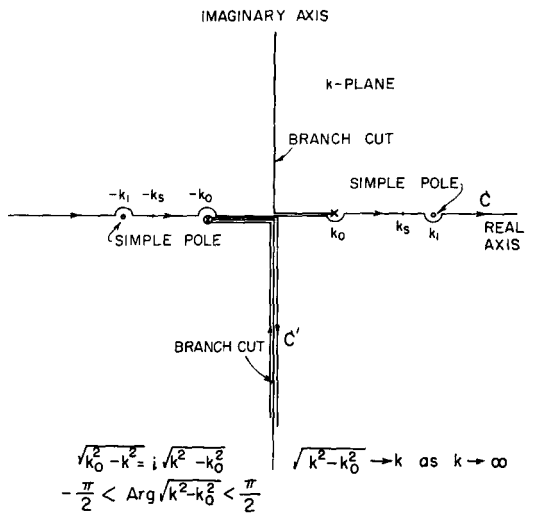


FIG. 2. Integration paths C, C' and singularities on k plane.

In region III, where $b < r < \infty$, it is

$$\bar{G}_3(k, r) = \frac{\xi\mu_0\epsilon_0 J_0(\xi a)H_0^{(1)}(\phi r)}{2\pi\phi b D(k)}, \quad (9)$$

where

$$D(k) = \xi J_0(\xi b)H_1^{(1)}(\phi b) - \epsilon_r\phi J_1(\xi b)H_0^{(1)}(\phi b). \quad (10)$$

In order to invert (8) from the k domain into the real z domain, the singularities of (8) on the complex k plane must be carefully investigated. The only two branch points are at $k = \pm k_0$. Points at $k = \pm k_1$ are not branch points, but are two simple poles. This can be recognized easily by employing a small-argument expansion for the Bessel functions with arguments ξa and ξb in (8). The leading term is

$$\lim_{k\to\pm k_1} \bar{G}_2(k, a) = \lim_{k\to\pm k_1} \frac{\mu_0\epsilon_r\phi H_0^{(1)}(\phi b)}{\pi b(k_1^2 - k^2)[2H_1^{(1)}(\phi b) - \epsilon_r\phi b H_0^{(1)}(\phi b)]}. \quad (11)$$

The numerator of (8) is the characteristic equation of a Goubau line; therefore, (8) has two zeros which are designated at $\pm k_s$. Other poles can be found by locating the zeros of the denominator. Note that $D(k) = 0$ is the characteristic equation of a dielectric waveguide. If the branch cuts are drawn in the manner shown in Fig. 2 and the same sequence of steps is followed that is described in two papers,[1,3] similar conclusions can be drawn. They are:

(a) There is no pole on the real axis in the range $k_1 < |k| < \infty$;

[3] C. Y. Ting, "A Theoretical Study of Dielectric-Coated Cylindrical Antenna," Cruft TR 506, Harvard University, 1966.

(b) There is no pole on the real axis in the range $k_0 < |k| < k_1$ for $(k_1^2 - k_0^2)^{\frac{1}{2}}b < 2.405$;

(c) There is no pole on either side of the branch cuts;

(d) There is no pole in the domain of very large $|k|$ including infinite;

(e) By taking the limit as $k_1 \to k_0$, it can be proved that there is no pole on the finite complex k plane.

With this information, and for $z > 0$, the Fourier-inverse contour can be closed in the lower half-plane and Green's function expressed as follows:

$$
\begin{aligned}
G_2(z, a) &= \frac{1}{2\pi}\int_c \bar{G}_2(k, a)e^{-ikz}\,dk \\
&= i\frac{\mu_0\epsilon_r(k_1^2 - k_0^2)^{\frac{1}{2}}K_0[(k_1^2 - k_0^2)^{\frac{1}{2}}b]e^{ik_1|z|}}{2\pi bk_1\{2K_1[(k_1^2 - k_0^2)^{\frac{1}{2}}b] + \epsilon_1(k_1^2 - k_0^2)^{\frac{1}{2}}bK_0[(k_1^2 - k_0^2)^{\frac{1}{2}}b]\}} \\
&\quad + i\int_0^{k_0} \frac{\mu_0\epsilon_r[J_0(Qa)]^2e^{ixz}\,dx}{\pi^3b^2\{[QJ_0(Qb)J_1(Pb) - \epsilon_rPJ_1(Qb)J_0(Pb)]^2 + [QJ_0(Qb)Y_1(Pb) - \epsilon_rPJ_1(Qb)Y_0(Pb)]^2\}} \\
&\quad + \int_0^{\infty} \frac{\mu_0\epsilon_r[J_0(Va)]^2e^{-vz}\,dy}{\pi^3b^2\{[VJ_0(Vb)J_1(Ub) - \epsilon_rUJ_1(Vb)J_0(Ub)]^2 + [VJ_0(Vb)Y_1(Ub) - \epsilon_rUJ_1(Vb)Y_0(Ub)]^2\}},
\end{aligned}
\tag{12}
$$

where

$$Q = (k_1^2 - x^2)^{\frac{1}{2}}, \quad P = (k_0^2 - x^2)^{\frac{1}{2}}, \quad U = (k_0^2 + y^2)^{\frac{1}{2}}, \quad V = (k_1^2 + y^2)^{\frac{1}{2}}.$$

The first term comes from the residue at $-k_1$; the second and the third terms come from the branch cut, as shown in Fig. 2.

Once the Green's function is known, it is possible to proceed to analyze the finite antenna. As shown in Fig. 1(b), a finite tubular dipole is imbedded in an infinite dielectric rod with a delta generator at $z = 0$. From the condition that the tangential electric field vanish on the surface of the antenna, Hállen's integral equation is obtained. It is

$$
\begin{aligned}
\frac{4\pi}{\mu_0}A_z &= \int_{-h}^{h} I(z')\kappa(z - z')\,dz' \\
&= \frac{i4\pi}{\zeta_1}\left[C\cos k_1z + \frac{V}{2}\sin k_1|z|\right],
\end{aligned}
\tag{13}
$$

where $\zeta_1 = (\mu_0/\epsilon_1)^{\frac{1}{2}}$, C is a constant to be determined by the condition that the current vanish at $z = \pm h$, and

$$
\begin{aligned}
\kappa(z - z') &= \frac{2}{\mu_0}\int_{-\infty}^{\infty} \bar{G}_2(k, a)e^{-ik(z-z')}\,dk \\
&= \frac{4\pi}{\mu_0}G_2(z - z', a).
\end{aligned}
\tag{14}
$$

The contour C of the Fourier-inverse integration can go either above $-k_1$ and below k_1 as shown in Fig. 2, or the other way around. The answer is the same. This has been discussed in an earlier paper.[3]

III. A NUMERICAL METHOD

Equation (13) is an exact integral equation for the model shown in Fig. 1(b). When $k_1b \ll 1$, the small-argument expansion of the Bessel functions can be used to approximate the kernel. This is discussed by Wu,[2] who predicted that the current distribution when both the antenna and the coating are very thin should be close to that in a free-space dipole. On the other hand, it is interesting to know the change in the current distribution when the coating gets thicker. Since no simple approximation can be made for the kernel, it is difficult to obtain even an approximate solution. The method employed here is a numerical one given by Young.[4,5] In his two papers, first integrals of the product of two functions $f(x)$ and $g(x)$ are expressed in the form

$$\int_a^b f(x)g(x)\,dx = \sum_{r=1}^{n}\gamma_r f(x_r) + R, \tag{15}$$

where $x_1, x_2, \cdots, x_n$ are the n abscissas with which are associated weights $\gamma_1, \gamma_2, \cdots, \gamma_n$, and R is a correction term. It has been shown that by expanding $f(x)$ in a Taylor's series about the midpoint of the interval between a and b and by equally spacing the n abscissas, i.e., $x_n - x_{n-1} = x_{n-1} - x_{n-2} = \cdots = x_2 - x_1 = t$, the γ's can be expressed in a matrix form. For instance, for $n = 3$, they are

$$\begin{bmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\end{bmatrix} = \frac{1}{2}\begin{bmatrix}0 & -1 & 1\\ 2 & 0 & -2\\ 0 & 1 & 1\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\end{bmatrix}, \tag{16}$$

where

$$\mu_s = \frac{1}{t^s}\int_a^b (x - x_2)^s g(x)\,dx, \quad s = 0, 1, 2.$$

[4] Andrew Young, Proc. Roy. Soc. (London) **A224**, 552 (1954).
[5] Andrew Young, Proc. Roy. Soc. (London) **A224**, 561 (1954).

The remainder term R is proportional to the fourth derivative of $f(x)$ within the interval. The next step is to apply the approximate product integration (15) to the numerical solution of integral equations. To begin with, because of the symmetry of the current $I(z) = I(-z)$, Eq. (13) may be rewritten as

$$\int_0^h I(z')[\kappa(z - z') + \kappa(z + z')]\,dz' = \frac{i4\pi}{\zeta_1}\left(C\cos k_1 z + \frac{V}{2}\sin k_1|z|\right). \quad (17)$$

The interval $(0, h)$ may be divided into l subintervals. Within each subinterval, an approximation of the type (15) is used. That is, by expanding the current $I(z)$ in each subinterval into a quadratic form (or $n = 3$) about the midpoint of the subinterval, the right-hand side of (17) becomes

$$\begin{aligned}\int_0^h I(z')[\kappa(z - z') + \kappa(z + z')]\,dz' &= \left\{\int_0^{2t} + \int_{2t}^{4t} + \int_{4t}^{6t} + \cdots + \int_{2(l-1)t}^{2lt}\right\} \\ &\quad \times I(z')[\kappa(z - z') + \kappa(z + z')]\,dz' \\ &= \sum_{j=1}^{l}\{\gamma_1^j(z)I[(2j - 2)t] \\ &\quad + \gamma_2^j(z)I[(2j - 1)t] + \gamma_3^j(z)I(2jt)\}, \quad (18)\end{aligned}$$

where $t = h/2l$. By defining

$$\begin{aligned}\mu_n(mt) &= \frac{1}{t^{n-1}}\int_{-t}^{t} z'^{n-1}\kappa(mt - z')\,dz' \\ &= (-1)^{n-1}\mu_n(-mt), \quad (19)\end{aligned}$$

with $n = 1, 2, 3$, and $m = 0, 1, \cdots, 4l - 1$, all the γ's in (18) can be expressed in terms of the μ's. According to the relation (16), they are

$$\begin{aligned}\gamma_1^j(z) &= \tfrac{1}{2}\{-\mu_2[z - (2j - 1)t] + \mu_2[z + (2j - 1)t] \\ &\quad + \mu_3[z - (2j - 1)t] + \mu_3[z + (2j - 1)t]\}, \quad (20a)\end{aligned}$$

$$\begin{aligned}\gamma_2^j(z) &= \mu_1[z - (2j - 1)t] + \mu_1[z + (2j - 1)t] \\ &\quad - \mu_3[z - (2j - 1)t] - \mu_3[z + (2j - 1)t], \quad (20b)\end{aligned}$$

$$\begin{aligned}\gamma_3^j(z) &= \tfrac{1}{2}\{\mu_2[z - (2j - 1)t] - \mu_2[z + (2j - 1)t] \\ &\quad + \mu_3[z - (2j - 1)t] + \mu_3[z + (2j - 1)t]\}. \quad (20c)\end{aligned}$$

Now let $z = mt$ in (17) together with (18), and let m change from 0 to $2l$. In this manner a set of $2l + 1$ linear equations are generated with $2l + 1$ unknowns. Since the current vanishes at $z = 2lt$, there are only $2l$ unknowns for the current plus an unknown constant C. In matrix notation,

$$[A][I] = [G], \quad (21)$$

where $[I]$ and $[G]$ are $2l + 1$ by 1 column matrices. Their transposed forms are

$$[I]^T = [I(0), I(t), I(2t), \cdots, I(2lt - t), C], \quad (22)$$

$$[G]^T = [0, \sin(k_1 t), \sin(2k_1 t), \cdots, \sin(2lk_1 t)]. \quad (23)$$

$[A]$ is a $2l + 1$ by $2l + 1$ square matrix whose elements $A_{p,q}$ are given by

$$q = 1: \qquad A_{p,1} = \gamma_1^1(pt), \quad (24a)$$

$$q = \text{even number}: \qquad A_{p,q} = \gamma_3^{q/2}(pt) + \gamma_1^{q/2+1}(pt), \quad (24b)$$

$$q = \text{odd number}: \qquad A_{p,q} = \gamma_2^{(q+1)/2}(pt), \quad (24c)$$

$$q = 2l + 1: \qquad A_{p,2l+1} = (i4\pi/\zeta_1)\cos(pk_1 t). \quad (24d)$$

The γ's are given by (20); they are all complex quantities. If the square matrix $[A]$ is inverted, the numerical value of the current and constant C are immediately obtained. Thus,

$$[I] = [A]^{-1}[G]. \quad (25)$$

It is noted that all of the constants μ given by (19) are in double-integral form. By interchanging the order of integration, one of them can be carried out easily and the other is left for the computer. Explicit formulas for the μ's can be found in the appendix of Ting's paper.[3]

IV. NUMERICAL RESULTS

Computations have been made with an IBM 7094 computer. Since many integrations of Bessel functions are involved in generating the constants μ and then the matrix elements $A_{p,q}$, a considerable length of time is required in order to achieve one curve of the current distribution. Fortunately, a way has been found which can save much computing time and give a number of curves simultaneously. Beginning with the longest antenna to be investigated, the length h is divided into l subdivisions as described before, and the $2l + 1$ by $2l + 1$ matrix $[A]$ is formed. Then, for shorter antennas with length $h - (h/l)n$, $n = 1, 2, \cdots, l - 1$, the matrix elements $A_{p,q}$ in each case are precisely the same as before except in the last column, which should always retain cosine terms. The only significant change is that the order of the matrix shrinks by two each time n is increased by 1. Consequently, once the $2l + 1$ by $2l + 1$ matrix is formed by redefining the last column each time, the current distributions for l different lengths are obtained almost simultaneously.

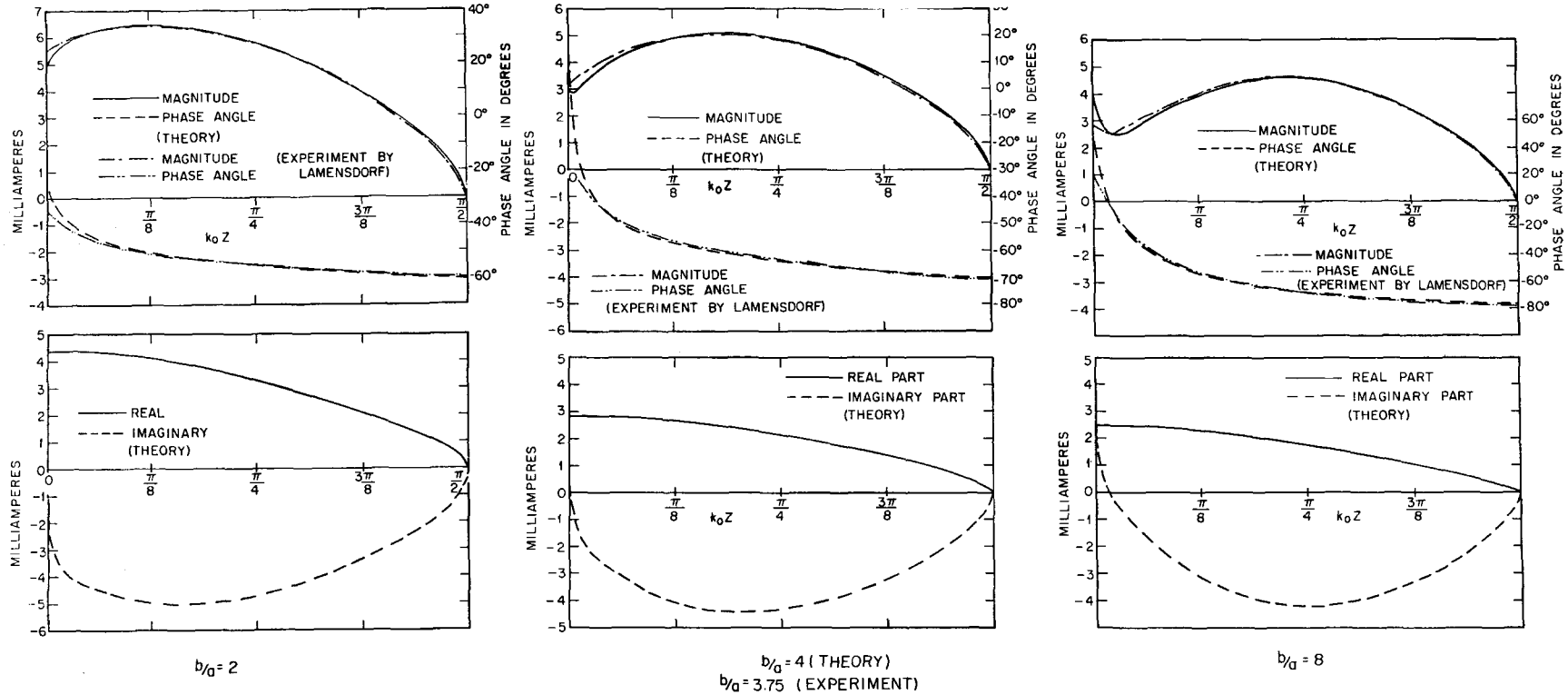


FIG. 3. Theoretical current distribution $\epsilon_r = 3.0$, $k_0a = 0.04$, $k_0h = \pi/2$, and experimental current distribution $\epsilon_r = 3.2$, $k_0a = 0.04$, $k_0h = \pi/2$.

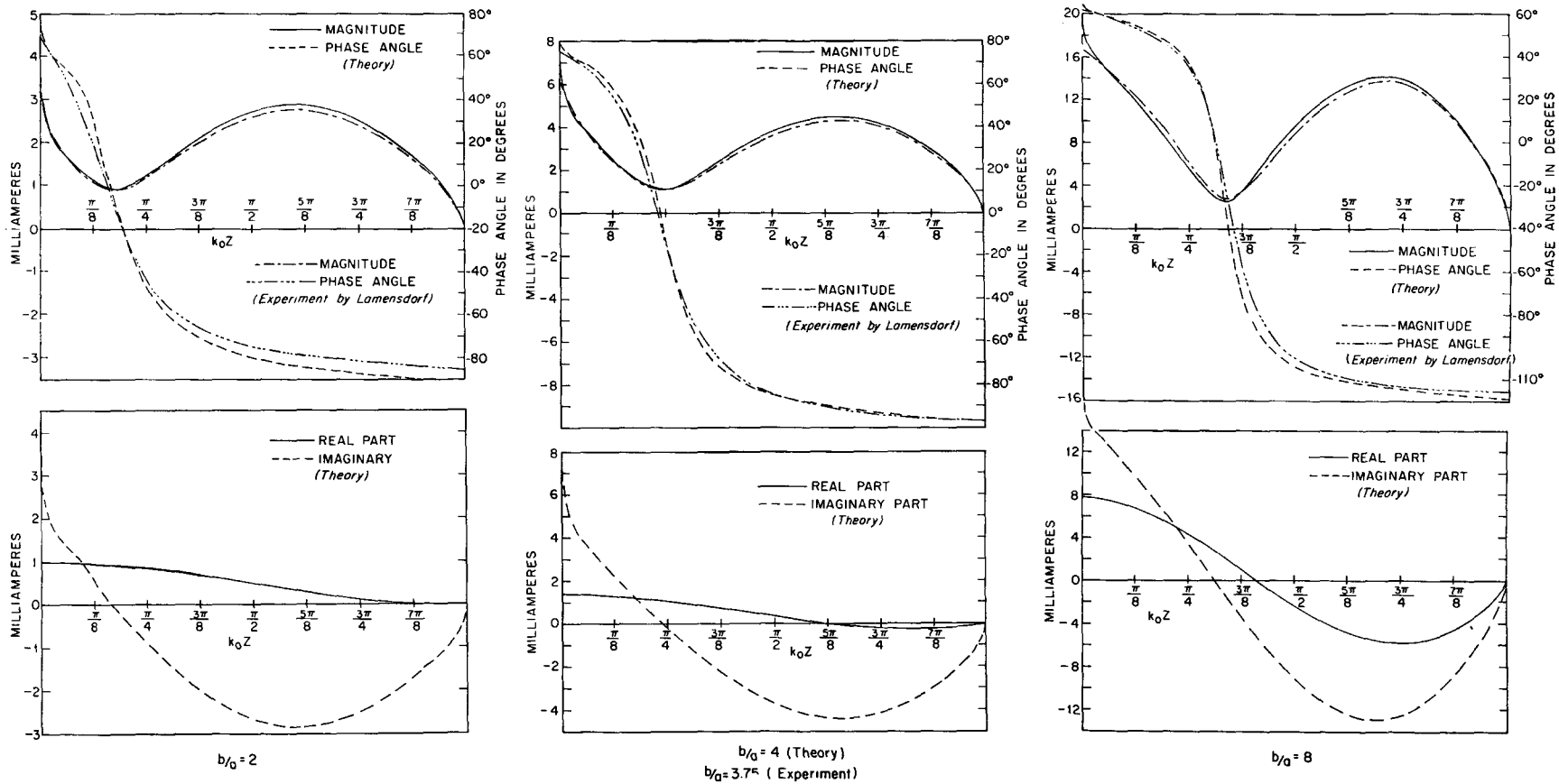


FIG. 4. Theoretical current distribution $\epsilon_r = 3.0$, $k_0a = 0.04$, $k_0h = \pi$, and experimental current distribution $\epsilon_r = 3.2$, $k_0a = 0.04$, $k_0h = \pi$.

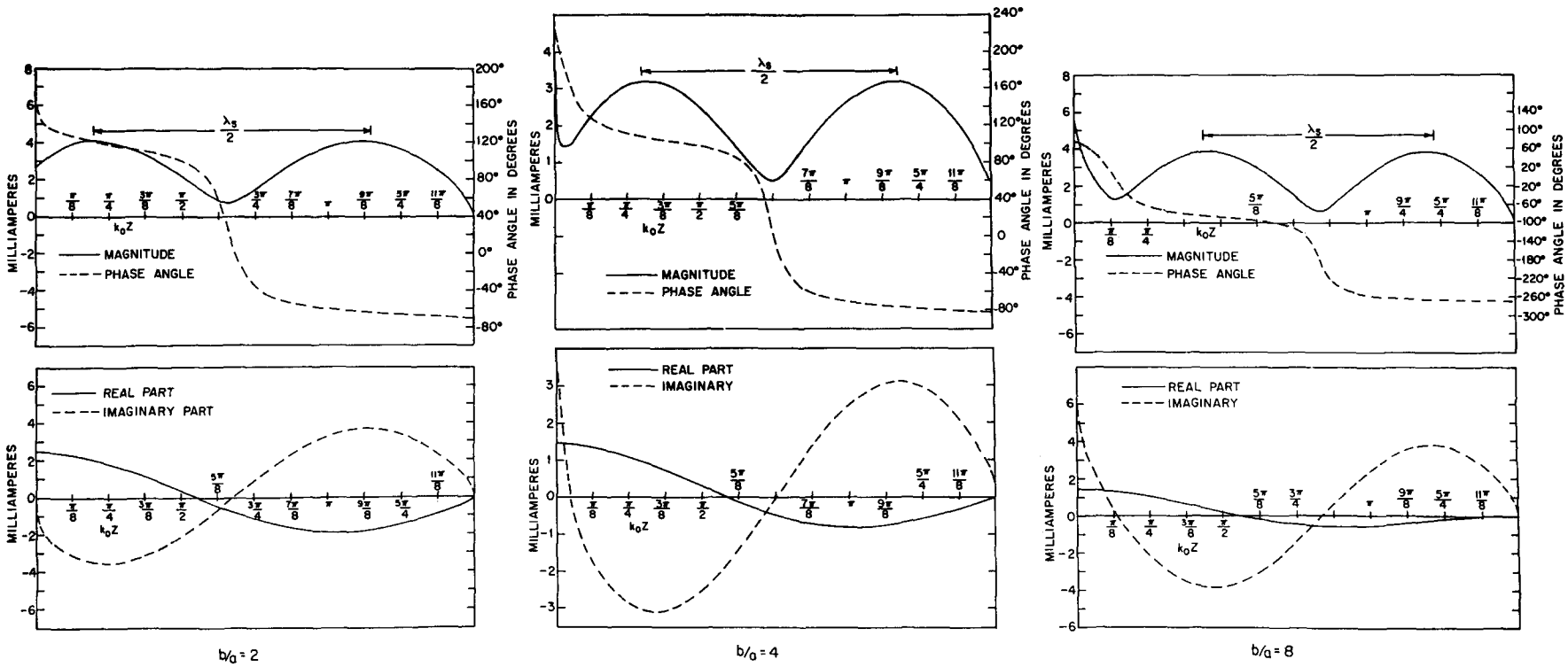


FIG. 5. Theoretical current distribution $\epsilon_r = 3.0$, $k_0a = 0.04$, $k_0h = \frac{13}{24}\pi$.

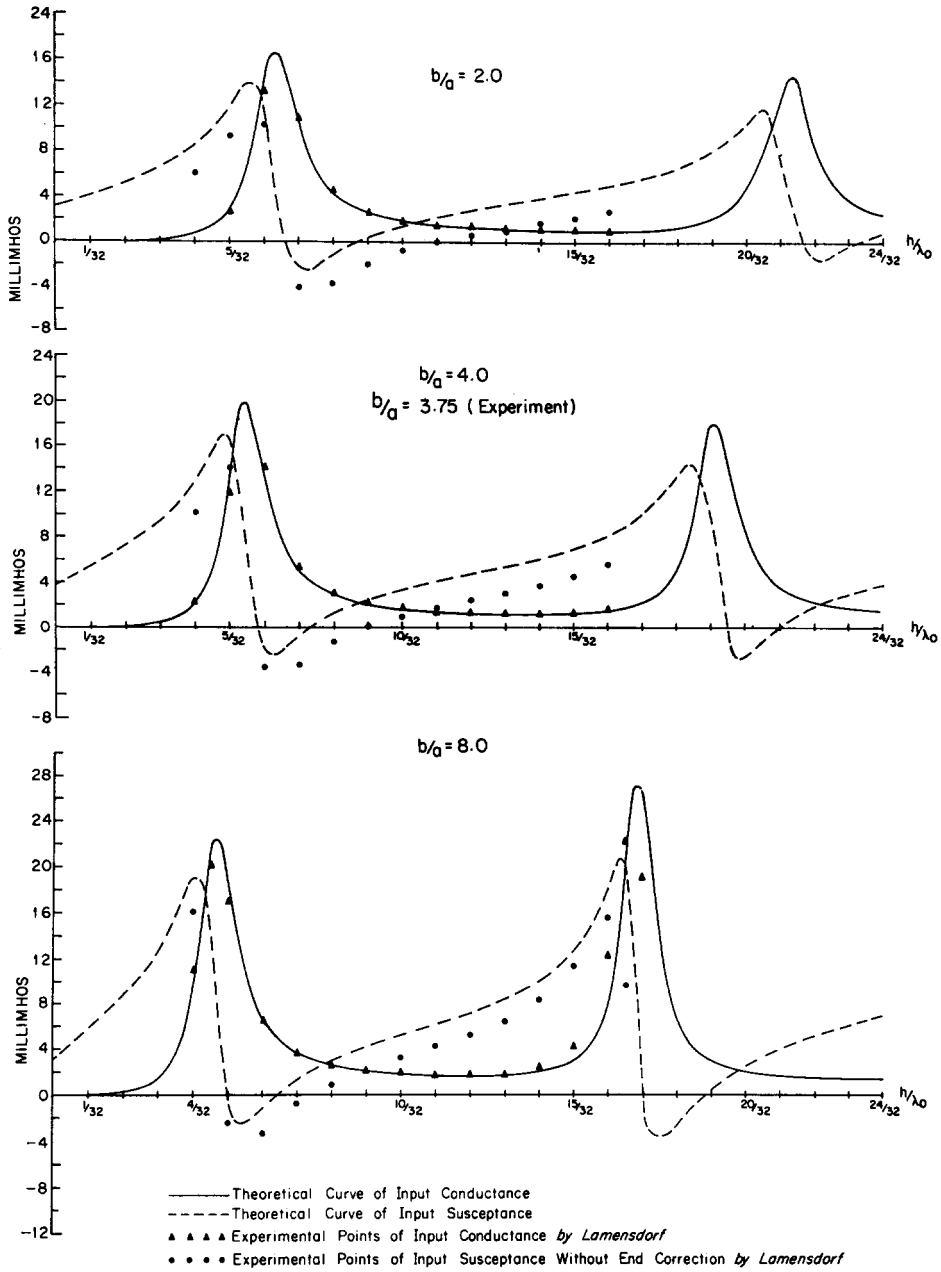


FIG. 6. Theoretical input admittance $\epsilon_r = 3.0, k_0a = 0.04$ and experimental input admittance $\epsilon_r = 3.2, k_0a = 0.04$.

Some typical results have been obtained for $\epsilon_r = 3.00$, $k_0a = 0.04$, $k_0h = 3\pi/2$, $l = 24$, and three different thicknesses of the dielectric coating, namely, $b/a = 2$, 4, and 8. For each case the current distributions for 24 different lengths have been obtained, of which only $k_0h = \frac{1}{2}\pi$, π, $\frac{3}{2}\pi$ are shown in Figs. 3–5. Also shown in Figs. 3 and 4 are the experimental curves by Lamensdorf.[6] They provide an excellent check on the theory. Note that when the antenna becomes longer, beautiful standing waves are formed along them as shown in Fig. 5. The wavelengths are close to the surface wavelength, especially when b/a is as large as predicted in Ting's paper.[1]

Another interesting part of the results is the input admittance. Since calculations are based upon the assumption that the voltage across the delta generator is 1, the real part of the current at $z = 0$ is the input conductance, and the imaginary part at $z = 0$ is the input susceptance. Figure 6 shows the curves of the input admittances as the length of the antenna changes.

[6] The experimental data are obtained from D. Lamensdorf, who used a dielectric sleeve which was much longer than the antenna itself, and in close contact with it. Experimental results show that the length of the sleeve is not important, so long as it is much longer than the antenna. (See Ref. 9.)

Experimental points by Lamensdorf[6] are superimposed on them. It is noted that the input conductances agree very well, but not the input susceptances. The reason is simple. Since a delta generator at $z = 0$ was assumed, the input susceptances should theoretically be infinite at $z = 0$. Therefore, the more points that are taken in the calculation of the current distribution, the higher the input susceptance is. Nevertheless, the general shape of the input susceptance curve, obtained in the manner previously described, is still good. If one point is calibrated the rest are known.

Comparisons with free-space dipoles[7] are also interesting. In general, the input conductances are larger for dielectric-coated antennas and the input susceptances are more inductive. This is because the antenna is effectively thicker in the dielectric rod than in free space. The resonant and antiresonant lengths are shorter for dielectric-coated antennas. In other words, the effective length of an antenna in a dielectric rod is greater than that in free space as was anticipated before performing any calculations.

There is an interesting characteristic for the case $b/a = 8$. Owing to the interaction of the two kinds of standing wave, i.e., the radiation and the transmission formed along the antenna, the second resonant peak is greater than the first resonant peak. This has also been verified by experiment.

V. FIELD PATTERNS

Once the current is known, the far-field pattern can be calculated easily. The transformed vector potential in region III, $\bar{G}_3(k, r)$, resulting from a ring delta source,is given by (9), so that the θ component of the magnetic field in this region resulting from this ring delta source can be expressed as follows:

$$-\frac{\partial \bar{G}_3(k, r)}{\partial r} = \frac{\xi\mu_0 J_0(\xi a)H_1^{(1)}(\phi r)}{2\pi b D(k)}. \tag{26}$$

The actual magnetic field due to this ring delta source is the inverse Fourier transform of (26). By superposition, the total magnetic field $B_{3\theta}$ due to current in the whole antenna is

$$B_{3\theta} = \int_{-h}^{h} I(z')\,dz' \times \int_c \frac{\xi\mu_0}{4\pi^2 b D(k)} J_0(\xi a)H_1^{(1)}(\phi r)e^{-ik(z-z')}\,dk. \tag{27}$$

It is now convenient to change to spherical coordinates (R, Θ, Φ), with $z = R\cos\Theta$, $r = R\sin\Theta$ in (27). Then, as $R \to \infty$, (27) becomes

$$\lim_{R\to\infty} B_{3\theta} = \lim_{R\to\infty} \int_{-h}^{h} I(z')\,dz' \int_c \frac{\xi\mu_0 J_0(\xi a)e^{ikz'}e^{-\frac{3}{4}i\pi}}{4\pi^2 b D(k)} \times \left(\frac{2}{\pi\phi R\sin\Theta}\right)^{\frac{1}{2}} e^{iR(\phi\sin\Theta - k\cos\Theta)}\,dk. \tag{28}$$

The method of steepest descents applies. The evaluation of the integral (28) at the saddle point $k = -k_0\cos\Theta$ gives

$$\lim_{k\to\infty} B_{3\theta} = \frac{e^{ik_0R}}{2\pi R}\left[\frac{-\xi\mu_0 J_0(\xi a)}{D(k)}\right] \int_{-h}^{h} I(z')e^{-ik_0z'\cos\Theta}\,dz', \tag{29}$$

where the square bracket is evaluated at $k = -k_0\cos\Theta$. The Poynting vector in the far field is, by definition,

$$S = (2\mu_0)^{-1}E_3 \times B_3^* = [2\mu_0(\mu_0\epsilon_0)^{\frac{1}{2}}]^{-1}\,|B_{3\theta}|^2, \tag{30}$$

if the field factor is defined as $(4\pi k^2 S)^{\frac{1}{2}}$ which can be expressed as follows:

$$F_r(\Theta) = (4\sqrt{15}\,\pi R/\mu_0)\,|B_{3\theta}|. \tag{31}$$

To use the numerical solution to calculate the field factor, use is again made of the approximate method described in Sec. III of dividing the antenna into l subdivisions (same number as before). Within each subdivision the current is approximated by a quadratic form, and the integral involved in (29) can be put in the form (18). The final result is

$$F_r(\Theta) = \left|\frac{2\sqrt{15}\,\xi t J_0(\xi a)}{\pi[\xi b J_0(\xi b)H_1^{(1)}(\phi b) - \epsilon_r\phi b J_1(\xi b)H_0^{(1)}(\phi b)]}\right|_{k=-k_0\cos\Theta} \times \left|\sum_{j=1}^{l}[\gamma_1^j I(2jt - 2t) + \gamma_2^j I(2jt - t) + \gamma_3^j I(2jt)]\right|, \tag{32}$$

where $t = h/2l$ as before and

$$\gamma_1^j = -\mu_2(j) + \mu_3(j), \quad \gamma_2^j = 2\mu_1(j) - 2\mu_3(j), \quad \gamma_3^j = \mu_2(j) + \mu_3(j), \tag{33}$$

$$\mu_1(j) = \omega^{-1}\{\sin(2j\omega) - \sin[(2j-2)\omega]\}, \tag{34a}$$

$$\mu_2(j) = \omega^{-2}\{\cos(2j\omega) + \omega\sin(2j\omega) - \cos[(2j-2)\omega] + \omega\sin[(2j-2)\omega]\}, \tag{34b}$$

$$\mu_3(j) = \omega^{-3}\{2\omega\cos(2j\omega) + (\omega^2 - 2)\sin(2j\omega) + 2\omega\cos[(2j-2)\omega] - (\omega^2 - 2)\sin[(2j-2)\omega]\}, \tag{34c}$$

$$\omega = tk_0\cos\Theta. \tag{35}$$

[7] R. W. P. King, *The Theory of Linear Antenna* (Harvard University Press, Cambridge, Mass., 1956).

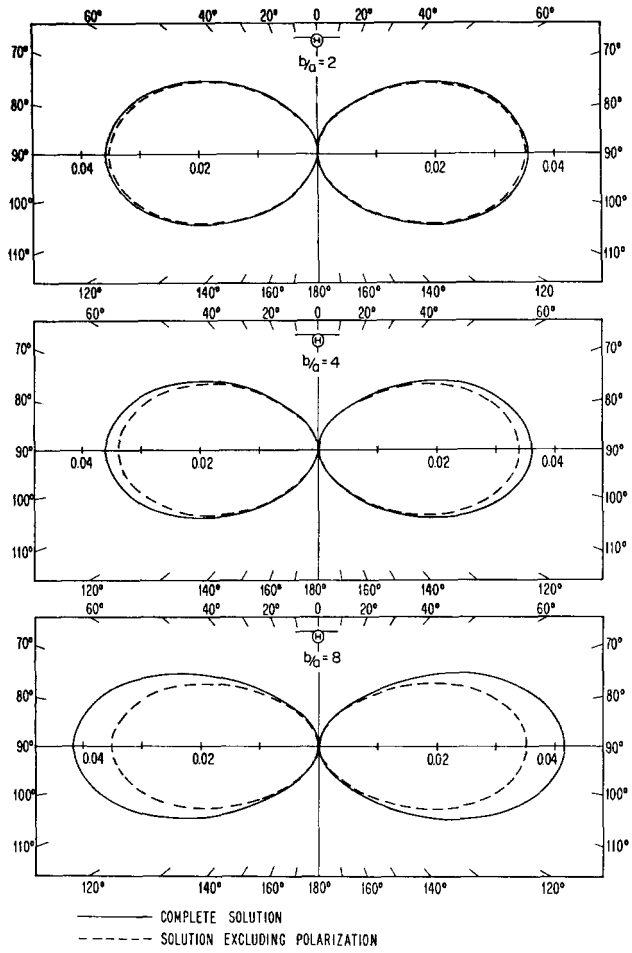


FIG. 7. Field patterns $\epsilon_r = 3.0$, $k_0a = 0.04$, $k_0h = \frac{3}{4}\pi$.

The use of the numerical value of $I(mt)$, $m = 0, 1, \cdots, 2l$, obtained from Sec. IV, in Eq. (32) yields the field pattern.

It is interesting to know the contribution from the conduction current on the conducting tube and the contribution from the polarization in the dielectric cylinder. The latter is excluded if the free-space Green's function is used; the far field can be expressed as

$$B_{3\theta}(R \to \infty) \sim \frac{-i\mu_0 k_0 \sin\Theta e^{ik_0R}}{4\pi R} \times \int_{-h}^{h} I(z')e^{-ik_0z'\cos\Theta}\,dz' \quad (36)$$

and the field factor is

$$F_r(\Theta) = \sqrt{15}\, k_0 t \sin\Theta \times \left| \sum_{j=1}^{l} [\gamma_1^j I(2jt - 2t) + \gamma_2^j I(2jt - t) + \gamma_3^j I(2jt)] \right|. \quad (37)$$

Numerical calculations have been carried out with the above two methods for $k_0h = \frac{3}{4}\pi$ and $b/a = 2$, 4, 8. Results are shown graphically in Fig. 7. The study of these curves shows that the field patterns obtained from (32) and (37) have the same shape but that their magnitudes differ somewhat, depending on the thickness of the dielectric cylinder. The field pattern of the dielectric-coated antenna has greater broadside characteristics than that of the free-space dipole, and this property becomes more prominent as the dielectric is made thicker. Although a part of the imaginary part of the current has a reversed sign, there is no minor lobe because the antenna is still shorter than one wavelength in free space. The contribution to the field by the time-varying polarization in the dielectric cylinder is very small compared with that by the current in the antenna itself [the difference between (32) and (37)], and this difference is roughly proportional to the thickness of the dielectric layer. If the dielectric layer is not extremely thick, as in cases previously discussed, the contribution by the polarization can be neglected for engineering purposes.

VI. LONG DIELECTRIC-COATED CYLINDRICAL ANTENNA

As the antenna gets longer, a larger computer is needed to solve the problem numerically and, also, the computing time becomes considerably longer. Fortunately, a new method has been developed with which the problem can be solved much more easily. It was shown in Ting's paper[1] that, in an infinite cylindrical antenna with a reasonably thick dielectric coating, the radiation current excited by a delta generator is much smaller than the transmission current except when very close to the generator. Also, its rate of decay is greater, initially, than an exponential rate; it becomes $1/z^2$ asymptotically. Therefore, if the antenna is long enough so that the radiation current can be neglected at the ends and if the reflection coefficient of the transmission current can be found, then the problem is solved.

In order to find the reflection coefficient of the transmission current, a model of Fig. 8 is considered.

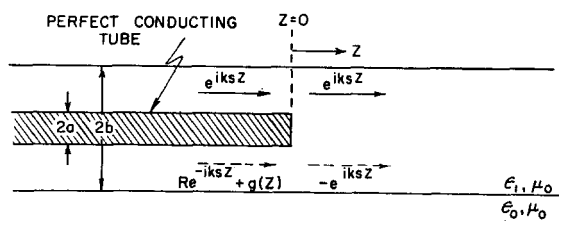


FIG. 8. A schematic diagram of a semi-infinite perfectly conducting tube in an infinite dielectric cylinder.

A semi-infinite perfectly conducting tube terminated at $z = 0$ and imbedded in an infinite concentric dielectric cylinder is used. Again the conducting tube has the radius a, and the dielectric cylinder has the radius b and the dielectric constant ϵ_1. A Wiener–Hopf method similar to the work of Levine and Schwinger[8] is used. Assume there is an incident transmission current $e^{+ik_s z}$ traveling from $z = -\infty$ toward $z = \infty$, where k_s is the Goubau surface wavenumber defined in the same manner as before. After the reflection of the incident current at $z = 0$, the scattered current can be expressed as follows:

$$I(z) = \begin{cases} R\,e^{-ik_s z} + g(z), & z \leq 0, \\ -e^{ik_s z}, & z \geq 0, \end{cases} \tag{38}$$

where R is the reflection coefficient of the transmission current, $g(z)$ is some unknown function which consists of the radiation current on the outside, and the attenuated waveguide-mode current on the inside of the tube, generated by the reflection of the incident transmission current. Both of these are assumed to suffer rapid attenuation; note that $-e^{ik_s z}$ cancels out the incident current $e^{ik_s z}$. Since there is no conducting tube on the side $z > 0$, there is no conduction current there. The boundary condition for a vanishing current at $z = 0$ requires $R = -[1 + g(0)]$. The Fourier transform of (38) is

$$\bar{I}(k) = \left[\frac{1}{i(k + k_s)}\right]_+ + \left[\frac{R}{i(k - k_s)} + \bar{G}_-(k)\right]_-, \tag{39}$$

where the plus and minus subscripts indicate the plus and minus functions defined by

$$\bar{F}_+(k) = \int_0^\infty F(z)\,e^{ikz}\,dz, \tag{40a}$$

$$\bar{F}_-(k) = \int_{-\infty}^0 F(z)\,e^{ikz}\,dz. \tag{40b}$$

$\bar{F}_+(k)$ is analytic in the upper half k plane and $\bar{F}_-(k)$ is analytic in the lower half k plane. They have a common analytic region which shrinks to the real axis.

The Fourier-transformed Green's function $\bar{K}(k)$ for the z component of the electric field at $r = a$ is

$$\bar{K}(k) = (i\omega\xi^2/k_1^2)\bar{G}_2(k, a), \tag{41}$$

where $\bar{G}_2(k, a)$ is given by (8). The singularities of $\bar{K}(k)$ are the same as in $\bar{G}_2(k, a)$; this is discussed in Sec. II. Note that two poles at $k = \pm k_1$ have been canceled out.

From the boundary condition that requires the tangential electric field to vanish on the surface of the perfect conductor, a Wiener–Hopf-type integral equation is formulated as follows:

$$\int_{-\infty}^{\infty} I(z')K(z - z')\,dz' = \begin{cases} 0, & z < 0, \\ E(z), & z > 0, \end{cases} \tag{42}$$

where $E(z)$ is an unknown function of the tangential electric field for $z > 0$. The Fourier transformation of both sides of (42), with the assumption that $E(z)$ is Fourier integrable, gives

$$\left[\frac{1}{i(k + k_s)}\right]_+ + \left[\frac{R}{i(k - k_s)} + \bar{G}_-(k)\right]_- = \frac{\bar{E}_+(k)}{\bar{K}(k)}. \tag{43}$$

$1/\bar{K}(k)$ has two branch points at $\pm k_0$, two simple poles at $\pm k_s$, and an infinite number of simple poles on the imaginary axis which makes $J_0(\xi a) = 0$. Let $1/\bar{K}(k)$ be split into a product of a plus and a minus function as follows:

$$\frac{1}{\bar{K}(k)} = \left[\frac{\bar{P}_+(k)}{k + k_s}\right]_+ \left[\frac{\bar{Q}_-(k)}{k - k_s}\right]_-, \tag{44}$$

where the two poles at $\pm k_s$ have been separated out for simplicity. $\bar{P}_+(k)$, $\bar{Q}_-(k)$ are given by

$$\bar{P}_+(k) = \exp\left[\frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{\ln\,[(\lambda^2 - k_s^2)/\bar{K}(\lambda)]}{\lambda - k}\,d\lambda\right], \quad \text{Im}\,\lambda > 0, \tag{45}$$

$$\bar{Q}_-(k) = \exp\left[\frac{-1}{2\pi i}\int_{-\infty}^{\infty} \frac{\ln\,[(\lambda^2 - k_s^2)/\bar{K}(\lambda)]}{\lambda - k}\,d\lambda\right], \quad \text{Im}\,\lambda < 0. \tag{46}$$

With (44), (43) becomes

$$\frac{k - k_s}{i(k + k_s)\bar{Q}_-(k)} + \left[\frac{R}{i\bar{Q}_-(k)} + \frac{(k - k_s)\bar{G}_-(k)}{\bar{Q}_-(k)}\right]_- = \left[\frac{\bar{E}_+(k)\bar{P}_+(k)}{k + k_s}\right]_+. \tag{47}$$

Then, by splitting the first term of (47) into a sum of a plus and a minus function, one obtains

$$\left[\frac{k - k_s}{i(k + k_s)\bar{Q}_-(k)} + \frac{2k_s}{i(k + k_s)\bar{Q}_-(-k_s)}\right]_- + \left[\frac{R}{i\bar{Q}_-(k)} + \frac{(k - k_s)\bar{G}_-(k)}{\bar{Q}_-(k)}\right]_- = \left[\frac{\bar{E}_+(k)\bar{P}_+(k)}{k + k_s}\right]_+ + \left[\frac{2k_s}{i(k + k_s)\bar{Q}_-(-k_s)}\right]_+. \tag{48}$$

The left-hand side of (48) is a minus function, the right-hand side is a plus function; therefore they must be equal to an entire function. From an investigation of the asymptotic behavior of the function on the

[8] H. Levine and J. Schwinger, Phys. Rev. 73, 383 (1948).

right-hand side in (48), it is easy to prove that the entire function is zero. Since, as $|k| \to \infty$, $\bar{K}(k) \sim k$, and $\bar{K}_+(k)$ is an even function of k, so that $\bar{P}_+(k) = \bar{Q}_-(-k) \sim \sqrt{k}$, and $\bar{E}_+(k)$ is no worse than a constant, the right-hand side of (48) tends to zero. It follows that

$$\frac{R}{i(k-k_s)} + \bar{G}_-(k) = \frac{-1}{k-k_s}\left[\frac{k-k_s}{i(k+k_s)} + \frac{2k_s\bar{Q}_-(k)}{k(k+k_s)\bar{Q}_-(-k_s)}\right]. \quad (49)$$

When the inverse Fourier transforms are taken, the residue contributions resulting from the simple pole at $k = k_s$ should be equated on both sides of the equation. This gives

$$R = -\bar{Q}_-(k_s)/\bar{Q}_-(-k_s), \quad (50)$$

which is a simple expression of the reflection coefficient of the transmission current. $\bar{Q}_-(k_s)$ and $\bar{Q}_-(-k_s)$ can be evaluated with (46). After simplification has been made on (50), a final form is

$$R = -\exp\left[\frac{-2k_s}{\pi i} P\int_0^\infty \frac{\ln\left[(\lambda^2 - k_s^2)/\bar{K}(\lambda)\right]}{\lambda^2 - k_s^2} d\lambda\right], \quad (51)$$

where P indicates the principal value. R as given in (51) can be computed numerically; in general it is a complex quantity.

After R is known, it can be used readily in the analysis of the finite long antenna shown in Fig. 1(b). Since the antenna is assumed long, the radiation current can be neglected at the ends of the antenna, and both the radiation current and the transmission current can be considered separately.

The total current is the sum of the radiation current, the infinite series of the multiply-reflected transmission currents, and the unknown reflected current $g(z)$. Mathematically it can be written in the form

$$I(z) = I_r(z) + G_s \frac{e^{ik_s z}[1 + R\,e^{i2k_s(h-z)}] + e^{ik_s h} g(z-h)}{1 - R\,e^{i2k_s h}}, \quad (52)$$

where $I_r(z)$ is the same as the radiation current of an infinite cylindrical dielectric-coated antenna, and G_s is its input transmission conductance. Both are given in Ting's paper,[1] and the driving voltage V is assumed to be 1. The current–standing-wave ratio is found to be $S = 1 + |R|/1 - |R|$. Let the input admittance be defined as the current at point $z = 0$, which from (52), is

$$Y_{\text{in}} = G_{\text{in}} + iB_{\text{in}} = G_r + iB_r + G_s \frac{1 + R\,e^{i2k_s h}}{1 - R\,e^{i2k_s h}}, \quad (53)$$

where G_{in}, B_{in} are the input conductance and input susceptance of the finite long antenna, and G_r, B_r are the input radiation conductance and the input radiation susceptance of the corresponding infinite antenna. It is seen from (53) that the locus of the input admittance is a circle.

If the resonant and antiresonant lengths are defined respectively at the maximum and minimum of the input conductance, they are given by

$$h = -\frac{1}{2k_s}\left(\tan^{-1}\frac{b}{a}\right), \quad (54)$$

and the maximum and minimum of the input susceptance occur at

$$h = \frac{1}{2k_s}\left(\cos^{-1}\frac{2|R|}{1+|R|^2} - \tan^{-1}\frac{b}{a}\right), \quad (55)$$

where $R = a + ib$. The corresponding maximum and minimum input conductances are $G_r + G_s \cdot S$, and $G_r + G_s/S$, respectively.

In the numerical calculation, the input radiation susceptance of an infinite cylindrical dielectric-coated antenna B_r due to the delta generator is infinite. One way to avoid this difficulty is to subtract the inside current from the outside current. Since the same logarithmic singularity occurs on both inside and outside surfaces near the driving point, they cancel when the two currents are subtracted and a finite value is obtained. This does not necessarily correspond exactly to the actual value for an infinite antenna with a certain gap, but it has been checked experimentally that they have the same order of magnitude.

The numerical values for the three cases $G_r = 3.0$, $b/a = 2, 4, 8$, have been calculated. The loci are shown graphically in Fig. 9. Superimposed on each circle is the input admittance curve calculated in Sec. IV for a relatively short antenna, and corrected (imaginary part) according to Lamensdorf's experiment.[9] It is interesting to note that as the length of the antenna increases, the admittance approaches the circle of (53). This differs from the bare long dipole antenna[10] for which the admittance ultimately converges to one point.

Typical current distributions, both in magnitude and phase, have been obtained from (52) for the above three cases with $h = 3\lambda_0$. They are shown in Fig. 10, in which $I_r(z)$ is obtained from Ting's paper.[1] Since $g(z)$ is unknown, the dotted lines at the ends are drawn arbitrarily.

A comparison of the current distribution obtained

[9] David Lamensdorf, "An Experimental Investigation of Dielectric-Coated Antennas," Cruft SR 13, Harvard University, 1966.

[10] Keigo Iizuka, R. W. P. King, and Sheila Prasad, Proc. Inst. Elec. Eng. **110**, (Feb. 1963), pp. 303–309.

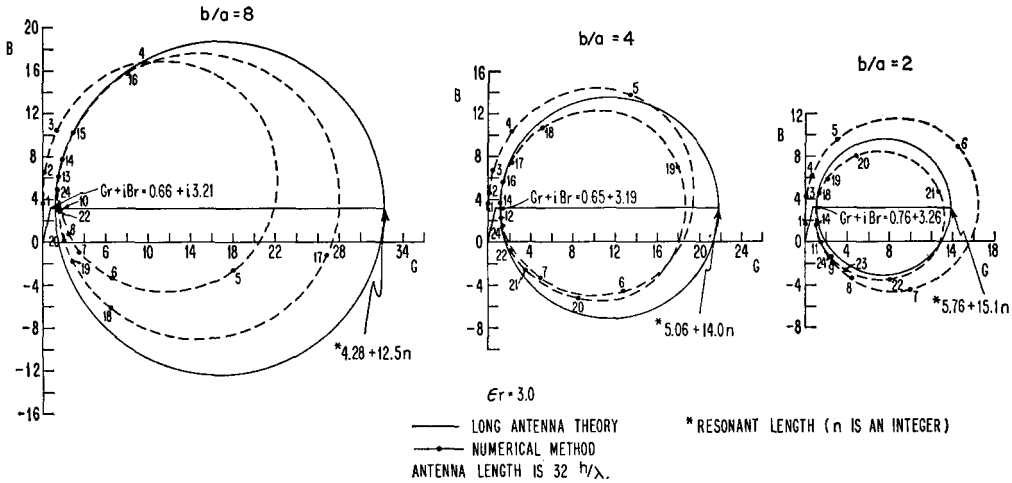


Fig. 9. Input admittance $Y = G + iB$ in millimhos.

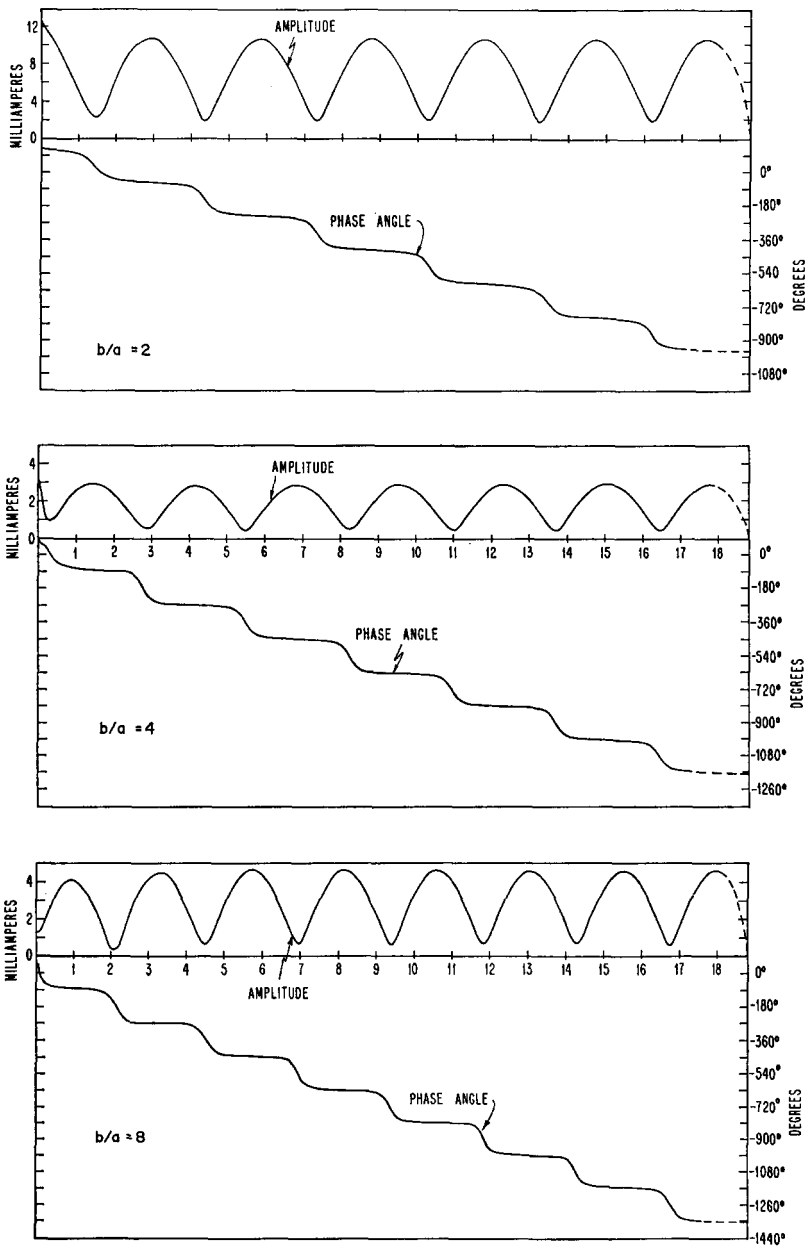


FIG. 10. Current distribution for

$\epsilon_r = 3.0$,
$k\,a = 0.04$,
$k_0 h = 6\pi$

by (52) with that obtained by numerical methods, can further confirm the theory. The longest structure calculated in Sec. IV is $h = \frac{3}{4}\lambda_0$. For the antenna with $b/a = 8$ which has the largest transmission current and the smallest radiation current among the three cases, the results are superimposed in Fig. 5. Even though $h = \frac{3}{4}\lambda_0$ is not really long, the experiment is not bad except near the end where the theory yields no answer.

VII. CONCLUSIONS

Two methods have been used in solving the problem of a finite dielectric-coated dipole antenna. The first is an entirely numerical method. Because of the limitation of the size of a computer and the computing time, it is useful only for relatively short antennas. Excellent results have been obtained when compared with experimental data. Actually, this method can be used to solve many kinds of problems that involve finite cylindrical antennas once the appropriate Green's function is known. The second method applies specifically when the antenna is sufficiently long. In general, the longer the antenna, the thicker the dielectric coating; and the higher the dielectric constant, the more accurate will the results be. The minimum length required before the theory can be applied depends on the desired accuracy. As the coating becomes thinner and thinner, the relative magnitude of the transmission current decreases, and the minimum length required becomes greater and greater. In the limit as the coating goes to zero, the transmission current vanishes and the theory ceases to exist.

Finally, no matter how complicated the mathematics may be, for engineering purposes a relatively short antenna with a dielectric coating of reasonable thickness can be treated simply as a free-space dipole with a modified wavenumber. The imaginary part of the current is well represented by a sine term, the real part by a shifted cosine term; however, the wavenumber is no longer that for free-space but close to that for the Goubou surface wave. The field pattern can be calculated with the free-space Green's function. In addition, the dielectric cylinder makes the antenna effectively longer and increases the radiation resistance of a very short dipole.

ACKNOWLEDGMENT

The author wishes to thank Professor R. W. P. King for his suggestion of this problem, his helpful discussion, for his reading and correcting of the manuscript, and above all, for his encouragement throughout the course of this research. The author also thanks Professor T. T. Wu, Dr. Chin-Lin Chen, and David Chang for their stimulating discussion and comments, and David Lamensdorf for his valuable experimental data.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

S-Matrix Singularity Structure in the Physical Region. I. Properties of Multiple 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)

We obtain the real singularities and corresponding discontinuities of a class of multiple integrals over real contours. Our aim is to give a unified treatment, obtaining, by elementary mathematical methods, both previously known results and some new generalizations. In a subsequent paper the results are applied to unitarity integrals.

1. INTRODUCTION

This paper is the first of a series which has the aim of showing that unitarity requires the physical amplitudes to have the Landau singularity structure characteristic of perturbation theory. The first step in this direction is to understand in some detail the singularities of unitarity integrals that occur for real values of the external momenta. To do this in full generality, new analytic machinery is needed. The aim of the present paper is to provide this machinery by simplifying and generalizing previous analyses of multiple integrals.[1-3]

The integrals $I(p)$ we consider are integrals over real values of k of an integrand possessing δ-function constraints $\delta(D_j(p, k))$ and singularities $S_i(p, k)$ with associated increments $i\epsilon_i$ that prescribe the distortion of the contour. It is found that the real singular points of $I(p)$ lie on certain arcs of Landau curves (Theorem 1). We classify three types of singularity mechanism:

(i) *Explicit;* which corresponds to the constraint surfaces D having linearly dependent normals at some point in k-space;

(ii) *Generative;* which corresponds to S and D surfaces having linearly dependent normals so that the contour is trapped; and

(iii) *Regenerative;* operating when the integrand itself has the singularity in question.

It is found that for generative and regenerative singularities, the integral is the limit onto real p of an analytic function, and that the sense of this limit is given by a simple rule (Theorem 2). In Sec. 3 we use a simple argument to evaluate discontinuities across generative singularities, paying particular attention to the over-all sign and region of integration of the resulting discontinuity integral. We are able to treat the case where there is no vanishing cycle and to extend the analysis to cover a variety of further situations.

In the applications which we shall discuss in later papers of this series, it is not, strictly speaking, the discontinuity which is important, but rather an expression denoted by $(I_> - I_{<\eta-i\epsilon})$ which gives the difference between I evaluated on one side of the Landau curve and a specific analytic continuation to the same p of the function obtained by evaluating I on the opposite side of the Landau curve. For an I which possesses a $+i\epsilon$ natural distortion (Sec. 2), this quantity is, in fact, the discontinuity, but we are also interested in other cases for which the two evaluations of I may not be continuations of each other, so that I does not, properly speaking, have a discontinuity round the Landau curve. In Theorem 4 we derive formulas for this important expression for all three classes of singularity occurring in various combinations.

In the final section we illustrate our theorems by considering Feynman integrals and briefly survey the properties of Landau singularities which will be used in subsequent papers.

2. THE INTEGRALS

Consider a multiple integral

$$I(p_1 \cdots p_n) = \int dk_1 \cdots dk_l f(p_1 \cdots p_n; k_1 \cdots k_l) \times \prod_j \delta(D_j(p_1 \cdots p_n; k_1 \cdots k_l)),$$

or, for short,

$$I(p) = \int dk\, f(p, k) \prod \delta(D). \tag{2.1}$$

* 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] For an account of the analytic properties of integrals and Landau curves together with a full list of references, see R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, London, 1966), Chap. 2.

[2] P. V. Landshoff and D. I. Olive, J. Math. Phys. 7, 1464 (1966); M. J. W. Bloxham, Nuovo Cimento **44**, 794 (1966); J. B. Boyling, Nuovo Cimento **44**, 379 (1966).

[3] An account of methods using homology theory is given in R. C. Hwa and V. L. Teplitz, *Homology and Feynman Integrals* (W. A. Benjamin, Inc., New York, 1966). We do not use these methods in this paper. Some of our theorems have recently been obtained by these methods: F. Pham, Ann. Inst. Henri Poincaré **6A**, 89 (1967).

The integrand $f(p, k)$ is supposed analytic for real values of p and k, satisfying the constraints

$$D_j(p, k) = 0, \quad j = 1, \cdots, r, \tag{2.2}$$

except for singularities

$$S_i(p, k) = 0, \quad i = 1, \cdots, m, \tag{2.3}$$

where S_i and D_j are real analytic functions. The integration contours are supposed real, except for small imaginary detours necessary in order to avoid the singularities. It is assumed that these can be specified by the sense along the imaginary normal to the S_i, that is, by deforming the contour near S_i by

$$\delta k = \frac{\partial S_i}{\partial k} i\epsilon_i, \tag{2.4}$$

where ϵ_i is a real increment of given sign. The corresponding variation in S_i is given by

$$dS_i = i\epsilon_i \sum \left(\frac{\partial S_i}{\partial k}\right)^2, \tag{2.5}$$

and we talk of an "$S_i + i\epsilon_i$" distortion of the contour. The generalized Landau curve in p space arising through the participation of some subset of these S's and D's is defined by the implicit equations

$$\begin{aligned} &\text{each participating } S(p, k) = 0, \\ &\text{each participating } D(p, k) = 0, \\ &\sum \sigma_i \frac{\partial S_i}{\partial k} + \sum \alpha_j \frac{\partial D_j}{\partial k} = 0, \end{aligned} \tag{2.6}$$

where α_j and σ_i are coefficients which are nonzero for the participating S's and D's. Before seeing that $I(p)$ can be singular only on certain arcs of these curves, we dismiss for the time being the case when the integrand $f(p, k)$ has a singularity $S(p)$ independent of k. The Landau equations are then trivially satisfied, and $I(p)$ has the same singularity, which is of the regenerative type.

The Singularity Theorem

Theorem 1[4]*:* For real values of p, the integral $I(p)$ can only be singular at points lying on the Landau curve (2.6) corresponding to real values of k and satisfying the further conditions

(i) for each participating S: sign $\sigma_i\epsilon_i$ = sign ϵ_{nat},

(ii) for each nonparticipating D: $D(p, k) = 0$. (2.7)

Thus the quantities $\sigma_i\epsilon_i$ must have a common sign, and ϵ_{nat} is a newly defined increment with that sign.

[4] See H. P. Stapp, Phys. Rev. **125**, 2139 (1962), Appendix H; I. T. Drummond, Nuovo Cimento **29**, 720 (1963); P. V. Landshoff and D. I. Olive, J. Math. Phys. **7**, 1464 (1966), Appendix.

Proof: $I(p)$ is analytic at p only if all its derivatives with respect to the p variables exist in the sense of complex variable theory. Let Δp be a small variation in p and consider

$$I(p + \Delta p) = \int f(p + \Delta p, k) \prod \delta(D(p + \Delta p, k))\, dk.$$

For each k satisfying (2.2), we can find a Δk such that

$$D_j(p + \Delta p, k) = D_j(p, k + \Delta k),$$

i e.,

$$\sum_{\alpha=1}^{n} \frac{\partial D_j}{\partial p_\alpha}(p, k)\Delta p_\alpha = \sum_{\lambda=1}^{l} \frac{\partial D_j}{\partial k_\lambda}(p, k)\Delta k_\lambda, \tag{2.8}$$

providing the $(r \times l)$ matrix $\partial D_j/\partial k_\lambda$ has rank r. This is certainly so if the r rows are linearly independent, that is, if the Landau equations are not satisfied with the participation of D's alone. If the equations are satisfied with the participation of D's alone, p lies on an explicit singularity, and we can proceed no further in attempting to prove analyticity at p. Away from an explicit singularity, $\Delta k(k, p, \Delta p)$ can be found as a linear function in Δp, and is analytic in k and p, as this is true of the D's. The transformation of variables $k + \Delta k = \hat{k}$ has Jacobian J equaling one plus a function linear in Δp and analytic in $(\hat{k}, p)$, and gives

$$I(p + \Delta p) = \int f(p + \Delta p, \hat{k} - \Delta k) J \prod \delta(D(p, \hat{k}))\, d\hat{k}.$$

We can now calculate the derivatives of I, providing we can calculate $f(p + \Delta p, k - \Delta k)$ for all points on the contour and for all differentials satisfying (2.8). In other words, $I(p)$ is analytic if f is analytic at all points of the contour for differentials lying in the constraint surfaces (2.2).

Since there may well be singularities of f on the real contour, the next step is to see under what circumstances we can free the contour of singularities by a small deformation. Let k be a real point satisfying (2.2), and lying on an intersection of certain of the singularity surfaces,[5] so that

$$S_i(p, k) = 0, \quad i = 1, \cdots, M, \quad i \leq M \leq m.$$

Consider the possibility of deforming the contour away from each impinging singularity at k via an imaginary increment δk so that, according to (2.5),

$$dS_i \equiv \sum_{\lambda=1}^{l} \frac{\partial S_i}{\partial k_\lambda} \delta k_\lambda = i\rho_i, \quad \text{where} \quad \rho_i\epsilon_i > 0, \quad i = 1, \cdots, M, \tag{2.9}$$

[5] In the applications of interest there are an infinite number of possible singularity surfaces, but only a finite number can participate in a pinch which corresponds to a solution of (2.6) giving a curve in p space.

while we keep

$$dD_j \equiv \sum_{\lambda=1}^{l} \frac{\partial D_j}{\partial k_\lambda} \delta k_\lambda = 0, \quad j = 1, \cdots, r. \quad (2.10)$$

The $M + r$ variations dS and dD are given in terms of the l quantities δk by these differential relations. If the $(r + M) \times l$ matrix relating the differentials has rank $r + M$, then we can always find a δk to give any chosen dS and dD (not all zero), and in particular to satisfy (2.9) and (2.10). If the rank is one less than this, the rows of the matrix are linearly dependent so that the Landau equations are satisfied by the S's and D's. Lower ranks would still correspond to intersections of Landau curves and will not be further discussed. Even if the rank is $r + M - 1$, it may still be possible to find differentials satisfying (2.9) and (2.10). Let us choose δk so that the latter equations are satisfied. Then, by (2.6) and (2.10),

$$\sum_{i=1}^{M} \sigma_i \, dS_i = 0$$

or, equivalently, if we can apply (2.9),

$$\sum_{i=1}^{M} \sigma_i \rho_i = 0, \quad \rho_i \epsilon_i > 0. \quad (2.11)$$

This is impossible if the quantities $\sigma_i \epsilon_i$ all have the same sign. Thus conditions (2.7), in addition to (2.6), imply that we cannot make a suitable distortion, and I is singular. Conversely, if we cannot distort contours as required, (2.7) must be satisfied in addition to (2.6). For suppose $\sigma_1 \epsilon_1 > 0$. Then we can choose δk so that (2.10) is satisfied, $\rho_1 \cdots \rho_{M-1}$ satisfy $\rho\epsilon > 0$, and also

$$\sum_{i=1}^{M-1} \sigma_i \rho_i > 0.$$

It follows from (2.11) that $\sigma_M \rho_M < 0$. By the hypothesis that we cannot find the desired deformation, $\rho_M \epsilon_M < 0$, and hence $\sigma_M \epsilon_M > 0$. The same argument can be repeated to show that each $\sigma_i \epsilon_i > 0$, which establishes the result.

The Landau Curves

If the Landau equations (2.6) are solved for k, σ, and α in terms of p, the Landau curve is obtained in the form

$$L(p) \equiv \sum \sigma S + \sum \alpha D = 0. \quad (2.12)$$

We can define a variable η measured along the normal to L at the point p by

$$d\eta = \frac{\partial L}{\partial p} \, dp = \left(\sum \sigma \frac{\partial S}{\partial p} + \sum \alpha \frac{\partial D}{\partial p} \right) dp, \quad (2.13)$$

using (2.6) to obtain the second equality. In general, we cannot show that the normal variable always exists or varies continuously, because there may be points where all $\partial L/\partial p$ vanish.

Singularity of the Landau curve can cease when one of the conditions (2.7) fails. One way of violating these conditions is for a σ to change sign. It must first vanish; and when it does so, the Landau equations for the lower-order curve with the corresponding S not participating are also satisfied. Furthermore, by (2.13), the normals coincide. Cessation of singularity at a point of tangency with a lower-order curve is called the *hierarchial effect*. As we see later in the study of unitarity integrals, singularity may also cease at a point of tangency with a higher-order curve when some conditions $D = 0$ cease to hold (an *anti-hierarchial effect*).

For the time being we are going to make the simplifying assumption that a particular Landau curve arises from a particular combination of participating S's and D's. Then we can prove[2] the following:

Theorem 2: Near a singular point of a Landau curve L where the normal exists, the two integrals $I_<$ and $I_>$, defined in the regions $\eta < 0$ and $\eta > 0$, are analytically related to each other by a path of continuation following an $\eta + i\epsilon_{\text{nat}}$ detour, with ϵ_{nat} defined by Eq. (2.7).

ϵ_{nat} is defined only when one or more S's participate, and the theorem applies only to this case, that is, to a generative singularity. In the case of an explicit singularity in which only D's participate so that no ϵ_{nat} is defined, the integrals $I_<$ and $I_>$ are not, in general, analytically related. This fact is very familiar: for example, the three-particle unitarity integral has an explicit singularity at the three-particle threshold and vanishes identically below the threshold. However, zero is not in general an analytic continuation of its value above.

We think of the integral as a limit as the ϵ_i's associated with the S_i's tend to zero. We may also consider the motion of $L = 0$ in p space as the ϵ_i are allowed to become small but finite. Variations $dS_i = i\epsilon_i$, $dD_j = 0$, lead to a displacement of L given by $d\eta = -i \sum \sigma_i \epsilon_i$. By (2.7) each term $\sigma_i \epsilon_i$ has the same sign, which is that of ϵ_{nat}. It is clear that the sense in which the integral is a boundary value must be opposite to that in which the singularity moves. We conclude, therefore, that the boundary value is an $\eta + i\epsilon_{\text{nat}}$ limit and refer to this as the *natural distortion*. This is equivalent to saying that $I_<$ and $I_>$ are analytically related by an $\eta + i\epsilon_{\text{nat}}$ path of analytic continuation, provided the mechanism under discussion is the only one associated with the singularity.

Since, in general, the normal to L is only defined almost everywhere on L, the integral is only a limit almost everywhere. Furthermore, as the normal may not vary continuously along L, the $i\epsilon$ prescription is local rather than global. Later we shall see that the $i\epsilon$ prescription is fully global for the amplitudes, but not, in general, for the unitarity integrals.

The natural distortion for a regenerative singularity is understood to be simply its $i\epsilon$ prescription so that it too is given by (2.7).

If I is now singular on L via several independent generative or regenerative mechanisms, each mechanism has its own natural distortion. These may well disagree and then, as for an explicit singularity, $I_>$ and $I_<$ are not analytically related.

3. DISCONTINUITIES OF MULTIPLE INTEGRALS

The Vanishing Cycle

Suppose that when $p = \bar{p}$, the surfaces S_i pinch at $k = \bar{k}$:

$$\sum \sigma_i \frac{\partial S_i}{\partial k} = 0. \tag{3.1}$$

When $\eta \neq 0$, define

$$V_{\gtrless} = \{k; k \text{ real}, \sigma_i S_i \gtrless 0, R(k) > 0\}, \tag{3.2}$$

where $R > 0$ is a small region surrounding $\bar{k}$. Then, if $V_> = 0$ in $\eta < 0$, we say it is a real vanishing cycle in $\eta > 0$. Similarly, if $V_< = 0$ in $\eta > 0$, it is a real vanishing cycle in $\eta < 0$.

We illustrate possible regions V by examples (Fig. 1). We see that in (a) there is only one vanishing cycle, $V_>$ in $\eta > 0$; in (b) there are two, $V_>$ in $\eta > 0$ and $V_<$ in $\eta < 0$; in (c) there are none.

These examples illustrate two general facts. First, when the number of surfaces is just one more than the dimensionality of the space in which they are linearly dependent, then there are two real vanishing cycles; second, when the number of surfaces is less than this and $\sigma_i S_i > 0$ are all convex at $\bar{k}$, then $V_>$ is a real vanishing cycle in $\eta > 0$.

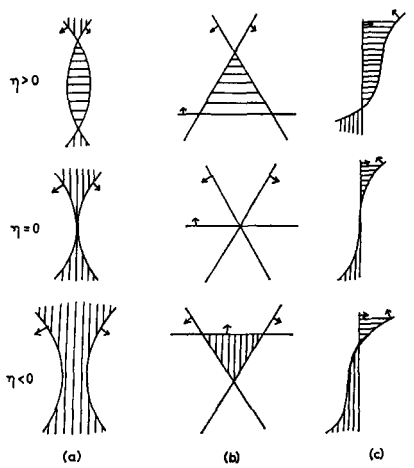


FIG. 1. The arrows indicate the normals into $\sigma_i S_i > 0$ for each S_i. It is these normals whose vector sum vanishes at the pinch point (3.1). $V_>$ and $V_<$ are, respectively, the regions shaded horizontally and vertically.

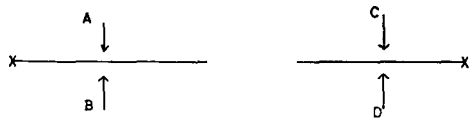


FIG. 2. $\text{dif}^> = A - B$, $\text{dif}^< = C - D$, for $\epsilon > 0$.

Definition of "dif"

Let $g(x)$ be a function with a branch point at $x = 0$, and

$$g = \operatorname*{Lt}_{\epsilon \to 0} g(x + i\epsilon)$$

where ϵ has some given sign. It is convenient to introduce the discontinuity forms

$$\left.\begin{matrix} \text{dif}_x^> g \\ \text{dif}_x^< g \end{matrix}\right\} \equiv g(x + i\epsilon) - g(x - i\epsilon), \tag{3.3}$$

where $g(x - i\epsilon)$ is defined by the cut being drawn along the positive real axis for $\text{dif}_x^>$ and along the negative axis for $\text{dif}_x^<$, as in Fig. 2. We note that $\text{dif}_x^>$ is zero in $x < 0$ and $\text{dif}_x^<$ zero in $x > 0$. Because of Theorem 2, for an integral with one generative mechanism producing a singularity at $\eta = 0$,

$$I(\eta) = \operatorname*{Lt}_{\epsilon_{\text{nat}} \to 0} I(\eta + i\epsilon_{\text{nat}})$$

and we can define $\text{dif}_\eta^> I$ and $\text{dif}_\eta^< I$. The latter is nonvanishing in $\eta < 0$, and if we continue it from there into $\eta > 0$ while swinging the cut round in the half-plane opposite the natural boundary value so as to preserve the natural distortion,

$$\text{dif}_\eta^< I \quad \text{in} \quad \eta < 0 \xrightarrow[\eta - i\epsilon_{\text{nat}}]{} -\text{dif}_\eta^> I \quad \text{in} \quad \eta > 0, \tag{3.4}$$

where the arrow denotes "analytically continues into," and, as indicated, the path of continuation passes below $\eta = 0$. If, with the same notation,

$$I_< \xrightarrow[\eta - i\epsilon_{\text{nat}}]{} I_{<\, \eta - i\epsilon_{\text{nat}}},$$

then in $\eta > 0$ we may write

$$\text{dif}_\eta^> I = I_> - I_{<\, \eta - i\epsilon_{\text{nat}}}.$$

The Basic Discontinuity Formula

For a singularity L generated by a pinch between S_i which form a real vanishing cycle $V_>$ in $\eta > 0$, Boyling[6] showed that

$$\text{dif}_\eta^> \int f\, dk = \int_{V_>} \prod \text{dif}_{S_i} f\, dk. \tag{3.5}$$

[6] J. B. Boyling, Nuovo Cimento **44**, 379 (1966). An intuitive statement of (3.5) was given in J. C. Polkinghorne, Nuovo Cimento **25**, 901 (1962), Appendix.

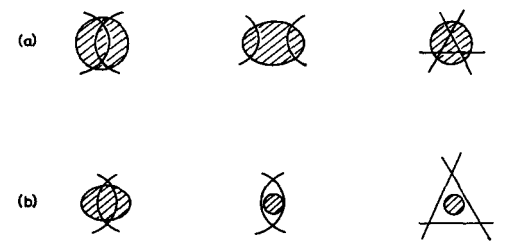


FIG. 3. The shaded regions are R-regions. Examples (a) are valid, but examples (b) are not, because as p approaches $\bar{p}$, pinches develop that involve $R = 0$.

An attractive point is the simplicity of the over-all factor, plus one. The result is unaffected if f includes delta-function constraints. We shall rederive this formula by elementary methods, generalizing the argument to treat the case where there is no real vanishing cycle. This is the usual circumstances for unitarity integrals on mixed α parts of Landau curves.

Consider an integrand f with singularities $S_1, \cdots, S_N$ (having associated distortions $i\epsilon_1, \cdots, i\epsilon_N$) that pinch at one point $k = \bar{k}$ to give a singularity at $p = \bar{p}$ on L. If p lies only on one Landau curve, it must be possible to choose a region $R(k) > 0$ containing $\bar{k}$ sufficiently small not to contain any other pinch configurations, nor allow its boundary to participate in any pinches with the S's when $p = \bar{p}$. We shall consider p in the neighborhood of $\bar{p}$ for which these statements remain true (see Fig. 3). If

$$A(\eta + i\epsilon_{\text{nat}}) = \int_{R>0} f\, dk,$$

then

$$\text{dif} \int f\, dk = \text{dif}\, A, \tag{3.6}$$

since $\int_{R<0} f\, dk$ is nonsingular at $\bar{p}$ because it omits the pinch point $\bar{k}$. By the definition of dif_{S_i} [Eq. (3.3) with ϵ_i for ϵ]

$$(S_1 + i\epsilon_1, S_2 + i\epsilon_2 \cdots) - f(S_1 - i\epsilon_1, S_2 + i\epsilon_2) = \text{dif}_{S_1}^{\gamma_1} f(S_1, S_2 + i\epsilon_2 \cdots),$$

where $\gamma_1 = \gtrless$ accordingly as we choose the cut attached to S_1 to lie in $S_1 > 0$ or $S_1 < 0$. To preserve the $S_1 + i\epsilon_1$ rule, the two choices of cut must be related by swinging the cut through the region $S_1 - i\epsilon_1$. Repeating for S_2,

$$f(S_1 + i\epsilon_1, S_2 + i\epsilon_2, \cdots) - f(S_1 + i\epsilon_1, S_2 - i\epsilon_2 \cdots) - f(S_1 - i\epsilon_1, S_2 + \epsilon_2) + f(S_1 - i\epsilon_1, S_2 - i\epsilon_2) = \text{dif}_{S_2}^{\gamma_2} \text{dif}_{S_1}^{\gamma_1} f(S_1 S_2 \cdots).$$

We can repeat for all $S_1 \cdots S_N$, providing we restrict ourselves to $\eta \neq 0$, since the normals to the S's are then not linearly dependent where S's intersect and in that case the various "difs" commute. Integrating the result over $R > 0$ yields

$$\int_{R>0} f(S_1 + i\epsilon_1, S_2 + i\epsilon_2 \cdots)\, dk + (-)^N \int_{R>0} f(S_1 - i\epsilon_1, S_2 - i\epsilon_2 \cdots)\, dk + R(\eta) = \int_{R>0} \prod \text{dif}_{S_i}^{\gamma_i} f\, dk. \tag{3.7}$$

$R(\eta)$ is made up of terms with mixtures of $S + i\epsilon$ and $S - i\epsilon$ and must be regular on the part of the Landau curve on which A is singular, by Theorem 1 (2.7), provided that we can assume that no subset of the S's can pinch to give the same L. Later in this section we discuss the case when this assumption is false.

The first term in (3.7) is A. The second we shall call B, and (by Theorems 1 and 2) it also is singular, with the opposite natural distortion $\eta - i\epsilon_{\text{nat}}$. There is no necessity for it to be a continuation of A. The term on the right we call C. $C_>$ and $C_<$, evaluated in $\eta > 0$ and $\eta < 0$, respectively, are not, in general, analytically related. Because of this we have two independent equations:

$$A(\eta + i\epsilon_{\text{nat}}) + B(\eta - i\epsilon_{\text{nat}}) + R(\eta) = C_>(\eta) \quad \text{in} \quad \eta > 0, \tag{3.8}$$

$$A(\eta + i\epsilon_{\text{nat}}) + B(\eta - i\epsilon_{\text{nat}}) + R(\eta) = C_<(\eta) \quad \text{in} \quad \eta < 0. \tag{3.9}$$

Continuing (3.9) into $\eta > 0$ following a $\eta - i\epsilon_{\text{nat}}$ detour,

$$A(\eta + i\epsilon_{\text{nat}}) \to A(\eta - i\epsilon_{\text{nat}}) \equiv A(\eta + i\epsilon_{\text{nat}}) - \text{dif}_\eta^> A,$$
$$B(\eta - i\epsilon_{\text{nat}}) \to B(\eta - i\epsilon_{\text{nat}}),$$
$$R(\eta) \to R(\eta),$$
$$C_<(\eta) \to (C_<(\eta))_{\eta - i\epsilon_{\text{nat}}},$$

and we obtain

$$A(\eta + i\epsilon_{\text{nat}}) - \text{dif}_\eta^> A + B(\eta - i\epsilon_{\text{nat}}) + R(\eta) = (C_<(\eta))_{\eta - i\epsilon_{\text{nat}}}, \quad \text{in} \quad \eta > 0.$$

Subtracting this from (3.8) and using (3.6), we have the result:

Theorem 3: In $\eta > 0$,

$$\text{dif}_\eta^> \int f\, dk = C_> - (C_<)_{\eta - i\epsilon_{\text{nat}}}; \tag{3.10}$$

similarly, in $\eta < 0$,

$$\text{dif}_\eta^< \int f\, dk = C_< - (C_>)_{\eta - i\epsilon_{\text{nat}}}; \tag{3.11}$$

where

$$C = \int_{R>0} \prod \mathrm{dif}_{S_i}^{\gamma_i} f\, dk \quad \text{for any choice of } \gamma_i. \tag{3.12}$$

If there is a real vanishing cycle $V_>$, Eq. (3.10) simplifies to Boyling's form (3.5) because there are σ_i such that $\sigma_i S_i > 0$ is empty in $\eta < 0$, so that for the choice $\gamma_i = \gtrless$ as $\sigma_i \gtrless 0$, $C_<$ has an empty integration region. Furthermore, for p sufficiently near to $\bar{p}$, $\sigma_i S_i > 0$, where the integrand does not vanish, is contained in R. This result shows the significance of the vanishing cycle. Theorem 3 applies even when there is no real vanishing cycle. In such a case our general answer apparently depends on the region $R > 0$. This cannot be so, as there is no such dependence in the original integral. In fact, when $C_<$ is continued and subtracted from $C_>$, part of the contours cancel. The part that remains is independent of R and, in general, is complex. As far as we know, it is difficult to specify it by simple rules.

Generalizations of the Discontinuity Theorem

We need some generalizations in order to apply our results to Feynman and unitarity integrals:

(1) The integrand f can include nonparticipating singularities. Since these do not enter $R > 0$, the argument is the same.

(2) δ functions $\delta(D_j)$ can be included in f whether or not they participate in the pinch. The argument is the same.

(3) The integration region may have boundaries $B_1, \cdots, B_M$ that participate in the pinch. We can choose the region $R > 0$ so that only these boundaries enter near $\bar{p}$.

Then, as $1 = \theta(B_1) + \theta(-B_1)$,

$$\int \theta(R)\theta(B_2)\cdots\theta(B_M) f\, dk = \int \theta(R)\theta(B_1)\theta(B_2)\cdots\theta(B_M) f\, dk + \int \theta(R)\theta(-B_1)\theta(B_2)\cdots\theta(B_M) f\, dk. \tag{3.13}$$

Providing f has no singularity coinciding with B_1, the left-hand side of this equation is regular at $\bar{p}$ since it lacks B_1 (which is vital to the pinch). Repeating this argument for each B in turn, we obtain

$$\mathrm{dif} \int \theta(R)\theta(B_1)\cdots\theta(B_M) f\, dk = \mathrm{dif} \int \theta(R) \prod \mathrm{sign}\,(\beta_j)\theta(\beta_j B_j) f\, dk$$

for any choice of β_j.

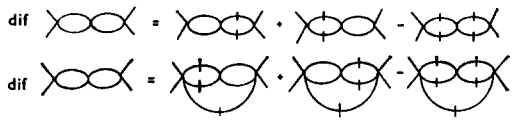


FIG. 4. Some more complicated dif formulas.

Repeating the argument leading to Theorem 3, we obtain (3.10) and (3.11), where, in place of (3.12), we have

$$C = \int_{R>0} \prod_j \mathrm{sign}\,(\beta_j)\theta(\beta_j B_j) \prod_i \mathrm{dif}_{S_i}^{\gamma_i} f\, dk.$$

This generalizes another of Boyling's conclusions.[6]

(4) Suppose the three subsets of S_i (defined by two basic subsets $i \in I_1$, $i \in I_2$, and their union $i \in I_1 \cup I_2$) each pinch to give the same L with the same natural distortion. An example of this is given by the double-loop self-energy graph of Fig. 4. Consider

$$h(p, k) \equiv \prod_{i\in I_1} \mathrm{dif}_{S_i}^{\gamma_i} f + \prod_{i\in I_2} \mathrm{dif}_{S_i}^{\gamma_i} f - \prod_{i\in I_1\cup I_2} \mathrm{dif}_{S_i}^{\gamma_i} f$$

with $S + i\epsilon_i$ understood for the S's without "difs." h expands to give a linear combination of f's with various $i\epsilon$ prescriptions for the S_i's. Consider the f's in this expansion which have the correct $i\epsilon$ prescription for all $i \in I_1$. In the expansion of the first term such an f appears only once, and has $S_i + i\epsilon_i$ for all the remaining i. In the second term such f's are found with all possible mixtures of prescriptions for the remaining i's, each combination appearing once. All these latter f's appear also in the third term, with the opposite sign, leaving only the $f(S_1 + i\epsilon_1 \cdots S_N + i\epsilon_N)$ from the first term. Similarly, the term $(-1)^N f(S_1 - i\epsilon_1 \cdots S_N - i\epsilon_N)$ appears just once, and no other term appears which, when integrated, can lead to singularity in accordance with Theorem 1. The argument leading to Theorem 3 then yields (3.10) and (3.11) with, in place of (3.12),

$$C = \int_{R>0} \prod_{I_1} \mathrm{dif}_{S_i}^{\gamma_i} f + \prod_{I_2} \mathrm{dif}_{S_i}^{\gamma_i} f - \prod_{I_1\cup I_2} \mathrm{dif}_{S_i}^{\gamma_i} f\, dk. \tag{3.14}$$

Similarly, with three basic subsets we find

$$C = \int_{R>0} \Big(\prod_{I_1} + \prod_{I_2} + \prod_{I_3} - \prod_{I_1\cup I_2} - \prod_{I_2\cup I_3} - \prod_{I_3\cup I_1} + \prod_{I_1\cup I_2\cup I_3} \Big) \mathrm{dif}_{S_i}^{\gamma_i} f\, dk,$$

and so forth.

We shall not discuss in this paper the cases where the sets I_1 and I_2 may have different natural distortions,

nor the case when similar problems arise for combinations of S's and D's.

(5) Finally we consider the possibility of an infinite degeneracy in k, that is, an infinite set of values of k, continuously connected, all of which satisfy the Landau equations for a given $\bar{p}$. As a simple example of this behavior we can take the case when the participating S's and D's are independent of one of the integration variables k_0. Denoting the remaining variables $\hat{k}$, consider $\hat{f}(\hat{k}) = \int dk_0 f(k_0, \hat{k})$. Each of the S's and D's occur regeneratively in f. dif_{S_i} commutes with the integration with respect to k_0 and $\mathrm{dif} \int f\,dk = \mathrm{dif} \int \hat{f}\,d\hat{k} = C_> - (C_<)_{\eta - i\epsilon \mathrm{nat}}$, where

$$C = \int_{R(\hat{k})>0} \prod \mathrm{dif}\, \hat{f}\, d\hat{k} = \int_{R(\hat{k})>0} \prod \mathrm{dif}\, f\, dk.$$

The only modification is the obvious one that R depends on $\hat{k}$ but not on k_0, so that it resembles a cylinder rather than a sphere in k space.

Different Mechanism in the Same Integral

So far we have considered integrals generating a given singularity L by only one mechanism. Suppose there are many mechanisms $a, b, \cdots$. Then, providing these are independent in the sense that the corresponding pinch points can be surrounded by disjoint regions R_a, R_b, we have

$$\int f\,dk = \int_{R_a} f\,dk + \int_{R_b} f\,dk + \cdots,$$

plus a function regular on L.

Each integral $\int_{R_i} f\,dk$ is singular on L by only one mechanism, but the different integrals may have different natural distortions, implying that no natural distortion can be given for the total integral $\int f\,dk$. If, on the other hand, the natural distortions of the various parts are the same, then $\mathrm{dif} \int f\,dk$ is well defined and is equal to the sum of the contributions from each mechanism taken independently. As we remarked in Sec. 1, in the unitarity analysis to which we subsequently apply our results it is not dif I which is the fundamental quantity, but rather $I_> - I_{<\eta - i\epsilon}$ and we now notice that this is a quantity which can still be evaluated simply even when natural distortions are mixed.

Suppose $\int_{R_i} f\,dk$ has a $+i\epsilon$ natural distortion. Then

$$\left(\int_{R_i} f\,dk\right)_> - \left(\int_{R_i} f\,dk\right)_{<\eta - i\epsilon} = \mathrm{dif}_\eta^> \int_{R_i} f\,dk = C_>(R_i) - C_<(R_i)_{\eta - i\epsilon},$$

while, if it has a $-i\epsilon$ natural distortion,

$$\left(\int_{R_i} f\,dk\right)_> - \left(\int_{R_i} f\,dk\right)_{<\eta - i\epsilon} = 0.$$

So for the whole integral

$$I_> - I_{<\eta - i\epsilon} = \left(\sum_+ C\right)_> - \left(\sum_+ C\right)_{<\eta - i\epsilon}, \quad (3.15)$$

where the sum $\sum$ covers the C integrals for the mechanisms with a $+i\epsilon$ natural distortion.

If I has in addition an explicit singularity corresponding to a pinch of constraint surfaces at $\bar{k}_E$, we surround this point with a region R and divide I into two parts $\int_{R>0} f\,dk$ and $\int_{R<0} f\,dk$. If R is sufficiently small, the first integral has no generative singularity, while the second has no explicit singularity so that the previous analysis applies. Equation (3.15) is unaltered if for an explicit singularity we understand the C function to be I itself, integrated over R, and require it to be included in both summations $\sum_+$ and $\sum_-$.

We now consider the case when I is singular by both a regenerative mechanism and a generative mechanism (but not also singular by an explicit mechanism). The natural distortion for the regenerative singularity is simply its $i\epsilon$ prescription (understanding S = positive number $\times$ L). We are concerned with evaluating $I_> - I_{<\eta - i\epsilon}$ and both mechanisms will only be nontrivially involved if they both have $+i\epsilon$ natural distortions. In this case

$$\begin{aligned} I_> - I_{<\eta - i\epsilon} &= \left(\int f(S + i\epsilon, S_i + i\epsilon_i)\, dk\right)_> \\ &\quad - \left(\int f(S + i\epsilon, S_i + i\epsilon_i)\, dk\right)_{<\eta - i\epsilon} \\ &= \int \mathrm{dif}_S^> f + J_> - J_{<\eta - i\epsilon}, \end{aligned}$$

where

$$J = \int f(S - i\epsilon, S_i + i\epsilon_i)\, dk$$

and we have used the identity of $S + i\epsilon$ and $S - i\epsilon$ in $\eta < 0$. Now the regenerative mechanism making J singular on L has a $-i\epsilon$ natural distortion, so, applying the previous analysis and using $f(S - i\epsilon) = f(S + i\epsilon) - \mathrm{dif}_S^> f$, we obtain

$$\begin{aligned} I_> - I_{<\eta - i\epsilon} &= \int \mathrm{dif}_S^> f(S, S_i + i\epsilon_i)\, dk \\ &\quad + \left(\int_{R>0} \prod \mathrm{dif}_{S_i}^{\gamma_i} f(S + i\epsilon, S_i)\, dk\right)_> \\ &\quad - (\text{the same})_{<\eta - i\epsilon} \\ &\quad - \int_{R>0} \mathrm{dif}_S^> \prod_i \mathrm{dif}_{S_i}^{\gamma_i} f(S, S_i)\, dk. \end{aligned} \quad (3.16)$$

In the situations to which we shall apply this result, the final multiple discontinuity will be zero, and we once more obtain a straightforward sum of two contributions.

Thus in these circumstances we arrive at our final theorem.

Theorem 4:

$$I_> - I_{<\eta-i\epsilon} = \left(\sum_+ C\right)_> - \left(\sum_+ C\right)_{<\eta-i\epsilon}, \quad (3.17)$$

where the sum runs over explicit mechanisms and over those generative and regenerative mechanism that have an $\eta + i\epsilon$ natural distortion, and

$$C = \int_{R_G>0} \prod \mathrm{dif}_{S_i}^{\gamma_i} f\, dk, \quad \text{for generative,}$$

$$C = \int \mathrm{dif}_S^{\geq} f\, dk, \quad \text{for regenerative,}$$

$$C = \int_{R_E>0} f\, dk, \quad \text{for explicit mechanisms.}$$

We have not proved this theorem for the case when all three mechanisms occur in a single integral, as this will not be so in any of the applications we have in mind.

Similarly,

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

where $\sum_-$ includes mechanisms with $\eta - i\epsilon$ natural distortions and explicit mechanisms.

4. APPLICATION TO FEYNMAN INTEGRALS

We now illustrate our work by considering Feynman integrals. They constitute a particularly simple example, provided one ignores, as we shall, the special features arising from the noncompactness of their integration regions.

In a theory with one spin-zero particle of mass m with Lagrangian terms of the form $g\varphi^n\varphi^{\dagger m}/n!\,m! +$ c.c., the contribution to an amplitude corresponding to a particular connected Feynman diagram F is given by the Feynman rules[7]

(i) g for each vertex,
(ii) $(q^2 - m^2 + i\epsilon)^{-1}$ for each line,
(iii) $\int i(2\pi)^{-4}\, d^4k$ for each loop,
(iv) $(n_F)^{-1}$, where n_F is the symmetry number of the diagram, that is, the number of permutations of internal lines which preserve F, holding external lines fixed. (4.1)

[7] T. T. Wu, Phys. Rev. **125**, 1436 (1962). These rules apply to the amplitude which is the connected part of the S matrix with the factor $-i(2\pi)^4\delta(P_i - P_f)$ divided out, taken between covariantly normalized states so that, e.g., $\langle p \mid p' \rangle = (2\pi)^3 2p^0\delta(\mathbf{p} - \mathbf{p}')$.

It is understood that the internal momenta q are expressed as linear combinations of the loop momenta k and the external momenta p. Such an integral has singularities $S_i(k, p) = q_i^2 - m^2$, each with an $S + i\epsilon$ prescription, and no D constraints. Applying Theorem 1, we find that the integral is singular for real p at points of the Landau curve:

each participating line:

$$q^2 - m^2 = 0; \quad (4.2)$$

each loop:

$$\sum_k \alpha q = 0, \quad (4.3)$$

where for each participating line sign $\alpha = \text{sign}\, \epsilon_{\text{nat}}$. $\sum_k$ denotes summation over the lines through which k runs. It is conventional to choose the over-all sign of the α's so that they are all positive at a singularity and talk of the positive α parts of the Landau curves.

Following (2.12) we can write a Landau curve as

$$L(p) \equiv \sum \alpha(q^2 - m^2) = 0 \quad (4.4)$$

and define a variable along the normal by

$$d\eta = \sum \frac{\partial L}{\partial p}\, dp = \sum \left(\sum_p \alpha q\right) dp. \quad (4.5)$$

It is a consequence of a theorem due to Pham[8] that the normal varies continuously along a positive α arc of a Landau curve. The theorem states that if $\bar{p}$ is a point on a positive α arc of a Landau curve, then all points p of L corresponding to q_i^0 with the same sign as at $\bar{p}$ lie to the same side of the tangent plane at $\bar{p}$. Thus, as long as we exclude parts of L corresponding to internal antiparticles rather than particles, L is convex at positive α points. Having chosen the α to be positive, we refer to $d\eta > 0$ as the inside of L. Furthermore, the integral has an $\eta + i\epsilon$ prescription, that is, it is the limit onto real η of a function analytic in Im $\eta > 0$. Each Landau singularity of a Feynman integral has associated with it a figure obtained from the Feynman diagram by contracting nonparticipating lines. According to the hierarchical effect of Sec. 2, a singularity associated with one such diagram is switched off where one or more α's vanish and this will be at a point of tangency (*effective intersection*) with the lower-order curve whose figure is obtained by contracting the relevant lines. There the boundary-value prescriptions for the integral on the two curves

[8] F. Pham, Ann. Inst. Henri Poincaré **6A**, 89 (1967). Similar results and generalizations to the mass shell appear in C. Chandler and H. P. Stapp, "Macroscopic Causality Conditions and Properties of Scattering Amplitudes," to be published in J. Math. Phys. **10** (1969).

agree, and (since also the sense of the boundary value varies continuously round each positive α arc) we can say that a Feynman integral possesses a "global $i\epsilon$ prescription." We shall discover in the next paper this is not something we can say of unitarity integrals, as these may be singular on mixed α arcs.

There are also more complicated hierarchical effects corresponding to the possibility that the contraction may yield a hinged graph, that is, two subgraphs joined by a common vertex. In this case the three normals of the corresponding Landau curves are linearly dependent.

The Discontinuity

For simplicity we only evaluate the discontinuity of a Feynman integral across its "leading singularity," in which all lines participate.

Let us assume first that the conditions of Theorem 3 are satisfied. By Pham's theorem[8] above, each region $\alpha(q^2 - m^2) > 0$ is convex at the pinch point. Hence there is a real vanishing cycle in $\eta > 0$, and $C_< = 0$. As each singularity is a pole and

$$\text{dif} \gtrless (q^2 - m^2 + i\epsilon)^{-1} = -2\pi i\delta(q^2 - m^2),$$

we find that $\text{dif}^> F$ is obtained from the Feynman integral by replacing each propagator as follows:

$$(q^2 - m^2 + i\epsilon)^{-1} \to -2\pi i\delta(q^2 - m^2). \quad (4.6)$$

This is the well-known Cutkosky rule.

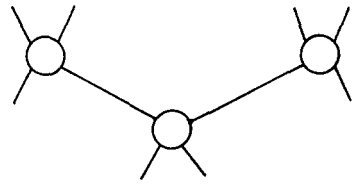


FIG. 5. A simple hinged diagram.

The straightforward Cutkosky rule requires modification when the Feynman integral satisfies the conditions of generalization (4), as do, for example, the graphs on the left-hand side of the equations of Fig. 4. In this case (3.14) is applicable and results in discontinuities such as those given in the figure, where the graphs are evaluated by the Feynman rules (4.1), except that —|— is to stand for $-2\pi i\delta(q^2 - m^2)$ while ——— represents the usual $(q^2 - m^2 + i\epsilon)^{-1}$ propagator.

Finally, we note that some diagrams give rise to two or more simultaneous singularities. A trivial example is the double pole of Fig. 5. We call such diagrams *hinged diagrams*. In such cases the multiple discontinuity across the simultaneous singularities is given by the Cutkosky rules.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

High-Energy Scattering Cross Section for Singular Potentials at Large Angles

RICHARD M. SPECTOR
Physics Department, Wayne State University, Detroit, Michigan

(Received 28 March 1968)

For scattering off singular repulsive potentials at high energy, the large-angle scattering is shown to be classical. For nonrelativistic singular potentials of the form $V(r) = gr^{-n}$ $(g > 0, n > 2)$, we give the explicit form of the differential cross section for large angles at high energy as a power series in $(\pi - \theta)$. The coefficients of the linear through cubic terms in $(\pi - \theta)$ are obtained.

In recent years a large amount of effort has gone into the study of singular potentials. In nonrelativistic theory, there have been studies of exact scattering solutions for special cases,[1] studies of special techniques of solution (peratization),[2] and studies of high-energy[3] and low-energy[4] phase shifts. In addition, various papers have dealt with general theoretical problems[5] and physical applications.[6,7]

Despite the many investigations in this area, there has been little work on differential cross sections with two main exceptions.[7] These two papers are not only very specialized but are very complicated formally. In this paper we demonstrate that for singular potentials the high-energy large-angle (nearly backward in the center-of-mass system) scattering is classical. This enables us to explicitly calculate the differential cross section by means of a power series in the classical impact parameter. This yields a power series in $(\pi - \theta)$ for the scattering for $\theta \approx \pi$. We explicitly evaluate the linear through cubic terms and indicate the region of validity of these terms. A brief abstract of this work has been previously reported.[8]

Our argument to establish the necessary classical behavior follows along lines given in Schiff[9] for the classical limit of Coulomb scattering. For high energies the quantum-mechanical scattering can be classical if the wavelength of the incoming particle is small compared to the classical distance of the closest approach. This requires that

$$1/k \ll r_0 \tag{1}$$

where k is the wavenumber and r_0 the closest-approach distance. For a singular potential of the form $V(r) = gr^{-n}$ $(g > 0, n > 2)$ we have

$$gr_0^{-n} = E = \hbar^2k^2/2m,$$
$$r_0 = (2mg/\hbar^2k^2)^{1/n}, \tag{2}$$

so that (1) gives

$$\frac{\hbar^{2/n}k^{-1+2/n}}{(2mg)^{1/n}} \ll 1. \tag{3}$$

For $n > 2$ there is some energy above which (3) becomes valid. In fact, at any energy (3) is true for a large enough mass so that there are nonrelativistic regions for which (3) is true. Actually, (1) is obviously fulfilled at high energy for any singular potential.[10] However, (3) alone is obviously not quite sufficient to make the scattering classical.

For example, the forward scattering for $n > 3$ in the quantum-mechanical case is well defined,[11] but classically it diverges. This occurs because the long tail of the potential falls off too slowly classically, giving rise to much small-angle scattering from large impact parameters. Quantum-mechanically this small-angle scattering cannot be resolved due to inherent quantum uncertainties.

For large-angle scattering at small impact parameters we may always obtain $l(\pi - \theta) \gg 1$ for high enough energies except only at $\theta = \pi$ where $l = 0$. Hence this condition, which is the usual classical limit requirement,[11] can be satisfied arbitrarily near π for high enough energy.

The classical formula for the scattering of a particle of energy E with impact parameter s is[11]

$$\theta = \pi - 2s\int_0^{u_0}\left[1 - \frac{V(u)}{E} - s^2u^2\right]^{-\frac{1}{2}}du, \tag{4}$$

[1] R. Spector, J. Math. Phys. **5**, 1185 (1965); E. Vogt and G. H. Wannier, Phys. Rev. **95**, 1190 (1954). References 1–7 are only representative, not comprehensive.

[2] G. Tiktopoulous and S. Treiman, Phys. Rev. **134**, B844 (1964); H. Aly, Riazuddin, and A. Zimmerman, *ibid.* **136**, B1174 (1964); R. Spector, J. Math. Phys. **7**, 2103 (1966); T. T. Wu, Phys. Rev. **136**, B1176 (1964).

[3] F. Calogero, Nuovo Cimento **27**, 261 (1963).

[4] T. O'Malley, L. Spruch, and L. Rosenberg, J. Math. Phys. **2**, 491 (1961).

[5] H. Cornille, Nuovo Cimento **38**, 1243 (1965); **39**, 557 (1965); **43**, 786 (1966).

[6] R. Spector and Ramesh Chand, Progr. Theoret. Phys. (Kyoto) **39**, 682 (1968).

[7] G. Tiktopoulous, Phys. Rev. **138**, B1550 (1965); C. B. Kouris, Nuovo Cimento **44**, 598 (1966).

[8] R. Spector, Bull. Am. Phys. Soc. **12**, 50 (1967).

[9] Leonard Schiff, *Quantum Mechanics* (McGraw-Hill Book Co., Inc., New York, 1955), 2nd ed., p. 120.

[10] We mean, in the usual terminology, any potential more singular than the inverse square at the origin.

[11] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon Press, Inc., London, 1958).

where $V(u) = gu^n = gr^{-n}$ and θ is the center-of-mass scattering angle. In (4), $u_0(s^2)$ is the value for which we have

$$1 - \frac{V(u_0)}{E} - s^2u_0^2 = 0. \tag{5}$$

We write (4) as

$$\chi = \pi - \theta = 2z^{\frac{1}{2}}\int_0^{u_0}[1 - \lambda u^n - zu^2]^{-\frac{1}{2}}\,du \tag{6}$$

with $z = s^2$ and $\lambda = g/E$. We wish to investigate small values of z which give rise to large-angle scattering. To this end we write

$$I(z) = \int_0^{u_0(z)}[1 - \lambda u^n - zu^2]^{-\frac{1}{2}}\,du, \tag{7}$$

where

$$I(0) = \int_0^{u_0(0)}[1 - \lambda u^n]^{-\frac{1}{2}}\,du, \quad \text{with} \quad u_0(0) = \lambda^{-1/n}.$$

Putting $t = \lambda u^n$ we immediately have, with $p = 1/n$,

$$I(0) = p\lambda^{-p}\int_0^1 t^{p-1}(1 - t)^{-\frac{1}{2}}\,dt, \tag{8}$$

where the integral is just the beta function and

$$\beta(p, \tfrac{1}{2}) = \frac{\Gamma(p)\Gamma(\frac{1}{2})}{\Gamma(p + \frac{1}{2})} = (\pi)^{\frac{1}{2}}\frac{\Gamma(p)}{\Gamma(p + \frac{1}{2})}. \tag{9}$$

The linear z term in $I(z)$ is not so easily found because z appears in both the upper limit of integration and the integrand. No expansion for small z is possible unless the limit [which is the root of Eq. (5)] and the integrand are consistently approximated at the same time.

We may evaluate the derivative of $I(z)$ at $z = 0$, however. Thus,

$$\frac{dI}{dz} = \lim_{R\to u_0}\left[\frac{du_0/dz}{[1 - \lambda R^n - zR^2]^{\frac{1}{2}}} + \frac{1}{2}\int_0^R \frac{u^2\,du}{[1 - \lambda u^n - zu^2]^{\frac{3}{2}}}\right], \tag{10}$$

where the limit is necessary since the two terms have mutually cancelling singularities. To circumvent this we may integrate by parts the integral in (10):

$$\frac{1}{2}\int_0^R \frac{(n\lambda + 2zu)}{(n\lambda + 2zu)}\frac{u^2}{[1 - \lambda u^n - zu^2]^{\frac{3}{2}}}\,du = \frac{u}{n\lambda u^{n-2} + 2z}\frac{1}{[1 - \lambda u^n - zu^2]^{\frac{1}{2}}}\Bigg|_0^R - \int_0^R \frac{1}{[1 - \lambda u^n - zu^2]^{\frac{1}{2}}}\frac{d}{du}\left[\frac{u}{n\lambda u^{n-2} + 2z}\right]du. \tag{11}$$

It is trivially shown by differentiating (5) with respect to z that

$$\frac{u_0}{n\lambda u_0^{n-2} + 2z} = -\frac{du_0}{dz}, \tag{12}$$

so that the upper limit of the integrated term in (11) exactly cancels the corresponding term in (10) in the limit $R \to u_0$. We have now that

$$\frac{dI}{dz}\bigg|_{z=0} = -\int_0^{u_0=\lambda^{-p}} \frac{1}{[1 - \lambda u^n]^{\frac{1}{2}}}\frac{d}{du}\left\{\frac{u^{3-n}}{n\lambda}\right\}du = \frac{(n-3)}{n\lambda}\int_0^{\lambda^{-p}}\frac{u^{2-n}}{[1 - \lambda u^n]^{\frac{1}{2}}}\,du. \tag{13}$$

This integral converges only for $n < 3$; however, the expression in (11) exists at $z = 0$ for all n. What has happened is that in going from (11) to (13) the lower limit of the integrated term in (11) has been lost since it is zero for $z \neq 0$. And the integral in (11) also exists for all n for $z \neq 0$. However, for $z = 0$ the integrated term and the integral in (11) have mutually cancelling singularities for $n \geq 3$. This is easily seen since near $u \approx 0$ we may neglect the square root and do the integral near its lower limit. The result exactly cancels the lower limit of the integrated term in (11) for all n, and for all z whether zero or not.

Since the value of dI/dz at $z = 0$ is well defined for all $n > 2$ and must agree with (13) for $n < 3$, we may evaluate (13) and use that value for all n.[12] Thus, with $t = \lambda u^n$ again,

$$\frac{dI}{dz}\bigg|_{z=0} = \lambda^{-3p}\left(\frac{n-3}{n^2}\right)\int_0^1 t^{3p-2}(1 - t)^{-\frac{1}{2}}\,dt = \lambda^{-3p}\left(\frac{n-3}{n^2}\right)\frac{\Gamma(3p-1)\Gamma(\frac{1}{2})}{\Gamma(3p - \frac{1}{2})} = -\lambda^{-3p}p\sqrt{\pi}\,\frac{\Gamma(3p)}{\Gamma(3p - \frac{1}{2})}. \tag{14}$$

We have that, since $I(z) \approx I(0) + [dI(0)/dz]z$,

$$\chi = A_1s - A_2s^3 \tag{15}$$

with

$$A_1 = 2(\pi)^{\frac{1}{2}}p\lambda^{-p}\frac{\Gamma(p)}{\Gamma(p + \frac{1}{2})},$$

$$A_2 = 2(\pi)^{\frac{1}{2}}p\lambda^{-3p}\frac{\Gamma(3p)}{\Gamma(3p - \frac{1}{2})}.$$

We must seek s as a function of χ to find the differential cross section. We put

$$s = B_1\chi + B_2\chi^2 + B_3\chi^3$$

[12] More explicitly, Eq. (13) is defined in a strip in the complex n plane which includes the real axis. Then (13) exists for all Re $n \geq 2$ at $z = 0$. Equation (14) must be true for Re $n \geq 3$ by analytic continuation. Actually (14) can be found by other lengthy means involving some manipulations on (13) which dispense with the sophisticated analytic-continuation argument used here. The answer has been found this way and is the same, as it must be.

and assuming χ is small, solve (15) to cubic order. This results in

$$s \approx \frac{1}{A_1}\chi + \frac{A_2}{A_1^4}\chi^3,$$
$$\frac{ds}{d\theta} \approx -\frac{1}{A_1} - \frac{3A_2}{A_1^4}\chi^2. \tag{16}$$

Using the standard formula we have

$$\sin\theta\sigma(\theta) = -s\frac{ds}{d\theta} \approx \frac{1}{A_1^2}\chi + \frac{4A_2}{A_1^5}\chi^3, \tag{17}$$

valid for high energies and $\chi \approx 0$ ($\theta \approx \pi$).

We may investigate the validity of (17) by requiring that

$$\chi^3\left|\frac{4A_2}{A_1^5}\right| \ll \frac{\chi}{A_1^2},$$
$$\chi^2\left|\frac{4A_2}{A_1^3}\right| \ll 1, \tag{18}$$
$$\left|\frac{\chi^2}{\pi p^2}\frac{\Gamma(3p)\Gamma^3(p+\frac{1}{2})}{\Gamma(3p-\frac{1}{2})\Gamma^3(p)}\right| \ll 1.$$

For n large enough (p small enough), we may get a ready evaluation of (18), since $\Gamma(x) \approx 1/x$ for small x, as

$$\chi = (\pi - \theta) \ll \sqrt{6}.$$

Hence, for large n, Eq. (17) is valid for θ deviating from π by, say, up to half a radian or so. We note that the validity of (17) as shown by (18) is independent of energy. However for energies too low, the expression (17), though a valid expression for the classical scattering, does *not* approximate the quantum-mechanical expression. It it necessary to keep in mind that (17) is a limit in two senses: large energy and large angle.

High-energy singular-potential scattering is of physical interest in the investigation of p–p and π–p scattering. Such potentials with complex strengths have been used by a number of workers.[7] An extensive discussion of the physical use of singular potentials may be found in a forthcoming review paper.[13]

ACKNOWLEDGMENTS

I am grateful to Dr. L. D. Favro and Dr. A. M. Saperstein for helpful discussions.

[13] W. M. Frank, D. J. Land, and R. M. Spector, "Singular Potentials," a review article to be published in 1969.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Mean Ergodic Theorems in Quantum Mechanics*

J. E. MOYAL
Argonne National Laboratory, Argonne, Illinois

(Received 4 April 1968)

An application of the abstract mean ergodic theorems to quantum systems is described, which is rather closely analogous to the application of these same theorems in classical statistical mechanics.

The abstract mean ergodic theorems[1] can be applied to algebras of quantum observables to yield results which are in closer analogy to the ergodic theorems of classical statistical mechanics than the quantal ergodic theorems of von Neumann and Pauli–Fierz (for the latter, see, e.g., Ref. 2 or 3).

In the classical case, if f is a function on the phase space Ω of a conservative mechanical system, if $\omega \to \omega_t$ is the temporal transformation in time t of a point $\omega \in \Omega$, if $U_t f(\omega) = f(\omega_t)$, and if

$$f_t = \frac{1}{t}\int_0^t U_s f(\omega)\, ds,$$

then the von Neumann or L_2 mean ergodic theorem states that, if $f \in L_2(\Omega, \mu)$ (μ the Lebesgue measure), then f_t converges in the $L_2(\mu)$-norm topology to a function $f_0 \in L_2(\Omega, \mu)$ which is *invariant* under the transformation U_t ($U_t f_0 \equiv f_0$). If ρ is a finite invariant measure on Ω and $f \in L_1(\Omega, \rho)$, then the individual ergodic theorem states that f_t converges pointwise almost everywhere (ρ) (i.e., except possibly on a set of ρ measure zero) and also in the $L_1(\rho)$-norm topology to an invariant function $f_0 \in L_1(\Omega, \rho)$; here, it is the convergence in L_1 norm (which we may call the L_1 mean ergodic theorem) which generalizes to the quantum case.

For quantum systems with a Hilbert state-space $\mathcal{H}$, the analogs of the classical L_2 and L_1 spaces above one, are, respectively, the Hilbert–Schmidt algebra $\mathcal{S}$ and the trace-class algebra $\mathcal{C}$ of compact linear operators on $\mathcal{H}$ (see Ref. 4 for the relevant theory). A compact operator T admits the polar representation

$$T = \sum_i \alpha_i \phi_t \otimes \psi_i^*, \tag{1}$$

where $\phi \otimes \psi^*$ is the tensor product of $\phi \in \mathcal{H}$ and $\psi^* \in \mathcal{H}^*$, the α_i are the eigenvalues of $|T| = (T^*T)^{\frac{1}{2}}$ (hence $\alpha_i \geq 0$), the ψ_i are the corresponding eigenfunctions and $\phi_i = W\psi_i$, where $T = W|T|$ and W is partially isometric. Then $\mathcal{S}$ consists precisely of those compact T such that $\sum_i \alpha_i^2 < \infty$, $\mathcal{C}$ of those such that $\sum_i \alpha_i < \infty$ (hence $\mathcal{C} \subset \mathcal{S}$), and both are Banach algebras when equipped, respectively, with the Hilbert–Schmidt norm $\|T\|_\sigma = (\sum \alpha_i^2)^{\frac{1}{2}} = [\mathrm{Tr}\,(T^*T)]^{\frac{1}{2}}$ or the trace norm $\|T\|_\tau = \sum \alpha_i = \mathrm{Tr}\,|T|$ (where Tr denotes the trace). Furthermore, $\mathcal{S}$ is also a Hilbert space when equipped with the inner product $(T, S) = \mathrm{Tr}\,(T^*S)$ (i.e., $\mathcal{S}$ is an H^* algebra).

Consider now a quantum system with Hamiltonian H. The temporal transformation of any $T \in \mathfrak{L}(\mathcal{H})$, the algebra of all bounded linear operators on $\mathcal{H}$, is (in the Heisenberg representation)

$$T_t = e^{itH}Te^{-itH} = U_t^* T U_t = K_t T. \tag{2}$$

Theorem 1: $\{K_t\}$ is a one-parameter isometric positive group of linear transformations on either $\mathcal{S}$ or $\mathcal{C}$: i.e., $\|K_tT\|_\sigma \equiv \|T\|_\sigma$, $\|K_tT\|_\tau \equiv \|T\|_\tau$ and $K_tT \geq 0$ for all t if $T \geq 0$.

Proof: It is obvious that $K_{t+s} = K_tK_s = K_sK_t$. The rest follows immediately from the polar representation (1) of T: thus, if $T \in \mathcal{S}$, then

$$K_tT = \sum_i \alpha_i (U_t\phi_i) \otimes (U_t\psi_i)^*;$$

since $\{U_t\psi_i\}$, $\{U_t\phi_i\}$ are orthonormal sets, $K_tT \in \mathcal{S}$ and $\|K_tT\|_\sigma = (\sum_i \alpha_i^2)^{\frac{1}{2}} = \|T\|_\sigma$. Similarly, if $T \in \mathcal{C}$, then $K_tT \in \mathcal{C}$ and $\|K_tT\|_\tau = \sum_i \alpha_i = \|T\|_\tau$. Furthermore, $T \geq 0$ if and only if $\phi_i = \psi_i$ in (1); hence if $T \geq 0$, then obviously $K_tT \geq 0$. Q.E.D.

Theorem 2: The one-parameter group $\{K_t\}$ is strongly continuous on either $\mathcal{S}$ or $\mathcal{C}$ (considered as Banach spaces).

Proof: Let us introduce the common notation $\|T\|_{\tau\sigma}$ to mean either $\|T\|_\tau$ or $\|T\|_\sigma$, while $\|T\|$ is the usual operator norm of T [considered as an element of $\mathfrak{L}(\mathcal{H})$]. It is sufficient to prove that $\|(K_t - I)T\|_{\tau\sigma} \to 0$ as $t \to 0$ when T is self-adjoint, for it is known (cf. Ref. 5) that this implies that $\|(K_{t+h} - K_t)T\|_{\tau\sigma} \to 0$ as

* Work performed under the auspices of the United States Atomic Energy Commission.

[1] P. R. Halmos, *Lectures on Ergodic Theory* (The Mathematical Society of Japan, Tokyo, 1956).

[2] R. Jancel, *Les fondements de la mécanique statistique classique et quantique* (Gauthiers-Villars, Paris, 1963).

[3] I. E. Farquhar, *Ergodic Theory in Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1964).

[4] R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, 1960).

[5] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* (American Mathematical Society, Providence, R.I., 1957).

$h \to 0$, and we can apply the result to the Hermitian components of T in the general case. Using the polar representation (1), we find that if $T^* = T$, then (cf. Ref. 4):

$$\begin{aligned}\|(K_t - I)T\|_{\tau\sigma} &\leq \|(U_t^* - I)T\|_{\tau\sigma}\,\|U_t\| + \|T(U_t - I)\|_{\tau\sigma} \\ &= 2\,\|(U_t^* - I)T\|_{\tau\sigma} \leq \sum_i \alpha_i\,\|(U_t^* - I)\phi_i \otimes \psi_i^*\|_{\tau\sigma} \\ &= \sum_i \alpha_i\,\|(U_t^* - I)\phi_i\|, \end{aligned} \tag{3}$$

because $\|\phi \otimes \psi^*\|_{\tau\sigma} = \|\phi\|\,\|\psi\| = 1$. Since $\{U_t\}$ is a strongly continuous one-parameter group of unitary transformations on $\mathcal{H}$, $\|(U_t - I)\phi_i\| \to 0$ as $t \to 0$. Since $\|(U_t - I)\phi_i\| \leq 2$, we can choose n such that for every $\epsilon > 0$,

$$\sum_{k \geq n} \alpha_k\,\|(U_t^* - I)\phi_k\| \leq 2\sum_{k \leq n} \alpha_k < \frac{\epsilon}{2}$$

uniformly in t, and then choose t_0 such that for $t \leq t_0$, $\sum_1^n \alpha_i\,\|(U_t^* - I)\phi_i\| < \epsilon/2$. This proves that the last expression in (3) converges to 0 as $t \to 0$. Q.E.D.

It follows that we can apply the theory of parametric semigroups of linear transformations on a Banach space (cf. Ref. 5) to $\{K_t\}$ operating on either $\mathcal{S}$ or $\mathcal{C}$. Thus, $\{K_t\}$ has an infinitesimal generator iD with dense domain $\mathfrak{D}(D)$: i.e., we can write

$$K_t = e^{itD}. \tag{4}$$

It follows at once from (2) that, for all $T \in \mathfrak{D}(D)$,

$$DT = HT - TH; \tag{5}$$

thus D is a *derivation*: i.e., if also $S \in \mathfrak{D}(D)$ and $TS \in \mathfrak{D}(D)$, then

$$D(TS) = (DT)S + T(DS). \tag{6}$$

We also know that the spectrum of D is real (because $\{K_t\}$ is a group: cf. Ref. 5).

The point spectrum of D may be characterized as follows (T below may be considered either as an element of $\mathcal{S}$ or of $\mathcal{C}$):

Theorem 3: T is an eigenfunction of D: i.e., $DT = \lambda T$, if and only if its polar representation is of the form

$$T = \sum_i \alpha_i \phi_i \otimes \psi_i^*, \tag{7}$$

where, for each i, ϕ_i and ψ_i are eigenfunctions of H corresponding respectively to eigenvalues E_i, E_i' such that $E_i - E_i' = \lambda$.

Proof: It follows from (5) that $DT^* = (DT)^*$; hence, from (6), if $DT = \lambda T$ and $DS = \eta S$, then $D(TS^*) = (\lambda - \eta)TS^*$. Therefore, if the polar representations of T is (7), then

$$D(TT^*) = \sum_i \alpha_i^2 D(\phi_i \otimes \phi_i^*) = 0$$

and

$$D(T^*T) = \sum_i \alpha_i^2 D(\psi_i \otimes \psi_i^*) = 0.$$

Hence $D(\phi_i \otimes \phi_i^*) = D(\psi_i \otimes \psi_i^*) = 0$ for all i, which implies that the ϕ_i and ψ_i are eigenfunctions of H. Let $H\phi_i = E_i\phi_i$ and $H\psi_i = E_i'\psi_i$; then

$$DT = \lambda T = \sum_i \alpha_i(E_i - E_i')\phi_i \otimes \psi_i^*.$$

Hence $DT\psi_i = \lambda\alpha_i\phi_i = \alpha_i(E_i - E_i')\phi_i$, which proves that $E_i - E_i' = \lambda$. The converse is obvious. Q.E.D.

We shall call T *invariant* if $K_tT \equiv T$ and denote by $\mathcal{N}(D)$ the null-space of D: i.e., $\mathcal{N}(D) = \{T \mid T \in \mathfrak{D}(D)$ and $DT = 0\}$. It follows immediately from Theorem 3 that:

Corollary 1: The following statements are equivalent:

(1) T is invariant.

(2) $T \in \mathcal{N}(D)$.

(3) $T = \sum_i \alpha_i\phi_i \otimes \psi_i^*$, where ϕ_i and ψ_i are eigenfunctions of H corresponding to the same eigenvalue E_i.

Corollary 2: If $T \geq 0$ (and in particular if $T = \rho$, a density matrix) then its polar representation is $T = \sum_i \alpha_i\phi_i \otimes \phi_i^*$, and T is invariant if and only if the ϕ_i are eigenfunctions of H.

Since $\mathcal{S}$ is a Hilbert space, von Neumann's mean ergodic theorem applies (see Ref. 1). Let $\overline{\mathcal{R}}(D)$ denote the closure in $\mathcal{S}$ of the range of D; then obviously $\mathcal{N}(D)$ and $\overline{\mathcal{R}}(D)$ are closed linear subspaces of $\mathcal{S}$.

Theorem 4: If $T \in \mathcal{S}$, then there exists a unique invariant element $T_0 \in \mathcal{N}(D)$ such that

$$\lim_{t \to \infty} \left\| \frac{1}{t}\int_0^t K_sT\,ds - T_0 \right\|_\sigma = 0.$$

The operator P defined by $PT = T_0$ is the projection on $\mathcal{N}(D)$: $P = P^2$, $\mathcal{R}(P) = \mathcal{N}(D)$ and $\mathcal{N}(P) = \overline{\mathcal{R}}(D)$; i.e., $\mathcal{S}$ admits the direct-sum decomposition

$$\mathcal{S} = \mathcal{N}(D) \oplus \overline{\mathcal{R}}(D).$$

In other words, the linear operator

$$\frac{1}{t}\int_0^t K_s\,ds,$$

on $\mathcal{S}$, converges strongly to P as $t \to 0$.

We now turn to the analog of the classical L_1 mean ergodic theorem. The role of the invariant measure ρ in the latter is played here by an invariant density matrix $\boldsymbol{\rho}$, and the classical convergence of the time average $\bar{T}_t$ in the $L_1(\Omega, \rho)$-norm topology to an invariant function f_0 is replaced by convergence in the trace-norm topology of $\bar{T}_t\boldsymbol{\rho}$ to $T_0\boldsymbol{\rho}$, where T is any bounded linear operator,

$$\bar{T}_t = t^{-1}\int_0^t K_s T\, ds,$$

and T_0 is invariant ($K_t T_0 \equiv T_0$); we make use in proving this result of the facts that, if $T \in \mathfrak{L}(\mathcal{H})$, then $T\boldsymbol{\rho} \in \mathfrak{T}$, and that, since $\boldsymbol{\rho}$ is invariant, $K_t(T\boldsymbol{\rho}) = (K_t T)\boldsymbol{\rho}$.

Theorem 5: Let $T \in \mathfrak{L}(\mathcal{H})$ and let $\boldsymbol{\rho}$ be any invariant density matrix. Then there exists an invariant element $T_0 \in \mathfrak{L}(\mathcal{H})$, such that

$$\lim_{t\to\infty} \left\| \left(\frac{1}{t}\int_0^t K_s T\, ds\right)\boldsymbol{\rho} - T_0\boldsymbol{\rho} \right\|_\tau = 0. \tag{8}$$

Furthermore, T_0 is uniquely defined modulo $\boldsymbol{\rho}$ in the sense that if T_0' is another such ergodic limit, then $(T_0 - T_0')\boldsymbol{\rho} = 0$ for all invariant $\boldsymbol{\rho}$; in fact, if P is the projection on the linear subspace of $\mathcal{H}$ spanned by the eigenfunctions of H, then $T_0 = T_0' P$.

Proof: We use the following terminology: if $\{T_n\} \subset \mathfrak{T}$, we say that $\{T_n\}$ converges (σ) to $T_0[T_n \to T_0(\sigma)]$ if $\|T_n - T_0\|_\sigma \to 0$, and that it converges (τ) to $T_0[T_n \to T_0(\tau)]$ if $T_0 \in \mathfrak{T}$ and $\|T_n - T_0\|_\tau \to 0$. Let $S_t = (1/t)\int_0^t K_s T\, ds$, let $\{\phi_i\}$ denote the system of all eigenfunctions of H and let $\pi_i = \phi_i \otimes \phi_i^*$. By Theorem 4, $S_t\pi_k \to S_k(\sigma)$, but $S_t\pi_k = S_t\pi_k^2 \to S_k\pi_k$, hence $S_k\pi_k = S_k$. Since $S_k \in \mathcal{N}(D)$, by Corollary 1 of Theorem 3, $S_k = \sum_i \gamma_i \psi_i \otimes \phi_i^*$, where ψ_i is an eigenfunction of H belonging to the same eigenmanifold as ϕ_i; hence $S_k\pi_k = \gamma_k\psi_k \otimes \phi_k^* = S_k$. If $\boldsymbol{\rho}$ is an invariant density matrix, then, by Corollary 2 of Theorem 3, $\boldsymbol{\rho} = \sum_i \alpha_i\pi_i$, and $S_t\boldsymbol{\rho} = \sum_i \alpha_i S_t\pi_i$, with the series converging (τ) uniformly in t, since

$$\left\| \sum_n^{n+k} \alpha_i S_t \pi_i \right\|_\tau \leq \sum_n^{n+k} \alpha_i \|T\|.$$

Hence,

$$\lim_{t\to\infty}(\tau) S_t\boldsymbol{\rho} = S(\boldsymbol{\rho}) = \sum_i \alpha_i \lim_{t\to\infty}(\tau) S_t\pi_i = \sum_i \alpha_i\gamma_i\psi_i \otimes \phi_i^*;$$

i.e., $S_t\boldsymbol{\rho}$ converges (τ) to $S(\boldsymbol{\rho})$ and, hence, $S(\boldsymbol{\rho}) = \sum_i \alpha_i\gamma_i\psi_i \otimes \phi_i^* \in \mathfrak{T}$, which implies that $\sum_i \alpha_i |\gamma_i| < \infty$. This must be true for every positive sequence $\{\alpha_i\}$ such that $\sum_i \alpha_i < \infty$, since $\boldsymbol{\rho}$ can be any positive element of $\mathcal{N}(D)$. Thus $\{\gamma_i\} \in l_1^* = l_\infty$, the dual of the sequence-space l_1, and is therefore bounded. Let $T_0 = \sum_i \gamma_i\psi_i \otimes \phi_i^*$: i.e., T_0 is the bounded linear operator such that, for every $\psi \in \mathcal{H}$,

$$T_0\psi = \sum_i \gamma_i(\phi_i, \psi)\psi_i$$

(cf. Ref. 4). It is then simple to verify that $T_0\boldsymbol{\rho} = S(\boldsymbol{\rho})$, which proves (8). It is also easy to see that T_0 is invariant: for it follows from the expression above for T_0 that, if $\psi \in \mathcal{H}$, then

$$\begin{aligned}(K_t T_0)\psi = e^{itH}T_0 e^{-itH}\psi &= \sum_j \gamma_j(\phi_j, e^{-itH}\psi)(e^{itH}\psi_j) \\ &= \sum_j \gamma_j(e^{itH}\phi_j, \psi)(e^{itH}\psi_j) \\ &= \sum_j \gamma_j(\phi_j, \psi)\psi_j = T_0\psi,\end{aligned}$$

since $e^{itH}\phi_j = e^{itE_j}\phi_j$ and $e^{itH}\psi_j = e^{itE_j}\psi_j$. Finally, if $S_t\boldsymbol{\rho} \to T_0'\boldsymbol{\rho}$ for all invariant $\boldsymbol{\rho}$, then clearly $(T_0 - T_0')\boldsymbol{\rho} = 0$ for all such $\boldsymbol{\rho}$ and, in particular, $(T_0 - T_0')\phi_i \otimes \phi_i^* = 0$ for all i; hence, if P is the projection defined in the Theorem 5, then

$$(T_0 - T_0')P = 0$$

and, since obviously $T_0 P = T_0$, we have $T_0 = T_0' P$. Q.E.D.

Another application of the abstract ergodic theorems, which we describe only briefly, is feasible in the case of algebras of quantum observables which admit a phase-space representation (cf. Ref. 6). The *Weyl correspondence*[7] W maps functions g on the classical phase space $\Omega = R^{2n}$ onto linear operators $G = W(g)$ in a suitable representation of the canonical commutation relations: e.g., as operators on $\mathcal{H}_\Omega = L_2(R^n)$. This correspondence can be extended to boson fields (cf. Refs. 8 and 9): for example, using the Fock–Cook representation (cf. Ref. 10), W maps symmetric functions g on $\Omega = \bigcup_{n=0} R^{6n}$ onto linear operators $G = W(g)$ on $\mathcal{H}_\Omega = \oplus_{n=0}^\infty \mathcal{H}^n$, where $\mathcal{H}^n$ is the n-fold tensor product with itself of $\mathcal{H} = L_2(R^3)$. (This representation is the natural setting for the use of the grand canonical ensemble.) The time transformation $K_t G = e^{itH}Ge^{-itH}$ induces a transformation (denoted by the same symbol K_t) on $g = W^{-1}(G)$, whose infinitesimal generator D can be written explicitly for a wide class of C^∞ functions (cf. Refs.

[6] J. E. Moyal, Proc. Cambridge Phil. Soc. **45**, 99 (1949).
[7] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Methuen & Co., Ltd., London, 1931).
[8] I. E. Segal, *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, R.I., 1963).
[9] D. Kastler, Commun. Math. Phys. **1**, 14 (1965).
[10] J. M. Cook, Trans. Am. Math. Soc. **74**, 222 (1953).

11 and 12) in the form

$$D_g = (2/\hbar)\sin(\hbar/2)\{\partial_{p_1}\partial_{q_2} - \partial_{q_2}\partial_{p_1}\}(\hat{H}g),$$

where $\hat{H} = W^{-1}(H)$ and $\{\partial_{p_1}\partial_{q_2} - \partial_{q_2}\partial_{p_1}\}(\hat{H}g) = \{\hat{H},g\}$, the classical Poisson bracket.

If $g \in L_2(\Omega)$, then it is easy to see that K_t acting on g satisfies the conditions of the classical L_2 mean ergodic theorem. However, it is known (cf. Ref. 13) that W maps $L_2(\Omega)$ isometrically (apart from a normalization constant) onto the Hilbert–Schmidt class $\mathcal{S}$, so that all one obtains in this case is again the statement of Theorem 4.

Suppose, however, that ρ is a finite probability measure on Ω which is invariant under K_t, and suppose that $g \in L_1(\Omega, \rho)$; then it can be seen that K_t satisfies the conditions of the individual ergodic theorem (and hence the L_1 mean ergodic theorem), so that

$$\frac{1}{t}\int_0^t K_s g\, ds$$

converges almost everywhere (ρ) to a limit $g_0 \in L_1(\Omega, \rho)$, and

$$\lim_{t\to\infty}\int_\Omega \left|\frac{1}{t}\int_0^t K_s g\, ds - g_0\right| d\rho(\omega) = 0.$$

This result is valid for boson fields using the Fock–Cook representation.

[11] T. F. Jordan and E. C. G. Sudarshan, Rev. Mod. Phys. 33, 515 (1961).
[12] A. Grossmann, G. Loupias, and E. M. Stein, "An Algebra of Pseudodifferential Operators and Quantum Mechanics in Phase-Space," Ann. Inst. Fourier (to be published).
[13] J. C. T. Pool, J. Math. Phys. 7, 66 (1966).

ACKNOWLEDGMENTS

I am much indebted to Dr. J. M. Cook and Dr. J. C. T. Pool for stimulating discussions and criticism.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Bound States of a New Kind in Quantum Field Theory*

ROBERT L. ZIMMERMAN
Institute of Theoretical Science and Department of Physics, University of Oregon, Eugene, Oregon

(Received 11 March 1968)

It is the purpose of this article to study the various kinds of bound states and their properties that can occur in quantum field theory. The ability of quantum field theory to create and destroy particles makes the picture of a bound state very complex. However, for this very reason it allows for a larger variety of bound states than can occur in nonrelativistic quantum mechanics. Not only do we consider those bound states that appear as poles in the tau functions,but also as branch points. The bound states that appear as branch points offer many new possibilities in constructing composite theories. An example of this type of bound state is given in which the fermion is a bound state of a boson. The many ambiguities that exist with this kind of bound state are illustrated. In discussing the bound states that appear as poles we have broken them into three subgroups: N-body bound states, mixed bound states, and self-interacting bound states. The N-body bound states are the most analogous to those occurring in nonrelativistic quantum mechanics. The mixed bound states have the peculiar property of not being composed of any unique group of elementary particles. The self-interacting bound states are those that are directly coupled with themselves. The characteristics of these subgroups of bound states are discussed and illustrated.

I. INTRODUCTION

In previous articles[1–3] a method was developed for constructing Lagrangians with composite fields that gave equivalent results to theories where the fields were elementary. This method is very useful in studying the various properties of composite fields since comparison can be made with elementary fields that have similar characteristics. It is the purpose of this paper to study the various kinds of bound states and their properties, relying strongly on this method of comparing them with elementary fields of a similar nature.

We restrict our discussion to Lagrangian field theories, implicitly assuming that it satisfies the assumptions of axiomatic field theory.

In the context of quantum field theory only a small class of bound states has been considered. Essentially, they are those bound states which manifest themselves as poles in the various tau functions. These are the most obvious kind because of the analogous type of

* Supported in part by the Air Force Office of Scientific Research, Grant No. AF-AFOSR 1296-67.
[1] R. L. Zimmerman, Phys. Rev. **141**, 1554 (1966).
[2] R. L. Zimmerman, Phys. Rev. **146**, 955 (1966).
[3] R. L. Zimmerman, Phys. Rev. **164**, 1945 (1967).

11 and 12) in the form

$$D_g = (2/\hbar)\sin(\hbar/2)\{\partial_{p_1}\partial_{q_2} - \partial_{q_2}\partial_{p_1}\}(\hat{H}g),$$

where $\hat{H} = W^{-1}(H)$ and $\{\partial_{p_1}\partial_{q_2} - \partial_{q_2}\partial_{p_1}\}(\hat{H}g) = \{\hat{H},g\}$, the classical Poisson bracket.

If $g \in L_2(\Omega)$, then it is easy to see that K_t acting on g satisfies the conditions of the classical L_2 mean ergodic theorem. However, it is known (cf. Ref. 13) that W maps $L_2(\Omega)$ isometrically (apart from a normalization constant) onto the Hilbert–Schmidt class $\mathcal{S}$, so that all one obtains in this case is again the statement of Theorem 4.

Suppose, however, that ρ is a finite probability measure on Ω which is invariant under K_t, and suppose that $g \in L_1(\Omega, \rho)$; then it can be seen that K_t satisfies the conditions of the individual ergodic theorem (and hence the L_1 mean ergodic theorem), so that

$$\frac{1}{t}\int_0^t K_s g\, ds$$

converges almost everywhere (ρ) to a limit $g_0 \in L_1(\Omega, \rho)$, and

$$\lim_{t\to\infty}\int_\Omega \left|\frac{1}{t}\int_0^t K_s g\, ds - g_0\right| d\rho(\omega) = 0.$$

This result is valid for boson fields using the Fock–Cook representation.

[11] T. F. Jordan and E. C. G. Sudarshan, Rev. Mod. Phys. 33, 515 (1961).
[12] A. Grossmann, G. Loupias, and E. M. Stein, "An Algebra of Pseudodifferential Operators and Quantum Mechanics in Phase-Space," Ann. Inst. Fourier (to be published).
[13] J. C. T. Pool, J. Math. Phys. 7, 66 (1966).

ACKNOWLEDGMENTS

I am much indebted to Dr. J. M. Cook and Dr. J. C. T. Pool for stimulating discussions and criticism.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Bound States of a New Kind in Quantum Field Theory*

ROBERT L. ZIMMERMAN
Institute of Theoretical Science and Department of Physics, University of Oregon, Eugene, Oregon

(Received 11 March 1968)

It is the purpose of this article to study the various kinds of bound states and their properties that can occur in quantum field theory. The ability of quantum field theory to create and destroy particles makes the picture of a bound state very complex. However, for this very reason it allows for a larger variety of bound states than can occur in nonrelativistic quantum mechanics. Not only do we consider those bound states that appear as poles in the tau functions,but also as branch points. The bound states that appear as branch points offer many new possibilities in constructing composite theories. An example of this type of bound state is given in which the fermion is a bound state of a boson. The many ambiguities that exist with this kind of bound state are illustrated. In discussing the bound states that appear as poles we have broken them into three subgroups: N-body bound states, mixed bound states, and self-interacting bound states. The N-body bound states are the most analogous to those occurring in nonrelativistic quantum mechanics. The mixed bound states have the peculiar property of not being composed of any unique group of elementary particles. The self-interacting bound states are those that are directly coupled with themselves. The characteristics of these subgroups of bound states are discussed and illustrated.

I. INTRODUCTION

In previous articles[1–3] a method was developed for constructing Lagrangians with composite fields that gave equivalent results to theories where the fields were elementary. This method is very useful in studying the various properties of composite fields since comparison can be made with elementary fields that have similar characteristics. It is the purpose of this paper to study the various kinds of bound states and their properties, relying strongly on this method of comparing them with elementary fields of a similar nature.

We restrict our discussion to Lagrangian field theories, implicitly assuming that it satisfies the assumptions of axiomatic field theory.

In the context of quantum field theory only a small class of bound states has been considered. Essentially, they are those bound states which manifest themselves as poles in the various tau functions. These are the most obvious kind because of the analogous type of

* Supported in part by the Air Force Office of Scientific Research, Grant No. AF-AFOSR 1296-67.
[1] R. L. Zimmerman, Phys. Rev. **141**, 1554 (1966).
[2] R. L. Zimmerman, Phys. Rev. **146**, 955 (1966).
[3] R. L. Zimmerman, Phys. Rev. **164**, 1945 (1967).

bound state that occurs in nonrelativistic quantum mechanics. In nonrelativistic quantum mechanics we can interpret these bound states as being composed of N elementary particles held together by an attractive force. The classical examples are atoms and molecules which are composed of the nucleus and electrons—the elementary constituents bound together by an attractive Coulomb force.

In quantum field theory the physical picture that a bound state is composed of N elementary particles bound by an attractive force is no longer valid. This is so for the following reasons:

(a) The binding force is due to the exchange of particles.

(b) The N elementary particles will each be engulfed in a virtual cloud of particles.

(c) The N elementary particles can virtually transform into other kinds of particles.

The picture of a bound state becomes very hazy. It is no longer a trivial task to tell whether or not a bound state is composed of N particles. For example, to say that the pion is a two-body bound state of a nucleon and an antinucleon, or that the deuteron is a two-body bound state of two nucleons, is no longer a trivial problem. This observation *cannot* be accomplished by observing which tau functions contain the various singularities that are due to the bound states.

The ability of quantum field theory to create and destroy particles has caused the interpretation of bound states to become more complex. However, for this same reason we can consider a much larger class of bound states that do not appear in nonrelativistic quantum mechanics. In particular, it may very well be the case that the bound states manifest themselves as branch points rather than poles in the tau functions. This new kind of bound state offers many new possibilities in constructing composite theories. Up to now, the spin-$\frac{1}{2}$ fields have occupied a unique position in that they were regarded as being more elementary than, for example, the boson fields. The reason for this is that one could in principle always make the integer and higher-spin fermi fields appear as bound states of the spin-$\frac{1}{2}$ field. To go the other way was not possible. One explicit example of such a theory is the Heisenberg nonlinear theory.[4] If one allows for this much wider class of bound states, then it is possible to construct theories where the spin-$\frac{1}{2}$ fields appear as bound states of, for example, spin-zero fields. This removes the spin-$\frac{1}{2}$ field from its unique position of being the most fundamental field. This is very appealing, since nature in no way has conveyed to us that the spin-$\frac{1}{2}$ fields should occupy such a prominent role.

Bound states in quantum field theory have been discussed from numerous points of view. Nishijima,[5] Zimmerman,[6] Haag,[7] and others have studied the Heisenberg field operators for the bound states. Gell-Mann and Low[8] constructed the Bethe–Salpeter amplitudes, following which numerous articles have been devoted toward understanding the properties of bound states. Recently the vanishing of Z_3 (i.e., the wavefunction renormalization constant) has also been linked to the study of bound states.[9] In all of these cases only those bound states that appeared as poles in the tau functions were considered. This is only a small class of bound states that can occur.

Before we continue, a few words should be said on what we mean by a composite particle. A composite or elementary particle is not a fundamental concept of nature. That is to say, nature does not have any intrinsic property that gives us an absolute definition of whether a particle is composite or elementary. The concept arises from the interpretation of the various theories and formalisms that are used to describe nature. Therefore the concepts are functions of the formalism and theories and not an inherent property of nature. It may very well be that there are many equivalent formalisms that can characterize our universe. In one formalism a particle could be considered elementary, whereas in another it may be composite. In the case of a Lagrangian field theory the meaning of an elementary and a composite field is very clear. We say a field is elementary if it explicitly appears in the Lagrangian. If these elementary fields do not form an irreducible set of operators, we must add some new fields to make the set irreducible. These additional fields we call the composite fields. This definition is unquestionably a function of the formalism, as it must be. In fact, it has been shown in a previous paper[1] that a particle may be considered elementary for a particular Lagrangian, whereas an equivalent theory could be constructed in which the particle is composite.

The outline of the paper is as follows. In Sec. II we define the notation and review some of the well-known relations needed in discussing the properties of composite particles. The most familiar class of bound states is discussed in Sec. III. We refer to this class as bound states of the first kind. This class of

[4] H. P. Durr, W. Heisenberg, H. Mitter, S. Schlieder, and K. Yamazaki, Z. Naturforsch. **14a**, 441 (1959).

[5] K. Nishijima, Progr. Theoret. Phys. (Kyoto) **10**, 549 (1953); **12**, 274 (1954); **13**, 305 (1955).

[6] W. Zimmerman, Nuovo Cimento **10**, 597 (1958).

[7] R. Haag, Phys. Rev. **112**, 669 (1958).

[8] M. Gell-Mann and F. E. Low, Phys. Rev. **84**, 350 (1951).

[9] B. Jouvet, Nuovo Cimento Suppl. **2**, 941 (1955).

bound states manifests itself as poles in the various tau functions. Included in this class are those bound states that are analogous to the ones in nonrelativistic quantum mechanics. However, the bound states of the first kind also contain less familiar types of bound states and are broken up into subclasses exhibiting common characteristics. Examples of Lagrangians and Heisenberg bound-state operators are given for the various kinds of bound states in these classes. In Sec. IV we discuss that class of bound states which we refer to as bound states of the Nth kind ($N > 1$). This class of bound states manifests itself as branch points in the tau functions. An example of this class of bound states is given, where the spin-$\frac{1}{2}$ particle appears as a bound state of a spin-zero particle. Section V concludes with the summary and conclusion.

II. NOTATION AND BASIC RELATIONS

In this section we introduce the notation and well-known relations among the interpolating and asymptotic fields. This is done for both elementary and composite fields.

Consider the elementary interpolating field $\phi_i(x)$. In order to relate $\phi_i(x)$ to an asyn.ptotic field we define

$$\phi_i^f(t) = i\int d^3x f^*(x,t)\overleftrightarrow{\partial}_0\phi_i(x,t), \tag{1a}$$

$$\phi_i^U(t) = i\int d^3x U_{PS}^\dagger(x,t)\phi_i(x,t), \tag{1b}$$

$$\phi_i^V(t) = i\int d^3x \phi_i^\dagger(x,t)V_{PS}(x,t), \tag{1c}$$

using (1a) if $\phi_i(x,t)$ is a scalar field and (1b) or (1c) if it is a spin-$\frac{1}{2}$ field. $\phi_i^\dagger(x,t)$ is the Hermitian conjugate of $\phi_i(x,t)$. $f_k(x,t)$, $U_{PS}(x,t)$, and $V_{PS}(x,t)$ are normalized solutions of the free field equations

$$(\Box + m_i^2)f_k(x,t) = 0, \tag{2a}$$

$$(i\not{\partial} - m_i)\begin{Bmatrix} U_{PS}(x,t) \\ V_{PS}(x,t)\end{Bmatrix} = 0. \tag{2b}$$

The $U_{PS}(x,t)$ corresponds to the positive-energy solutions and $V_{PS}(x,t)$ corresponds to the negative-energy ones. For all practical purposes we can limit ourselves to the plane-wave solutions

$$f_k(x,t) = [(2\pi)^3 2\omega_k]^{-\frac{1}{2}}e^{-ik^\mu x_\mu}, \tag{3a}$$

$$U_{PS}(x,t) = [m_i/(2\pi)^3 E_p]^{\frac{1}{2}}u(P,S)^{-ip^\mu x_\mu}, \tag{3b}$$

$$V_{PS}(x,t) = [m_i/(2\pi)^3 E_p]^{\frac{1}{2}}v(P,S)^{+ip^\mu x_\mu}, \tag{3c}$$

where

$$(\not{P} - m_i)u(P,S) = 0,$$
$$(\not{P} + m_i)v(P,S) = 0, \tag{4}$$

and

$$\omega_k = (\mathbf{k}^2 + m_i^2)^{\frac{1}{2}}.$$

Assuming the asymptotic condition to be valid, for any two arbitrary eigenstates $|\alpha\rangle$ and $|\beta\rangle$ we have

$$\lim_{t\to\{^{+\infty}_{-\infty}\}}\langle\alpha|\,\phi_i^f(t)\,|\beta\rangle = \sqrt{Z_i}\,\langle\alpha|\,\phi^f_{i\{^{\text{in}}_{\text{out}}\}}\,|\beta\rangle, \tag{5}$$

where Z_i is the wavefunction renormalization constant of $\phi_i(x,t)$. $\phi^f_{i\{^{\text{in}}_{\text{out}}\}}$ are defined by the relation

$$\phi^f_{i\{^{\text{in}}_{\text{out}}\}} = i\int d^3x f^*(x,t)\overleftrightarrow{\partial}_0\phi_{i\{^{\text{in}}_{\text{out}}\}}(x,t) \quad \text{(scalar fields)}, \tag{6a}$$

$$\phi^U_{i\{^{\text{in}}_{\text{out}}\}} = \int d^3x U^\dagger_{PS}(x,t)\phi_{i\{^{\text{in}}_{\text{out}}\}}(x,t) \quad \text{(spin-}\tfrac{1}{2}\text{ fields)},$$

$$\phi^V_{i\{^{\text{in}}_{\text{out}}\}} = \int d^3x \phi^\dagger_{i\{^{\text{in}}_{\text{out}}\}}(x,t)V_{PS}(x,t) \quad \text{(spin-}\tfrac{1}{2}\text{ fields)}, \tag{6b}$$

where the in and out fields $\phi_{i\{^{\text{in}}_{\text{out}}\}}(x,t)$ satisfy the free field equations with the physical mass of the particles

$$(\Box + m_i^2)\phi_{i\{^{\text{in}}_{\text{out}}\}}(x,t) = 0 \quad \text{(scalar field)}, \tag{7a}$$

$$(i\not{\partial} - m_i)\phi_{i\{^{\text{in}}_{\text{out}}\}}(x,t) = 0 \quad \text{(spin-}\tfrac{1}{2}\text{ field)}. \tag{7b}$$

Equivalently, we may express the fields $\phi_{i\{^{\text{in}}_{\text{out}}\}}(x,t)$ by the relation

$$\phi_{i\{^{\text{in}}_{\text{out}}\}}(x,t) = \int d^3k[\phi^f_{i\{^{\text{in}}_{\text{out}}\}}f_k(x,t) + \phi^{f+}_{i\{^{\text{in}}_{\text{out}}\}}f_k^*(x,t)] \quad \text{(scalar fields)}, \tag{8a}$$

$$\phi_{i\{^{\text{in}}_{\text{out}}\}}(x,t) = \int d^3P\Sigma_s[\phi^U_{i\{^{\text{in}}_{\text{out}}\}}U_{P,S}(x,t) + \phi^{V+}_{i\{^{\text{in}}_{\text{out}}\}}V_{P,S}(x,t)] \quad \text{(spin-}\tfrac{1}{2}\text{ fields)}. \tag{8b}$$

The eigenvectors of the physical states can be constructed by operating on the vacuum with the creation operators $\phi^{f+}_{i\{^{\text{in}}_{\text{out}}\}}$. At the loss of no generality we will work with the in states and let η denote in general either f, U, or V. Defining the physical vacuum state $|0\rangle$ in the usual manner, the one-particle state is

$$|a_i\rangle = \phi^{\eta+}_{i\,\text{in}}|0\rangle. \tag{9}$$

In general the N-particle state can be constructed

$$\phi^{\eta i+}_{j\,\text{in}}\cdots\phi^{\eta i+}|0\rangle. \tag{10}$$

If the set of operators $\{\phi^{\eta+}_{i\,\text{in}}\}$ are irreducible, the eigenstates constructed from the irreducible set will be complete. As a consequence of the completeness of the

eigenstates, we can construct the identity operator

$$I = \sum_{N=0} |a_1, \cdots, a_N\rangle\langle a_N, \cdots, a_1|, \tag{11}$$

where the sum is over the complete set of eigenstates. In this case there are no bound states. However, if such a completeness relation cannot be constructed from the elementary fields, i.e., the set $\{\phi_{\text{in}}^{\eta+}\}$ is not irreducible, then we must add a set of eigenstates $|b_i\rangle$, $|b, b_j\rangle$, $|b_i, a_j, b, \cdots\rangle$, etc., which will give us the identity operator, i.e., a complete set of states. The construction of such eigenstates is the essential problem in understanding bound states.

The eigenstates containing composite particles can be created by operating on the elementary eigenstates with the bound-state creation operator $B_{i\,\text{in}}^{\eta+}$. The creation operator $B_{i\,\text{in}}^{\eta+}$ is related to an interpolating bound-state field $B_i(x, t; \zeta_1 \cdots \zeta_n)$ by an asymptotic condition. A critical assumption is that the bound-state operator $B_i(x, t; \zeta_1 \cdots \zeta_n)$ can be expressed in terms of some function of the elementary fields $\phi_j(x, t)$. It is the form of this functional relation between $B_i(x, t; \zeta_1 \cdots \zeta_n)$ and $\phi_j(x, t)$ that will determine the properties of the bound states.

Let us now consider the relation between the bound-state operators $B(x, t; \zeta_1 \cdots \zeta_n)$ and their asymptotic fields as we have done for the elementary fields $\phi_i(x, t)$.

In terms of the interpolating fields $B_i(x, t; \zeta_1 \cdots \zeta_n)$ we define $B_i^f(t; \zeta_1 \cdots \zeta_n)$ by the equation

$$B_i^f(t; \zeta_1, \cdots, \zeta_n) = i\int d^3x f_k^*(x, t)\partial_0 B_i(x, t; \zeta_1, \cdots, \zeta_n), \tag{12a}$$

$$B_i^U(t; \zeta_1, \cdots, \zeta_n) = \int d^3x U_{P,S}^\dagger(x, t) B_i(x, t; \zeta_1, \cdots, \zeta_n),$$

$$B_i^V(t; \zeta_1, \cdots, \zeta_n) = \int d^3x B_i^\dagger(x, t; \zeta_1, \cdots, \zeta_n) V_{PS}(x, t). \tag{12b}$$

We use (12a) when bound state is spin-zero and (12b) for spin-$\frac{1}{2}$ bound states. Higher-spin bound states can be treated in a similar manner.

The asymptotic condition for the bound states becomes

$$\lim_{t\to\{^{+\infty}_{-\infty}\}} \langle\alpha|\, B_i^\eta(t)\,|B\rangle = \sqrt{Z}_i\,\langle\alpha|\, B_{i\{^{\text{in}}_{\text{out}}\}}^f\,|\beta\rangle, \tag{13}$$

whereas Z_i is the wavefunction renormalization constant for the bound-state field $B_i(x, t)$; i.e.,

$$(Z_i)^{\frac{1}{2}} = (2\pi)^{\frac{3}{2}}(2P_0)^{\frac{1}{2}}\,\langle 0|\, :B_i(0):\,|b\rangle. \tag{14}$$

$B_{i\{^{\text{in}}_{\text{out}}\}}^f$ satisfy the equation

$$B_{i\{^{\text{in}}_{\text{out}}\}}^f(\zeta_1, \zeta_2, \cdots, \zeta_N) = i\int d^3x f^*(x, t)\partial_0 B_{i\{^{\text{in}}_{\text{out}}\}}(x, t; \zeta_1, \cdots, \zeta_N)$$

(spin-zero fields), (15a)

$$B_{i\{^{\text{in}}_{\text{out}}\}}^U(\zeta_1, \cdots, \zeta_n) = \int d^3x U_{P,S}^\dagger(x, t) B_{i\{^{\text{in}}_{\text{out}}\}}(x, t; \zeta_1, \cdots, \zeta_n),$$

$$B_{i\{^{\text{in}}_{\text{out}}\}}^V(\zeta_1, \cdots, \zeta_n) = \int d^3x B_{i\{^{\text{in}}_{\text{out}}\}}^\dagger(x, t; \zeta_1, \cdots, \zeta_n) V_{P,S}(x, t)$$

(spin-$\frac{1}{2}$ fields), (15b)

and the $B_{i\{^{\text{in}}_{\text{out}}\}}$ satisfy the free field equations with the mass of the bound state

$$(\Box + m_i^2) B_{i\{^{\text{in}}_{\text{out}}\}}(x, t; \zeta_1, \cdots, \zeta_n) = 0$$

(scalar field), (16a)

$$(i\not{\nabla} - m_i) B_{i\{^{\text{in}}_{\text{out}}\}}(x, t; \zeta_1, \cdots, \zeta_n) = 0$$

(spin-$\frac{1}{2}$ field). (16b)

III. BOUND STATES OF THE FIRST KIND

The most familiar and, in fact, the only type of bound state that has been studied, with the possibility of one or two exceptions,[1,10] consists of those that manifest their presences as poles in the various tau functions of the elementary fields. We devote this section to the investigation of this class of bound states and refer to them as bound states of the first kind.

To illustrate the appearance of the pole, which is a general characteristic of this class of bound states, let us consider an explicit example. Assume that we have a local bound state of two neutral, possibly degenerate, spin-zero particles. Let us now explicitly illustrate the appearance of the pole in the simplest tau function

$$K(x_1, \cdots, x_4) = \langle 0|\, T(\phi_1(x_1)\phi_2(x_2)\phi_1(x_3)\phi_2(x_4))\,|0\rangle, \tag{17}$$

where ϕ_1 and ϕ_2 correspond to real spin-zero fields. Using the completeness relation in Eq. (11), we can expand this tau function, so that

$$K(x_1, \cdots, x_4) = \begin{cases} \sum_N \langle 0|\, T[\phi_1(x_1)\phi_2(x_2)]\,|N\rangle\langle N|\, T[\phi_1(x_1)\phi_2(x_2)]\,|0\rangle, \\ \qquad \text{for} \quad x_{10}, x_{20} > x_{30}, x_{40}, \\ \sum_N \langle 0|\, T[\phi_1(x_3)\phi_2(x_4)]\,|N\rangle\langle N|\, T[\phi_1(x_1)\phi_2(x_2)]\,|0\rangle, \\ \qquad \text{for} \quad x_{30}, x_{40} > x_{10}, x_{20}. \end{cases} \tag{18}$$

[10] B. Jouvet and J. C. LeGuillou, "Particules Autocomposées," Preprint from Laboratoire de Physique atomique et moléculaire, College de France, Paris.

If there exists a single-particle composite eigenstate $|b_i m_b, \mathbf{P}, q\rangle$ with the quantum numbers of the fields ϕ_1 and ϕ_2, then for $x_{10}, x_{20} > x_{30}, x_{40}$ we can separate the contribution from this intermediate state and Eq. (18) becomes

$$\begin{aligned} &K(x_1, x_2, x_3, x_4) \\ &= \sum_q \int d^3P \langle 0| T[\phi_1(x_1)\phi_2(x_2)] |b; q\rangle \\ &\quad \times \langle q; b| T[\phi_1(x_3)\phi_2(x_4)] |0\rangle \\ &\quad + \sum_{N \neq b} \langle 0| T[\phi_1(x_1)\phi_2(x_2)] |N\rangle \langle N| T[\phi_1(x_3)\phi_2(x_4)] |0\rangle, \end{aligned} \tag{19}$$

where q denotes the possible degenerate states. Introducing the notation for the Bethe–Salpeter amplitude

$$X_q(x_1, x_2) = \langle 0| T[\phi_1(x_1)\phi_2(x_2)] |b; q\rangle, \tag{20}$$

we have

$$X_q(x_1, x_2) = e^{-iP^\mu \bar{X}_u} F_q(x_1 - x_2), \tag{21}$$

where

$$\bar{X}_u = (m_1 x_1 + m_2 x_2)_u / (m_1 + m_2)$$

and m_1 and m_2 are the physical masses of the particles associated with the elementary fields ϕ_1 and ϕ_2, respectively.

Using the property of translational invariance on the tau functions and substituting Eq. (21) into Eq. (19) we have

$$\begin{aligned} &K(P; x, y) \\ &= \sum_q [f_{P,q}(x) f^\dagger_{P,q}(y)] / \{P_0 - [\mathbf{P}^2 + (M_0 - i)^2]^{\frac{1}{2}}\} \\ &\quad + \text{terms regular at } P_0 = [\mathbf{P}^2 + M_0^2]^{\frac{1}{2}}, \end{aligned} \tag{22}$$

where

$$K(x_1, x_2, x_3, x_4) = \int d^4P e^{iP(x-y)_u} K(P; x, y). \tag{23}$$

We see that a simple pole will occur in $K(P; x, y)$ at $P_0 = (\mathbf{P}^2 + m^2)^{\frac{1}{2}}$ if $X_q(x, y)$ is not identically zero for some q. As a consequence of this pole we can obtain an explicit equation for the Bethe–Salpeter amplitude. This follows immediately from the integral equation

$$\begin{aligned} &K(x_1, x_2, x_3, x_4) \\ &\quad = \Delta(x_1, x_3)\Delta(x_2, x_4) + \Delta(x_1, x_4)\Delta(x_2, x_3) \\ &\qquad + \int d^4\omega_1 \cdots d^4\omega_4 \Delta(x_1, \omega_1)\Delta(x_2, \omega_2) \\ &\qquad \times G(\omega_1, \omega_2, \omega_3, \omega_4) K(\omega_3, \omega_4, x_3, x_4), \end{aligned} \tag{24}$$

where

$$\Delta(x_1, x_2) = \langle 0| T(\phi_1(x_1)\phi_2(x_2)) |0\rangle \tag{25}$$

and $G(\omega_1, \omega_2, \omega_3, \omega_4)$ is the Bethe–Salpeter interaction kernel.

Multiplying Eq. (24) by $P_0 - \mathbf{P}^2 + m^2$ and letting $P_0 \to \mathbf{P}^2 + m^2$, we get a homogeneous equation for the Bethe–Salpeter amplitude

$$\begin{aligned} X(x_1, x_2) = &\int d^4\omega_3 \, d^4\omega_4 \Delta(x_1, \omega)\Delta(x_2, \omega_2) \\ &\times G(\omega_1, \omega_2, \omega_3, \omega_4) X(\omega_3, \omega_4). \end{aligned} \tag{26}$$

If we know (a) the Bethe–Salpeter amplitude which is obtained from a solution of Eq. (26), (b) the solutions to all the tau functions for the elementary fields, and (c) that we have a local two-body bound state composed of the particles associated with the fields ϕ_1 and ϕ_2 (this we *assumed* to be true)—then we can, in principle, determine all the matrix elements for the bound states. The mass of the bound state is ascertainable from the position of the pole in the tau function. Consequently, knowing (a), (b), and (c), we can calculate all the properties of the bound states. In general, we do not know (c) and, consequently, cannot determine all the properties of the bound states. The form of the bound-state operator cannot be ascertainable by merely observing the properties of the tau functions, but is a consequence of additional assumptions. In the previous example we assumed that we had a local two-body bound state. The determination of the form of the bound-state operator will be discussed in a future paper, so no more will be said about it here. We do, however, discuss the various possible forms that can occur and their properties.

In the remainder of this section we divide the class of bound states of the first kind into subgroups with similar properties. These subgroups are characterized by the form of their bound-state operators. We consider the following subgroups:

(A) N-body bound states,
(B) mixed bound states,
(C) self-interacting bound states.

A. N-Body Bound States

This subclass is characterized by its bound-state operator

$$\begin{aligned} B(x) = &\int d^4\omega_1, \cdots, d^4\omega_N h(x; \omega_1, \omega_2, \cdots, \omega_N) \\ &\times :\phi_i(\omega_1), \cdots, \phi_l(\omega_N):, \end{aligned} \tag{27}$$

where $2 \leq N < \infty$.

Because of the kernel $h(x; \omega_1, \cdots, \omega_N)$, this bound-state operator is generally nonlocal. For the special case that the operator is restricted to be local we get the familiar form of an N-body bound-state operator

$$B(x) = \frac{:\phi_i(x), \cdots, \phi_1(x):}{(2\pi)^{\frac{3}{2}}(2P_0)^{\frac{1}{2}} \langle 0| :\phi_i(0), \cdots, \phi_1(0):|b\rangle}. \tag{28}$$

The C number $(2\pi)^{\frac{3}{2}}(2P_0)^{\frac{1}{2}}\langle 0|:\phi_i(0),\cdots,\phi_1(0):|b\rangle$ is included in order to make the bound-state field properly normalized.

The normal ordering sign : : is needed in order to assure $\langle 0|B(x)|0\rangle = 0$. The product of the operators on the right-hand side of Eq. (28) is at the same space–time point, and in order to make them well-defined it is implicitly implied that a limiting process should be taken.

From the form of the bound-state operator in Eqs. (27) or (28), it is obvious why we say that this bound state is composed of N elementary particles described by the N fields $\phi_i(x),\cdots,\phi_1(x)$. It is this subclass which is analogous to the familiar bound state of atoms and molecules that occur in quantum mechanics.

If we explicitly assume that we have an N-body bound state, then we can determine which N fields form the bound states by judiciously observing the poles in the various tau functions. Its mass can be determined from the position of the poles. If the form of $h(x;\omega_1,\cdots,\omega_N)$ is known, all other properties can be established by solving the Bethe-Salpeter amplitude and the tau functions for the elementary fields.

In order to illustrate this type of bound state let us construct a Lagrangian containing an N-body bound state. We will do this by considering a Lagrangian with an elementary particle and construct an equivalent Lagrangian in which the particle is composite.

Consider the case of a pseudoscalar meson interacting with a fermion by means of a Yukawa interaction. The meson and fermion are both considered to be elementary and the Lagrangian density is of the form

$$L(\phi,\psi,\bar{\psi}) = \bar{\psi}(x)(\overline{\nabla} + m_0)\psi(x) + \tfrac{1}{2}\phi(x)(\Box - \mu_0^2)\phi(x) + g_0\bar{\psi}(x)\gamma_5\psi(x)\phi(x). \quad (29)$$

The tau functions for the fermi fields contain the well-known poles at the physical mass of the pion. By the method in constructed Refs. 1, 2, and 3, a Lagrangian which will give the same value for the tau functions of the fermi and antifermi fields and not contain the boson is

$$L_\phi(\psi,\bar{\psi}) = \bar{\psi}(x)(\overline{\nabla} + m_0)\psi(x) + \tfrac{1}{2}(g_0^2) \times \int \bar{\psi}(x)\gamma_5\psi(x)\Delta(x - x^1)\bar{\psi}(x)\gamma_5\psi(x^1)\,d^4x^1, \quad (30)$$

where

$$\Delta(x - x^1) = \frac{1}{(2\pi)^4}\int \frac{e^{iP^\mu(x-x^1)_\mu}}{P^2 + \mu_0^2 - i\epsilon}.$$

The Lagrangian density in Eq. (30) does not contain the boson field, so by our definition it is a bound state relative to this Lagrangian. Its presence is manifested by the appearance of the poles in the tau functions of the elementary ψ, $\bar{\psi}$ fields. A nonlocal bound-state operator for boson can be constructed as explained in Ref. 3, such that

$$B(x) = \frac{\int \Delta(x - y)\,:\bar{\psi}(y)\gamma_5\psi(y):\,d^4y}{(2\pi)^{\frac{3}{2}}(2P_0)^{\frac{1}{2}}\int \Delta(-y)\langle 0|:\bar{\psi}(y)\gamma_5\psi(y):|b_i\rangle\,d^4y}. \quad (31)$$

For special values of the bare mass and bare coupling constants the Lagrangian density in Eq. (31) can be made local, and in this case the bound-state operator becomes

$$B(x) = \frac{:\bar{\psi}(x)\gamma_5\psi(x):}{(2\pi)^{\frac{3}{2}}(2P_0)^{\frac{1}{2}}\langle 0|:\bar{\psi}(0)\gamma_5\psi(0):|b\rangle}. \quad (32)$$

We therefore see that the Lagrangian density in Eq. (30) can describe a boson as being a two-particle bound state of a fermion and antifermion. Whether or not the form of this bound-state operator is unique is an interesting question which will be discussed in a future article.

B. Mixed Bound States

Let us consider that subgroup of bound states which are characterized by the operator

$$B(x) = \sum_{j=1}^{N} h_j(x;\omega_1,\cdots,\omega_j)\,:\phi_i(\omega_1),\cdots,\phi_1(\omega_j):, \quad (33)$$

where $2 \leq N < \infty$.

This type of bound state is a generalization of the N-body bound state. It is no longer possible to say that the bound particle is composed of N elementary particles, but it has become a mixture of various combinations.

As is characteristic of the class of bound states of the first kind, the mixed bound states will manifest their presences as poles in the various tau functions. The appearance of the singularities in many cases will be identical to that of the N-body bound state, and consequently by observing these singularities in the tau function it is in general impossible to distinguish between N-body bound states and mixed bound states. The knowledge of the bound-state operator is necessary to describe all the properties of the bound state. Therefore, to determine all the characteristics of the bound state some additional information other than the singularities of the tau functions is needed.

In order to illustrate a Lagrangian having a mixed bound state let us use the method of constructing equivalent field theories.

Consider the Lagrangian density

$$L(\phi_1, \cdots, \phi_4) = \sum_{i=1}^{4} \tfrac{1}{2}\phi_i(x)(\Box + m_i^2)\phi_i(x) + g_1\phi_1(x)[\phi_2(x)\phi_3(x) + g_2\phi_2(x)\phi_3(x)\phi_4(x)], \quad (34)$$

where $\phi_i(x)$ is a real scalar field. We can now construct an equivalent theory where the particle associated with $\phi_1(x)$ is a bound state, i.e., the field $\phi_1(x)$ does not appear in the Lagrangian.

For a particular choice of the coupling constants, the Lagrangian density describing the composite system is

$$L_{\phi_1}(\phi_2, \phi_3, \phi_4) = \sum_{i=2}^{4} \tfrac{1}{2}\phi_i(x)(\Box + m_i^2)\phi_1(x) + \lambda[\phi_2(x)\phi_3(x) + g\phi_2(x)\phi_3(x)\phi_4(x)]^2. \quad (35)$$

Since the tau functions of the elementary fields ϕ_2, ϕ_3, and ϕ_4 are equivalent, whether Eq. (34) or Eq. (35) is used to evaluate them, the composite particle will appear as a pole. Its renormalized bound-state operator is

$$B(x) = \frac{:[\phi_2(x)\phi_3(x) + g\phi_2(x)\phi_3(x)\phi_4(x)]:}{(2\pi)^{\frac{3}{2}}(2P_0)^{\frac{1}{2}}\langle 0|\,:[\phi_2(0)\phi_3(0) + g\phi_2(0)\phi_3(0)\phi_4(0)]:|0\rangle}. \quad (36)$$

We see that the bound state is of the mixed type.

C. Self-Interacting Bound States

This type of bound state is characterized by the Heisenberg bound-state operator

$$B(x) = \sum_{j=1} h_j(x; \omega_1, \cdots, \omega_j) :\phi_i(\omega_1), \cdots, \phi_1(\omega_j):. \quad (37)$$

This is just the generalization of the mixed bound state with an infinite number of terms. This type of bound state is unique from the subgroups A and B in that the bound state effectively couples directly with itself. To illustrate this direct coupling consider the Lagrangian density

$$L(\phi, \psi, \bar{\psi}) = \tfrac{1}{2}\phi(x)(\Box - \mu_0^2)\phi(x) - \bar{\psi}(x)(i\overline{\nabla} + m_0)\psi(x) - ig_0\bar{\psi}(x)\gamma_5\psi(x)\phi(x) - \lambda\phi^4(x). \quad (38)$$

This Lagrangian density describes an elementary meson, fermion, and antifermion. The elementary boson is directly coupled with itself by means of the Matthews term $\lambda\phi(x)^4$. It is this direct coupling that sets this type of bound state from the previous examples.

The construction of the equivalent theory where the meson appears as a bound state has been done by the author.[2] The Lagrangian density corresponding to the composite meson is

$$L_\phi(\psi, \bar{\psi}) \approx -\bar{\psi}(x)(i\overline{\nabla} + m_0)\psi(x) - (+g_1^2/3g_2)^{\frac{1}{2}}\bar{\psi}\gamma_5\psi\cos(\tfrac{1}{3}\alpha + \tfrac{1}{3}\pi) - \frac{4}{9g_2}\cos^4(\tfrac{1}{3}\alpha + \tfrac{1}{3}\pi) \quad (39)$$

in the weak coupling limit, where

$$\alpha = \arccos[(27g_1^2g_2)^{\frac{1}{2}}\bar{\psi}(x)\gamma_5\psi(x)]. \quad (40)$$

In the strong coupling limit we get

$$L_\phi(\bar{\psi}, \psi) \approx -\bar{\psi}(x)(i\overline{\nabla} + m_0)\psi(x) + (g_1^2/12g_2)^{\frac{1}{2}}\bar{\psi}(x)\gamma_5\psi(x) \times \{\cos[2\alpha(x)] + i\sqrt{3}\cot[2\alpha(x)]\} - \frac{1}{36g_2}\{\csc[2\alpha(x)] + i\sqrt{3}\cot[2\alpha(x)]\}^4, \quad (41)$$

where

$$\tan\alpha(x) = [\tan\tfrac{1}{2}\beta(x)]^{\frac{1}{3}} \quad (42)$$

and

$$\sin\beta(x) = [4/(27g_2g_1)]^{\frac{1}{2}}[\bar{\psi}(x)\gamma_5\psi(x)]^{-1}. \quad (43)$$

The form of the bound-state operator in the weak coupling limit is

$$B(x) \approx :\left\{\text{const} \times \cos[\tfrac{1}{3}\alpha(x) + \tfrac{1}{3}\pi] + \frac{\text{const}}{\bar{\psi}(x)\gamma_5\psi(x)} \times \cos^4[\tfrac{1}{3}\alpha(x) + \tfrac{1}{3}\pi]\right\}:, \quad (44)$$

where $\alpha(x)$ is defined in Eq. (40). Notice that for $\lambda \to 0$, Eq. (44) for $B(x)$ reduces to Eq. (32).

In the strong coupling limit the bound-state operator is of the form

$$B(x) : \text{const} \times \cos[2\alpha(x)] + i\sqrt{3}\cot[2\alpha(x)] - : \text{const}\,[\bar{\psi}(x)\gamma_5\psi(x)]^{-1}\{\csc[2\alpha(x)] + 3\cot[2\alpha(x)]\}^4, \quad (45)$$

where $\alpha(x)$ is defined by Eqs. (42) and (43).

The bound-state operators in Eqs. (44) and (45) are both of the form of Eq. (37). The meson will manifest its presence as poles in the tau functions of the elementary fields. As before, it is *not* possible to construct the

form of the bound-state operator by merely observing the singularities in the tau functions. The only reason that we were able to construct the bound-state operators was because of the privileged knowledge of both the elementary and composite Lagrangian in addition to some implicit assumptions that were not discussed.

From what we have said we can epitomize the class of bound states of the first kind by saying:

(a) The set of elementary fields $\phi_1, \phi_2, \cdots, \phi_N$ that appear in the Lagrangian do not form an irreducible set.

(b) The bound-state operators $B_1, B_2, \cdots, B_M$ that are needed to make the eigenstates complete can be expressed in terms of the elementary fields.

(c) At least one Bethe–Salpeter amplitude is not identically zero:

$$\langle 0|\, T(\phi_1(x_1), \cdots, \phi_k(x_N))\, |b\rangle \not\equiv 0$$

for some combination of fields.

(d) As a consequence of (c) the bound state will manifest itself by the appearance of poles in various tau functions. Consequently, a homogeneous equation for the Bethe–Salpeter amplitude can be derived.

(e) The form of the bound-state operator can *not* be ascertained from a knowledge of the tau functions alone.

IV. BOUND STATES OF THE *N*th KIND

In the previous section we considered the class of bound states that appeared as poles in the tau functions, i.e., bound states of the first kind. Because of the pole we could isolate its residue and obtain an explicit equation for the Bethe–Salpeter amplitude [cf. Eq. (26)]. Knowing the Bethe–Salpeter amplitude, the tau functions for the elementary fields, and the form of the bound-state operator, we could, in principle, calculate all the properties of the bound state. For the pole to exist it was necessary that

$$\langle 0|\, T(\phi_1(x_1), \cdots, \phi_k(x_S))\, |b\rangle \not\equiv 0 \qquad (46)$$

for some combination of fields $\phi_i(x)$. If Eq. (46) is identically zero, the pole does not appear but the bound state may still exist. It now manifests its presence as a branch point in the tau functions. It is this class of bound states which appear as branch points in the tau functions that we consider in this section.

The bound states that appear as branch points are much more complicated than bound states of the first kind. Apart from their difficulties, these types of bound states are extremely interesting because of the possible new kinds of theories that can be considered. For example, with this type of bound state it is possible to construct theories where a spin-$\frac{1}{2}$ particle appears as a bound state of an integer- or zero-spin field. The customary procedure is just the opposite, i.e., to construct integer-spin particles from spin-$\frac{1}{2}$ particles. It is for this reason that these bound states are very enticing regardless of their complexity.

These bound states that appear as branch points can most readily be broken into classes with similar properties. These classes, being generalizations of bound states of the first kind, are here designated as bound states of the second kind, the third kind, or in general the *N*th kind.

Let us begin with a discussion of bound states of the second kind. As the bound states of the first kind were characterized by the statements (a)–(e) at the end of Sec. III, the bound states of the second kind are characterized by a similar set of requirements:

(a) The set of elementary fields $\phi_1, \cdots, \phi_N$ that appear in the Lagrangian do not form an irreducible set.

(b) The bound-state operators $B_1, B_2, \cdots, B_M$ that are needed to make the eigenstates complete can be expressed in terms of the elementary fields.

(c) For all eigenstates $|b_i\rangle$ and all combinations of elementary fields

$$\langle 0|\, T(\phi_1(x_1), \cdots, \phi_k(x_n))\, |b_i\rangle\, 0.$$

(d) For some combination of elementary fields and some two-particle eigenstate $|b_i, b_j\rangle$ we have

$$\langle 0|\, T(\phi_1(x_1), \cdots, \phi_k(x_m))\, |b_i, b_j\rangle \not\equiv 0.$$

As a consequence of these properties, in particular (c) and (d), the bound state of the second kind will manifest itself as a branch point rather than a pole. This fact is easily illustrated by the example in Eq. (22). If $X_q^{(b)}(x_1, x_2) \equiv 0$, then the next contribution to $K(x_1, x_2, x_3, x_1)$ is the two-particle intermediate state. The contribution is proportional to

$$\langle 0|\, T(\phi_1(x_1)\phi_2(x_2))\, |b_i, b_j\rangle$$

lying in the continuum and giving rise to a branch point, and the position of the branch point is dependent on the mass of the bound state.

The difficulties of the bound state of the second kind, and so for that matter of the bound state of the *N*th kind ($N \geq 2$), are immediately apparent when one considers the calculation of a general matrix element

$$\langle \alpha|\, T(\phi_1(x_1), \cdots, \phi_k(x_k))\, |\beta\rangle, \qquad (47)$$

where $\langle \alpha|$ and $|\beta\rangle$ are two arbitrary states. In the case where we had only bound states of the first kind, this

matrix element could be evaluated, since the bound state produces a pole which allows one to explicitly express the general matrix element [Eq. (47)] in terms of the Bethe–Salpeter amplitude and tau functions for the elementary fields.[11] For the bound states of the second kind we have no pole term and consequently cannot carry out the analogous calculation. Closely related to this difficulty is the ambiguity that arises in constructing the properties of the one-particle eigenstate $|b\rangle$, since, in general, the only nonzero matrix elements that appear are those with an even number of bound states, e.g.,

$$\begin{aligned} \langle 0|\, T(\phi(x)\phi(y))\, |b\rangle &\equiv 0, \\ \langle 0|\, T(\phi(x)\phi(y))\, |b, b\rangle &\not\equiv 0, \\ \langle 0|\, T(\phi(x)\phi(y))\, |b, b, b\rangle &\equiv 0, \\ \langle 0|\, T(\phi(x)\phi(y))\, |b, b, b, b\rangle &\not\equiv 0, \end{aligned} \tag{48}$$

et cetera.

This will, in general, be due to the conservation numbers that force the matrix element

$$\langle 0|\, T(\phi(x)\phi(y))\, |b\rangle$$

to vanish. Therefore the quantum numbers of the bound states are only determined for, at most, the product of two bound-state operators.

In order to illustrate these ambiguities and difficulties we consider an explicit Lagrangian density with a bound state of the second kind. We can do this as before by means of constructing an equivalent field theory. The example we do consider is a spin-$\frac{1}{2}$ fermion as a bound state of the scalar π meson. The Lagrangian density in which both the fermion and π meson appear as elementary particles is assumed to be

$$\begin{aligned} L(\phi, \psi, \bar{\psi}) &= -\bar{\psi}(x)(i\not{\nabla} + m_0)\psi(x) + \tfrac{1}{2}\phi(x)(\Box - \mu_0^2)\phi(x) \\ &\quad + g_0\bar{\psi}(x)\Gamma\psi(x), \end{aligned} \tag{49}$$

where Γ is either a scalar or pseudoscalar gamma matrix. As explained in Ref. 1, the equivalent Lagrangian $L_{\psi,\bar{\psi}}(\phi)$, where the spin-$\frac{1}{2}$ field appears as a bound state, can be obtained from the expression

$$\begin{aligned} \int d^4x L_{\psi,\bar{\psi}}(\phi) &= \int d^4x \tfrac{1}{2}\{\phi(x)(\Box - \mu_0^2)\phi(x)\} \\ &\quad - \tfrac{1}{2}\,\mathrm{tr}\,\ln(1 + 2g_0 S\phi), \end{aligned} \tag{50}$$

where

$$S(x^1 - x) = (i\not{\nabla}_{x^1} + m_0)\Delta(x^1 - x).$$

The Lagrangians in Eqs. (49) and (50) yield the same tau functions for the scalar field $\phi(x)$. However, in Eq. (50) the fermi field does not appear and consequently by our definition is designated as a bound state. It is also true that if $|b\rangle$ represents the one-particle eigenstate of the bound fermi field then

$$\langle 0|\, T(\phi(x_1), \cdots, \phi(x_n))\, |\alpha\rangle = 0 \tag{51}$$

for all n and arbitrary eigenstates $|\alpha\rangle$ that contain an odd number of fermions. This is just due to the fact that $|\alpha\rangle$ must have a noninteger-spin eigenstate and ϕ is a scalar or pseudoscalar field. On the other hand, the eigenstate for the product of two spin-$\frac{1}{2}$ fields $|b, b\rangle$ has a scalar representation and consequently

$$\langle 0|\, T(\phi(x_1), \cdots, \phi(x_N))\, |bb\rangle \not\equiv 0 \tag{52}$$

if all the other quantum numbers are in agreement. Therefore, the Lagrangian density in Eq. (50) contains a bound state of the second kind.

Notice that just given a Lagrangian with a bound state of the second kind we have no way of determining the quantum numbers of the one-particle eigenstate $|b\rangle$. The quantum numbers are only determined for the two-particle eigenstates $|b, b\rangle$, the matrix elements containing one-particle eigenstates identically vanish. For example, in the above illustration the only way we know that $|b\rangle$ was an eigenstate of spin $\frac{1}{2}$ is our privileged knowledge of its equivalent field theory described by Eq. (49). In general, this information is not available.

We can now generalize the previous definitions to include bound states of the Nth kind. We say that a bound state belongs to the Nth kind if:

(a) The set of elementary fields $\phi_1, \cdots, \phi_N$ that appear in the Lagrangian do not form an irreducible set.

(b) The bound-state operators $B_1, \cdots, B_M$ that are needed to make the eigenstates complete can be expressed in terms of the elementary fields.

(c) For all eigenstates $|b_i\rangle$ and all combinations of elementary fields,

$$\langle 0|\, T(\phi_1(x_1), \cdots, \phi_k(x_n))\, |b_i\rangle \equiv 0;$$

also for all combinations of two-particle eigenstates $|b_i, b_j\rangle$ and all combinations of elementary fields,

$$\langle 0|\, T(\phi_1(x_1), \cdots, \phi(x_n))\, |b_i, b_j\rangle \equiv 0;$$

and in general for all combinations of eigenstates $|b_i, \cdots, b_j\rangle$ up to $N - 1$ particles,

$$\langle 0|\, T(\phi_1(x_1), \cdots, \phi_s(x_n))\, |b_i, \cdots, b_j\rangle \equiv 0$$

for all combinations of elementary fields.

(d) For some combination of elementary fields and some N-particle eigenstate $|b_i, \cdots, b_j\rangle$, we have

$$\langle 0|\, T(\phi_1(x_1), \cdots, \phi_s(x_n))\, |b_i, \cdots, b_j\rangle \not\equiv 0.$$

[11] S. Mandelstam, Proc. Roy. Soc. (London) **A233**, 248 (1955).

The bound states of the Nth kind ($N \geq 3$) exhibit the same difficulties as bound states of the second kind, yet with even greater ambiguities.

A further exploration of the properties of the bound states of the Nth kind ($N \geq 2$) does not seem worth the effort at the present time until the more obvious difficulties are better understood.

V. CONCLUSION

In the previous sections we have discussed the properties and kinds of bound states that can occur in quantum field theory. We do not intend to imply that we have exhausted all possibilities; rather, only the more obvious types. Also we have by no means exhausted the characteristics and problems associated with the bound states; quite to the contrary, we have raised more questions than we have solved. In particular for the bound states of the second and higher kind, we could have broken these classes up into subclasses as was done for the bound states of the first kind. The complexity of the problem did not warrant such a discussion until some of the more practical difficulties with these bound states are pursued.

Under the bound states of the first kind we considered three subclasses. The first class, N-body bound states, are the most analogous to ordinary quantum mechanics. Although the classical picture of N elementary particles being bound by an attractive force is not valid, we can identify the N-body composite state to be composed of those N particles that are associated with the Heisenberg bound-state operator [cf. Eq. (27)]. In the case of the mixed bound state we can no longer say that it corresponds to a bound state of N particles. It is a mixture of various combinations of particles, the last subclass being the self-interacting bound state. If it is indeed the case that the π meson is to be treated as a bound state and has a self-interaction (i.e., the Matthews term) then this subclass is very important. We see, however, that the problem becomes much more complex for self-interacting bound states.

It is important to notice that just given the knowledge that the bound state appears as a pole in the tau function is not enough to determine all of its properties. That is, one cannot tell to which subclass the bound state belongs, and consequently one cannot determine all of its properties. To construct such a bound-state operator is not a trivial problem but a necessary task, if one is to have full knowledge of the composite particle such as is needed in discussing composite-particle scattering. This problem will be studied in a future article.

There are a few special examples in which the construction of such a bound-state operator is trivial, that is, in those cases where the bound-state theories have been constructed from an equivalent theory where the particle was elementary. This was the case for all the examples considered in this paper. However, given a theory with a composite state that has not been constructed from an elementary state, the form of the bound-state operator is not apparent and it is by no means a trivial task to determine it. Such examples are the form of the bound-state operator describing the $V - \theta$ bound state[12] in the Lee model or the deuteron bound state. It is always assumed that the deuteron is a two-nucleon bound state, but this need not be the case.

Taylor[13] and Weinberg[14] have both given procedures for replacing composite particles by elementary ones. They have limited themselves to bound states of the first kind. Furthermore they have assumed that the composite state belongs to the N-body subclass. This assumption is completely arbitrary, and in order to replace the composite state by an elementary one the form of the bound-state operator must be known.

The bound states of the second and higher kind allow us, in principle, to construct bound-state theories of a larger class. We need no longer assume that the fermi fields occupy a unique fundamental position. A position taken, for instance, in constructing the Heisenberg nonlinear theory but in no way dictated to us from the intrinsic properties of nature.

It is apparent that the Lagrangian for these higher kinds of bound states are very complex. However, it is not apparent that they cannot be simplified in such a way as not to destroy the characteristics of the bound state. This task must surely be accomplished not only for the higher kinds of bound states, but also for bound states of the first kind if field theory is to be of any utility in constructing composite theories.

[12] R. D. Amado, Phys. Rev. **122**, 696 (1961).
[13] M. M. Broido and J. G. Taylor, Phys. Rev. **147**, 993 (1966).
[14] S. Weinberg, *Proceedings of the 1962 High Energy Physics Conference*, CERN, Geneva.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Partial-Wave Analysis in Terms of $SL(2, C)/SU(2)$ Harmonic Functions

AWELE MADUEMEZIA*
International Centre for Theoretical Physics, Trieste, Italy

(Received 2 August 1968)

A new scheme for analyzing the scattering amplitude in the s channel of a two-body elastic collision in terms of the representation functions of the covering group of the homogeneous Lorentz group is presented. The scheme uses representation functions defined on the homogeneous space $SL(2, C)/SU(2)$, which for the equal-mass case considered here is the same as the hyperquadric $\mathbf{q}^2 - q_0^2 = -m^2$, where m is the particle mass, $\mathbf{q}$ is the relative center-of-mass momentum (in the initial or final state), and $2q_0 = s^{\frac{1}{2}}$ is the center-of-mass energy. The corresponding representations are multivalued and belong to the so-called degenerate (but not most-degenerate) series. The scattering amplitude derived has the correct threshold behavior in the s variable.

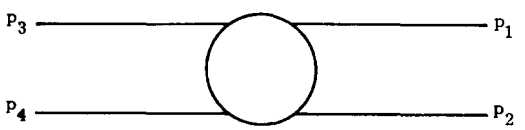
1. INTRODUCTION

Recently, a number of authors[1] have generalized Toller's idea[2] of using the representation of the Lorentz group instead of those of the rotation group for partial-wave analysis. The idea of the generalization is to make Toller's scheme applicable to directions of scattering other than the forward.

One feature common to the classic Jacob–Wick expansion and these extensions is the use of the scalar product between "plane-wave" states (or linear-momentum states) and the angular-momentum states. This scalar product turns out to be expressible in terms of the representation *matrices* of the rotation group in the Jacob–Wick case and of the homogeneous Lorentz group in the Toller-like extensions. It appears that one could sometimes dispense with this scalar product and derive an expansion of the Jacob and Wick amplitude in terms of the *basis functions* of a unitary irreducible representation of the Lorentz group or, rather, of its covering group.

In this paper we propose a simple scheme which literally translates the physical situation into mathematical language.

We consider a two-body elastic collision



with four-momenta p_1, p_2, incoming, and $-p_3, -p_4$, outgoing, and with all masses equal to m for simplicity.

The scattering amplitude may be written $f(\theta, \varphi)$, as in Jacob and Wick's paper,[3] where θ, φ are the usual spherical angles with φ azimuthal. We note that these authors expand this amplitude in terms of the functions $d^j_{\lambda\mu}(2\theta)$, which are the harmonic functions for the group $SU(2)$,[4] the covering group of the rotation group. One expects then that a natural extension of this expansion to one that uses the representation of the Lorentz group would involve the use of the harmonic functions for the group $SL(2, C)$, the covering group of the Lorentz group.

As far as partial-wave analysis is concerned, a simple chain of inclusions, $P_{lm}(\theta) \subset d^j_{\lambda\mu}(2\theta) \subset \cdots$, *within the family of the harmonic functions* would then be set up.

Now the (physical) s channel is unambiguously defined by the (mathematical) relation

$$s = (p_1 + p_2)^2 = (2q_0)^2 > 4m^2,$$
$$t = (p_1 + p_3)^2 < 0, \tag{1.1}$$

where $2q_0$ is the center-of-mass energy *and is positive.*

For the equal-mass case which we consider we have

$$s = 4(\mathbf{q}^2 + m^2), \tag{1.2}$$

where $\mathbf{q}$ is the relative center-of-mass momentum.

The relation $s < 0$ has a two-fold ambiguity only to the extent that it refers to both the t and the u channels.

In the physical channel one has to use wavefunctions for which $s = (2q_0)^2 > 0$. One also constructs the scattering amplitude to be a Lorentz-invariant. Put in mathematical language, the scattering amplitude is a linear combination of functions defined on the space $s = (2q_0)^2$. Such a space is easily constructed. The corresponding functions span a Hilbert space. They are the so-called harmonic functions of which $P_{lm}(\cos\theta)$ and $d^j_{\lambda\mu}(2\theta)$ are the first two examples.

* On leave from Department of Physics, University of Ghana, Legon.

[1] G. Domokos and G. L. Tindle, Phys. Rev. **165**, 1906 (1968); R. Delbourgo, Abdus Salam, and J. Strathdee, Phys. Rev. **164**, 1981 (1967).

[2] M. Toller, Nuovo Cimento **37**, 631 (1965); Internal Reports, University of Rome, Nos. 76 (1965); 84 (1965); A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967).

[3] M. Jacob and G. C. Wick, Ann. Phys. (N.Y.) **7**, 404 (1959).

[4] R. Rączka and J. Fischer, Commun. Math. Phys. **3**, 233 (1966); M. A. B. Bég and H. Ruegg, J. Math. Phys. **6**, 677 (1965).

2. THE s CHANNEL AS A HOMOGENEOUS SPACE

From the relation

$$s = 4(\mathbf{q}^2 + m^2),$$

we have

$$\mathbf{q}^2 - q_0^2 = -m^2, \quad q_0 > 0. \tag{2.1}$$

With m fixed, (2.1) allows for the joint variation of $|\mathbf{q}|$ and s.

We denote the hypersurface defined by (2.1) by S. It is probably well known that this surface is homeomorphic to the homogeneous space

$$SL(2, C)/SU(2).$$

Nevertheless, it may be instructive to prove this here, more so as our proof is really simple.

Suppose f, g are two matrices of $SL(2, C)$, that is, matrices of the form

$$\begin{pmatrix} a & b \\ c & \dfrac{1+bc}{a} \end{pmatrix},$$

where a, b, c are arbitrary complex numbers and $a \neq 0$. Then the elements of $SL(2, C)/SU(2)$ are the cosets of $SU(2)$ in $SL(2, C)$. These are characterized as follows: f, g lie in the same coset of $SU(2)$ if and only if $f^{-1}g \in SU(2)$, i.e.,

$$\text{if and only if} \quad \beta^{\mathrm{T}} f^{\mathrm{T}} \beta g \in SU(2),$$

where f^{T} denotes the transpose of f and

$$\beta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.2}$$

It follows that

$$(\beta^{\mathrm{T}} f^{\mathrm{T}} \beta g)^{\dagger} \beta^{\mathrm{T}} f^{\mathrm{T}} \beta g = 1, \tag{2.3}$$

where $\dagger$ is the adjoint (or Hermitian conjugate) operation.

Hence

$$g^{\dagger} \beta^{\mathrm{T}} f^{\mathrm{T}\dagger} f^{\mathrm{T}} \beta g = 1,$$

i.e.,

$$ff^{\dagger} = gg^{\dagger} = K, \quad \text{say}, \tag{2.4}$$

where K is a Hermitian matrix

$$\begin{pmatrix} k_1 & k_2 \\ k_2^* & k_3 \end{pmatrix},$$

where k_1, k_3 real *and positive* and k_2 complex.

Conversely, if $ff^{\dagger} = gg^{\dagger}$, then f, g belong to the same coset of $SU(2)$.

The cosets of $SU(2)$, i.e., the elements of $SL(2, C)/SU(2)$, are therefore represented by the points k_1, k_2, k_3. These numbers are, however, not independent because K must be unimodular. Thus we must have

$$k_1 k_3 = |k_2|^2 + 1. \tag{2.5}$$

Suppose now that y, z are two complex numbers and q_0, q_2 two real numbers and m a real and positive number.

Let us put

$$\begin{aligned} \operatorname{Re} k_2 &= |y|/m \equiv q_1/m, \quad \text{say}; \\ \operatorname{Im} k_2 &= |z|/m \equiv q_3/m, \quad \text{say}; \\ mk_1 &= q_0 - q_2; \\ mk_3 &= q_0 + q_2. \end{aligned} \tag{2.6}$$

Then (2.5) becomes

$$\begin{aligned} \mathbf{q}^2 - q_0^2 &= -m^2, \\ 2q_0 &= m(k_1 + k_3) > 0. \quad \text{Q.E.D.} \end{aligned} \tag{2.7}$$

In deriving Eq. (2.7) for the hypersurface $S \approx SL(2, C)/SU(2)$, we have allowed y, z to be complex *in order to reduce the degeneracy of the representations later to be constructed on* S. It is, however, to be understood that the physical components of momentum are given by $q_1 = |y|$, q_2 and $q_3 = |z|$. The phases introduced by the complexification of y and z are therefore to be regarded as unphysical.

The space S is conveniently parametrized as follows:

$$\begin{aligned} m^{-1}y &= x_1 + ix_2 = 2^{-\frac{1}{2}} e^{i\varphi_1} \sin\theta \sinh\alpha, \\ m^{-1}q_2 &= x_5 = 2^{-\frac{1}{2}} \sin\theta \sinh\alpha, \\ m^{-1}z &= x_3 + ix_4 = e^{i\varphi_2} \cos\theta \sinh\alpha, \\ m^{-1}q_0 &= x_6 = \cosh\alpha, \\ 0 \leq \varphi_1 \leq 2\pi, &\quad 0 \leq \varphi_2 < 2\pi, \quad 0 \leq \theta < \pi/2, \\ 0 \leq \alpha < \infty. & \end{aligned} \tag{2.8}$$

Mathematically, θ is a free parameter and simply measures the inclination of $\mathbf{q}$ with respect to arbitrarily chosen axes. With a judicious choice of these axes, we may (physically) identify θ with the s-channel center-of-mass scattering angle, usually written θ_s. Another way of putting this is to note that $\mathbf{q}^2$ is the same for initial and final states, so that in (1.2) we could substitute $(\mathbf{q}')^2$ for $\mathbf{q}^2$, where $\mathbf{q}'$ is the final state relative center-of-mass momentum. This operation then ties up the *dynamics* of scattering which involves the momentum-transfer variable t, with the *kinematics* supplied by s and other initial conditions.

The metric on S is induced by the metric $g_{\alpha\beta}(E_5^1)$—for a positive-definite line element $ds^2 = x_\alpha g_{\alpha\beta} x_\beta$—on the six-dimensional pseudo-Euclidean space E_5^1. This is given by

$$g_{\alpha\beta}(S) = \sum_{k,l=1}^{6} g_{kl}(E_5^1) \partial_\alpha x^k \partial_\beta x^l, \quad \alpha, \beta = 1, 2, 3, 4, \tag{2.9}$$

where

$$\partial_\beta\,(\beta = 1, 2, 3, 4) = (\partial_{\varphi_1}, \partial_{\varphi_2}, \partial_{\theta_s}, \partial_\alpha)$$

and

$$g_{kl}(E_5^1) = \begin{pmatrix} -1 & & & & \\ & -1 & & 0 & \\ & & -1 & & \\ & & & -1 & \\ & 0 & & & -1 \\ & & & & & 1 \end{pmatrix}. \tag{2.10}$$

Hence,

$$-g(S) = \begin{pmatrix} \frac{1}{2}\sin^2\theta_s \sinh^2\alpha & 0 & 0 & 0 \\ 0 & \cos^2\theta_s \sinh^2\alpha & 0 & 0 \\ 0 & 0 & \sinh^2\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.11}$$

$$(\bar{g})^{\frac{1}{2}} = (|\det g(S)|)^{\frac{1}{2}} = \tfrac{1}{2}\sin\theta_s \cos\theta_s \sinh^3\alpha. \tag{2.12}$$

3. THE HARMONIC FUNCTIONS

The harmonic functions are eigenfunctions of the so-called Laplace–Beltrami operator

$$\Delta(S) = (\bar{g})^{-\frac{1}{2}}\partial_\alpha g^{\alpha\beta}(S)(\bar{g})^{\frac{1}{2}}\partial_\beta, \tag{3.1}$$

where the $g^{\alpha\beta}(S)$ are the matrix elements of $g(S)^{-1}$. From (2.10), (2.11), (2.12), and (3.1) we obtain

$$\Delta(S) = -\frac{1}{\sinh^3\alpha}\frac{\partial}{\partial\alpha}\sinh^3\alpha\frac{\partial}{\partial\alpha} - \frac{\delta}{\sinh^2\alpha}, \tag{3.2}$$

where

$$\delta = \frac{1}{\sin\theta_s\cos\theta_s}\frac{\partial}{\partial\theta_s}\sin\theta_s\cos\theta_s\frac{\partial}{\partial\theta_s} + \frac{1}{\cos^2\theta_s}\frac{\partial^2}{\partial\varphi_2^2} + \frac{2}{\sin^2\theta_s}\frac{\partial^2}{\partial\varphi_1^2}. \tag{3.3}$$

As usual, we chose the φ dependence of the eigenfunctions of Δ in the form

$$\exp\,[i(\mu_1\varphi_1 + \mu_2\varphi_2)].$$

The constant μ_2 we take as an integer m_2, so that the eigenfunction is single-valued in φ_2. With regard to φ_1, however, we are forced by (later) physical considerations to choose $\mu_1 = m_1/(2)^{\frac{1}{2}}$, where m_1 is an integer. The eigenfunctions are therefore multivalued in φ_1. This is why we must work in the range $0 \leq \varphi_1 < 2\pi$. The numbers μ_1, μ_2 are subject to a constraint $\mu_1 + \mu_2 = M$, where M is an arbitrary number.

The operator δ then has the eigenvalues $-2j(2j+2)$, where $2j$ is an integer, and eigenfunctions

$$\Psi(\theta_s, \varphi_1, \varphi_2) = Ne^{i[m_1(2)^{-\frac{1}{2}}\varphi_1 + m_2\varphi_2]} d^j_{\frac{1}{2}(m_1-m_2),\frac{1}{2}(m_1+m_2)}(2\theta_s) \tag{3.4}$$

where $d^j_{\lambda\mu}(2\theta_s)$ is defined as in Ref. 5, i.e., in terms of Jacobi polynomials $P_j^{\alpha\beta}(\cos\theta_s)$:

$$d^j_{m',m}(2\theta_s) = \left[\frac{\Gamma(j+m'+1)\Gamma(j-m'+1)}{\Gamma(j+m+1)\Gamma(j-m+1)}\right]^{\frac{1}{2}} \times (\cos\theta_s)^{m'+m}(\sin\theta_s)^{m'-m} \times P_{j-m'}^{m'-m,m'+m}(\cos 2\theta_s). \tag{3.5}$$

Finally, we have to solve the eigenvalue equation

$$\Delta V(\alpha) = \left[-\frac{1}{\sinh^3\alpha}\frac{d}{d\alpha}\sinh^3\alpha\frac{d}{d\alpha} + \frac{2j(2j+2)}{\sinh^2\alpha}\right]V(\alpha) = \lambda V(\alpha). \tag{3.6}$$

An equation of this form has been studied by Limić, Niederle, and Rączka[6] in connection with the continuous most-degenerate representations of the group $SO(4, 1)$. Their results indicate that Δ has only a continuous spectrum:

$$\lambda = -\Lambda^2 - \tfrac{9}{4}, \quad 0 \leq \Lambda < \infty. \tag{3.7}$$

The (normalized) eigenfunction V_j^Λ is given by

$$V_j^\Lambda(\alpha) = K^{-\frac{1}{2}}\tanh^{2j}\alpha\cosh^{i\Lambda-\frac{3}{2}}\alpha \times {}_2F_1(j - \tfrac{1}{2}(i\Lambda - \tfrac{3}{2}), j - \tfrac{1}{2}(i\Lambda - \tfrac{5}{2}); 2j+2; \tanh^2\alpha), \tag{3.8}$$

where

$$K = \left|\frac{(2\pi)^{\frac{1}{2}}\Gamma(i\Lambda)\Gamma(2j+2)}{\Gamma(j+\frac{1}{2}(i\Lambda+\frac{3}{2}))\Gamma(j+\frac{1}{2}(i\Lambda+\frac{5}{2}))}\right|^2. \tag{3.9}$$

The normalization of the functions $\Psi(\theta_s, \varphi_1\varphi_2)$, $V_j^\Lambda(\alpha)$ is carried out with respect to the invariant measures $\sin\theta_s\cos\theta_s\,d\theta_s\,d\varphi_1\,d\varphi_2$ and $\sinh^3\alpha\,d\alpha$, respectively. The harmonic functions form an orthonormal set of functions.

4. THE SCATTERING AMPLITUDE

For a two-particle system with incoming helicities λ_a, λ_b and outgoing helicities λ_c, λ_d the scattering amplitude is a square-integrable junction on S, provided we assume a finite total cross section which goes to zero at $s \to \infty$. It may then be written as a linear combination of the harmonic functions on S:

$$f_{\lambda_c\lambda_d,\lambda_a\lambda_b}(\theta_s\alpha) = \frac{1}{|\mathbf{q}|}\int_0^\infty d\Lambda \sum_{j=0}^\infty (j+\tfrac{1}{2})\,\langle\lambda_c\lambda_\alpha|\,C^{j\Lambda}\,|\lambda_a\lambda_b\rangle \times V_j^\Lambda(\alpha)\,d^j_{\frac{1}{2}(m_1-m_2),\frac{1}{2}(m_1+m_2)}(2\theta_s), \tag{4.1}$$

[5] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1957).

[6] N. Limić, J. Niederle,and R. Rączka, J. Math. Phys. 7, 2026 (1966).

where

$$m_1 = \lambda_a - \lambda_d + \lambda_c - \lambda_b,$$
$$m_2 = \lambda_c - \lambda_a + \lambda_b - \lambda_d,$$
$$\sinh\alpha = |\mathbf{q}|/m, \quad \cosh\alpha = s^{\frac{1}{2}}/2m,$$

and j is the angular momentum. The quantity $C^{j\Lambda}$ does not depend on s.

We have factored out $(j+\frac{1}{2})/|\mathbf{q}|$ so as to put the expansion in a form analogous to the Jacob–Wick expansion.

We have also averaged over the unphysical angles φ_1, φ_2.

It is the fact that m_1, m_2 turn out to be linear combinations of the helicities that constrained us to use the multi-valued representations.

If we compare our expansion (4.1) with the Jacob–Wick expansion

$$f_{\lambda_c\lambda_d,\lambda_a\lambda_b}(E,\theta) = \frac{1}{|\mathbf{q}|}\sum_j (j+\tfrac{1}{2})\,\langle\lambda_c\lambda_d|\,T^j(E)\,|\lambda_a\lambda_b\rangle\, d^j_{\lambda\mu}(\theta),$$
$$\lambda = \lambda_a - \lambda_b, \quad \mu = \lambda_c - \lambda_d, \quad (4.2)$$

we see that, assuming that the order of integration and summation may be interchanged in (4.1),

$$T^j(E) = \int_0^\infty d\Lambda\, C^{j\Lambda} V_j^\Lambda(\alpha). \quad (4.3)$$

We thus have an explicit analytic form for the energy dependence of the scattering amplitude. The function $C^{j\Lambda}$ is of the form of a form factor in Λ space. It must be of such a form as to ensure the convergence of the integral.

Let us take a quick glance at the analytic structure of $V_j^\Lambda(\alpha)$ as a function of s.

First we note that its ${}_2F_1$ component is a single-valued analytic function of $\tanh^2\alpha$ $(= 1 - 4m^2/s)$ in the whole (complex) $\tanh^2\alpha$ plane with a branch cut along the positive real axis from $+1$ to $+\infty$. This cut thus corresponds to $s < 0$. This is the usual left-hand cut of the elastic scattering amplitude. One would obtain the right-hand cut by analytic continuation into the complex s plane. For the analytically continued amplitude, one expects a cut from $s = +4m^2$ to $s = +\infty$. Our amplitude $T^j(E)$ would be the boundary value (from above) of this analytically continued amplitude.

Let us also consider the asymptotic behavior of $T^j(E)$ as s goes to infinity,[7] i.e., as $\tanh^2\alpha \to 1$.

In this limit, the series for ${}_2F_1(a, b; c; \tanh^2\alpha)$ is absolutely convergent if and only if

$$\mathrm{Re}\,(a + b - c) < 0.$$

For the function $V_j^\Lambda(\alpha)$ this corresponds to

$$\mathrm{Re}\,(j - \tfrac{1}{2}(i\Lambda - \tfrac{3}{2}) + j - \tfrac{1}{2}(i\Lambda - \tfrac{5}{2}) - 2j - 2) = \mathrm{Re}\,(-i\Lambda) = 0,$$

irrespective of the value of j. Thus using the formula

$${}_2F_1(a, b; c; 1-\epsilon) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(a)}\sum_{n=0}^{\infty}\frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)}\frac{(1-\epsilon)^n}{n!},$$
$$1 > \mathrm{Re}\,(a + b - c) \geq 0,$$
$$c \neq 0, -1, -2, \cdots, \epsilon > 0,$$

we have

$$T^j(E) \underset{s\to\infty}{\sim} \left(\frac{s}{4m^2}\right)^{-\frac{3}{4}}\int_0^\infty d\Lambda\, C^{j\Lambda}\left(\frac{s}{4m^2}\right)^{i\Lambda/2} \times \frac{\Gamma(2j+2)\Gamma(4+i\Lambda)}{\Gamma(j+\frac{11}{4}+i\Lambda/2)\Gamma(j+\frac{13}{4}+i\Lambda/2)}. \quad (4.4)$$

That is, $T^j(E)$ goes to zero at least as $s^{-\frac{3}{4}}$. For the reason given in Footnote 7, this is only of mathematical and not of experimental interest.

What is of experimental interest is the behavior at small s, i.e., the so-called threshold behavior.

Thus for $s = 4m^2(1+\epsilon)$, ϵ a small parameter,

$$\tanh^2\alpha = 1 - 4m^2/s \approx \epsilon \quad (4.5)$$

and

$${}_2F_1(a, b; c; \tanh^2\alpha) \approx 1 + (ab/c)\tanh^2\alpha. \quad (4.6)$$

Then,

$$T^j(E) = \int_0^\infty d\Lambda\, C^{j\Lambda} K^{-\frac{1}{2}}(1 - 4m^2/s)^{2j}(s/4m^2)^{i\Lambda/2-\frac{3}{4}} \times \left\{1 + (1 - 4m^2/s) \times \frac{[j - \frac{1}{2}(i\Lambda - \frac{3}{2})][j - \frac{1}{2}(i\Lambda - \frac{5}{2})]}{2j+2}\right\} \quad (4.7)$$

$$= x^{-\frac{3}{4}-2j}(x-1)^{2j}\int_0^\infty d\Lambda\, F(j\Lambda)\exp\left(\frac{i\Lambda}{2}\ln x\right) \times \left(1 + x^{-1}(x-1) \times \frac{(j+1)^2 - \frac{1}{16} - \Lambda^2/4 - i(j+1)\Lambda}{2(j+1)}\right), \quad (4.8)$$

where

$$x = s/4m^2 = 1 + \epsilon$$

and

$$F(j\Lambda) = C^{j\Lambda}K^{-\frac{1}{2}} \quad (4.9)$$

Since x is close to 1, $\exp(\frac{1}{2}\Lambda \ln x)$ is a slowly varying

[7] This is to be considered a mathematical limit, since physically such a limiting process would open up inelastic channels and our expansion would break down.

[8] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, Berlin, 1966), p. 37.

function of Λ; its period $4\pi/\ln x$ is almost infinite! We may therefore set it equal to 1 as a first approximation. Furthermore, $F(j\Lambda)$ is of the nature of a form factor in Λ space. We cannot determine its exact form within the context of this general theory. To do this one has to bring in such other constraints as unitarity and crossing on the scattering amplitude. A (mathematically) desirable form which it should have would be the Gaussian form

$$\exp\left[-\Lambda^2/\alpha(j)\right], \tag{4.10}$$

where $\alpha(j)$ is some function of j. The Λ integration would then be trivial. In any case, with the $\exp(-i\Lambda \ln x)$ term disposed of as we indicated, we can study the s behavior of $T^j(E)$ and hence of the differential cross section

$$\frac{d\sigma}{d\Omega} \sim \left|\sum_j T^j(E)\right|^2 \mathbf{q}^{-2}$$

without carrying out the Λ integration. The two terms of (4.7) give a variation of $T^j(E)$, with s of the forms

$$y_1 = x^{-\frac{3}{4}-2j}(x-1)^{2j}$$

and

$$y_2 = x^{-\frac{3}{4}-2j-1}(x-1)^{2j+1},$$

respectively. Now,

$$\frac{dy_1}{dx} = \left[\frac{-(\frac{3}{4}+2j)}{x} + \frac{2j}{x-1}\right] x^{-\frac{3}{4}-2j}(x-1)^{2j}. \tag{4.11}$$

Thus, since $x > 1$,

$$\frac{dy_1}{dx} \gtrless 0$$

according as

$$\frac{2j}{x-1} \gtrless \frac{\frac{3}{4}+2j}{x},$$

i.e., according as

$$x \lesseqgtr 1 + \tfrac{8}{3}j. \tag{4.12}$$

Similarly

$$\frac{dy_2}{dx} \gtrless 0$$

according as

$$x \lesseqgtr 1 + \tfrac{4}{3}(2j+1). \tag{4.13}$$

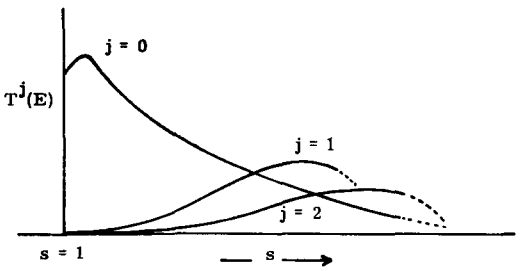


FIG. 1. General form of s dependence of partial-wave amplitude $T^j(E)$. The hump in the $j = 0$ curve would be absent if we used only the leading term in (4.8).

It follows that the scattering amplitude has the threshold behavior indicated in Fig. 1, which is the correct one. One cannot make any definite dynamical predictions without studying the detailed form of $F(j, \Lambda)$ and without effecting the Λ integration. This problem is being studied.

Let us conclude with a remark on crossing and reggeization. When one goes to the crossed channels, one has $s < 0$, $t > 0$; $s < 0$, $u > 0$. One would again derive an expansion in terms of harmonic functions as before, using spaces defined by $t > 0$ and by $u > 0$, respectively. Analytic continuation from one channel amplitude to another is then quite straightforward. If one then wishes to "reggeize" one would do so simply by executing the integral $\int_0^\infty d\Lambda$ *as a contour integral.* One does so by suitably closing the contour in a complex Λ plane. One would then pick up poles of T in the Λ plane. These poles are expected to be subject to some constraints of the form $f(\Lambda, j) = 0$.

ACKNOWLEDGMENTS

The author is deeply indebted to Professor E. J. Squires and Professor A. O. Barut for reading the manuscript and for suggestions leading to its improvement. He wishes to thank Professor Abdus Salam and Professor P. Budini for hospitality at the International Centre for Theoretical Physics, Trieste, and also the IAEA for an associateship that made his stay at the Centre possible.

Analytic Continuation of Appell's Hypergeometric Series F_2 to the Vicinity of the Singular Point $x = 1, y = 1$

G. E. HAHNE
Ames Research Center, NASA, Moffett Field, California

(Received 14 August 1968)

The infinite series, absolutely convergent if $|x| + |y| < 1$, for Appell's $F_2(\alpha, \beta, \beta', \gamma, \gamma'; x, y)$ is analytically continued into a linear combination of four infinite series in powers of $(x - 1)$ and $(y - 1)$; each of the latter four series is absolutely convergent if $|x - 1| + |y - 1| < 1$. The analytic continuation is carried out by manipulation of the Mellin–Barnes integral representations for the hypergeometric functions appearing in the course of the calculation.

I. INTRODUCTION

Appell's hypergeometric series F_2, which is a function of two complex arguments x and y, and of five complex parameters α, β, β', γ, and γ' ($\gamma, \gamma' \neq 0, -1, -2, \cdots$), is defined as follows[1,2]:

$$F_2(\alpha, \beta, \beta', \gamma, \gamma'; x, y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{(\gamma)_m(\gamma')_n m!\, n!} x^m y^n, \tag{1}$$

where the symbols of the type $(\delta)_m$ are given by

$$(\delta)_m = \frac{\Gamma(\delta + m)}{\Gamma(\delta)} = \begin{cases} 1, & \text{if } m = 0, \\ \delta(\delta + 1) \cdots (\delta + m - 1), & \\ & \text{if } m = 1, 2, \cdots. \end{cases}$$

The series (1) converges absolutely for $|x| + |y| < 1$ and, in general, diverges for $|x| + |y| > 1$.[1,2]

In calculations related to certain physical problems,[3,4] the need may arise for analytic continuation of F_2 into a series that is convergent in a neighborhood of the point $x = 1, y = 1$. Borngässer[5] obtained series expansions, in powers of $(1 - x)$ and $(1 - y)$, of the four linearly independent functions that are solutions of the system of partial differential equations satisfied by F_2. Formulas for the analytic continuation of Appell's F_2 to a neighborhood of $x = 1, y = 1$ were first given by Olsson.[6] Subsequently, more results were presented by Almström and Olsson.[7] The continuations of Olsson were derived with the aid of manipulations of series and of the use of an integral representation of Euler's type for a function appearing in the calculation. In the present work, the vehicle for carrying out the desired continuation of F_2 is the Mellin–Barnes type of integral representation for hypergeometric functions, and an application of Barnes's lemma provides the essential simplification that makes the calculation feasible.

II. ANALYTIC CONTINUATION OF F_2 TO THE VICINITY $x = 1, y = 1$

Appell's hypergeometric function F_2 may be expressed in terms of a Mellin–Barnes contour integral as follows[8]:

$$F_2(\alpha, \beta, \beta', \gamma, \gamma'; x, y) = \frac{\Gamma(\gamma')}{\Gamma(\alpha)\Gamma(\beta')} \frac{1}{2\pi i} \int_{-k-i\infty}^{-k+i\infty} F(\alpha + t, \beta; \gamma; x) \times \frac{\Gamma(\alpha + t)\Gamma(\beta' + t)}{\Gamma(\gamma' + t)} \Gamma(-t)(-y)^t\, dt, \tag{2}$$

where $F(\alpha + t, \beta; \gamma; x)$ is Gauss's hypergeometric function.[9] In (2), the contour in the t plane parallels the imaginary axis, except that, where necessary, it is indented so that the poles of $\Gamma(\alpha + t)\Gamma(\beta' + t)$ all lie on the left-hand side of the contour, and the poles of $\Gamma(-t)$ all lie on the right-hand side. From the asymptotic behavior of $\Gamma(c + t)$ for large $|t|$ and fixed c,[10] and of $F(\alpha + t, \beta; \gamma; x)$ for large $|t|$, and for fixed α, β, γ, x, and $\arg(t)$ (see Appendix A), it may be concluded that the integral in (2) converges absolutely if $y \neq 0$, $x \neq 1$, and if $|\arg(-y)| < \pi$, $|\arg(1 - x)| < \pi$, and $|\arg(-y) - \arg(1 - x)| < \pi$. All of the above conditions can be satisfied if both $\mathrm{Im}(x) > 0$ and $\mathrm{Im}(y) > 0$, or if both $\mathrm{Im}(x) < 0$ and $\mathrm{Im}(y) < 0$; one of these two sets of conditions is assumed to hold until further notice.

[1] P. Appell and J. Kampé de Fériet, *Fonctions hypergéométriques et hypersphériques; polynomes d'Hermite* (Gauthier-Villars et Cie., Paris, 1926), pp. 13–19.
[2] *Higher Transcendental Functions, Vol. 1*, A. Erdélyi, Ed. (McGraw-Hill Book Company, Inc., New York, 1953), pp. 222–229.
[3] L. C. Biedenharn, J. L. McHale, and R. M. Thaler, Phys. Rev. **100**, 376 (1955).
[4] K. Alder, A. Bohr, T. Huus, B. Mottelson, and A. Winther, Rev. Mod. Phys. **28**, 432 (1956).
[5] Ludwig Borngässer, "Über hypergeometrische Funktionen zweier Veränderlichen," Dissertation, Technische Hochschule Darmstadt, 1933, pp. 42–45.
[6] P. O. M. Olsson, Arkiv Fysik **30**, 187 (1965); **29**, 459 (1965).
[7] H. Almström and P. Olsson, J. Math. Phys. **8**, 2013 (1967).
[8] Reference 1, p. 40.
[9] Reference 2, p. 56.
[10] Reference 2, p. 47.

Barnes[11,12] has established the following integral representation for Gauss's hypergeometric function:

$$F(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c-a)\Gamma(c-b)} \frac{1}{2\pi i} \int_{-l-i\infty}^{-l+i\infty} \Gamma(a+s) \times \Gamma(b+s)\Gamma(c-a-b-s)\Gamma(-s)(1-z)^s \, ds, \quad (3)$$

where $|\arg(1-z)| < 2\pi$, and where the poles of $\Gamma(a+s)\Gamma(b+s)$ are on the left, and the poles of $\Gamma(c-a-b-s)\Gamma(-s)$ are on the right of the contour in the s plane. If a is replaced by $\alpha + t$, b by β, c by γ, and z by x, Eq. (3) then gives an integral representation which may replace the hypergeometric function inside the integral of formula (2). Suitable restrictions on the parameters make it possible to interchange the order of integrations in the resulting iterated integral which yields

$$F_2(\alpha, \beta, \beta', \gamma, \gamma'; x, y) = \frac{\Gamma(\gamma)\Gamma(\gamma')}{\Gamma(\alpha)\Gamma(\beta')\Gamma(\gamma-\beta)\Gamma(\beta)} \frac{1}{2\pi i} \times \int_{-l-i\infty}^{-l+i\infty} \Gamma(\beta+s)\Gamma(-s)(1-x)^s \times \left[\frac{1}{2\pi i} \int_{-k-i\infty}^{-k+i\infty} \frac{\Gamma(\alpha+s+t)\Gamma(\beta'+t)}{\Gamma(\gamma-\alpha-t)\Gamma(\gamma'+t)} \times \Gamma(\gamma-\alpha-\beta-s-t)\Gamma(-t)(-y)^t \, dt\right] ds. \quad (4)$$

It remains to transform the inner integral on the rhs of (4) into a form involving a power of $1-y$. This calculation may be begun with an application of Barnes's lemma,[13] according to which

$$\frac{1}{2\pi i} \int_{-m-i\infty}^{-m+i\infty} \frac{\Gamma(\alpha+s+u)\Gamma(\beta'+u)\Gamma(t-u)\Gamma(\gamma'-\beta'-\alpha-s-u)}{\Gamma(\gamma'-\beta')\Gamma(\gamma'-\alpha-s)} \, du = \frac{\Gamma(\alpha+s+t)\Gamma(\beta'+t)}{\Gamma(\gamma'+t)}, \quad (5)$$

provided that the poles of $\Gamma(\alpha+s+u)\Gamma(\beta'+u)$ are on the left, and the poles of $\Gamma(t-u)\Gamma(\gamma'-\beta'-\alpha-s-u)$ are on the right of the u contour. Note that only one of the arguments of the gamma functions inside the integral in (5) involves t. The expression on the rhs of (5) appears as a factor inside the inner integral in (4). If the integral representation (5) is substituted for this ratio of gamma functions in (4), and the order of performing the operations of integration along the u and along the t contours in the resulting integral is reversed, one obtains for the *inner* integral in (4):

$$\frac{1}{2\pi i} \int_{-m-i\infty}^{-m+i\infty} \frac{\Gamma(\alpha+s+u)\Gamma(\beta'+u)\Gamma(\gamma'-\beta'-\alpha-s-u)}{\Gamma(\gamma'-\beta')\Gamma(\gamma'-\alpha-s)} \times \left[\frac{1}{2\pi i} \int_{-k-i\infty}^{-k+i\infty} \frac{\Gamma(-u+t)\Gamma(\gamma-\beta-\alpha-s-t)\Gamma(-t)}{\Gamma(\gamma-\alpha-t)} (-y)^t \, dt\right] du. \quad (6)$$

If $|y| > 1$, the inner integral (including the factor $1/2\pi i$) of (6) is equal to the sum of the residues of the integrand at the poles of $\Gamma(-u+t)$,[14] i.e., equals

$$\frac{\Gamma(\gamma-\beta-\alpha-s-u)\Gamma(-u)}{\Gamma(\gamma-\alpha-u)} (-y)^u F\left(\gamma-\beta-\alpha-s-u, -u; \gamma-\alpha-u; \frac{1}{y}\right),$$

which, by Euler's transformation,[15] becomes

$$\frac{\Gamma(\gamma-\beta-\alpha-s-u)\Gamma(-u)}{\Gamma(\gamma-\alpha-u)} (-y)^u \left(\frac{y-1}{y}\right)^u F\left(\beta+s, -u; \gamma-\alpha-u; \frac{1}{1-y}\right).$$

The ratio $(-y)^u/y^u$ is $e^{\mp i\pi u}$, where the upper, or lower, sign in the exponent corresponds to $\operatorname{Im}(y) > 0$, or to $\operatorname{Im}(y) < 0$, respectively. The above expression may again be given in terms of a Mellin–Barnes contour integral

$$e^{\mp i\pi u} \frac{\Gamma(\gamma-\beta-\alpha-s-u)}{\Gamma(\beta+s)} \frac{1}{2\pi i} \int_{-n-i\infty}^{-n+i\infty} \frac{\Gamma(-u+v)\Gamma(\beta+s+u-v)\Gamma(-v)}{\Gamma(\gamma-\alpha-v)} (y-1)^v \, dv, \quad (7)$$

where $|\arg(y-1)| < \pi$ and where the contour divides the v plane so as to separate the poles of

[11] E. W. Barnes, Proc. London Math. Soc., Ser. 2, 6, 141 (1908).
[12] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, London, 1927), 4th ed., p. 290.
[13] Reference 11, p. 155; Ref. 2, p. 50, Eq. (8).
[14] This assertion is closely related to one proved in Ref. 12, Sec. 14.5.
[15] Reference 2, p. 109, Eq. (6).

$\Gamma(-u+v)$ from those of $\Gamma(\beta+s+u-v)\Gamma(-v)$; the proof of this assertion lies in assuming that $|y-1|>1$ and then evaluating the integral as $2\pi i$ times the sum of the residues of the integrand at the poles of $\Gamma(-u+v)$. If (7) is now substituted for the expression within brackets in (6), and the order of integrations in the resulting double integral is reversed, one obtains

$$\frac{1}{\Gamma(\gamma'-\beta')\Gamma(\gamma'-\alpha-s)\Gamma(\beta+s)}\frac{1}{2\pi i}\int_{-n-i\infty}^{-n+i\infty}\frac{\Gamma(-v)}{\Gamma(\gamma-\alpha-v)}(y-1)^v \times\left[\frac{1}{2\pi i}\int_{-m-i\infty}^{-m+i\infty}\Gamma(\alpha+s+u)\Gamma(\beta'+u)\Gamma(\beta+s-v+u)\Gamma(\gamma-\beta-\alpha-s-u) \times\Gamma(\gamma'-\beta'-\alpha-s-u)\Gamma(v-u)e^{\mp i\pi u}\,du\right]dv. \quad (8)$$

The expression in brackets in (8) will be called M_- or M_+, the subscript corresponding to the sign in the exponent. $M_\mp$ is evaluated in terms of ${}_3F_2$ generalized hypergeometric series of unit argument in Appendix B.[16]

If now the expression for $M_\mp$ obtained in Appendix B [formulas (B9) and (B10)] is substituted into (8) and the resulting integral replaces the inner integral on the rhs of (4), the following integral representation for an analytic continuation of Appell's F_2 is obtained:

$$F_2(\alpha,\beta,\beta',\gamma,\gamma';x,y)=\frac{\Gamma(\gamma)\Gamma(\gamma')e^{\pm i\pi\alpha}}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\beta')}\left[\frac{e^{\pm i\pi(\beta'-\gamma')}}{\Gamma(\gamma-\beta)}\frac{1}{(2\pi i)^2}\int_{-l-i\infty}^{-l+i\infty}\int_{-n-i\infty}^{-n+i\infty}L_1(\alpha,\beta,\beta',\gamma,\gamma';s,v) \times\frac{\Gamma(\gamma'-\beta'-\alpha+\beta-v)\Gamma(-v)\Gamma(-s)}{\Gamma(\gamma-\alpha-v)}(x-1)^s(y-1)^v\,dv\,ds + \frac{e^{\pm i\pi(\beta-\gamma)}}{\Gamma(\gamma'-\beta')}\frac{1}{(2\pi i)^2}\int_{-l-i\infty}^{-l+i\infty}\int_{-n-i\infty}^{-n+i\infty}L_2(\alpha,\beta,\beta',\gamma,\gamma';s,v) \times\frac{\Gamma(\gamma-\beta-\alpha+\beta'-s)\Gamma(-s)\Gamma(-v)}{\Gamma(\gamma'-\alpha-s)}(x-1)^s(y-1)^v\,dv\,ds\right]. \quad (9)$$

The upper or lower signs, respectively, in the exponents in (9) correspond to $\mathrm{Im}\,(x)>0$ and $\mathrm{Im}\,(y)>0$, or, respectively, to $\mathrm{Im}\,(x)<0$ and $\mathrm{Im}\,(y)<0$. Use has been made of the fact that $e^{\pm i\pi s}(1-x)^s\doteq(x-1)^s$, where in either case $|\arg(x-1)|<\pi$. An asymptotic bound for large $|s|$ and/or large $|v|$ of the ${}_3F_2$'s of unit argument present implicitly in the integrands on the rhs of (9) is derived in Appendix C. From this asymptotic approximation it may be inferred that both double integrals in (9) converge absolutely, if $x\neq 1$, $y\neq 1$, and

$$\begin{aligned}|\arg(x-1)|&<\pi,\\ |\arg(y-1)|&<\pi,\\ |\arg(x-1)-\arg(y-1)|&<\pi;\end{aligned} \quad (10)$$

hence, the previously assumed conditions on x and y may be relaxed to the weaker restrictions (10); that is, (9) immediately provides a further analytic continuation of F_2.

The final step in the derivation is to evaluate the rhs of (9) in terms of infinite series in powers of $(x-1)$ and $(y-1)$. To accomplish this, it is sufficient to assume temporarily that $|x-1|<\frac{1}{2}$, $|y-1|<\frac{1}{2}$, $|\arg(x-1)|<\pi/2$, and $|\arg(y-1)|<\pi/2$. Let C_1 be the contour in the s plane consisting of the vertical straight-line segment $s=-l+iy_1$, $R\geq y_1\geq -R$, and of the semicircle $s=-l+Re^{i\varphi_1}$, $-\pi/2\leq\varphi_1\leq+\pi/2$, and let C_2 be a similar contour in the v plane (one may restrict the parameter values so that these vertical straight lines divide the increasing from the decreasing sequences of poles of the integrand). Choose R so that neither C_1 nor C_2 passes through any of the singularities of the integrands of the rhs of (9). Then as $R\to+\infty$ (in a manner so that the distance from the contours to the singularities is bounded away from zero), one has

$$\begin{aligned}&\int_{-l-i\infty}^{-l+i\infty}ds\int_{-n-i\infty}^{-n+i\infty}dv\\ &=\lim_{R\to\infty}\left\{\oint_{C_1}ds\oint_{C_2}dv+\int_{-R}^{R}(i\,dy_1)\int_{-\pi/2}^{\pi/2}iRe^{i\varphi_2}\,d\varphi_2\right.\\ &\quad+\int_{-\pi/2}^{\pi/2}iRe^{i\varphi_1}\,d\varphi_1\int_{-R}^{R}(i\,dy_2)\\ &\quad\left.-\int_{-\pi/2}^{\pi/2}iRe^{i\varphi_1}\,d\varphi_1\int_{-\pi/2}^{\pi/2}iRe^{i\varphi_2}\,d\varphi_2\right\}. \quad (11)\end{aligned}$$

[16] Note that the s and v contours divide the increasing from the decreasing sequences of poles of M_+ and M_-.

It can be shown, with the aid of the results of Appendix C, that, if the rhs of (9) is written as the limit of four integrals according to (11), then each of the latter three integrals tends to zero as $R \to \infty$; at the same time, the first integral may be evaluated with the residue calculus. Applying this process to (9), one obtains the desired series[17]:

$$
\begin{aligned}
F_2(\alpha, \beta, \beta', \gamma, \gamma'; x, y) = {} & \frac{\Gamma(\gamma)\Gamma(\gamma')\Gamma(\gamma - \beta - \gamma' + \beta')\Gamma(\gamma' - \beta' - \alpha + \beta)}{\Gamma(\beta')\Gamma(\gamma - \beta)\Gamma(\gamma' - \beta' + \beta)\Gamma(\gamma - \alpha)} e^{\pm i\pi(\alpha+\beta'-\gamma')} \\
& \times \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\alpha)_{m+n}(\beta)_m(\alpha + 1 - \gamma)_n}{m!\, n!\, (\gamma' - \beta' + \beta)_m(\alpha + 1 - \gamma' + \beta' - \beta)_n} \\
& \times {}_3F_2\left[\begin{matrix}\alpha + 1 - \gamma + \beta - \beta' + m, \gamma' - \beta' - \alpha + \beta - n, \gamma' - \beta'; \\ \gamma' - \beta' + \beta + m, 1 - \gamma + \beta + \gamma' - \beta'\end{matrix}\ 1\right](1 - x)^m(1 - y)^n \\
& + \frac{\Gamma(\gamma)\Gamma(\gamma')\Gamma(\alpha - \gamma' + \beta' - \beta)}{\Gamma(\alpha)\Gamma(\beta')\Gamma(\gamma - \beta)} e^{\pm i\pi(\alpha+\beta'-\gamma')}(y - 1)^{\gamma'-\beta'-\alpha+\beta} \\
& \times \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\gamma' - \beta' + \beta + m)_n(\beta)_m(1 - \gamma + \beta + \gamma' - \beta')_n}{m!\, n!\, (1 + \gamma' - \beta' - \alpha + \beta)_n} \\
& \times {}_3F_2\left[\begin{matrix}\alpha + 1 - \gamma + \beta - \beta' + m, -n, \gamma' - \beta'; \\ \gamma' - \beta' + \beta + m, 1 - \gamma + \beta + \gamma' - \beta'\end{matrix}\ 1\right](1 - x)^m(1 - y)^n \\
& + \frac{\Gamma(\gamma)\Gamma(\gamma')\Gamma(\gamma' - \beta' - \gamma + \beta)\Gamma(\gamma - \beta - \alpha + \beta')}{\Gamma(\beta)\Gamma(\gamma' - \beta')\Gamma(\gamma - \beta + \beta')\Gamma(\gamma' - \alpha)} e^{\pm i\pi(\alpha+\beta-\gamma)} \\
& \times \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\alpha)_{m+n}(\beta')_n(\alpha + 1 - \gamma')_m}{m!\, n!\, (\gamma - \beta + \beta')_n(\alpha + 1 - \gamma + \beta - \beta')_m} \\
& \times {}_3F_2\left[\begin{matrix}\alpha + 1 - \gamma' + \beta' - \beta + n, \gamma - \beta - \alpha + \beta' - m, \gamma - \beta; \\ \gamma - \beta + \beta' + n, 1 - \gamma' + \beta' + \gamma - \beta\end{matrix}\ 1\right](1 - x)^m(1 - y)^n \\
& + \frac{\Gamma(\gamma)\Gamma(\gamma')\Gamma(\alpha - \gamma + \beta - \beta')}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma' - \beta')} e^{\pm i\pi(\alpha+\beta-\gamma)}(x - 1)^{\gamma-\beta-\alpha+\beta'} \\
& \times \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\gamma - \beta + \beta' + n)_m(\beta')_n(1 - \gamma' + \beta' + \gamma - \beta)_m}{m!\, n!\, (1 + \gamma - \beta - \alpha + \beta')_m} \\
& \times {}_3F_2\left[\begin{matrix}\alpha + 1 - \gamma' + \beta' - \beta + n, -m, \gamma - \beta; \\ \gamma - \beta + \beta' + n, 1 - \gamma' + \beta' + \gamma - \beta\end{matrix}\ 1\right](1 - x)^m(1 - y)^n. \qquad (12)
\end{aligned}
$$

The rhs of (12) consists of four terms, each containing a doubly infinite sum; the third and fourth terms differ from the first and second, respectively, by simultaneous interchange of x and y, β and β', γ and γ'. The cases $\gamma - \beta - \gamma' + \beta' =$ integer, $\gamma' - \beta' - \alpha + \beta =$ integer, and $\gamma - \beta - \alpha + \beta' =$ integer have been avoided in obtaining (12); these special cases aside, the restrictions put on the parameters are no longer needed, and the validity of (12) may be extended, by analytic continuation in the parameters, to almost all sets of values of α, β, β', γ, and γ'. With the aid of the results of Appendix C, it may be concluded that all four series on the rhs of (12) will converge absolutely if $|x - 1| + |y - 1| < 1$. Thus, Eq. (12) is the essential result of this paper.

It is possible to write the second and the fourth term on the rhs of (12) in the form of Appell's hypergeometric function F_3; this will now be done. Consider the hypergeometric function[18]

$$
\begin{aligned}
& x^{-\beta}y^{\beta'-\gamma'}F_3\left(\beta, \gamma' - \beta', \beta + 1 - \gamma, 1 - \beta', \gamma' - \beta' + \beta - \alpha + 1; \frac{1 - y}{x}, \frac{y - 1}{y}\right) \\
& \qquad = \sum_{m=0}^{\infty} \frac{(\beta)_m(\beta + 1 - \gamma)_m}{(\gamma' - \beta' + \beta - \alpha + 1)_m m!} {}_2F_1\left[\begin{matrix}\gamma' - \beta', 1 - \beta'; \\ \gamma' - \beta' + \beta - \alpha + 1 + m\end{matrix}\ (y - 1)/y\right]\left(\frac{1 - y}{x}\right)^m x^{-\beta}y^{\beta'-\gamma'}. \qquad (13)
\end{aligned}
$$

[17] The function F_2 and its analytic continuations will be singular only on the planes $x = 0$, $y = 0$, $x = 1$, $y = 1$, $x = \infty$, $y = \infty$, $x + y = 1$ (cf. Ref. 1, pp. 42–49). The choice of signs in the exponents of (12) corresponds to a choice of paths for the analytic continuation described in the text.

[18] If this function is multiplied by $(1 - y)^{\gamma'-\beta'-\alpha+\beta}$, it will be a solution to the system of partial differential equations satisfied by F_2. Compare P. O. M. Olsson, Arkiv Fysik 25, 473 (1964), especially p. 480.

If one applies Euler's transformation to the hypergeometric function under the summation sign, the rhs of (13) becomes

$$\sum_{m=0}^{\infty} \frac{(\beta)_m(\beta+1-\gamma)_m}{(\gamma'-\beta'+\beta-\alpha+1)_m m!}\, {}_2F_1\left[\begin{matrix}\gamma'-\beta', \gamma'+\beta-\alpha+m; 1-y \\ \gamma'-\beta'+\beta-\alpha+1+m\end{matrix}\right] x^{-\beta-m}(1-y)^m.$$

If $|1-x|+|1-y|<1$, one may write the above sum as a series in powers of $(1-x)$ and $(1-y)$ by expressing the hypergeometric function as an infinite series in powers of $(1-y)$, expanding $x^{-\beta-m} = [1-(1-x)]^{-\beta-m}$ as an infinite series in powers of $(1-x)$, interchanging the order of summation, and collecting terms. The following series is then obtained:

$$\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \frac{(\gamma'+\beta-\alpha)_q(\beta)_p(\gamma'-\beta')_q}{(\gamma'-\beta'+\beta-\alpha+1)_q p!\, q!}\, {}_3F_2\left[\begin{matrix}\beta+p, \beta+1-\gamma, -q; 1 \\ \beta'+1-\gamma'-q, \gamma'+\beta-\alpha\end{matrix}\right](1-x)^p(1-y)^q. \tag{14}$$

At this point, the terminating series ${}_3F_2$ in the summation (14) is transformed into another terminating ${}_3F_2$ series as follows[19]:

$${}_3F_2\left[\begin{matrix}\beta+p, \beta+1-\gamma, -q; 1 \\ \beta'+1-\gamma'-q, \gamma'+\beta-\alpha\end{matrix}\right]$$
$$= \frac{(1-\gamma+\beta+\gamma'-\beta')_q(\gamma'-\beta'+\beta+p)_q}{(\gamma'-\beta')_q(\gamma'+\beta-\alpha)_q}$$
$$\times {}_3F_2\left[\begin{matrix}\alpha+1-\gamma+\beta-\beta'+p, -q, \gamma'-\beta'; 1 \\ \gamma'-\beta'+\beta+p, 1-\gamma+\beta+\gamma'-\beta'\end{matrix}\right]. \tag{15}$$

If the rhs of (15) is used in (14), it is clear that the double sum obtained is identical with the second double sum of the rhs of (12). With appropriate changes, the fourth double sum in (12) may also be written as an F_3. The first and third terms of (12) are both regular at the point $x=y=1$; no hypergeometric form seems to exist for them in general.

If one of the Eulerian transformations is first applied to the F_2, for example,[20]

$$F_2(\alpha,\beta,\beta',\gamma,\gamma';x,y) = (1-x)^{-\alpha} \times F_2(\alpha,\gamma-\beta,\beta',\gamma,\gamma';x/(x-1),y/(1-x)), \tag{16}$$

and the process leading to (12) is then applied to the F_2 on the rhs of (16), an analytic continuation of F_2 to a different region of x, y space results. In the example, four double series in the variables

$$1-\frac{x}{x-1} = \frac{-1/x}{1-(1/x)},$$
$$1-\frac{y}{1-x} = \frac{(1/x)+(1/y)-(1/xy)}{[1-(1/x)](1/y)}$$

appear. These double series converge provided that $(1/x, 1/y)$ is in the region

$$\left|\frac{1/x}{1-(1/x)}\right| + \left|\frac{(1/x)+(1/y)-(1/xy)}{(1/y)[1-(1/x)]}\right| < 1.$$

The region is shaped roughly like the interior of a truncated cone for which the apex is $(1/x, 1/y) = (0,0)$, and the "axis" is the surface $(1/x)+(1/y)-(1/xy)=0$.[21]

A final remark: Since both types of confluent hypergeometric function $\Phi(b,c;x)$ and $\Psi(b,c;x)$ possess integral representations of the Mellin–Barnes type,[22] it is clear that the functions

$$F_P(a,b_1,b_2,c_1,c_2,x_1,x_2)$$

and

$$F_R(a,b_1,b_2,c_1,c_2,x_1,x_2)$$

defined by Olsson[23] both may be represented by Mellin–Barnes double integrals. Manipulations of these integrals similar to those carried out earlier in the present work yield series expansions in powers of $(1-x_1)$ and $(1-x_2)$ for the functions F_P or F_R, which results have already been obtained by Olsson[23] (who used other methods).

APPENDIX A

An asymptotic approximation to the hypergeometric function $F(a+\lambda, b; c; z)$ for large λ may be found by a method outlined by Watson.[24] The result is as follows[25]:

$${}_2F_1\left[\begin{matrix}a+\lambda, b; z \\ c\end{matrix}\right]$$
$$\sim \frac{\Gamma(c)}{\Gamma(c-b)}(-z)^{-b}\lambda^{-b}\left[1+O\left(\frac{1}{|\lambda|}\right)\right] + \frac{\Gamma(c)}{\Gamma(b)}(1-z)^{c-a-b-\lambda}z^{b-c}\lambda^{b-c}\left[1+O\left(\frac{1}{|\lambda|}\right)\right]$$
as $|\lambda|\to\infty$.

[19] L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge University Press, London, 1966), p. 120; the formula given here is that relating $Fp(0;4,5)$ to $Fn(5;1,2)$.

[20] Reference 1, p. 25; Ref. 2, p. 240.

[21] Compare A. Erdélyi, Acta Math. **83**, 131 (1950).

[22] Reference 2, p. 256, Eqs. (4) and (5).

[23] P. O. M. Olsson: Ref. 6 and Arkiv Fysik **28**, 113 (1965). Olsson shows that these functions satisfy the same system of partial equations as is satisfied by F_2.

[24] G. N. Watson, Trans. Cambridge Phil. Soc. **22**, 277 (1918), in particular, see Sec. 17.

[25] The asymptotic approximations for large $|b|$ to $F(a,b;c;z)$, given in Ref. 2, p. 77, formulas (13), (14), and (15), are incorrect, for example, in the case that $a=c$.

The above formula holds if, but not only if,

$$|\arg(\lambda)| \leq \pi/2, \quad |\arg(z)| < \pi, \quad |\arg(-z)| < \pi,$$

and

$$|\arg(1-z)| < \pi.$$

APPENDIX B

Consider the integrals I_- and I_+, where

$$I_\mp = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \Gamma(a+s)\Gamma(b+s)\Gamma(c+s) \times \Gamma(1-d-s)\Gamma(1-f-s)\Gamma(-s)e^{\mp i\pi s}\,ds; \tag{B1}$$

the contour is supposed to be indented, where necessary, so that the set of poles of $\Gamma(a+s)\times\Gamma(b+s)\Gamma(c+s)$ lies on its left, and so that the poles of $\Gamma(1-d-s)\Gamma(1-f-s)\Gamma(-s)$ lie on its right. Let $K(a,b,c,d,f)$ be defined as follows:

$$K(a,b,c,d,f) = \frac{1}{2\pi i}e^{i\pi(1-d)}I_- - \frac{1}{2\pi i}e^{-i\pi(1-d)}I_+; \tag{B2}$$

combining (B1) and (B2), one has

$$K(a,b,c,d,f) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\Gamma(a+s)\Gamma(b+s)\Gamma(c+s)}{\Gamma(d+s)} \times \Gamma(1-f-s)\Gamma(-s)\,ds. \tag{B3}$$

The reflection formula for the gamma function[26] was used to obtain (B3).

Using (B2), one may solve for I_- and I_+ in terms of $K(a,b,c,d,f)$ and $K(a,c,b,f,d)$, with the following result ($d-f \neq$ integer):

$$I_\mp = K(a,b,c,d,f)\frac{\pi e^{\mp i\pi(1-f)}}{\sin\pi(f-d)} + K(a,c,b,f,d)\frac{\pi e^{\mp i\pi(1-d)}}{\sin\pi(d-f)}. \tag{B4}$$

The integral (B3) for $K(a,b,c,d,f)$ may be expressed in terms of ${}_3F_2$ generalized hypergeometric series with unit argument as follows[27]:

$$K(a,b,c,d,f) = \frac{\Gamma(1-f+a)\Gamma(1-f+b)\Gamma(a)\Gamma(b)\Gamma(c)}{\Gamma(1-f+a+b)\Gamma(d)} \times {}_3F_2\left[\begin{matrix}a, b, d-c;\\ 1-f+a+b, d\end{matrix}\,1\right]. \tag{B5}$$

While, in general, the series

$${}_3F_2\left[\begin{matrix}\alpha_1, \alpha_2, \alpha_3;\\ \beta_1, \beta_2\end{matrix}\,1\right] = \sum_{n=0}^{\infty}\frac{(\alpha_1)_n(\alpha_2)_n(\alpha_3)_n}{(\beta_1)_n(\beta_2)_n n!}$$

converges only for $\mathrm{Re}\,(\beta_1+\beta_2-\alpha_1-\alpha_2-\alpha_3) > 0$, it may be shown that the complete analytic continuation of

$$\frac{1}{\Gamma(\beta_1)\Gamma(\beta_2)\Gamma(\beta_1+\beta_2-\alpha_1-\alpha_2-\alpha_3)} \times {}_3F_2\left[\begin{matrix}\alpha_1, \alpha_2, \alpha_3;\\ \beta_1, \beta_2\end{matrix}\,1\right]$$

is an entire analytic function of all five complex parameters α_1, α_2, α_3, β_1, and β_2. In this sense, the ${}_3F_2$ has a value for any set of values of the five parameters, except only those sets corresponding to poles of $\Gamma(\beta_1)\Gamma(\beta_2)\Gamma(\beta_1+\beta_2-\alpha_1-\alpha_2-\alpha_3)$.

It will be convenient to apply one of the transformation formulas of Thomae[28] to the ${}_3F_2$ on the rhs of (B5):

$${}_3F_2\left[\begin{matrix}\alpha_1, \alpha_2, \alpha_3;\\ \beta_1, \beta_2\end{matrix}\,1\right] = \frac{\Gamma(\beta_2)\Gamma(\beta_1+\beta_2-\alpha_1-\alpha_2-\alpha_3)}{\Gamma(\beta_2-\alpha_3)\Gamma(\beta_1+\beta_2-\alpha_1-\alpha_2)} \times {}_3F_2\left[\begin{matrix}\alpha_3, \beta_1-\alpha_1, \beta_1-\alpha_2;\\ \beta_1, \beta_1+\beta_2-\alpha_1-\alpha_2\end{matrix}\,1\right]. \tag{B6}$$

Using (B5) and (B6), one obtains the following expression for $K(a,b,c,d,f)$:

$$K(a,b,c,d,f) = \frac{\Gamma(a+1-f)\Gamma(b+1-f)\Gamma(c+1-f)\Gamma(a)\Gamma(b)}{\Gamma(1-f+a+b)\Gamma(1-f+d)} \times {}_3F_2\left[\begin{matrix}d-c, 1-f+b, 1-f+a;\\ 1-f+a+b, 1-f+d\end{matrix}\,1\right]. \tag{B7}$$

The integrals whose evaluation in terms of ${}_3F_2$'s of unit argument is the goal of this appendix are M_- and M_+, where[29]

$$M_\mp = e^{\mp i\pi v}\frac{1}{2\pi i}\int_{-p-i\infty}^{-p+i\infty}\Gamma(\alpha+s+v+w) \times \Gamma(\beta+s+w)\Gamma(\beta'+v+w) \times \Gamma(\gamma-\beta-\alpha-s-v-w) \times \Gamma(\gamma'-\beta'-\alpha-s-v-w) \times \Gamma(-w)e^{\mp i\pi w}\,dw. \tag{B8}$$

Taking $a = \alpha+s+v$, $b = \beta+s$, $c = \beta'+v$, $d = \alpha+1-\gamma+\beta+s+v$, $f = \alpha+1-\gamma'+\beta'+s+v$, and using (B4) and (B7) to evaluate (B8), one

[26] Reference 2, p. 3, formula (6).

[27] Reference 19, p. 112.

[28] J. Thomae, J. Reine Angew. Math. **87**, 26 (1879); F. J. W. Whipple, Proc. London Math. Soc., Ser. 2, **23**, 104 (1925); see also Ref. 19, p. 116; in Whipple's notation, (B6) is stated as follows: $Fp(0;4,5) = Fp(0;3,4)$.

[29] Note that the rhs of (B8) differs from the corresponding expression in (8) in the text by a change of the variable of integration from u to w, where $u = v+w$.

obtains

$$M_{\mp} = e^{\pm i\pi s}[e^{\mp i\pi(\gamma'-\beta'-\alpha)}\Gamma(\gamma'-\alpha-s)\Gamma(\gamma'-\beta'-\alpha+\beta-v)\Gamma(\gamma'-\beta')L_1(\alpha,\beta,\beta',\gamma,\gamma';s,v) + e^{\mp i\pi(\gamma-\beta-\alpha)}\Gamma(\gamma-\alpha-v)\Gamma(\gamma-\beta-\alpha+\beta'-s)\Gamma(\gamma-\beta)L_2(\alpha,\beta,\beta',\gamma,\gamma';s,v)], \quad \text{(B9)}$$

where

$$L_1(\alpha,\beta,\beta',\gamma,\gamma';s,v) = \frac{\Gamma(\gamma-\beta-\gamma'+\beta')\Gamma(\alpha+s+v)\Gamma(\beta+s)}{\Gamma(\gamma'-\beta'+\beta+s)} \times {}_3F_2\left[\begin{matrix}\alpha+1-\gamma+\beta-\beta'+s,\ \gamma'-\beta'-\alpha+\beta-v,\ \gamma'-\beta';\\ \gamma'-\beta'+\beta+s,\ 1-\gamma+\beta+\gamma'-\beta'\end{matrix}\ 1\right],$$

$$L_2(\alpha,\beta,\beta',\gamma,\gamma';s,v) = L_1(\alpha,\beta',\beta,\gamma',\gamma;v,s). \quad \text{(B10)}$$

Note that neither $L_1(\alpha,\beta,\beta',\gamma,\gamma';s,v)$ nor $L_2(\alpha,\beta,\beta',\gamma,\gamma';s,v)$ has an increasing sequence of poles in either the s plane of the v plane.

To obtain (B4), it was assumed that $d-f\neq$ integer. This implies that $\beta-\gamma-\beta'+\gamma'\neq$ integer in (B9). An expression for $M_{\mp}$ when $\beta-\gamma-\beta'-\gamma'=$ integer may be obtained as a limit of (B9), or by other means; neither these nor other special values of the parameters will be considered in this paper.

APPENDIX C

In order to obtain (12) from (9) in the text and in order to ascertain the region of convergence of the double series in (12), it is useful to establish that, for appropriate ranges of values of σ and τ, one has the inequality

$$\left|\frac{1}{\Gamma(f)}\,{}_3F_2\left[\begin{matrix}a-\sigma,\ b+\tau,\ c;\\ d+\tau,\ f\end{matrix}\ 1\right]\right| \leq K(1+|\sigma|)^P(1+|\tau|)^Q; \quad \text{(C1)}$$

here a, b, c, d, and f are given parameters, while K, P, and Q are positive numbers whose values may depend on a, b, c, d, and f, but are independent of σ and τ. The values of the parameters and the ranges of values of σ and τ which must be considered fall into two cases: (1) $\mathrm{Re}\,(d+f-a-b-c)>0$, $\mathrm{Re}\,(d)>0$, $\mathrm{Re}\,(d-b)>0$, while σ and τ range over all values such that $\mathrm{Re}\,(\sigma)\geq 0$, $\mathrm{Re}\,(\tau)\geq 0$; (2) a, b, c, d, f take essentially arbitrary given values (except that neither d nor $d+f-a-b-c$ should be zero or a negative integer) while both σ and τ range over all nonnegative integer values.

It can be shown, with the aid of the contiguous function relations for the hypergeometric series ${}_3F_2$ of unit argument,[30] that the demonstration of the inequality (C1) for both of the above cases reduces to proving that (C1) holds under the circumstances that

$$\mathrm{Re}\,(d+f-a-b-c)>0, \quad \mathrm{Re}\,(d)>0,$$
$$\mathrm{Re}\,(d-b)>0, \quad \mathrm{Re}\,(b)>0,$$
$$\mathrm{Re}\,(f-c)>0, \quad \mathrm{Re}\,(c)>0,$$

while σ and τ take on all values for which $\mathrm{Re}\,(\sigma)\geq 0$, $\mathrm{Re}\,(\tau)\geq 0$. With these restrictions the ${}_3F_2$ of unit argument is given by a convenient integral representation:

$$\frac{1}{\Gamma(f)}\,{}_3F_2\left[\begin{matrix}a-\sigma,\ b+\tau,\ c;\\ d+\tau,\ f\end{matrix}\ 1\right] = \frac{\Gamma(d+\tau)}{\Gamma(d-b)\Gamma(b+\tau)\Gamma(f-c)\Gamma(c)} \times \int_0^1\int_0^1 u^{b+\tau-1}(1-u)^{d-b-1} \times v^{c-1}(1-v)^{f-c-1}(1-uv)^{-a+\sigma}\,du\,dv. \quad \text{(C2)}$$

To see that the rhs of (C2) equals the lhs, one may expand $(1-uv)^{-a+\sigma}$ in powers of uv and integrate the resulting series term by term.[31]

The proof of (C1) by means of (C2) depends on a preliminary result concerning $|\Gamma(d+\tau)/\Gamma(b+\tau)|$. Let r be the least positive integer such that

$$\mathrm{Re}\,(b-d+r)>0;$$

then one has[31]

$$\frac{\Gamma(d+\tau)}{\Gamma(b+\tau)} = \frac{(b+\tau)_r}{\Gamma(b-d+r)}\int_0^1 u^{d+\tau-1}(1-u)^{b-d+r-1}\,du.$$

Taking absolute values of both sides of this equation one obtains

$$\left|\frac{\Gamma(d+\tau)}{\Gamma(b+\tau)}\right| \leq \frac{|(b+\tau)_r|}{|\Gamma(b-d+r)|} \times \int_0^1 u^{\mathrm{Re}\,(d)-1}(1-u)^{\mathrm{Re}\,(b-d)+r-1}\,du \leq K_1(1+|\tau|)^r, \quad \text{(C3)}$$

[30] E. D. Rainville, Bull. Am. Math. Soc. **51**, 714 (1945).

[31] Reference 2, p. 9, Eqs. (1) and (5).

where K_1 is positive and can be chosen so as not to vary with τ, but so that (C3) holds for all τ such that Re $(\tau) \geq 0$.

With this result the proof of (C1) is immediate, for if one takes the absolute value of both sides of (C2) one obtains

$$\left|\frac{1}{\Gamma(f)}\,{}_3F_2\left[\begin{matrix}a-\sigma, b+\tau, c; 1\\ d+\tau, f\end{matrix}\right]\right| \leq \left|\frac{1}{\Gamma(d-b)\Gamma(f-c)\Gamma(c)}\right|\left|\frac{\Gamma(d+\tau)}{\Gamma(b+\tau)}\right| \times \int_0^1\int_0^1 u^{\mathrm{Re}\,(b)-1}(1-u)^{\mathrm{Re}\,(d-b)-1}\,v^{\mathrm{Re}\,(c)-1}(1-v)^{\mathrm{Re}\,(f-c)-1}(1-uv)^{-\mathrm{Re}\,(a)}\,du \leq K_2(1+|\tau|)^r, \quad \text{(C4)}$$

where K_2 is positive and is fixed at so large a value that (C4) holds for all τ such that Re $(\tau) \geq 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Finite Transformations of $SU3$

DOUGLAS FRANCIS HOLLAND
Physics Department, University of Arizona, Tucson, Arizona

(Received 4 August 1968)

The properties of $SU3$ finite transformations are investigated. These transformations on the defining three-dimensional complex space are parameterized in a form employing three factors, two of which are the Euler parameterization of an $SU2$ subgroup. The irreducible representations of the factored parameterization are found explicitly. The volume element is calculated and the orthogonality relation is verified. Spherical harmonic basis states are derived as a specialization of the transformation matrix. Another result is a definition of triality and a simple proof that it is additive modulus three.

I. INTRODUCTION

For some time now, the group $SU3$ has been thought to carry the symmetry of the elementary particles. Considerable work has been done concerning the infinitesimal generators of the group. As an alternative mathematical technique, we wish to investigate the global properties.

A parameterization of all unitary groups was given by Murnaghan.[1] We derive the parameterization used by Nelson[2] and set the minimum ranges. This is of special interest as it employs two factors which are $SU2$ transformations in the Euler form.

Chacon and Moshinsky[3] derive the IR's in Murnaghan's parameterization by extensive use of Weyl reflections. Nelson,[2] using his parameterization, restricted the representation matrices to a particular right-hand state and so derived a single column of the matrix which acts as a set of spherical harmonic basis states. Both Nelson and previously Beg and Ruegg[4] employ differential operators to represent the infinitesimal generators in their derivations. We derive the complete matrix for each IR; the $SU2$ factors are known and the third factor is evaluated by employing tensor basis states. By applying a finite transformation to these tensors the spherical harmonic basis states are derived.

Weyl[5] and Murnaghan[1] discuss the dependence of the volume element on the class parameters for integration concerning the characters of the group. We evaluate the complete dependence of the volume element on all parameters for Nelson's parameterization.

Symmetries of the transformations were investigated, and one result is a natural definition of triality. Baird and Biedenharn,[6] and Hagan and Macfarlane[7] prove triality is additive modulus three. We provide a proof that follows very simply from our definition.

In this paper we are guided by the strong analogy that exists between our parameterization and that for $SU2$. We follow closely the same procedures used in $SU2$ to derive the corresponding results for $SU3$.

[1] Francis D. Murnaghan, *The Unitary and Rotation Group* (Spartan Books, Washington, D.C., 1962).
[2] T. J. Nelson, J. Math. Phys. **8**, 857 (1967).
[3] E. Chacon and M. Moshinsky, Phys. Letters **23**, 567 (1966).
[4] M. Beg and H. Ruegg, J. Math. Phys. **6**, 677 (1965).
[5] Herman Weyl, *Classical Groups* (Princeton University Press, Princeton, New Jersey, 1946).
[6] G. C. Baird and L. C. Biedenharn, *Proceedings of the Coral Gables Conference on Symmetry Principles at High Energy*, B. Kursunoglu, Ed. (W. H. Freeman and Co., San Francisco, 1964).
[7] C. R. Hagan and A. J. Macfarlane, J. Math. Phys. **5**, 1335 (1964).

where K_1 is positive and can be chosen so as not to vary with τ, but so that (C3) holds for all τ such that Re $(\tau) \geq 0$.

With this result the proof of (C1) is immediate, for if one takes the absolute value of both sides of (C2) one obtains

$$\left|\frac{1}{\Gamma(f)}\,{}_3F_2\left[\begin{matrix}a-\sigma, b+\tau, c; 1\\ d+\tau, f\end{matrix}\right]\right| \leq \left|\frac{1}{\Gamma(d-b)\Gamma(f-c)\Gamma(c)}\right|\left|\frac{\Gamma(d+\tau)}{\Gamma(b+\tau)}\right| \times \int_0^1\int_0^1 u^{\mathrm{Re}\,(b)-1}(1-u)^{\mathrm{Re}\,(d-b)-1}\,v^{\mathrm{Re}\,(c)-1}(1-v)^{\mathrm{Re}\,(f-c)-1}(1-uv)^{-\mathrm{Re}\,(a)}\,du \leq K_2(1+|\tau|)^r, \quad \text{(C4)}$$

where K_2 is positive and is fixed at so large a value that (C4) holds for all τ such that Re $(\tau) \geq 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Finite Transformations of $SU3$

DOUGLAS FRANCIS HOLLAND
Physics Department, University of Arizona, Tucson, Arizona

(Received 4 August 1968)

The properties of $SU3$ finite transformations are investigated. These transformations on the defining three-dimensional complex space are parameterized in a form employing three factors, two of which are the Euler parameterization of an $SU2$ subgroup. The irreducible representations of the factored parameterization are found explicitly. The volume element is calculated and the orthogonality relation is verified. Spherical harmonic basis states are derived as a specialization of the transformation matrix. Another result is a definition of triality and a simple proof that it is additive modulus three.

I. INTRODUCTION

For some time now, the group $SU3$ has been thought to carry the symmetry of the elementary particles. Considerable work has been done concerning the infinitesimal generators of the group. As an alternative mathematical technique, we wish to investigate the global properties.

A parameterization of all unitary groups was given by Murnaghan.[1] We derive the parameterization used by Nelson[2] and set the minimum ranges. This is of special interest as it employs two factors which are $SU2$ transformations in the Euler form.

Chacon and Moshinsky[3] derive the IR's in Murnaghan's parameterization by extensive use of Weyl reflections. Nelson,[2] using his parameterization, restricted the representation matrices to a particular right-hand state and so derived a single column of the matrix which acts as a set of spherical harmonic basis states. Both Nelson and previously Beg and Ruegg[4] employ differential operators to represent the infinitesimal generators in their derivations. We derive the complete matrix for each IR; the $SU2$ factors are known and the third factor is evaluated by employing tensor basis states. By applying a finite transformation to these tensors the spherical harmonic basis states are derived.

Weyl[5] and Murnaghan[1] discuss the dependence of the volume element on the class parameters for integration concerning the characters of the group. We evaluate the complete dependence of the volume element on all parameters for Nelson's parameterization.

Symmetries of the transformations were investigated, and one result is a natural definition of triality. Baird and Biedenharn,[6] and Hagan and Macfarlane[7] prove triality is additive modulus three. We provide a proof that follows very simply from our definition.

In this paper we are guided by the strong analogy that exists between our parameterization and that for $SU2$. We follow closely the same procedures used in $SU2$ to derive the corresponding results for $SU3$.

[1] Francis D. Murnaghan, *The Unitary and Rotation Group* (Spartan Books, Washington, D.C., 1962).
[2] T. J. Nelson, J. Math. Phys. **8**, 857 (1967).
[3] E. Chacon and M. Moshinsky, Phys. Letters **23**, 567 (1966).
[4] M. Beg and H. Ruegg, J. Math. Phys. **6**, 677 (1965).
[5] Herman Weyl, *Classical Groups* (Princeton University Press, Princeton, New Jersey, 1946).
[6] G. C. Baird and L. C. Biedenharn, *Proceedings of the Coral Gables Conference on Symmetry Principles at High Energy*, B. Kursunoglu, Ed. (W. H. Freeman and Co., San Francisco, 1964).
[7] C. R. Hagan and A. J. Macfarlane, J. Math. Phys. **5**, 1335 (1964).

We use the following set of infinitesimal generators:

$$I_1=\begin{pmatrix}0&1&0\\1&0&0\\0&0&0\end{pmatrix},\quad I_2=\begin{pmatrix}0&-i&0\\i&0&0\\0&0&0\end{pmatrix},$$
$$I_3=\begin{pmatrix}1&0&0\\0&-1&0\\0&0&0\end{pmatrix},\quad I_4=\begin{pmatrix}0&0&1\\0&0&0\\1&0&0\end{pmatrix},$$
$$I_5=\begin{pmatrix}0&0&-i\\0&0&0\\i&0&0\end{pmatrix},\quad I_6=\begin{pmatrix}0&0&0\\0&0&1\\0&1&0\end{pmatrix},\tag{1}$$
$$I_7=\begin{pmatrix}0&0&0\\0&0&-i\\0&i&0\end{pmatrix},\quad I_8=\begin{pmatrix}1/\sqrt{3}&0&0\\0&1/\sqrt{3}&0\\0&0&-2/\sqrt{3}\end{pmatrix},$$
$$\mathrm{Tr}\,(I_i\ \ I_j)=2\delta_{ij}.$$

The basis states are chosen such that I_3 and I_8 are diagonal in all representations, and thus serve to partially label the states

$$\bar{Y}\equiv\frac{1}{\sqrt{3}}I_8,\quad \bar{Y}\Psi\begin{pmatrix}Y\\M\end{pmatrix}=Y\Psi\begin{pmatrix}Y\\M\end{pmatrix},$$
$$\bar{M}\equiv\tfrac{1}{2}I_3,\quad \bar{M}\Psi\begin{pmatrix}Y\\M\end{pmatrix}=M\Psi\begin{pmatrix}Y\\M\end{pmatrix}.\tag{2}$$

The eigenvalues (Y, M) of the states within an IR are displayed in two-dimensional weight diagrams.

The addition of $I^2=\frac{1}{4}(I_1^2+I_2^2+I_3^2)$ forms a complete set of commuting operators which serve to label the states

$$I^2\Psi\begin{pmatrix}IY\\M\end{pmatrix}=I(I+1)\Psi\begin{pmatrix}IY\\M\end{pmatrix}.\tag{3}$$

II. PARAMETERIZATION OF GROUP TRANSFORMATIONS

We now derive the following parameterization:

$$U(\alpha)=T_2T_3T_2',$$

where

$$T_2=e^{-i\alpha I_3/2}e^{-i\beta I_2/2}e^{-i\gamma I_3/2},\quad T_2'\text{ with primed variables},$$
$$T_3=e^{-i\rho I_8/\sqrt{3}}e^{-i\nu I_7}.\tag{4}$$

Transformations in $SU3$ map $Z\to Z'$, where Z and Z' are vectors in three-dimensional complex space such that the norm of Z' equals the norm of Z. We set $Z^\dagger Z=Z'^\dagger Z'=1$.

Beg and Ruegg[4] show that Z and Z' can be parameterized as follows:

$$Z=\begin{pmatrix}e^{i\Phi_1}\cos\theta\\e^{i\Phi_2}\sin\theta\cos\Phi\\e^{i\Phi_3}\sin\theta\sin\Phi\end{pmatrix},\quad\begin{matrix}0\le\Phi_i\le 2\pi,\\0\le\theta\le\pi/2,\\0\le\Phi\le\pi/2.\end{matrix}\tag{5}$$

By expansion of the exponentials, one can show that a solution for $(\alpha', \beta', \gamma')$ exists such that

$$T_2'Z=\begin{pmatrix}0\\e^{i\Phi_3}\cos\Phi''\\e^{i\Phi_3}\sin\Phi''\end{pmatrix},\quad 0\le\Phi''\le\pi/2.\tag{6}$$

The minimum ranges required are

$$0\le\alpha',\gamma'\le 4\pi,\quad 0\le\beta'\le\pi.$$

Therefore, if

$$e^{\frac{2}{3}i\rho}=e^{-i\Phi_3},\qquad 0\le\rho\le 3\pi,\tag{7a}$$
$$\nu=\pi/2-\Phi'',\quad 0\le\nu\le\pi/2,\tag{7b}$$

then

$$T_3T_2'Z=\begin{pmatrix}0\\0\\1\end{pmatrix}.\tag{8}$$

T_2 is the most general special unitary matrix with (1) in the (3, 3) position. Let U be a general matrix of $SU3$. From Eq. (8), and since the matrices are unitary, we have

$$T_3T_2'U^{-1}=\begin{pmatrix}u_{11}'&u_{12}'&0\\u_{21}'&u_{12}'&0\\0&0&1\end{pmatrix}\equiv T_2^{-1}.\tag{9a}$$

Therefore,

$$U=T_2T_3T_2'.\tag{9b}$$

We note that

$$e^{-i\pi I_3/2}I_7e^{i\pi I_3/2}=I_6\tag{10}$$

and another redefinition of γ and α' to absorb $e^{i\pi I_3/2}$ would allow the replacement of I_7 by I_6. Also, an interchange of the role of the first and second components of Z, would allow the replacement of I_6 or I_7 by I_4 or I_5. We choose I_4 to be consistent with Nelson's[2] work.

III. EXPLICIT DETERMINATION OF THE TRANSFORMATION MATRICES

We now seek to generalize our result for the defining three-dimensional representation to all IR's. In what follows, we use the integers (λ, μ) to denote an IR (see Ref. 8), and (I, Y, M) to denote a state within the IR. The symbols (λ, μ) are suppressed unless needed.

From the commutation rules, it follows that an $SU2$ subalgebra exists, composed of I_1, I_2, and I_3. Hence the basis states are chosen such that

$$I^2\equiv I_1^2+I_2^2+I_3^2,\tag{11a}$$

[8] G. C. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963).

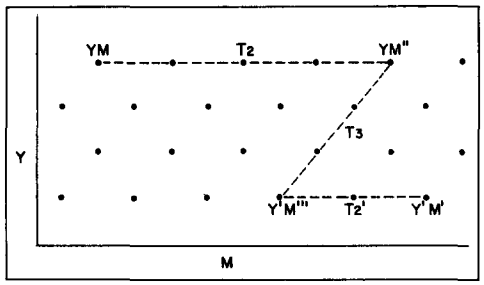


FIG. 1. Weight diagram for derivation of the representations of $SU3$.

I_3 and I_8 are diagonal, and

$$\left\langle \begin{matrix} IY \\ M \end{matrix} \middle| e^{-i\alpha I_3/2} e^{-i\beta I_2/2} e^{-i\gamma I_3/2} \middle| \begin{matrix} I'Y' \\ M' \end{matrix} \right\rangle = D^I_{MM'}(\alpha, \beta, \gamma)\delta_{YY'}\delta_{II'}, \quad (11b)$$

$$\left\langle \begin{matrix} IY \\ M \end{matrix} \middle| e^{-i\rho I_8/\sqrt{3}} \middle| \begin{matrix} I'Y' \\ M' \end{matrix} \right\rangle = e^{-i\rho Y}\delta_{YY'}\delta_{II'}\delta_{MM'}. \quad (11c)$$

Now consider a weight diagram (Fig. 1) which displays the basis states according to the (Y, M) values and consider

$$\left\langle \begin{matrix} IY \\ M \end{matrix} \middle| T_2 T_3 T_2' \middle| \begin{matrix} I'Y' \\ M \end{matrix} \right\rangle.$$

T_2 and T_2' connect states in the same horizontal line, while T_3 connects states on the diagonal shown. This gives a relation between M''' and M'':

$$M'' \equiv M''' + \tfrac{1}{2}(Y - Y'). \quad (12)$$

Summarizing the above,

$$\left\langle \begin{matrix} IY \\ M \end{matrix} \middle| D(\alpha) \middle| \begin{matrix} I'Y' \\ M' \end{matrix} \right\rangle = \sum_{M''} e^{i\rho Y} D^I_{MM''}(\alpha, \beta, \gamma) \times \left\langle \begin{matrix} IY \\ M'' \end{matrix} \middle| e^{-i\nu I_4} \middle| \begin{matrix} I'Y' \\ M''' \end{matrix} \right\rangle D^{I'}_{M''M'}(\alpha', \beta', \gamma'). \quad (13)$$

One notices that the undetermined matrix

$$\left\langle \begin{matrix} IY \\ M'' \end{matrix} \middle| e^{-i\nu I_4} \middle| \begin{matrix} I'Y' \\ M''' \end{matrix} \right\rangle$$

plays a role analogous to $d^j_{mm'}(\beta)$ of $SU2$. In $SU2$, the $d^j_{mm'}(\beta)$ matrix is determined by applying $e^{-i\beta J_2}$ to a suitable polynomial basis state representation. A search for basis states in $SU3$ led to a paper by Mukunda and Pandit.[9] Their basis states Ψ^{IY}_M are given in terms of tensors $T^{m\cdots m_\lambda}_{n\cdots n_\mu}$ which transform under the three-dimensional defining representation and its complex conjugate.

For low-dimensional representations or the important special case where $I' = M' = 0$, the transformation of interest $e^{-i\nu I_4}$ can be applied directly to these tensors. The results for the eight-dimensional representation are displayed in Table 1.

A general result for all IR's would be very difficult with this method. An easier way presents itself by noting that

$$e^{i\pi I_6/2} I_4 e^{-i\pi I_6/2} = I_2.$$

Therefore we have only to determine matrices of the type

$$\left\langle \begin{matrix} I_\alpha Y_\alpha \\ M_\alpha \end{matrix} \middle| e^{i\pi I_6/2} \middle| \begin{matrix} IY \\ M \end{matrix} \right\rangle$$

as

$$\left\langle \begin{matrix} IY \\ M'' \end{matrix} \middle| e^{-i\nu I_4} \middle| \begin{matrix} I'Y' \\ M''' \end{matrix} \right\rangle = \sum_{\alpha\beta} \left\langle \begin{matrix} IY \\ M'' \end{matrix} \middle| e^{-i\pi I_6/2} \middle| \begin{matrix} I_\beta Y_\beta \\ M_\beta \end{matrix} \right\rangle \delta_{I_\beta I_\alpha} \times d^{I_\beta}_{M_\beta M_\alpha}(2\nu)\delta_{Y_\beta Y_\alpha} \left\langle \begin{matrix} I_\alpha Y_\alpha \\ M_\alpha \end{matrix} \middle| e^{i\pi I_6/2} \middle| \begin{matrix} I'Y' \\ M''' \end{matrix} \right\rangle. \quad (14)$$

Using the relations connecting the states Ψ^{IY}_M and tensors, we finally obtain

$$\left\langle \begin{matrix} I_\alpha Y_\alpha \\ M_\alpha \end{matrix} \middle| e^{i\pi I_6/2} \middle| \begin{matrix} IY \\ M \end{matrix} \right\rangle = \sum_{m_1 m_2} (i)^{\lambda - j_1^{\min} - m_1} (-i)^{\mu - j_2^{\min} + m_2} C(j_1^{\min} j_2^{\min} I; m_1 m_2 M) \times C(\tfrac{1}{2}(\lambda - j_1^{\min} + m_1), \tfrac{1}{2}(\mu - j_2^{\min} - m_2), I_\alpha; \tfrac{1}{2}(3j_1^{\min} + m_1 - \lambda), \tfrac{1}{2}(-3j_2^{\min} + m_2 + \mu), M_\alpha) \times N_2[\tfrac{1}{2}(\lambda - j_1^{\min} + m_1), \tfrac{1}{2}(\mu - j_2^{\min} - m_2), I_\alpha] N_3(I_\alpha Y_\alpha)^{-1} N_3(IY), \quad (15)$$

where

$$N_2(j_1 j_2 I) = \left[\frac{(2I+1)!\,(\lambda - j_1 + j_2 - I)!\,(\mu - j_2 + j_1 - I)!}{(j_1 + j_2 - I)!\,(j_1 + j_2 + I + 1)!\,(\lambda - 2j_1)!\,(\mu - 2j_2)!} \right]^{\frac{1}{2}},$$

$$N_3(IY) = \left\{ \frac{(2I+1)!\,(\lambda + \mu + 1)!}{[\tfrac{1}{3}(\lambda + 2\mu) + I + \tfrac{1}{2}Y + 1]!\,[\tfrac{1}{3}(2\lambda + \mu) + I - \tfrac{1}{2}Y + 1]!} \right\}^{\frac{1}{2}},$$

$$j_1^{\min} = \tfrac{1}{2}I + \tfrac{1}{4}Y + \tfrac{1}{6}(\lambda - \mu), \quad j_2^{\min} = \tfrac{1}{2}I - \tfrac{1}{4}Y - \tfrac{1}{6}(\lambda - \mu).$$

[9] N. Mukunda and L. K. Pandit, J. Math. Phys. 6, 746 (1965).

TABLE I. $\left\langle \begin{matrix} IY \\ M \end{matrix} \middle| e^{-i\nu I_4} \middle| \begin{matrix} I'Y' \\ M' \end{matrix} \right\rangle$ matrix for $SU3$.

IY \ $I'Y'$ / M \ M'	$\frac{1}{2}$ 1 / $\frac{1}{2}$	$\frac{1}{2}$ 1 / $-\frac{1}{2}$	1 1 / 1	0 1 / 0	1 0 / -1	0 0 / 0	$\frac{1}{2}$ -1 / $\frac{1}{2}$	$\frac{1}{2}$ -1 / $-\frac{1}{2}$
$\frac{1}{2}$ 1 / $\frac{1}{2}$	$\cos^2\nu$			$-(i\sqrt{2})\sin\nu \times \cos\nu$		$-i(\sqrt{\frac{3}{2}})\sin\nu \times \cos\nu$		$-\sin^2\nu$
$\frac{1}{2}$ 1 / $-\frac{1}{2}$		$\cos\nu$			$-\sin\nu$			
1 0 / 1			$\cos\nu$				$-i\sin\nu$	
1 0 / 0	$-(i/\sqrt{2})\sin \times \cos\nu$			$\frac{1}{2}(1+\cos^2\nu)$		$-(\frac{1}{2}\sqrt{3})\sin^2\nu$		$-(i/\sqrt{2})\sin\nu \times \cos\nu$
1 0 / -1		$-i\sin\nu$			$\cos\nu$			
0 0 / 0	$-i(\sqrt{\frac{3}{2}})\sin\nu \times \cos\nu$			$-(\frac{1}{2}\sqrt{3})\sin^2\nu$		$-\frac{1}{2}\sin^2\nu+\cos^2\nu$		$-i(\sqrt{\frac{3}{2}})\sin\nu \times \cos\nu$
$\frac{1}{2}$ -1 / $\frac{1}{2}$			$-i\sin\nu$				$\cos\nu$	
$\frac{1}{2}$ -1 / $-\frac{1}{2}$	$-\sin^2\nu$			$(-i/\sqrt{2})\sin\nu \times \cos\nu$		$-i(\sqrt{\frac{3}{2}})\sin\nu \times \cos\nu$		$\cos^2\nu$

IV. SPHERICAL HARMONIC BASIS STATES

For the group $SU2$, the spherical harmonic basis states were obtained by specializing to the right-hand state $m' = 0$ in $D^j_{mm'}(\alpha, \beta, \gamma)$. In this way the dependence on γ was eliminated. Comparing this to $SU3$ we define

$$\Psi^{IY}_M \equiv \left\langle \begin{matrix} IY \\ M \end{matrix} \middle| T_2T_3T_2' \middle| I = 0 \right\rangle = \left\langle \begin{matrix} IY \\ M \end{matrix} \middle| T_2T_3 \middle| I = 0 \right\rangle \quad \text{for } SU3. \quad (16)$$

The choice of $I = 0$ for the right-hand state eliminates three variables.

We now derive the explicit form of this matrix. From a weight diagram we see that

$$\Psi^{IY}_M = e^{i\rho Y} D^I_{MM'}(\alpha, \beta, \gamma) \left\langle \begin{matrix} IY \\ M' \end{matrix} \middle| e^{-i\nu I_4} \middle| \begin{matrix} 0Y_0 \\ 0 \end{matrix} \right\rangle. \quad (17)$$

We can evaluate the last factor by applying this transformation $e^{-i\nu I_4}$ to the tensor basis states.

The basis states are

$$\Psi^{IY}_M = \frac{-(-i)^{2I}(2I+1)^{\frac{1}{2}}}{(\lambda+1)^{\frac{1}{2}}(\mu+1)^{\frac{1}{2}}} \csc\nu D^I_{M,\frac{1}{2}Y+\frac{1}{3}(\lambda-\mu)}(\alpha, \beta, \gamma) \times d^{\frac{1}{2}(\lambda+\mu+1)}_{\frac{1}{2}(\mu-\lambda+3Y+6I+3),\frac{1}{2}(\mu-\lambda+3Y-6I-3)}(2\nu) e^{i\rho Y}. \quad (18)$$

Nelson[2] derived these functions by constructing differential operators representing the infinitesimal generators. With the following substitutions these results agree to within a phase factor:

$$\lambda \to m, \quad \mu \to n, \quad Y \to u, \quad \rho \to \beta, \quad \beta \to \alpha_2, \quad \gamma \to \gamma, \quad \alpha \to \alpha_3.$$

V. VOLUME ELEMENT

Murnaghan[1] derives the following relation for the volume element:

$$U^{-1}_{(\alpha)} \frac{\partial}{\partial\alpha_i} U(\alpha) = \sum_j C^{(\alpha)}_{ij} I_j, \quad \rho(\alpha) \equiv |C^{(\alpha)}_{ij}|. \quad (19)$$

The determinant is unaffected by a unitary similarity transformation. Therefore we may as well evaluate

$$T_3T_2'\left[U^{-1}\frac{\partial}{\partial\alpha_i}U\right]T_2'^{-1}T_3^{-1} = T_2^{-1}\frac{\partial}{\partial\alpha_i}T_2, \quad \text{for } \alpha, \beta, \gamma, \quad (20)$$

$$= \left[\frac{\partial}{\partial\alpha_i}T_3T_2'\right]T_2'^{-1}T_3^{-1}, \quad \text{for other variables.} \quad (21)$$

With a little manipulation, Eq. (20) decomposes the 8×8 determinant into a 3×3 (equal to the $SU2$ volume element $\sin\beta$) and a 5×5 determinant independent of α, β, γ. By the same argument, the contribution from the other $SU2$ factor is $\sin\beta'$. The dependence on the other factor is easily found by evaluation of the 5×5 determinant with $\alpha = \gamma = \alpha' = \gamma' = 0$, $\beta = \beta' = \pi/2$.

The volume element is arbitrary in regard to the coefficient. This constant is usually chosen such that $\int_R \rho(\alpha)\, d(\alpha) = 1$. We also wish the volume element corresponding to the $SU2$ parts of the parameterization to have the correct coefficient for $SU2$. We therefore have

$$\rho(\alpha)\, d(\alpha) = \sin\beta \sin\beta' \sin(2\nu) \sin^2\nu \times \frac{d\rho}{3\pi}\frac{d\alpha}{4\pi}\frac{d\beta}{2}\frac{d\gamma}{4\pi}(d2\nu)\frac{d\alpha'}{4\pi}\frac{d\beta'}{2}\frac{d\gamma'}{4\pi}. \quad (22)$$

VI. ORTHOGONALITY RELATIONS

The orthogonality relations are derived by using a procedure similar to the one outlined in Gottfried[10] for $SU2$. A problem in phases arises in connection with the D^* appearing in the orthogonality relation. We leave it to the reader to add the appropriate phase factors where needed (see Ref. 11, pp. 46–47).

Using the Clebsch–Gordan (CG) coefficients to reduce the direct product, we have

$$D^{\mu_1^*}_{\nu_1'\nu_1}(\alpha) D^{\mu_2}_{\nu_2'\nu_2}(\alpha) = \sum_{\mu_\gamma \nu'\nu} \begin{pmatrix} \mu_1^* & \mu_2 & \mu_\gamma \\ -\nu_1 & \nu_2 & \nu \end{pmatrix} \begin{pmatrix} \mu_1^* & \mu_2 & \mu_\gamma \\ -\nu_1' & \nu_2' & \nu' \end{pmatrix} D^{\mu_\gamma}_{\nu'\nu}(\alpha), \quad -\nu_1 \equiv (I_1, -Y_1, -M_1). \quad (23)$$

We now show that

$$\int_R D^{\mu}_{\nu'\nu}(\alpha)\rho(\alpha)\, d(\alpha) = \delta_{\mu 1}\delta_{\nu' 0}\delta_{\nu 0},$$

where $\mu \equiv 1$ for the one-dimensional representation.

From Eqs. (13) and (18), and using the result from $SU2$:

$$\int_R D^{I}_{M_1 M_2}(\alpha, \beta, \gamma)\, dR(\alpha, \beta, \gamma) = \delta_{I0}\delta_{M_1 0}\delta_{M_2 0}. \quad (24)$$

We have also

$$\int_R D^{\mu}_{\nu'\nu}(\alpha)\rho(\alpha)\, d(\alpha) = \delta_{\lambda_1\mu}\delta_{Y0}\delta_{I0}\delta_{M0}\delta_{I'0}\int_0^{\pi/2} \left\langle \begin{matrix} 0 & 0 \\ & 0 \end{matrix} \middle| e^{-i\nu I_4} \middle| \begin{matrix} 0 & 0 \\ & 0 \end{matrix} \right\rangle \times \sin(2\nu)\sin^2\nu\, d(2\nu) = \delta_{\mu 1}\delta_{\nu' 0}\delta_{\nu 0}. \quad (25)$$

[10] Kurt Gottfried, *Quantum Mechanics, Vol. I* (W. A. Benjamin, Inc., New York, 1966).

[11] P. Carruthers, *Introduction to Unitary Symmetry, Interscience Tracts on Physics and Astronomy No. 27* (Interscience Publishers, Inc., New York, 1966).

We use this result in integrating Eq. 23 as follows:

$$\int D^{\mu_1^*}_{\nu_1'\nu_1}(\alpha) D^{\mu_2}_{\nu_2'\nu_2}(\alpha)\rho(\alpha)\, d(\alpha) = \begin{pmatrix} \mu_1^* & \mu_2 & 1 \\ -\nu_1 & \nu_2 & 0 \end{pmatrix} \begin{pmatrix} \mu_1^* & \mu_2 & 1 \\ -\nu_1' & \nu_2' & 0 \end{pmatrix}. \quad (26)$$

Using results derived by de Swart[12] for the $SU3$ CG coefficients, we have the desired result:

$$\int_R D^{\mu_1^*}_{\nu_1'\nu_1}(\alpha) D^{\mu_2}_{\nu_2'\nu_2}(\alpha)\rho(\alpha)\, d(\alpha) = \frac{\delta_{\mu_1\mu_2}\delta_{\nu_1\nu_2}\delta_{\nu_1'\nu_2'}}{d_1}. \quad (27)$$

VII. TRIALITY

We see from Eq. (13) that

$$D^{\mu}_{\nu_1\nu_2}(\rho + \pi, \alpha + 2\pi, \alpha_i) = \exp[-i\pi(\bar{Y} + 2\bar{M})] D^{\mu}_{\nu_1\nu_2}(\rho, \alpha, \alpha_i). \quad (28)$$

$\exp[-i\pi(\bar{Y} + 2\bar{M})]$ is a scalar matrix (it commutes with all matrices in the defining representation) and therefore $\exp[-i\pi(\bar{Y} + 2\bar{M})] = \exp(2\pi i t/3) I$, where t is an integer from the set $\{0, 1, 2\}$.

This provides a convenient way to define t, the triality of a representation. It reduces to the standard definition, since, when $I = 0$, $M = 0$, then $Y_0 = \frac{2}{3}(\mu - \lambda)$ and therefore $t = \lambda - \mu \bmod 3$. The representations can be classified according to the value of t. In the decomposition of direct products we find that

$$D^{\mu}_{\nu'\nu} = \sum_{\nu_1'\nu_1\nu_2'\nu_2} \begin{pmatrix} \mu_1 & \mu_2 & \mu \\ \nu_1' & \nu_2' & \nu' \end{pmatrix} D^{\mu_1}_{\nu_1'\nu_1} D^{\mu_2}_{\nu_2'\nu_2} \begin{pmatrix} \mu_1 & \mu_2 & \mu \\ \nu_1 & \nu_2 & \nu \end{pmatrix}$$

implies that

$$\exp(i\tfrac{2}{3}\pi t) = \exp(i\tfrac{2}{3}\pi t_1)\exp(i\tfrac{2}{3}\pi t_2). \quad (29)$$

Therefore, for the triality we find $t = t_1 + t_2 \bmod 3$.

ACKNOWLEDGMENTS

I wish to thank Professor H. Mahmoud for his stimulating discussions and Professor J. Lomont for mathematical guidance.

[12] J. J. de Swart, Rev. Mod. Phys. 35, 916 (1963).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Korteweg–de Vries Equation and Generalizations. III.* Derivation of the Korteweg–de Vries Equation and Burgers Equation

C. H. Su† AND C. S. GARDNER‡
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey

(Received 22 April 1968)

The Korteweg–de Vries equation and the Burgers equation are derived for a wide class of nonlinear Galilean-invariant systems under the weak-nonlinearity and long-wavelength approximations. The former equation is shown to be a limiting form for nonlinear dispersive systems while the latter is a limiting form for nonlinear dissipative systems.

For a wide class of nonlinear Galilean-invariant systems, if the nonlinearity is weak and if one makes the long-wavelength approximation, the governing equations can be reduced to either the Korteweg–de Vries equation

$$n_\tau + nn_\xi + \delta n_{\xi\xi\xi} = 0 \tag{1}$$

or the Burgers equation

$$n_\tau + nn_\xi - \nu n_{\xi\xi} = 0, \tag{2}$$

depending on whether the system is dispersion or dissipation dominated. Here we propose to derive these equations for a class of nonlinear systems characterized by the state variables (n, u, f) which are governed by

$$n_t + (nu)_x = 0, \tag{3}$$

$$(nu)_t + (nu^2 + P)_x = 0, \tag{4}$$

$$P = P(f, n, u, f_i, n_i, u_i, f_{ij}, n_{ij}, u_{ij}, \cdots), \tag{5}$$

$$F(f, n, u, f_i, n_i, u_i, f_{ij}, n_{ij}, u_{ij}, \cdots) = 0, \tag{6}$$

where the subscripts i and j denote differentiation with respect to the space and time variables x and t. Equation (3) is the familiar law of conservation of particles if we interpret n and u as the number density and particle velocity, respectively. Equation (4) is then the law of momentum conservation with (5) defining the generalized stress force P as a function of the state variables (n, u, f) and their derivatives. The state variable f here serves as a parametric function in (5) and (6), which defines P as a functional of n, u, and all their derivatives. It is known that (1) and (2) describe small (nonlinear) perturbations from a uniform equilibrium state. We assume that both P and F can be expanded as Taylor series around such a uniform state.

To give some idea of possible forms for P and F we list several examples of physical interest:

(1) *Gas dynamics.* Here f stands for the thermodynamic pressure. For P and F we have

$$P = m^{-1}[p - \mu(\partial u/\partial x)], \quad F \equiv p - An^\gamma = 0, \tag{7}$$

where μ = viscosity coefficient, γ = ratio of specific heats, and A is a constant of proportionality.

(2) *Waves in shallow water.*[1] The number density n in our equation now stands for h, the elevation of the water surface above the bottom of a channel. In this case the state is defined by only two functions h, u. Thus (6) is not needed and we have

$$P = \tfrac{1}{2}gh^2 - \tfrac{1}{3}h^3(u_{xt} + uu_{xx} - u_x^2). \tag{8}$$

(3) *Hydromagnetic waves in cold plasma.*[2] Here f stands for the magnetic field $B(x, t)$ and we have

$$P = \tfrac{1}{2}B^2, \quad F \equiv B - n - (B_x/n)_x = 0. \tag{9}$$

(4) *Ion-acoustic waves in cold plasma.*[3] Here f stands for the electrostatic potential $\psi(x, t)$ and we have

$$P = e^\psi - \tfrac{1}{2}\psi_x^2, \quad F \equiv n - e^\psi + \psi_{xx} = 0. \tag{10}$$

At equilibrium, all the derivatives in P and F drop out and we leave out the dependence of P and F on u to preserve Galilean invariance of the system, i.e.,

$$P = P(f, n) \quad \text{and} \quad F(f, n) = 0. \tag{11}$$

Equation (4) can then be written as follows

$$u_t + uu_x + (a^2/n)n_x = 0, \tag{12}$$

* Papers I and II were published in J. Math. Phys. **9**, 1202, 1204 (1968).

† Present address: Division of Applied Mathematics, Brown University, Providence, Rhode Island.

‡ Present address: Department of Applied Mathematics, University of Texas, Austin, Texas.

[1] See the appendix for a derivation of the correction to the usual shallow-water equations.

[2] C. S. Gardner and G. K. Morikawa, Courant Institute of Mathematical Sciences, New York University Report NYU-9082, 1960 (unpublished).

[3] H. Washimi and T. Taniuti, Phys. Rev. Letters **17**, 996 (1966).

where $a^2 \equiv [P_n - (F_n/F_f)P_f]$ with the subscripts denoting partial differentiations. If $a^2 > 0$, then (3) and (12) define a hyperbolic system of equations with the two characteristic directions given by

$$dx/dt = u \pm a. \tag{13}$$

Therefore, a as defined above is the speed of wave propagation. Furthermore, in the limit of infinitesimal perturbations around a uniform state, one obtains the wave equation with a constant speed of propagation:

$$u_{tt} - a_0^2 u_{xx} = 0, \tag{14}$$

were a_0 is the wave speed of the uniform state. Any solution of Eq. (14) consists of two form-preserving waves: one is right-going and the other left-going.

We view the Korteweg–de Vries and Burger equations as designed to describe the slow change of one of these two waves due to both nonlinear and dispersive (or dissipative) effects characterized by the dependence of P and F on the derivatives. We therefore change (3)–(6) to a frame of reference which moves with one of these waves, say to the right at speed a_0 (a similar result can be obtained for the left-going wave). To account for the slow variation of the waveform, we introduce, after Gardner and Morikawa,[2] a scale transformation of the independent variables; i.e.,

$$\xi = \epsilon^{\alpha}(x - a_0 t), \tag{15}$$

$$\tau = \epsilon^{\alpha+1} t, \tag{16}$$

where ϵ denotes the amplitude of the initial disturbance and is assumed to be small compared with unity. The exponent $\alpha > 0$ is a number to be determined such that the time variation of a state variable (in the waveframe) is balanced by both nonlinear and dispersive (or dissipative) effects. Using Eqs. (15) and (16), in (3) and (4), we obtain

$$\epsilon n_\tau + (u - a_0)n_\xi + n u_\xi = 0, \tag{17}$$

$$\epsilon u_\tau + (u - a_0)u_\xi + n^{-1}P_\xi = 0. \tag{18}$$

We now assume that the state variables n, f, and u can be represented asymptotically as series in powers of ϵ about an equilibrium state $\mathbf{A} = (n, f, u) = (n_0, f_0, 0)$; i.e.,

$$n = n_0 + \epsilon n^{(1)} + \epsilon^2 n^{(2)} + \cdots, \tag{19}$$

$$f = f_0 + \epsilon f^{(1)} + \epsilon^2 f^{(2)} + \cdots, \tag{20}$$

$$u = 0 + \epsilon u^{(1)} + \epsilon^2 u^{(2)} + \cdots. \tag{21}$$

These expansions and the transformations (15) and (16) are then substituted into the Taylor series of P and F around the equilibrium state $\mathbf{A}_0$. In the first order of approximation, all the derivatives of the state variables with respect to x and t are dropped because of (15) and (16), and we have

$$P = P_0 + P_{f0}(f - f_0) + P_{n0}(n - n_0) + P_{u0}(u - u_0) + O(\epsilon^2), \tag{22}$$

$$F = F_0 + F_{f0}(f - f_0) + F_{n0}(n - n_0) + F_{u0}(u - u_0) + O(\epsilon^2). \tag{23}$$

To preserve the Galilean invariance of the system, the u dependence of P and F must be in the form

$$H_t + uH_x, \tag{24}$$

where H is a function of the state variables and their derivatives. The dependence of H on u must be again of that form. Thus in (22) and (23), P_{u0} and F_{u0}, which are evaluated at the equilibrium, vanish identically. From (22) and (23) we obtain

$$\frac{\partial P^{(1)}}{\partial \xi} = \left(P_{n0} - \frac{F_{f0}}{F_{n0}} P_{f0}\right)\frac{\partial n^{(1)}}{\partial \xi} \equiv a_0^2 \frac{\partial n^{(1)}}{\partial \xi}. \tag{25}$$

The leading approximation to (17) and (18), then, is

$$a_0 n_\xi^{(1)} = n_0 u_\xi^{(1)},$$

$$a_0 u_\xi^{(1)} = (a_0^2/n_0) n_\xi^{(1)}.$$

Integrating these and noting the boundary condition for $n^{(1)}$ and $u^{(1)}$ at $\xi \to \pm\infty$ we have

$$a_0 n^{(1)} = n_0 u^{(1)}. \tag{26}$$

This permits reducing the leading-order problem to one variable, say $n^{(1)}$. Our objective now is to derive an evolution equation for $n^{(1)}$. Corrections to (22) and (23) can now be obtained from (5) and (6) with the help of (26). Within the order of our approximation, we obtain

$$P_\xi^{(2)} \approx a_0^2 n_\xi^{(2)} + A n^{(1)} n_\xi^{(1)} + \epsilon^{\alpha-1} B n_{\xi\xi}^{(1)} + \epsilon^{2\alpha-1} C n_{\xi\xi\xi}^{(1)}, \tag{27}$$

where A, B, and C (like a_0) are constants depending on the partial derivatives of P and F evaluated at the equilibrium. Here we have kept the term $\epsilon^{2\alpha-1}\eta_{\xi\xi\xi}^{(1)}$ although it is small in comparison with $\epsilon^{\alpha-1}\eta_{\xi\xi}^{(1)}$ ($\alpha \geq 0$) for cases where $B \neq 0$. For the examples we listed above, the constants a_0^2, A, B, and C are given in the following table:

	a_0^2	A	B	C
Gas dynamics	$\gamma KT/m$	0	$-va_0$	0
Water waves	gh_0	0	0	$\frac{1}{3}gh_0^3$
Hydromagnetic	B_0	1	0	1
Ion-acoustic	1	0	0	1

In the next order approximation of (17) and (18), we have

$$n_\tau^{(1)} + 2\frac{a_0}{n_0} n^{(1)} n_\xi^{(1)} - a_0 n_\xi^{(2)} + n_0 u_\xi^{(2)} = 0, \quad (28)$$

$$\frac{a_0}{n_0} n_\tau^{(1)} + \frac{A}{n_0} n^{(1)} n_\xi^{(1)} + \epsilon^{\alpha-1}\frac{B}{n_0} n_{\xi\xi}^{(1)} + \epsilon^{2\alpha-1}\frac{C}{n_0} n_{\xi\xi\xi}^{(1)} - a_0 u_\xi^{(2)} + \frac{a_0^2}{n_0} n_\xi^{(2)} = 0. \quad (29)$$

We now eliminate $n_\xi^{(2)}$ and $u_\xi^{(2)}$ between (28) and (29) and obtain the evolution equation for $n^{(1)}$, i.e.,

$$n_\tau^{(1)} + \left(\frac{A}{2a_0} + \frac{a_0}{n_0}\right) n^{(1)} n_\xi^{(1)} + \epsilon^{\alpha-1}\frac{B}{2a_0} n_{\xi\xi}^{(1)} + \epsilon^{2\alpha-1}\frac{C}{2a_0} n_{\xi\xi\xi}^{(1)} = 0. \quad (30)$$

If $B \neq 0$ (for a dissipative system $B < 0$), we set $\alpha = 1$ and neglect the last term in (30). The resulting equation is the Burgers equation (2) (except for some constant coefficients which can be scaled out). On the other hand if $B = 0$, i.e., if the system is nondissipative, we set $\alpha = \frac{1}{2}$ and obtain the Korteweg–de Vries equation (1). Evolution equations for higher-order quantities can be obtained in a similar fashion.

ACKNOWLEDGMENTS

One of the authors (C. H. Su) wishes to acknowledge helpful discussions with Dr. M. D. Kruskal. He also wishes to thank Dr. R. M. Miura for his critical reading of the manuscript and Dr. Y. P. Pao for discussion of the work in the Appendix.

This work has been supported by the U.S. Air Force under Contract AF 49(638)-1555.

APPENDIX: DERIVATION OF THE CORRECTION EQUATION TO THE SHALLOW-WATER THEORY

We start with the two-dimensional incompressible inviscid hydrodynamic equations

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -p_x, \quad (A1)$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -p_v - g, \quad (A2)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \quad (A3)$$

where we have taken the fluid density to be 1. The vertical component of velocity v on the bottom of the channel (for brevity, we assume a horizontal bottom) and the pressure on the free surface are both assumed to be zero. We also need the kinematic free-surface condition

$$v_s = \frac{\partial h}{\partial t} + u_s\frac{\partial h}{\partial x}, \quad (A4)$$

where $h(x, t)$ is the height of the water surface above the bottom and the subscript s denotes quantities evaluated at the free surface.

We want to find a set of equations which governs the evolution of $h(x, t)$ and the average horizontal velocity $\bar{u}$ defined by

$$\bar{u}(x, t) = \frac{1}{h(x, t)}\int_0^{h(x,t)} u(x, y, t)\, dy, \quad (A5)$$

where $y = 0$ is taken to be the bottom of the channel. Integrating (A3) with respect to y, one obtains

$$v = -\int_0^y \frac{\partial u}{\partial x}\, dy. \quad (A6)$$

Combining (A6) and (A4), we have

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(h\bar{u}) = 0. \quad (A7)$$

This gives the evolution of $h(x, t)$, provided we know the evolution of $h\bar{u}$. We integrate (A1) from $y = 0$ to $y = h(x, t)$ and obtain

$$\frac{\partial}{\partial t}(h\bar{u}) + \frac{\partial}{\partial x}[h(\overline{u^2} + \bar{p})] = 0, \quad (A8)$$

where

$$h\overline{u^2} = \int_0^h u^2\, dy, \quad h\bar{p} = \int_0^h p\, dy. \quad (A9)$$

The average pressure in (A9) is obtained by integrating (A2), i.e.,

$$p(x, y, t) = -\int_h^y dy\left(g + \frac{dv}{dt}\right) + p_s(x, t) = \int_y^h dy\left(g + \frac{dv}{dt}\right),$$

since $p_s(x, t) = 0$, where

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}.$$

Then,

$$h\bar{p} = \int_0^h p(x, y, t)\, dy = \frac{gh^2}{2} + \int_0^h dy\, y\frac{dv}{dt}. \quad (A10)$$

Equation (A8) can now be written as

$$\frac{\partial}{\partial t}(h\bar{u}) + \frac{\partial}{\partial x}\left[h\bar{u}^2 + \frac{gh^2}{2}\right] + \frac{\partial}{\partial x}\left[\int_0^h dy\, y\frac{dv}{dt} + h(\overline{u^2} - \bar{u}^2)\right] = 0. \quad (A11)$$

Neglecting the last term in (A11), the remaining equation and (A7) consist of the familiar shallow-water equations. We now want to include the effect of the last term in (A11) under the long-wavelength approximation. Using (A6) and the condition of irrotationality

$$\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x},$$

it is readily shown that the leading approximations to the last term in (A11) is

$$\int_0^h dy\, y \frac{dv}{dt} = -\frac{h^3}{3}(\bar{u}_{xt} + \bar{u}\bar{u}_{xx} - \bar{u}_x^2), \quad \text{(A12)}$$

and

$$\overline{u^2} - \bar{u}^2 = \frac{h^5}{45}\bar{u}_{xx}^2, \quad \text{(A13)}$$

where the subscripts represent the derivative of the average velocity. The correction as given by (A13) is of higher order compared to (A12). Using (A12) we obtain the P given by Eq. (8) and

$$\frac{\partial}{\partial t}(h\bar{u}) + \frac{\partial}{\partial x}\left[h\bar{u}^2 + \frac{gh^2}{2} - \frac{h^3}{3}(\bar{u}_{xt} + \bar{u}\bar{u}_{xx} - \bar{u}_x^2)\right] = 0. \quad \text{(A14)}$$

It is of interest to point out that the improved equations derived above contain the stationary waves in the form of the well-known solitary waves and cnoidal waves. By dropping the time derivatives in (A7) and (A14), one can readily integrate these equations to give the solitary wave solution

$$h(x) = 1 + (j^2 - 1)\,\mathrm{sech}^2\{\tfrac{3}{4}(1 - j^{-2})^{\frac{1}{2}}x\},$$

where we have set $g = 1$, and $j = h\bar{u}$ is the flux per unit width of the channel with $\bar{u}$ normalized by the wave speed $(gh_0)^{\frac{1}{2}}$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Second-Order Dispersion-Energy Series for Axially Symmetric Molecules

JOHANNES H. VAN DER MERWE
Merensky Institute for Physics, University of Stellenbosch, South Africa
AND
ALWYN J. VAN DER MERWE*
Department of Physics, University of Denver, Denver, Colorado

(Received 14 August 1968)

With the aid of a model representing each molecule as a set of semi-isotropic dipole oscillators, the London–van der Waals interaction energy of a pair of unexcited axially symmetric molecules is calculated, in the second order of perturbation theory, as an infinite series in negative powers of the intermolecular separation.

1. INTRODUCTION

In order to estimate the dispersion forces between two atoms, arising from their mutual polarization, it has long been customary to treat the electrostatic interaction between them as a perturbation and to identify the resultant displacement of their energy levels, computed with the aid of either Schrödinger perturbation theory or the variational method, as the potential of the interatomic force.[1]

With regard to the classical aspect of this problem, it has been found advantageous to express the potential energy V of the electrostatic interaction as a sum of inverse powers of the internuclear distance R, whose terms one can easily recognize as furnishing the interaction of the various electric multipole moments of the one atom with those of the other. If the atoms are electrically neutral, the interactions involving monopoles vanish and, thus, the dipole–dipole potential energy, proportional to R^{-3}, becomes the leading term of this series. Margenau[2,3] gave the first few terms of the series in Cartesian coordinates, while Heller[4] provided the higher multiple terms up to the sixteenth. The complete expansion of V was furnished independently by Carlson and Rushbrooke,[5] using spherical coordinates, and by Rose,[6] in irreducible tensor form. Both expansions presuppose

* Work of this author was supported by the National Aeronautics and Space Administration under Grant NsG-518.

[1] F. London, Z. Physik **63**, 245 (1930).

[2] H. Margenau, Phys. Rev. **38**, 747 (1931).

[3] H. Margenau, Rev. Mod. Phys. **11**, 1 (1939).

[4] R. Heller, J. Chem. Phys. **9**, 156 (1941).

[5] B. C. Carlson and G. S. Rushbrooke, Proc. Cambridge Phil. Soc. **46**, 626 (1950).

[6] M. E. Rose, J. Math. & Phys. **37**, 215 (1958).

Neglecting the last term in (A11), the remaining equation and (A7) consist of the familiar shallow-water equations. We now want to include the effect of the last term in (A11) under the long-wavelength approximation. Using (A6) and the condition of irrotationality

$$\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x},$$

it is readily shown that the leading approximations to the last term in (A11) is

$$\int_0^h dy\, y \frac{dv}{dt} = -\frac{h^3}{3}(\bar{u}_{xt} + \bar{u}\bar{u}_{xx} - \bar{u}_x^2), \quad \text{(A12)}$$

and

$$\overline{u^2} - \bar{u}^2 = \frac{h^5}{45}\bar{u}_{xx}^2, \quad \text{(A13)}$$

where the subscripts represent the derivative of the average velocity. The correction as given by (A13) is of higher order compared to (A12). Using (A12) we obtain the P given by Eq. (8) and

$$\frac{\partial}{\partial t}(h\bar{u}) + \frac{\partial}{\partial x}\left[h\bar{u}^2 + \frac{gh^2}{2} - \frac{h^3}{3}(\bar{u}_{xt} + \bar{u}\bar{u}_{xx} - \bar{u}_x^2)\right] = 0. \quad \text{(A14)}$$

It is of interest to point out that the improved equations derived above contain the stationary waves in the form of the well-known solitary waves and cnoidal waves. By dropping the time derivatives in (A7) and (A14), one can readily integrate these equations to give the solitary wave solution

$$h(x) = 1 + (j^2 - 1)\,\mathrm{sech}^2\{\tfrac{3}{4}(1 - j^{-2})^{\frac{1}{2}}x\},$$

where we have set $g = 1$, and $j = h\bar{u}$ is the flux per unit width of the channel with $\bar{u}$ normalized by the wave speed $(gh_0)^{\frac{1}{2}}$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 3 MARCH 1969

Second-Order Dispersion-Energy Series for Axially Symmetric Molecules

JOHANNES H. VAN DER MERWE
Merensky Institute for Physics, University of Stellenbosch, South Africa
AND
ALWYN J. VAN DER MERWE*
Department of Physics, University of Denver, Denver, Colorado

(Received 14 August 1968)

With the aid of a model representing each molecule as a set of semi-isotropic dipole oscillators, the London–van der Waals interaction energy of a pair of unexcited axially symmetric molecules is calculated, in the second order of perturbation theory, as an infinite series in negative powers of the intermolecular separation.

1. INTRODUCTION

In order to estimate the dispersion forces between two atoms, arising from their mutual polarization, it has long been customary to treat the electrostatic interaction between them as a perturbation and to identify the resultant displacement of their energy levels, computed with the aid of either Schrödinger perturbation theory or the variational method, as the potential of the interatomic force.[1]

With regard to the classical aspect of this problem, it has been found advantageous to express the potential energy V of the electrostatic interaction as a sum of inverse powers of the internuclear distance R, whose terms one can easily recognize as furnishing the interaction of the various electric multipole moments of the one atom with those of the other. If the atoms are electrically neutral, the interactions involving monopoles vanish and, thus, the dipole–dipole potential energy, proportional to R^{-3}, becomes the leading term of this series. Margenau[2,3] gave the first few terms of the series in Cartesian coordinates, while Heller[4] provided the higher multiple terms up to the sixteenth. The complete expansion of V was furnished independently by Carlson and Rushbrooke,[5] using spherical coordinates, and by Rose,[6] in irreducible tensor form. Both expansions presuppose

* Work of this author was supported by the National Aeronautics and Space Administration under Grant NsG-518.

[1] F. London, Z. Physik **63**, 245 (1930).

[2] H. Margenau, Phys. Rev. **38**, 747 (1931).

[3] H. Margenau, Rev. Mod. Phys. **11**, 1 (1939).

[4] R. Heller, J. Chem. Phys. **9**, 156 (1941).

[5] B. C. Carlson and G. S. Rushbrooke, Proc. Cambridge Phil. Soc. **46**, 626 (1950).

[6] M. E. Rose, J. Math. & Phys. **37**, 215 (1958).

charge distributions that should not overlap (to ensure the convergence of the infinite series) but are otherwise arbitrary. They are thus valid for molecules as well as for atoms. However, neither of the results has a form suitable for our purposes; the first task of our paper is to disclose the form that is appropriate.

The quantum-mechanical aspect of the dispersion-force problem is vastly more complicated than the classical, due mainly to the fact that the unperturbed wavefunctions required in our perturbation calculations are unknown except in a few simple cases. One is forced, in order to discover results of some generality, to operate with models. In this connection, it has been noted often, starting with London,[7] that the Drude model[8] offers an especially simple way of deriving the dipole–dipole contribution to the dispersion energy in the second order $E^{(2)}$; the first-order contribution vanishes because of the spherical symmetry of the wavefunctions employed. This model represents each molecule as an assembly of three-dimensional isotropic harmonic dipole oscillators of a definite frequency—whose value is suggested by the optical dispersion curve of the chemical species. With its aid, London found the simple and familiar result

$$E^{(2)} = -\tfrac{3}{4}h\nu\alpha^2R^{-6},$$

ν frequency, α molecular polarizability,

for two similar molecules in their ground states.[7] Margenau,[9] carrying the use of the isotropic oscillator model further, applied it also to the dipole–quadrupole and quadrupole–quadrupole terms of V to get further terms of $E^{(2)}$, proportional to R^{-8} and R^{-10}, respectively. Heller[4] did the same for unlike interacting molecules up to terms of the order R^{-12}. Finally, the complete energy series

$$E^{(2)} = \sum_{n=2}^{\infty} C_n R^{-2n-2}$$

was established by Brooks[10] (using the Carlson–Rushbrooke expansion of V) for like molecules and by Fontana[11] (using the Rose expansion) for unlike molecules. These series, which should be applied only to atoms and spherical molecules, were demonstrated to provide an asymptotic expansion of the dispersion energy,[10,12] and are rather useful and reliable if suitably truncated.[11,12]

[7] F. London, Z. Physik. Chem. (Leipzig) **B11**, 222 (1930).
[8] See, for example, J. Hirschfelder, C. Curtiss, and R. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), p. 956.
[9] H. Margenau, J. Chem. Phys. **6**, 896 (1938).
[10] F. C. Brooks, Phys. Rev. **86**, 92 (1952).
[11] P. R. Fontana, Phys. Rev. **123**, 1865 (1961).
[12] A. Dalgarno and J. T. Lewis, Proc. Phys. Soc. (London) **A69**, 57 (1956).

Our main object in the present paper is to discover an analog of the Brooks–Fontana energy series applicable to nonspherical molecules possessing axial symmetry. In doing this, we neglect, as did these authors, third- and higher-order perturbation contributions; remarks on the importance of such corrections have appeared elsewhere.[12,13] The model which suggests itself as appropriate to axially symmetric molecules is one in which the forementioned isotropic oscillators are replaced by oscillators having a particular frequency along the molecular axis and a different one in the directions perpendicular thereto. This semi-isotropic oscillator model also originated with, and was employed by, London who quoted the solution of our problem as it concerns the dipole–dipole term of V, which is his well-known R^{-6} formula with its characteristic dependence on the relative orientations of the molecular axes.[14] The steps leading from London's model to his formula were made explicit in a recent paper by one of us.[15] Our present method is essentially a generalization of the one employed there; but, while the physical basis remains unchanged, our mathematical attack will perforce proceed on an entirely different level.

2. EXPANSION OF COULOMB POTENTIAL RELATIVE TO BODY AXES OF MOLECULES

As we are interested in the electrostatic interaction of two neutral charge distributions, we may initially confine ourselves to a study of the Coulomb interaction potential V between two finite electric dipoles, each composed of the charges e, $-e$, and having the moments $e\mathbf{r}'$ and $e\mathbf{r}''$, respectively. If $\mathbf{R}$ denotes the vector stretching from the positive charge (origin) of the first dipole to that of the second dipole, then, obviously,

$$V = e^2(R^{-1} + |\mathbf{R} - \mathbf{r}' + \mathbf{r}''|^{-1} - |\mathbf{R} - \mathbf{r}'|^{-1} - |\mathbf{R} + \mathbf{r}''|^{-1}). \quad (1)$$

From here on we conveniently set $e = 1$ until further notice.

The Taylor expansion of V is easily seen to be

$$V = \sum_{n=2}^{\infty} V_{n-1}, \quad (2)$$

with

$$V_{n-1} = \sum_{r=1}^{n-1} \frac{(-1)^r}{r!\,(n-r)!} [(\mathbf{r}'' \cdot \nabla)^{n-r}(\mathbf{r}' \cdot \nabla)^r] \frac{1}{R}, \quad (3)$$

and it converges if both r' and r'' are less than R.

[13] W. L. Bade, J. Chem. Phys. **27**, 1280 (1957).
[14] F. London, J. Phys. Chem. **46**, 305 (1942).
[15] A. J. van der Merwe, Z. Physik **196**, 212 (1966).

In order to obtain results in irreducible tensor form,[16] we resolve all vectors and vector operators in their spherical components which are defined for any vector $\mathbf{A} = (A_x, A_y, A_z)$ by

$$A_{\pm 1} = \mp 2^{-\frac{1}{2}}(A_x \pm iA_y), \quad A_0 = A_z.$$

In terms of their spherical components, the scalar product of any two vectors or vector operators $\mathbf{A}$ and $\mathbf{B}$ becomes

$$\mathbf{A} \cdot \mathbf{B} = \sum_{\mu} (-1)^{\mu} A_{\mu} B_{-\mu}, \quad \mu = 0, \pm 1,$$

a result we use forthwith to further develop Eq. (3).

First we find the effect of applying an operator $(\mathbf{a} \cdot \nabla)^p$, with $\mathbf{a}$ any constant vector and p an integer, to the irregular solid harmonic

$$\Upsilon_s^m(\mathbf{R}) = R^{-s-1} Y_s^m(\hat{\mathbf{R}}), \tag{4}$$

$\hat{\mathbf{R}}$ being a unit vector in the direction of $\mathbf{R}$. We invoke to this end the formula

$$\partial_\mu \Upsilon_s^m = (-1)^{\mu}[(s+1)(2s+1)]^{\frac{1}{2}} \times \langle s+1 \quad 1 \quad m+\mu \quad -\mu \mid s \quad m \rangle \Upsilon_{s+1}^{m+\mu}, \tag{5}$$

which is a special case of the well-known gradient formula[17]; the ∂_μ signify the spherical components of the operator ∇, and a Clebsch–Gordan coefficient appears on the right. Repeated application of (5) leads, when the tabulated values of the Clebsch–Gordan coefficients are inserted, to

$$\partial_0^k \Upsilon_s^m = (-1)^k \left[\frac{(2s+1)(k+s-m)!\,(k+s+m)!}{(2k+2s+1)(s-m)!\,(s+m)!}\right]^{\frac{1}{2}} \Upsilon_{s+k}^m \tag{6}$$

and

$$\partial_{\pm 1}^k \Upsilon_s^m = (-1)^k \left[\frac{(2s+1)(2k+s \pm m)!}{2^k(2k+2s+1)(s \pm m)!}\right]^{\frac{1}{2}} \Upsilon_{s+k}^{m \pm k}, \tag{7}$$

cf. Ref. 6. From Eqs. (6) and (7) it follows, in a straightforward manner, that

$$(\mathbf{a} \cdot \nabla)^p \Upsilon_s^m(\mathbf{R}) = \sum \left[\frac{(2s+1)(p+s-m-2k)!\,(p+s+m+2k)!}{2^l(2p+2s+1)(s+m)!\,(s-m)!}\right]^{\frac{1}{2}} \frac{(-1)^{p-l} p!\, a_0^{p-l} a_{-1}^{\frac{1}{2}l+k} a_1^{\frac{1}{2}l-k}}{(p-l)!\,(\frac{1}{2}l+k)!\,(\frac{1}{2}l-k)!} \Upsilon_{s+p}^{m+2k}(\mathbf{R}), \tag{8}$$

wherein the summation indices assume the values

$$l = 0, 1, \cdots, p; \quad k = -\tfrac{1}{2}l, -\tfrac{1}{2}l + 1, \cdots, \tfrac{1}{2}l.$$

The correctness of this equation may be verified by induction.

We employ formula (8) in the evaluation of V_{n-1}, Eq. (3), by identifying the vector $\mathbf{a}$ successively with the displacement vectors $\mathbf{r}'$ and $\mathbf{r}''$ and by observing that

$$R^{-1} = (4\pi)^{\frac{1}{2}} \Upsilon_0^0. \tag{9}$$

In so doing, we suppose that $\mathbf{r}'$ and $\mathbf{r}''$ are resolved in spherical components x'_μ and x''_μ, respectively, relative to a "primed" coordinate system S', associated with the first dipole, and a "doubly primed" system S'', associated with the second dipole. These systems are fully defined below, but at present it suffices to imagine that a rotation through the Euler angles α, β, γ carries the axes of S'' over into those of S'. If the polar angles of $\mathbf{R}$ are θ', ϕ' in the system S' and θ'', ϕ'' in S'', then the matrix $\mathfrak{D}(\alpha\beta\gamma)$ representing this rotation is determined by the transformation equation[18]

$$\Upsilon_r^{2k_1}(\theta', \phi') = \sum_m \mathfrak{D}^r_{m\,2k_1}(\alpha\beta\gamma) \Upsilon_r^m(\theta'', \phi''). \tag{10}$$

By virtue of Eqs. (8)–(10), one can now write

$$(\mathbf{r}' \cdot \nabla)^r R^{-1} = \sum \left[\frac{4\pi(r-2k_1)!\,(r+2k_1)!}{2^{l_1}(2r+1)}\right]^{\frac{1}{2}} \frac{(-1)^{r-l_1} r!\, x_0'^{r-l_1} x_{-1}'^{\frac{1}{2}l_1+k_1} x_1'^{\frac{1}{2}l_1-k_1}}{(r-l_1)!\,(\frac{1}{2}l_1+k_1)!\,(\frac{1}{2}l_1-k_1)!} \Upsilon_r^{2k_1}(\theta', \phi') \tag{11}$$

and

$$(\mathbf{r}'' \cdot \nabla)^{n-r} \Upsilon_r^{2k_1}(\theta', \phi') = \sum (-1)^{n-r-l_2} (n-r)! \frac{x_0''^{n-r-l_2} x_{-1}''^{\frac{1}{2}l_2+k_2} x_1''^{\frac{1}{2}l_2-k_2}}{(n-r-l_2)!\,(\frac{1}{2}l_2+k_2)!\,(\frac{1}{2}l_2-k_2)!} \times \left[\frac{(2r+1)(n-m-2k_2)!\,(n+m+2k_2)!}{2^{l_2}(2n+1)(r+m)!\,(r-m)!}\right]^{\frac{1}{2}} \mathfrak{D}^r_{m\,2k_1}(\alpha\beta\gamma) \Upsilon_n^{m+2k_2}(\theta'', \phi''). \tag{12}$$

[16] See, for example, M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957), Chap. 4.
[17] Reference 16, Eq. (6.42).
[18] We follow the notation (except for typographical variations) and conventions of Ref. 16.

Substitution of the last two equations in Eq. (3) transforms it into

$$V_{n-1}=\sum(-1)^{n+r-l_1-l_2}\left(\frac{x_0'^{r-l_1}x_{-1}'^{\frac{1}{2}l_1+k_1}x_1'^{\frac{1}{2}l_1-k_1}}{(r-l_1)!\,(\frac{1}{2}l_1+k_1)!\,(\frac{1}{2}l_1-k_1)!}\right)\left(\frac{x_0''^{n-r-l_2}x_{-1}''^{\frac{1}{2}l_2+k_2}x_1''^{\frac{1}{2}l_2-k_2}}{(n-r-l_2)!\,(\frac{1}{2}l_2+k_2)!\,(\frac{1}{2}l_2-k_2)!}\right)$$
$$\times\left[\frac{4\pi(r-2k_1)!\,(r+2k_1)!\,(n-m-2k_2)!\,(n+m+2k_2)!}{2^{l_1+l_2}(2n+1)(r+m)!\,(r-m)!}\right]^{\frac{1}{2}}\mathfrak{D}^r_{m\,2k_1}(\alpha\beta\gamma)\,\Upsilon^{m+2k_2}_n(\theta'',\phi''),\quad(13)$$

whereby we have expressed the Coulomb potential energy V relative to two arbitrary coordinate frames. For this result to become pertinent to our physical problem, it is necessary to go one step further and relate the coordinate axes to the symmetry directions of axially symmetric molecules.

We achieve this by requiring, in the first place, that the axes of z' and z'', which by earlier implication make the angles θ' and θ'' with $\mathbf{R}$, coincide with the symmetry axes of the first and second molecules, respectively. Moreover, we choose the remaining axes such that the x' axis is coplanar with the z' axis and $\mathbf{R}$ and make an angle $\theta'+\frac{1}{2}\pi$ with the latter, while the y' axis completes the right-handed rectangular set of axes. The axes of x'' and y'' are specified in a like manner; and the $z'x'$ and $z''x''$ planes, intersecting along $\mathbf{R}$, include the angle φ between them.

It will be realized, on some reflection, that the matrix representing the rotation between the systems S'' and S' just introduced is supplied by

$$\mathfrak{D}^r_{mn}(\alpha\beta\gamma)=\sum_{m'}\mathfrak{D}^r_{mm'}(0-\theta''-\varphi)\mathfrak{D}^r_{m'n}(0\;\theta'\;0)$$
$$=\sum_{m'}d^r_{m'm}(\theta'')\,d^r_{m'n}(\theta')e^{im'\varphi},\quad(14)$$

where the matrix elements of d^r are defined by[19]

$$d^r_{mn}(\theta)=[(r+n)!\,(r-n)!\,(r+m)!\,(r-m)!]^{\frac{1}{2}}$$
$$\times\sum_s\frac{(-1)^s(\cos\tfrac{1}{2}\theta)^{2r+n-m-2s}(-\sin\tfrac{1}{2}\theta)^{m-n+2s}}{(r-m-s)!\,(r+n-s)!\,(s+m-n)!\,s!},\quad(15)$$

the sum being extended over all integral values of s for which the factorial arguments are greater than or equal to zero. We observe furthermore that $\phi''=\pi$, by definition of the x'' axis, and that[20]

$$Y^m_n(\theta,\pi)=(-1)^m\left(\frac{2n+1}{4\pi}\right)^{\frac{1}{2}}d^n_{m0}(\theta).\quad(16)$$

Insertion of Eqs. (14) and (16) in Eq. (13) finally leads to

$$V_{n-1}=\frac{1}{R^{n+1}}\sum(-1)^{n+r-l_1+m}\left(\frac{x_0'^{r-l_1}x_{-1}'^{\frac{1}{2}l_1+k_1}x_1'^{\frac{1}{2}l_1-k_1}}{(r-l_1)!\,(\frac{1}{2}l_1+k_1)!\,(\frac{1}{2}l_1-k_1)!}\right)\left(\frac{x_0''^{n-r-l_2}x_{-1}''^{\frac{1}{2}l_2+k_2}x_1''^{\frac{1}{2}l_2-k_2}}{(n-r-l_2)!\,(\frac{1}{2}l_2+k_2)!\,(\frac{1}{2}l_2-k_2)!}\right)$$
$$\times\left[\frac{(r-2k_1)!\,(r+2k_1)!\,(n-m-2k_2)!\,(n+m+2k_2)!}{2^{l_1+l_2}(r-m)!\,(r+m)!}\right]^{\frac{1}{2}}$$
$$\times\left[\sum_{m'}d^r_{m'm}(\theta'')\,d^r_{m'2k_1}(\theta')e^{im'\varphi}\right]d^n_{m+2k_2\,0}(\theta''),\quad(17)$$

wherein the summation indices vary by integral steps between the following limits: $r=1,\cdots,n-1$; $l_1=0,\cdots,r$; $k_1=-\frac{1}{2}l_1,\cdots,\frac{1}{2}l_1$; $l_2=0,\cdots,n-r$; $k_2=-\frac{1}{2}l_2,\cdots,\frac{1}{2}l_2$; $m,m'=-r,\cdots,r$. This formula for the electrostatic interaction, which refers (on substitution of Cartesian for spherical components) to the physically important directions of the molecules, has a form tailored to our needs.

3. QUANTUM-MECHANICAL INTERACTION ENERGY

In the anisotropic oscillator model we treat each molecule as consisting effectively of a certain number f of electrons which independently execute simple harmonic motions about a common fixed point, with one frequency for the axial direction of the molecule and another for all the transverse directions.[15] Accordingly, the unperturbed wavefunction of a molecule is taken to be a product of wavefunctions in which each electron is represented by three one-dimensional oscillator wavefunctions. For simplicity, we deal initially with only one electron.

In view of the form of the interaction Hamiltonian (17), it is clear that our perturbation calculation of the ground-state dispersion energy requires a knowledge of matrix elements of the kind $\langle 0|\,\xi^s\,|m\rangle$, with $s,m=0,1,2,\cdots$, between the lowest and the mth energy eigenstates of a linear harmonic oscillator having the coordinate ξ ($=x$, y, or z). These matrix elements are nonzero only if $s+m$ is even and $s\geq m$; and then they are given, for normalized oscillator functions, by

$$\langle 0|\,\xi^s\,|m\rangle$$
$$=s!\,(s-m\mp 1)!!/(s-m+1)!\,(2^s m!\,\beta^s)^{\frac{1}{2}},\quad(18)$$

[19] Reference 16, Eq. (4.13).
[20] Reference 16, Eq. (4.30).

where $\beta = \mu\omega/\hbar = (\mu c/\hbar^2)^{\frac{1}{2}}$ (μ and ω being the mass and frequency of the oscillator) measures the stiffness c of the oscillator in the ξ direction. More conveniently, one may write

$$\langle 0|\ \xi^{m+2p}\ |m\rangle = (m+2p)!/(2^m m!)^{\frac{1}{2}} 2^{2p} p!\ \beta^{\frac{1}{2}m+p}, \quad (19)$$

with $p = 0, 1, 2, \cdots$.

Employing the notation $|m_1 m_2\rangle = |m_1\rangle\,|m_2\rangle$ for the product of state vectors referring to the x and y directions, respectively, we introduce the quantities

$$S^{lk}_{m_1 m_2} = \langle 00|\ x_{-1}^{\frac{1}{2}l+k} x_1^{\frac{1}{2}l-k}\ |m_1 m_2\rangle/(\tfrac{1}{2}l+k)!\,(\tfrac{1}{2}l-k)! \quad (20)$$

whose relevant properties are

$$S^{lk*}_{m_1 m_2} = (-1)^{2k} S^{l-k}_{m_1 m_2} \quad (21)$$

and $S^{lk}_{m_1 m_2} = 0$ unless $l + m_1 + m_2$ is even (i.e., $m_1 + m_2 \pm 2k$ is even) and $l \geq m_1 + m_2$. In terms of the S's and the matrix elements (18), the second-order perturbation energy corresponding to the interaction (17) could now be written down as a complicated function of $\beta'_x, \beta'_y, \cdots, \beta''_z$, the stiffness parameters for the axes of $x', y', \cdots, z''$. However, for axially symmetric molecules the simplifying assumptions

$$\beta'_x = \beta'_y = \beta'_\perp \quad \text{and} \quad \beta''_x = \beta''_y = \beta''_\perp$$

obtain; and in that case the discussion can proceed more economically by way of the compound quantities

$$X^{lk\lambda\kappa}_m = \sum_{m_1} S^{lk}_{m_1\, m-m_1} S^{\lambda-\kappa}_{m_1\, m-m_1} \quad (22)$$

which, at least when they are not zero, obey the formula

$$X^{lk\lambda\kappa}_m = X^{l-k\lambda-\kappa}_m = \delta_{k\kappa}/(\tfrac{1}{2}l - \tfrac{1}{2}m)!\,(\tfrac{1}{2}\lambda - \tfrac{1}{2}m)!\times(\tfrac{1}{2}m-k)!\,(\tfrac{1}{2}m+k)!\,(-2\beta_\perp)^{\frac{1}{2}l+\frac{1}{2}\lambda}. \quad (23)$$

The result (23) one derives by expanding the basis vectors $|m_1 m_2\rangle$ (for the Cartesian coordinate representation) linearly in terms of the basis vectors appropriate to the representation in plane polar coordinates,[21] calculating matrix elements, and then applying the addition theorem for binomial coefficients. The necessary conditions for $X^{lk\lambda\kappa}_m$ to be nonzero are: $l+m$, $\lambda+m$, $m \pm 2k$, $m \pm 2\kappa$ all even, l, $\lambda \geq m$, $k = \kappa$, $|k| \leq \frac{1}{2}m$; the last two requirements are expressed by formula (23), while those remaining are obvious from the definition (22) and the previously mentioned properties of the S's.

We introduce, furthermore, the abbreviations

$$Z^s_{m_3} = (2^{\frac{1}{2}s}/s!)\langle 0|\ z_s\ |m_3\rangle = \left(\frac{1}{m_3!\,\beta_z^s}\right)^{\frac{1}{2}} \frac{(s-m_3+1)!!}{(s-m_3+1)!} \quad (24)$$

and

$$|M\rangle = |m'_1\rangle\,|m'_2\rangle\,|m'_3\rangle\,|m''_1\rangle\,|m''_2\rangle\,|m''_3\rangle,$$

where the state vectors on the right pertain to the axes of x', y', etc. Then the second-order perturbation energy for two interacting molecules consisting of f' and f'' charged oscillators,

$$E^{(2)} = -f'f'' \sum_{M\neq 0} \frac{\langle 0|\ V\ |M\rangle\langle M|\ V\ |0\rangle}{E_M - E_0},$$

becomes, by Eqs. (2) and (17) and on restoring the symbol e for the electronic charge,

$$E^{(2)} = -\left(\frac{\mu e^4 f' f''}{\hbar^2}\right) \sum\sum \left(\frac{K^n_{r k_1 k_2} K^\nu_{\rho -k_1 -k_2}}{2^{(n+\nu)/2} R^{2+n+\nu}}\right) \left(\frac{X^{l_1 k_1 \lambda_1 k_1}_{m'_\perp} X^{l_2 k_2 \lambda_2 k_2}_{m''_\perp} Z^{r-l_1}_{m'_3} Z^{\rho-\lambda_1}_{m'_3} Z^{n-r-l_2}_{m''_3} Z^{\nu-\rho-\lambda_2}_{m''_3}}{m'_\perp \beta'_\perp + m'_3 \beta'_z + m''_\perp \beta''_\perp + m''_3 \beta''_z}\right), \quad (25)$$

wherein

$$K^n_{r k_1 k_2} = (-1)^{2k_1+2k_2} K^{n*}_{r -k_1 -k_2} = \sum_m (-1)^m \left[\frac{(r-2k_1)!\,(r+2k_1)!\,(n-m-2k_2)!\,(n+m+2k_2)!}{(r-m)!\,(r+m)!}\right]^{\frac{1}{2}} \times \left[\sum_l d^r_{lm}(\theta'')\, d^r_{l 2k_1}(\theta') e^{il\varphi}\right] d^n_{m+2k_2\, 0}(\theta''). \quad (26)$$

The first summation sign in (25) denotes sums over the indices n, r, l_1, l_2 and their Greek counterparts, with only even values of $n+\nu$, $r+\rho$, $l_1+\lambda_1$, $l_2+\lambda_2$ contributing, as well as over k_1 and k_2, with the $-k_1$, $-k_2$ and k_1, k_2 terms being complex conjugate to one another. The second sum extends over values of $m'_\perp \equiv m'_1 + m'_2$, $m''_\perp \equiv m''_1 + m''_2$, m'_3, m''_3 such that $m'_\perp + m''_\perp + m'_3 + m''_3 > 0$ equals $n, n-2, n-4, \cdots$ (if $n < \nu$) or $\nu, \nu-2, \nu-4, \cdots$ (if $n > \nu$).

Equation (25), whose R^{-6} term recovers London's formula,[14,15] furnishes the dispersion energy to any order between unexcited axially symmetric molecules on the basis of the Drude–London model. In applying our formula to physical problems, it is generally necessary to use a computer, but the procedure should be straightforward in view of the care we have taken

[21] See J. L. Powell and B. Crasemann, *Quantum Mechanics* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1965), Chap. 7.

in completely defining our quantities. Also, we ought to point out that such computations simplify because of the equality

$$\hat{K}{}^{n}_{rk_1k_2}(\theta', \theta'', \varphi) = (-1)^{2k_1+2k_2} K^{n*}_{n-r\,k_2k_1}(\theta'', \theta', \varphi), \quad (27)$$

a formula which is true on account of the fact that the rotation inverse to the one studied in Sec. 2 (i.e., with θ'', θ', $-\varphi$ replacing θ', θ'', φ) must produce results identical with Eqs. (17) and (25).

4. INTERFERENCE AND SPHERICALLY SYMMETRIC SYSTEMS

A bothersome feature of the energy formula (25) is the presence of summands with $n \neq \nu$, that is, joint contributions of different multipole terms V_{n-1} in the expansion of the classical potential energy; the lowest such term, stemming from the actions of both V_1 and V_3, occurs when n, $\nu = 2, 4$. It should however be observed that in the special instance of molecules having spherical symmetry all these "interference" terms disappear as we now demonstrate.

In Eq. (25), consider the sum over l_1, say, all other summation indices being held fixed. It follows from the definitions (22) and (24) that nonzero terms arise only for $l_1 = m'_\perp$, $m'_\perp + 2, \cdots (\leq r)$, and that this sum, apart from a proportionality constant, is

$$1 - \left(\frac{\beta'_z}{\beta'_\perp}\right)\frac{(r - m'_\perp - m'_3)}{2} + \left(\frac{\beta'_z}{\beta'_\perp}\right)^2 \frac{(r - m'_\perp - m'_3)(r - m'_\perp - m'_3 - 2)}{2.4} - \cdots, \quad (28)$$

wherein r takes on the values $r = m'_\perp + m'_3$, $m'_\perp + m'_3 + 2, \cdots (\leq n - 1)$. Invoking the binomial theorem, one easily sees that the sum (28) reduces to zero if $\beta'_\perp = \beta'_z$, unless $r = m'_\perp + m'_3$, which identifies the nonvanishing terms in the summation over r in (25) when the "first" molecule has spherical symmetry. The sums over l_2, λ_1, λ_2 in (25) may be subjected to similar analysis; we find if both molecules have spherical symmetry, i.e., $\beta'_\perp = \beta'_z = \beta'$ and $\beta''_\perp = \beta''_z = \beta''$, that $r = \rho = m'_\perp + m'_3$ and $n - r = \nu - \rho = m''_\perp + m''_3$. This means that, in the summations over n and ν in Eq. (25), contributions to the dispersion energy are made only by the terms $n = \nu = m'_\perp + m'_3 + m''_\perp + m''_3$, which proves our initial assertion.

One may also note that in the sum (28) it is then the first term, corresponding to $l_1 = m'_\perp$, that survives, and similar conclusions hold for the summations over l_2, λ_1, and λ_2. In total it emerges that only $l_1 = \lambda_1 = m'_\perp$ and $l_2 = \lambda_2 = m''_\perp$ represent terms contributing to $E^{(2)}$ in the present case.

There remains finally the task to derive the specialized formula for $E^{(2)}$ that obtains under the foregoing simplifying conditions—which reduce (25) to a sum over n, the k's and m's alone—and to show that it recovers previously known results. It should be noted, to begin with, that, since the dispersion energy for the spherically symmetric case must be independent of the angles θ', θ'', φ, we can put them all equal to zero, and Eq. (26) simplifies to

$$K^n_{rk_1k_2} = (-1)^{2k_1} n!\, \delta_{k_1,-k_2}. \quad (29)$$

Next we carry out the sum over k_1 and k_2 in (25), with the aid of the addition theorem for binomial coefficients, while keeping in mind that, as $k_1 = -k_2$ alone counts (i.e., $l_1 + l_2$ and $\lambda_1 + \lambda_2$ are even), only terms with $m'_\perp + m''_\perp$ even need be taken into account. Finally, recalling that $m'_\perp + m'_3 = r$ and $m''_\perp + m''_3 = n - r$, we sum over $m'_\perp$, using the same addition theorem and appealing to the equation

$$\sum_{q=0}^{\leq n/2} [4^q (q!)^2 (n - 2q)!]^{-1} = \frac{(2n)!}{2^n (n!)^3} \quad (30)$$

in which $2q = m'_\perp + m''_\perp$. The outcome is

$$E^{(2)} = -\frac{\mu e^4 f' f''}{\hbar^2 R^2} \times \sum_{n=2}^{\infty} \sum_{r=1}^{n-1} \frac{(2n)!}{(2R)^{2n} \beta'^r \beta''^{n-r} r!\,(n - r)!\,[r\beta' + (n - r)\beta'']}. \quad (31)$$

For the special case of similar molecules, $\beta' = \beta'' = \beta$, the summation over r can be performed forthwith, and we are left with

$$E^{(2)} = -\frac{\mu e^4 f' f''}{\beta \hbar^2 R^2} \sum_{n=2}^{\infty} \frac{(2n)!\,(2^n - 2)}{(2R)^{2n} n!\, n\beta^n}. \quad (32)$$

Formulas (31) and (32) are exactly equivalent to those established by Fontana[11] and Brooks,[10] respectively, for molecules possessing spherical symmetry.

ACKNOWLEDGMENT

The authors are indebted to Professor E. Tuttle for suggesting a proof of Eq. (30).