JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 3, NUMBER 1 JANUARY–FEBRUARY, 1962

Some Properties of the Five-Point Function in Perturbation Theory*

L. F. COOK, JR.
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

AND

JAN TARSKI
Institute for Advanced Study, Princeton, New Jersey

(Received May 26, 1961)

A preliminary investigation of the five-point function in its dependence on two complex variables is presented. Only single-loop diagrams are examined. The approach involves a determination of the singularity curves and of their regular and singular parts. The geometrical properties of singularity curves are described in detail; in particular, a method for determining the tangency of two curves is given. The following general conclusions are drawn: First, real and complex vertex singularities are near the physical regions and, therefore, can produce significant experimental effects. On the other hand, scattering singularities and five-point poles seem to be further removed from the physical regions. Second, it is not likely that a simple scheme can be found for the description of the analytic properties of the five-point function. A short discussion of scattering singularities involving unstable masses is also given.

1. INTRODUCTION

IN a recent paper[1] we described and discussed some of the results of an investigation concerning the singularities of production amplitudes. We emphasized in that paper the relevance of our results to the interpretation of experimental data. In particular, we discussed some of the processes which are being analyzed for the purpose of studying the π-π interaction, and the possible experimental consequences of the five-point poles.[2] We also mentioned the relevance of our results to the treatment of three-particle intermediate states in dispersion theory.

In this paper we present the details of our investigation. We studied, in effect, the analyticity of fifth-order Feynman amplitudes as functions of two complex invariants, and the results discussed in reference 1 were not the only goal of our investigation.

It may be appropriate at this point to contrast our work with previous studies of production amplitudes. One-dimensional dispersion relations have been studied by a number of authors,[3] and especially by Logunov and his collaborators.[4] More recently, Kim[5] has also investigated such dispersion relations. Various special aspects of the analyticity of production amplitudes have been discussed by Ascoli and his collaborators,[6] by Fowler, Landshoff, and Lardner,[7] by Lardner,[8] by Landshoff and Treiman,[9] and by Wu.[10] However, from

* Supported in part by the Atomic Energy Commission and in part by the U. S. A. F., Office of Scientfiic Research, Air Research and Development Command.

[1] L. F. Cook, Jr., and J. Tarski, Phys. Rev. Letters **5**, 585 (1960).

[2] R. E. Cutkosky, J. Math. Phys. **1**, 429 (1960).

[3] J. C. Polkinghorne, Nuovo cimento **4**, 216 (1956); T. W. B. Kibble, Proc. Roy. Soc. (London) **A244**, 355 (1958); G. R. Screaton, Nuovo cimento **11**, 229 (1959); G. Mohan, *ibid.* **19**, 331 (1961).

[4] A. A. Logunov and I. T. Todorov, Nuclear Phys. **10**, 552 (1959); A. A. Logunov, *ibid.* **10**, 71 (1959); further references are given in these papers.

[5] Y. S. Kim, Phys. Rev. Letters **6**, 313 (1961); Phys. Rev. **124**, 1241 and 1632 (1961).

[6] R. Ascoli, Nuovo cimento **18**, 744 and 754 (1960); R. Ascoli and A. Minguzzi, Phys. Rev. **118**, 435 (1960); R. Ascoli, A. Bottino, and A. Molinari, Nuovo cimento **19**, 687 (1961).

[7] M. Fowler, P. V. Landshoff, and R. W. Lardner, Nuovo cimento **17**, 956 (1960).

[8] R. W. Lardner, Nuovo cimento **19**, 77 (1961).

[9] P. V. Landshoff and S. B. Treiman, Nuovo cimento **19**, 1249 (1961).

[10] T. T. Wu, Phys. Rev. **123**, 678 (1961).

the recent work on scattering amplitudes it has become apparent that a utilization of additional analytic properties, i.e., in terms of additional complex variables, leads to more powerful methods for analyzing physical problems, and our investigation was carried out with this in mind.

One could argue that, since a production process which involves five particles has five kinematical invariants other than the external masses,[11] the study of the amplitude as a function of only two complex invariants can hardly be adequate. We should therefore like to justify our limited study with the following observations: First, the analytic properties for two complex invariants are easier to visualize, and to understand, because of our ability to draw diagrams in the real region (plane). Second, we expected, and our results support this, that the prominent features of the singularities already manifest themselves if two invariants are taken as complex. Third, it is not clear at the present time which regions in the space of the five complex invariants might be particularly significant. (A detailed study of the entire space would certainly be very involved, and may be unnecessary for applications.)

In short, we attempted to make a preliminary study which could serve as a basis for further work. Moreover, we did not make a systematic study even for the case of two complex invariants. Instead, we selected four diagrams corresponding to specific processes, and one configuration of momenta for each process. We then determined, for each case, analyticity of the amplitude as a function of two invariants, with the other invariants having those values which are associated with the physical process.

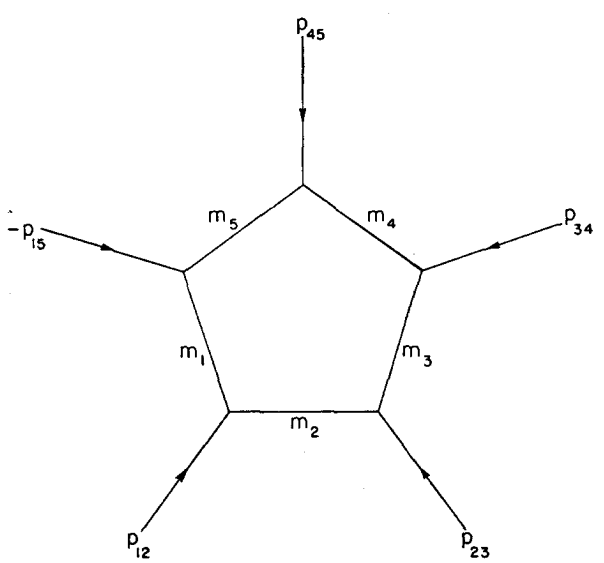


FIG. 1. Feynman diagram defining the amplitude M_5.

[11] We neglect spin dependence. As is well known, this dependence has no effect on analytic properties.

As was done in a number of earlier paper,[12–14] we restricted our investigation to single-loop diagrams. With regard to singularities which are intrinsic to more complicated diagrams, we may say this: It has been shown for some cases by Eden[15] and others that the essential properties of these singularities are also present in the single-loop diagrams for a given process, and that these singularities are further removed from the region of interest.[16] One may expect that similar conclusions are also valid for production processes. Diagrams with a simpler internal structure can be readily analyzed on the basis of previous studies,[12–14] and are discussed briefly in the sequel.

In order to give a concise description of our paper, it is convenient to review first certain definitions and facts. Our methods are based primarily on those of reference 14, which will be referred to in this paper as I. As in I, we consider a fifth-order, single-loop diagram as in Fig. 1, with the momenta and masses indicated there. The corresponding amplitude will be denoted throughout by M_5. This function will be called *the five-point function*, in analogy with the usual nomenclature. Further, we define the quantities[17] p_{ij} and x_{ij} by

$$p_{ij}=p_{i,i+1}+\cdots+p_{j-1,j}, \tag{1.1}$$

$$p_{ij}^2=m_i^2+m_j^2+2m_im_jx_{ij} \quad \text{for} \quad i<j, \tag{1.2a}$$

$$x_{ji}=x_{ij}, \quad x_{ii}=-1. \tag{1.2b}$$

The quantities p_{ij}^2, or the quantities x_{ij}, form a set of kinematical invariants[11] of the process. These sets of invariants have the feature that the five external momenta are involved in a symmetric way, and further, the quantities x_{ij} are required for the use of the established methods. However, we shall also refer to certain other sets of invariants in Sec. 5.

A convenient description of the possible singularities of M_5 is given by the following theorem (see Appendix B of I; this theorem has also been given by Polkinghorne and Screaton[18]):

Theorem 1.1: All singularities of M_5 lie on hypersurfaces defined by setting the determinants of the principal minors of the matrix (x_{ij}) equal to zero.

The curves and surfaces specified in this theorem will be called in the sequel *singularity curves* and

[12] R. Karplus, C. M. Sommerfield, and E. H. Wichmann, Phys. Rev. **111**, 1187 (1958).

[13] R. Karplus, C. M. Sommerfield, and E. H. Wichmann, Phys. Rev. **114**, 376 (1959).

[14] J. Tarski, J. Math. Phys. **1**, 149 (1960). (Referred to as I in the text.)

[15] R. J. Eden, Phys. Rev. **119**, 1763 (1960).

[16] As we were writing this paper, we were informed of results which invalidate this general statement. See R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, J. Math. Phys. **2**, 656 (1961).

[17] In references 13 and 14 quantities y_{ij} were used, where $y_{ij}=-x_{ij}$, rather than the x_{ij}. The x_{ij} preserve the signs of the quantities p_{ij}^2, and were first used by S. Mandelstam.

[18] J. C. Polkinghorne and G. R. Screaton, Nuovo cimento **15**, 289 (1960).

surfaces, without regard as to whether a given curve or surface may be singular or regular.

We select two of the invariants, x_{13} and x_{25}, to be our complex variables, and the other invariants will be considered to be parameters. We shall arrange the momenta for the processes in such a way that x_{13} is always related to the total incoming four-momentum squared, and x_{25} is related to the square of the four-momentum transfer of the heaviest particles, if possible. (We can arrange the momenta in this way since we do not study crossed diagrams. A brief discussion of this point is given in Sec. 5.) A reason for choosing x_{25} as one of the variables is given in reference 1, in connection with the extrapolation of experimental data.

One of the methods which we use for the determination of singularities of M_5 depends on the variation of parameters. To use this method we first choose values of the parameters for which the analytic properties of M_5 can be determined directly. In particular, we made a choice such that there exists a region in the real (x_{13},x_{25}) plane where M_5 is analytic and real (aside from spin factors), and the analytic properties of M_5 can then be studied by previous methods.[13,14] We then vary the parameters to the values associated with the physical processes in question, and we use methods which are discussed in I (and in references 2, 15, and 18) to follow the behavior of singularities. In the sequel we shall refer to this method as the continuation procedure. We also use certain other techniques which are less laborious but which are more limited in scope.

We remark that our emphasis in this paper is on singularities for which $\det(x_{ij})=0$, since the other singularities are also associated with lower order diagrams, and are reasonably well understood. Further, most of our discussion is confined to the real (x_{13},x_{25}) plane. This may appear at first sight to be a rather severe limitation, but in actual fact it lays the groundwork for the study of complex singularities. Our treatment of the complex singularities is somewhat brief, but we feel it is adequate for a preliminary study.

The geometrical properties of the singularity curves and surfaces, defined by Theorem 1.1, are discussed in Sec. 2, and in Sec. 3 we give a general discussion of the analytic properties of these curves and surfaces. Most of the discussion of these two sections consists of relating the existing techniques[13–15] to the amplitude M_5. However, we give a detailed discussion of the problem of tangency of curves: In Sec. 2 we describe a method for determining such points of tangency, and in Sec. 3 we discuss the analytic behavior near them. These considerations are essential for the determination of singularities. In Sec. 4 we determine the singularities for the desired values of the parameters, and we illustrate in detail the continuation procedure as well as other methods. A discussion of our results is given in Sec. 5. We discuss there briefly some aspects of our technique, we comment on the validity of dispersion relations and on complex singularities, and we also supplement the discussion of reference 1 on the theoretical and possible experimental consequences of our results.

Finally, in the Appendix we summarize the basic properties of the singularities of vertex functions and scattering amplitudes. We also discuss there briefly these singularities for the case of unstable external particles, since a familiarity with such singularities is needed for the main part of the paper.

2. GEOMETRICAL PROPERTIES OF SINGULARITY CURVES

In this section we will discuss some of the geometrical properties of the singularity curves, i.e., of the curves which are specified in Theorem 1.1. In accordance with our remarks in the Introduction, most of this section deals with curves in the real (x_{13},x_{25}) plane. However, a few remarks concerning the complex surfaces are also included.

A. General Remarks

We begin by defining our notation. The determinant of the principal submatrix of (x_{ij}) obtained by deleting rows and columns with indices $k, \cdots, l$ will be denoted by $K_{k,\cdots,l}$, and further, let

$$\Delta=\det(x_{ij}). \tag{2.1}$$

The cofactor of the (k,l) entry of (x_{ij}) will be denoted by B_{kl} (note that $B_{kl}=B_{lk}$, and $B_{ll}=K_l$). For example,

$$B_{24}=\det\begin{pmatrix} -1 & x_{12} & x_{13} & x_{15} \\ x_{13} & x_{23} & -1 & x_{35} \\ x_{14} & x_{24} & x_{34} & x_{45} \\ x_{15} & x_{25} & x_{35} & -1 \end{pmatrix}. \tag{2.2}$$

As in references 13 and 14 we introduce the angles θ_{ij} as follows: If $-1\leq x_{ij}\leq 1$, then θ_{ij} is defined by

$$\cos\theta_{ij}=-x_{ij}, \quad 0\leq\theta_{ij}\leq\pi. \tag{2.3}$$

We also adopt the convention that indices are to be taken modulo 5. Thus $\theta_{i,i+1}$ refers also to θ_{15} (since $x_{15}=x_{51}$).

It is important to realize that the condition $-1\leq x_{ij}\leq 1$ is related to stability.[13,14] In general, the four-momentum p_{ij} can create a real state with mass m_i+m_j if $x_{ij}\geq 1$. In particular, the external mass $M_{i,i+1}{}^2=p_{i,i+1}{}^2$ is stable if

$$x_{i,i+1}<1. \tag{2.4}$$

The conditions for the stability of the internal mass m_i are

$$x_{i-1,i},\ x_{i,i+1}>-1. \tag{2.5}$$

For purposes of orientation, we will review briefly certain properties of contracted diagrams. We recall[19]

[19] L. D. Landau, Nuclear Phys. 13, 181 (1959).

that the singularities associated with a given diagram include the singularities associated with the contracted diagrams (obtained by removing certain internal lines and joining their endpoints), and the situation with singularity curves is analogous. In particular, the singularity curve defined by $K_{k,\cdots,l}=0$ is also associated with the diagram obtained from that of Fig. 1 by contracting internal lines with indices $k, \cdots, l$. We shall therefore speak of curves defined by equations of the form $K_i=0$ and $K_{jk}=0$ as *scattering curves* and *vertex curves*, respectively. The singularity curve which is not associated with any contracted diagram is called the *leading curve* of the diagram. In the case of M_5, the leading curve is given by $\det(x_{ij})=0$.

Let us now consider those principal minors of (x_{ij}) which contain at least one of the variables x_{13} and x_{25}. There are fourteen minors [including (x_{ij}) itself] to be mentioned.

There are two 2×2 minors, which yield,

$$K_{245}=0 \quad \text{or} \quad x_{13}=\pm1, \tag{2.6a}$$

$$K_{134}=0 \quad \text{or} \quad x_{25}=\pm1. \tag{2.6b}$$

We see that these minors are independent of the parameters. There are six 3×3 minors. Of these, three involve x_{13}, and yield

$$K_{24}=0, \quad K_{45}=0, \tag{2.7a,b}$$

$$K_{25}=0, \tag{2.7c}$$

while the three remaining involve x_{25}, and yield,

$$K_{14}=0, \quad K_{34}=0, \tag{2.8a,b}$$

$$K_{13}=0. \tag{2.8c}$$

The expansion of these minors is given by the equation

$$K_{45}=-1+x_{12}^2+x_{13}^2+x_{23}^2+2x_{12}x_{13}x_{23}, \tag{2.9}$$

and its permutations. Further, the equation $K_{45}=0$ has the solution

$$x_{13}=-x_{12}x_{23}\pm[(1-x_{12}^2)(1-x_{23}^2)]^{\frac{1}{2}}, \tag{2.10}$$

and this can be written more compactly in terms of the angles θ_{ij} as follows:

$$x_{13}=-\cos(\theta_{12}\pm\theta_{23}). \tag{2.11}$$

Of the equations involving the 4×4 minors, there are two which involve only x_{13},

$$K_2=0 \quad \text{and} \quad K_5=0, \tag{2.12a,b}$$

and two which involve only x_{25},

$$K_1=0 \quad \text{and} \quad K_3=0. \tag{2.13a,b}$$

The remaining 4×4 minor yields

$$K_4=0. \tag{2.14}$$

This equation involves both x_{13} and x_{25}, and therefore represents a curve in the (x_{13},x_{25}) plane. We shall call this curve γ. The equation

$$\Delta=0 \tag{2.15}$$

similarly involves both variables, and the resulting curve will be called α. Finally, we shall also discuss in the sequel the complex surfaces defined by Eqs. (2.14) and (2.15), and these surfaces will be called σ_γ and σ_α, respectively.

The above determinants involve the variables x_{13} and x_{25} quadratically, and the most significant quantities associated with a quadratic expression are the discriminant and the coefficient in the quadratic term. The discriminants which we need are described concisely in the following theorem, which is proved in Appendix B of I. This theorem relates the above determinants, and provides the basis for much of the remainder of Sec. 2.

Theorem 2.1: Let (v_{ij}) be a symmetric matrix. Let V_{kl} be the cofactor of the (k,l) entry of (v_{ij}), and let $V_{kk}=L_k$. Let v_{pq} be a variable distinct from all other v_{ij} (except v_{qp}), and let us write $\det(v_{ij})=Av_{pq}^2+Bv_{pq}+C$. Then

$$B^2-4AC=4L_pL_q. \tag{2.16}$$

The coefficients of the quadratic and linear terms, A and B, also have simple expressions. Let Λ_{jk} be the determinant of the submatrix of (v_{ij}) obtained by deleting rows and columns with indices j and k. Then

$$A=-\Lambda_{pq}. \tag{2.17}$$

Further,

$$-(Av_{pq}+\tfrac{1}{2}B)=V_{pq};\ B=-2V_{pq}|_{v_{pq}=0}. \tag{2.18a,b}$$

We shall apply this theorem to both the determinant Δ and the determinants of its various submatrices. In the case of Δ, we have $V_{pq}=B_{pq}$, $L_p=K_p$ and $\Lambda_{pq}=K_{pq}$. In the case of K_j we have $L_p=K_{pj}$, etc. As we shall see, Theorem 2.1, together with Eq. (2.17), allows us to study the singularity curves in great detail: The discriminants, Eq. (2.16), determine the tangents to the curve, and the quadratic coefficients, Eq. (2.17), the asymptotes.

For completeness we also give some formulas relevant to the determinants K_i and Δ. The determinant K_5, e.g., has the following expansion [see I, Eq. (2.7)]:

$$\begin{aligned}K_5=1-&\sum_{1\le i<j\le4}x_{ij}^2-2\sum_{1\le i<j<k\le4}x_{ij}x_{jk}x_{ik}\\&+x_{12}^2x_{34}^2+x_{13}^2x_{24}^2+x_{14}^2x_{23}^2-2x_{12}x_{13}x_{24}x_{34}\\&-2x_{12}x_{14}x_{23}x_{34}-2x_{13}x_{14}x_{23}x_{24}.\end{aligned} \tag{2.19}$$

The equation $K_5=0$ may be solved for, e.g., x_{13} [Eq. (29) of reference 13]:

$$\begin{aligned}x_{13}=[1/(1-x_{24}^2)][&x_{12}x_{23}+x_{14}x_{34}\\&+x_{24}(x_{12}x_{34}+x_{14}x_{23})\pm(K_{15}K_{35})^{\frac{1}{2}}].\end{aligned} \tag{2.20}$$

We spare the reader the sight of the expansion of Δ, but the solution of the equation $\Delta=0$ may be expressed as follows:

$$x_{ij}=(-1/K_{ij})[B_{ij}|_{x_{ij}=0}\pm(K_iK_j)^{\frac{1}{2}}]. \quad (2.21)$$

We conclude this section with the following remark: Let two singularity curves be given, one of these being the leading curve of an arbitrary Feynman diagram, and the other, the leading curve of a diagram obtained from the original by a single contraction. Then it follows from Landau's equations[19] and the theory of envelopes that the two curves are tangent when they meet (excepting special limiting cases). Theorem 2.1 states essentially the same property, but only for single-loop diagrams. However, in this paper we shall make direct use of Eq. (2.16). A derivation of Eq. (2.16) from Landau's equations seems to be quite involved, since a detailed study of the reducibility of determinants is apparently required.

B. The Curve γ

In this section we shall discuss some of the geometrical properties of scattering curves which are defined by an equation $K_i=0$. For definiteness we shall consider the curve γ, since this curve is one of our singularity curves. Its properties are formally the same as those of the curve Γ discussed in detail in I. Here our emphasis is to review those properties of scattering curves which are also basic for the study of the curve α.

We consider the real (x_{13},x_{25}) plane, and we write K_4 as follows:

$$K_4=Ax_{13}^2+Bx_{13}+C=ax_{25}^2+bx_{25}+c. \quad (2.22)$$

Theorem 2.1 states that

$$B^2-4AC=4K_{14}K_{34}, \quad b^2-4ac=4K_{24}K_{45}. \quad (2.23a,b)$$

These equations imply that the values of x_{25} satisfying $K_{14}=0$ or $K_{34}=0$ define tangents parallel to the x_{13} axis, and the values of x_{13} satisfying $K_{24}=0$ or $K_{45}=0$ define tangents parallel to the x_{25} axis [cf. Eqs. (2.7a,b) and (2.8a,b)]. We obtain in this way at most four vertical and four horizontal tangent lines to γ, and there are no other vertical or horizontal tangent lines. These tangent lines partition the (x_{13},x_{25}) plane into regions, and in each region each of the discriminants (2.23a,b) is either positive or negative. If at least one of them is negative in a given region, there cannot be a real branch of γ in that region. We shall call such regions *forbidden*, and the remaining regions, *allowed*.

We next consider the asymptotes to γ. These are determined by the condition that $A=0$ or $a=0$ in Eq. (2.22), and are given [see Eqs. (2.6a,b) and (2.17)] by $x_{13}=\pm1$, $x_{25}\pm1$. The existence of these real asymptotes and the fact that the equation $K_4=0$ is quadratic in each of the variables easily imply that there are no asymptotes to γ which are at an angle to the coordinate axes, and that an allowed region which extends to infinity contains a branch of the curve if and only if it also contains an asymptote which extends to infinity in that region. A typical configuration of γ is given in Fig. 2. Such a configuration occurs when the parameters x_{ij}, with $i,\ j\neq4$, are such that $-1<x_{ij}<1$ [cf. I, Fig. 3; cf. also Eqs. (2.4, 5)]. Certain other configurations are given in the Appendix, and by Fowler *et al.*[7]

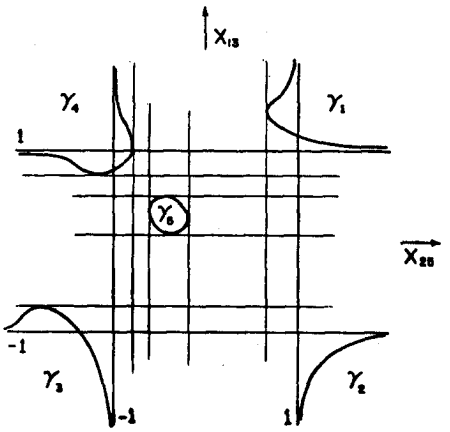


FIG. 2. The curve γ.

If the branches of γ are arranged as in Fig. 2, we shall label them $\gamma_1,\ \ldots,\ \gamma_5$, as on the figure. If there is no central oval, but the arrangement is otherwise analogous, we shall label the branches as $\gamma_1,\ \ldots,\ \gamma_4$.

It is sometimes useful to know which branch of γ is tangent to a given tangent line. For definiteness, let the line in question be given by $x_{13}=-\cos(\theta_{12}\pm\theta_{23})$, so that $K_{45}=0$ [Eqs. (2.9–11)]. We solve $K_4=0$ for x_{25} [Eq. (2.20)], eliminate x_{13}, and obtain for the point of tangency [Eqs. (27) and (28) of reference 13],

$$x_{25}=-(\cos\theta_{35}\sin\theta_{12}\pm\cos\theta_{15}\sin\theta_{23})/\sin(\theta_{12}\pm\theta_{23}). \quad (2.24)$$

The permutations of this equation give the other points of tangency. This equation has the following important property: If one of the parameters, x_{12} or x_{23}, is varied in such a way that the tangent line, $x_{13}=-\cos(\theta_{12}\pm\theta_{23})$, moves to one of the asymptotes $x_{13}=\pm1$ and back, then the denominator $\sin(\theta_{12}\pm\theta_{23})$ passes through zero and changes sign. Since in general the numerator will remain different from zero during such a variation, we conclude that the x_{25} coordinate of the point of tangency will approach $\pm\infty$, and then recede from $\mp\infty$. Moreover, in general, different branches of γ will be tangent to the line before and after the variation.

We conclude this section with a brief discussion of the complex surface σ_γ, defined by $K_4=0$. As in I, we consider the system of equations (μ, ν, and λ real; $\mu,\ \nu\neq0$):

$$K_4=0, \quad \mu x_{13}+\nu x_{25}=\lambda. \quad (2.25a,b)$$

This is a fourth-degree system, and has four points as solutions. Figure 2 shows that two of the solutions are

always real, and the other two may be real or complex. (However, distortions of the figures, made for clarity, might lead to six real roots.) Since every complex point (x_{13},x_{25}) satisfies Eq. (2.25b) for some real μ, ν, and λ, we conclude that the complex solutions of the system (2.25a,b) generate the entire complex surface σ_γ. An examination of the system (2.25a,b) also shows how a given arc of γ extends into the complex region along σ_γ, and how a given section of σ_γ joint two disconnected branches of γ. In particular, we make the following observation: If a given arc of γ satisfies $dx_{13}/dx_{25}>0$, then it extends to σ_γ in such a way that $\mathrm{Im}x_{13}$ and $\mathrm{Im}x_{25}$ are of the same sign there. If $dx_{13}/dx_{25}<0$, then $\mathrm{Im}x_{13}$ and $\mathrm{Im}x_{25}$ are of opposite signs on the complex extension.

C. The Curve α

We now consider the curve α. The determination of this curve is in principle analogous to that given for γ. In this case, the tangent lines are determined by the equations

$$K_1(x_{25})=0, \quad K_3(x_{25})=0, \qquad (2.26a,b)$$

$$K_2(x_{13})=0, \quad K_5(x_{13})=0, \qquad (2.27a,b)$$

where we have indicated explicitly the dependence on the variables. These tangent lines partition the (x_{13},x_{25}) plane into allowed and forbidden regions, just as is the case with γ.

We see from Eq. (2.17) that the asymptotes are given by the equations

$$K_{13}(x_{25})=0, \quad K_{25}(x_{13})=0. \qquad (2.28a,b)$$

In contrast to the case of scattering curves, e.g., γ, these asymptotes depend on the other parameters and consequently may be complex. This fact allows α to have a more diversified character than that of γ. In particular, while γ is necessarily unbounded in both x_{13} and x_{25} (the asymptotes are always real), α may be unbounded in both variables, bounded in only one variable, or bounded in both variables, depending on whether the asymptotes are real or complex.

The conditions for a pair of asymptotes to be real or complex may be readily obtained from Eq. (2.10), with indices permuted. For example, the asymptotes defined by $K_{13}=0$ are real if

$$|x_{14}|, |x_{34}| \leq 1 \quad \text{or} \quad |x_{14}|, |x_{34}| \geq 1 \qquad (2.29)$$

and complex otherwise. [cf. the stability conditions (2.4, 5). We assume that x_{14} and x_{34} are real.]

Of course, the absence of asymptotes parallel to coordinate axes does not immediately imply the boundedness of α. However, this boundedness can be established directly as follows: The expansion given in Eq. (2.19) shows that, if the conditions (2.29) are both violated, then $K_1K_3<0$ for large $|x_{25}|$, and α is bounded in x_{25}. This boundedness is also implied by the considerations that follow.

The tangent lines to α, particularly their variation as the parameters are varied, can be studied in the following manner. Let us consider a specific pair of tangent lines, say those given by $K_3=0$. This equation defines two values of x_{25}, which determine the tangent lines in question. On the other hand, the equation $K_3=0$ also defines a curve in the real (x_{14},x_{25}) plane which has all the properties of a scattering curve such as γ in Fig. 2. [We examine the (x_{14},x_{25}) plane, since the corresponding curve in a plane (x_{2i},x_{25}) or (x_{j5},x_{25}) would not be analogous to γ, and the discussion of Sec. 2(B) would not be applicable.] Intersections between this curve and a line $x_{14}=$constant will yield values of x_{25} which give the tangent lines to α, and of course, depend on the value chosen for x_{14}. While the explicit solution of the equation $K_3=0$, cf. Eq. (2.20), is not easy to handle, one can obtain in a semi-quantitative way the behavior of these lines, as a given parameter is varied, by examining the curve given by the equation $K_3=0$ in the real (x_{14},x_{25}) plane.

As in the case of γ, the curve $K_3=0$ is delimited by tangent lines which are determined by 3×3 determinants, and which can be easily obtained [see Eqs. (2.10, 11)]. Additional information about the curve $K_3=0$ is provided by its asymptotes, and by the points of tangency to the tangent lines [Eq. (2.24)].

Figure 3 shows how the considerations just described can be applied. We determine there the tangent lines for six values of one parameter, x_{45}, and, on this basis, we construct α for each set of values. Figure 3 shows three types of changes of α: one is from case (a) to (b), another from (c) to (d), and the third, from (e) to (f). We may note that the asymptotes to α are included among the tangent lines to the curves $K_i=0$, and therefore one can easily determine whether a given tangent line to α is inside or outside the asymptotes. It is also desirable to know which branch of α is tangent to a given tangent line, and we shall discuss this question in Sec. 2(D). We should say that an understanding of transitions such as shown in Fig. 3 is essential if one wishes to follow the behavior of singularities with the variation of a parameter.

We wish to conclude this section with two remarks. The first of these deals with the determination of the complex surface σ_α, which is the complex extension of α. We only note here that, once the real form of α is known, we can use the method which was given in Sec. 2(B) for the determination of σ_γ.

Our second remark is that in our discussion of γ and of α we have mentioned each of the singularity curves listed in Sec. 2(A). It is helpful to note, for orientation, that every one of these curves falls into one of two following classes:

(1) the curve γ together with its tangent lines and asymptotes, and
(2) the curve α together with its tangent lines and asymptotes.

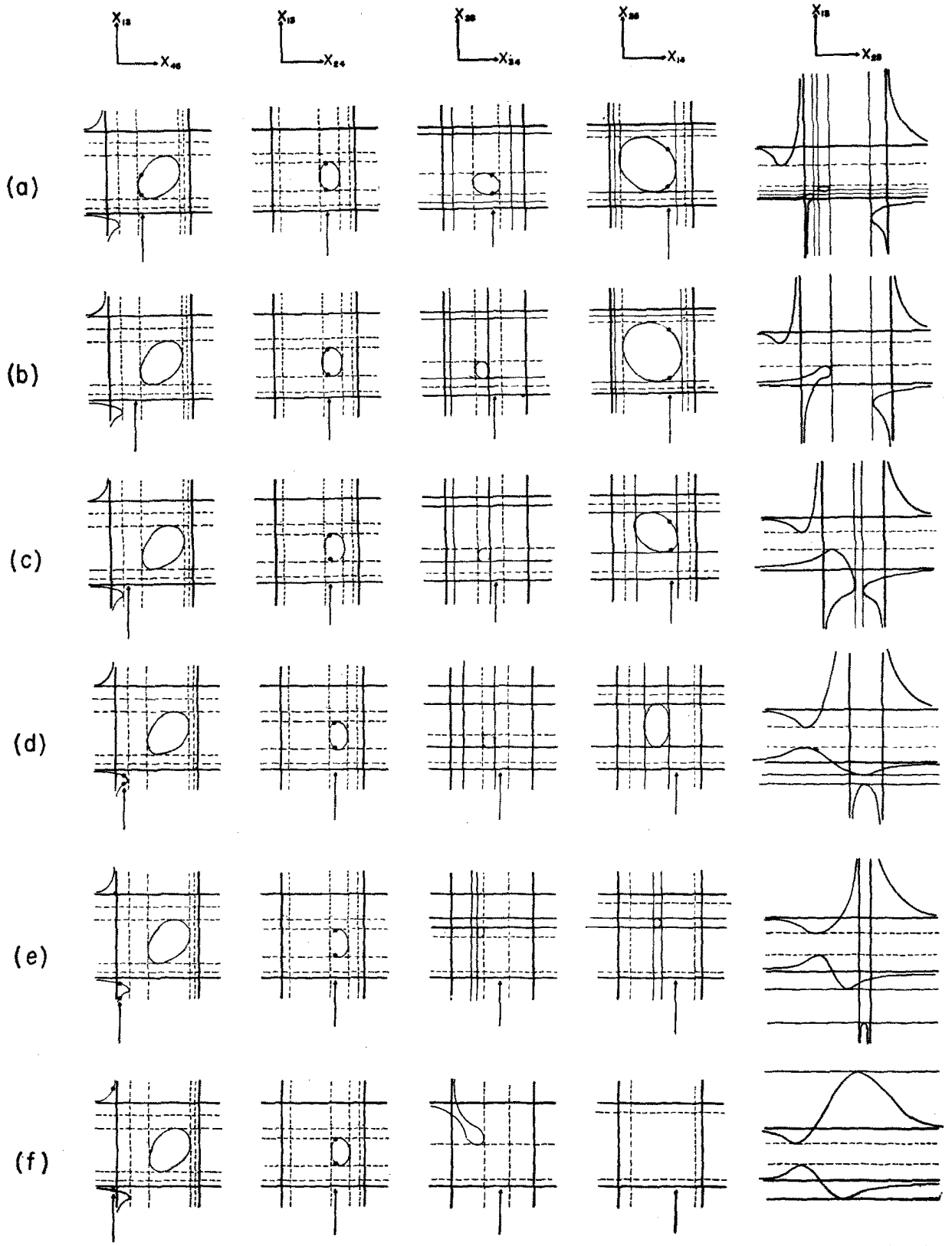


FIG. 3. The deformations of α as a parameter is varied. The rows (a)–(f) correspond to different values of x_{45}, with the other parameters fixed. The tangent lines which we dashed are not functions of x_{45}, and do not move as x_{45} is varied. The intersections in the various scattering diagrams are marked by points and yield the tangent lines for α. In the last column the resulting configurations of α are given.

D. Properties of Tangent Lines to α

There are some properties of the tangent lines to α, and of the associated points of tangency, which are requisite for an understanding of the continuation procedure. An examination of the cases of Fig. 3 and of the indicated transitions leads us to the following observation: If in the variation of a parameter a pair of tangent lines parallel to, say, the x_{25} axis appears or disappears at one value of the parameter, then a pair of tangent lines parallel to the x_{13} axis also appears or disappears at the same value of the parameter. (But this is not true for asymptotes.) This behavior can be understood precisely as follows: Consider the pair of tangent lines parallel to the x_{13} axis and given by $K_3=0$. If these two lines are to coincide then either $K_{23}=0$ or $K_{35}=0$. But if $K_{23}=0$, then the tangent lines given by $K_2=0$ coincide (these are parallel to the x_{25} axis), and if $K_{35}=0$, then those given by $K_5=0$ coincide.

We may observe, more generally, that if any two tangent lines coincide, then the number and continuity of solutions of $\Delta=0$ implies that the two points of tangency must also coincide (but we no longer have tangency in the usual sense). We then conclude that two perpendicular tangent lines must also coincide, and that the four tangent points to α all coincide at the intersection of the tangent lines.

Let us now consider the case when two nonpaired tangent lines coincide, say, one of which is determined by $K_1=0$, and the other, by $K_3=0$. Then we conclude from the preceding discussion that one of the tangent lines determined by $K_2=0$ and one determined by $K_5=0$ must also coincide. However, we are unable to give a more direct way of showing this. [In case of the scattering curves, some simple conditions for the coincidence of nonpaired lines have been obtained: $\Sigma\theta_{ij}=2\pi$; $\Sigma\theta_{ij}=2\pi+2\min(\theta_{ij})$; see Appendix C of I. The conditions for the appearance or disappearance of tangent lines to scattering curves are the following: $x_{ij}=\pm 1$, or $\theta_{ij}=0$ or π, for one of the parameters.]

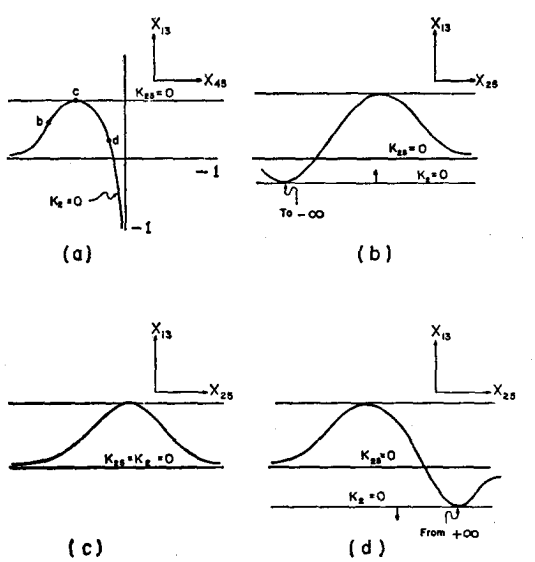


FIG. 4. The behavior of one of the points of tangency of α as a specific parameter is varied. This variation is such that a tangent line approaches and recedes from an asymptote. Figures (b)–(d) correspond to the values of x_{45} which are associated with points b–d of Fig. (a), respectively.

The considerations just presented lead to another, which concerns the orderings of the tangent lines. We easily conclude that the orderings of the tangent lines parallel to the x_{13} axis and the orderings of the tangent lines parallel to the x_{25} axis are analogous. For example, if the strip between the two tangent lines determined by $K_1=0$ contains both of the lines $K_3=0$, then either the strip between the two lines $K_2=0$ contains both of the lines $K_5=0$, or the strip between the two lines $K_5=0$ contains both of the lines $K_2=0$. (See e.g., Fig. 2 of I.) However, if one pair of asymptotes is real and the other complex, then this example requires a trivial modification.

We still have to discuss the location of the points of tangency between α and its tangent lines. For the case of the scattering curves, we have Eq. (2.24) which describes explicitly the dependence of the point of tangency on the parameters. For α we obtain easily the following expression for the point of tangency:

$$x_{25}=-B_{25}|_{x_{25}=0}/K_{25}. \tag{2.30}$$

The variable x_{13} is to be eliminated from the rhs of this equation by the condition $K_2=0$ or $K_5=0$, depending on the tangent line [cf. Eqs. (2.20, 21)]. The elimination of x_{13} in closed form appears to be a formidable task, and we shall therefore attempt to exploit the properties of Eq. (2.30) in order to determine the behavior of the point of tangency as a parameter is varied.

We explained in the foregoing the behavior of a point of tangency if the tangent line comes into coincidence with another tangent line. The other case that requires an explanation is that of the tangent line coming into coincidence with an asymptote. Let us therefore assume for definiteness that the tangent line in question is given by $K_2(x_{13})=0$, that we vary x_{14} in such a way that the tangent line approaches and recedes from the asymptote, and that $-1<x_{13}, x_{14}, x_{34}<1$ in the region of interest, so that the corresponding angles θ_{ij} are real. Let $\theta_{14}=\theta_{14}^0$ when the two lines in question coincide. The solution (2.11) of Eq. (2.9) leads to the following relation:

$$K_{25}=-4\sin\tfrac{1}{2}(\theta_{13}+\theta_{14}+\theta_{34})\sin\tfrac{1}{2}(\theta_{13}+\theta_{14}-\theta_{34}) \times\sin\tfrac{1}{2}(\theta_{13}-\theta_{14}+\theta_{34})\sin\tfrac{1}{2}(-\theta_{13}+\theta_{14}+\theta_{34}). \tag{2.31}$$

Excluding limiting cases, if $K_{25}=0$ then one and only one of the factors in the rhs of Eq. (2.31) vanishes. Moreover, if we require that $K_2=0$ then θ_{13} is a function of θ_{14}, and

$$\theta_{13}'(\theta_{14}^0)=0. \tag{2.32}$$

This relation can be obtained by examining Fig. 3:

The equation $K_{25}=0$ defines two of the tangent lines to the scattering curve $K_2=0$, and this tangency between the scattering curve $K_2=0$ and the line $K_{25}=0$ easily proves the assertion. We also conclude, using Eq. (2.32), that K_{25} has a simple zero when $\theta_{14}=\theta_{14}^0$.

We now conclude that the rhs of Eq. (2.30) has a simple pole when $\theta_{14}=\theta_{14}^0$. By the preceding paragraph this pole is not of a higher order, while the possibility that x_{25}, as given by Eq. (2.30), approaches a finite limit as $\theta_{14}\to\theta_{14}^0$ is excluded by a consideration of the number and continuity of roots of the equation $\Delta=0$. Furthermore, we conclude, just as in the case of the scattering curve, that the point of tangency recedes from infinity with a sign *opposite* to that with which it approached infinity. Figure 4 shows the situation here described, as well as the corresponding scattering curve which determines the tangent line. In other cases, as e.g., in Fig. 3, the point of tangency would recede from infinity along a different branch than that along which it approached infinity.

E. Tangencies between the Curves γ and α

In our approach to the five-point function a complication arises which has not been encountered in previous studies. As we remarked previously, it follows from Landau's equations, and also from Theorem 2.1, that if the curves γ and α have a point in common, then they are tangent there. (Certain exceptions to this occur, but these are limiting cases which do not require a special treatment.) A knowledge of such points of tangency is essential for a determination of the singular nature of α.

In our investigation the curves α and γ were obtained by computation, but such numerical methods are not adequate to determine the presence of a tangent point. We therefore need a suitable method for the study of the points of tangency.

We use Theorem 2.1 and the subsequent discussion, and obtain

$$\Delta=-K_{24}x_{24}^2+bx_{24}+c$$
$$=-K_{24}^{-1}[(-K_{24}x_{24}+\tfrac{1}{2}b)^2-K_2K_4], \quad (2.33)$$

and therefore,

$$K_2K_4-K_{24}\Delta-B_{24}^2=0. \quad (2.34)$$

The determinant B_{24} is, we recall, the cofactor of x_{24} in (x_{ij}), and is displayed in Eq. (2.2). Equation (2.34) shows that, as long as $K_2\neq0$ and $K_{24}\neq0$, any two of the three conditions

$$K_4=0, \quad \Delta=0, \quad B_{24}=0 \quad (2.35)$$

imply the third, and that equation also proves that α and γ are tangent. Now the curve $B_{24}=0$, in general, intersects α and γ, and these intersections yield the points of tangency between α and γ. We have thus reduced the determination of points of tangency to the determination of points of intersection. However,

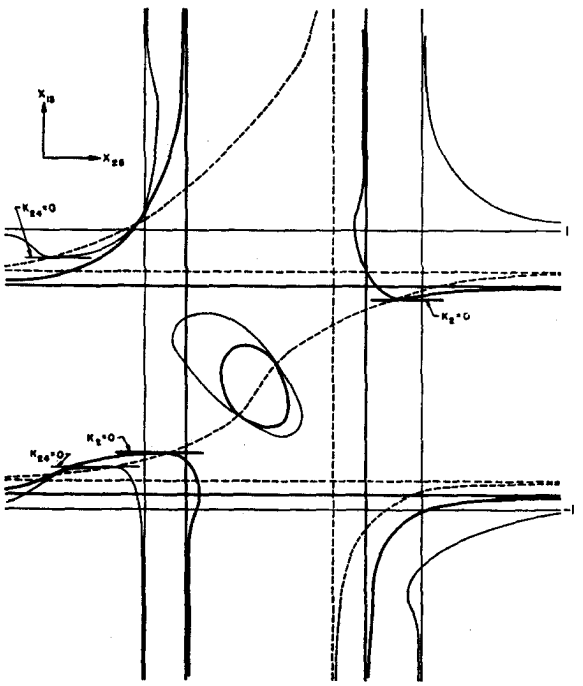


FIG. 5. The curves α, β, γ for a set of parameters. The curve α is heavy, β is dashed, and γ is light. The asymptotes to these curves are shown in a similar manner. (The set of parameters which we used corresponds to the initial process described in Sec. 4, but distortions were made for clarity.)

intersections of curves are in principle easy to determine, and graphical methods can always be used.

The curves defined by the equation $B_{24}=0$ will be called β. Equation (2.34) shows that β passes through the point of tangency of γ to the line $K_{24}=0$, and through the point of tangency of α to the line $K_2=0$. All other intersections of β and γ are also intersections of β and α (and conversely), and are also the points of tangency between α and γ. Figure 5 shows the there curves for a set of parameters.

Before we turn to a discussion of the properties of β, we should like to make three remarks. The first of these deals with the choice of x_{24} in Eq. (2.30). This choice was arbitrary, and we can easily show that any other x_{i4} would lead to the same results. Indeed, it is easy to see that any two equations $B_{i4}=0$, $B_{j4}=0$ imply that $B_{k4}=0$ for all k. [We recall that $B_{44}=K_4$. We also note that the condition that $B_{k4}=0$ for all k is necessary and sufficient for the 4×5 matrix, obtained from (x_{ij}) by deleting column 4, to have rank 3 or less.]

The second remark follows directly from the first. If two vertical tangent lines coincide, then, as we observed in Sec. 2(D), two horizontal tangent lines also coincide, and α passes through the intersection of these tangent lines. At this point of intersection we have, e.g., $K_1=0$, $K_2=0$, and $\Delta=0$. Then $B_{14}=0$, $B_{24}=0$, and consequently $K_4=B_{44}=0$. We see that when two branches of α merge, they pinch the curve β and also the curves $B_{i4}=0$, including γ.

The third remark concerns our choice of variables as x_{13} and x_{25}. If we had chosen our variables so that they had an index in common, say x_{24} and x_{25}, then we would have had to study the points of tangency between the curves $K_{13}=0$, $K_1=0$, $K_3=0$, and $\Delta=0$. Our choice of variables without a common index, like x_{13} and x_{25}, seems to be the more convenient one.

F. Properties of the Curve β

As we concluded in Sec. 2(E), graphical methods can be used to determine the curve β and its relevant intersections. However, it is also very desirable to obtain a few properties of β which would enable one to construct a rough sketch of this curve without extensive computations.

The most important property of the equation $B_{24}=0$ is that it is linear in x_{25} and quadratic in x_{13} [see Eq. (2.2)]. We may solve $B_{24}=0$ for x_{25}, and the solution can be expressed as follows:

$$x_{25}=-(x_{24}x_{13}^2+Rx_{13}+S)/Q, \tag{2.36}$$

where

$$Q=x_{45}x_{13}^2-(x_{14}x_{35}+x_{15}x_{34})x_{13} -(x_{45}+x_{14}x_{15}+x_{34}x_{35}), \tag{2.37}$$

and where R and S are polynomials in the parameters and are independent of x_{13} and x_{25}.

We next consider the asymptotes and tangents to β. There is one asymptote parallel to the x_{13} axis, and it is obtained from the quotient of the leading coefficients in the rhs of Eq. (2.36):

$$x_{25}=-x_{24}/x_{45}. \tag{2.38}$$

There are also two asymptotes parallel to the x_{25} axis, and these are obtained by solving the equation $Q=0$. These latter asymptotes may be real or complex, and in the latter case β extends only over a finite range of values of x_{25}. Moreover, there are two tangent lines to β which are parallel to the x_{13} axis, and these also may be real or complex. However, we have not been able to find a reasonably simple method for the determination of these tangent lines. Figure 6 shows various configurations of β. We have also included in this figure certain limiting configurations which are needed for an understanding of the continuity among the cases.

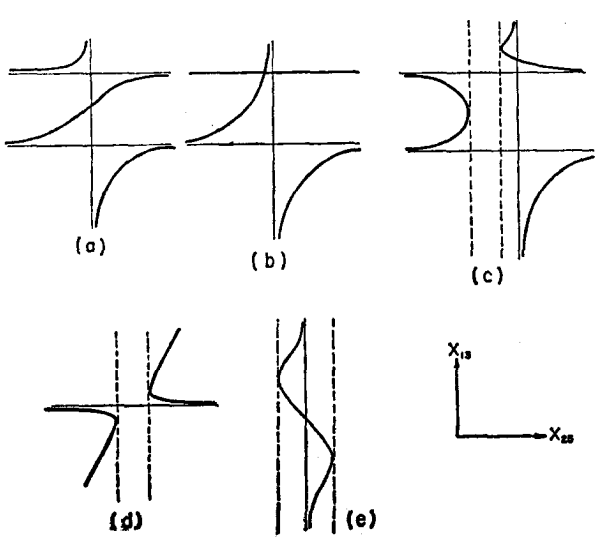


FIG. 6. Various configurations of the curve β.

We can add to the above conclusions a few further observations which are helpful in determining the approximate form of β. We recall from Sec. 2(E) that β passes through the points of tangency of γ to the lines $K_{24}=0$, and through the points of tangency of α to the lines $K_2=0$. The former of these points can be obtained from Eq. (2.24) (with indices permuted), and the latter can be obtained approximately if we know the shape of α. Also, we discussed in Sec. 2(E) the behavior of β when some of the tangent lines to α coincide.

Another property of β which should like to mention concerns some of the limiting cases. We easily see that if $x_{45}=0$ then β is a hyperbola, such as shown in Fig. 6(d). The case of Fig. 6(b), where β consists of a hyperbola and of a line parallel to the x_{25} axis, say $x_{13}=x_{13}^0$, occurs when $K_2=K_{24}=0$ for $x_{13}=x_{13}^0$. (The equations $K_2=K_{24}=0$ are sufficient, but not necessary, for the configuration of β in question.) The case $K_2=K_{24}=0$ corresponds to the degenerate case (3) discussed in Appendix C of I. From the discussion of I we conclude that this occurs when

$$\theta_{15}+\theta_{35}=\pi \quad \text{and} \quad \theta_{14}+\theta_{34}+\pi, \tag{2.39a}$$

and then $x_{13}^0=1$, or else, when

$$\theta_{15}=\theta_{35} \quad \text{and} \quad \theta_{14}=\theta_{35}, \tag{2.39b}$$

and then $x_{13}^0=-1$.

The last property of β which we will discuss is the number of its intersections with γ and with α. Consider the system

$$B_{24}=0, \quad K_4=0. \tag{2.40a,b}$$

By substituting x_{25} as given by Eq. (2.36) into $K_4=0$, we are led to a sixth-degree equation in x_{13}, and consequently the system (2.40a,b) has six points (x_{13},x_{25}) as solutions. At each such point $\Delta=0$ or $K_{24}=0$. At two of these points $K_{24}=0$, and at the remaining points necessarily $\Delta=0$, so that α and γ are tangent there. Of course, a consideration of the system

$$B_{24}=0, \quad \Delta=0 \tag{2.41a,b}$$

leads to analogous conclusions.

From these considerations we conclude that if there are four real tangent points between α and γ, then there are no tangencies between the complex surfaces σ_α and σ_γ. This is, e.g., the case in Fig. 5. We shall not discuss further the general behavior of such complex tangencies, but it can be helpful to know when they do not occur.

3. GENERAL REMARKS ON THE ANALYTICITY OF M_5

We know from Theorem 1.1 that all the singularities of M_5 are restricted to the singularity curves and surfaces, which we discussed in Sec. 2. We must still determine, however, which parts of these curves and surfaces are regular and which are singular. Since M_5 is a multivalued function and its singularities are not necessarily the same on different Riemann sheets, we must consider how to construct a suitable definition of the physical sheet of M_5. If such a definition is given, then one is concerned usually, but not always,[20] with singularities on the boundary of this physical sheet.

In this section we shall give a general discussion of these problems.

A. Review of Basic Methods

For the convenience of the reader we will give a brief summary of some of the methods and results which have been used in discussing singularities of Feynman amplitudes. Further discussion can be found in I, especially in the Appendices, and also in references 2, 15, 18, and 19.

We first present two integral expressions for the amplitude M_5. In terms of Feynman variables, M_5 is given by

$$M_5=\int_0^1 d\alpha_1\cdots\int_0^1 d\alpha_5\frac{F_0(\alpha_1,\cdots,\alpha_5)\delta(1-\sum\alpha_i)}{D_0^3}, \quad (3.1)$$

where F_0 is a polynomial which includes the spin dependence of M_5, and where

$$D_0=\sum_{i=1}^5\alpha_i m_i^2-\sum_{i<j}\alpha_i\alpha_j p_{ij}^2. \quad (3.2)$$

It is convenient to make the following change of variables:

$$u_k=\alpha_k m_k/\sum_{j=i}^5\alpha_j m_j, \quad (3.3)$$

and we obtain

$$M_5=\int_0^1 du_1\cdots\int_0^1 du_5\frac{F(u_1,\cdots,u_5)\delta(1-\sum u_i)}{D^3}. \quad (3.4)$$

The Jacobian of the transformation is included in the function F, which is again a polynomial. The denominator now reads

$$D=\sum_{i=1}^5 u_i^2-2\sum_{i<j}u_iu_jx_{ij}=-\sum_{i,j}u_iu_jx_{ij}. \quad (3.5)$$

Next we list the methods which can be used for the determination of singularities of integrals such as in Eq. (3.4). Some of these methods will be discussed more fully in the remainder of Sec. 3. We start with the necessary and sufficient conditions for the existence of a singularity.

(1) A necessary condition for the function

$$\omega(\zeta)=\int_A f(z,\zeta)dz \quad (3.6)$$

to be singular at $\zeta=\zeta_0$ is that $f(z,\zeta_0)$, as a function of z, be singular at an endpoint of the contour A, or that $f(z,\zeta)$ have two singularities, one on each side of the contour, which tend to coincidence as $\zeta\to\zeta_0$. (This latter case is sometimes called the case of pinching singularities.)

(2) Ordinarily, and in particular for M_5, the fulfillment of one of the two conditions in (1) is sufficient for the existence of a singularity.

(3) Singularities of M_5 on the curve or surface $K_{i\cdots j}=0$ are such that there are singularities of the integrands at the endpoints $u_i=\cdots=u_j=0$, while the other integrations give coincident singularities.

We next describe the methods for analytic continuation of integral expressions.

(4) If analytic continuation of an integral such as in Eq. (3.4) is to be carried out, one has to deform the contours so that no singularity of the integrand crosses the contours.

In particular, if we move along a singularity surface, then certain singularities in the successive integrations stay coincident. These can be separated from the contour, or can be made to pinch the contour only through an endpoint ($u_l=0$), and we have the following conclusion [see also Sec. 3(C)]:

(5) The singularity curve or surface $K_{i\cdots j}=0$ can change its singular nature only at a tangency with the singular curve or surface $K_{i\cdots jl}=0$. (In special cases the tangency becomes an intersection.) On the other hand, intersections with the surface $K_{i\cdots jlm}=0$, or more generally $K_{k\cdots m}=0$, are not relevant here, but one has to be careful not to cross a branch cut.

(6) In general, the singular nature of the curve or surface $K_{i\cdots jl}=0$ will change upon continuation around the branch cuts at $K_{i\cdots j}=0$.

(7) If along a path of continuation of M_5, $D\neq 0$ for $0\leq u_i\leq 1$, $\Sigma u_i=1$, then no deformation of the contour is necessary, and M_5 remains analytic.

(8) The singular nature of the curve or surface $K_{i\cdots j}=0$ depends only on the continuation of the integrations over $0\leq u_l\leq 1$, $\Sigma u_l=1$, with $l\neq i,\ \cdots,\ j$, and does not depend on the chosen contours for integrals over $u_i,\ \cdots,\ u_j$.

The following assertion describes a very useful condition for the existence of a singularity.

(9) Let $D\neq 0$ when $0\leq u_i\leq 1$, $\Sigma u_i=1$, if the x_{ij} are restricted to a real region R. Let $D=0$ somewhere in the region of integration when $x_{ij}=x_{ij}^0$, the point (x_{ij}^0) lying on the boundary of R. Then M_5 is singular at (x_{ij}^0).

[20] For a discussion of the continuations to unphysical sheets see, e.g., R. Blankenbecler, M. L. Goldberger, S. W. MacDowell, and S. B. Treiman, Phys. Rev. 123, 692 (1961).

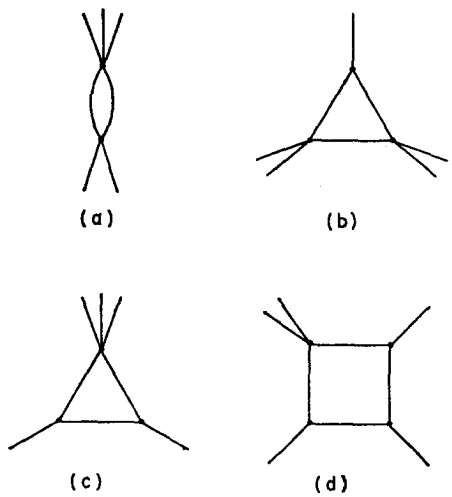


FIG. 7. Production diagrams which can be obtained from the single-loop diagram by contraction.

It is convenient to make here a few remarks concerning the contraction of Feynman diagrams [see also Sec. 2(A)]. Let us start by considering the two conditions of assertion (1). The presence of end-point singularities in some of the successive integrations for M_5 means that the singularity in question is also the leading singularity of a suitably contracted diagram; this is contained in the assertion (8). The contracted diagrams which are relevant for us are shown in Fig. 7, and they provide a convenient basis for the classification of singularities of M_5, as follows: normal thresholds, vertex singularities, scattering singularities, and five-point singularities or five-point poles. In I these singularities are classified as singularities of types $(E^{4-k}C^k)$, $k=1, \cdots, 4$, respectively. Here k is the number of integrations with pinching singularities, and there are also $4-k$ end-point singularities. We shall also use this notation occasionally in the sequel. Furthermore, we shall use geometrical shapes on figures to indicate, in an obvious way, the various kinds of singularities.

The quantity which takes the place of an external mass in a contracted diagram such as in Fig. 7 is the total center-of-mass energy of all the particles meeting at an external vertex. It is obvious that in general the stability conditions are violated in these diagrams. We therefore discuss the lower-order diagrams with unstable external particles in the Appendix, and we also review there, for convenience, those conditions for singularities which have been obtained in references 7, 12, 13, and 14.

We conclude this section with a discussion of the types of singularities associated with the various curves and surfaces. It follows from Landau's analysis[19] that if a given singularity is the leading singularity of a contracted diagram, then its type does not depend on the original diagram. Thus for normal thresholds, and vertex and scattering singularities, the singularities are such as given in I. Also, the five-point singularities are simple poles.[2] We summarize these considerations below. The notation is easily understandable, and z is any convenient variable which vanishes at the singularity:

$$\text{at normal thresholds,} \quad M_5 \sim z^{\frac{1}{2}}, \tag{3.7a}$$

$$\text{at vertex singularities,} \quad M_5 \sim \log z, \tag{3.7b}$$

$$\text{at scattering singularities,} \quad M_5 \sim z^{-\frac{1}{2}}, \tag{3.7c}$$

$$\text{at five-point singularities,} \quad M_5 \sim z^{-1}. \tag{3.7d}$$

It is significant that all of the singularities other than the five-point singularities are branch points.

B. Physical Regions and the Physical Sheet

The amplitude M_5 is a multivalued function, as we have pointed out before, and one would like to be able to describe the various possibilities of analytic continuation in an orderly manner. A natural starting point is a physical region, in which M_5 is to be considered as an amplitude which governs a physical process in some approximation.

A physical region is a region in which all of the momenta p_{ij} are real. The amplitude which describes the physical process in question is determined by Feynman's prescription: We replace each internal mass m^2 by $m^2-i\epsilon$, and pass to the limit $\epsilon \to 0+$. We see from the structure of the denominator D_0, Eq. (3.2), that this is equivalent to replacing p_{ij}^2 by $p_{ij}^2+i\epsilon$, or replacing x_{ij} by $x_{ij}+i\epsilon$, and passing to the same limit. In the sequel we will refer to such limits as *corresponding limits*. (Note that in I this term is used in a somewhat different way.) Of course, an amplitude in general has associated with it a number of physical regions, corresponding to the various processes which it describes.

An important property of an amplitude in a physical region is given by the following theorem due to Eden[15]:

Theorem 3.1: The only singularities of a scattering or production amplitude in a physical region (and evaluated in the specified limits) are the normal thresholds.

This theorem is valid to all orders of perturbation theory. While Eden's proof is given with reference to the scattering amplitudes, it applies equally well to production processes.

With reference to the previous remarks, we now wish to give a brief discussion of the physical sheet. In general, the physical sheet is a complex region into which an amplitude can be continued analytically from a physical region without becoming multivalued. Conventionally, the physical sheet is delimited by cuts which contain the physical regions. It is particularly desirable in applications to have the physical sheet connect the different physical regions which are associated with the amplitude in question. For con-

venience in discussing the physical sheet, we shall think of the five invariants other than the masses as complex variables.

We first note that M_5 is analytic if these invariants x_{ij} have nonzero imaginary parts of the same sign, and if the external masses satisfy the stability conditions, Eq. (2.4). To see this, we replace the quantity $x_{ij}+i\epsilon$ for each invariant x_{ij} in D [Eq. (3.5)] and then,

$$\mathrm{Im}D=-2\epsilon(u_1u_3+u_1u_4+\cdots+u_3u_5). \tag{3.8}$$

If we use the contours $0\leq u_i\leq 1$, $\Sigma u_i=1$, as in the determination of the amplitudes for physical processes, then, in order to have $\mathrm{Im}D=0$, we must have $u_i=u_{i+1}=u_{i+2}=0$ for some i, but in this case $\mathrm{Re}D>0$ because of the stability condition $x_{i+3,i+4}<1$. An entirely analogous argument shows that there is a region R in the space of the real invariants where M_5 is analytic and real, aside from spin factors. The existence of the region R implies that M_5 is a real function, i.e., that it takes complex conjugate values at complex conjugate points, aside from spin factors. [We shall discuss this region R more fully in Sec. 3(D)].

We now conclude that M_5 can be continued in a natural way from the physical regions to the complex region where $\mathrm{Im}x_{ij}>0$ (for the invariants), to the real region R, and from R to the complex region where $\mathrm{Im}x_{ij}<0$. Accordingly, we shall impose the following conditions on the physical sheet of M_5: The physical sheet must contain the specified regions, and M_5 is to be analytic and real in the region R, aside from spin factors. The physical sheet is to be limited by the branch cuts $x_{ij}>1$, i.e., the physical cuts (and also by other branch cuts).

The possibility of analytic continuation of M_5 to the specified regions is a great help in studying the analytic properties. However, these considerations cannot be easily extended so as to lead to an adequate definition of the physical sheet, for these reasons:

First, M_5 has complex singularities if the invariants have imaginary parts of different signs, and great care must be used in describing the analytic continuation to these regions. Second, other diagrams which contribute to the same processes as M_5 have singularities in the region $\mathrm{Im}x_{ij}>0$ and in the region $\mathrm{Im}x_{ij}<0$, and this leads to further complications if one is interested in the entire amplitude. [For a discussion of the analogous problem for scattering amplitudes, see reference 15, Sec. 2II(A).]

C. Behavior of M_5 Near Singular Tangent Points

As we pointed out in Sec. 3(A), the singular nature of the singularity curve $\Delta=0$ (i.e., α), or $K_{i\ldots j}=0$, can change only at a point of tangency to a curve $K_l=0$, or $K_{i\ldots jl}=0$, respectively. Further, at such a point of tangency the singular nature of the curve, in general, will change, as the following theorem shows. This theorem was suggested by an analysis of scattering amplitudes (see the end of Sec. 3 of I, and, for a more general analysis, Theorem 7B of reference 15).

Theorem 3.2: Let α be tangent to the curve $K_i=0$, which is singular in a neighborhood of the point of tangency. Then α is regular on one side of the point of tangency and singular on the other side. An analogous conclusion holds for the (real) curve $K_{i\ldots j}=0$ which is tangent to a singular section of $K_{i\ldots jl}=0$. (We have to approach α, or the curve $K_{i\ldots j}=0$, on both sides of the point of tangency from the same complex region. We also have to exclude certain limiting cases, as usual. We are restricting ourselves to real curves, since we have not attempted to study complex tangencies in this paper.)

Proof: This theorem is based on the fact that those coincident singularities which lead to singularities of α, or of the curve $K_{i\ldots j}=0$, either fall off the contour or else approach the contour from opposite sides at the point of tangency. We shall now give a detailed demonstration of this fact.

For definiteness let us consider the case where α is tangent to γ. Let us assume that we have performed all of the integrations for M_5 except that over u_4, so that

$$M_5=\int_0^1 du_4 f(u_4; x_{ij}). \tag{3.9}$$

We shall consider those two singularities, u_4' and u_4'', of f which arise out of coincident singularities in each of the previous integrations, i.e., of type (C^3). If the quantities x_{ij} define a singular point of γ, then $u_4'=0$ or $u_4''=0$. At the point of tangency, necessarily $u_4'=u_4''=0$. We conclude that if a point is slightly separated from the point of tangency, then the singularities of u_4' and u_4'' are not displaced to another sheet of f, but are near the end point $u_4=0$. (Excep tin limiting cases, there are no other singularities in the u_4 plane near $u_4=0$.) In particular, at a point of α, we have $u_4'=u_4''=u_4^0$.

It is easy to show that as long as the curves are real, the points $u_j=u_j^0$, which define the coincident singularities, are also real. The quantities u_j^0 satisfy the following equation given by Fowler *et al.*[7] [Eq. (10)].

$$u_i^0/u_l^0=B_{ik}/B_{lk}, \tag{3.10a}$$

where k is arbitrary. In particular,

$$u_4^0/u_2^0=K_4/B_{24}. \tag{3.10b}$$

Now, as we move along α through the point of tangency, K_4 goes to zero but does not change sign [see Sec. 2(E)], B_{24} goes through zero and changes sign, and $u_2^0\neq 0$ except in limiting cases (i.e., on the lines $K_2=0$). We conclude that u_4^0 changes sign. To complete the argument we must still show that the singularities u_4', u_4'' of f do not drag the contour with them in such a way as to nullify the effect of going through $u_4^0=0$.

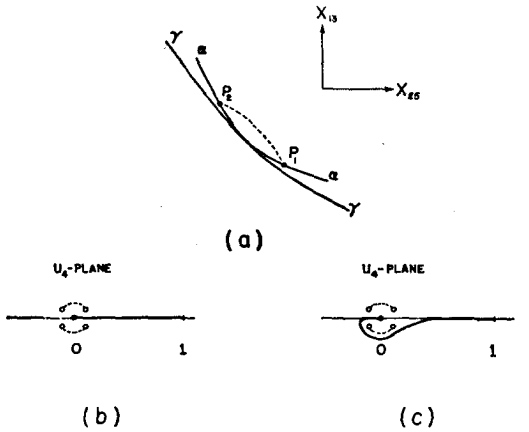


FIG. 8. (a) Continuation in the (x_{13},x_{25}) plane near a point of tangency of α to a singular branch of γ. (b) and (c); Two alternatives for the resulting displacement of singularities of f [Eq. (3.9)] in the u_4 plane.

If we are not a a point of α, but the x_{ij} are real, then $u_4' \neq u_4''$, and either u_4' and u_4'' are both real or they are complex conjugates. (cf. the remark that the u_j^0 are all real.) If we make a continuation such as shown in Fig. 8(a), then the first alternative is not possible, since $u_4'=u_4''<0$ at P_1 and $u_4'=u_4''>0$ at P_2, or vice versa, and u_4', $u_4'' \neq 0$ as long as we do not reach γ. We now conclude that u_4', u_4'' are displaced as shown on Fig. 8(b) or 8(c). The theorem follows.

We remark that, as in the case of scattering amplitudes, it is possible for the curve $K_i=0$ to be regular near the point of tangency, and for α to be singular there. See I, end of Sec. 3.

There is one further characteristic of the behavior of M_5 near such a point of tangency, and this has to do with analytic continuation around the branch points on the singular tangent curve. Of course, such a continuation may lead us out of the physical sheet, but it may be of interest even in these cases.

Theorem 3.3: Let α be tangent to a singular section of the surface $K_i=0$. Let P be a point of α sufficiently near the point of tangency. Let M_5 be singular (or regular) if P is reached by a given continuation. Then M_5 will be regular (or singular) if we continue it around the branch points of $K_i=0$ and back to P. As before, the point of tangency is assumed to be away from all other singularity curves $K_j=0$, $K_{ij}=0$, etc. An analogous conclusion holds for the curve $K_{k\ldots l}=0$, in place of α.

We assume here less than for Theorem 3.2. For instance, Theorem 3.2 does not apply to the case where two points of tangency between α and γ merge, while Theorem 3.3 does apply. Furthermore, we do not need to assume reality of curves, and this theorem is also applicable to a continuation around a complex branch cut. We assume that P is sufficiently near the point of tangency in order to avoid complications with other singularities.

This theorem is an elaboration of statement (6) of Sec. 3(A), and is analogous to Lemma 2 of I, Sec. 3. (However, the proof of that lemma as given in I is incomplete.) This theorem depends on the fact that the function f of Eq. (3.9) has singularities at $u_4=u_4'$, u_4'' which pinch the contour if M_5 is singular at P, but if, say, u_4' is continued around the endpoint $u_4=0$, the singularities no longer pinch.

Proof: For definiteness, let us assume tangency between α and γ, as in the proof of Theorem 3.2, and let us carry out the continuation around the branch points by displacing x_{13}. We will also use the form (3.9) for M_5. As previously, the function f has two singularities of type (C^3). The equation which yields these singularities can be obtained by computing successive discriminants, as described in I, Appendix B. We will write this equation as

$$A(x_{13}; x_{ij})u_4^2+B(x_{13}; x_{ij})u_4+C(x_{13}; x_{ij})=0. \quad (3.11)$$

Let us now consider the continuation of M_5. We continue from α to a neighborhood of a point of γ, having, say, $x_{13}=x_{13}^0$. This point is to be near the tangent point, but not the tangent point itself. By considering the types of singularities involved we conclude that $u_4=0$ is one (and only one) solution of Eq. (3.11) for $x_{13}=x_{13}^0$, and one can extend this solution to a neighborhood of $x_{13}=x_{13}^0$ by a power series:

$$u_4=a_1(x_{ij})(x_{13}-x_{13}^0)+a_2(x_{ij})(x_{13}-x_{13}^0)^2+\cdots. \quad (3.12)$$

We will show below that $a_1(x_{ij}) \neq 0$, and this immediately gives us the theorem. For, to one rotation of x_{13} around x_{13}^0 then corresponds one rotation of u_4 around zero, and the continuation of M_5 once around a branch point leads to a displacement of one singularity once around the endpoint. We now conclude as in Appendix A of I that the singular nature of M_5 will be changed when we continue back to α.

We still have to show that $a_1 \neq 0$. We first note that we get C of Eq. (3.11) by calculating the successive discriminants with the condition $u_4=0$, and therefore we can use Lemma 1B of I:

$$C=(\text{const})A_{j_1}{}^3A_{j_2}{}^2A_{j_3}K_4. \quad (3.13)$$

Here the A_{jk} are the leading coefficients in the successive discriminants. These are nonzero except in special cases, in which various other coefficients vanish as well. These cases can therefore be ignored, and the conditions $C=0$ and $K_4=0$ are equivalent. Now, if $a_1=0$, then Eq. (3.12) as an equation in x_{13} has a double root for $u_4=0$, and the same holds for Eq. (3.11). This implies a double root of $C=0$, or of $K_4=0$. This in turn implies $K_{j4}=0$ for some j, and our hypothesis is contradicted. The proof is complete.

D. Real Region of Analyticity

Real regions of analyticity, in which amplitudes are real (except for spin factors), are prominent in the

study of scattering amplitudes. These regions are also likely to be prominent in the study of production amplitudes. Moreover, as we pointed out in the Introduction, these regions provide a convenient starting point for the continuation procedure.

The amplitude M_5 is real and analytic if the quantities x_{ij} are real and if $D\neq 0$ throughout the region of integration. [cf. Eqs. (3.4, 5). The integration is to be carried out over the contours $0\leq u_i\leq 1$, $\Sigma u_i=1$, as in Sec. 3(B).] If D vanishes somewhere in the region of integration, then in general M_5 will have a nonzero imaginary part. The points of the boundary of the real region are singularities of M_5.

A necessary condition for D to be different from zero throughout the region of integration (actually, for $D>0$) is that every invariant x_{ij} be below its physical threshold: $x_{ij}<1$. Moreover, these conditions, taken with reference to the external masses, are also sufficient for the existence of a real region of analyticity of M_5.

Theorem 3.4: If the external masses $x_{i,i+1}$ satisfy the stability conditions, $x_{i,i+1}<1$, then M_5 is real and analytic provided the other invariants, $x_{i,i+2}$ and $x_{i,i+3}$, are less than or equal to -1.

Proof: We rearrange the terms in D as follows:

$$\begin{aligned}D=&2[u_1u_2(1-x_{12})+\cdots+u_4u_5(1-x_{45})+u_1u_5(1-x_{15})]\\&-2[u_1u_3(1+x_{13})+u_1u_4(1+x_{14})+\cdots+u_3u_5(1+x_{35})]\\&+(u_1-u_2+u_3-u_4+u_5)^2\\&+4(u_1u_4+u_2u_5-u_1u_5).\end{aligned}\tag{3.14}$$

Each of the first two terms, i.e., in the brackets, is positive or zero. The first term can vanish only if three of the u_i are zero, say u_1, u_3, and u_4. Next, the last term can be negative only if

$$u_1-u_2=\delta>0,\quad u_5-u_4=\delta'>0.\tag{3.15}$$

But then the last two terms give

$$(\delta+\delta'+u_3)^2+4u_2u_4-4\delta\delta'.\tag{3.16}$$

This expression is non-negative since $(\delta+\delta')^2-4\delta\delta'\geq 0$. Now, $D>0$ for, say, $u_1=u_3=u_4=0$ and $x_{25}\leq 1$, and the theorem follows.

Let us now consider the region R in the real (x_{13},x_{25}) plane where M_5 is real and analytic. As in the theorem, we impose stability conditions on the external masses. Then, if the other parameters satisfy $x_{ik}\leq -1$, the region R includes the region

$$R_0=\{x_{13}\leq -1,\ x_{25}\leq -1\}.\tag{3.17}$$

Let us for the moment restrict our attention to R_0, and let us continue M_5 analytically in the parameters other than the masses. The parameters are restricted to be real. Therefore our continuation will be limited by the normal thresholds $x_{ik}=1$, and it may be limited also by the vertex singularities $x_{ik}=-\cos(\theta_{ij}+\theta_{jk})$. On the other hand, an examination of the scattering singularities shows that, as long as $(x_{13},x_{25})\in R_0$, these singularities do not hinder the continuation. (See Appendix A.)

We still have to investigate the possibility that the five-point poles delimit the continuation. This is of interest, even though M_5 does not acquire an imaginary part on account of these poles. To study this question we follow Appendix B of reference 13. The poles in question arise when D vanishes inside the region of integration, but not on any boundary. The necessary conditions for this are, in analogy with Eqs. (B3a–c) of reference 13,

$$K_{ijk}>0,\quad -K_{ij}>0,\quad K_i>0,\tag{3.18a–c}$$

$$-\Delta<0.\tag{3.18d}$$

(Our choice of the matrix entries x_{ij} requires that the determinants of odd rank be multiplied by -1.) These inequalities show that the five-point poles do not delimit the continuation if $(x_{13},x_{25})\in R_0$. We summarize our conclusion in a lemma.

Lemma 3.5: A necessary and sufficient condition for the existence of a region of analyticity in the real (x_{13},x_{25}) plane, where M_5 is real, is that the parameters x_{lm} satisfy $x_{ik}<x_{ik}^0(x_{ij},x_{jk})$ for any triple (i,j,k). Here $x_{ik}^0=-\cos(\theta_{ij}+\theta_{jk})$ if this point is singular, otherwise $x_{ik}^0=1$. This region of analyticity, if it exists, includes R_0.

Let us consider further the problem of determining the boundaries of the region of analyticity. For this purpose we may start with R_0, and continue in x_{13} and x_{25} until we come upon a singularity. The vertex and scattering singularities can be readily determined (Appendix A). The condition for the presence of the five-point poles on the boundary of the region of analyticity can be obtained as follows. The inequalities (3.18a–d) show that such a boundary is possible only if there is a central oval of α, analogous to γ_5. The analysis of Sec. 3(C) then can be used to determine which parts of the central oval are singular and which

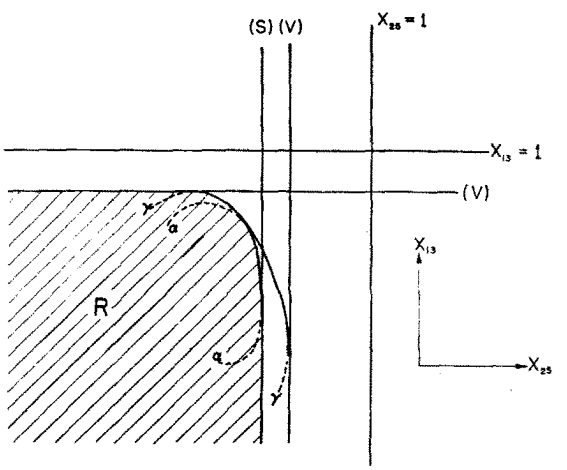


FIG. 9. A region R which is bounded by sections of the following curves: α, γ, a tangent line to α, and a tangent line to γ.

are regular. [It can be helpful to use the fact that, if (x_{13},x_{25}) belongs to the region, then so does $(x_{13}+\delta_1, x_{25}+\delta_2)$ for any $\delta_1, \delta_2 \geq 0$.] We remark in conclusion that here a more diversified type of boundary is possible than in the case of scattering curves. Figure 9 shows a boundary that consists of α, of γ, of a tangent line to α, and of a tangent line to γ.

We note in conclusion that Wu[10] has considered the real region of analyticity for certain production amplitudes which include the contribution of crossed diagrams.

E. Discussion of Methods for Determining Singularities

In Sec. 3(A) we listed some general methods which are basic for the determination of singularities, and in Secs. 3(B–D) we discussed some properties of M_5 which can be conveniently used for the same purpose. In this section we shall describe, in a general way, how these methods and properties can be combined. We obtain in this way some methods which are quite useful in the study of M_5. The actual applications of these methods to specific problems form the contents of Sec. 4.

(1) The first method which we mention depends on the continuation of M_5 from the real region of analyticity, R, along the complex singularity surfaces. The function M_5 can be continued in this way to those parts of singularity curves which lie on the cuts, i.e., on the boundary of the physical sheet. The information so obtained can be supplemented by means of the discussion of Sec. 3(C). This method was used extensively in I, and is also of some use in the case of M_5. However, its use in the case of M_5 is limited by the common occurrence of complex singularities, and also by the common nonexistence of the region R in the (x_{13},x_{25}) plane.

(2) The second method depends on the fact that M_5 is analytic if the invariants (other than the masses) satisfy $\mathrm{Im}x_{ij}>0$. Now, if along an arc of γ or α the condition $dx_{13}/dx_{25}>0$ is fulfilled, then the complex extension has imaginary parts of the same sign. We easily conclude that the arc in question is regular if the limits $\mathrm{Im}x_{ij} \to 0+$, i.e., corresponding limits, are taken for all of the invariants. This information can be again supplemented by the discussion of Sec. 3(C).

We should point out that the method just described is not always adequate, even for the corresponding limits, since there may be branches of α or of γ where $dx_{13}/dx_{25}<0$ everywhere. Furthermore, the method just described cannot be easily extended to the case of other limits. However, the singularities which occur for corresponding limits are probably the more significant ones, as far as the experimental effects are concerned.

It is of interest that this method can sometimes be used to establish tangency between α and γ. Indeed, a closed arc of α must necessarily have an even number of tangencies to those scattering curves which are singular in the corresponding limits.

(3) The third method is perhaps the most important one of the three, since it is of widest applicability. It is the method of the variation of parameters, or, as we shall sometimes call it, the continuation procedure. We start with certain values of the parameters, for which the configuration of singularities is understood in detail. Then, as we vary a parameter slightly, certain changes occur, but large parts of the singularity curves do not change in a significant way, and we can be sure of singularities in these parts. Of course, this procedure amounts simply to continuing M_5 along a singularity curve, but with respect to other variables, or, to continuing M_5 along a singularity surface in a three-dimensional real region.

We should emphasize that we shall apply this method only to the real region. While this method is applicable in principle to the complex region as well, it requires as a first step a detailed description of complex singularity surfaces, their singular sections, their tangencies, and their associated branch cuts. Such a description would be quite involved, and for this reason our investigation of complex singularities is very limited.

The fact that we are restricting ourselves to the real region has certain shortcomings, namely, that the method is not applicable if the variation of a parameter leads to the appearance of a new branch of a curve, or to a change of configuration at infinity. In such cases special methods have to be used. Again, the discussion of Sec. 3(C) is often helpful.

Another shortcoming arises in connection with our inadequate familiarity of the curve β. In particular, we cannot follow the points of tangency between α and γ in all cases. However, the complications which arise in the example of Sec. 4(D) can be easily resolved, as we shall see.

4. SINGULARITIES FOR SPECIFIC PROCESSES

In this section we will determine the singularities of M_5 for several specific processes. We will make extensive use of the general considerations of the last two sections, and in particular, we will apply the methods of Sec. 3(E) to specific examples. Method (1) will be used in Sec. 4(B), method (2) in Sec. 4(C), and method (3) in Secs. 4(D) and (E).

A. Description of the Processes

The processes for which we shall determine the singularities of M_5 are the following:

(a) $\pi+N \to N+\pi+\pi$,

(b) $p+d \to \mathrm{He}^3+\pi+\pi$,

(c) $\pi+d \to d+\pi+\pi$,

(d) $\bar{d}+d \to \pi+\pi+\pi$.

Our reasons for selecting these particular production

processes have been given in reference 1, and they can be summarized as follows: Process (a) and (b) have been investigated in recent experiments,[21,22] and they are of interest in the study of the π-π interaction. Process (c) was suggested for a possible experimental investigation of the five-point poles, since there are no single-particle poles in the amplitude. Process (d) is of interest in connection with the deuteron form factor.

As we stated in the Introduction, we consider for each process only one single-loop diagram and also a single configuration of momenta. The respective diagrams are shown in Fig. 10. The configurations of momenta are defined by the following equations, in which we use the labeling of momenta shown in Fig. 1:

$$p_{12}=(E_{12},q,0,0); \qquad -p_{34}=(E_{34},0,-\tfrac{1}{2}p,p);$$
$$p_{23}=(E_{23},-q,0,0); \qquad -p_{45}=(E_{45},0,-\tfrac{1}{2}p,-p);$$
$$p_{15}=(E_{15},0,p,0). \tag{4.1}$$

These configurations correspond to momenta p_{12} and p_{23} incoming, and the others, outgoing. The momenta p_{12} and p_{15} refer to the heaviest particles, except in process (d). The invariants p_{ij}^2, Eq. (1.1), are determined by Eqs. (4.1) and the incident laboratory energies, which were given the following values: for process (a), $T_\pi=670$ Mev; for process (b), $T_p=648$ Mev; for process (c), $T_\pi=650$ Mev; and for process (d), $T_d=500$ Mev. [For processes (a) and (b), these are the experimental energies.]

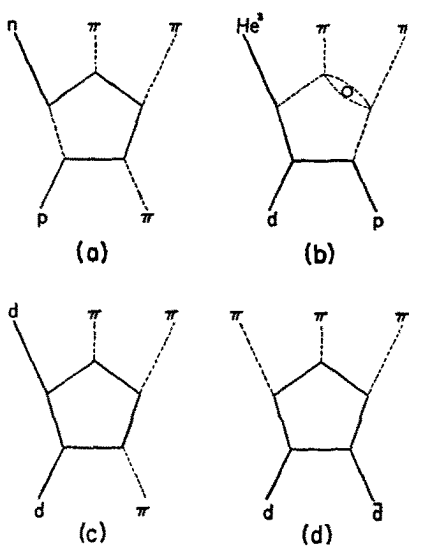


FIG. 10. Single-loop diagrams for the processes (a)–(d). These diagrams correspond to the amplitudes which were studied in detail. In diagram (b) we assume that two of the internal pion lines are in resonance and behave as a single particle of mass $3.5m_\pi$. (This diagram has been suggested by R. J. Eden, *Proceedings of the 1960 Annual International Conference on High-Energy Physics at Rochester*, p. 219.)

[21] B. C. Barish, R. J. Kurz, P. G. McManigal, V. Perez-Mendez, and J. Solomon, Phys. Rev. Letters **6**, 297 (1961); J. A. Anderson, V. X. Bang, P. G. Burke, D. D. Carmony; N. Schmitz, Phys. Rev. Letters **6**, 365 (1961).

[22] A. Abashian, N. E. Booth, and K. M. Crowe, Phys. Rev. Letters **5**, 258 (1960); **7**, 35 (1961).

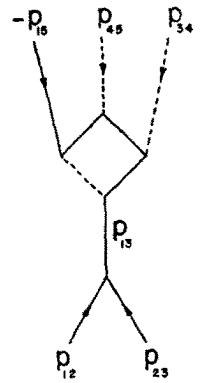


FIG. 11. A Feynman diagram for process (a). Except for the single-particle pole, the singularities of the corresponding amplitude are included among the singularities of M_5.

The values of the quantities x_{ij} which correspond to the invariants p_{ij}^2 [Eq. (1.2a)] in the processes (a)–(d) are given in Table I, columns (a)–(d), respectively. [The quantities x_{13} and x_{25} are our variables, but it is still of interest to know their actual values as given by Eq. (4.1).] Column (e) of this table gives another set of values for the quantities x_{ij} other than x_{13} and x_{25}, i.e., for the parameters. This set was chosen as the starting point for the continuation procedure. We shall refer to this set of values as the *initial process*. Of course, this set of values was not intended to correspond to any physical process, but our use of the word *process* for this set of values is analogous to our use of the word elsewhere, and is convenient.

To conclude this section we should like to discuss briefly two kinds of diagrams which we do not analyze in detail. The first of these consists of those fifth-order diagrams which exhibit single-particle intermediate states. These diagrams have singularities which either are poles corresponding to single-particle intermediate states, or else are actually included among the singularities of the single-loop diagrams. For example, the diagram of Fig. 11 has a pole at $p_{13}^2=m_N^2$, and has those singularities of M_5 which correspond to the internal momentum 2 contracted (i.e., those singularities which lie on one of the hypersurfaces $K_2=0$, $K_{2i}=0$, and $K_{2ij}=0$).

The second kind of diagrams consists of single-loop diagrams with crossed lines. We note here only that

TABLE I. Values of the quantities x_{ij}.

	Process				
x_{ij}	(a)	(b)	(c)	(d)	(e)
x_{12}	−0.075	−0.995	+0.995	+0.995	−0.1
x_{23}	−0.989	−0.075	−0.989	+0.995	−0.9
x_{34}	−0.989	−1.750	−0.989	−0.989	+0.6
x_{45}	−0.989	−1.750	−0.989	−0.989	+0.8
x_{15}	−0.075	−0.025	+0.995	−0.989	−0.2
x_{14}	+1.676	+0.351	+1.737	+2.889	+0.5
x_{24}	−1.145	−3.171	−1.196	−2.118	+0.3
x_{35}	−0.835	+2.900	−0.797	+2.460	−0.6
x_{13}	+5.868	+3.172	+2.691	+8.041	
x_{25}	−1.156	−3.277	−1.098	−1.792	

our method can be applied equally well to these diagrams.

B. Singularities of M_5 for the Initial Process

We shall now describe the singularity curves for the initial process, and determine their regular and singular parts. To begin with, we recall from the end of Sec. 2(C) that each of the singularity curves falls into one of these two classes: (1) the curve γ with its tangent lines and asymptotes, and (2) the curve α with its tangent lines and asymptotes. For the curves in class (1) the description given in Sec. 2(B) is adequate. We shall therefore proceed with an examination of the curves in class (2). In describing these curves we shall follow the discussion of Sec. 2(C). We first consider the asymptotes, which are given by the equations $K_{13}=0$ and $K_{25}=0$ [see Eqs. (2.4–11)]. We see from Table I that these asymptotes are all real, and that the values which determine them lie between -1 and 1.

We consider next the tangent lines, and these are given by the equations $K_i=0$ for $i=1, 2, 3$, and 5. Let us interpret these equations as curves in various planes, as in Fig. 3. These curves determine the tangent lines as intersections, and the chosen set of parameters has the property that all of these intersections are on the central ovals. Furthermore, the equations $K_{13}=0$, $K_{25}=0$ define tangent lines to the curves $K_i=0$, and in each case the oval lies between the two tangent lines in question. We conclude that the tangent lines to α all lie between the asymptotes to α. The configuration of α is therefore quite similar to that given in Fig. 3(a). The configuration of α for the initial process is shown in Fig. 5, on which some distortions were made. This figure was given in Sec. 2(E) to illustrate the use of the curve β.

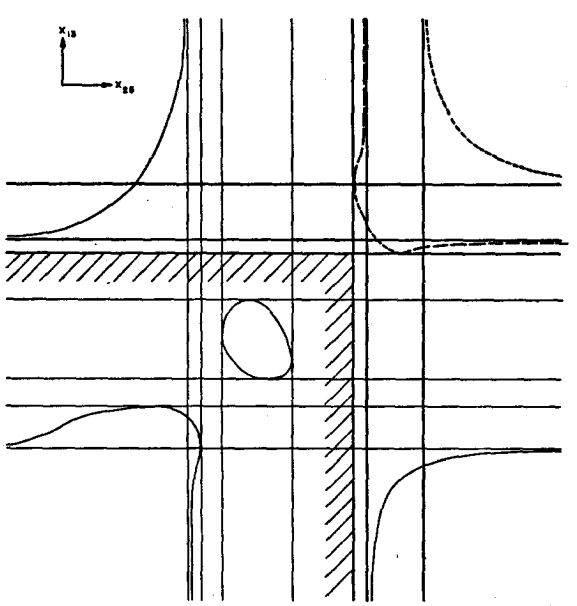


FIG. 12. The singularities for the initial process. Lines which are singular for all limits of x_{13} and x_{25} are shown as heavy solid. Curves which are singular for corresponding limits are shown with short dashes, while those which are singular for opposite limits are shown with long dashes. The regular parts of α and its regular tangents and asymptotes are also shown for orientation. These are shown as light solid.

The one remaining aspect of the configuration which still has to be discussed is the position of the tangencies, both to the tangent lines and to γ. Here we relied on a numerical determination[23] of the curves α and β (and also, for convenience, of γ) to obtain the positions of the tangent points. However, the analytic properties of the singularity curves can also give some information about these positions; see below.

We are now in a position to discuss the singular and the regular parts of the singularity curves. These singular parts are shown in Fig. 12. We have also included there for orientation the regular parts of α, and all of its tangent lines and asymptotes, even if regular. To establish the facts shown on this figure, we begin with the singularities arising from diagrams of order four or less. For the curves in class (1), we find that we have just those singularities which arise under case (i) for scattering diagrams (see Appendix A). For the curves in class (2), we find that two of the asymptotes and two of the tangent lines are singular.

We conclude from Sec. 3(D) and Fig. 12 that M_5 is real and analytic in the region

$$R=\{x_{13}<x_{13}^0,\ x_{25}<x_{25}^0\}, \tag{4.2}$$

where $x_{13}=x_{13}^0$ and $x_{25}=x_{25}^0$ are the two singular tangent lines. We also see that the configuration of α and of its regular and singular tangent lines and asymptotes is entirely analogous to that of case (ii) for scattering diagrams. We may therefore apply the same arguments which were used in case (ii) of I in order to draw the conclusions which are displayed in Fig. 12. [These arguments are summarized as method (1) in Sec. 3(E). However, method (2) together with Sec. 3(C) lead to these conclusions even more directly.]

It is worthwhile to make a few observations about the configuration of Fig. 12. The most striking fact is that no singularities of M_5 extend into the complex region, so that we have a Mandelstam-type representation. Further, the actual arrangement of tangencies of α to the singular tangent lines and to the singular branch γ_1 is the only arrangement that is consistent with the general considerations of Sec. 3.

The last observation which we shall make does not directly concern the situation here, but still is of some interest. The fact that there are no complex singularities arising from diagrams of order four or less implies that the singular nature of the surface σ_α cannot change in the complex region. But even if there were complex singularities, e.g., we could continue M_5 to an unphysical sheet, a similar conclusion could still

[23] This computation was carried out by Miss E. Williams on the IBM 650 computer at Lawrence Radiation Laboratory. We wish to thank her for her assistance.

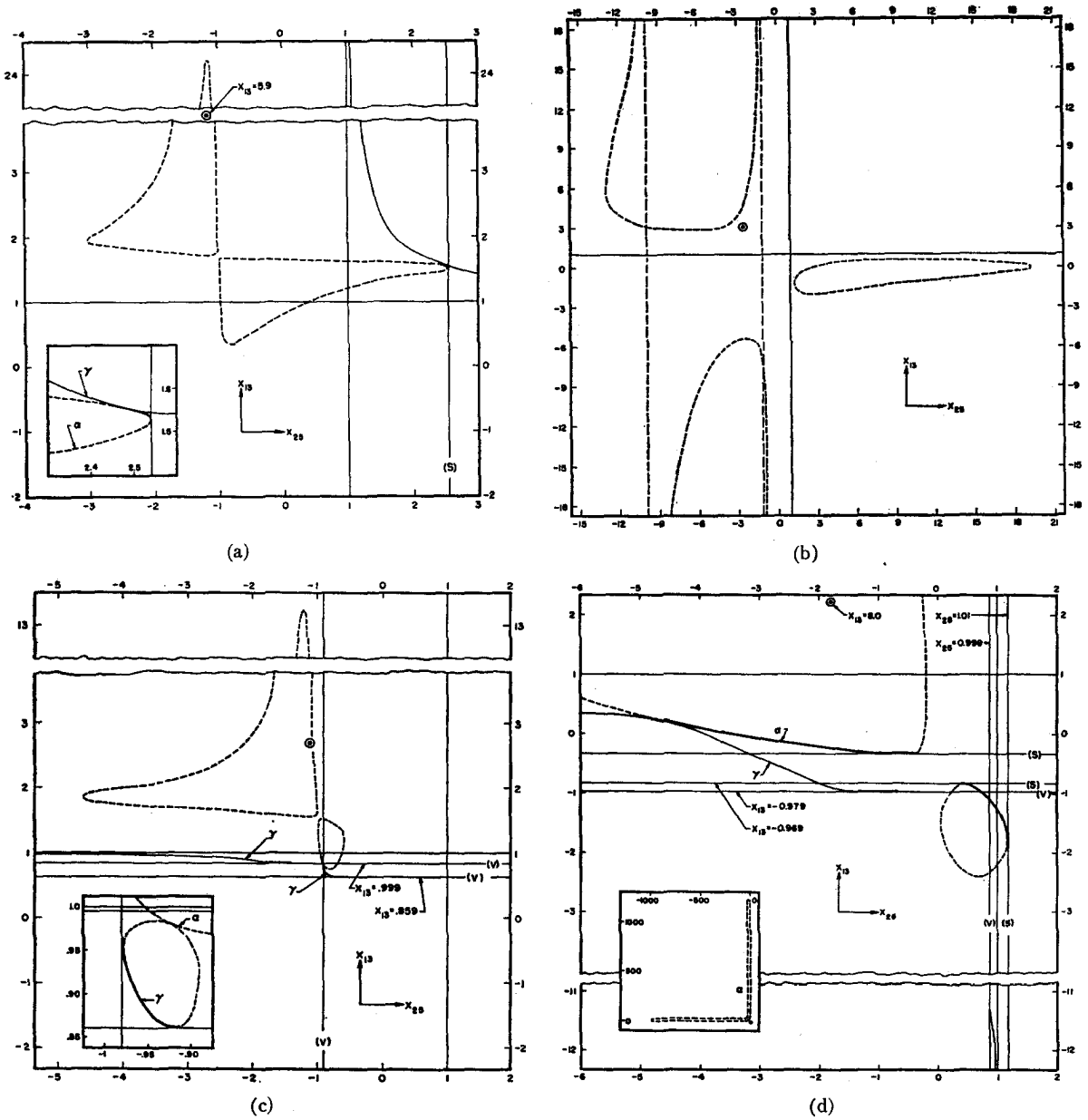


FIG. 13. Singularities corresponding to the processes (a)–(d), respectively. Only the singularities from corresponding limits are indicated. The entire curve α is shown: singular arcs are heavy solid, and regular arcs are dashed. Other singular lines and curves are light. The circled dots indicate the points associated with the momenta of Eqs. (4.1). The asymptotes in Fig. (b) are regular. A few minor distortions were made in the figures.

be valid. We would have to take care so as not to cross a branch cut, but the absence of complex tangencies with the surface σ_γ [see the end of Sec. 2(F)] would simplify the analysis.

C. Singularities for Processes (a)–(d) for Corresponding Limits

In Sec. 4(B) we were able to determine completely the singularities for the initial process, and we note that the existence of the real region R of analyticity greatly simplified the problem. On the other hand, we see from Table I that in each of the processes (a)–(d) at least one parameter is greater than one, and this implies that there is no analogous real region of analyticity. Consequently method (1) of Sec. 3(E) cannot be used here, and the complete determination of singularities is much more involved.

In this section we shall attempt to give the reader an orientation in the singularities corresponding to the processes (a)–(d). We shall therefore consider only the singularities in the real region, and in the corresponding

limits, i.e., $\mathrm{Im}x_{ij} \to 0+$. For these limits method (2) of Sec. 3(E) is applicable and convenient.

One complication arises here, namely, that the singular nature of tangent lines is determined by scattering diagrams with unstable external masses. Such singularities have not been studied before, and therefore we give a short discussion of them in Appendix B. In particular, examples of scattering diagrams which have been used in this connection are given later in Fig. 25. We shall not discuss these singularities further in the text, but we shall only state the results as they are needed.

The geometric configurations of α, and for convenience also those of γ, were determined on a computer.[23] The determination of the singularities by method (2) presents no difficulties, and the singularities for each process, together with the entire curve α, are shown in Fig. 13. We shall now complete the discussion of these singularities with a few comments.

One striking fact about these configurations is that in each case α passes quite near to the point (x_{13},x_{25}) determined by Eqs. (4.1). One would expect that α passes through the physical region. If this were so, then the regularity of α would also be implied by Theorem 3.1. However, we have not determined the physical region, and therefore we cannot be sure if α passes through it.

We may note that in Figs. 13(a) and (d) the tangency between α and a singular arc of γ follows in fact from the rest of the available information [see Sec. 3(E)]. Finally, we see that for process (c) there are no five-point poles for corresponding limits, and only one small arc of γ is singular. For process (b) the only singularities which are shown are the normal thresholds. However, all four processes have singularities for other limits, and also have complex singularities, so one should not be misled by this regularity for corresponding limits.

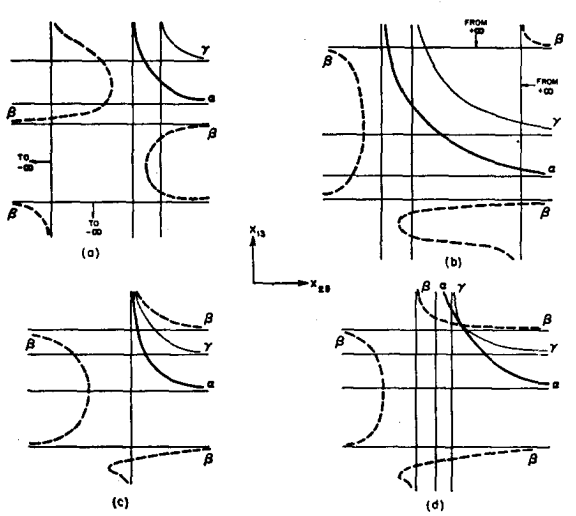


FIG. 14. The development of tangency between α and γ_1 as x_{45} is varied from its value for the initial process to its value for process (a).

D. Example of the Continuation Procedure

We described in Sec. 3(E) a method, the continuation procedure, and we asserted that this method is of wider applicability than the other two methods given there. We will now illustrate how this method can be applied by means of a specific example. For simplicity we will consider a situation in which both the beginning and the end are already familiar to us: We shall consider the variation of parameters from their values for the initial process to their values for process (a), and we shall only consider corresponding limits. In the next section, however, we shall give examples of other limits, and of complex singularities.

We note that we also make use of the continuation procedure in Appendix B; however, the continuation of this section is much more involved than any of the examples in this appendix. Moreover, the continuation of this section requires special considerations in two places: the formation of tangency between α and γ_1, and the appearance of an oval.

An examination of the parameters x_{ij} where $i, j \neq 4$, shows that their values for the initial process differ but little from their values for process (a). One would therefore expect that there is no significant change in the singularity curves as these parameters are varied from one set of values to the other. On the other hand, the parameters x_{i4} or x_{45}, have to be varied over an extended range of values. We choose somewhat arbitrarily the following procedure We first vary the parameters x_{ij}, where $i, j \neq 4$, to their physical values, i.e., to their values for process (a). Next we vary x_{45}. This variation is just that which is shown in Fig. 3, parts (a)–(e). [Part (f) shows the variation to $x_{45} < -1$, while the physical value is -0.989.] We next vary x_{34} to its physical value. Then we vary x_{24} to $-1+\epsilon$ and x_{14} to $1-\epsilon$, and finally we vary x_{14} and x_{24} to their physical values.

Let us now consider the singularities of the curve α when $x_{45}=0.989$, i.e., in the last diagram of Fig. 3(e). If we vary x_{45} further so that $x_{45} < -1$, as in Fig. 3(f), then two of the asymptotes become complex, while remaining regular, and two branches of α join. To avoid a contradiction, the two branches of α must be both singular or both regular as they join. However, one of the branches of α is singular for the initial process, and it can become regular in part only by tangency to the singular part of γ [Sec. 3(C)]. Let us now show that this tangency indeed occurs.

The singularity curves for the initial process are given in Figs. 5 and 12, and these curves are not changed in any significant way if we vary the parameters x_{ij}, where $i, j \neq 4$, to their physical values. Moreover, the resulting configuration of γ is the same as the configuration of γ for the initial process, since γ does not depend on x_{i4}, or on x_{45}.

The parameter x_{45} is varied next, and let us follow the behavior of the curve β as this variation takes

place. The general observations described in Sec. 2(F) are very helpful here. When $x_{45}=0$, then β is a hyperbola, as in Fig. 6(d) but reflected in a coordinate axis. We conclude that for $x_{45}=\eta_1$ and $x_{45}=-\eta_2$, with $\eta_i>0$ and small, we have the configurations shown in Fig. 14, parts (a) and (b), respectively. The next interesting change takes place when $x_{45}=-x_{24}$. Then the asymptotes to α, to β, and to γ all coincide at $x_{25}=1$, and the three curves approach one another, and their common asymptote, as $x_{13}\to\infty$. This configuration is shown in Fig. 14(c). It is also of interest that this is just the place where the asymptote to α, which we are considering, becomes regular. As x_{45} is decreased further, we find for one asymptote to β that $x_{25}<1$, while for one other, $x_{13}>1$; see Fig. 14(d). It follows that β and γ necessarily intersect. Then the arc of α under consideration becomes in part regular, as Fig. 14(d) shows. Further variation of x_{45} does not lead to any essential changes, and the tangency which we mentioned above does indeed occur.

We now consider the continuation of α as the other parameters are varied. This continuation is shown in Fig. 15, and we only need to add a few remarks. The starting point of this figure, part (a), corresponds to x_{45} at its physical value. We note that the point of tangency which we just discussed does not undergo any further significant displacements, and in each of the diagrams (a)–(e) of Fig. 15 the corresponding arc of α is singular between the point of tangency and infinity. Moreover, in the diagrams (a)–(e) all the tangent lines to α are regular, and there are no other tangencies of α to the singular arc γ_1 to alter the singular nature of α. In particular, we conclude, e.g., by method (2) of Sec. 3(E), that the oval which appears upon the variation from Fig. 15(d) to Fig. 15(e) is regular.

The last important change takes place as x_{14} is varied from $1-\epsilon$ to its physical value ($x_{14}>1$). Then one tangent line is first displaced along the x_{25} axis to $-\infty$, and then recedes from $+\infty$ as a singular line, bounding the singular section of α. This last configuration is shown in Fig. 15(f), and also in Fig. 13(a).

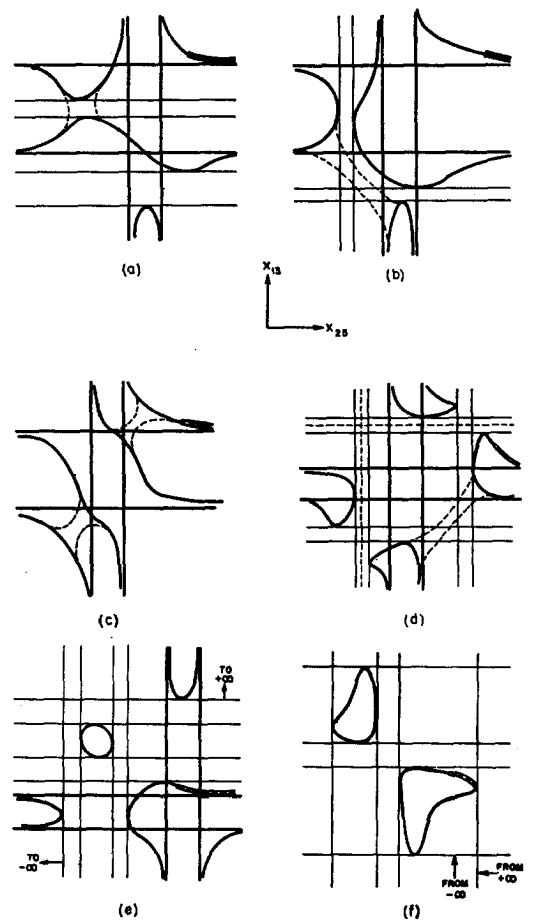


FIG. 15. The displacement of α as the parameters are varied. Figure (a) corresponds to the parameters x_{14}, x_{24}, and x_{34} at their values for the initial process, and the other parameters at their physical values for process (a). Figures (b)–(d) show the displacements as x_{34} is varied to its physical value. In Fig. (e), $x_{14}=1-\eta$ and $x_{24}=-1+\eta'$, with η, $\eta'>0$ and small. In Fig. (f) all parameters have their physical values [cf. Fig. 13(a)]. In each figure the singular arc of α is indicated by the heavy dashed line. The light dashed lines indicate how the branches merge and separate upon variation to the next figure, or that an oval will appear.

E. Illustrations of Mixed Limits and of Complex Singularities

We now wish to give a few illustrations of analytic continuation to the region where the imaginary parts have different signs, and in particular, to the real branch cuts but with limits other than corresponding, i.e., with mixed limits. A systematic study of these regions would require a discussion of a large number of separate cases, and for this reason we give only a few examples to illustrate the general pattern.

We will begin by giving a systematic discussion for process (a), and we consider real singularities first. Since x_{14} is on its branch cut, we will assume the limit $\mathrm{Im}x_{14}\to 0+$, but all the combinations of limits, where independently $\mathrm{Im}x_{13}\to 0\pm$ and $\mathrm{Im}x_{15}\to 0\pm$, must be considered separately. The three limits which have not been considered in Secs. 4(C,D) are:

$$\mathrm{Im}x_{14}\to 0+,\quad \mathrm{Im}x_{13}\to 0+,\quad \mathrm{Im}x_{25}\to 0-; \tag{4.3a}$$

$$\mathrm{Im}x_{14}\to 0+,\quad \mathrm{Im}x_{13}\to 0-,\quad \mathrm{Im}x_{25}\to 0-; \tag{4.3b}$$

$$\mathrm{Im}x_{14}\to 0+,\quad \mathrm{Im}x_{13}\to 0-,\quad \mathrm{Im}x_{25}\to 0+. \tag{4.3c}$$

We may also note that if we were to take $\mathrm{Im}x_{14}\to 0-$, we would be dealing with a case which is complex conjugate to one of the four that we just mentioned.

In the limit combination (4.3a) there are no singular tangencies, since both γ and the tangent lines are all regular (this can be easily verified). Furthermore, there are no branch cuts due to complex vertex singularities, and we conclude that the entire curve α is regular.

On the other hand, in combination (4.3b) the branch γ_1 is singular. The tangent lines are regular, but we have a branch cut extending from the complex region, and a careful treatment is therefore needed. One way to discuss this case is to examine the continuation of M_5 as x_{14} is varied from $1-\eta$ to its actual value, $x_{14}=1.676+i\epsilon$ (taken in the limit). In effect, we need to examine Fig. 14(e), and to vary the parameters from there. Let us therefore follow the behavior of the singular part of α as the parameters are varied. In particular, we will first vary x_{14} to $1+\eta$. The tangent line ($K_3=0$) that recedes from infinity is now regular, and we may continue M_5 along α until we come upon the branch cut that comes from the complex region. Similarly we may follow the regular part of α up to the complex branch cut. This is quite analogous to the discussion of case (2) in Appendix B, and the fact that in one case the two branches merge at infinity, and in the other, at a finite point, is not significant.

Once the transition from $x_{14}=1-\eta$ to $1+\eta$ is understood, the rest of the variation presents no problem. Figure 16 shows the singularities here described. The singular arc of α extends from the tangency with γ_1 to the branch cut, and we see that with the cuts placed as indicated, a part of the singular arc of α extends to a complex singular surface σ_α having $\mathrm{Im}x_{13}$, $\mathrm{Im}x_{25}<0$.

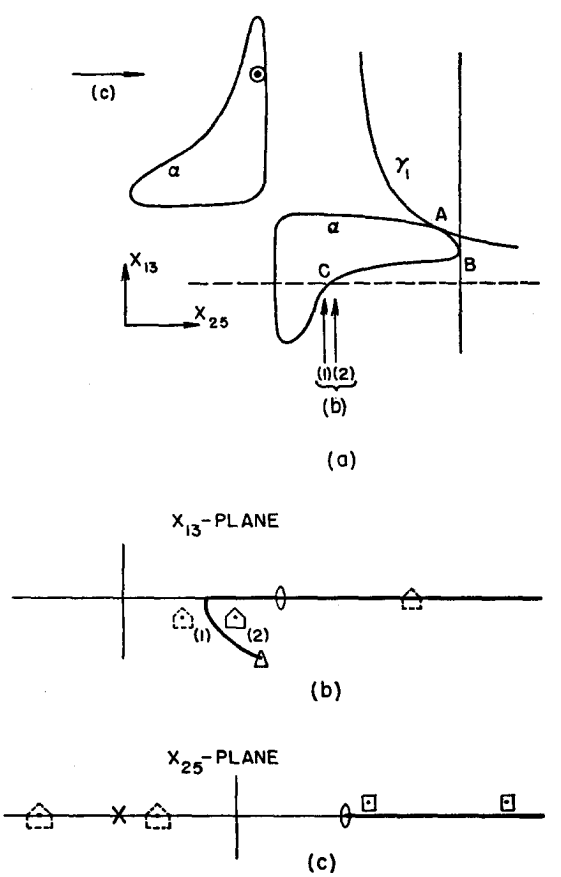


FIG. 16. Singularities for process (a). Figures (b) and (c) show the complex planes indicated in Fig. (a), and the singularities, branch cuts, and regular points of α are indicated in an obvious way. In Fig. (a) the circled dot corresponds to the physica point, while in Fig. (c) the analogous value of x_{25} is indicated by a cross.

We should point out that Fig. 3(a) of reference 1 is analogous to Fig. 15(c). The continuation to the five-point poles which is indicated in Fig. 3(a) (of reference 1) follows easily from Theorem 3.3.

The final limit combination, (4.3c), is quite analogous to (4.3b). Here the singular arc of α is bounded by the singular tangent line ($K_3=0$) and by the branch cut.

Figure 17 gives singularities for processes (b) and (c), and Fig. 18, for process (d). The assertions shown on these figures can be established without difficulty. (As usual, we assume corresponding limits for the parameters.)

Figure 18 presupposes that the complex vertex branch cuts do not join the real axes near the region of interest; otherwise the figure would have to be modified appropriately. We note that a part of the singular arc of α extends to a singular surface σ_α having $\mathrm{Im}x_{13}$ and $\mathrm{Im}x_{25}$ of opposite signs. This singular surface is necessarily delimited in the complex region by branch cuts generated by vertex or scattering singularities.

Figure 18(e) corrects Fig. 3(d) of reference 1; there is an error in the latter figure. However, the observations which were made in reference 1 in connection with this figure remain valid.

5. DISCUSSION

In this section we will first make some general remarks on the properties of singularity curves. These remarks are of course based on the analysis given in the previous sections. This is followed by a discussion of dispersion relations for production amplitudes and of complex singularities and, finally, we will make some observations, which supplement the discussion of reference 1, on the possible experimental consequences of our results.

A. General Observations Concerning the Singularity Curves

As we have mentioned in the Introduction, the investigation given here is a preliminary one, and we have, by no means, made an attempt at completeness. However, if the analytic properties of production amplitudes are to be used to analyze experimental data, a more extensive knowledge of these properties is definitely needed. We would therefore like to make some recommendations which may serve as a basis for further studies.

As a starting point we present a resume and criticism of our analysis. We began by finding a specific choice of parameters, the initial process, for which we could discuss completely the analyticity by recognizing the analogy between this process and the scattering amplitudes. However, this particular choice of parameters is

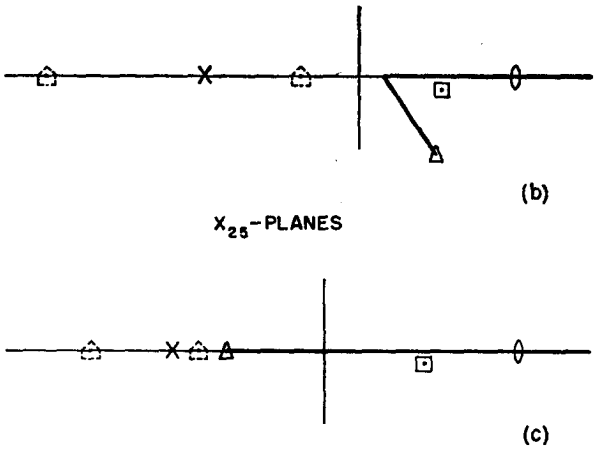


FIG. 17. The complex x_{25} planes for processes (b) and (c), respectively. The value of x_{13} in each figure is the value associated with the physical process.

of little interest in itself, since this process was not intended to correspond to any physical process, and the range of parameters leading to such analogies is quite limited.

Nevertheless, the initial process served as an orientation point and also as the starting point for the continuation procedure.

The continuation procedure is a tedious one and is not at all suited to an examination of a great number of cases. Because of this, we have largely restricted our investigation to those analytic properties which could be determined by more direct methods. We saw that we were able to obtain a significant amount of information by means of more direct methods, and we feel that it would facilitate further investigations of the five-point function if additional special properties could be established.

In order to illustrate such properties we will list some examples which we consider relevant, but which we have not examined in detail. We observe that there appears to be a parallel in the behavior of the tangent lines and the asymptotes to α. We will give two illustrations of this. First, consider a pair of tangent lines that are merging and becoming complex. If one of the merging lines is singular and one of the branches is singular, from some set of limits, then the singular nature of the combined branch will depend on the continuation around the singular complex tangent line. (This is shown explicitly for scattering singularities in Appendix B.) If two asymptotes are merging and becoming complex we have a similar situation: If one of the merging asymptotes and one of the merging branches is singular, then the singular nature of the combined branch will depend on a continuation around the complex singularity [Sec. 3(E)]. The interesting aspect of the situation depends on the fact that Theorem 3.3 applies to the case of merging tangent times, but not to the case of merging asymptotes.

The second illustration concerns the behavior of α near a singular asymptote. This behavior is related to that described in Theorem 3.2, viz., that α is singular (for some of the limits) for sufficiently large values in one direction, and regular for sufficiently large values in the opposite direction, near the asymptote. This pattern of behavior occurs in each of the cases studied, and may have a general validity.

Another example that can be given is a practical remark concerning the merging of tangents, asymptotes, and curves. It appears that nearly all of the significant changes occur when the parameters are varied in the interval $-1-\eta \leq x_{ij} \leq 1+\eta$, where $\eta > 0$ and is small. In particular, an essential change occurs at the physical threshold, i.e., at $x_{ij}=1$. This latter change can be compared to the behavior of one-dimensional dispersion relations in the momentum transfer for scattering

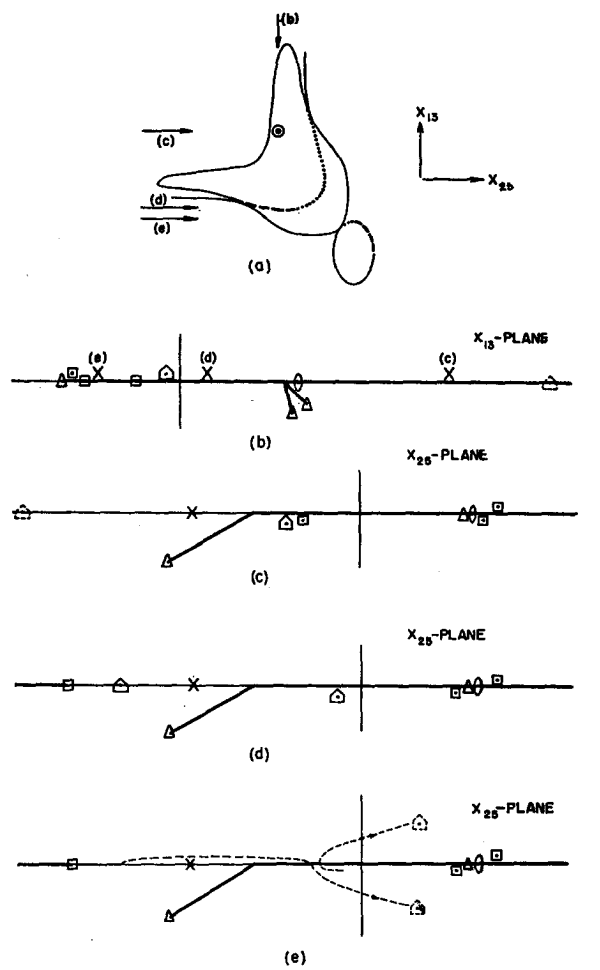


FIG. 18. Singularities for process (d). The organization of this figure is the same as that of Fig. 16. However, these differences should be noted: The crosses in Figs. (b)–(e) do not necessarily refer to physical processes. The dashed arcs in Fig. (a) indicate singularities from corresponding limits, while the dotted arcs indicate those from opposite limits, i.e., for x_{13} and x_{25}. In Fig. (e) the dashed lines represent the path along which the five-point poles are displaced as one varies x_{13} from the value (d) to (e).

processes, when the incident energy changes from a value below the threshold to a value above the threshold. However, in the regions $x_{ij}<1$ and $x_{ij}<-1$, very large variations in x_{ij} can often be made without markedly affecting the singular nature of the curves.

There are also certain situations which could be investigated to advantage, but which we have not analyzed. For example, it would be very advantageous to have more detailed information concerning the properties of the curve β, and in particular, the conditions for the transitions between the cases shown in Fig. 6. Also, the conditions for the crossing of nonpaired tangent lines [see Sec. 2(D)] would be informative, as such situations constitute transitions between cases.

This point brings us to a general discussion of the five-point function. The ultimate aim of investigating the five-point function is a description of the various configurations, with respect to both geometric and analytic properties. This description necessarily has to be broken into cases, just as has been done for the scattering and vertex amplitudes.

In general, the transitions between the cases correspond to the coincidences of tangent lines or asymptotes. For those transitions which are determined by the coincidence of paired tangent lines, or of a tangent line and an asymptote, the conditions can be easily specified [Sec. 2(D)]. On the other hand, transitions which are characterized by the crossing of nonpaired tangent lines seem to be much more difficult to determine. This is to be contrasted with the four-point function, where the conditions for all transitions can be easily specified. The transitions in which nonpaired tangent lines cross are particularly significant for the four-point function, but it is difficult to say how significant they may be for the five-point function.

Finally, we may also remark that the number of cases which may be needed for an adequate description of the five-point function is definitely larger than that needed for the four-point function, and this in itself is of course an additional difficulty.

B. Analytic Properties and Dispersion Relations

A primary application of an investigation of the five-point function concerns the validity of dispersion relations for production amplitudes. We remark first that there is a variety of dispersion relations which can be considered, depending on the choice of independent variables and on the number of these variables which are dispersed. In contrast to the scattering amplitudes, here several invariants are on their physical cuts, and as we shall see this appears to be the source of the difficulty that one encounters in constructing dispersion relations for production amplitudes. In the following we will consider dispersion relations as valid only if a representation involving real contours exists in the considered variables, i.e., if there are no complex singularities.

It is natural to begin our discussion with the variables x_{ij}. For process (a) we see from Figs. 16(b) and (c) that a one-dimensional dispersion relation in x_{13} is not valid, but one in x_{25} is valid. An examination of Fig. 16(b) shows that the dispersion relation fails because of a complex vertex-type singularity, which arises when $x_{14}>1$, i.e., when x_{14} is on its physical cut. This condition, however, is fulfilled whenever p_{45} and $-p_{15}$ are real outgoing momenta. On the other hand, it is easy to see that for a dispersion relation in x_{25}, a necessary condition is that $x_{35}<1$. The conclusion of interest is that in the two cases only a vertex singularity is complex, and that the presence of this singularity is related directly to the fact that one of the invariants is on its physical cut.

However, it should be emphasized that the diagram for process (a) is somewhat special in that it produces a threshold for the two outgoing pions, i.e., $x_{35}=1$, at $4m_N^2$, rather than at $4m_\pi^2$ (where $x_{35}=-0.957$). It is this fact that allows a dispersion relation in x_{25}. However, there exist diagrams which will yield a threshold at $4m_\pi^2$, so that complex vertex singularities are present, and dispersion relations for the *total* production amplitude in x_{25} are not possible. It is clear that the argument just presented is quite general, since, with our choice of independent variables, three of them will always be on their physical cuts. We conclude that, in general, one-dimensional dispersion relations are not valid for our choice of the scalar invariants.

These remarks can be applied to Lardner's[8] analysis of the three-particle intermediate states occurring in scattering amplitudes. On the assumption that one-dimensional dispersion relations exist for production amplitudes in what is essentially x_{25}, Lardner shows that the Mandelstam representation is valid with the contributions of the three-particle states included. The relevant point here concerning his investigations is the validity of the assumption. We wish to point out that ordinarily this assumption is violated (see also reference 9). The difficulty arises because of complex vertex singularities.

However, it is important to realize that one-dimensional dispersion relations, in variables other than the invariants used here, have been constructed, with certain restrictions.[3–5] In fact, the results of our investigation suggest that the invariant momenta are not suitable for a representation involving real contours, and perhaps some effort should be devoted to the selection of invariants with the help of which complex singularities would be suppressed. This remark is also suggested by the work of Landshoff and Treiman.[9]

With regard to the results of Kim,[5] we may note that they apply to a restricted class of Feynman diagrams involving electromagnetic interactions, and are valid for only a limited range of his variables. Nevertheless, while his results are expressed in terms of variables other than ours, these results can be

extended by means of an analysis analogous to that presented here.

With regard to dispersion relations of higher dimensions we will make only two remarks: First, a two-dimensional representation, analogous to that of Mandelstam for the four-point function, is valid for the initial process, but is not valid for processes (a)–(d). In process (a) there are complex vertex singularities and complex five-point poles, while for processes (b)–(d) there are complex scattering singularities as well. Second, it follows from Appendix B that a five dimensional representation is not valid, even if all the mass parameters $x_{i,i+1}$ are less than -1; in this case there are complex scattering singularities. In other cases there are complex vertex singularities.

Let us now turn to a discussion of complex singularities. The better part of the effort devoted so far to studies of dispersion relations has been directed toward establishing the absence of complex singularities. However, throughout our paper and, in particular, in the preceding discussion, we have been vary careful to note the type of complex singularity that arises. Our motivation for this is the possibility that if the nature of the complex singularities is understood, it may be possible to exploit them and obtain useful results nevertheless. Examples of this occur in studies of partial-wave amplitudes[24] and in an analysis of the scalar form-factor of the nucleon.[25]

Of the complex singularities which occur in our investigation, those of the vertex type are apparently the most significant and also the easiest to handle. In case of these singularities, there are explicit formulas which are simple in form. This fact suggests that some consideration should be given to the utilization of dispersion relations with complex contours as an alternative to the construction of dispersion relations with real contours but involving a more complicated set of invariants.

One general property of the complex singularities can be expressed most conveniently if we consider the five invariants (other than the masses) as complex variables. The hypersurfaces containing the singularities in this five-dimensional complex space have the property that they are all connected to the real axis. This follows from the structure of singularity surfaces and the properties of continuation of the function along the surfaces. This means that all of the complex branch cuts in a given complex plane can be connected to the real axis in a natural way.

This property has some bearing on a recent paper of Blankenbecler.[26] The approach of this paper requires a continuation of a production amplitude from a physical region, where the corresponding limits $(\mathrm{Im}x_{ij} \to 0+)$ are taken, to a region where $\mathrm{Im}x_{kl} \to 0-$, for some particular k, l, corresponding to the dispersion variable. The fact that the complex branch cuts arrange themselves in a natural way is an indication as to how the continuation is to be carried out. However, this is an important point, and a more careful study is definitely needed.

We remark, finally, that we have ignored the effect of crossed diagrams. These diagrams will, in general, introduce singularities into the region where $\mathrm{Im}x_{ij}>0$. The analyticity in this region has been a simplifying circumstance, and the corresponding nonanalyticity, in the general case, illustrates a further difficulty in the study of production amplitudes.

C. Discussion of Experimental Consequences

As we mentioned in the Introduction, one of our motivations was concerned with the experimental consequences of intrinsic five-point singularities,[1] in particular, with regard to two actual experiments. Before we discuss these and other experiments in detail, two remarks concerning the general validity of our conclusions should be made.

First, as we pointed out in Sec. 4(A), we confined our investigation to only one configuration of momenta for each process. However, even if moderate changes in these configurations are made, we feel that the essential aspects of the arrangement of singularities will remain unchanged. Moreover, the contributions from certain extreme configurations to cross sections are not expected to be significant.

Of course, to verify the validity of these preceding remarks it is necessary to have a complete knowledge of the physical region in terms of the five invariants. In fact, such knowledge is essential for a complete understanding of production amplitudes. (Further, if such knowledge were available, it could also be used to establish the regularity of certain singularity curves; see Theorem 3.1.)

Second, we have not considered the effects of crossed diagrams. However, it seems to be a general conclusion that the singularities of crossed diagrams are further removed from the physical region than those that we have considered in this paper. This is suggested by the fact that for these diagrams more of the vectors p_{ij} are spacelike, and consequently some of the invariants x_{ij} are smaller. In general, this produces a situation in which fewer of the invariants are on their physical cuts, and consequently some of the anomalous singularities will be displaced to an unphysical sheet. However, it is difficult to convince oneself that the singularities peculiar to crossed diagrams are never of experimental importance. Thus all statements that are made in the following are subject to this reservation.

Let us now turn to the applications of our investigation to experiments. These applications are more direct than those discussed in Sec. 5(B).

One such application concerns the extrapolation of

[24] J. G. Taylor and E. A. Warburton, Phys. Rev. **120**, 1506 (1960).
[25] R. Blankenbecler and J. Tarski, Phys. Rev. **125**, 782 (1962).
[26] R. Blankenbecler, Phys. Rev. **122**, 983 (1961).

experimental data. It must be emphasized that in this application the choice of variables is of great importance. Extrapolations are customarily performed by fixing all variables except the extrapolating variable. In general, selecting different variables corresponds to extrapolating along different directions in the real region of the invariants, and to different points. It follows that when one extrapolates a given set of data, the validity of the extrapolation is dependent on the choice of variables.

Of the processes that we have studied in detail, the preceding remarks apply directly only to process (a), since the extrapolation is of immediate interest only in this amplitude. Figure 16(b) shows that the contemplated extrapolation, in terms of our variables, should yield the correct amplitude at the final extrapolation point, namely $p_{25}^2=m_\pi^2$, since there are no nearby singularities. (One may also check that, as long as our variables are adhered to, diagrams with internal pion lines also do not lead to singularities near the pole.)

A contrary conclusion was obtained by Landshoff and Treiman,[8] who use variables previously introduced by Ascoli.[6] They found that vertex singularities occur in the neighborhood of the pion pole, showing that the extrapolation procedure was invalid. It thus appears that for the intended extrapolation the variables x_{ij} are more satisfactory than those of Ascoli. However, a detailed study is needed in order to draw a more definite conclusion.

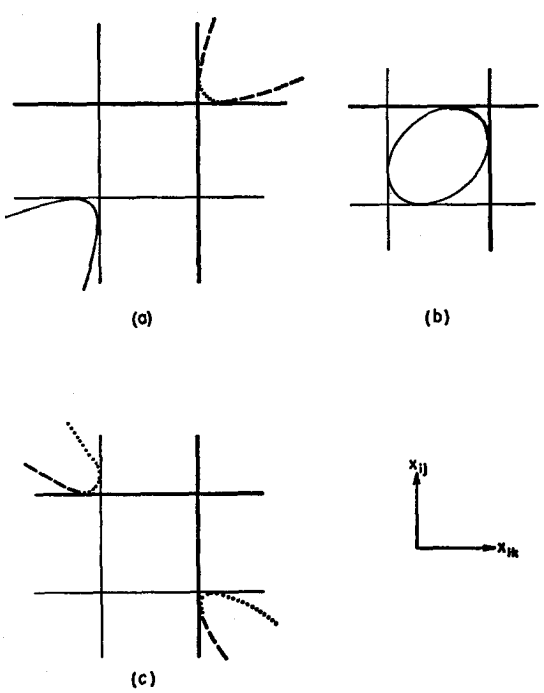


FIG. 19. Vertex singularity curves. For Fig. (a), $x_{jk}<-1$; for Fig. (b), $-1<x_{jk}<1$; and for Fig. (c), $x_{jk}>1$. Singular parts are shown heavy, and regular parts, light. In Figs. (a) and (c) dashes indicate singularities for opposite limits, and dots, singularities for corresponding limits.

We also remark that there are apparently no experimental effects directly due to the five-point poles in process (a).

In process (b), the anomalous peak found by Abashian, Booth, and Crowe[22] is not a direct effect of the five-point poles. It is also very doubtful that the vertex branch cut near the physical point can produce this effect. The effect of such a branch cut would be, in general, a smoother distribution.

In connection with process (c), our configuration of momenta does not lead to direct effects of five-point poles. We have not made an attempt to find a possible momentum configuration which might lead to such direct effects. However, for the chosen configuration of momenta five-point poles do appear on the physical sheet, but are separated from the physical region by the vertex branch cut. This suggests that a different choice for the momentum configuration would lead to five-point poles which could have direct experimental effects. We should emphasize that the search for such a configuration would be greatly aided by the knowledge of the physical region. On the other hand, we remark that the chosen configuration should lead to direct effects of the vertex singularities.

With regard to process (d), we remark that the amplitude has many singularities, including the five-point poles, near the region of interest, i.e., at energies near $9m_\pi^2$. These singularities evidently cannot be ignored in an evaluation of the contribution of the three-pion intermediate state to the deuteron form factor.

We also carried out a determination of singularities for the process $N\bar{N}\to 3\pi$ [diagram for process (a) but with p_{12} and $-p_{15}$ incoming], with $T_{\bar{N}}=350$ Mev. We concluded that the only singularities on the physical sheet are the normal thresholds and the arc γ_1. In particular, there are no five-point poles on the physical sheet.

The last two processes which we have mentioned support a familiar conclusion: The presence of external vertices which are near instability leads to singularities, other than normal thresholds, on the physical sheet.

ACKNOWLEDGMENTS

We are indebted to Dr. D. L. Judd for his hospitality at the Lawrence Radiation Laboratory, where the early part of this work was carried out. One of us (J.T.) would like to thank Professor J. R. Oppenheimer for his hospitality at the Institute for Advanced Study.

APPENDIX. SINGULARITIES FOR DIAGRAMS OF ORDER FOUR OR LESS

In this appendix we discuss the main properties of lower order singularities, i.e., of normal thresholds, vertex singularities, and scattering singularities. In Appendix A we consider normal thresholds, vertex singularities, and scattering singularities with stable

external masses. These singularities have been discussed in other works, and we summarize their properties for the convenience of the reader. In Appendix B we consider scattering singularities involving unstable external masses. The relevance of these singularities, and of vertex singularities with unstable masses, was pointed out in Sec. 3(A): Contracting the single-loop diagram leads to diagrams which are analogous to lower order diagrams but with unstable external masses.

A. Summary of Basic Facts

Most of the facts here presented are established, explicitly or implicitly, in references 7, 12, 13, and 14; the others follow trivially.

Normal Thresholds: In case of M_5, the singularity surfaces associated with the normal thresholds are given by the equations $x_{ij}=\pm 1$. The points for which $x_{ij}=1$ are singular, while those for which $x_{ij}=-1$ are regular.

Vertex Singularities: Let the vertex singularities under discussion involve the quantities x_{ij}, x_{jk}, and x_{ik}. The associated singularity surface is then given by the equation $K_{lm}=0$, where $i, \cdots, m$ are all different.

We first consider x_{ij} and x_{jk} to be parameters less than one, and we have then these two cases:

(1) If $x_{ij}\leq -1$ or $x_{jk}\leq -1$, then there are no vertex singularities on the physical sheet.

(2) If $-1\leq x_{ij}, x_{jk}<1$, then the equation $K_{lm}=0$ has two real solutions; $x_{ik}=-\cos(\theta_{ij}\pm\theta_{jk})$. The points where $x_{ik}=-\cos(\theta_{ij}-\theta_{jk})$ are regular. The points where $x_{ik}=-\cos(\theta_{ij}+\theta_{jk})$ are singular if and only if

$$\theta_{ij}+\theta_{jk}\geq\pi. \tag{A1}$$

The case where one of the masses is unstable, i.e., where $x_{ij}>1$ or $x_{jk}>1$, is most conveniently discussed with the help of singularity curves. We therefore consider these next.

We now take x_{ij} and x_{ik} to be variables, and x_{jk}, a parameter. Figure 19 shows three configurations of the curve $K_{lm}=0$ in the real (x_{ij},x_{ik}) plane. The singularities indicated in Fig. 19(b) exhibit the information expressed under case (2) above. The singular behavior of the curves, as shown in Fig. 19(a), can be easily established by continuation from the real region. Figure 19(c) shows information which is also displayed in Fig. 19(a), but all of the figures are included for completeness.

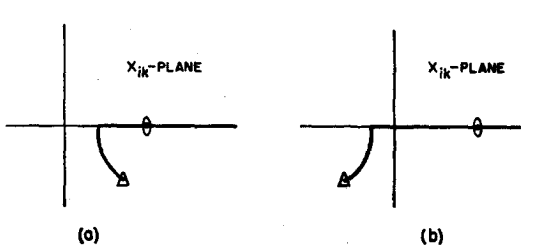


FIG. 20. Complex vertex singularities. We take $x_{ij}>1$. Then Fig. (a) corresponds to $-1<x_{jk}<0$, and Fig. (b), to $0<x_{jk}<1$. The selection of the branch cut is explained in the text.

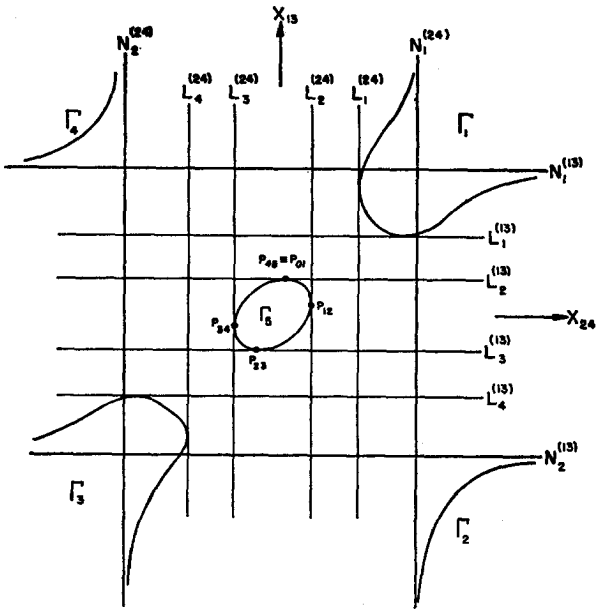


FIG. 21. The curve Γ, with the labeling of relevant lines and points indicated.

The singularities for the case $x_{ij}>1$ can now be easily determined. If $x_{jk}<-1$ or $x_{jk}>1$ then the values of x_{ik} corresponding to the singularities are real, and the behavior of these singularities can be obtained from Fig. 19(a) or (c). On the other hand, if $-1<x_{jk}<1$, then the values of x_{ik} corresponding to the singularities are complex. Furthermore, if we specify the limit $\mathrm{Im}x_{ij}\to 0+$, then those singularities are on the physical sheet for which $\mathrm{Im}x_{ik}<0$. This is shown in Fig. 20. The branch cuts can of course be chosen arbitrarily, at least in a perturbation-theoretic treatment. We have chosen for each cut the path along which the singularity is displaced as the parameter x_{ij} is varied from one to the selected value.

Scattering Singularities for Stable Masses: For definiteness we take x_{13} and x_{24} as the complex variables, and the quantities x_{12}, x_{23}, x_{34}, and x_{14} as parameters. We consider first the case where

$$-1<x_{12}, x_{23}, x_{34}, x_{14}<1. \tag{A2}$$

A typical configuration of the singularity curve Γ, determined by the equation $K_5=0$, is given in Fig. 21. (The curves Γ and γ are analogous, and Fig. 2 gives another possible configuration of Γ.) We shall also use the labeling shown in Fig. 21 for sections of Γ, and for its tangents and asymptotes. The complex singularity surface will be called Σ. This surface consists of four parts, $\Sigma_1, \cdots, \Sigma_4$, where Σ_i is the complex extension of the arc $P_{i-1,i}P_{i,i+1}$ of Γ_5.

We define

$$\Theta=\theta_{12}+\theta_{23}+\theta_{34}+\theta_{14}, \tag{A3}$$

and we shall call two angles adjacent if they have an index in common. Also, we shall refer to the limits $\mathrm{Im}x_{13}\to 0\pm$, $\mathrm{Im}x_{24}\to 0\pm$ as the limits from corresponding half-planes, or as corresponding limits. [The limits $\mathrm{Im}x_{13}$, $\mathrm{Im}x_{24}\to 0-$ are complex conjugate to the limits $\mathrm{Im}x_{13}$, $\mathrm{Im}x_{24}\to 0+$, and therefore this usage of *corresponding limits* is consistent with the usage introduced in Sec. 3(B).] The limits $\mathrm{Im}x_{13}\to 0\pm$, $\mathrm{Im}x_{24}\to 0\mp$ will be called limits from opposite half-planes, or opposite limits.

General conclusions are the following: The real region of analyticity $R^{(4)}$ always includes the region which is below the line $L_3{}^{(13)}$ and to the left of the line $L_3{}^{(24)}$, and consequently the curve Γ_3 is always regular. The complex surfaces Σ_2, Σ_3, and Σ_4 are always regular. Any point of Γ_1 other than a point of tangency is always singular for limits of one kind, and regular for limits of the other kind.

There are four cases, and we do not discuss the transitions between them.

Case (i): $\Theta<2\pi$, and the sum of any two adjacent angles is less than π. Then $R^{(4)}$ is bounded by the lines $N_1{}^{(13)}$ and $N_1{}^{(24)}$. The curves Γ_2, Γ_4, and Γ_5, and the surface Σ_1 are regular. The curve Γ_1 is singular for corresponding limits.

Case (ii): $\Theta<2\pi$, and the sum of one or two pairs of adjacent angles is greater than π. Then at least one of the two lines $L_1{}^{(13)}$, $L_1{}^{(24)}$ bounds the region $R^{(4)}$. The curves Γ_2, Γ_4, and Γ_5, and the surface Σ_1 are regular, as in case (i). The curve Γ_1 is singular for corresponding limits where $dx_{13}/dx_{24}<0$, and is singular for opposite limits where $dx_{13}/dx_{24}>0$.

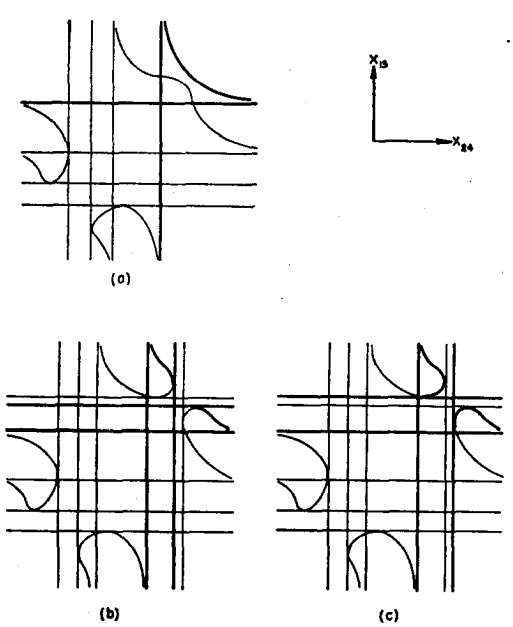


FIG. 22. The curve Γ. Figures (b) and (c) correspond to case (1), while in Fig. (a) one is approaching case (1). Singular parts are indicated by heavy lines, and regular parts, by light lines. In Fig. (b) the limits $\mathrm{Im}x_{13}$, $\mathrm{Im}x_{24}\to 0+$ are taken, and in Fig. (c), the limits $\mathrm{Im}x_{13}$, $\mathrm{Im}x_{24}\to 0-$.

Case (iii): $2\pi<\Theta<2\pi+2\min(\theta_{ij})$. Then $R^{(4)}$ is bounded by the lines $L_2{}^{(13)}$, $L_2{}^{(24)}$, and the arc $P_{01}P_{12}$. This arc is singular, the rest of Γ_5 is regular, Σ_1 is singular, Γ_1 is singular for opposite limits, Γ_2 and Γ_4 are singular at those points which are joined to Σ_1 and are regular elsewhere.

Case (iv): $\Theta>2\pi+2\min(\theta_{ij})$. Then $R^{(4)}$ is bounded by the lines $L_3{}^{(13)}$ and $L_3{}^{(24)}$. The arc $P_{23}P_{34}$ of Γ_5 is singular for corresponding limits, the arcs $P_{01}P_{12}$, $P_{12}P_{23}$, and $P_{34}P_{45}$ are singular for opposite limits, and otherwise Γ_5 is regular. The singularities of Γ_1, Γ_2, Γ_4, and Σ_1 are as in case (iii).

The following rule is very useful for determining which branch of Γ is tangent to each of the lines $L_1{}^{(13)}$ and $L_1{}^{(24)}$: If $\Theta<2\pi$, then Γ_1 is tangent to the line $L_1{}^{(ik)}$ if and only if this line is singular; if $\Theta>2\pi$, then Γ_1 is tangent to the line $L_1{}^{(ik)}$ if and only if this line is regular.

If some x_{ij} is less than -1, then the singular parts of Γ are entirely analogous to those for $x_{ij}=-1$. In particular, Σ_1 disappears, and there can be no complex singularities, even if the sum of the three remaining angles is greater than 2π.

B. Scattering Singularities for Unstable Masses

We will not attempt to give here a systematic discussion of this subject but will merely give a few examples. These examples are intended primarily to suggest the type of analysis that may be appropriate, and the type of answers that one can expect. The method that can be conveniently used here is the continuation procedure.

We will be dealing here with the curve Γ and the surface Σ, as in Appendix A. The parameter x_{14} will refer to an unstable mass: $x_{14}>1$, and, unless an explicit remark to the contrary is made, the remaining masses will be assumed to be stable:

$$x_{12},\ x_{23},\ x_{34}<1. \tag{B1}$$

We could consider eight limit combinations, since we may have independently $\mathrm{Im}x_{14}\to 0\pm$, $\mathrm{Im}x_{13}\to 0\pm$, and $\mathrm{Im}x_{24}\to 0\pm$. However, four of these are complex conjugates of the others. Therefore there is no loss of generality in considering only the limit

$$\mathrm{Im}x_{14}\to 0+, \tag{B2}$$

and this limit will always be understood.

We conclude from Sec. 3(B) that there are no singularities in the region $\mathrm{Im}x_{13}$, $\mathrm{Im}x_{24}>0$. However, the considerations of Sec. 3(B) now do not apply to the region $\mathrm{Im}x_{13}$, $\mathrm{Im}x_{24}<0$, and we may have singularities there.

We will first consider in detail three special cases, and then we will give examples showing how one determines the singular nature of the tangent lines.

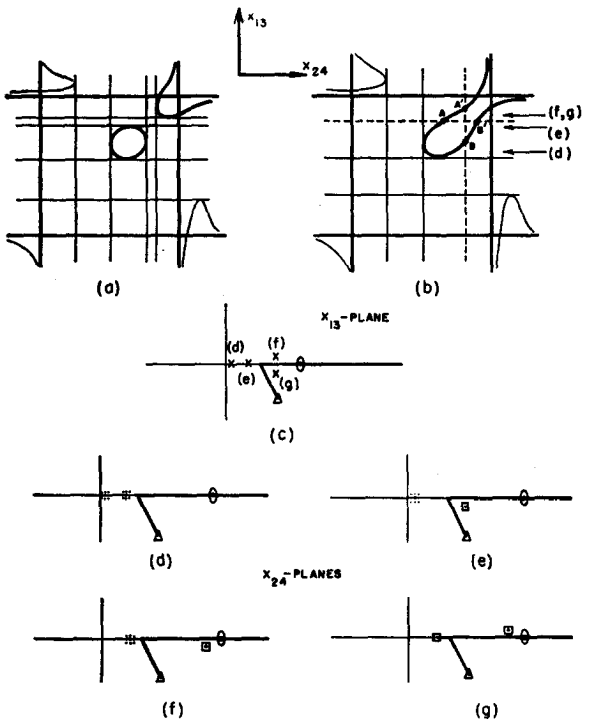


FIG. 23. The curve Γ and selected complex planes for case (2). Figure (b) corresponds to case (2) while in Fig. (a) one is approaching case (2). In these two figures, heavy lines indicate the relevant branches of Γ. In Figs. (b) and (c) the values indicated by (d)–(g) show the values of x_{13} associated with Figs. (d)–(g), respectively. The singularities, branch cuts and the points of Γ which are regular are indicated in an obvious manner.

In the discussion of these cases η will refer to a number which is positive and sufficiently small.

(1) The first case is defined by

$$x_{12},\ x_{23},\ x_{34}<-1. \tag{B3}$$

We see from Fig. 19(a) that all the equations $K_{i5}=0$ determine real tangent lines. Consequently there are no complex vertex singularities, and this fact greatly simplifies the analysis. Figure 22 shows the singularity curves for $x_{14}=1-\eta$ and $x_{14}=1+\eta$. The singularities for two limit combinations are indicated, and we see from Fig. 22(c) that there are complex scattering singularities in the region $\mathrm{Im}x_{13}$, $\mathrm{Im}x_{24}<0$.

This case is not needed for an analysis of processes (a)–(d), but it is referred to in Sec. 5(B).

(2) The second case is defined by $-1<x_{i,i+1}<1$ and

$$\theta_{12}+\theta_{23}+\theta_{34}<\pi. \tag{B4}$$

The singularity curves for $x_{14}=1-\eta$ and $x_{14}=1+\eta$ are shown in Fig. 23(a) and (b), respectively. For $x_{14}=1-\eta$ we have case (ii) described in Appendix A. However, for $x_{14}=1+\eta$, $K_{25}=0$ and $K_{35}=0$ determine complex branch points, and these are shown, together with branch cuts and the other singularities, in Figs. 23(d)–(g). Figure 23(c) shows the values of x_{13} which correspond to the x_{24}-planes of Figs. 23(d)–(g).

Figure 23(b) shows the lines along which the complex branch cuts approach the real region. The exact location of these lines is of course arbitrary. The complex planes, Fig. 23(d)–(g), show the singular nature of Γ_1, and we see that this is related to the location of these lines. To establish the singular behavior given here, we argue as follows: We observe that there are parts of Γ_1 and Γ_5 which are not affected by the variation of x_{14} from $1-\eta$ to $1+\eta$, and therefore we know the singular behavior there. It is then easy to make a continuation from these parts of the curve to the two arcs AA' and BB'. For example, to determine the singular nature of the arc BB' for the limit $\mathrm{Im}x_{24}\to 0-$ or for $\mathrm{Im}x_{13}\to 0+$, we can continue from the arc near the asymptote $x_{13}=1$, and for $\mathrm{Im}x_{24}\to 0+$ or for $\mathrm{Im}x_{13}\to 0-$, from the real region of analyticity. (cf. Theorem 3.3)

An examination of the singular parts of Γ in Fig. 23 shows that the singular part of Γ does not extend to the complex region. We conclude that the only complex singularities are of vertex type. This can be compared with case (1), where the only complex singularities are of scattering type. In each of these cases, we see that only one type of complex singularities arises, and this greatly simplifies the description of singularities. In general, this is not the case, as we will see in the following example.

(3) The third case is defined by $-1<x_{i,i+1}<1$ and

$$\theta_{12}+\theta_{23}-\theta_{34}>\pi. \tag{B5}$$

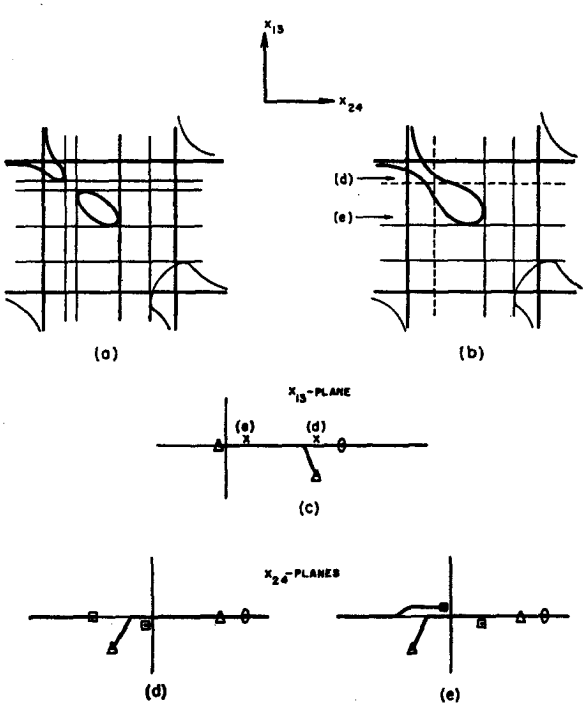


FIG. 24. The curve Γ and selected complex planes for case (3). The remarks given in Fig. 23 also apply here.

This case occurs for a number of scattering diagrams that are relevant for us. The singularity curves for $x_{14}=1-\eta$ and $x_{14}=1+\eta$ are shown in Fig. 24(a) and (b), respectively. For $x_{14}=1-\eta$ we have case (iv) described in Appendix A, and for $x_{14}=1+\eta$ we have complex singularities, which are shown in Figs. 24(c)–(e). Figure 24(c) shows the values of x_{13} which correspond to the x_{24} planes of Figs. 24(d) and (e).

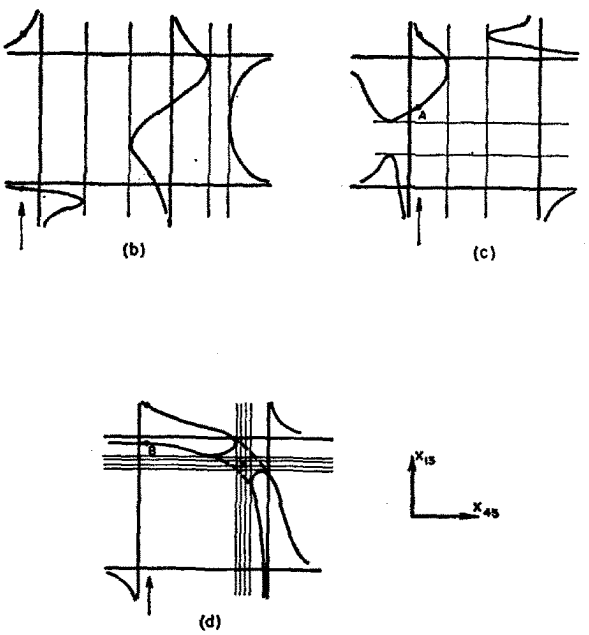


FIG. 25. Examples of scattering curves. These examples apply for the determination of the singular nature of the tangent lines $K_2=0$, in processes (b)–(d), respectively. Figure (d) corresponds to the situation in which two parameters approach instability.

The determination of singularities here can be carried out just as in case (2). These points, however, should be noted. First, the part of Γ_4 that extends to infinity along the asymptote $x_{13}=1$ is singular. Since, as an external mass is increased, the singular points of this curve arise for increasing values of x_{25}, the natural way for placing the associated branch cut is as shown in Fig. 24(d) or (e). Second, the overlapping branch cuts shown in Fig. 24(d) do not hinder the possibility of analytic continuation from one half-plane to the other. Third, the surface Σ_4 disappears as x_{14} is varied from $1-\eta$ to $1+\eta$, and in its place a surface appears that has $\mathrm{Im}x_{13}$ and $\mathrm{Im}x_{24}$ of opposite signs. This surface is in part singular and in part regular, as the analytic behavior of Γ_4 shows. The singular and the regular parts of this surface are separated by the complex branch cuts.

We conclude this appendix with three further examples of scattering curves. We have used these curves for determining the singular nature of some of the tangent lines for processes (b)–(d). These curves are shown in Fig. 25. Figure 25(c) is an example of case (3). The point A is regular for corresponding limits. Figure 25(d) shows the configuration as two invariants, which are in fact unstable for process (d), approach instability. The point B is singular for corresponding limits. The singular nature of these points in the case of opposite limits depends on the location of the complex branch cuts, and the same holds for the two other points marked on diagrams (c) and (d).

Two Folk Lemmas on the Expansion of the S Matrix or the Out Field in Normal Ordered In Fields*

O. W. GREENBERG†
Lawrence Radiation Laboratory,‡ University of California, Berkeley, California
(Received July 27, 1961)

We prove that if the out field or the S matrix is expanded in terms of normal ordered products of the in field, then either the expansion has infinite degree or it is the trivial case $A^{\text{out}}=A^{\text{in}}$, $S=1$. From this fact it follows that any field theory model in which the Heisenberg field (local or not) has a terminating normal ordered expansion in terms of a (generalized) free field cannot provide a nontrivial unitary S matrix.

1. INTRODUCTION

WE prove two folk lemmas[1] concerning the impossibility that (1) a (nontrivial) finite degree expansion of the out field in terms of the in field, or (2) a (nontrivial) finite degree normal ordered expansion of the S matrix in terms of the in field, can satisfy unitarity. The phrase "nontrivial" is inserted to exclude the equivalent trivial possibilities (1) $A^{\text{out}}(x)=A^{\text{in}}(x)$, and (2) $S=1$. These lemmas remain valid if the in and out fields are replaced by a pair of generalized free fields[2] with the same Lehmann weight and the same relativistic no particle state, and the S matrix is replaced by any unitary operator.

From these lemmas it follows that any field theory model in which the Heisenberg field (local or not) has a finite degree normal ordered expansion in terms of a free field or a generalized free field either has $S=1$ or violates unitarity. This situation with respect to unitarity is in contrast with the possibility of constructing finite degree models with nontrivial locality.[2-4]

These lemmas are proved in Sec. 2; remarks about their physical content are made in Sec. 3.

2. IMPOSSIBILITY THAT A NONTRIVIAL FINITE DEGREE NORMAL ORDERED EXPANSION OF A^{out} OR S CAN SATISFY UNITARITY

Lemma 1.[5] If A^{out} and A^{in} both belong to the usual irreducible respresentation of the mass m free-field commutation relations,

$$[A^{\text{in}}(x),A^{\text{in}}(y)]=[A^{\text{out}}(x),A^{\text{out}}(y)]=i\Delta_{m^2}(x-y), \quad (1)$$

$$A^{\text{in}(+)}(x)|0\rangle=A^{\text{out}(+)}(x)|0\rangle=0, \quad (2)$$

then either $A^{\text{out}}=A^{\text{in}}$, or the normal ordered expansion of A^{out} in terms of A^{in} (or vice versa) has infinite degree.

Proof. Equations (1) and (2) imply

$$A^{\text{in}(-)}(x)|0\rangle=A^{\text{out}(-)}(x)|0\rangle. \quad (3)$$

If A^{out} has a finite degree normal ordered expansion in terms of A^{in}, this expansion must have the form

$$A^{\text{out}}(x)=A^{\text{in}}(x)+\sum_{n=2}^{N}\int d^4y_1\cdots d^4y_n \times g^{(n)}(x-y_1,\cdots x-y_n):A^{\text{in}}(y_1)\cdots A^{\text{in}}(y_n): \quad (4)$$

where $g^{(n)}(y_1,\cdots y_n)$ is a (real, for neutral fields) symmetric, Lorentz invariant function of its arguments, and Eq. (3) requires that (a) the leading term be $A^{\text{in}}(x)$ and (b) the Fourier transform $f^{(n)}(k_1,\cdots k_n)$ of $g^{(n)}(y_1,\cdots y_n)$ vanish if all n k_i are in the same cone. Neutrality requires that

$$\bar{f}^{(n)}(-k_1,\cdots -k_n)=f^{(n)}(k_1,\cdots k_n).$$

It is convenient to introduce Fourier transformed fields and to work in momentum space. Then the expansion has the form

$$\tilde{A}^{\text{out}}(k)\delta_m(k) =\tilde{A}^{\text{in}}(k)\delta_m(k)+\sum_{n=2}^{N}(2\pi)^{4n}\int d^4k_1\cdots d^4k_n \times\delta(k-\textstyle\sum_1^n k_i)f^{(n)}(k_1,\cdots k_n):\tilde{A}^{\text{in}}(k_1)\cdots\tilde{A}^{\text{in}}(k_n)\delta_m(k_1)\cdots\delta_m(k_n): \quad (5)$$

where

$$A^{\text{out,in}}(x)=\int d^4k e^{-ikx}\tilde{A}^{\text{out,in}}(k)\delta_m(k),$$

$$\delta_m(k)\equiv\delta(k^2-m^2), \quad \epsilon\delta_m(k)\equiv\epsilon(k)\delta(k^2-m^2),$$

and

$$g^{(n)}(y_1,\cdots y_n)=\int d^4k_1\cdots d^4k_n f^{(n)}(k_1,\cdots k_n) \times\exp(-i\sum_1^n k_jy_j).$$

In momentum space, the requirement of unitarity,

---

* Supported in part by the United States Atomic Energy Commission.

† National Science Foundation Postdoctoral Fellow.

‡ Now at Physics Department, University of Maryland, College Park, Maryland.

[1] These folk lemmas have been proved by many people, but as far as we know do not appear in the literature. The present author proved Lemma 1 in 1956.

[2] O. W. Greenberg, Ann. Phys. **16**, 158 (1961).

[3] A. S. Wightman, "Problèmes mathematique de la théorie quantique des champs," University of Paris Lecture Notes (1957), pp. 57-64.

[4] K. Bardakci and E. C. G. Sudarshan, Nuovo cimento **21**, 722 (1961.).

[5] We consider only neutral scalar fields. However with appropriate changes, such as using anticommutators for Fermi fields rather than commutators, the results obtained can be extended straightforwardly to a finite number of charged and neutral fields with arbitrary spin and internal quantum numbers.

Eq. (1), reads

$$[\tilde{A}^{\text{out}}(p)\delta_m(p),\tilde{A}^{\text{out}}(q)\delta_m(q)] = [\tilde{A}^{\text{in}}(p)\delta_m(p),\tilde{A}^{\text{in}}(q)\delta_m(q)] = (2\pi)^{-3}\epsilon\delta_m(p)\delta(p+q). \quad (6)$$

Our proof consists in inserting the expansion Eq. (5) for $\tilde{A}^{\text{out}}$ into the commutation relation Eq. (6) and showing that because the term in the commutator with $2N-2$ normal ordered $\tilde{A}^{\text{in}}$ operators must vanish, the coefficient $f^{(N)}(k_1,\cdots k_N)$ of the last term in the assumed expansion [Eq. (5)] must vanish. Repetition of this argument leads to the conclusion that $f^{(n)}=0$, $2\leqslant n\leqslant N$, and only the trivial case $\tilde{A}^{\text{out}}(k)\delta_m(k) = \tilde{A}^{\text{in}}(k)\delta_m(k)$ remains. We hope that the simplicity of this argument will not be obscured by the combinatorics associated with $2N-2$ normal ordered operators.

The term with $2N-2$ normal ordered operators in the commutator of the out fields, C_{2N-2}, is

$$C_{2N-2}=(2\pi)^{8N-3}N^2\int d^4p_1\cdots d^4p_{N-1}d^4q_1\cdots d^4q_{N-1} \times f^{(N)}(p_1,\cdots p_{N-1},\ p-\textstyle\sum_1^{N-1}p_i) \times f^{(N)}(q_1,\cdots q_{N-1},\ q-\sum_1^{N-1}q_i)\epsilon\delta_m(p-\sum_1^{N-1}p_i) \times\delta(p-\textstyle\sum_1^{N-1}p_i+q-\sum_1^{N-1}q_i) \times:\prod_{i=1}^{N-1}\tilde{A}^{\text{in}}(p_i)\delta_m(p_i)\tilde{A}^{\text{in}}(q_i)\delta_m(q_i):,$$

where we have made use of the symmetry of $f^{(N)}$; and at least one annihilator and one creator occurs in the normal ordered product. The condition on the function $f^{(N)}$ which follows from the vanishing of the operator C_{2N-2} can be found by taking the appropriate matrix element. We consider $M_{2N-2}\equiv\langle k_1,\cdots k_{N-1}|C_{2N-2}\times|k_N,\cdots k_{2N-2}\rangle$, where

$$|k_1,\cdots k_s\rangle=(s!)^{-\frac{1}{2}}\prod_1^s\tilde{A}^{\text{in}*}(k_i)\delta_m(k_i)|0\rangle.$$

After doing some combinatorics, making use of the symmetry of $f^{(N)}$ to combine terms where possible, and performing the dp and dq integrations with the delta functions which result from the commutators of the in fields, we find

$$M_{2N-2}=(2\pi)^{2N+3}N^2(N-1)!\prod_{n=1}^{2N-2}\theta(k_n)\delta_m(k_n)\delta(p+q-\textstyle\sum_1^{N-1}k_i+\sum_1^{N-1}k_{N-1+i})\{f^{(N)}(k_1,\cdots k_{N-1},\ p-\sum_1^{N-1}k_i)$$

$$\times f^{(N)}(-k_N,\cdots -k_{2N-2},\ q+\textstyle\sum_1^{N-1}k_{N-1+i})\epsilon\delta_m(p-\sum_1^{N-1}k_i)$$

$$+\sum_{\alpha=1}^{N-1}\sum_{\beta=1}^{N-1}f^{(N)}(k_1,\cdots k_{\alpha-1},-k_{N-1+\beta},k_{\alpha+1},\cdots k_{N-1},\ p-\sum_1^{\prime\alpha}k_i+k_{N-1+\beta})$$

$$\times f^{(N)}(-k_N,\cdots -k_{N-1+\beta-1},k_\alpha,-k_{N-1+\beta+1},\cdots k_{2N-2};\ q+\sum_1^{N-1\prime\beta}k_{N-1+i}-k_\alpha)\epsilon\delta_m(p-\sum_1^{N-1\prime\alpha}k_i+k_{N-1+\beta})+\cdots$$

$$+\sum_{\alpha_1>\cdots>\alpha_s=1}^{N-1}\ \sum_{\beta_1>\cdots>\beta_s=1}^{N-1}f^{(N)}(k_{\alpha_{s+1}},\cdots k_{\alpha_{N-1}},-k_{\beta_1},\cdots -k_{\beta_s},\ p-\sum_{s+1}^{N-1}k_{\alpha_i}+\sum_1^s k_{\beta_i})$$

$$\times f^{(N)}(-k_{\beta_{s+1}},\cdots -k_{\beta_{N-1}},k_{\alpha_1},\cdots k_{\alpha_s},\ q+\sum_{s+1}^{N-1}k_{\beta_i}-\sum_1^s k_{\alpha_i})\epsilon\delta_m(p-\sum_{s+1}^{N-1}k_{\alpha_i}+\sum_1^s k_{\beta_i})+\cdots$$

$$+f^{(N)}(-k_N,\cdots -k_{2N-2},\ p+\textstyle\sum_1^{N-1}k_{N-1+i})f^{(N)}(k_1,\cdots k_{N-1},\ q-\sum_1^{N-1}k_i)\epsilon\delta_m(p+\sum_1^{N-1}k_{N-1+i})\}. \quad (7)$$

Here the notation $\sum'^{\alpha}$ means that $i=\alpha$ is omitted from the sum. In the general term, the momenta $(k_1,\cdots k_{N-1})$ are divided into two groups $(k_{\alpha_{s+1}},\cdots k_{\alpha_{N-1}})$ and $(k_{\alpha_1},\cdots k_{\alpha_s})$, where $\alpha_1,\cdots\alpha_{N-1}$ are some permutation of $1,\cdots N-1$, in all possible combinations, but disregarding permutations which do not exchange momenta between the two groups. A similar division is performed on the momenta $(k_N,\cdots k_{2N-2})$. The total number of terms in which both sets of $N-1$ momenta are divided in groups of $N-1-s$ and s is

$$[(N-1)!]^2\left[\frac{(N-1)!}{(N-1-s)!s!}\right]^2,$$

where the factor $[(N-1)!]^2$ represents the number of terms which are equivalent since they differ only by permutation of the first $N-1$ arguments in each $f^{(N)}$. One of these factors $(N-1)!$ is removed by the normalization of the states used in forming the matrix element M_{2N-2}; the other such factor appears as a common factor on the right-hand side of Eq. (7).

From the commutation rules of A^{out}, Eq. (1), we know that

$$M_{2N-2}=0. \quad (8)$$

We will prove that Eq. (8) requires $f^{(N)}=0$. Our first step is to show that the cases with different numbers of the momentum arguments of $f^{(N)}$ in each cone can be

treated separately. In Eq. (7) there are N different sets of terms corresponding to the values $s=0, 1, \cdots N-1$. Each set of terms has as a factor a mass shell delta function whose argument is $p-\sum_{s+1}^{N-1} k_{\alpha_i}+\sum_1^s k_{\beta_i}$. Since for all k_i, $k_i^2=m^2$, $k_i^0>0$, these mass shell delta functions do not, in general, contribute simultaneously for different values of s except on a set of lower dimension in the space of the k_i (i.e., except on a set of measure zero). Sets of measure zero can be neglected since the $f^{(N)}$, like S-matrix elements,[6] must be finite and thus cannot contain delta functions. Thus, effectively, the mass shell delta functions in Eq. (7) isolate terms with different values of s, i.e., with different numbers of the momenta in $f^{(N)}$ in each cone.

This paragraph gives the argument that

$$f^{(N)}(k_1, \cdots k_{N-1}, -k_N)=0,$$

$k_i^0>0$, $1\leqslant i\leqslant N$, for the special case when exactly $N-1$ of the momentum arguments of $f^{(N)}$ are in one cone. Consider the case $s=0$, for which there is just one term. Eliminate q using the four-dimensional momentum conservation delta function, choose $k_1=k_N$, $k_2=k_{N+1}$, $\cdots k_{N-1}=k_{2N-2}$, and use the neutrality condition $f^{(N)}(k_i)=\bar{f}^{(N)}(-k_i)$. Then, dropping irrelevant factors, we find

$$|f^{(N)}(k_1, \cdots k_{N-1}, p-\textstyle\sum_1^{N-1} k_i)|^2 \times\epsilon\delta_m(p-\sum_1^{N-1} k_i)=0. \quad (9)$$

Since $f^{(N)}$ vanishes if all k_i are in the same cone, we choose p so that (a) $p^2=m^2$, $p^0>0$, (b) the delta function in Eq. (9) contributes, and (c) the ϵ gives a negative sign. Since we can obtain any $k_N=-(p-\sum_1^{N-1} k_i)$, $k_N^2=m^2$, $k_N^0>0$ in this way, we conclude that

$$f^{(N)}(k_1, \cdots k_{N-1}, -k_N)=0.$$

Consideration of the next case, in which there are $N-2$ k_i in one cone and 2 in the other, leads us to the terms with $s=1$. Here there are $(N-1)^2$ different terms instead of just one. After repeating the considersations above Eq. (9), we find that the following sum of terms must vanish:

$$\sum_{\alpha=1}^{N-1} |f^{(N)}(k_1, \cdots k_{\alpha-1}, -k_\alpha, k_{\alpha+1}, \cdots k_{N-1}, p-\sum_1^{\prime\alpha} k_i+k_\alpha)|^2\epsilon\delta_m(p-\sum_1^{\prime\alpha} k_i+k_\alpha)$$
$$+\sum_{\alpha\neq\beta=1}^{N-1} f^{(N)}(k_1, \cdots k_\beta, \cdots k_{\alpha-1}, -k_\beta, k_{\alpha+1}, \cdots k_{N-1}, p-\sum_1^{\prime\alpha} k_i+k_\beta)$$
$$\times f^{(N)}(k_1, \cdots k_\alpha, \cdots k_{\beta-1}, -k_\alpha, k_{\beta+1}, \cdots k_{N-1}, p-\sum_1^{\prime\alpha} k_i+k_\beta)\epsilon\delta_m(p-\sum_1^{\prime\alpha} k_i+k_\beta)=0. \quad (10)$$

If the left-hand side of Eq. (10) consisted of a sum, with negative coefficients, of terms of the form $|f^{(N)}|^2$, we could conclude that each $f^{(N)}$ in Eq. (10) vanishes. However the terms in Eq. (10) having the form $f^{(N)}\bar{f}^{(N)}$ with different arguments upset this conclusion and demand further study. Since these terms contain k_α and $-k_\alpha$ in a single $f^{(N)}$, they are a special case of the term with $N-2$ creators and 2 annihilators (or vice versa) which we are now considering.

We can examine the terms we get if we choose, for example, $k_1=k_2$. We then find a sum of terms, including terms of the types

$$|f^{(N)}(k_1, -k_1, k_3, \cdots k_{N-1}, p-\textstyle\sum_3^{N-1} k_i)|^2,$$
$$|f^{(N)}(k_1, k_1, -k_3, k_4, \cdots k_{N-1}, p-\sum_4^{N-1} k_i-2k_1+k_3)|^2,$$
$$f^{(N)}(k_1, k_1, k_3, -k_3, k_5, \cdots k_{N-1}, p-\textstyle\sum_5^{N-1} k_i-2k_1)$$
$$\times\bar{f}^{(N)}(k_1, k_1, k_4, -k_4, k_5, \cdots k_{N-1}, p-\sum_5^{N-1} k_i-2k_1),$$

and

$$f^{(N)}(k_1, k_3, -k_3, k_4, \cdots k_{N-1}, p-\textstyle\sum_4^{N-1} k_i-k_1)$$
$$\times\bar{f}^{(N)}(k_1, k_1, -k_1, k_4, \cdots p-\sum_4^{N-1} k_i-k_1).$$

Notice that the $f^{(N)}$ which previously appeared in a term of the type $f^{(N)}\bar{f}^{(N)}$ appears as $|f^{(N)}|^2$; however there are still terms of the type $f^{(N)}\bar{f}^{(N)}$. These last terms contain $f^{(N)}$ or $\bar{f}^{(N)}$ evaluated at a still more special set of arguments than any of the earlier terms which we have encountered.

Rather than continuing our discussion by setting more and more sets of momenta k_i equal to each other, we go at once to the extreme case and set all the k_i equal to k. We then find the equation

$$(N-1)^2|f^{(N)}(k, \cdots k, -k, p-(N-3)k)|^2 \times\epsilon\delta_m(p-(N-3)k)=0,$$

and conclude that

$$f^{(N)}(k, \cdots k, -k, -k')=0.$$

Having shown that this most special case of $f^{(N)}$ vanishes, we now allow more and more of the k_i to differ and obtain a set of equations in which at each step the "special" terms which do not contain $|f^{(N)}|^2$ have already been shown to vanish so that we can conclude that the $f^{(N)}$ which occur in absolute values squared vanish. Finally we again reach Eq. (10), this time having proved earlier that the terms in the $\sum_{\alpha\neq\beta}$ vanish, and conclude that

$$f^{(N)}(k_1, \cdots k_{N-2}, -k_{N-1}, -k_N)=0.$$

The argument for the other cases $s=2, 3, \cdots N-2$,

[6] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo cimento 6, 319 (1957).

corresponding to the other possible distributions of the k_i between the two cones, proceeds in analogy to that of the case $s=1$ above. Thus we can conclude that

$$f^{(N)}(k_1, \cdots k_{N-1-s}, -k_{N-s}, \cdots -k_N)=0, \quad 0\leqslant s\leqslant N-2,$$

and that the last term in the expansion of Eq. (5) vanishes. Repetition of our entire argument for $n=N-1$, $n=N-2$, $\cdots n=2$, leads us to conclude that only $\tilde{A}^{\text{out}}(k)\delta_m(k)=\tilde{A}^{\text{in}}(k)\delta_m(k)$ is consistent with the commutation relations (i.e., with unitarity) and completes the proof of Lemma 1.

Lemma 2. If the S operator [which relates A^{out} and A^{in} by $A^{\text{out}}(x)=S^{-1}A^{\text{in}}(x)S$] is unitary, then either $S=1$, or the normal ordered expansion of S in terms of A^{in} (or A^{out}) has infinite degree.

Proof. We deduce Lemma 2 as a corollary to Lemma 1. If S is unitary and has a finite degree normal ordered expansion in terms of A^{in} then $A^{\text{out}}(x)=S^*A^{\text{in}}(x)S$ would have a finite degree normal ordered in field expansion. However Lemma 1 excludes this possibility except for the trivial case $A^{\text{out}}=A^{\text{in}}$ which corresponds to $S=e^{i\varphi}1$. The requirement $S|0\rangle=|0\rangle$ fixes $\varphi=0$, which completes the proof of Lemma 2.

From the methods of proof of these lemmas, it is clear that they remain valid for generalized free fields. We state them in a form appropriate for this case.

Lemma 1a. If ϕ_1 and ϕ_2 are generalized free fields which both have the same relativistic no particle state[7]

$$\phi_1^{(+)}(x)|0\rangle=\phi_1^{(s)}(x)|0\rangle=\phi_2^{(+)}(x)|0\rangle =\phi_2^{(s)}(x)|0\rangle=0, \quad (11)$$

and the same Lehmann weight

$$[\phi_1(x),\phi_1(y)]=[\phi_2(x),\phi_2(y)] =i\int da^2\rho(a^2)\Delta_{a^2}(x-y), \quad (12)$$

then either $\phi_2=\phi_1$, or the normal ordered expansion of ϕ_2 in terms of ϕ_1 (or vice versa) has infinite degree.

Lemma 2a. If a unitary operator[8] U relates ϕ_2 and ϕ_1 by $\phi_2(x)=U^{-1}\phi_1(x)U$ then either $U=1$, or the normal ordered expansion of U in terms of ϕ_1 (or ϕ_2) has infinite degree.

Finally, these lemmas provide a proof that any field theory model[2-4] in which the Heisenberg field has a finite degree normal ordered expansion in terms of a (generalized) free field cannot have a non-trivial unitary S matrix. No assumption about the locality of the Heisenberg field is necessary for this conclusion.

3. REMARKS

We make some remarks on the physical content of these lemmas. If Lemma 2 were not true then it would be possible that there be, for example, elastic scattering between pairs of particles, but, for n greater than some finite N, no contribution at all to the n-particle elastic scattering amplitude in which each particle scatters elastically with all the other particles. In fact, there would be no contribution at all to the n-particle scattering amplitude coming from a totally connected Feynman diagram. Although in practice such a contribution from totally connected diagrams might be small, it should not be absent entirely. On intuitive grounds, or on the basis of perturbation theory and Feynman diagrams, we expect that n incoming particles must at least produce that elastic scattering which would result from all combinations of the elastic scattering between all pairs of incoming particles. Thus Lemma 2 seems obvious intuitively.

Since the demonstrations of lemmas 1 and 2 require no statements about the interpolating Heisenberg field, these lemmas are independent of the assumption of locality. It is an open question whether an S matrix which allows only a finite set of intrinsic processes is consistent with locality; clearly such a possibility *is* consistent with unitarity alone. For example, an S operator of the form $S=e^{i\eta}$,

$$\eta=\int d^4k_1\cdots d^4k_4\delta(k-\textstyle\sum_1^4 k_i)f(k_1,\cdots k_4) \times:\tilde{A}^{\text{in}}(k_1)\delta_m(k_1)\cdots\tilde{A}^{\text{in}}(k_4)\delta_m(k_4):, \quad (13)$$

where $f(k_i)=\bar{f}(-k_i)$, and f is totally symmetric, leads to elastic scattering only. Such an S operator is unitary since η has been chosen Hermitian. Lemma 2 is not violated since the normal ordered expansion of this S operator does not terminate. We do not know whether local field theory allows such an S matrix. It is interesting that the elastic scattering amplitude which follows from Eq. (13) cannot have the form of the Mandelstam representation.[9]

ACKNOWLEDGMENTS

We thank Professor A. S. Wightman for suggesting the consideration of Lemma 1 and for several helpful discussions, and Professor P. G. Federbush, Professor M. T. Grisaru, and Professor F. E. Low for asking critical questions about this work. It is a pleasure to thank Professor G. F. Chew for the hospitality of the Lawrence Radiation Laboratory.

[7] $\phi^{(s)}(x)$ is the Fourier transform of $\tilde{\phi}(k)$ restricted to space-like momenta.

[8] In reference 2 it was proved that Eqs. (11) and (12) uniquely determine all the vacuum expectation values of a generalized free field. Therefore there exists such a unitary operator relating ϕ_2 and ϕ_1.

[9] We are indebted to M. Froissart and A. Scotti for this remark.

Hamiltonian Formulation of Action-at-a-Distance in Electrodynamics

EDWARD H. KERNER
*Physics Department, University of Buffalo, Buffalo, New York**
(Received March 7, 1961)

The Wheeler-Feynman scheme of classical electrodynamics, using half-retarded+half-advanced fields between pairs of interacting charges, is written in a differential form which expresses the motion of the charges at any inertial observer's single present time through infinite-order differential equations of motion. The conversion of field variables to mechanical ones in this way does *not* imply an infinite number of degrees of freedom; rather it is shown how to construct a Hamiltonian, depending solely on ordinary particle-position coordinates and certain canonically conjugate momenta, which is both the energy and the generator of the "correct" motions. The latter are taken to be those continuously developable from free-particle motions as the strength of interaction increases from the value zero. It is just this criterion that establishes particle-position+conjugate momentum to be sufficient for the description of motion. The highness of the order of the equations of motion apparently gets to be expressed through high powers of momenta in the Hamiltonian. An example is sketched, illustrating a general algorithm, in which the particles are treated nonrelativistically and their interactions are treated to order e^4 and c^{-4}, corresponding to fourth-order equations of motion. The general form of the Hamiltonian is a double power series expansion in e^2 and c^{-1}; when the c^{-1} expansion is terminated covariance is sacrificed, but the possibility of collecting all powers of e^2 is gained.

INTRODUCTION

THE idea of action-at-a-distance in electrodynamics goes back many years, but not until recent times has it been put into a complete and definitive form by Wheeler and Feynman.[1] These authors showed that the usual scheme of: purely retarded interactions between charges, plus the Lorentz-Dirac self-force for the individual charges describing radiation damping, is equivalent to: half-retarded plus half-advanced interactions amongst charges, together with similar interactions between the charges and a perfect absorber surrounding them. When advanced and retarded fields emanating from any charge onto another are represented by the Lienard-Wiechert potentials, the scheme of Wheeler-Feynman becomes a pure particle dynamics responding directly to the premise that fields, after all, are just auxiliaries introduced to describe forces on charges. It is expressible through a single action principle, due to the over-all symmetry between past and future. In effect the dissipative character of pure retardation and radiation damping is brought under a regime that, owing to the absorber, is conservative on the whole. One looks then only at the motions of charges, never of fields. "The" field as such is in fact without meaning. And there is no such thing as self-interaction.

We should like to sketch here, in purely classical terms, a way to a Hamiltonian formulation of the Wheeler-Feynman theory. The main point is a showing that, despite the customary attitude that interacting charges require for a description of their motions an infinite number of degrees of freedom, *it is quite sufficient to deal solely with the ordinary position coordinates of the charges and certain momenta canonically conjugate to them.* This general viewpoint goes back several years and has recently been broached again,[2] but thus far without providing any comprehensive formulation of electrodynamics and without reference to Wheeler-Feynman's theory, which is the only one as yet capable of explicitly and satisfactorily supporting action-at-a-distance notions. The discussion will be limited in its practical aspects to a nonrelativistic treatment of particle kinetic energies and to the beginning terms of an infinite series expressing through mechanical variables the interaction of charges. This much is illustrative, but what procedurally is embodied in it has general import and in itself it has seemed convincing already with regard to the full problem. Beyond this we present a very simple physical argument for the quite general validity of the mechanized electrodynamics; a conclusive mathematical proof is still wanting.

With physical actions propagated at a finite velocity, the paradox of the italicized statement needs explanation. How, after all, can a single time description of the relativistic many-body problem and a strictly Newtonian outlook that limits itself to specifying all the motion, through only a statement of initial particle positions and momenta be able to cope with delayed (or advanced) effects for whose accounting during the finite time interval between radiation transfers the whole edifice of field theory had to be specifically invented?

The answer is broadly as follows. The Lienard-Wiechert potentials at charge e_1 due to charge e_2, as has been known for years and as will be shown shortly, can be written by means of a Taylor expansion in terms of the present position $\mathbf{r}_2(t)$ of e_2 together with

* The preliminary stages of this work were started at the Brookhaven National Laboratory, to whom thanks are due for the opportunity of a visit.

[1] J. A. Wheeler and R. P. Feynman, Revs. Modern Phys. **21**, 425 (1949).

[2] P. A. M. Dirac, Revs. Modern Phys. **21**, 392 (1949); L. H. Thomas, Phys. Rev. **85**, 868 (1952); L. H. Thomas and B. Bakamjian, Phys. Rev. **92**, 1300 (1953); P. Havas and J. Plebanski, Bull. Am. Phys. Soc. **5**, 433 (1960); B. Bakamjian, Phys. Rev. **121**, 1849 (1961); L. L. Foldy, Phys. Rev. **122**, 275 (1961).

all its derivatives $d^n\mathbf{r}_2(t)/dt^n$. Thence, using a "differential" form of Wheeler-Feynman's action principle, the motions of both charges can be specified through a joint Lagrangian containing all derivatives of both positions. In this way, or else directly, the equations of motion are set forth as infinite-order differential equations; the infinitude of "field" coordinates are seemingly converted to an infinity of mechanical ones. So to speak, the high derivatives and fine curvatures of orbits record at one instant of time the finitely delayed and advanced effects and electrodynamics is made over into just dynamics. Thereafter it is easy to construct an integral of motion identifiable as the energy. This can indeed be cast into Hamiltonian form using the century-old method of Ostrogradsky[3] for higher-order equations of motion, in which $\mathbf{r}_i$, $\dot{\mathbf{r}}_i$, $\ddot{\mathbf{r}}_i$, $\cdots$ are reckoned as generalized coordinates and a system of canonically conjugate momenta $\mathbf{P}_i^{(0)}$, $\mathbf{P}_i^{(1)}$, $\mathbf{P}_i^{(2)}$, $\cdots$ is constructed.

The number of degrees of freedom remains as large as ever. But the scheme, though formally complete, is unsatisfactory: (a) Far too broad a class of motions is encompassed, (b) if we glance ahead to the simplest questions of quantization we get the intolerable dilemma that $\mathbf{r}_i$, $\dot{\mathbf{r}}_i$, $\ddot{\mathbf{r}}_i$, $\cdots$ are all instantaneously and accurately specifiable (so the whole motion is accurately specifiable), notwithstanding the indeterminacy of the $\mathbf{P}_i^{(n)}$ (these are not independent remarks). The quantum theory of higher-order equations of motion is in fact nearly totally obscure.[4]

Under (a) we must consider that as e_1e_2 is allowed to vanish, we must end up with free-particle motions forming a twelve-parameter family $\mathbf{r}_i(0)$, $\mathbf{p}_i(0)$; conversely, if we turn on e_1e_2 starting from the value zero, we can ask for those motions which alone are continuously developable from the free-particle ones and are perforce a twelve-parameter set, not an infinite-parameter one. It is clear that the Ostrogradsky scheme, and the equations of motion, are over-complete. The correct motion must of course satisfy these equations but not every solution of them can be admitted as a physically realizable motion. A similar over-completeness probably marks the usual field-theory description of motion.

How can we generate the 'correct' motions and only those? The power of Hamiltonian methods is far greater than what is spanned in Ostrogradsky's method and we shall try to indicate how they may be equal to this question. Notice first the example of the properties of the simple classical $H(x,p)=ap^2+bp^3+V(x)$. The Poisson bracket (x,H) gives $\dot{x}$ as a quadratic in p, and then $(\dot{x},H)$ gives $\ddot{x}$ as linear in p. This removes the ambiguity in the physical meaning of p, giving by $p(\ddot{x})=(p,H)$ a well-defined *third*-order differential equation for $x(t)$ instead of the multiple second-order equations coming from the branches of $p=p(\dot{x})$. Yet, as always, H is the generator of infinitesimal contact transformations evolving $x(t)$, $p(t)$ from only the *pair* of initial values $x(0)$, $p(0)$. Thus, the high order of an equation of motion does not have to signify more degrees of freedom but may only mean higher powers of p in the Hamiltonian. Next, observe how Hamiltonian methods may be brought into the practical example of the nonrelativistic linear motion of a charge in a prescribed force-field with radiation damping: $-\alpha d^3x/dt^3+m\ddot{x}=-\partial V/\partial x$ $(\alpha=\frac{2}{3}e^2/c^3)$. Bhabha[5] has remarked that the runaway solutions are characterized by an essential singular behavior at $\alpha=0$, so that the meaningful motions are those marked as Maclaurin-expansible in α. This is an invitation to attempt a Hamiltonian $H_0+\alpha H_1+\alpha^2H_2+\cdots$. Then if derivatives are calculated as Poisson brackets, the equation of motion itself dictates a succession of partial differential equations for H_1, H_2, $\cdots$, starting from $H_0=p^2/2m+V(x)$ (for instance $m(((x,H_0),H_1)+((x,H_1),H_0))=(((x,H_0),H_0),H_0)$); these are solvable as power series in p and in other ways.[6]

Now coming back to the main problem, the unambiguously defined energy of the system of two charges can be written directly in terms of particle positions and their derivatives, $E=E_0(\mathbf{v}_1,\mathbf{v}_2)+e_1e_2E_1\equiv E_0+\epsilon E_1$. Here E_1, the part involving the high derivatives, proceeds as a power series in c^{-1} The opening terms are the Coulomb (c^0) and Darwin (c^{-2}) interactions. As these depend only upon positions and velocities, this much of E, taken by itself, can be cast into Hamiltonian form. Is there a systematic way for so casting the next (c^{-4}) term and all terms, in short to extend the Darwin interaction to arbitrarily high precision? (This was the starting point of the present work.) To see that an affirmative answer is possible, we assume at once that the correct Hamiltonian must be $H_0+\epsilon H_1+\epsilon^2H_2+\cdots$ so as to guarantee a motion continuously connected to free-particle motion. Then any derivative, including particularly $\mathbf{v}_i$, can be written via Poisson brackets as a series in ϵ. The H_k are all of them $H_k(\mathbf{r}_i,\mathbf{p}_i)$, where $\mathbf{p}_i$ has no other meaning than that it is the canonical mate to $\mathbf{r}_i$. The statement that $E=E_0+\epsilon E_1$ then writes itself as a succession of inhomogeneous differential equations coupling H_k to itself and to the earlier H's, beginning with the known H_0. These are all basically very simple, owing to E_0's simplicity, and

[3] E. T. Whittaker, *Analytical Dynamics* (Cambridge University Press, New York, 1937), 4th edition, Chap. X.

[4] A. Pais and G. E. Uhlenbeck, Phys. Rev. **79**, 145 (1950); W. Waldmann, Z. Naturforsch. **8a**, 329 (1953); P. Caldirola, Nuovo cimento Suppl. **3**, 297 (1956).

[5] H. J. Bhabha, Phys. Rev. **70**, 759 (1946).

[6] There is a considerable variety of equations of motion, including dissipative ones, as above, which may be cast into Hamiltonian form directly, without the mediation of a Lagrangian (which may not exist). The higher Hamiltonians H_1, H_2, $\cdots$ above each involve all derivatives of $V(x)$, signifying a 'nonlocal' character of the motion that accords with Dirac's remark [P. A. M. Dirac, Proc. Roy. Soc. (London) **A167**, 148 (1938)] that the regular motions need a specification of initial position and velocity and some *final* acceleration.

give H_k as successively higher degree polynomials in $\mathbf{p}_1, \mathbf{p}_2$.

There results finally a double-power-series expansion of H according to powers of both ϵ and c^{-1}. What must be asked of the Hamiltonian is that it be not only the energy, which it is by construction, but the generator of motion. By explicit calculation of H to order ϵ^2 and c^{-4} and even beyond, we find an identity in $\mathbf{r}_i$, $\mathbf{p}_i$ up to terms of similar order in the raw equations of motion when derivatives in it are calculated Poisson bracket-wise from H. The intricacy of this calculation is appreciable; enough so, that it becomes hard to believe that it could be only accidentally correct and that in higher orders the scheme will somehow fail. The internally consistent way by which derivatives in H are self-calculated would seem to be some warrant that these same derivatives placed into the equations of motion upon which H is itself founded will come to satisfy these equations. Indeed the physical identity of H with the Ostrogradsky Hamiltonian $H^{\text{ost}}(\mathbf{r}_i, \dot{\mathbf{r}}_i, \cdots; \mathbf{P}_i^{(0)}, \mathbf{P}_i^{(1)}, \cdots)$ turns out virtually to assure this in a general fashion.

The Newtonian and Hamiltonian viewpoint we are taking gives in principle a complete description of what the charges are doing at any inertial observer's single present time. The noncovariance of simultaneity is no hindrance at all so long as the observer's frame is arbitrary; over-all covariance is in fact built into the theory, even if it is not very manifest. Upon truncation of the c^{-1} series, covariance is of course lost. But there is a compensation for this: It is the possibility of collecting all orders of ϵ; for the double series in ϵ and c^{-1} can be rearranged as $\sum c^{-2n} \times$ (finite polynomial in ϵ). This nonrelativistic attitude, here and elsewhere, could be of practical use insofar as it relies not so much on the smallness of ϵ but rather the largeness of c. Of course, increasing the strength ϵ of interaction, and hence the particle velocities, is an order to keep more powers of c^{-1}, but it will be seen that the successive powers of c^{-1} are accompanied by powers of (particle masses)$^{-1}$, in effect being an expansion in the charge radii $e_i^2/m_i c^2$. On the other hand, it appears likely that the truncation of the series in ϵ, with retention of all powers of c^{-1}, is step-by-step covariant, as in quantum electrodynamics. No assertion is made regarding convergence of any series, though there is reason to question seriously any ϵ expansion no matter how arranged. This does not vitiate the requirement that the motion and the Hamiltonian be ϵ expansible so as to have the free-particle limit in view, but only raises the question of what the series sums up to and what its convergence domain may be.

There are interesting points in considering the translation of the differential action-at-a-distance scheme into quantum theory, about which our remarks in this primarily classical exposition must be counted as tentative. It is first of all fairly clear that, without great ambiguity, the translation is basically feasible owing to the central role given to Poisson brackets as the agent for computing time derivatives; that is, the translation is from classical to quantal differential calculus, from Poisson-bracket to commutator. In fact the classical Hamiltonians we shall produce as examples are transcribable as quantal operators, even after the use of classical Poisson brackets, when attention is paid to questions of ordering of classical factors and of hermiticity. The question of how to deal with spinning charges rather than spinless ones, or with both together, seems not insuperable according to preliminary study. It turns out to be possible to set up the energy in a way that makes it natural to extract a Dirac-type square root for the spinning charge and that allows clear identification of the unperturbed Hamiltonian.

On the nonrelativistic view which builds upon the series in c^{-1}, we are taking as objects of study complete dynamical orbits, like Kepler orbits but corrected to any specified order for high velocities, and not the fits and jumps compelled by strict covariance with its vast array of 'photons'; this is, in quantization, a rending of $e^2/\hbar c$ into separate pieces. We attain on this basis only a kind of asymptotic approach to covariance, though still reserving the alternative of the expansion in ϵ with collection of all powers of c^{-1}.

In certain problems, like the elastic, nonradiative collision of two charges, the Wheeler-Feynman scheme of half-retarded, half-advanced potentials should entirely suffice, for then the charges are in effect set out as complete absorbers for each other. Explicitly 'radiative' processes like the Compton effect should be similarly analyzable by examining the motion of a charge e_s that is the source of the 'incident photon,' and a charge e_a that absorbs the 'scattered photon,' e_s and e_a then comprising a sufficient Wheeler-Feynman absorber. That is, external photon lines, like Faraday lines of force, are to begin and end on charges, certain of which are labeled "absorber." It can be seen that such an effect as light-light scattering, put down as a "vacuum-polarization" effect, already has classical meaning in the action-at-a-distance theory. Plainly at least four charges must be surveyed; a source e_s and absorber e_a, and a second source $e_{s'}$ and absorber $e_{a'}$, all separated by large distances. Looking at the charges instead of the light, it is clear that the $e_s - e_a$ interaction cannot be independent of the $e_{s'} - e_{a'}$ one, but that all pairs must be taken to be in interaction. The "fields" for $e_s - e_a$ transmission and for $e_{s'} - e_{a'}$ transmissions when traced back to their sources, cannot be separately considered and thought to glide harmlessly past each other, for the cross-coupling between e_s or e_a and $e_{s'}$ or $e_{a'}$, which must be admitted if the $e_s - e_a$, $e_{s'} - e_a$, ones are, means e_s, e_a get shaken by $e_{s'}$'s and $e_{a'}$'s motions, and vice versa, so each transmission is affected by the other, which is only a wordy action-at-a-distance way of saying that light interacts with light; the shake involves $e_a e_{s'}$ and then gets incorporated into the otherwise private lines of the $e_s - e_a$ and $e_{s'} - e_{a'}$

transmissions, which themselves proceed by an amount $e_a e_a$ and $e_{a'} e_{a'}$, so the whole effect is of order e^4 at least. The four photon lines in the well-known lowest-order square diagram representation of this process simply have to go *somewhere*.

INSTANTANEOUS EQUATIONS OF MOTION

As already indicated we shall consider only two charges, the pair being the basic element in the complete Wheeler-Feynman theory. The absorber charges can be brought in ultimately once the time-symmetrical interaction of a single pair is understood. In short, we deal just with Fokker's[7] electrodynamics, the precursor to Wheeler-Feynman's.

Consider the retarded scalar potential at position $\mathbf{r}_1$, time t produced by e_2 located at $\mathbf{r}_2(t)$:

$$\varphi(\mathbf{r}_1,t)=\int \frac{\rho_2(\mathbf{r}_2,\, t-r_{12}/c)}{r_{12}}d\mathbf{r}_2.$$

Expand in a Taylor series about the present time t, supposing the motion to be sufficiently regular to do this:

$$\varphi(\mathbf{r}_1,t)=\int \frac{d\mathbf{r}_2}{r_{12}}\sum\frac{(\partial/\partial t)^n}{n!}\rho_2(\mathbf{r}_2,t)(-r_{12}/c)^n$$

$$=e_2\sum\frac{(-D_2)^n}{n!c^n}r_{12}{}^{n-1}.$$

Here D_2 signifies time-differentiation of only $\mathbf{r}_2$, for example

$$D_2 r_{12}=\frac{d}{dt}|\mathbf{r}_1-\mathbf{r}_2(t)|=\frac{-\mathbf{v}_2\cdot\mathbf{r}}{r}(\mathbf{r}\equiv\mathbf{r}_1-\mathbf{r}_2).$$

The order of $\sum$ and $\int$ has been reversed and ρ_2 written as $e_2\delta[\mathbf{r}_2(t)]$. This is but another way[8] of writing the usual Lienard-Wiechert $e_2/(r-(\mathbf{v}_2/c)\cdot\mathbf{r})_{\text{ret}}$. For the advanced potential, replace $-D_2$ by $+D_2$, and for the half-retarded plus half-advanced potential keep only the even powers of $D_2{}^n$. In the same way, the retarded vector potential is

$$\mathbf{A}(\mathbf{r}_1,t)=\frac{e_2}{c}\sum\frac{(-D_2)^n}{n!c^n}\mathbf{v}_2 r_{12}{}^{n-1}.$$

Now at $\mathbf{r}_1$ place e_1. The Lagrangian for e_1's motion is, with half-retarded+half-advanced fields from e_2,

$$L_1(\mathbf{r}_1,\mathbf{v}_1,t)=-m_1c^2\left(1-\frac{\mathbf{v}_1{}^2}{c^2}\right)^{\frac{1}{2}}$$

$$-e_1e_2\sum_{p=0}^{\infty}\frac{D_2{}^{2p}}{2p!c^{2p}}\left(1-\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}\right)r^{2p-1}.$$

Reversing the roles of 1 and 2 gives e_2's Lagrangian similarly. The gauge condition $\operatorname{div}\mathbf{A}+(1/c)(\partial\varphi/\partial t)=0$ is maintained term-by-term as written, but not if $\mathbf{v}_1\cdot\mathbf{v}_2 r^{2p-1}/c^{2p+2}$ is lumped with r^{2p+1}/c^{2p+2} into one term of the same order in c^{-1}. The lumping is not useless. To secure it, note that the operator of total differentiation $D\equiv d/dt$ is identical to D_1+D_2. Then $D_2{}^{2p}=D_2{}^{2p-1}\times(D-D_1)=D_2{}^{2p-2}(-D_1D_2)+$an exact derivative which in L_1 may be dropped. Since

$$D_1D_2r^{2p-1}=-(2p-1)r^{2p-3}\mathbf{v}_1\cdot\mathbf{v}_2$$

$$-(2p-1)(2p-3)r^{2p-5}\mathbf{v}_1\cdot\mathbf{r}\mathbf{v}_2\cdot\mathbf{r},$$

a Lagrangian dynamically equivalent to L_1 is

$$L_1'=-m_1c^2\left(1-\frac{\mathbf{v}_1{}^2}{c^2}\right)^{\frac{1}{2}}-\frac{e_1e_2}{r}+e_1e_2\sum D_2{}^{2p}\omega_{2p}$$

$$\omega_{2p}\equiv\frac{1}{2p!c^{2p+2}}\left(\frac{2p+1}{2p+2}\mathbf{v}_1\cdot\mathbf{v}_2r^{2p-1}-\frac{2p-1}{2p+2}\mathbf{v}_1\cdot\mathbf{r}\mathbf{v}_2\cdot\mathbf{r}r^{2p-3}\right),$$

and similarly for L_2'.

Many mathematical questions arise: How far can the Taylor series be trusted to converge and properly to translate field variables into mechanical ones? What meaning would there be to breaking off the infinite series, i.e., how much of the true character of the motion is approachable? How is it approached when the high derivatives are included only in succession? These and other questions have no really clear answers, though there is indication that the Taylor expansion might be viewed as a temporary expedient whose use can perhaps be ameliorated eventually by summing the series in some sense. When r_{12}/c is sufficiently small, corresponding to physically interesting conditions, the the term-wise consideration of the series must be expected to be useful. We can but proceed heuristically at this point, acting as though we had mathematical legitimacy but remembering that we do not.

The individual Lagrangians are to give way to a single joint Lagrangian describing the motions of both charges together. Using the basic rule $D=D_1+D_2$,

$$D_2{}^{2p}=D_2{}^p(D-D_1)^p=(-D_1D_2)^p+\text{terms in } D,$$

so that

$$L_1=-m_1c^2\left(1-\frac{\mathbf{v}_1{}^2}{c^2}\right)^{\frac{1}{2}}-e_1e_2\sum\frac{(-D_1D_2)^p}{2p!c^{2p}}$$

$$\times\left(1-\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}\right)r^{2p-1}+\text{exact derivatives.}$$

It is not entirely obvious that exact derivatives may be freely discarded, but it is not difficult to prove this by induction on n when $D_2{}^{2p}$ in L_1 is replaced by $D_2{}^{2p-n}\times(-D_1)^n$; in fact $D_2{}^{2p}$ can be so replaced for all $n\leq 2p$. Because of the symmetry of the interaction term

[7] A. D. Fokker, Z. Physik 58, 386 (1929).
[8] L. Page, Phys. Rev. 24, 296 (1924).

in e_1e_2, a joint Lagrangian clearly is

$$L=-m_1c^2\left(1-\frac{\mathbf{v}_1^2}{c^2}\right)^{\frac{1}{2}}-m_2c^2\left(1-\frac{\mathbf{v}_2^2}{c^2}\right)^{\frac{1}{2}}$$
$$-e_1e_2\sum\frac{(-D_1D_2)^p}{2p!c^{2p}}\left(1-\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}\right)r^{2p-1}.$$

Equally good is

$$-m_1c^2\left(1-\frac{\mathbf{v}_1^2}{c^2}\right)^{\frac{1}{2}}-m_2c^2\left(1-\frac{\mathbf{v}_2^2}{c^2}\right)^{\frac{1}{2}}$$
$$-e_1e_2\sum\frac{\frac{1}{2}(D_1^{2p}+D_2^{2p})}{2p!c^{2p}}\left(1-\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}\right)r^{2p-1}$$

or still other joint Lagrangians, including L' that follows from a similar symmetrization of L_1'. The equations of motion

$$\frac{\partial L}{\partial \mathbf{r}_i}-D\frac{\partial L}{\partial \mathbf{v}_i}+D^2\frac{\partial L}{\partial \dot{\mathbf{v}}_i}-\cdots=0$$

are just the same, but for a rearrangement, as those from the individual L_i, $\partial L_i/\partial \mathbf{r}_i-D\partial L_i/\partial \mathbf{v}_i=0$.

These considerations are not new. It turns out that Hargreaves[9] discussed them more than forty years ago in a remarkable and unremarked paper that anticipated Fokker's and others' action-at-a-distance ideas.

The Wheeler-Feynman variational principle is now readily stated in the present language when the absorber charges also are gathered as above, pair-wise with each other, and with e_1, e_2 under one grand Lagrangian.

The invariance of L against time and space displacements and space rotations must have the usual conservation laws as consequences. First we make the transition to a Hamiltonian by Ostrogradsky's[3] method for a Lagrangian containing higher derivatives. Let p in $\sum_p$ extend only to n for a moment, so that L has derivatives up to $\mathbf{v}_1^{(n)}$, $\mathbf{v}_2^{(n)}$. Then the canonically conjugate pairs

$$\mathbf{r}_i,\ \mathbf{P}_i^{(0)}=\frac{\partial L}{\partial \mathbf{v}_i}-D\frac{\partial L}{\partial \dot{\mathbf{v}}_i}+\cdots+(-1)^nD^n\frac{\partial L}{\partial \mathbf{v}_i^{(n)}},$$

$$\mathbf{v}_i,\ \mathbf{P}_i^{(1)}=\frac{\partial L}{\partial \dot{\mathbf{v}}_i}-D\frac{\partial L}{\partial \ddot{\mathbf{v}}_i}+\cdots+(-1)^{n-1}D^{n-1}\frac{\partial L}{\partial \mathbf{v}_i^{(n)}},$$

$$\cdots$$

$$\mathbf{v}_i^{(n-1)},\ \mathbf{P}_i^{(n)}=\frac{\partial L}{\partial \mathbf{v}_i^{(n)}},$$

provide the Hamiltonian

$$H_n^{\text{ost}}=\mathbf{P}_1^{(0)}\cdot\mathbf{v}_1+\mathbf{P}_1^{(1)}\cdot\dot{\mathbf{v}}_1+\cdots+\mathbf{P}_1^{(n)}\cdot\mathbf{v}_1^{(n)}$$
$$+\mathbf{P}_2^{(0)}\cdot\mathbf{v}_2+\mathbf{P}_2^{(1)}\cdot\dot{\mathbf{v}}_2+\cdots+\mathbf{P}_2^{(n)}\cdot\mathbf{v}_2^{(n)}-L,$$

where $\mathbf{v}_1^{(n)}$, $\mathbf{v}_2^{(n)}$ are to be introduced in terms of $\mathbf{P}_2^{(n)}$, $\mathbf{P}_1^{(n)}$ through $\mathbf{P}_i^{(n)}=\partial L/\partial \mathbf{v}_i^{(n)}$. That H_n^{ost} is independent of time means energy conservation, $H_n^{\text{ost}}=\text{const}$. The equations of motion

$$\frac{d\mathbf{P}_1^{(0)}}{dt}=\frac{\partial L}{\partial \mathbf{r}_1}=\mathbf{F}(1,2),\quad \frac{d\mathbf{P}_2^{(0)}}{dt}=\frac{\partial L}{\partial \mathbf{r}_2}=\mathbf{F}(2,1)$$

give $\mathbf{P}_1^{(0)}+\mathbf{P}_2^{(0)}=\textbf{const}$, owing to $\mathbf{F}(1,2)$ being $-\mathbf{F}(2,1)$ which follows simply from the symmetry of L. This rejuvenates the Newtonian equality of action and reaction, though the "force" $\mathbf{F}$ is noncentral. What then generalizes angular-momentum conservation is

$$\mathbf{r}_1\times\mathbf{P}_1^{(0)}+\mathbf{r}_2\times\mathbf{P}_2^{(0)}+\sum_0^{n-1}\mathbf{v}_1^{(i)}\times\mathbf{P}_1^{(i+1)}$$
$$+\mathbf{v}_2^{(i)}\times\mathbf{P}_2^{(i+1)}=\textbf{const}$$

as is provable by induction.

Notice that H_n^{ost} has a term $\sim\mathbf{P}_1^{(n)}\cdot\mathbf{P}_2^{(n)}/e_1e_2$ which nearly forbids any useful study of the vitally important limit $e_1e_2\to 0$. Nonetheless the extension $n\to\infty$,

$$H_\infty^{\text{ost}}=\sum_0^\infty\mathbf{P}_1^{(n)}\cdot\mathbf{v}_1^{(n)}+\mathbf{P}_2^{(n)}\cdot\mathbf{v}_2^{(n)}-L,$$

is formally feasible, which dissolves this difficulty, but creates the new one that for $e_1e_2\to 0$ the system of two noninteracting particles still gets to be described by an infinite number of degrees of freedom. One supposes here that in the present mechanical language the 'vacuum field' of field theory has made its appearance. The apparition is dispersed by the merest glance at L, which shows only $\mathbf{P}_1^{(0)}$, $\mathbf{P}_2^{(0)}$ to be nonvanishing. However, for the elaboration of electrodynamics by this Ostrogradsky scheme, or by its field-theoretic counterpart, this glance is not at all permissible; the ghost must be accepted at least on a provisional basis. All the same, H_∞^{ost} identifies physically what is the conserved energy, its mathematical representation through a specious set of canonical variables being set aside.

HAMILTONIZATION

The energy may be constructed a little differently but equivalently. Connected to the separate Lagrangians L_1, L_2 are separate Hamiltonians $H_i=\mathbf{\Pi}_i\cdot\mathbf{v}_i-L_i$, $\mathbf{\Pi}_i\equiv\partial L_i/\partial \mathbf{v}_i$ for the motion of each charge due to the fields of the other, taken as prescribed fields. Then the calculation of the rate of change of H_1+H_2 gives

$$D(H_1+H_2)$$
$$=(d\mathbf{\Pi}_1/dt)\cdot\mathbf{v}_1+\mathbf{\Pi}_1\cdot\dot{\mathbf{v}}_1-\left(\frac{\partial L_1}{\partial \mathbf{r}_1}\cdot\mathbf{v}_1+\frac{\partial L_1}{\partial \mathbf{v}_1}\cdot\dot{\mathbf{v}}_1+D_2L_1\right)$$
$$+(d\mathbf{\Pi}_2/dt)\cdot\mathbf{v}_2+\mathbf{\Pi}_2\cdot\dot{\mathbf{v}}_2-\left(\frac{\partial L_2}{\partial \mathbf{r}_2}\cdot\mathbf{v}_2+\frac{\partial L_2}{\partial \mathbf{v}_2}\cdot\dot{\mathbf{v}}_2+D_1L_2\right)$$

[9] D. Hargreaves, Proc. Cambridge Phil. Soc. 22, 191 (1917).

or, because of the equations of motion $d\mathbf{\Pi}_i/dt=\partial L_i/\partial \mathbf{r}_i$,

$$D(H_1+H_2)=-(D_2L_1+D_1L_2)$$
$$=e_1e_2\sum (D_1^{2p+1}+D_2^{2p+1})\frac{(1-\mathbf{v}_1\cdot\mathbf{v}_2/c^2)}{2p!c^{2p}}r^{2p-1}.$$

Now $Z_{2p+1}\equiv D_1^{2p+1}+D_2^{2p+1}$ is an exact derivative,

$$Z_1=D_1+D_2=D$$
$$Z_3=D_1^2(D-D_2)+D_2^2(D-D_1)=D(Z_2-\Delta)$$
$$\cdots$$
$$Z_{2p+1}=D(Z_{2p}-\Delta Z_{2p-2}+\Delta^2Z_{2p-4}-\cdots+(-1)^p\Delta^p)$$
$$\equiv D(Z_{2p}-Y_{2p})$$

where $\Delta\equiv D_1D_2$, so that it is meaningful to write for H

$$H=H_1+H_2-e_1e_2\sum (Z_{2p}-Y_{2p})\frac{(1-\mathbf{v}_1\cdot\mathbf{v}_2/c^2)}{2p!c^{2p}}r^{2p-1}.$$

It is not difficult to show with a little calculation that H^{ost} reduces to this when the $\mathbf{P}_i^{(n)}$ are written out properly, or directly when DL is suitably calculated from $L=L_1+(L-L_1)=L_2+(L-L_2)$. It is tempting to try to associate the 'private' derivatives D_1, D_2 with the 'private' Hamiltonians H_1, H_2 and to calculate D_iU as (U,H_i), but this must fail because it violates $D=D_1+D_2$; nor does there seem to be any valid way of otherwise decomposing the whole H into pieces relevant to the separate motions. The compartmentalized knowledge we have of $\mathbf{\Pi}_i$ as canonical momenta apparently cannot be made use of in the consolidated problem.

The true position of the strength of interaction e_1e_2 seems rather to be obscured by nearly any plausible attempt to introduce conjugate variables directly into H. The role of e_1e_2 is best seen by simply writing the energy in terms of the particle positions and their derivatives,

$$H=\frac{m_1\mathbf{v}_1^2}{(1-\mathbf{v}_1^2/c^2)^{\frac{1}{2}}}+m_1c^2(1-\mathbf{v}_1^2/c^2)^{\frac{1}{2}}+\frac{m_2\mathbf{v}_2^2}{(1-\mathbf{v}_2^2/c^2)^{\frac{1}{2}}}$$
$$+m_2c^2(1-\mathbf{v}_2^2/c^2)^{\frac{1}{2}}+e_1e_2\sum\frac{Z_{2p}}{2p!c^{2p}}\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}r^{2p-1}$$
$$+\frac{Y_{2p}}{2p!c^{2p}}\left(1-\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}\right)r^{2p-1}.$$

At this point we call a halt to generalities and do a small theoretical experiment to illustrate and clarify what is proposed to be a general procedure.

First write the kinetic terms in H nonrelativistically as $\frac{1}{2}m_1\mathbf{v}_1^2+\frac{1}{2}m_2\mathbf{v}_2^2$, and then keep only the first two terms of the two series to give, after rearrangement,

$$H^{\text{exp}}=\tfrac{1}{2}m_1\mathbf{v}_1^2+\tfrac{1}{2}m_2\mathbf{v}_2^2+\epsilon/r+(\epsilon/2c^2)(\mathbf{v}_1^2\cdot\mathbf{v}_2/r$$
$$+\mathbf{v}_1\cdot\mathbf{r}\mathbf{v}_2\cdot\mathbf{r}/r^3)+\epsilon/2c^4(D_1^2+D_2^2-D_1D_2)\mathbf{v}_1\cdot\mathbf{v}_2r.$$

To order everything strictly according to powers of c^{-1} up to the fourth would mean including $\epsilon Y_4r^3/24c^4$ as well as $\frac{3}{8}m_i\mathbf{v}_i^4/c^2+\frac{5}{16}m_i\mathbf{v}_i^6/c^4$; this merely complicates things here without illuminating them. The point of the experiment is to see how to write H^{exp} in terms of canonical conjugates $\mathbf{r}_i$, $\mathbf{p}_i$ alone so as to generate the correct higher-order differential equations of motion. These are the ones following from the Lagrangian[10]

$$L^{\text{exp}}=\tfrac{1}{2}m_1\mathbf{v}_1^2+\tfrac{1}{2}m_2\mathbf{v}_2^2-\epsilon(1-\mathbf{v}_1\cdot\mathbf{v}_2/c^2)/r$$
$$+(\epsilon/2c^2)D_1D_2(1-\mathbf{v}_1\cdot\mathbf{v}_2/c^2)r,$$

that corresponds to H^{exp}, and are of the fourth order.

The use of the differential operators D_1, D_2 needs some elaboration. They can of course be moved into an operand by explicit calculation (for instance giving above in $D_1^2+D_2^2-D_1D_2$ fifteen distinct terms) to the point where they can be called D, but this is much too clumsy. If U is some $U(\mathbf{r}_1,\mathbf{r}_2,\mathbf{p}_1,\mathbf{p}_2)$ and H is some $H(\mathbf{r}_1,\mathbf{r}_2,\mathbf{p}_1,\mathbf{p}_2)$ we can understand D_kU to be

$$D_2U=U_{\xi_i}H_{\rho_i}-U_{\rho_i}H_{\xi_i},\quad D_1U=U_{x_i}H_{p_i}-U_{p_i}H_{x_i}.$$

Here the labels $\mathbf{r}_1=(x_1,x_2,x_3)$, $\mathbf{p}_1=(p_1,p_2,p_3)$, $\mathbf{r}_2=(\xi_1,\xi_2,\xi_3)$, $\mathbf{p}_2=(\rho_1,\rho_2,\rho_3)$ are used, the subscripts mean partial derivatives, and a repeated index means summation from 1 to 3. Then D_1D_2U is apparently

$$D_1D_2U=U_{x_i\xi_j}H_{p_i}H_{\rho_j}-U_{p_i\xi_j}H_{x_i}H_{\rho_j}+U_{p_i\rho_j}H_{x_i}H_{\xi_j}$$
$$-U_{x_i\rho_j}H_{p_i}H_{\xi_j}+U_{x_i}H_{p_i\xi_j}H_{\rho_j}-U_{p_i}H_{x_i\xi_j}H_{\rho_j}$$
$$+U_{p_i}H_{x_i\rho_j}H_{\xi_j}-U_{x_i}H_{p_i\rho_j}H_{\xi_j}.$$

The first four terms are symmetric to the interchange of 1 and 2, and the last four are unsymmetric, so D_1D_2 comes out different from D_2D_1, which it never can do according to direct calculation referring to its primitive meaning. The difficulty comes from too literal a Poisson-bracket use of the D_k in succession, as can be seen by writing

$$D_1D_2U=D_1(U_{\xi_i}\dot{\xi}_i+U_{\rho_i}\dot{\rho}_i)$$
$$=U_{\xi_ix_j}\dot{x}_j\dot{\xi}_i+U_{\xi_ip_j}\dot{\xi}_i\dot{p}_j+U_{\rho_ix_j}\dot{x}_j\dot{\rho}_i+U_{\rho_ip_j}\dot{p}_j\dot{\rho}_i$$
$$+U_{\xi_i}D_1\dot{\xi}_i+U_{\rho_i}D_1\dot{\rho}_i.$$

The $D_1\dot{\xi}_i$, $D_1\dot{\rho}_i$ are *1*-type derivatives of *2*-type variables and must be placed equal to zero under the meaning of D_1 as a private derivative with respect to *1* variables only. These null terms are just the unsymmetric ones of the preceding equation while the first four are the symmetric ones in it. In effect the multiple derivatives of H, like $H_{p_i\xi_j}$ above, mixed in the *1* and *2* variables, are to be discarded. The simplest systematic way for preparing all differential operators for their operands is to express them now in combinations of D (full Poisson bracket) and $D_1D_2\equiv\Delta$ (semi-bracket as just

[10] An interesting sidelight on the significance of the ϵ expansion is gained by dropping the c^{-4} term here and treating the remaining exactly. The momenta $\mathbf{p}_i$ are $m_i\mathbf{v}_i+(\epsilon/2c^2)(\mathbf{v}_j/r+\mathbf{r}\mathbf{v}_j\cdot\mathbf{r}/r^3)$ and inverting for $\mathbf{v}=\mathbf{v}(\mathbf{p})$ produces a Hamiltonian in which every term but ϵ/r has a factor $(1-a_1a_2/r^2)^{-2}$ of $(1-a_1a_2/4r^2)^{-2}$, a_i being the charge radii e_i^2/m_ic^2. (The expansion by powers of ϵ of this Hamiltonian coincides with the direct calculation from $E=E_0+\epsilon E_1$ by the method about to be described.) This would seem to be the essence of a warning and a clue as to how the charge radii enter electrodynamics generally and limit either the physical theory itself, or its mathematical representation, or both.

described) instead of D_1 and D_2, writing for instance $D_1^2+D_2^2-D_1D_2$ as $D^2-3\Delta$.

Let us now admit a Hamiltonian $H_0+\epsilon H_1+\epsilon^2H_2+\cdots$ which is both the energy and a function of canonical variables that gives it power to generate derivatives. Using the abbreviations $\mathbf{r}_{1k}=(\mathbf{r}_1,H_k)$, $\mathbf{r}_{2k}=(\mathbf{r}_2,H_k)$ we find by equating like powers of ϵ in $H^{\exp}$,

$$H_0=\tfrac{1}{2}m_1\mathbf{r}_{10}^2+\tfrac{1}{2}m_2\mathbf{r}_{20}^2$$

$$H_1=m_1\mathbf{r}_{10}\cdot\mathbf{r}_{11}+m_2\mathbf{r}_{20}\cdot\mathbf{r}_{21}+\frac{1}{r}+\frac{1}{2c^2}\left(\frac{\mathbf{r}_{10}\cdot\mathbf{r}_{20}}{r}+\frac{\mathbf{r}_{10}\cdot\mathbf{r}\mathbf{r}_{20}\cdot\mathbf{r}}{r^3}\right)$$
$$+\frac{1}{2c^4}\{[(\mathbf{r}_{10}\cdot\mathbf{r}_{20}r,H_0),H_0]-3(\mathbf{r}_{10}\cdot\mathbf{r}_{20}r;H_0,H_0)\}$$

$$H_2=m_1(\mathbf{r}_{10}\cdot\mathbf{r}_{12}+\tfrac{1}{2}\mathbf{r}_{11}^2)+m_2(\mathbf{r}_{20}\cdot\mathbf{r}_{22}+\tfrac{1}{2}\mathbf{r}_{21}^2)$$
$$+\frac{1}{2c^2}\left(\frac{\mathbf{r}_{10}\cdot\mathbf{r}_{21}+\mathbf{r}_{11}\cdot\mathbf{r}_{20}}{r}+\frac{\mathbf{r}_{10}\cdot\mathbf{r}\mathbf{r}_{21}\cdot\mathbf{r}+\mathbf{r}_{11}\cdot\mathbf{r}\mathbf{r}_{20}\cdot\mathbf{r}}{r^3}\right)$$
$$+\frac{1}{2c^4}\{(([\mathbf{r}_{10}\cdot\mathbf{r}_{21}+\mathbf{r}_{11}\cdot\mathbf{r}_{20}]r,H_0),H_0)$$
$$-3([\mathbf{r}_{10}\cdot\mathbf{r}_{21}+\mathbf{r}_{11}\cdot\mathbf{r}_{20}]r;H_0,H_0)$$
$$+((\mathbf{r}_{10}\cdot\mathbf{r}_{20}r,H_0),H_1)+((\mathbf{r}_{10}\cdot\mathbf{r}_{20}r,H_1),H_0)$$
$$-3(\mathbf{r}_{10}\cdot\mathbf{r}_{20}r;H_0,H_1+H_1,H_0)\}$$

etc.

The notation $(U;H,H)$ is used for the semi-bracket, and the last semi-bracket corresponding to different orders of H in the D_1- and D_2-parts of D_1D_2 means

$$(U;H_0,H_1+H_1,H_0)$$
$$\equiv(\partial/\partial\epsilon)(U;H_0+\epsilon H_1,H_0+\epsilon H_1)|_{\epsilon=0}$$
$$=U_{x_i\xi_j}(H_{0p_i}H_{1\rho_j}+H_{1p_i}H_{0\rho_j})+U_{p_i\rho_j}(H_{0x_i}H_{1\xi_j}$$
$$+H_{1x_i}H_{0\xi_j})-U_{p_i\xi_j}(H_{0x_i}H_{1\rho_j}+H_{1x_i}H_{0\rho_j})$$
$$-U_{x_i\rho_j}(H_{0p_i}H_{1\xi_j}+H_{1p_i}H_{0\xi_j}),$$

with a similar meaning for an arbitrary semi-bracket $(U;H_k,H_l+H_l,H_k)$ that is mixed in the H-orders of D_1 and D_2.

The known $H_0=\mathbf{p}_1^2/2m_1+\mathbf{p}_2^2/2m_2$ of course satisfies

$$H_0=\tfrac{1}{2}m_1(\mathbf{r}_1,H_0)^2+\tfrac{1}{2}m_2(\mathbf{r}_2,H_0)^2$$
$$=\tfrac{1}{2}m_1\left(\frac{\partial H_0}{\partial\mathbf{p}_1}\right)^2+\tfrac{1}{2}m_2\left(\frac{\partial H_0}{\partial\mathbf{p}_2}\right)^2,$$

and gives $\mathbf{r}_{10}=\mathbf{p}_1/m_1$, $\mathbf{r}_{20}=\mathbf{p}_2/m_2$. For H_1 we have

$$H_1=\mathbf{p}_1\cdot\frac{\partial H_1}{\partial\mathbf{p}_1}+\mathbf{p}_2\cdot\frac{\partial H_1}{\partial\mathbf{p}_2}+\frac{1}{r}+\frac{1}{2c^2m_1m_2}\left(\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{r}+\frac{\mathbf{p}_1\cdot\mathbf{r}\mathbf{p}_2\cdot\mathbf{r}}{r^3}\right)$$
$$+\frac{1}{2c^4m_1m_2}\left\{\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{m_1^2}\left(\frac{\mathbf{p}_1^2}{r}-\frac{[\mathbf{p}_1\cdot\mathbf{r}]^2}{r^3}\right)+\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{m_2^2}\left(\frac{\mathbf{p}_2^2}{r}-\frac{[\mathbf{p}_2\cdot\mathbf{r}]^2}{r^3}\right)\right.$$
$$\left.+\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{m_1m_2}\left(\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{r}-\frac{\mathbf{p}_1\cdot\mathbf{r}\mathbf{p}_2\cdot\mathbf{r}}{r^3}\right)\right\}.$$

The inhomogeneous parts of this differential equation for H_1 are symmetric polynomials in $\mathbf{p}_1$, $\mathbf{p}_2$ homogeneous of degrees 0, 2, 4. Since $\mathbf{p}_1\cdot\partial/\partial\mathbf{p}_1+\mathbf{p}_2\cdot\partial/\partial\mathbf{p}_2$ is linear and is homogeneous of degree 0, H_1 will be, to within an ignorable solution of the homogeneous equation, a linear combination of the same polynomials:

$$H_1=\frac{1}{r}-\frac{1}{2c^2m_1m_2}\left(\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{r}+\frac{\mathbf{p}_1\cdot\mathbf{r}\mathbf{p}_2\cdot\mathbf{r}}{r^3}\right)$$
$$-\frac{1}{6c^4m_1m_2}\left\{\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{m_1^2}\left(\frac{\mathbf{p}_1^2}{r}-\frac{[\mathbf{p}_1\cdot\mathbf{r}]^2}{r^3}\right)+\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{m_2^2}\left(\frac{\mathbf{p}_2^2}{r}-\frac{[\mathbf{p}_2\cdot\mathbf{r}]^2}{r^3}\right)\right.$$
$$\left.+\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{m_1m_2}\left(\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{r}-\frac{\mathbf{p}_1\cdot\mathbf{r}\mathbf{p}_2\cdot\mathbf{r}}{r^3}\right)\right\}.$$

The computation of H_2 is already very lengthy; for instance the last of the inhomogeneous terms in the H_2 equation, $(\mathbf{r}_{10}\cdot\mathbf{r}_{20}r;H_0,H_1+H_1,H_0)$, is

$$\frac{\partial}{\partial\mathbf{r}_1}\left[\frac{\partial}{\partial\mathbf{r}_2}(\mathbf{r}_{10}\cdot\mathbf{r}_{20}r)\cdot\frac{\mathbf{p}_2}{m_2}\right]\cdot\frac{\partial H_1}{\partial\mathbf{p}_1}+\frac{\partial}{\partial\mathbf{r}_2}\left[\frac{\partial}{\partial\mathbf{r}_1}(\mathbf{r}_{10}\cdot\mathbf{r}_{20}r)\cdot\frac{\mathbf{p}_1}{m_1}\right]\cdot\frac{\partial H_1}{\partial\mathbf{p}_2}$$
$$-\frac{\partial}{\partial\mathbf{p}_1}\left[\frac{\partial}{\partial\mathbf{r}_2}(\mathbf{r}_{10}\cdot\mathbf{r}_{20}r)\cdot\frac{\mathbf{p}_2}{m_2}\right]\cdot\frac{\partial H_1}{\partial\mathbf{r}_1}$$
$$-\frac{\partial}{\partial\mathbf{p}_2}\left[\frac{\partial}{\partial\mathbf{r}_1}(\mathbf{r}_{10}\cdot\mathbf{r}_{20}r)\cdot\frac{\mathbf{p}_1}{m_1}\right]\cdot\frac{\partial H_1}{\partial\mathbf{r}_2},$$

which when written out in full gives initially some 68 pieces. If we limit things systematically to order c^{-4}, only the c^0 terms need be kept here and throughout the curly bracket of which it is a part. Altogether for H_2 we find, up to c^{-4} order,

$$H_2=\mathbf{p}_1\cdot\frac{\partial H_2}{\partial\mathbf{p}_1}+\mathbf{p}_2\cdot\frac{\partial H_2}{\partial\mathbf{p}_2}+\frac{1}{m_1m_2c^4}\left\{-\frac{5}{8}\left(\frac{\mathbf{p}_1^2}{m_1r^2}+\frac{\mathbf{p}_2^2}{m_2r^2}\right)\right.$$
$$+\frac{13}{8}\left[\frac{(\mathbf{p}_1\cdot\mathbf{r})^2}{m_1r^4}+\frac{(\mathbf{p}_2\cdot\mathbf{r})^2}{m_2r^4}\right]$$
$$\left.+\left(\frac{1}{m_1}+\frac{1}{m_2}\right)\left(\frac{\mathbf{p}_1\cdot\mathbf{p}_2}{r^2}-\frac{1}{2}\frac{\mathbf{p}_1\cdot\mathbf{r}\mathbf{p}_2\cdot\mathbf{r}}{r^4}\right)\right\},$$

whence H_2 is just the negative of the inhomogeneous part.

The conclusion of the experiment is this, that directly in the equations of motion

$$D\frac{\partial L^{\exp}}{\partial\mathbf{v}_i}-D^2\frac{\partial L^{\exp}}{\partial\dot{\mathbf{v}}_i}=\frac{\partial L^{\exp}}{\partial\mathbf{r}_i},$$

when all derivatives are calculated as Poisson-brackets from $H_0+\epsilon H_1+\epsilon^2H_2$ there results an identity in $\mathbf{r}_1$, $\mathbf{p}_1$ up to terms in ϵ^2 and c^{-4}. If c^{-1} ordering is ignored and

H_2 is calculated fully, the same thing holds (to order ϵ^2). This calculation is not light though it is straightforward. Pragmatically, one concludes that H built as above is bound to be right as far as the equations of motion are concerned.

To see in a general way that this is so, notice that H, computed to any order, is symmetric to $1 \leftrightarrows 2$, as it must be. This means that $\mathbf{p}_1+\mathbf{p}_2=\textbf{const.}$, since $\dot{\mathbf{p}}_i = -\partial H/\partial \mathbf{r}_i$ and $\partial H/\partial \mathbf{r}_1$ equals $-\partial H/\partial \mathbf{r}_2$. That is, the $\mathbf{p}_i$ are the 'right,' conserved, momenta. They must therefore be physically the same thing as the conserved Ostrogradsky $\mathbf{P}_i^{(0)}$. But the equations of motion $d\mathbf{P}_i^{(0)}/dt = -\partial H^{\text{ost}}/\partial \mathbf{r}_i$ are only the full Lagrangian equations of motion

$$\partial L/\partial \mathbf{r}_i - D\partial L/\partial \mathbf{v}_i + D^2\partial L/\partial \dot{\mathbf{v}}_i - \cdots = 0$$

in the thinnest of disguises; and $H^{\text{ost}}(\mathbf{r}_i,\dot{\mathbf{r}}_i,\cdots;\ \mathbf{P}_i^{(0)}, \mathbf{P}_i^{(1)},\cdots)$ is physically the same as $H(\mathbf{r}_i,\mathbf{p}_i)$, both being $E(\mathbf{r}_i,\mathbf{v}_i,\dot{\mathbf{v}}_i,\cdots)$ dressed up in different mathematical clothing. This is to say that *the $H(\mathbf{r}_i,\mathbf{p}_i)$ equations of motion, $d\mathbf{p}_i/dt = -\partial H/\partial \mathbf{r}_i$, have exactly the same physical contents as the primitive Lagrangian equations of motion.* In short, so long as we self-consistently construct E in the $\mathbf{r}_i$, $\mathbf{p}_i$ language we succeed in writing the equations of motion in the same language.

The physical transparency of this result contrasts with a certain inaccessability of its strict mathematical statement. What is involved in the latter would appear to be not basically different from that in the case of the simple Hamiltonian $ap^2+bp^3+V(x)$ mentioned earlier; namely, that $\mathbf{p}_i$ is not defined in terms of the physical ultimates $\mathbf{r}_i$, $\dot{\mathbf{r}}_i$, $\ddot{\mathbf{r}}_i$, $\cdots$ through $\dot{\mathbf{r}}_i=(\mathbf{r}_i,H)$, as it is in ordinary particle mechanics via $\mathbf{p}=\mathbf{p}(\mathbf{r},\mathbf{v})$, but becomes so defined only after $\dot{\mathbf{r}}_i$, $\ddot{\mathbf{r}}_i$, $\cdots$ are reckoned as multiple Poisson brackets from H; when H is truncated and contains only finite powers of $\mathbf{p}_i$, the proper assembling of a finite number of $\mathbf{r}_i$, $\dot{\mathbf{r}}_i(\mathbf{r}_i,\mathbf{p}_i)$, $\ddot{\mathbf{r}}_i(\mathbf{r}_i,\mathbf{p}_i)$, $\cdots$ is then to be capable of isolating $\mathbf{p}_i$ alone, and not any power of it, the assemblage of particle-position derivatives being, according to the preceding physical argument, the same assemblage defining the correct Ostrogradsky $\mathbf{P}_i^{(0)}$; and when H is not truncated the assemblage is to involve all derivatives, as does $\mathbf{P}_i^{(0)}$.

Let us recapitulate finally what explicit assumptions have gone into the present statement of electrodynamics as action-at-a-distance: (a) That position and momentum are sufficient dynamical variables for the description of motion of charged as of uncharged particles, being the action-at-a-distance way of putting both on the same footing when any description of fields is eschewed; (b) that Bhabha's criterion selects out of the infinite-parametered class of electrodynamic motions, set out initially through infinite-order differential equations, a finite-parametered class, $\mathbf{r}_i(0)$, $\mathbf{p}_i(0)$, which alone represent the physically realizable motions; (c) that this criterion is expressed through the existence of a Hamiltonian $H=H_0(\mathbf{r},\mathbf{p})+\epsilon H_1(\mathbf{r},\mathbf{p})+\cdots$ generating the admissible motions; and for whose calculation, (d), the algorithm

$$H_0+\epsilon H_1+\cdots = E(\mathbf{r}_i,\dot{\mathbf{r}}_i,\ddot{\mathbf{r}}_i,\cdots) = E(\mathbf{r}_i,(\mathbf{r}_i,H),((\mathbf{r}_i,H),H),\cdots)$$

suffices by dint of H's dual role as the physically identifiable energy E and as Poisson-bracket generator of time derivatives.

The nature of the multiple series in ϵ and $1/c^2$ for H remains unknown. The algorithm giving H has not been shown conclusively, but only plausibly, to embody the primitive action-at-a-distance equations of motion.

Field-Theory Analogs of the Lagrange and Poincaré Invariants*

P. A. STURROCK
Microwave Laboratory, W. W. Hansen Laboratories of Physics, Stanford University, Stanford, California
(Received September 9, 1960)

The Lagrange differential invariant and the Poincaré integral invariant of classical dynamics have as their analogs in Lagrangian field theory a "differential divergence-free vector" and an "integral divergence-free vector."

The former, which is expressible as a divergence-free vector-bracket expression, may be used to derive conservation relations associated with the transformation properties of a given system. It is not necessary that these transformations should be infinitesimal; by way of example, conservation theorems are established for systems which are periodic and for systems which are invariant under spatial inversion. The differential divergence-free vector may also be used to establish reciprocity and orthogonality relations: simple examples which are here discussed are Betti's reciprocal theorem of elasticity and Lorentz's reciprocal relation of electromagnetic theory. An extended form of the differential divergence-free vector allows for variation not only of the dependent variables but also of the independent variables.

The integral divergence-free vector associates a conserved quantity with any closed one-parameter family of solutions of the field equations. As examples, we derive the "equation of conservation of probability" of quantum mechanics, and a classical form of the relation between the momentum and wave vectors for a plane wave in a propagating medium.

The theorem of classical dynamics relating a complete set of Poisson brackets to a complete set of Lagrange brackets cannot be extended to the present formalism. The formula which represents the obvious extension of the classical formula for the Poisson bracket is of no interest since it can be shown not to be canonically invariant.

I. INTRODUCTION

THE Lagrange differential invariant[1] occupies an important position in the logical structure of Classical Dynamics. The formula for this invariant represents one of the simplest statements of the laws of dynamics since it involves only the dynamical variables and their conjugate momenta. One may readily derive from this invariant the existence and properties of the Lagrangian and Hamiltonian functions, of the various characteristic functions and of the generators of canonical transformations. The Poincaré[2] and adiabatic invariants[3] may be regarded as derivative of the Lagrange invariant.

In this article we look for the analogs in Lagrangian field theory, by which we mean the theory of fields whose behavior may be derived from an action principle, of the Lagrange differential invariant and of the Poincaré integral invariant of classical dynamics. Each takes the form of a divergence-free vector, the former constructed from two independent perturbations of field variables and their conjugate momenta, the latter from an integral over a cyclic parameter. The former leads immediately to a divergence-free vector bracket which one may ascribe to a two-parameter family of solutions of the field equations.

The differential divergence-free vector and the integral divergence-free vector find application in establishing conservation relations whenever the symmetry and transformation properties of the Lagrangian function, or the properties of some class of solutions of the field equations, make it possible to associate an appropriate family of solutions with any one particular solution. We shall see that, in the particular case that the field equations are linear, it is not necessary that the transformations from which conservation relations are to be derived should be infinitesimal. One may, for instance, associate a conservation relation with invariance under spatial inversion.

The Lagrange invariant is a source of reciprocal theorems in both dynamics[4] and optics.[5] We find that the differential divergence-free vector also is a ready source of reciprocal theorems, of which two familiar examples will be given. Reciprocal and orthogonality theorems are closely related, and we shall see that the latter also may be derived from the differential divergence-free vector.

Since Poisson brackets,[6] which are closely connected with Lagrange brackets,[7] assumed an important role in the early development of quantum mechanics,[8] it is natural to inquire whether these quantities also have vector analogs in the Lagrangian theory of fields. We shall see, in the last section, that it is difficult to find the appropriate analog, if such exists: the obvious formula for such a generalization of the Poisson bracket is not canonically invariant and so is of no interest.

* The research reported in this document was supported jointly by the U. S. Army Signal Corps, the U. S. Air Force, and the U. S. Navy (Office of Naval Research).

[1] E. T. Whittaker, *Analytical Dynamics* (Cambridge University Press, London, 1937), 4th ed., pp. 298–9.

[2] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1950), pp. 247 ff.

[3] E. T. Whittaker, *A History of Theories of Aether and Electricity* (Thomas Nelson and Sons, New York, 1953), pp. 122 ff.

[4] Reference 1, pp. 304–5.

[5] M. Herzberger, *Strahlenoptik* (Julius Springer Verlag, Berlin, Germany, 1931), pp. 22 ff.

[6] Reference 1, pp. 299 ff.

[7] Reference 1, pp. 298–9.

[8] P. A. M. Dirac, *Quantum Mechanics* (Oxford University Press, London, 1947), 3rd ed., pp. 84 ff.

II. DIFFERENTIAL DIVERGENCE-FREE VECTOR

We consider a field described by dynamical variables $\phi_\alpha(x)$ which are functions of the spatial variables x^μ; the index α may enumerate scalar quantities or the components of vectors, spinors, etc. We assume that the field equations are represented by a variation principle

$$\delta\int \mathcal{L}dx=0, \tag{2.1}$$

where $\mathcal{L}(\phi_\alpha,\phi_{\beta;\mu},x^\nu)$ is the relevant Lagrangian function, dx is an abbreviation for $dx^1dx^2\cdots$, and

$$\phi_{\alpha;\mu}=d\phi_\alpha/dx^\mu, \quad \text{etc.} \tag{2.2}$$

It is convenient to adopt the notation df/dx^μ for a partial derivative when the total dependence of the function on the independent variables is taken into account; we reserve the notation $\partial f/\partial x^\mu$, etc., for the case where the dependent variable is expressed in a certain functional form so that the relevant derivative may not comprise the total variation of the function with any one of the independent variables.

To each dynamical variable ϕ_α corresponds a canonical momentum vector $\pi^{\alpha\mu}$ defined by

$$\pi^{\alpha\mu}=\partial\mathcal{L}/\partial\phi_{\alpha;\mu}. \tag{2.3}$$

The Euler-Lagrange equations[9] derivable from (2.1) are

$$\frac{d}{dx^\mu}\left(\frac{\partial\mathcal{L}}{\partial\phi_{\alpha;\mu}}\right)=\frac{\partial\mathcal{L}}{\partial\phi_\alpha}. \tag{2.4}$$

Here, and henceforth, the summation convention is employed. A more compact expression of (2.3) and (2.4) is provided by the differential relation

$$\delta\mathcal{L}=\frac{d}{dx^\mu}(\pi^{\alpha\mu}\delta\phi_\alpha)\equiv\pi^{\alpha\mu}{}_{;\mu}\delta\phi_\alpha+\pi^{\alpha\mu}\delta\phi_{\alpha;\mu}, \tag{2.5}$$

since if $\delta\mathcal{L}$ is expressed as

$$\delta\mathcal{L}=\frac{\partial\mathcal{L}}{\partial\phi_\alpha}\delta\phi_\alpha+\frac{\partial\mathcal{L}}{\partial\phi_{\alpha;\mu}}\delta\phi_{\alpha;\mu}, \tag{2.6}$$

we regain (2.3) and (2.4) on equating coefficients of the independent variations $\delta\phi_\alpha$ and $\delta\phi_{\alpha;\mu}$.

Suppose that the field variables are now subject to an additional variation $d\phi_\alpha$. The product variation of $\mathcal{L}$ is seen from (2.5) to be given by

$$d\delta\mathcal{L}=(d/dx^\mu)(d\pi^{\alpha\mu}\delta\phi_\alpha+\pi^{\alpha\mu}d\delta\phi_\alpha). \tag{2.7}$$

However, reversal of the order in which the variations are effected shows that

$$\delta d\mathcal{L}=(d/dx^\mu)(\delta\pi^{\alpha\mu}d\phi_\alpha+\pi^{\alpha\mu}\delta d\phi_\alpha). \tag{2.8}$$

Since the variation $d\delta\mathcal{L}$ is independent of the order in which the constituent variations are carried out, the order of d and δ may be interchanged. On subtracting (2.8) from (2.7), we now see that

$$(d/dx^\mu)(\delta\pi^{\alpha\mu}d\phi_\alpha-d\pi^{\alpha\mu}\delta\phi_\alpha)=0. \tag{2.9}$$

This is the equation which takes the place, in field theory, of the equation expressing the constancy of the Lagrange invariant in classical dynamics.[1] The expression within brackets will be referred to as the "differential divergence-free vector."

We may see that (2.9) is a statement of the canonical nature of the field by rearranging it in the form

$$(-\delta\pi^{\alpha\mu}{}_{;\mu}d\phi_\alpha+\delta\phi_{\alpha;\mu}d\pi^{\alpha\mu}) -(-d\pi^{\alpha\mu}{}_{;\mu}\delta\phi_\alpha+d\phi_{\alpha;\mu}\delta\pi^{\alpha\mu})=0. \tag{2.10}$$

This differential relation expresses the condition that the vector $(-\pi^{\alpha\mu}{}_{;\mu},\phi_\beta)$, regarded as a function of the variables $(\phi_\alpha,\pi^{\beta\mu})$, should be integrable. That is, it represents the condition that there should exist a function (the Hamiltonian function) $H(\pi^{\alpha\mu},\phi_\beta,x^\nu)$, with the property

$$\delta H=\phi_{\alpha;\mu}\delta\pi^{\alpha\mu}-\pi^{\alpha\mu}{}_{;\mu}\delta\phi_\alpha. \tag{2.11}$$

If the system is such that the variables ϕ_α and $\pi^{\beta\mu}$ may be varied independently, this leads to the canonical equations

$$\frac{d\phi_\alpha}{\partial x^\mu}=\frac{\partial H}{\partial\pi^{\alpha\mu}}, \quad \frac{d\pi^{\alpha\mu}}{dx^\mu}=-\frac{\partial H}{\partial\phi_\alpha}, \tag{2.12}$$

If variations of variables ϕ and $\pi^{\alpha\mu}$ are not independent, (2.12) should be modified appropriately.[10] We may recover (2.5), expressing the properties of the Lagrangian function, from (2.11) by the Legendre transformation[11]

$$\mathcal{L}=\pi^{\alpha\mu}\phi_{\alpha;\mu}-H. \tag{2.13}$$

By analogy with classical dynamics,[12] we may adopt as a criterion for a canonical transformation, from variables $\pi^{\alpha\mu}$, ϕ_β to $\Pi^{\gamma\nu}$, Φ_δ, for instance, that the form of the differential divergence-free vector should be preserved:

$$\delta\Pi^{\alpha\mu}d\Phi_\alpha-d\Pi^{\alpha\mu}\delta\Phi_\alpha=\delta\pi^{\alpha\mu}d\phi_\alpha-d\pi^{\alpha\mu}\delta\phi_\alpha. \tag{2.14}$$

This leads to the representation of canonical transformations by generating functions.[13] For instance, if (2.14) is rearranged as

$$(\delta\Pi^{\alpha\mu}d\Phi_\alpha-\delta\pi^{\alpha\mu}d\phi_\alpha)-(d\Pi^{\alpha\mu}\delta\phi_\alpha-d\pi^{\alpha\mu}\delta\phi_\alpha)=0, \tag{2.15}$$

and if the variables are so related that Φ_α and ϕ_β may be varied independently, then there exists a function $G^\mu(\Phi_\alpha,\phi_\beta,x^\nu)$ such that

$$\Pi^{\alpha\mu}=\partial G^\mu/\partial\Phi_\alpha, \quad \pi^{\alpha\mu}=-\partial G^\mu/\partial\phi_\alpha. \tag{2.16}$$

If variations of Φ_α and ϕ_β are not independent, (2.16)

[9] Reference 2, p. 38.

[10] Reference 1, p. 295.
[11] Reference 2, pp. 215 ff.
[12] Reference 1, p. 297.
[13] Reference 2, pp. 239 ff.

must be replaced by slightly more complicated expressions.[10]

The relation (2.9) leads to the field-theory analog of the Lagrange bracket.[7] If the variables $\phi_\alpha(x)$ and the concomitant momentum vectors $\pi^{\alpha\mu}(x)$ are expressed as functions which involve two or more parameters $A, B, \cdots$, (2.9) leads to the relation

$$(d/dx^\mu)\{A,B\}^\mu=0, \tag{2.17}$$

where $\{A,B\}^\mu$, the divergence-free vector bracket, is defined by

$$\{A,B\}^\mu=\frac{\partial\pi^{\alpha\mu}}{\partial A}\frac{\partial\phi_\alpha}{\partial B}-\frac{\partial\pi^{\alpha\mu}}{\partial B}\frac{\partial\phi_\alpha}{\partial A}. \tag{2.18}$$

Although, for simplicity, we refer to (2.18) as a "vector," it should be noted that the parameters A, B, etc. may on occasion be formed from the components of a tensor or spinor, in which case the transformation properties of (2.18) will not be those of a simple vector.

In the important special case that the field equations (2.3) and (2.4) are linear and homogeneous, we may form a two-parameter family of solutions from any two distinct particular solutions by varying the amplitudes of the solutions. Hence, from the distinct solutions $\pi^{\alpha\mu}(x)$, $\phi_\beta(x)$ and $\tilde{\pi}^{\alpha\mu}(x)$, $\tilde{\phi}_\beta(x)$, we may form the set $A\pi^{\alpha\mu}(x)$, $A\phi_\beta(x)$; $B\tilde{\pi}^{\alpha\mu}(x)$, $B\tilde{\phi}_\beta(x)$. The bracket expression of (2.18) now simplifies so that (2.17) becomes

$$(d/dx^\mu)(\pi^{\alpha\mu}\tilde{\phi}_\alpha-\tilde{\pi}^{\alpha\mu}\phi_\alpha)=0. \tag{2.19}$$

III. CONSERVATION THEOREMS

One would expect, from the form of (2.9) and (2.17), that the differential divergence-free vector would enable one to derive conservation theorems for fields with appropriate transformation properties. We shall first investigate simple familiar examples.

If the Lagrangian function is independent of one or more dynamical variables $\phi_{\alpha'}$, then, according to (2.3), $\pi^{\alpha\mu}$ is independent of $\phi_{\alpha'}$. Hence, if $\delta\phi_{\alpha'}$ is a variation which is arbitrary but independent of x, then $\delta\pi^{\alpha\mu}=0$. It now follows from (2.9) that

$$(d/dx^\mu)d\pi^{\alpha'\mu}=0, \tag{3.1}$$

from which we infer that $\pi^{\alpha'\mu}{}_{;\mu}$ is a function of x only. Hence, with this approach, we need additional information such as the existence of a "null" solution for which $\pi^{\alpha\mu}(x)=0$ in order to obtain the appropriate conservation theorem

$$d\pi^{\alpha'\mu}/dx^\mu=0 \tag{3.2}$$

which we might have obtained more directly from (2.4).

We must expect to obtain an incomplete theorem from (2.9) whenever we are considering invariance under a single transformation. If, for instance, the Lagrangian function has translational symmetry, we may adopt as one set of variations

$$\delta\phi_\alpha=\delta x^\mu\phi_{\alpha;\mu},\quad \delta\pi^{\alpha\nu}=\delta x^\mu\pi^{\alpha\nu}{}_{;\mu}, \tag{3.3}$$

where δx^μ is here independent of x. We now see from (2.9) that

$$(d/dx^\nu)(\pi^{\alpha\nu}{}_{;\mu}d\phi_\alpha-d\pi^{\alpha\nu}\phi_{\alpha;\mu})=0. \tag{3.4}$$

On noting (2.5) and the identity

$$(d/dx^\nu)\{(\pi^{\alpha\nu}d\phi_\alpha)_{;\mu}-(\pi^{\alpha\sigma}d\phi_\alpha)_{;\sigma}\delta_\mu{}^\nu\}=0 \tag{3.5}$$

and subtracting (3.4) from (3.5), we see that

$$(d/dx^\nu)dT_\mu{}^\nu=0, \tag{3.6}$$

where

$$T_\mu{}^\nu=\pi^{\alpha\nu}\phi_{\alpha;\mu}-\mathfrak{L}\delta_\mu{}^\nu. \tag{3.7}$$

It is now necessary to assume that there is a "zero-stress solution," for which $T_\mu{}^\nu(x)=0$, in order to obtain from (3.6) the conservation theorem

$$dT_\mu{}^\nu/dx^\nu=0. \tag{3.8}$$

As we may anticipate from our remarks at the end of the previous section, the above difficulty does not arise in the case where the field equations are linear and homogeneous. As an example, we consider the gauge invariance[14] of the source-free electromagnetic field, for which ϕ_μ is the vector potential and $\pi^{\mu\nu}$ the field tensor. From any solution we may form a second solution by the transformation

$$\tilde{\phi}_\mu=\phi_\mu+\chi_{;\mu},\quad \tilde{\pi}^{\mu\nu}=\pi^{\mu\nu}, \tag{3.9}$$

where $\chi(x)$ is an arbitrary function of x. Hence, (2.19) shows that

$$\pi^{\mu\nu}\chi_{;\mu\nu}+\pi^{\mu\nu}{}_{;\nu}\chi_{;\mu}=0, \tag{3.10}$$

from which it follows that $\pi^{\mu\nu}$ is antisymmetric and satisfies the conservation equation

$$d\pi^{\mu\nu}/dx^\nu=0. \tag{3.11}$$

One of the most interesting properties of the differential divergence-free vector is that, for linear homogeneous field equations, it enables one to derive a conservation theorem for finite transformations, that is from transformations which are not, and cannot be constructed from, infinitesimal transformations. Suppose, for example, that the Lagrangian function is *periodic* in the independent variables so that

$$\mathfrak{L}(\phi_\alpha,\phi_{\beta;\mu},x^\nu+\xi^\nu)\equiv\mathfrak{L}(\phi_\alpha,\phi_{\beta;\mu},x^\nu). \tag{3.12}$$

Then from any one solution $\pi^{\alpha\mu}(x)$, $\phi_\beta(x)$ of the field equations, we may form a second solution as follows:

$$\tilde{\pi}^{\alpha\mu}(x)=\pi^{\alpha\mu}(x+\xi),\quad \tilde{\phi}_\alpha(x)=\phi_\alpha(x+\xi). \tag{3.13}$$

Equation (2.19) now becomes

$$(d/dx^\mu)\{\pi^{\alpha\mu}(x)\phi_\alpha(x+\xi)-\pi^{\alpha\mu}(x+\xi)\phi_\alpha(x)\}=0, \tag{3.14}$$

so that we have constructed a relation analogous to (3.8) even though the Lagrangian function is not invariant under infinitesimal displacements.

[14] L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1951), pp. 47–8.

A further interesting application of (2.19) is that of associating a conservation equation with reflection symmetry, that is, invariance under inversion with respect to one or more coordinates. Suppose, for instance, that

$$\mathfrak{L}(\phi_\alpha\phi_{\beta;\mu},\ -x^\nu)\equiv\mathfrak{L}(\phi_\alpha,\phi_{\beta;\mu},x^\nu). \tag{3.15}$$

Then from any one solution we may form a second solution as follows:

$$\tilde{\phi}_\alpha(x)=\phi_\alpha(-x),\quad \tilde{\pi}^{\alpha\mu}(x)=-\pi^{\alpha\mu}(-x). \tag{3.16}$$

The minus sign occurs in the second equation because

$$\tilde{\phi}_{\alpha;\mu}(x)=-\phi_{\alpha;\mu}(-x).$$

Hence, (2.19) becomes

$$(d/dx^\mu)\{\pi^{\alpha\mu}(x)\phi_\alpha(-x)+\pi^{\alpha\mu}(-x)\phi_\alpha(x)\}=0. \tag{3.17}$$

If, as another example, μ enumerates time (x^0) and space coordinates (x^1,x^2,x^3), and if the Lagrangian function is invariant under inversion with respect to the spatial coordinates alone, we may derive an equation similar to (3.17). On integration over the spatial coordinates, we then obtain the invariant

$$\int d^3x\{\pi^{\alpha0}(t,\mathbf{x})\phi_\alpha(t,-\mathbf{x})-\pi^{\alpha0}(t,-\mathbf{x})\phi_\alpha(t,\mathbf{x})\}=\text{const.}, \tag{3.18}$$

where the time and space variables have been written alternatively as $t,\mathbf{x}$. Hence it is possible to introduce a "parity invariant"[15] in purely classical theory.

IV. RECIPROCITY AND ORTHOGONALITY RELATIONS

In this section we shall consider the particular case that the field equations are linear and homogeneous, but we may at any time remove this restriction by replacing relations such as (2.19) by the differential form (2.9).

Equation (2.19) may be integrated over any volume bounded by one or more closed surfaces. We then obtain

$$\int dS_\mu\pi^{\alpha\mu}\tilde{\phi}_\alpha=\int dS_\mu\tilde{\pi}^{\alpha\mu}\phi_\alpha. \tag{4.1}$$

The variables ϕ_α will normally be "extensive quantities," and the variables $\pi^{\alpha\mu}$ "intensive quantities." Hence (4.1) has an interpretation of the following form: The surface integral of one set of "forces" times the "response" due to a second set of "forces" is equal to the surface integral of the second set of "forces" times the "response" due to the first set of "forces."

As a simple example we may consider the small-amplitude static theory of elasticity. The indices μ and α each take the values 1, 2, 3 and will be replaced by $r(s$, etc.); $\phi_r(x)$ now denotes the displacement of the point of the elastic body which is at x in the stress-free state, and the Lagrangian function is the negative of the potential energy density $\Upsilon(\phi_r,\phi_{s;t},x_\mu)$.[16] Since Υ is symmetrical in $\phi_{r;s}$ and its transpose $\phi_{s;r}$, the conjugate variable, now defined by

$$\pi_{rs}=-\partial\Upsilon/\partial\phi_{r;s}, \tag{4.2}$$

is symmetrical in its indices. Equation (4.1) now becomes

$$\int dS_r\pi_{rs}\tilde{\phi}_s=\int dS_r\tilde{\pi}_{rs}\phi_s. \tag{4.3}$$

Since π_{rs} is the stress tensor for the elastic medium, (4.3) has the following interpretation: the work done in performing the first displacement of the surface of the volume against the elastic forces produced by the second displacement is equal to the work done in performing the second displacement against the elastic forces produced by the first displacement. This relation is known as Betti's reciprocal theorem of elasticity.[17]

Relation (4.1) may be given an interpretation analogous to that of Helmoltz's reciprocal theorem[4] if we assume that $\pi^{\alpha\mu}$ is nonzero on the bounding surface only in the neighborhood of a point x, and that $\tilde{\pi}^{\alpha\mu}$ is nonzero on the bounding surface only in the neighborhood of a point $\tilde{x}$. Then (4.1) states that the "force" $\int dS_\mu\pi^{\alpha\mu}$ at x times the "response" $\tilde{\phi}_\alpha$ at that point due to a second "force" at $\tilde{x}$ is equal to the "force" $\int dS_\mu\tilde{\pi}^{\alpha\mu}$ at $\tilde{x}$ times the "response" ϕ_α at that point due to the "force" at x. If we consider two infinitesimal closed surfaces surrounding the points $x,\tilde{x}$, we may introduce a Green function as follows:

$$\phi_\alpha(x)=G_{\alpha\beta}(x,\tilde{x})\int dS_\mu\tilde{\pi}^{\beta\mu}. \tag{4.4}$$

We may now express the reciprocal relation (4.1) in terms of the Green function[18]:

$$G_{\alpha\beta}(x,\tilde{x})=G_{\beta\alpha}(\tilde{x},x). \tag{4.5}$$

The Green function which we have introduced has been defined by discussion of linear homogeneous field equations, whereas Green functions are usually introduced by discussion of inhomogeneous field equations.[18] We may relate the two approaches by interpreting the flux across the infinitesimal surface surrounding the point $\tilde{x}$ as due to a forcing term derived from a contribution $\phi_\beta\delta(x-\tilde{x})$ to the Lagrangian function. This gives rise to an inhomogeneous contribution to the Euler-

[15] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), p. 139.

[16] A. E. H. Love, *Mathematical Theory of Elasticity* (Cambridge University Press, London, 1927), 4th ed., pp. 166 ff.

[17] Reference 16, pp. 173–4.

[18] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), pp. 882–3, 1770.

Lagrange equation in that (2.4) is changed to

$$\frac{d\pi^{\alpha\mu}}{dx^\mu}-\frac{\partial\mathfrak{L}}{\partial\phi_\alpha}=\delta_\beta{}^\alpha\delta(x-\tilde{x}). \tag{4.6}$$

The function $G_{\alpha\beta}(x,\tilde{x})$ originally defined by (4.4) is now seen to be the solution $\phi_\alpha(x)$ (usually restricted by convenient subsidiary boundary conditions) of (4.6).

Equations of the form (2.19) are sometimes referred to as "reciprocal relations." For instance, in the theory of the electromagnetic field, (2.19) may be expressed in three-dimensional tensor notation as

$$\frac{1}{c}\frac{d}{dt}(E_r\tilde{A}_r-\tilde{E}_rA_r)+\frac{d}{dx_r}[E_r\tilde{\phi}-\tilde{E}_r\phi+\epsilon_{rst}(H_s\tilde{A}_t-\tilde{H}_sA_t)]=0, \tag{4.7}$$

where ϵ_{rst} is the alternating tensor. This is a generalization of the Lorentz reciprocal relation[19]

$$\epsilon_{rst}\frac{d}{dx_r}(H_s{}^*E_t+H_sE_t{}^*)=0, \tag{4.8}$$

which we may obtain from (4.7) by considering two solutions of the field equations of the same frequency, which we represent as

$$Er(x,t)\rightarrow E_r{}^*(x)e^{-i\omega t},\quad \text{etc.},$$
$$\tilde{E}_r(x,t)\rightarrow\tilde{E}_r(x)e^{i\omega t},\quad \text{etc.}, \tag{4.9}$$

and adopting the gauge in which $\phi=0$.

We shall now show, by means of an example, how a reciprocal relation may yield an orthogonality theorem. We consider a waveguide which is uniform in the z direction but has arbitrary cross section in the $x-y$ plane, and consider two normal modes of the system by replacing (4.9) by

$$\begin{aligned}&E_x(x,y,z,t)\rightarrow E_x{}^*(x,y)\exp[-i(k^*z+\omega t)],\quad \text{etc.},\\ &\tilde{E}_x(x,y,z,t)\rightarrow E_x(x,y)\exp[i(\tilde{k}z+\omega t)],\quad \text{etc.}\end{aligned} \tag{4.10}$$

We now find, on integrating (4.7) over the cross section of the waveguide, that

$$\frac{d}{dz}\Big[\exp[i(\tilde{k}-k^*)z]\int\int dxdy(H_x{}^*\tilde{E}_y-H_y{}^*\tilde{E}_x+\tilde{H}_xE_y{}^*-\tilde{H}_yE_x{}^*)\Big]=0 \tag{4.11}$$

since, at a metallic surface, E_r and H_r must be normal and tangential to the surface, respectively. It follows that

$$\int\int dxdy(H_x{}^*\tilde{E}_y-H_y{}^*\tilde{E}_x+\tilde{H}_xE_y{}^*-\tilde{H}_yE_x{}^*)=0 \quad \text{if } \tilde{k}\neq k^*, \tag{4.12}$$

which is in the form of an orthogonality theorem.[20]

V. EXTENSION OF THE DIFFERENTIAL DIVERGENCE-FREE VECTOR

The equation of the Lagrange differential invariant of classical mechanics may be extended by allowing for variations of the independent variable (time) as well as of the dynamical (dependent) variables. The analogous divergence-free vector of field theory may be extended in the same way.

We consider the integral of $\mathfrak{L}$ over a volume V and suppose that, under perturbation, the volume of integration is changed to V'. We characterize this perturbation by specifying that to each point x^μ of V corresponds a point $x^\mu+\delta x^\mu$ of V', where δx^μ is a prescribed function of x^ν. We assume, in addition, that the field variables are perturbed, but we characterize this perturbation by relating the value of the perturbed field variable $\phi_\alpha(x)$ at the "displaced" point $x^\mu+\delta x^\mu$ to the value of the unperturbed field variable at x^μ. Thus we write $\phi_\alpha(x)+\delta\phi_\alpha(x)$ for the value of $\phi_\alpha(x)$ at $x^\mu+\delta x^\mu$ under the perturbation. The value of $\phi_\alpha(x)$ at the point x is therefore $\phi_\alpha(x)+\delta\phi_\alpha(x)-\phi_{\alpha;\mu}(x)\delta x^\mu$.

Under perturbation, the value of the action integral is changed, due in part to the perturbation of the field, and in part to the perturbation of the volume of integration. The former may be evaluated by means of (2.5) if we replace $\delta\phi_\alpha$ by $\delta\phi_\alpha-\phi_{\alpha;\mu}\delta x^\mu$. The latter may be evaluated as a surface integral. Hence we find that

$$\delta\int\mathfrak{L}dx\rightarrow\int\frac{d}{dx^\mu}\{\pi^{\alpha\mu}(\delta\phi_\alpha-\phi_{\alpha;\nu}\delta x^\nu)\}dx+\int\mathfrak{L}\delta x^\mu dS_\mu, \tag{5.1}$$

where the first integral on the right-hand side is over the volume V and the second integral is over the surface bounding V. We now see that the total effect of the perturbation may be ascribed to an equivalent perturbation of $\mathfrak{L}$ given by

$$\delta\mathfrak{L}=(d/dx^\mu)(\pi^{\alpha\mu}\delta\phi_\alpha-T_\nu{}^\mu\delta x^\nu), \tag{5.2}$$

where the stress tensor $T_\mu{}^\nu$ has already been defined by (3.7).

It follows immediately from (5.2) that the extended form of the differential divergence-free vector, and the equation which it satisfies, are as follows:

$$\frac{d}{dx^\mu}\{(\delta\pi^{\alpha\mu}d\phi_\alpha-\delta T_\nu{}^\mu dx^\nu)-(d\pi^{\alpha\mu}\delta\phi_\alpha-dT_\nu{}^\mu\delta x^\nu)\}=0. \tag{5.3}$$

The corresponding form of the divergence-free vector

[19] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 479.

[20] G. S. Kino, *Proceedings of the Symposium on Electronic Waveguides* (Polytechnic Institute of Brooklyn, 1958), pp. 269–81.

bracket is

$$\{A,B\}^\mu=\left(\frac{\partial\pi^{\alpha\mu}}{\partial A}\frac{\partial\phi_\alpha}{\partial B}-\frac{\partial T_\nu{}^\mu}{\partial A}\frac{\partial x^\nu}{\partial B}\right)-\left(\frac{\partial\pi^{\alpha\mu}}{\partial B}\frac{\partial\phi_\alpha}{\partial A}-\frac{\partial T_\nu{}^\mu}{\partial B}\frac{\partial x^\nu}{\partial A}\right). \quad (5.4)$$

VI. INTEGRAL DIVERGENCE-FREE VECTOR

We now look for a vector which is related to the differential divergence-free vector in the same way as the Poincaré invariant[2] is related to the Lagrange invariant.[1] If the increments appearing in (2.9) are assumed to be due to an increment δA of the parameter A and an increment dB of the parameter B, we find that

$$\oint\pi^{\alpha\mu}d\phi_\alpha=\delta\pi^{\alpha\mu}d\phi_\alpha-d\pi^{\alpha\mu}\delta\phi_\alpha=\{A,B\}^\mu\delta A\,dB, \quad (6.1)$$

where the contour integral is taken around the parallelogram with vertices (A,B), $(A+\delta A,B)$, $(A+\delta A,B+dB)$, $(A,B+dB)$. The last term of (6.1) may now be integrated over a finite region of the $A-B$ plane; since each contribution to the integral satisfies (2.17), so does the complete integral. Hence we arrive at the result

$$\frac{d}{dx^\mu}\oint\pi^{\alpha\mu}\frac{\partial\phi_\alpha}{\partial\kappa}d\kappa=0, \quad (6.2)$$

where we have introduced κ as the cyclic parameter enumerating points on the boundary of the region of integration of the $A-B$ plane. The vector appearing in (6.2) is referred to as the "integral divergence-free vector." The relation (6.2) may be obtained more directly by integrating (2.5) over a closed ensemble.

The Schrödinger equation, which may be derived from the Lagrangian function[21]

$$\mathfrak{L}=\tfrac{1}{2}i\hbar(\phi^*\phi_{;0}-\phi^*{}_{;0}\phi)-\frac{\hbar^2}{2m}\phi^*{}_{;r}\phi_{;r}-V\phi^*\phi, \quad (6.3)$$

offers an interesting example of the application of the integral divergence-free vector. We write x_0 for t and treat all indices as suffixes; ϕ and ϕ^* are to be regarded as independent dynamical variables. The momentum vectors are found from (2.3) to be

$$\pi_0=\tfrac{1}{2}i\hbar\phi^*,\quad \pi_r=-\frac{\hbar^2}{2m}\phi^*{}_{;r},$$
$$\pi^*{}_0=-\tfrac{1}{2}i\hbar\phi,\quad \pi^*{}_r=-\frac{\hbar^2}{2m}\phi_{;r}. \quad (6.4)$$

We now note that the dynamical variables admit the following gauge transformation:

$$\phi\rightarrow\phi e^{-i\kappa},\quad \phi^*\rightarrow\phi^*e^{i\kappa}, \quad (6.5)$$

[21] Reference 18, p. 314.

and construct the integral divergence-free vector associated with the phase angle κ:

$$\left.\begin{aligned}P_0&=\oint d\kappa\left(\pi_0\frac{\partial\phi}{\partial\kappa}+\pi_0{}^*\frac{\partial\phi^*}{\partial\kappa}\right),\\ P_r&=\oint d\kappa\left(\pi_r\frac{\partial\phi}{\partial\kappa}+\pi_r{}^*\frac{\partial\phi^*}{\partial\kappa}\right).\end{aligned}\right\} \quad (6.6)$$

Evaluation of these integrals shows that

$$P_0=h\rho,\quad P_r=hj_r, \quad (6.7)$$

where

$$\rho=\phi^*\phi,\quad j_r=\frac{1}{2}\frac{i\hbar}{m}(\phi^*{}_{;r}\phi-\phi^*\phi_{;r}), \quad (6.8)$$

so that (6.2) yields the "equation of conservation of probability"[22]:

$$\partial\rho/\partial t+(\partial j_r/\partial x_r)=0. \quad (6.9)$$

We may obtain an extension of the integral divergence-free vector analogous to that of the differential divergence-free vector derived in Sec. V. If we provide for variations of the independent variables as well as for variations of the dependent variables, we find from (5.4) that the appropriate extension of (6.2) is

$$\frac{d}{dx^\mu}\oint d\kappa\left\{\pi^{\alpha\mu}\frac{\partial\phi_\alpha}{\partial\kappa}-T_\nu{}^\mu\frac{\partial x^\nu}{\partial\kappa}\right\}=0. \quad (6.10)$$

As an interesting example of the application of this relation, we shall derive the relationship between the energy-momentum vector and the wave vector for a plane wave of a homogeneous Hamiltonian wave-propagating system.

We consider a solution of the Euler-Lagrange equations for which $\pi^{\alpha\mu}$ and ϕ_α are expressible as functions of θ alone, where

$$\theta=k_\mu x^\mu. \quad (6.11)$$

We consider (6.10) as it applies to averages over the spatial variables x^1, x^2, x^3:

$$\oint d\kappa\left\{\left\langle\pi^{\alpha0}\frac{\partial\phi_\alpha}{\partial\kappa}\right\rangle-\left\langle T_\nu{}^0\frac{\partial x^\nu}{\partial\kappa}\right\rangle\right\}=\text{const.} \quad (6.12)$$

The parameter κ must enumerate a cyclic family of wave functions (the analog of a "tube" of trajectories in dynamics), but the precise way in which κ performs this enumeration may vary in time. We consider the particular case that all wave functions enumerated by κ are identical (corresponding to a tube collapsed on itself). We assume that κ merely determines a local shift of the spatial components of the coordinate system,

$$x^r\rightarrow x^r+\kappa\xi^r(x^0), \quad (6.13)$$

[22] Reference 15, p. 24.

but that there is no concomitant change of the field. It is sufficient, for our purposes, that ξ^r should depend only on x^0.

It is clear, from our definition of κ, that the integral (6.12) is zero. This may otherwise be established by evaluating (6.12) at a value of x^0 for which $\xi^r=0$. We now evaluate (6.12) for a value of x^0 for which $\xi^r=0$ if $r\neq r'$, but $\xi^{r'}=k_{r'}{}^{-1}$. We note from (6.11) that, since there is no variation of the wave function in the fixed coordinate system,

$$\theta\rightarrow\theta+\kappa. \tag{6.14}$$

Hence, if the integration in (6.12) is taken over the range 0 to 2π, we obtain

$$\oint d\theta\left\langle\pi^{\alpha 0}\frac{\partial\phi_\alpha}{\partial\theta}\right\rangle-\oint d\kappa\langle T_{r'}{}^0\rangle k_{r'}{}^{-1}=0, \tag{6.15}$$

which may be rewritten as

$$\langle T_r{}^0\rangle=J^0k_r, \tag{6.16}$$

where we now take note of the fact that r' is arbitrary, and where $2\pi J^0$ is to be interpreted as the "action density":

$$J^0=\frac{1}{2\pi}\oint d\theta\pi^{\alpha 0}\frac{\partial\phi_\alpha}{\partial\theta}. \tag{6.17}$$

The integration over θ, which occurs in (6.17), ensures that J^0 is independent of x. It is worth noticing that we have nowhere assumed that the system under discussion is described by linear field equations.

Further discussion of the stress tensor for plane waves in Hamiltonian media has been given elsewhere.[23] This example was introduced here merely to demonstrate that certain properties of this stress tensor may be derived by means of the extended integral divergence-free vector.

VII. DISCUSSION

We saw, in the example concerning the Schrödinger equation in Sec. VI, that we may set up a conservation relation whenever symmetry properties of the field equations enable one to associate a cyclic family of wave-functions with any particular wave-function. This is as one would expect. However the divergence-free vector bracket (2.18) appears to associate a "conservation theorem" with every pair of infinitesimal transformations, whereas one expects just one conservation quantity to be associated with invariance under any one infinitesimal transformation. It has been found, in examples which have so far been investigated, that each of the "surplus" conservation relations is a repetition of a relation associated with a single differential transformation, or is such that each component of the vector vanishes identically, or has a form which one would consider inappropriate for a conservation relation because, for instance, the components involve second, or higher, spatial derivatives of the field variables. It seems that the conventional understanding of what constitutes a conservation relation is such that just one such relation is associated with invariance under any one infinitesimal transformation.

The close relationship which has been noted elsewhere[24] between the Poincaré invariant and adiabatic invariants leads one to expect that the integral divergence-free vector discussed in Sec. VI should lead to the concept of an "adiabatically divergence-free vector." Consider a wave-like solution of the field equations in an inhomogeneous medium; we may establish a formula for the vector (6.2), where κ is now a phase parameter, on the assumption that the medium is locally uniform. We should now expect that the flux, over a closed volume, of the vector given by this formula would approximately vanish even in an inhomogeneous medium provided that the medium varies only slowly by comparison with the periodicity of the wave function.

Since we have established a field-theory analog of the Lagrange bracket, one is tempted to look for a field-theory analog of the Poisson bracket[6] also. In classical dynamics, the family of "trajectories" which may be followed by a given system may be enumerated by a finite number of parameters; a set of Lagrange brackets is then determined, which are constant in time. Since the "state" of the system at any time determines its preceding and subsequent behavior, these parameters are also expressible as functions of the coordinates and momenta describing the system; the set of Poisson brackets which one may associate with these functions form a matrix which is inverse to the matrix formed by the Lagrange brackets, so that the Poisson brackets also are constant in time.

The preceding argument does not seem to lend itself to generalization for field theory. The totality of possible wave functions may be enumerated by a suitable set of parameters, but these parameters are not determined by, and so cannot be expressed in terms of, the "state" of the system at one point of space-time. Hence one cannot expect the divergence-free vector bracket to be of any help in establishing a field-theory analog of the Poisson bracket.

The familiar formula for the Poisson brackets suggests the following field-theory analog: To any functions F,G of the field variables $\pi^{\alpha\mu}$, ϕ_α, we associate a vector

$$[F,G]_\mu=\frac{\partial F}{\partial\pi^{\alpha\mu}}\frac{\partial G}{\partial\phi_\alpha}-\frac{\partial F}{\partial\phi_\alpha}\frac{\partial G}{\partial\pi^{\alpha\mu}}. \tag{7.1}$$

This quantity, being a function of $\pi^{\alpha\mu}$, ϕ_α, is thereby expressible as a function of x. Since Poisson brackets are constant in time, one would hope that the expres-

[23] P. A. Sturrock, Phys. Rev. 121, 18 (1961).

[24] P. A. Sturrock, *Static and Dynamic Electron Optics* (Cambridge University Press, London, 1955), pp. 159 ff., 196 ff.

sions (7.1) would be divergence-free vectors; this, however, is not likely, since (7.1) has the form of a covariant vector rather than a contravariant vector. One may readily verify from this formula and (2.12) that the spatial derivative of a function is not expressible as the bracket combination of this function with the Hamiltonian. These facts suggest that the expression (7.1) is of no interest; this is confirmed conclusively by the demonstration, which we shall now give, that (7.1) is not canonically invariant.

We consider an infinitesimal canonical transformation

$$\left.\begin{aligned}\pi^{\alpha\mu} &\rightarrow \Pi^{\alpha\mu}=\pi^{\alpha\mu}+\Delta\pi^{\alpha\mu},\\ \phi_\alpha &\rightarrow \Phi_\alpha=\phi_\alpha+\Delta\phi_\alpha.\end{aligned}\right\} \tag{7.2}$$

The change in the bracket expression (7.1) may now be evaluated by noting that

$$\left.\begin{aligned}\frac{\partial}{\partial\Pi^{\alpha\mu}}&=\frac{\partial}{\partial\pi^{\alpha\mu}}-\frac{\partial\Delta\pi^{\beta\nu}}{\partial\pi^{\alpha\mu}}\frac{\partial}{\partial\pi^{\beta\nu}}-\frac{\partial\Delta\phi_\beta}{\partial\pi^{\alpha\mu}}\frac{\partial}{\partial\phi_\beta},\\ \frac{\partial}{\partial\Phi_\alpha}&=\frac{\partial}{\partial\phi_\alpha}-\frac{\partial\Delta\pi^{\beta\nu}}{\partial\phi_\alpha}\frac{\partial}{\partial\pi^{\beta\nu}}-\frac{\partial\Delta\phi_\beta}{\partial\phi_\alpha}\frac{\partial}{\partial\phi_\beta}.\end{aligned}\right\} \tag{7.3}$$

If we also note from (2.14) that the infinitesimal canonical transformation (7.2) may be derived from a generating function $\Delta V^\mu(\pi^{\alpha\nu},\phi_\beta,x)$ according to

$$\Delta\pi^{\alpha\mu}=\frac{\partial\Delta V^\mu}{\partial\phi_\alpha},\quad \Delta\phi_\alpha\delta_\nu{}^\mu=-\frac{\partial\Delta V^\mu}{\partial\pi^{\alpha\nu}}, \tag{7.4}$$

we find that the transformation (7.2) leads to the following change in the bracket expression (7.1):

$$\Delta[F,G]_\mu=\frac{\partial^2\Delta V^\nu}{\partial\pi^{\alpha\mu}\partial\pi^{\beta\nu}}\left\{\frac{\partial F}{\partial\phi_\alpha}\frac{\partial G}{\partial\phi_\beta}-\frac{\partial F}{\partial\phi_\beta}\frac{\partial G}{\partial\phi_\alpha}\right\}$$
$$-\frac{\partial^2\Delta V^\nu}{\partial\phi_\alpha\partial\phi_\beta}\left\{\frac{\partial F}{\partial\pi^{\alpha\mu}}\frac{\partial G}{\partial\pi^{\beta\nu}}-\frac{\partial F}{\partial\pi^{\beta\nu}}\frac{\partial G}{\partial\pi^{\alpha\mu}}\right\}. \tag{7.5}$$

We note that, if μ and ν took only one value, (7.5) would vanish in agreement with our knowledge that the expression (7.1) would then be canonically invariant. However, let μ, ν, take the values 1,2; α,β will take one value only so that these indices may be discarded. If we now choose F,G to be π^1, π^2, and adopt $\Delta V^1=\epsilon\phi^2$, $\Delta V^2=0$, we find from (7.5) that

$$\Delta[F,G]_1=0,\quad \Delta[F,G]_2=2\epsilon. \tag{7.6}$$

Hence we have proved, by means of a counter-example, that the expression (7.1) is not canonically invariant.

The terminology "differential divergence-free vector" and "integral divergence-free vector" is rather cumbersome. A nomenclature which would be briefer and emphasize the connection with the concepts of classical dynamics would be to name these quantities "Lagrange vector" and "Poincaré vector," respectively.

ACKNOWLEDGMENTS

I am indebted to Dr. D. Finkelstein and Dr. J. M. Jauch for their criticism of an earlier draft of this article.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 1 JANUARY-FEBRUARY, 1962

Spin and Statistics with an Electromagnetic Field

DAVID G. BOULWARE*
Harvard University, Cambridge, Massachusetts
(Received August 3, 1961)

The connection between spin and statistics is proved in the presence of an electromagnetic field. Two functions, corresponding to the Wightman functions for the two orderings of Lorentz gauge operators, are exhibited as expectation values of radiation gauge operators. These are shown to be Lorentz invariant and have the requisite positive frequency characteristics. The radiation gauge commutation relations on an $x^0=$const surface imply that the corresponding Lorentz gauge functions vanish for spacelike separations. Assuming the "wrong" connection between spin and statistics, the anticommutator functions are shown to be zero and continued to the origin. At the origin, the Lorentz gauge functions are simply the radiation gauge operator expectation values. Hence, the wrong commutation relations imply the vanishing of the radiation gauge field anticommutator at the origin, which is inconsistent with a positive definite metric and a non-vanishing field.

I. INTRODUCTION

THE proofs of the connection between spin and statistics have depended on the assumption of manifest Lorentz covariance and a positive definite metric in Hilbert space[1] or on somewhat restrictive assumptions as to the nature of the dynamics.[2] The presence of an electromagnetic field requires that we either use an indefinite metric or drop manifest covariance. In either case, the proofs[1] are not valid. On the other hand, Pauli's[2] work is restricted to the case of

* National Science Foundation Predoctoral Fellow.

[1] G. Luders and B. Zumino, Phys. Rev. **110**, 1450 (1958); N. Burgoyne, Nuovo cimento **8**, 607 (1958).

[2] J. Schwinger, Proc. Natl. Acad. Sci. U. S. **44**, 223, 617 (1958); W. Pauli, Phys. Rev. **58**, 716 (1940).

sions (7.1) would be divergence-free vectors; this, however, is not likely, since (7.1) has the form of a covariant vector rather than a contravariant vector. One may readily verify from this formula and (2.12) that the spatial derivative of a function is not expressible as the bracket combination of this function with the Hamiltonian. These facts suggest that the expression (7.1) is of no interest; this is confirmed conclusively by the demonstration, which we shall now give, that (7.1) is not canonically invariant.

We consider an infinitesimal canonical transformation

$$\left.\begin{aligned}\pi^{\alpha\mu} &\rightarrow \Pi^{\alpha\mu}=\pi^{\alpha\mu}+\Delta\pi^{\alpha\mu},\\ \phi_\alpha &\rightarrow \Phi_\alpha=\phi_\alpha+\Delta\phi_\alpha.\end{aligned}\right\} \tag{7.2}$$

The change in the bracket expression (7.1) may now be evaluated by noting that

$$\left.\begin{aligned}\frac{\partial}{\partial\Pi^{\alpha\mu}}&=\frac{\partial}{\partial\pi^{\alpha\mu}}-\frac{\partial\Delta\pi^{\beta\nu}}{\partial\pi^{\alpha\mu}}\frac{\partial}{\partial\pi^{\beta\nu}}-\frac{\partial\Delta\phi_\beta}{\partial\pi^{\alpha\mu}}\frac{\partial}{\partial\phi_\beta},\\ \frac{\partial}{\partial\Phi_\alpha}&=\frac{\partial}{\partial\phi_\alpha}-\frac{\partial\Delta\pi^{\beta\nu}}{\partial\phi_\alpha}\frac{\partial}{\partial\pi^{\beta\nu}}-\frac{\partial\Delta\phi_\beta}{\partial\phi_\alpha}\frac{\partial}{\partial\phi_\beta}.\end{aligned}\right\} \tag{7.3}$$

If we also note from (2.14) that the infinitesimal canonical transformation (7.2) may be derived from a generating function $\Delta V^\mu(\pi^{\alpha\nu},\phi_\beta,x)$ according to

$$\Delta\pi^{\alpha\mu}=\frac{\partial\Delta V^\mu}{\partial\phi_\alpha},\quad \Delta\phi_\alpha\delta_\nu{}^\mu=-\frac{\partial\Delta V^\mu}{\partial\pi^{\alpha\nu}}, \tag{7.4}$$

we find that the transformation (7.2) leads to the following change in the bracket expression (7.1):

$$\Delta[F,G]_\mu=\frac{\partial^2\Delta V^\nu}{\partial\pi^{\alpha\mu}\partial\pi^{\beta\nu}}\left\{\frac{\partial F}{\partial\phi_\alpha}\frac{\partial G}{\partial\phi_\beta}-\frac{\partial F}{\partial\phi_\beta}\frac{\partial G}{\partial\phi_\alpha}\right\}$$
$$-\frac{\partial^2\Delta V^\nu}{\partial\phi_\alpha\partial\phi_\beta}\left\{\frac{\partial F}{\partial\pi^{\alpha\mu}}\frac{\partial G}{\partial\pi^{\beta\nu}}-\frac{\partial F}{\partial\pi^{\beta\nu}}\frac{\partial G}{\partial\pi^{\alpha\mu}}\right\}. \tag{7.5}$$

We note that, if μ and ν took only one value, (7.5) would vanish in agreement with our knowledge that the expression (7.1) would then be canonically invariant. However, let μ, ν, take the values 1,2; α,β will take one value only so that these indices may be discarded. If we now choose F,G to be π^1, π^2, and adopt $\Delta V^1=\epsilon\phi^2$, $\Delta V^2=0$, we find from (7.5) that

$$\Delta[F,G]_1=0,\quad \Delta[F,G]_2=2\epsilon. \tag{7.6}$$

Hence we have proved, by means of a counter-example, that the expression (7.1) is not canonically invariant.

The terminology "differential divergence-free vector" and "integral divergence-free vector" is rather cumbersome. A nomenclature which would be briefer and emphasize the connection with the concepts of classical dynamics would be to name these quantities "Lagrange vector" and "Poincaré vector," respectively.

ACKNOWLEDGMENTS

I am indebted to Dr. D. Finkelstein and Dr. J. M. Jauch for their criticism of an earlier draft of this article.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 1 JANUARY-FEBRUARY, 1962

Spin and Statistics with an Electromagnetic Field

DAVID G. BOULWARE*
Harvard University, Cambridge, Massachusetts
(Received August 3, 1961)

The connection between spin and statistics is proved in the presence of an electromagnetic field. Two functions, corresponding to the Wightman functions for the two orderings of Lorentz gauge operators, are exhibited as expectation values of radiation gauge operators. These are shown to be Lorentz invariant and have the requisite positive frequency characteristics. The radiation gauge commutation relations on an $x^0=$const surface imply that the corresponding Lorentz gauge functions vanish for spacelike separations. Assuming the "wrong" connection between spin and statistics, the anticommutator functions are shown to be zero and continued to the origin. At the origin, the Lorentz gauge functions are simply the radiation gauge operator expectation values. Hence, the wrong commutation relations imply the vanishing of the radiation gauge field anticommutator at the origin, which is inconsistent with a positive definite metric and a non-vanishing field.

I. INTRODUCTION

THE proofs of the connection between spin and statistics have depended on the assumption of manifest Lorentz covariance and a positive definite metric in Hilbert space[1] or on somewhat restrictive assumptions as to the nature of the dynamics.[2] The presence of an electromagnetic field requires that we either use an indefinite metric or drop manifest covariance. In either case, the proofs[1] are not valid. On the other hand, Pauli's[2] work is restricted to the case of

* National Science Foundation Predoctoral Fellow.

[1] G. Luders and B. Zumino, Phys. Rev. **110**, 1450 (1958); N. Burgoyne, Nuovo cimento **8**, 607 (1958).

[2] J. Schwinger, Proc. Natl. Acad. Sci. U. S. **44**, 223, 617 (1958); W. Pauli, Phys. Rev. **58**, 716 (1940).

free fields, and, while it contains all the essential points, is not readily generalizable. Schwinger,[2] while going to interacting fields, is forced to assume either that the kinematic terms of the Lagrangian dominate the spectrum or that the different charge degrees of freedom are not kinematically coupled.

In the present work, we show that the usual connection between spin and statistics follows from the basic assumptions of the theory, a knowledge of the electromagnetic interaction, and the assumption that the theory is manifestly Lorentz covariant with the exception of the electromagnetic field. This last assumption will be discussed at greater length below.

In Sec. II, we exhibit a radiation gauge operator product (which is related to the Lorentz gauge Green's function) and prove that it is manifestly Lorentz covariant. Section III is devoted to showing that the operator product is a positive frequency function. Then, using analyticity properties and assuming canonical- (anti-) commutation relations for the independent field components, we show in Sec. IV that the function corresponding to the Lorentz gauge anticommutator must vanish everywhere. The exponentials in the operator product all go to zero at the origin and the anticommutator function becomes simply the expectation value of the anticommutator of the fields. The vanishing of this anticommutator is then shown, assuming a positive definite metric in Hilbert space, to be equivalent to the vanishing of the field. In the Appendixes we give a derivation of the relation between the Lorentz gauge Green's functions and radiation gauge Green's functions and give a short discussion of the electromagnetic field commutation properties.

II. LORENTZ TRANSFORMATIONS

Since we must make detailed use of the electromagnetic field coupling, we introduce the Lagrangian to begin with. As the addition of an anomalous magnetic moment[3] involves no new ideas, only additional complication of the constraint equations, we assume only a minimal coupling. We then take as our Lagrangian

$$\mathfrak{L}=-\tfrac{1}{4}\chi a^\mu(1/i)\partial_\mu\chi+\tfrac{1}{4}(\partial_\mu\chi)a^\mu(1/i)\chi-H(\chi) -\tfrac{1}{2}F^{\mu\nu}(\partial_\mu A_\nu-\partial_\nu A_\mu)+\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu}+j_\mu A^\mu, \quad (1)$$

where, formally $j^\mu=\frac{1}{2}e\chi a^\mu q\chi$.

The χ's are Hermitian field operators; H is restricted only in that it be Hermitian and lead to a positive definite energy spectrum and charge conservation. The matrices a^μ must be Hermitian and, since we assume that the fields obey either commutation or anticommutation relations, they must be reducible into symmetric and antisymmetric parts referring to Fermi and Bose statistics, respectively.[4] The charge matrix q is imaginary, antisymmetric and, assuming that the fields only carry charges ±1 or 0; $q^3=q$. The metric tensor has diagonal elements $(-1,1,1,1)$.

Subsequently, we will need the commutator of the charge density j^0 with the field. To derive that commutator, consider the variation

$$\delta\chi=-ieq\chi(x)\delta\lambda(x).$$

Then,

$$\delta W=G_2-G_1=\delta\left[\int_{\sigma_1}^{\sigma_2}dx\mathfrak{L}\right]$$

$$=\int_{\sigma_1}^{\sigma_2}dx(\partial_\mu\delta\lambda(x))j^\mu(x)$$

$$G_2=\int_{\sigma_2}d\sigma_\mu j^\mu(x)\delta\lambda(x)$$

and

$$-ieq\chi(x)\delta\lambda(x)=(1/i)[\chi(x),G] =-i\int_\sigma d\sigma_\mu'[\chi(x),j^\mu(x')]\delta\lambda(x').$$

Hence,

$$\delta(x_0-x_0')[\chi(x),j^0(x')]=eq\chi(x)\delta(x-x'), \quad (2)$$

where we have taken our quantization on an $x^0=$const surface.

Now, we must consider the electromagnetic field in greater detail. The equations of motion are given by:

$$F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu$$

$$\partial_\nu F^{\mu\nu}=j^\mu,$$

which must be rewritten in terms of the independent components, ${}^TA^k$ and ${}^TF^{0k}$, where ${}^TB^k$ means the transverse part of B^k, and a latin index means only spatial components. Then

$$\begin{aligned} A^0&=\partial^0\Lambda-(1/\nabla^2)j^0\\ A^k&={}^TA^k+\partial^k\Lambda\\ F^{kl}&=\partial^k{}^TA^l-\partial^l{}^TA^k\\ {}^LF^{0l}&=+\partial^l(1/\nabla^2)j^0 \end{aligned} \quad (3)$$

are the constraint equations and the equations of motion are:

$$\begin{aligned} \partial_0{}^TF^{0k}&=-\nabla^2{}^TA^k-{}^Tj^k\\ \partial_0{}^TA^k&=-{}^TF^{0k}. \end{aligned} \quad (4)$$

The gauge Λ is undetermined and will be taken to be zero. However, if we consider a variation in the gauge $\delta\Lambda$, the generator turns out to be

$$G=-\int d\sigma_\mu j^\mu(x)\delta\Lambda(x),$$

which only generates a change in the charged fields as we saw before. Then, the generator of the combined transformation $\delta\chi=-ieq\delta\Lambda\chi$; $\delta A^\mu=\partial^\mu\delta\Lambda$ vanishes. This

[3] Lowell Brown, Ph.D. thesis (unpublished).
[4] J. Schwinger, Phys. Rev. **91**, 713 (1953).

is necessary for the coupling of a gauge transformation to the Lorentz transformation. The Lorentz transformation,

$$\begin{aligned}\delta\chi&=\tfrac{1}{2}(x^\mu\partial^\nu-x^\nu\partial^\mu+iS^{\mu\nu})\epsilon_{\mu\nu}\chi\\ \delta A^\lambda&=\tfrac{1}{2}\epsilon_{\mu\nu}[(x^\mu\partial^\nu-x^\nu\partial^\mu)A^\lambda-g^{\lambda\nu}A^\mu+g^{\lambda\mu}A^\nu],\end{aligned}\tag{5}$$

under which the Lagrangian is formally invariant, yields the generators $J^{\mu\nu}$. However, the assumed variation of the fields is inconsistent with the transverse nature of the vector potential A^k. Hence, a gauge transformation is induced. Now, for consistency, the actual variation of the fields must also yield the generators $J^{\mu\nu}$. Thus, the charged fields must undergo the corresponding gauge transformation, as we will see.

We can now make explicit our assumption that "the theory is manifestly Lorentz covariant with the exception of the electromagnetic field." We assume that either there are no further invariance operations whose generators vanish, or that, if they exist, they are not induced by the Lorentz transformation.

It is straightforward to calculate the infinitesimal gauge transformation from the condition that the transversality of the vector potential be maintained. However, it is somewhat easier and perhaps enlightening to observe that any physically observable quantity such as the fields $F^{\mu\nu}$ and the current density j^μ must be manifestly Lorentz covariant. Then

$$(1/i)[j^\mu,J^{0k}]=(x^0\partial^k-x^k\partial^0)j^\mu-(g^{\mu k}j^0-g^{\mu 0}j^k).\tag{6}$$

Now, we may use the constraint equations to find

$$\begin{aligned}&(1/i)[A^0,J^{0k}]\\ &\quad=(x^0\partial^k-x^k\partial^0)A^0-\partial^0[x^k,(1/\nabla^2)]j^0+(1/\nabla^2)j^k,\end{aligned}$$

which can be rewritten, using (3) and (4), as

$$\begin{aligned}&\frac{1}{i}[A^0,J^{0k}]=(x^0\partial^k-x^k\partial^0)A^0-A^k-\partial^0\Lambda^{0k}\\ &\Lambda^{0k}(x)=\int\frac{d^3\tau'}{4\pi r'}[F^{0k}+x'^kj^0](\mathbf{r}'+\mathbf{r},x^0).\end{aligned}\tag{7}$$

Then, since

$$\begin{aligned}(1/i)[F^{\mu\nu},J^{0k}]&=(x^0\partial^k-x^k\partial^0)F^{\mu\nu}\\ &\quad-(g^{\mu k}F^{0\nu}-g^{\mu 0}F^{k\nu}+g^{\nu k}F^{\mu 0}-g^{\nu 0}F^{\mu k}),\end{aligned}$$

and $F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu$ we must have

$$(1/i)[A^l,J^{0k}]=(x^0\partial^k-x^k\partial^0)A^l-g^{kl}A^0-\partial^l\Lambda^{0k},\tag{8}$$

and in accordance with our previous arguments;

$$(1/i)[\chi,J^{0k}]=(x^0\partial^k-x^\lambda\partial^0+iS^{0k})\chi+ieq\Lambda^{0k}\cdot\chi,\tag{9}$$

where the $\cdot$ denotes a symmetrized product. If the gauge Λ^{0k} in Eq. (9) were not equal to that in Eq. (7) and (8) then inserting these expressions into the Action principle would yield the generator

$$G=J^{0k}+\int d\tau j^0(\Lambda-\Lambda'),$$

and an inconsistency, since J^{0k} alone generates the Lorentz transformation.

Before introducing the various operator products, it is convenient to define uncharged Hermitian and charged nonHermitian fields

$$\begin{aligned}\varphi&=(1-q^2)\chi=\varphi^\dagger\\ \psi&=\frac{q^2+q}{\sqrt{2}}\chi=\left[\frac{q^2-q}{\sqrt{2}}\chi\right]^\dagger.\end{aligned}\tag{10}$$

Then, the following products vanish identically

$$\langle\varphi\psi\rangle\quad\langle\varphi\psi^\dagger\rangle\quad\langle\psi\psi\rangle$$

along with their complex conjugates

$$\langle\psi^\dagger\varphi\rangle\quad\langle\psi\varphi\rangle\quad\langle\psi^\dagger\psi^\dagger\rangle$$

from charge conservation.

$\langle B\rangle=\langle 0|B|0\rangle$ where $\langle 0|$ is assumed to be the nondegenerate Lorentz invariant ground state.

The time ordered operator products:

$$\begin{aligned}&G^{(L)}(x,x')\\ &=\theta(x^0-x'^0)G_>^{(L)}(x,x')\pm\theta(x'^0-x^0)G_<^{(L)}(x,x')\\ &=\begin{cases}\langle(\varphi(x)\varphi(x'))_+\rangle\epsilon(x,x') & \text{for uncharged fields}\\ e^{-\zeta(x-x')}\left\langle\left(\exp\left[-ie\int f(x,x',y)A^0(y)d^4y\right]\right.\right.\\ \qquad\left.\left.\times\psi(x)\psi^\dagger(x')\right)_+\right\rangle\epsilon(x,x') & \text{for charged fields.}\end{cases}\end{aligned}$$

$$f(x,x',y)=\lim_{\epsilon\to 0}\int\frac{d^4k}{(2\pi)^4}\frac{ik^0}{k^2-i\epsilon}[e^{ik(x-y)}-e^{ik(x'-y)}]$$

$$\zeta(x-x')=\lim_{\epsilon\to 0}\int\frac{d^4k}{(2\pi)^4}\frac{ie^2k^{0^2}}{(k^2-i\epsilon)^2\mathbf{k}^2}[1-e^{ik(x-x')}]$$

$$\epsilon(y)=\begin{cases}+1 & \text{Bose fields}\\ +1 & \text{Fermi fields}\quad y>0\\ -1 & \text{Fermi fields}\quad y<0\end{cases}\tag{11}$$

are shown in the Appendix to be related to the four dimensional transverse Lorentz gauge Green's functions and, for independent components of the fields, they are equal to the Lorentz gauge Green's functions.

The work of Johnson and Zumino[5] indicates that Green's functions may not exist for an arbitrary Lorentz gauge; hence we must consider the possibility of further gauge transformations. Such a Lorentz invariant transformation can only add a scalar positive frequency

[5] K. Johnson and B. Zumino, Phys. Rev. Letters 3, 351 (1959); J. Schwinger, Phys. Rev. 117, 1407 (1960).

function of $(x-x')^2$, zero for $(x-x')^2=0$, to ζ, which in no way affects the Lorentz transformation or spectral properties of the functions. Thus, the necessity of such a transformation cannot affect the validity of the results, since all the arguments remain valid under the transformation.

We now show that the Lorentz invariance of the Green's functions follows directly from the nature of the operator products. The derivation in Appendix I indicates that this must be so; however, we want to show that the Lorentz invariance follows directly from the properties of the operators. The general statement of Lorentz invariance for a function of n variables and the associated spin indices is:

$$\sum_{j=1}^{n}(x_j^{\mu}\partial_j^{\nu}-x_j^{\nu}\partial_j^{\mu}+iS_j^{\mu\nu})f(x_1\cdots x_n)=0.$$

The function (11) is trivially invariant under spatial rotations since they do not induce gauge transformations, and A^0 is unchanged. Now, since the vacuum is invariant, and considering first the uncharged fields

$$0=\langle(1/i)[(\varphi(x)\varphi(x'))_+,J^{0k}]\rangle\epsilon(x^0-x'^0)$$
$$=\langle\{[x^0\partial^k-x^k\partial^0+iS^{0k}]\varphi(x)\varphi(x')+\varphi(x)[x'^0\partial'^k-x'^k\partial'^0+iS^{0k}]\varphi(x')\}_+\rangle\epsilon$$
$$=[x^0\partial^k-x^k\partial^0+x'^0\partial'^k-x'^k\partial'^0+i(S^{0k}+S'^{0k})]\times\langle(\varphi(x)\varphi(x'))_+\rangle\epsilon-(x^k-x'^k)\delta(x^0-x'^0)\times\langle(\varphi(x),\varphi(x'))_{\mp}\rangle.$$

The last term (arising from commuting the time derivative with time ordering), is zero unless $x^0=x'^0$. However, since we will be interested only in the unordered products, we may drop this term, and observe that the ordered product is Lorentz invariant, provided the times are not equal. This apparently paradoxical statement is a consequence of the necessity of quantizing on a particular space like surface. Also, the independent components of the fields obey canonical commutation relations, thus the term vanishes since $x\delta(x)\equiv 0$.

We have considered the uncharged field in such detail so that when the same situation arises with the charged fields, we will be able to extract this particular complication.

We have, for the charged fields,

$$0=e^{-i\zeta}\left\langle(1/i)\left[\left(\exp\left[-ie\int fA^0\right]\psi(x)\psi^{\dagger}(x')\right)_+,J^{0k}\right]\right\rangle\epsilon$$
$$=e^{-i\zeta}\Bigg\langle-ie\int f(x,x',y)dy\Bigg\{[y^0\partial^k-y^k\partial^0]\left(\exp\left[-ie\int fA^0\right]A^0(y)\psi\psi^{\dagger}\right)_+-\left(\exp\left[-ie\int fA^0\right]A^k(y)\psi\psi^{\dagger}\right)_+$$
$$-\partial_y{}^0\left(\exp\left[-ie\int fA^0\right]\Lambda^{0k}(y)\psi\psi^{\dagger}\right)_+-(y^k-x^k)\delta(y^0-x^0)\left(\exp\left[-ie\int fA^0\right][A^0,\psi]\psi^{\dagger}\right)_+$$
$$-(y^k-x'^k)\delta(y^0-x'^0)\left(\exp\left[-ie\int fA^0\right]\psi[A^0,\psi^{\dagger}]\right)_+\Bigg\}\Bigg\rangle\epsilon(x^0-x'^0)+[x^0\partial^k-x^k\partial^0+x'^0\partial'^k-x'^k\partial'^0+i(S^{0k}+S'^{0k})]$$
$$\times G^{(L)}(x,x')+ie e^{-i\zeta}\left\langle\left(\exp\left[-ie\int fA^0\right][\Lambda^{0k}(x)\cdot\psi(x)\psi^{\dagger}(x')-\psi(x)\Lambda^{0k}(x')\cdot\psi^{\dagger}(x')\right)_+\right\rangle\epsilon$$
$$-\delta(y^0-x^0)e^{-i\zeta}\left\langle\left(\exp\left[-ie\int fA^0\right][\Lambda^{0k}(y),\psi(x)]\psi^{\dagger}(x')\right)_+\right\rangle\epsilon$$
$$-\delta(y^0-x'^0)e^{-i\zeta}\left\langle\left(\exp\left[-ie\int fA^0\right]\psi(x)[\Lambda^{0k}(y),\psi^{\dagger}(x')]\right)_+\right\rangle\epsilon. \quad (12)$$

The delta function terms come from commuting time derivatives with time ordering and the remaining terms from a straightforward application of the commutator of J^{0k} with the various fields. The x and x' derivatives refer only to the dependence of ψ and $\psi^{\dagger}$, not to the dependence of f on x and x'. We have neglected a term in $\delta(x^0-x'^0)$ coming from the commutator of ψ and $\psi^{\dagger}$ just as with the uncharged fields.

The following properties of f, Λ^{0k} and ζ may be readily verified:

$$\int f(x,x',y)\partial^0 dy g(y)=g(x)-g(x')-\int\left(\frac{1}{x\partial^2}-\frac{1}{x'\partial^2}\right)\nabla^2 g$$
$$(-\nabla^2\Lambda^{0k})(y)=F^{0k}(y)+\partial\partial^k A^0(y)=\partial^0 A^k+\partial^k A^0$$
$$\int f(x,x',y)(y^0\partial_y{}^k-y^k\partial_y{}^0)dy g(y)=[x^0\partial^k-x^k\partial^0+x'^0\partial'^k-x'^k\partial'^0]\int fg+\left[\frac{1}{x\partial^2}-\frac{1}{x'\partial^2}\right]\partial^k g$$
$$(\xi^0\partial^l-\xi^l\partial^l)\zeta(\xi)=2ie^2\int\frac{dk}{(2\pi)^4}\frac{k^0k^l}{k^2(\mathbf{k}^2)^2}[1-e^{ik\xi}].$$

We also need the commutator of j^0 with ψ and $\psi^\dagger$. Using Eqs. (10) and (2) we find

$$\delta(x^0-x'^0)[\psi(x),j^0(x')]=e\psi(x)\delta(x-x')$$
$$\delta(x^0-x'^0)[\psi^\dagger(x),j^0(x')]=-e\psi^\dagger(x')\delta(x-x').$$

Applying these results to Eq. (12) yields

$$0=[x^0\partial^l-x^l\partial^0+x'^0\partial'^l-x'^l\partial'^0+i(S^{0l}+S'^{0l})]G^{(L)}(x,x')$$
$$-ie^2\Big[\int f(x,x',y)dy\Big\{\frac{\delta(y^0-x^0)}{8\pi|\mathbf{y}-\mathbf{x}|}(y^l-x^l)$$
$$-\frac{\delta(y^0-x'^0)(y^l-x'^l)}{4\pi|\mathbf{y}-\mathbf{x}'|}\Big\}-2\int\frac{dk}{(2\pi)^4}\frac{k^0k^l}{k^2(\mathbf{k}^2)^2}$$
$$\times[1-e^{ik(x-x')}]\Big]G^{(L)}(x,x'),$$

where the x and x' derivatives now act on all f and ζ dependence. But

$$\int fdy\Big\{\frac{\delta(y^0-x^0)}{8\pi|\mathbf{y}-\mathbf{x}|}(y^l-x^l)-\frac{\delta(y^0-x'^0)(y^l-x'^l)}{8\pi|\mathbf{y}-\mathbf{x}'|}\Big\}$$
$$=\tfrac{1}{2}\int\frac{dkdk'dyik^0}{(2\pi)^8k^2}[e^{ik(x-y)}-e^{ik(x'-y)}]$$
$$\times[e^{ik'(y-x)}-e^{ik'(y-x')}]i\frac{\partial}{\partial k'^l}\frac{1}{\mathbf{k}'^2}$$
$$=2\int\frac{dk}{(2\pi)^4}\frac{k^0k^l}{k^2(\mathbf{k}^2)^2}[1-e^{ik(x-x')}].$$

Thus, considering $G_>{}^{(L)}$ and $G_<{}^{(L)}$ separately, and invoking translational invariance,

$$(\xi^0\partial^k-\xi^k\partial^0)(G_\gtrless{}^{(L)}(\xi)+i[S^{0k}G_\gtrless{}^{(L)}+G_\gtrless{}^{(L)}S^{0kT}]=0 \quad (13)$$

for $\xi^0\gtrless 0$. We have written the two point function as a matrix.

III. SPECTRAL PROPERTIES

The assumption of a ground state implies that vacuum expectation values of radiation gauge operators contain only positive frequencies. However, it must be shown that the Lorentz functions are also positive frequency functions. Their representation as Lorentz gauge operator expectation values is not sufficient since unphysical states are introduced which have no restrictions on their spectral forms.

The uncharged field, $G_>{}^{(L)}$ is easily shown to be a positive frequency function of $x-x'$:

$$\langle 0|\varphi(x)\varphi(x')|0\rangle=\int dp'\sum_n\langle 0|\varphi|np'\rangle\langle np'|\varphi|0\rangle e^{ip'\xi}$$
$$p^\mu|np'\rangle=p'^\mu|np'\rangle,$$

where $p'^0\geq 0$ and Lorentz invariance implies $p_n{}^2\leq 0$.

We consider the function $G_>{}^{(L)}(\xi)$ for charged fields. For $\xi^0>0$ we have the representation of Eq. (11), which we may rewrite as

$$G_>{}^{(L)}(\xi)=e^{-\zeta(\xi)}\Big\langle\Big(\exp\Big[-ie\int_0^\infty f(0,-\xi,y)A^0(y)\Big]\Big)_+$$
$$\times\psi(0)e^{ip\xi}\Big(\exp\Big[-ie\int_0^{\xi^0}f(\xi,0,y)A^0(y)\Big]\Big)_+$$
$$\times\psi^\dagger(0)\Big(\exp\Big[-ie\int_{-\infty}^0 f(\xi,0,y)A^0(y)\Big]\Big)_+\Big\rangle. \quad (14)$$

The functions $\zeta(\xi)$, $f(\xi,0,-y)$, and $f(0,-\xi,+y)$ are all positive frequency functions for $\xi^0>0$ and $y^0>-\xi^0$, as may be seen from their explicit forms in Eq. (11). Expanding the exponentials in a power series leaves only the dependence of the middle exponential on ξ^0, where the integration limit could give trouble. The presence of the $e^{iP\xi}$, however, insures that the function:

$$\Big\langle p'\Big|e^{iP\xi}\Big(\exp\Big[-ie\int_0^{\xi^0}f(\xi,0,y)A^0(y)\Big]\Big)_+\Big|p''\Big\rangle,$$

which consists of terms of the form

$$F_n(\xi,p',p'')$$
$$=\frac{(-ie)^n}{n!}\Big\langle p'\Big|e^{iP\xi}\int_0^{\xi^0}dy_1\cdots dy_nf(\xi,0,y_1)\cdots f(\xi,0,y_n)$$
$$\times[A^0(y_1)\cdots A^0(y_n)]_+\Big|p''\Big\rangle$$
$$=(-ie)\sum\langle p'|A^0(0)|p'''\rangle e^{ip'\xi}\int_0^{\xi^0}dyf(\xi,0,y)$$
$$\times F_{n-1}(y,p''',p'')e^{i(p'''-p')y}$$

is a positive frequency function. To prove this, we note that F_1 is a positive frequency function

$$F_1(\xi,p',p'')$$
$$=e^{ip'\xi}\langle p'|A^0(0)|p''\rangle\int_0^{\xi^0}dy\int\frac{dk}{(2\pi)^4}\frac{e^{ik\xi}-1}{k^2-i\epsilon}k^0e^{i(p''-p'-k)y}$$
$$=\frac{e}{2}\langle p'|A^0(0)|p''\rangle\Big\{\frac{\exp(i\mathbf{p}''\cdot\xi)[e^{-i(k+p'^0)\xi^0}-e^{-ip''^0\xi^0}]}{k+p'^0-p''^0}$$
$$+\frac{\exp(i\mathbf{p}'\cdot\xi)(e^{-ip'^0\xi^0}-e^{-i(k+p''^0)\xi^0})}{k+p''^0-p'^0}\Big\}\quad k=[(\mathbf{p}'-\mathbf{p}'')^2]^{\frac{1}{2}}$$

which is manifestly positive frequency.

Then, assuming F_{n-1} is a positive frequency function we may write it:

$$F_{n-1}(\xi,p',p'')=\int dK\theta(K^0)e^{iK\xi}\bar{F}_{n-1}(K,p',p'')$$

and

$$F_n(\xi,p',p'')$$
$$=e\sum_{p'''}\langle p'|A^0(0)|p'''\rangle e^{ip'\xi}\int_0^{\xi^0}dy\int\frac{dk}{(2\pi)^4}\frac{k^0}{k^2-it}$$
$$\times[e^{ik\xi}-1]e^{-iky}\int dKe^{iKy}\theta(K^0)\bar{F}_{n-1}(K,p''',p'')$$
$$=\int dKe^{iK\xi}\theta(K^0)\bar{F}_n(K,p',p'')$$

$$\bar{F}_n(K,p',p'')$$
$$=\sum_{p'''}\int dK'\tfrac{1}{2}e\langle p'|A^0(0)|p'''\rangle\bar{F}_{n-1}(K',p''',p'')$$
$$\times\left\{\frac{\delta(\mathbf{K}-\mathbf{p}'-\mathbf{K}')}{K'^0-|\mathbf{K}'|}[\delta(K^0-p'^0-|\mathbf{K}'|)\right.$$
$$-\delta(K^0-p'^0-K'^0)]+\frac{\delta(\mathbf{K}-\mathbf{p}'')}{K'^0+|\mathbf{K}'|}$$
$$\left.\times[\delta(K^0-p'^0)-\delta(K^0-p'^0-K'^0-|\mathbf{K}'|)]\right\}.$$

By induction, all the F_n's and hence $G_>^{(L)}$ are positive frequency functions. Similar arguments show that $G_<^{(L)}(\xi)$ is a negative frequency function.

Equation (13) and the assumption that only a finite number of finite dimensional representations of the Lorentz group occur allows us to write:

$$G_{\gtrless}^{(L)\lambda\sigma}(\xi)=\sum_{m=1}^{n}{}^{(m)}T^{\lambda\sigma}{}_{\mu_1\cdots\mu_{p_m}}\xi^{\mu_1}\cdots\xi^{\mu_{p_m}}f_{\gtrless}^{(m)}(\xi^2),\quad(15)$$

where

$$i(S^{\mu\nu}{}^{(m)}T_{\mu_1\cdots\mu_{p_m}}+{}^{(m)}T_{\mu_1\cdots\mu_{p_m}}S^{\mu\nu T})^{\lambda\sigma}$$
$$+\sum_{j=1}^{p_m}{}^{(m)}T^{\lambda\sigma}{}_{\mu_1\cdots\mu_{j-1}\nu_j\mu_{j+1}\cdots\mu_{p_m}}(g^{\nu i\nu}\delta_{\mu_j}{}^{\mu}-g^{\nu i\mu}\delta_{\mu_j}{}^{\nu})=0.$$

λ and σ are field induces, ${}^{(m)}T_{\mu_1}\cdots\mu_{p_m}$ a matrix. Then, assuming the existence, in the sense of generalized functions, of the $f^{(m)}$ for real ξ, the frequency conditions imply that $f_>^{(m)}(z)$ is analytic for $z=(\xi+i\eta)^2$ and $f_<^{(m)}(z)$ is analytic for $z=(\xi-i\eta)^2$ for $\eta^2<0$, $\eta^0<0$. The functions $G_{\gtrless}^{(L)}(\xi\pm i\eta)$ are known for $\eta=0$, $\xi^0\gtrless 0$; then we may continue analytically to find $G_{\gtrless}^{(L)}$ for all ξ.[6]

[6] D. Hall and A. S. Wightman, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **31**, No. 5 (1957).

IV. COMMUTATION RELATIONS

In this section, we will give expressions for the charged fields only. The uncharged fields are trivially analogous.

For $x^0=x'^0$, we assume that the fields obey Fermi or Bose commutation relations. Then, for $\xi^0=0$, $\xi^2>0$

$$0=e^{-t}\left\langle\left(\exp\left[-ie\int_0^{\infty}fA^0\right]\right)_+[\psi(x),\psi^\dagger(x')]_{\mp}\right.$$
$$\left.\times\left(\exp\left[-ie\int_{-\infty}^{0}fA^0\right]\right)_+\right\rangle$$
$$=G_>^{(L)}(\xi)\mp G_<^{(L)}(\xi).\quad(16)$$

But, since we have Lorentz invariance, the equation is true for all $\xi^2>0$.

Now, we must consider the integer and half integer spin fields separately. For the integer spin fields these may be taken as tensor fields, and, considering only fields within the same irreducible representation of the Lorentz group, there are an even number of tensor indices. Hence, in Eq. (15) p_m must be even for all m. Then, for integer spin, we must have

$$G_<^{(L)\lambda\sigma}(\xi)=G_<^{(L)\lambda\sigma}(-\xi).\quad(17)$$

For half-integer spin fields, we use the Rarita-Schwinger formalism[7] in which the fields are spinors with tensor indices. Again the tensor indices of the field give an even number of tensor indices to $G_<^{(L)}$; however, we must consider the spinor indices of the function as well. The quantity with positive definite characteristics is found by taking the trace over spinor indices.

The spinor matrices occur as:

$$\beta\,\gamma^\mu\beta\,\tfrac{1}{2}[\gamma^\mu,\gamma^\nu]\beta\,\gamma^5\gamma^\mu\beta\,\gamma^5\beta$$

and have the transformation properties indicated by the tensor indices. The extra factor of β is a consequence of using $\psi\dagger$ rather than $\bar{\psi}$. Then, the trace extracts the $\gamma^0\beta$ term, giving $\mathrm{Tr}G_{\gtrless}^{(L)\lambda\sigma}$ the transformation properties of the time component of a vector as well as the tensor indices $\lambda\sigma$. Since we restrict ourselves to (as far as the tensor indices are concerned) commutators of fields within the same irreducible representation, we find that p_m must be odd. Hence, for $\frac{1}{2}$ integer spin fields,

$$\mathrm{Tr}G_<^{(L)\lambda\sigma}(\xi)=-\mathrm{Tr}G_<^{(L)\lambda\sigma}(-\xi).\quad(18)$$

Now, assume that an integer spin field obeys Fermi statistics. Then using Eqs. (16) and (17)

$$G_>^{(L)\lambda\sigma}(\xi)+G_<^{(L)\lambda\sigma}(-\xi)=0.\quad(19)$$

A half-integer spin field which obeys Bose statistics yields, using Eqs. (16) and (18)

$$\mathrm{Tr}[G_>^{(L)\lambda\sigma}(\xi)+G_<^{(L)\lambda\sigma}(-\xi)]=0.\quad(20)$$

[7] W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).

However, Eqs. (19) and (20) are the values of Lorentz invariant functions of $z=\xi+i\eta$ analytic for $\eta^2<0$, $\eta^0<0$, and which are zero for $\xi^2>0$. These may be continued to be zero[6] for all z, hence for $z=0$ and we have

$$\begin{aligned} G_>^{(L)\lambda\sigma}(0)+G_<^{(L)\lambda\sigma}(0)&=0 \quad \text{integer spin,} \\ \mathrm{Tr}[G_>^{(L)\lambda\sigma}(0)+G_<^{(L)\lambda\sigma}(0)]&=0 \quad \tfrac{1}{2}\text{ integer spin.} \end{aligned} \tag{21}$$

Since the functions are well defined everywhere, we may set $x=x'$ in the expression for $G_>^{(L)}$ and $G_<^{(L)}$ to find that

$$\begin{aligned} \langle\{\psi^\lambda(x),\psi^{\dagger\sigma}(x)\}\rangle&=0 \quad \text{integer spin,} \\ \mathrm{Tr}\langle\{\psi^\lambda(x),\psi^{\dagger\sigma}(x)\}\rangle&=0 \quad \tfrac{1}{2}\text{ integer spin.} \end{aligned}$$

The exponentials drop out since $\zeta(0)=f(x,x,y)=0$.

We have a positive definite metric in Hilbert space and the form $\langle\psi\dagger\psi\rangle$ is a positive definite quantity; thus we may drop the trace and summarize the results of the preceding discussion by the statement that any field which has the "wrong" connection between spin and statistics also satisfies the relation

$$\langle\{\psi^\lambda(x),\psi^{\dagger\lambda}(x)\}\rangle=0,$$

or, since

$$\begin{aligned} \langle\psi^\lambda(x)\psi^{\lambda\dagger}(x)\rangle&\geq 0, \\ \langle\psi^\lambda(x)\psi^{\lambda\dagger}(x)\rangle&=0. \end{aligned} \tag{22}$$

But $\langle\psi^\lambda(x)\psi^{\dagger\lambda}(x')\rangle$ may be written as

$$\int dp\theta(p^0)\sum_n |\langle 0|\psi^\lambda(0)|np\rangle|^2 e^{ip\xi} = \int dp\theta(p^0)\rho(p)e^{ip\xi}\rho(p)\geq 0.$$

Equation (22) implies

$$\int dp\theta(p^0)\rho(p)=0$$

thus $\rho\equiv 0$ and $\langle\psi^\lambda(x)\psi^{\dagger\lambda}(x')\rangle\equiv 0$ or $\psi^{\lambda\dagger}(x)|0\rangle=0$.

But no field which annihilates the vacuum can play any role in the theory, since the theory is determined by its vacuum expectation values.[8] If a vacuum expectation value contains such a field, we use the equations of motion and constraint to express all the fields in terms of independent fields at the same time. Then, if there are no other fields with the wrong connection in the product, we commute the field in question to the vacuum and the expectation value must vanish. If there are such fields, we take the field closest to the vacuum, and commute it to the vacuum with the same result. There is no possiblility of the equations of contraint allowing dependent components of the field to exist, since it is impossible to build up fields with the 'wrong' connection between spin and statistics from fields with the right connection. Hence any field with the wrong connection must vanish identically.

[8] A. S. Wightman, Phys. Rev. **101**, 860 (1956).

ACKNOWLEDGMENTS

It is a pleasure to thank Professor W. Gilbert for many discussions and suggestions, and for his continuing support and encouragement. I would also like to thank Dr. Bruno Zumino for interesting discussions on this paper.

APPENDIX I

The most natural way to derive the results of Sec. II is to use the Action Principle in conjunction with external sources.[9] For this, we use the Lagrangian

$$\mathfrak{L}=\mathfrak{L}_0+\eta(x)\chi(x)+J_\mu(x)A^\mu(x),$$

where $\mathfrak{L}_0$ is the Lagrangian of Eq. (1), and η and J_μ are external sources. J_μ and the sources for Bose fields commute with everything. The sources for Fermi fields anti-commute with each other and their fields, and commute with everything else.

The equations of motion now become:

$$a^\mu[(1/i)\partial_\mu - eqA_\mu]\chi(x)+\frac{\delta_l H(\chi)}{\delta\chi(x)}=\pm\eta(x)$$

$$\partial_\mu F^{\nu\mu}=j^\nu+J^\nu$$

$$F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu.$$

The + or − refers to Bose or Fermi fields respectively.

The equations of contraint and motion for the electromagnetic field may be combined to give:

$$\partial^2 A=-\left(1-\frac{\partial\nabla}{\nabla^2}\right)\left(1-\frac{\nabla\partial}{\nabla^2}\right)(j+J),$$

where the vector index has been suppressed. The Green's functional is

$$G^{(R)}[\eta,J]=\left\langle\left(\exp\left[i\int_{-0}^{\infty}[\eta(x)\chi(x)+J_\mu(x)A^\mu(x)]dx\right]\right)_+\right\rangle,$$

where the fields obey the equations of motion without sources.

Then, the Action principle yields

$$\begin{aligned} \frac{1}{i}\frac{\delta_l}{\delta\eta(x)}G^{(R)}[\eta,J]&=\langle\chi(x)\rangle^{\eta,J} \\ &=\left\langle\left(\chi\exp\left[i\int\eta\chi+JA\right]\right)_+\right\rangle \\ \frac{1}{i}\frac{\delta}{\delta J_\mu(x)}G^{(R)}[\eta,J]&=\langle A^\mu(x)\rangle^{\eta,J}. \end{aligned} \tag{A1}$$

[9] For similar treatments see B. Zumino, J. Math. Phys. **1**, 1 (1960) and J. Schwinger, Phys. Rev. **115**, 721 (1959).

The Green's functions are now obtained by taking an appropriate number of variational derivatives; the one of particular interest to us is

$$G^{(R)}(x,x')=\frac{1}{i}\frac{\delta_l}{\delta\eta(x)}\frac{1}{i}\frac{\delta_l}{\delta\eta(x')}G^{(R)}[\eta,J]=\epsilon\langle[\chi(x)\chi(x')]_+\rangle+\frac{1}{i}\left\langle\frac{\delta\chi(x')}{\delta\eta(x)}\right\rangle.$$

The ϵ comes from commuting variations with the fields. The additive term comes from the fact that the dependent components of the fields have explicit dependences on the source functions. We saw in Sec. II that additional terms were needed to give Lorentz invariance at equal times. This term will provide the necessary terms.

The equations of motion now become Green's functional equations:

$$\left[a^\mu\left(\frac{1}{i}\partial_\mu-eq\frac{1}{i}\frac{\delta}{\delta J^\mu(x)}\right)\frac{1}{i}\frac{\delta_l}{\partial\eta(x)}+\frac{\delta_l H}{\delta\chi(x)}\left(\frac{1}{i}\frac{\delta_l}{\delta\eta}\right)\mp\eta(x)\right]G^{(R)}[\eta,J]=0 \quad \text{(A2a)}$$

$$\partial^2\frac{1}{i}\frac{\delta}{\delta J(x)}G^{(R)}[\eta,J]=-\left(1-\frac{\partial\nabla}{\nabla^2}\right)\left(1-\frac{\nabla\partial}{\nabla^2}\right)\times\left(\frac{e}{\partial}\frac{1}{i}\frac{\delta_l}{\delta\eta(x)}aq\frac{1}{i}\frac{\delta}{\delta\eta(x)}+J\right)G^{(R)}. \quad \text{(A2b)}$$

Now, however, we want to calculate the Lorentz gauge Green's functions in terms of the radiation gauge functions. We must pick a particular Lorentz gauge in order to get explicit expressions. The transverse gauge is perhaps the most natural, so we use that gauge, without loss of generality since we may still make Lorentz invariant gauge transformations.

Wherever $1/i(\delta/\delta J)$ appears, there is an implied $[1-(\partial\nabla/\nabla^2)]1/i(\delta/\delta J)G$ since we are in radiation gauge. To go to Lorentz gauge, this must become $[1-(\partial\partial/\partial^2)]1/i(\delta/\delta J)G$. In Eq. (A2a) this may be achieved by considering $\bar{G}[\eta,J]=G^{(R)}[\eta\ \exp[-eq\partial/\partial^2\times(\delta/\delta J)],J]$ where the $\delta/\delta J$ lies to the left of the J dependence. Then

$$\frac{\delta_l}{\delta\eta}\bar{G}=\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right)\frac{\delta_l}{\delta\left[\eta\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right)\right]}\times G^{(R)}\left[\eta\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right),J\right],$$

and $\bar{G}$ obeys:

$$\left\{a\left[\frac{1}{i}\partial-\left(1-\frac{\partial\partial}{\partial^2}\right)eq\frac{1}{i}\frac{\delta}{\delta J}\right]\frac{1}{i}\frac{\delta}{\delta\eta}+\frac{\delta_l H}{\delta\chi}\mp\eta\right\}\bar{G}=0, \quad \text{(A3)}$$

where we have used $\eta\exp[-eq\partial/\partial^2(\delta/\delta J)]=\exp[eq\times\partial/\partial^2(\delta/\delta J)]\eta$ and the gauge invariance of H. Then Eq. (A3) is Lorentz invariant since

$$\left(1-\frac{\partial\partial}{\partial^2}\right)\left(1-\frac{\partial\nabla}{\nabla^2}\right)=\left(1-\frac{\partial\partial}{\partial^2}\right).$$

Now, we must consider Eq. (A2b). This equation becomes:

$$\partial^2\frac{1}{i}\frac{\delta}{\delta J}\bar{G}[\eta,J]=-\left(1-\frac{\partial\nabla}{\nabla^2}\right)\left(1-\frac{\nabla\partial}{\nabla^2}\right)\left(\frac{e}{\partial}\frac{1}{i}\frac{\delta_l}{\delta\eta}aq\frac{1}{i}\frac{\delta_l}{\delta\eta}\right)\bar{G}-\left(1-\frac{\partial\nabla}{\nabla^2}\right)G^{(R)}\left[\eta\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right),J\right]J,$$

since $\partial_\mu J^\mu=0$. But

$$G^{(R)}\left[\eta\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right),J\right]J=J\bar{G}+e\left(\frac{\partial}{\partial^2}-\frac{\nabla}{\nabla^2}\right)\eta\times\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right)q\frac{\delta_l}{\delta\eta\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right)}G^{(R)},$$

and

$$\partial\left(\frac{e}{\partial}\frac{1}{i}\frac{\delta}{\delta\eta}aq\frac{1}{i}\frac{\delta}{\delta\eta}\right)\bar{G}=-e\eta\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right)q\frac{\delta_l}{\delta\eta\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right)}G^{(R)}.$$

The $\partial/\partial^2-\nabla/\nabla^2$ occurs because

$$\frac{\partial}{\partial^2}\frac{\delta}{\delta J}=\left(\frac{\partial}{\partial^2}-\frac{\nabla}{\nabla^2}\right)\frac{\delta}{\delta J}.$$

Then

$$\partial^2\frac{1}{i}\frac{\delta}{\delta J}\bar{G}=-\left(1-\frac{\partial\nabla}{\nabla^2}\right)\left\{\left(1-\frac{\partial\partial}{\partial^2}\right)\left(\frac{e}{2}\frac{1}{i}\frac{\delta_l}{\delta\eta}aq\frac{1}{i}\frac{\delta_l}{\delta\eta}\right)+J\right\}\bar{G}.$$

Now, if we require that $J=[1-(\partial\partial/\partial^2)]J$ which is just a change in the independent components of J, we have

$$G^{(L)}[\eta,J]=\bar{G}[\eta,J]J=\left(1-\frac{\partial\partial}{\partial^2}\right)J=G^{(R)}\left[\eta\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right),J\right]J=\left(1-\frac{\partial\partial}{\partial^2}\right)J. \quad \text{(A4)}$$

And $G^{(L)}$ obeys:

$$\left[a^\mu\left(\frac{1}{i}\partial_\mu-eq\frac{1}{i}\frac{\delta}{\delta J^\mu}\right)\frac{1}{i}\frac{\delta_l}{\delta\eta}+\frac{\delta_l H}{\delta\chi}\mp\eta\right]G^{(L)}=0$$

$$\partial^2\frac{1}{i}\frac{\delta}{\delta J}G^{(L)}=-\left(1-\frac{\partial\partial}{\partial^2}\right)\left[\frac{e}{2}\frac{1}{i}\frac{\delta}{\delta\eta}aq\frac{1}{i}\frac{\delta}{\delta\eta}+J\right]G^{(L)}.$$

And, assuming the Lorentz invariance of $\delta_l H/\delta\chi$, we have manifestly Lorentz invariant equations. Then, the Green's functions of interest to us are:

$$G^{(L)}(x,x')$$
$$=\frac{1}{i}\frac{\delta_l}{\delta\eta(x)}\frac{1}{i}\frac{\delta_l}{\delta\eta(x')}G^{(L)}[\eta,J]\Big|_{\eta=J=0}$$
$$=\frac{1}{i}\frac{\delta_l}{\delta\eta(x)}\frac{1}{i}\frac{\delta_l}{\delta\eta(x')}G^{(R)}\left[\eta\exp\left(-eq\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right),J\right]\Big|_{\eta=J=0}$$
$$=\exp\left[-e\left(q_x\frac{\partial}{\partial^2}+q_{x'}\frac{\partial}{\partial^2}\frac{\delta}{\delta J}\right)\right]G^{(R)}(x,x',J)\Big|_{J=0},$$

where $G^{(R)}(x,x',J)$ is the radiation gauge Green's function in the presence of an external current J. Since $\nabla\cdot 1/i(\delta/\delta J)G^{(R)}=0$, only the $\delta/\delta J^0$ term survives. Now, in general,

$$\frac{1}{i}\frac{\delta}{\delta J_0(y)}\langle(F(A^0)\cdots)_+\rangle$$
$$=\langle(A^0(y)F(A^0)\cdots)_+\rangle+i\int\frac{dz\delta(y^0-x^0)}{4\pi|\mathbf{y}-\mathbf{z}|}\left\langle\left(\frac{\delta F}{\delta A^0(z)}\right)_+\right\rangle.$$

Then, the Lorentz gauge Green's functions may be written:

$$G^{(L)}(x,x')=\left\langle\left(\exp\left\{-ie\int f(x,x',y)\right.\right.\right.$$
$$\left.\left.\times dy\left[A^0(y)-\frac{i}{\nabla^2}\frac{\delta}{\delta A^0}\right]\right\}\chi(x)\chi(x')\right)_+\Big\rangle_\epsilon$$
$$+\frac{1}{i}\left\langle\left(\exp[\quad]\frac{\delta_l\chi(x')}{\delta\eta(x)}\right)_+\right\rangle,$$

where

$$f(x,x',y)=q\frac{\partial_0}{\partial^2}+q'\frac{\partial_0'}{\partial'^2}=\int\frac{dk}{(2\pi)^4}\frac{ik^0}{k^2}[qe^{ik(x-y)}+q'e^{ik(x'-y)}].$$

In general $[q,p]=i$ implies

$$e^{-i(\alpha q+\beta p)}=e^{-i\alpha q}e^{-i\beta p}e^{i\alpha\beta/2}.$$

Then the Green's function may be written:

$$G^{(L)}(x,x')$$
$$=e^{-\zeta(x-x')}\left[\left\langle\left\{\exp\left(-ie\int fA^0\right)\chi(x)\chi(x')\right\}_+\right\rangle\epsilon(x,x')\right.$$
$$\left.+\left\langle\left\{\exp\left(-ie\int fA^0\right)\frac{1}{i}\frac{\delta_l\chi(x')}{\delta\eta(x)}\right\}_+\right\rangle\right]$$
$$\zeta(\xi)=ie^2\int\frac{dk}{(2\pi)^4}\frac{k^{02}}{(k^2)^2\mathbf{k}^2}[1+qq'e^{ik\xi}]q^2q'^2.$$

APPENDIX II

We have not included the electromagnetic field in our considerations. The assumed form of the Lagrangian implies that the fields obey Bose statistics directly. Alternatively, we may consider the Lorentz gauge Green's functions which may be derived from Eq. (A4) to give

$$\mathcal{G}^{(R)\mu\nu}(x,x')=\langle(A^\mu(x)A^\nu(x'))_+\rangle-i\frac{\delta_0{}^\mu\delta_0{}^\nu\delta(x^0-x'^0)}{4\pi|r-r'|}$$
$$=\left(1-\frac{\partial\nabla}{\nabla^2}\right)^\mu\mathcal{G}^{(L)}(x,x')\left(1-\frac{\nabla\partial}{\nabla^2}\right)^\nu.$$

Then, the vanishing of $\mathcal{G}^{(L)}$ implies the vanishing of $\mathcal{G}^{(R)}$ and the usual arguments hold for the independent components A^k.

Absorption of Light by Light*

HOWARD R. REISS
U. S. Naval Ordnance Laboratory, Silver Spring, Maryland, Department of Physics, University of Maryland, College Park, Maryland
(Received April 21, 1961)

The mutual absorption of two nonparallel classical photon fields through the production of electron pairs is calculated. One of the photon fields is treated accurately in the sense that wave functions are employed which are solutions of the Dirac equation containing the potential of this photon field in the Hamiltonian. The second photon field is introduced as a first-order perturbation to the system containing the first field. An interesting feature is the appearance of an index related to the number of photons in the first field, though this field is not second-quantized. It is shown that the calculation of the absorption process reduces to known results in the limit of small field strength and in the limit of small energy of the photons in the first field. The general results of the theory can be applied to examine the validity of a perturbation expansion.

INTRODUCTION

RELATIVELY little work has been reported on the mutual absorption of two or more electromagnetic fields. Breit and Wheeler[1] calculated the fundamental process of the production of an electron pair by photon-photon collision. Toll and Wheeler[2] investigated the absorption of photons by a constant electromagnetic field. In both of these investigations the calculation was performed in the lowest order of perturbation theory. In the present paper, the mutual absorption of two plane wave fields is considered in such fashion that one of the fields is treated accurately, and the other field is viewed as a perturbation. Hence the Breit-Wheeler limit results when both fields are taken to be weak, and the Toll-Wheeler limit obtains in the limit of vanishing frequency of the first field. The present investigation thus gives the general case which yields these two rather diverse limits; but of perhaps more importance, it gives an accurate result in one of the fields which can then be examined for its analytic properties to determine the convergence of a perturbation expansion in that field strength parameter.[3,4]

CALCULATION

The calculation of the interaction of two oscillatory electromagnetic fields is performed within the framework of Dirac theory, i.e., a non-second-quantized theory. Of the two interacting fields, the first is called the background field; and the second, referred to hereafter as the incoming field, is treated as a perturbation to the system containing the background field. The background field is treated accurately in the sense that the perturbation expansion is in terms of the solutions of the Dirac equation containing this field in the Hamiltonian. A system of units will be employed in which $\hbar=1$, $c=1$.

Let the background field be a monochromatic plane polarized wave given by the vector potential,

$$\mathbf{A}=\hat{e}_1 a\cos\varphi=\hat{e}_1 a\cos[\omega(t-x_3)],$$

where $\hat{e}_1$ is a unit vector in the x_1 direction. A gauge has been chosen such that the scalar potential vanishes. An additive phase factor is of no significance and will be neglected in both the background and incoming fields. The solution of the Dirac equation,

$$i(\partial/\partial t)\psi=[-i\boldsymbol{\alpha}\cdot(\partial/\partial\mathbf{x})+e\boldsymbol{\alpha}\cdot\mathbf{A}+\beta m]\psi \qquad (1)$$

where β and the components of $\boldsymbol{\alpha}$ are the usual anti-commuting Dirac matrices, is[5]

$$\psi=L^{-\frac{3}{2}}[1+\epsilon^2/(\rho E)]^{-\frac{1}{2}}(1-\gamma\cos\varphi)^{\frac{1}{2}}(1-\Gamma)\times\exp[-i(S+Et-\mathbf{p}\cdot\mathbf{x})] \qquad (2)$$

with the notation

$$\epsilon=\tfrac{1}{2}ea$$
$$\rho=E-p_3$$
$$\gamma=(\epsilon/\rho)\alpha_1(1+\alpha_3)$$
$$\Gamma=E^{-1}(\boldsymbol{\alpha}\cdot\mathbf{p}+\beta m)$$
$$S=(2p_1\epsilon/\omega\rho)\sin\varphi+(\epsilon^2/2\omega\rho)(2\varphi+\sin 2\varphi).$$

L is the length of the side of the cube within which the normalization is carried out, and E and the components of $\mathbf{p}$ are constants which satisfy the relation $E^2=\mathbf{p}^2+m^2$. L is restricted to be an integer multiple of $2\pi/\omega$, so that ψ satisfies periodic boundary conditions at the edges of the normalization volume. It should be noted that Eq. (2) is a square matrix of rank four which contains all four of the single column matrix solutions of the Dirac equation.

The wave function as given by Eq. (2) is an eigenfunction of the operators $-i(\partial/\partial x_1)$ and $-i(\partial/\partial x_2)$ with eigenvalues p_1 and p_2, respectively. Since both operators commute with the Hamiltonian of Eq. (1), they are constants of the motion, and p_1, p_2 may be

* Based on portions of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Maryland (1958).

[1] G. Breit and J. A. Wheeler, Phys. Rev. **46**, 1087 (1934).

[2] J. S. Toll and J. A. Wheeler (to be published); J. S. Toll, Ph.D. Dissertation, Princeton University (1952, unpublished).

[3] H. R. Reiss, Thesis, University of Maryland (1958); NAVORD Report 6180, U. S. Naval Ordnance Laboratory (1958).

[4] H. R. Reiss, J. Math. Phys. (to be published).

[5] D. M. Volkov, Z. Physik **94**, 250 (1935); N. D. Sengupta, Bull. Calcutta Math. Soc. **39**, 147 (1947); A. H. Taub, Ann. Math. **40**, 937 (1939), Revs. Modern Phys. **21**, 388 (1949).

interpreted as momentum components in the x_1, x_2 directions. Another operator which commutes with the Hamiltonian is $i(\partial/\partial t+\partial/\partial x_3)$, which has the eigenvalue $\rho=E-p_3$. E and p_3 by themselves have no special significance except in the limit $a\rightarrow 0$, where E would be the energy and p_3 the x_3 component of momentum. It is convenient to introduce a box size index given by $L=2\pi n/\omega$, and then to replace the set of values p_1, p_2, ρ by the set of independent quantum numbers n_1, n_2, n_3. These quantum numbers are introduced by the requirement of periodicity of ψ in the cube of side L, which can be accomplished by

$$p_1=2\pi n_1/L=n_1\omega/n;\quad p_2=2\pi n_2/L=n_2\omega/n.$$

These last relations give

$$\rho=\pm[(n_1^2+n_2^2+n_3^2)\omega^2/n^2+m^2+2\epsilon^2]^{\frac{1}{2}}-n_3\omega/n, \quad (3)$$

where the plus sign corresponds to positive energies and the minus sign to negative energies. The analogy of Eq. (3) to the field free case is evident, for if $\epsilon=0$ then Eq. (3) gives

$$\rho=\pm|E|-p_3.$$

The orthonormality of the wave function, Eq. (2), can be expressed as

$$\int\psi^\dagger\psi' d^3x=\tfrac{1}{2}(1-\Gamma)\delta_{n_1n_1'}\delta_{n_2n_2'}\delta_{n_3n_3'}\delta_\pm,$$

where the $\delta_\pm$ is included to indicate that not only must n_1, n_2, n_3 be the same for ψ and ψ', but ρ and ρ' must both have the same choice of the ambiguous sign in Eq. (3). The matrix $\frac{1}{2}(1-\Gamma)$ is idempotent and has a trace equal to two.

The system containing the background field is to be perturbed by another plane electromagnetic wave of frequency $\tilde\omega$. If this new field propagates in the same positive x_3 direction as the background field, there can be no interaction between them. If the incoming field is not propagating parallel to the background field, then it is always possible to find a Lorentz transformation to a system in which the incoming field is anti-parallel to the background field. The two special cases in which the incoming field is plane polarized parallel or perpendicular to the background field may be superposed to obtain any arbitrary relative polarization of the fields. For polarization of the incoming field perpendicular to the background field, the perturbing Hamiltonian is

$$H_\perp=e\tilde a\alpha_2\cos\tilde\varphi=\tilde\epsilon\alpha_2[\exp(i\tilde\varphi)+\exp(-i\tilde\varphi)],$$

where $\tilde\epsilon=\frac{1}{2}e\tilde a$ and $\tilde\varphi=\tilde\omega(t+x_3)$. For parallel polarization

$$H_{||}=\tilde\epsilon\alpha_1[\exp(i\tilde\varphi)+\exp(-i\tilde\varphi)].$$

Consider the complete set of solutions ψ_n to Eq. (1), where the ψ_n are given by Eq. (2). Then the solution of

$$(H+H_\perp)\psi=i\partial\psi/\partial t$$

can be expanded in a series of the matrix functions ψ_n as

$$\psi=\textstyle\sum_n\psi_nA_n(t), \quad (4)$$

where the A_n are time dependent matrix expansion coefficients. The A_n are given by[3]

$$A_m(t)=-i\int_0^t(\psi_m,H_\perp\psi_i)dt'+\delta_{nm}A_n(0), \quad (5)$$

with the initial condition

$$A_n(0)=\tfrac{1}{2}(1-\Gamma_n)\delta_{ni}.$$

For a final state ψ_f, the quantity $(\psi_f,H_\perp\psi_i)$ is given by

$$(\psi_f,H_\perp\psi_i)=\frac{\tilde\epsilon}{L}\left(\frac{1-\Gamma_f}{2}\right)\frac{\alpha_2\exp[i(\rho_f-\rho_i-2\tilde\omega)t]}{[1+\epsilon^2/(\rho_fE_f)]^{\frac{1}{2}}[1+\epsilon^2/(\rho_iE_i)]^{\frac{1}{2}}}$$
$$\times\int_{-\frac{1}{2}L}^{\frac{1}{2}L}dx_3\left[1+\epsilon\alpha_1(1+\alpha_3)\left(\frac{1}{\rho_f}-\frac{1}{\rho_i}\right)\cos\varphi\right]$$
$$\times\exp[i(A\sin\varphi+B\sin2\varphi+C\varphi)]\left(\frac{1-\Gamma_i}{2}\right), \quad (6)$$

where

$$\begin{aligned}A&=(1/\rho_f-1/\rho_i)2\epsilon p_1/\omega\\ B&=(1/\rho_f-1/\rho_i)\epsilon^2/2\omega\\ C&=(1/\rho_f-1/\rho_i)\epsilon^2/\omega+(p_{3f}-p_{3i})/\omega+\tilde\omega/\omega\\ &=(n_{3f}-n_{3i})/n+\tilde\omega/\omega.\end{aligned} \quad (7)$$

The integral over x_3 becomes independent of t if $C=k/n$ for some integer k. The implication of this is that $\tilde\omega L=2\pi j$ for some integer j (to be compared with $\omega L=2\pi n$), so that $\tilde\omega$ must satisfy the same periodicity conditions as ω. If the x_3 integral is transformed to the variable φ, and the condition $C=k/n$ is imposed, it then becomes convenient to make a further transformation to another variable θ, where $\theta=\varphi-2l\pi$. The integral then takes on the form

$$\sum_{l=0}^{n-1}e^{i2\pi kl/n}\int_0^{2\pi}d\theta\exp[i(A\sin\theta+B\sin2\theta+k\theta/n)].$$

The summation index l is entirely outside the integral, and the summation is readily performed to give

$$\sum_{l=0}^{n-1}e^{i2\pi kl/n}=n\delta_{k,nq},$$

where q is an integer. Now if $(\psi_f,H_\perp\psi_i)$ is written as

$$(\psi_f,H_\perp\psi_i)=H_{\perp fi}\exp[i(\rho_f-\rho_i-2\tilde\omega)t]$$

then

$$H_{\perp fi}=(\tilde\epsilon/2\pi)\tfrac{1}{2}(1-\Gamma_f)\alpha_2[1+\epsilon^2/(\rho_fE_f)]^{-\frac{1}{2}}[1+\epsilon^2/(\rho_iE_i)]^{-\frac{1}{2}}$$
$$\times[I_0+\epsilon\alpha_1(1+\alpha_3)(1/\rho_f-1/\rho_i)I_1]\tfrac{1}{2}(1-\Gamma_i),$$

$$I_n\equiv\int_0^{2\pi}d\theta(\cos n\theta)\exp[i(A\sin\theta+B\sin2\theta+q\theta],$$

with A and B given by Eq. (7).

The total probability per unit time per unit volume for creating a pair is

$$W_\perp=\frac{1}{L^3}\lim_{t\to\infty}\frac{1}{t}\sum_{(n_1,n_2,n_3)f(n_1,n_2,n_3)i}\mathrm{Tr}(A_f^\dagger A_f).$$

It has already been established that the matrix element will vanish unless

$$n_{1f}=n_{1i}$$
$$n_{2f}=n_{2i}$$
$$n_{3f}=n_{3i}+nq-j,$$

so the subscripts f and i on n_1 and n_2 may be dropped, and

$$W_\perp=\frac{1}{L^3}\lim_{t\to\infty}\frac{1}{t}\sum_{n_1,\,n_2,\,n_{3i},\,q}\mathrm{Tr}(A_f^\dagger A_f). \tag{8}$$

From Eq. (5),

$$A_f(t)=-iH_{\perp fi}\int_0^t\exp[i(\rho_f-\rho_i-2\tilde\omega)t']dt',$$

so that

$$A_f^\dagger A_f=2H_{\perp fi}^\dagger H_{\perp fi}\times\{1-\cos[(\rho_f-\rho_i-2\tilde\omega)t]\}(\rho_f-\rho_i-2\tilde\omega)^{-2}.$$

This expression is to be substituted into Eq. (8), where the limit $t\to\infty$ is taken. Hence the expression for $A_f^\dagger A_f$ contains a representation of the Dirac delta function. If the normalization volume L^3 is now taken to expand to ∞, so the summations over n_1, n_2, n_{3i} are replaced by integrations, Eq. (8) then appears as

$$W_\perp=L^{-3}\sum_q\int dn_1\int dn_2\int dn_{3i}(2\pi/\tilde\omega)\times\mathrm{Tr}(H_{\perp fi}^\dagger H_{\perp fi})\delta[(\rho_f-\rho_i)/\tilde\omega-2]. \tag{9}$$

The n_{3i} integral in Eq. (9) is conveniently accomplished by a transformation to the variable

$$\bar\rho=(\rho_f-\rho_i)/\tilde\omega-2.$$

Since $\bar\rho$ is a double-valued function of n_{3i}, the n_{3i} integral is first split into two parts corresponding to values of n_{3i} less than or greater than $n_{3i}(\bar\rho_{\min})$, and distinguished by superscripts as

$$n_{3i}^-\lessgtr n_{3i}(\bar\rho_{\min})\lessgtr n_{3i}^+.$$

Now $W_\perp$ becomes

$$W_\perp=L^{-3}\sum_q\int dn_1\int dn_2(2\pi/\tilde\omega)\times\{[(\partial n_{3i}/\partial\bar\rho)\,\mathrm{Tr}(H_{\perp fi}^\dagger H_{\perp fi})]_{n_{3i}^+}-[(\partial n_{3i}/\partial\bar\rho)\,\mathrm{Tr}(H_{\perp fi}^\dagger H_{\perp fi})]_{n_{3i}^-}\} \tag{10}$$

where the condition $\bar\rho=0$ is understood.

The calculation of the trace is lengthy and gives the final result

$$\mathrm{Tr}(H_{\perp fi}^\dagger H_{\perp fi})=(2\tilde\epsilon^2\tilde\omega^2/\pi^2)(\rho_f^2+\zeta)^{-1}(\rho_i^2+\zeta)^{-1}\times[(\zeta-2\epsilon^2-p_2^2\zeta/q\omega\tilde\omega)I_0^2+4\epsilon p_1I_0I_1+4\epsilon^2I_1^2],$$

where

$$\zeta=p_1^2+p_2^2+m^2+2\epsilon^2$$

and

$$\rho_f=\tilde\omega[1\pm(1-\zeta/q\omega\tilde\omega)^{\frac12}],$$
$$\rho_i=\tilde\omega[-1\pm(1-\zeta/q\omega\tilde\omega)^{\frac12}]. \tag{11}$$

In Eq. (11) the upper sign is associated with n_{3i}^- and the lower sign is associated with n_{3i}^+. With these results substituted into Eq. (10), $W_\perp$ now takes the form

$$W_\perp=L^{-2}\sum_q\int dn_1\int dn_2(\tilde\epsilon^2/2\pi^2)(1-\zeta/q\omega\tilde\omega)^{-\frac12}\times[(1-2\epsilon^2/\zeta-p_2^2/q\omega\tilde\omega)I_0^2+(4\epsilon p_1/\zeta)I_0I_1+(4\epsilon^2/\zeta)I_1^2]. \tag{12}$$

The coefficients A and B [Eq. (7)] in the integrals, I_0, I_1, become proportional to q when $\bar\rho=0$, so I_0 and I_1 may be written in the combined form

$$I_n=\int_0^{2\pi}d\theta(\cos n\theta)\exp[iq(b_1\sin\theta+\tfrac12 b_2\sin2\theta+\theta)], \tag{13}$$

where

$$b_1=4\epsilon p_1/\zeta,\quad b_2=2\epsilon^2/\zeta. \tag{14}$$

If a new integral I_2 is introduced by means of Eq. (13), an integration by parts of I_0 yields

$$I_0+b_1I_1+b_2I_2=0.$$

Then for the combination which occurs in Eq. (12),

$$(1-b_2)I_0^2+b_1I_0I_1+2b_2I_1^2=b_2(2I_1^2-I_0^2-I_0I_2)=-b_2\int_0^{2\pi}d\theta\int_0^{2\pi}d\theta'(\cos\theta-\cos\theta')^2\times\exp\{iq[b_1(\sin\theta+\sin\theta')+\tfrac12 b_2(\sin2\theta+\sin2\theta')+\theta+\theta']\}.$$

Define a new function I_{11}^2 by

$$I_{11}^2=2I_1^2-I_0^2-I_0I_2. \tag{15}$$

This notation is chosen to show the analogy of this function to I_0^2, I_0I_1 and I_1^2; but it should be remarked that I_{11}^2 is not a simple product of integrals in the θ, θ' variables.

A transformation of the integration variables n_1, n_2

to p_1, ζ yields the final result for $W_\perp$,

$$W_\perp=\frac{\tilde{\epsilon}^2}{8\pi^4}\sum_{q=[(m^2+2\epsilon^2)/\omega\tilde{\omega}]}^{\infty}\int_{m^2+2\epsilon^2}^{q\omega\tilde{\omega}}\frac{d\zeta}{(1-\zeta/q\omega\tilde{\omega})^{\frac{1}{2}}}\times\int_{-(\zeta-m^2-2\epsilon^2)^{\frac{1}{2}}}^{(\zeta-m^2-2\epsilon^2)^{\frac{1}{2}}}\frac{dp_1}{(\zeta-m^2-2\epsilon^2-p_1^2)^{\frac{1}{2}}}\times\left[\frac{2\epsilon^2}{\zeta}I_{11}^2-\frac{(\zeta-m^2-2\epsilon^2-p_1^2)}{q\omega\tilde{\omega}}I_0^2\right]. \quad (16)$$

The lower limit of the q sum

$$q_m=[(m^2+2\epsilon^2)/\omega\tilde{\omega}] \quad (17)$$

is defined to mean the smallest integer containing $(m^2+2\epsilon^2)/\omega\tilde{\omega}$.

The calculation of W_{11} is similar to the $W_\perp$ case, and yields the result

$$W_{11}=\frac{\tilde{\epsilon}^2}{8\pi^4}\sum_{q=q_m}^{\infty}\int_{m^2+2\epsilon^2}^{q\omega\tilde{\omega}}\frac{d\zeta}{(1-\zeta/q\omega\tilde{\omega})^{\frac{1}{2}}}\times\int_{-(\zeta-m^2-2\epsilon^2)^{\frac{1}{2}}}^{(\zeta-m^2-2\epsilon^2)^{\frac{1}{2}}}\frac{dp_1}{(\zeta-m^2-2\epsilon^2-p_1^2)^{\frac{1}{2}}}\times\left[\left(1-\frac{\zeta}{q\omega\tilde{\omega}}\right)\frac{2\epsilon^2}{\zeta}I_{11}^2+\frac{(\zeta-p_1^2-2\epsilon^2)}{q\omega\tilde{\omega}}I_0^2\right]. \quad (18)$$

The above calculation contains some interesting qualitative features. If one considers the creation of an electron pair by the collision of a photon of energy $\tilde{\omega}$ with q' photons of energy ω moving in the opposite direction, then energy conservation requires

$$q'\omega+\tilde{\omega}\geq 2m.$$

In the center-of-mass frame of the electron pair, momentum conservation demands

$$q'\omega=\tilde{\omega}.$$

These two conditions can be combined to give

$$q'\omega\tilde{\omega}\geq m^2. \quad (19)$$

Since the scalar product of the propagation four-vectors of the two plane waves,

$$k_\mu\tilde{k}^\mu=2\omega\tilde{\omega},$$

is a relativistic invariant, then condition (19) holds in any relativistic frame of reference. The present theory involves a summation over an index q which starts with the value defined by Eq. (17). Hence

$$q\omega\tilde{\omega}\geq m^2+2\epsilon^2. \quad (20)$$

In the limit of small field strength q thus has the significance of q', the number of particles from the background field contributing to the production of the pair. In this limit then, the sum over q in Eqs. (16) and (18) can be interpreted as a sum over the number of photons participating, although the photon field is not quantized in the present investigation. Hence this interpretation cannot be maintained strictly except in the weak field limit.

The distinction between the threshold conditions, Eqs. (19) and (20), lies in the fact that Eq. (19) presumes the electron pair to be created with no energy above the rest energy, while Eq. (20) evidently requires that the electron pair must have an interaction energy with the background field as soon as it is created. This interpretation is further supported by the fact that neither E nor p_3 of the electron is a constant of the motion in the present theory, but only $E-p_3$. This is suggestive of a continual interchange of momentum between the photon field (propagating in the x_3 direction) and the x_3 momentum component of the electron in such fashion that $E-p_3$ is preserved.

The threshold condition, Eq. (20), can be obtained in another way. No real electron pairs can be produced unless the argument of the delta function in Eq. (9) has at least one zero. It can be seen that $\bar{\rho}\to\infty$ for $n_{3f}\to\pm\infty$, and that $\bar{\rho}$ has exactly one minimum located at

$$n_{3f}=-\tfrac{1}{2}(j-nq).$$

The requirement, $\bar{\rho}_{\min}\leqslant 0$, then leads to

$$q\geqslant\zeta/\omega\tilde{\omega}$$

or

$$q\omega\tilde{\omega}\geqslant p_1^2+p_2^2+m^2+2\epsilon^2,$$

which is an elaboration of Eq. (20) when the members of the electron pair are permitted to have momenta in the x_1, x_2 directions. The analog of Eq. (19) results from considering q' photons from the background field combining with one photon from the incoming field under the threshold condition that $p_3=0$. The result is just

$$q\omega\tilde{\omega}\geqslant p_1^2+p_2^2+m^2.$$

WEAK-FIELD LIMIT

The Eqs. (16) and (18) apply for background field strength $a\omega=2\epsilon\omega/e$ which is arbitrarily large, apart from the restriction that the absorption which occurs is sufficiently small that attenuation of the field may be neglected. Hence the two photon pair production process first calculated by Breit and Wheeler should be obtained in the limit of small background field strength.

In the limit as $\epsilon\to 0$, consider a fixed value of q and let ϵ be sufficiently small that $qb_1\ll 1$. Then

$$I_0\approx-\pi b_1,\quad I_1\approx\pi$$

for $q=1$, and both functions are zero to lowest order in ϵ for other positive values of q. Also, then

$$I_{11}^2\approx 2\pi^2.$$

Since the sum over q reduces to the single term $q=1$,

$$W_\perp \approx \frac{\bar{\epsilon}^2}{8\pi^4}\int_{m^2}^{\omega\tilde{\omega}} \frac{d\zeta}{\zeta(1-\zeta/\omega\tilde{\omega})^{\frac{1}{2}}} \int_{-(\zeta-m^2)^{\frac{1}{2}}}^{(\zeta-m^2)^{\frac{1}{2}}} \frac{dp_1}{(\zeta-m^2-p_1^2)^{\frac{1}{2}}} \times\left[2\pi^2\frac{2\epsilon^2}{\zeta}-\frac{(\zeta-m^2-p_1^2)}{\omega\tilde{\omega}}\pi^2\frac{16\epsilon^2 p_1^2}{\zeta^2}\right].$$

The integrations are straightforward, and yield the final result

$$W_\perp \approx (\epsilon^2\bar{\epsilon}^2/2\pi)\{-(1+m^2/2\omega\tilde{\omega})(1-m^2/\omega\tilde{\omega})^{\frac{1}{2}} +2[1+(m^2/\omega\tilde{\omega})-(m^4/4\omega^2\tilde{\omega}^2)]\tanh^{-1}(1-m^2/\omega\tilde{\omega})^{\frac{1}{2}}\}.$$

As should be anticipated, this expression is symmetrical in background and incoming field quantities. To convert this into terms of reaction cross section σ,

$$\sigma_\perp = W_\perp/N_\omega N_{\tilde{\omega}},$$

where N_ω and $N_{\tilde{\omega}}$ are the densities of background and incoming photons. The energy density of the field is $a^2\omega^2/(8\pi)$, and the energy of a single photon is ω, so

$$N_\omega = a^2\omega/(8\pi) = \epsilon^2\omega/(2e^2\pi).$$

Hence

$$\sigma_\perp = (2\pi e^2/\omega\tilde{\omega})\{-(1+m^2/2\omega\tilde{\omega})(1-m^2/\omega\tilde{\omega})^{\frac{1}{2}} +2[1+(m^2/\omega\tilde{\omega})-(m^2/2\omega\tilde{\omega})^2]\tanh^{-1}(1-m^2/\omega\tilde{\omega})^{\frac{1}{2}}\},$$

which agrees exactly with the Breit-Wheeler result.[1] For the case of parallel polarization it is found that

$$\sigma_{\parallel} = (2\pi e^2/\omega\tilde{\omega})\{-(1+3m^2/2\omega\tilde{\omega})(1-m^2/\omega\tilde{\omega})^{\frac{1}{2}} +2[1+(m^2/\omega\tilde{\omega})-3(m^2/2\omega\tilde{\omega})^2\tanh^{-1}(1-m^2/\omega\tilde{\omega})^{\frac{1}{2}}\},$$

which also agrees with Breit and Wheeler.

LOW-FREQUENCY LIMIT

In the limit as the frequency of the background field approaches zero, the results for general photon energy given in Eqs. (16) and (18) should reduce (in a particular sense to be described) to the constant field case calculated by Toll and Wheeler.[2]

The Toll-Wheeler results are expressed in terms of a parameter χ, where χ is defined by

$$\chi = (\tilde{\omega}/m)(B/B_{\text{crit}}). \tag{21}$$

B is the field strength and B_{crit}, called the critical field, is a unit of field strength given by $B_{\text{crit}}=m^2/e$. The primary validity of the Toll-Wheeler work is in the small χ limit, where their limiting result converted into a transition probability per unit time per unit volume is

$$W_\perp = (m^2\bar{\epsilon}^2/4\pi)(\tfrac{3}{2})^{\frac{1}{2}}\chi e^{-4/3\chi} \tag{22}$$

and

$$W_{\parallel} = \tfrac{1}{2}W_\perp.$$

To compare $W_\perp$ for an oscillatory field of very low frequency with $W_\perp$ for a constant field, treat the oscillatory field as if it were a constant field over any small portion of a wavelength, and describe it by Eq. (22) with χ given by the local value. Then to obtain an average over a wavelength, set $\chi=\chi_0\cos\psi$. Hence

$$(\chi e^{-4/3\chi})_{\text{av}} = (2/\pi)\int_0^{\pi/2} d\psi\chi_0\cos\psi e^{-4/(3\chi_0\cos\psi)}.$$

With the change of variables

$$u = (4/3\chi_0)(1+1/\cos\psi)$$

and the assumption that $\chi_0 \ll 1$, the result is obtained that

$$(\chi e^{-4/3\chi})_{\text{av}} = (3\chi_0^3/2\pi)^{\frac{1}{2}}e^{-4/3\chi_0}.$$

So from Eq. (22), if the result,

$$W_\perp = \tfrac{3}{8}m^2\bar{\epsilon}^2(\chi_0/\pi)^{\frac{1}{2}}e^{-4/3\chi_0}, \tag{23}$$

is achieved, it can be concluded that the proper constant field limit obtains.

To examine Eqs. (16) and (18) in the constant field limit, certain detailed properties of the I_0^2 and I_{11}^2 functions must be established. Appendix A contains a demonstration that I_0^2 and I_{11}^2 are both even functions of the parameter b_1 which occurs in the exponent. Then, since Eqs. (16) and (18) have integrands which are even in p_1, and the even property of I_0^2 and I_{11}^2 in b_1 makes them even in p_1, the symmetrical region of integration in p_1 may be halved and the result from the reduced region doubled. Appendix B contains the results of integration of I_0^2 and I_{11}^2 in steepest descent approximation based upon the assumption of very large values of q. From Eq. (17) it is seen that for sufficiently small ω the minimum value of q becomes large, so that as $\omega \to 0$ the results of Appendix B are appropriate. The very large values that q_m attains as $\omega \to 0$ means that the discrete property of q_m as expressed by Eq. (17) loses its significance, and q_m may be adequately expressed by Eq. (17) without the brackets. Also, q may be viewed as a continuous variable and the sum over q replaced by an integral in the constant field limit. Qualitatively, the situation is now one in which the energy of each background photon is so small that no pairs can be produced unless a great many background photons participate. When more than the threshold number of photons are involved, it becomes relatively unimportant whether q photons or $q+1$ photons participate, so the discrete quality of the process is lost.

The calculation of I_0^2 and I_{11}^2 in Appendix B is most conveniently performed in terms of variables ξ, η in place of the ζ, p_1 variables of (16) and (18). The transformations are

$$\zeta = 2\epsilon^2(1+\eta), \quad p_1 = \tfrac{1}{2}\epsilon[(2+\eta)^2-\xi^2]^{\frac{1}{2}}, \quad \xi = (2\epsilon^2)^{-1}[(\zeta+2\epsilon^2)^2-(4\epsilon p_1)^2]^{\frac{1}{2}}, \quad \eta = (2\epsilon^2)^{-1}(\zeta-2\epsilon^2), \tag{24}$$

with the Jacobian

$$\frac{\partial(\zeta,p_1)}{\partial(\eta,\xi)} = -\frac{\xi\epsilon^3}{[(2+\eta)^2-\xi^2]^{\frac{1}{2}}}.$$

With the notation

$$z=2\epsilon^2/m^2, \quad w=\omega\bar{\omega}/m^2,$$

Eq. (16) becomes

$$W_\perp=\frac{m^2z\bar{\epsilon}^2}{4\pi^4}\int_{(1+z)/w}^{\infty}dq\int_{1/z}^{(qw/z)-1}\frac{d\eta}{[1-(\eta+1)z/qw]^{\frac{1}{2}}}\times\int_{[(2-\eta)^2+8/z]^{\frac{1}{2}}}^{2+\eta}\frac{\xi d\xi}{[(2+\eta)^2-\xi^2]^{\frac{1}{2}}[\xi^2-(2-\eta)^2-8/z]^{\frac{1}{2}}}\times\left\{\frac{1}{\eta+1}I_{11}^2-\frac{z}{8qw}\left[\xi^2-(2-\eta)^2-\frac{8}{z}\right]I_0^2\right\}. \quad (25)$$

From Eq. (B7)

$$(\eta+1)^{-1}I_{11}^2=(\sqrt{2}\pi/q)[\xi^2-(2-\eta)^2]^{\frac{1}{2}}\xi^{-\frac{1}{2}}\exp(q\mathcal{E}),$$

where $\mathcal{E}$ is a function of ξ and η defined by Eq. (B4). I_0^2 contains terms of the type in I_{11}^2, and some additional terms which are trigonometric with arguments proportional to q [see Eq. (B6)]. The trigonometric terms will be neglected as compared to the nontrigonometric one, and this action will be justified later. Then I_0^2 reduces to

$$I_0^2=(4\sqrt{2}\pi/q)(\eta+1)[\xi^2-(2-\eta)^2]^{-\frac{1}{2}}\xi^{-\frac{1}{2}}\exp(q\mathcal{E}).$$

It is convenient to perform the q integral first, so after the interchange of integrals Eq. (25) is

$$W_\perp\approx\frac{m^2z\bar{\epsilon}^2}{2\sqrt{2}\pi^3}\int_{1/z}^{\infty}d\eta\times\int_{[(2-\eta)^2+8/z]^{\frac{1}{2}}}^{2+\eta}d\xi\frac{\xi^{\frac{1}{2}}[\xi^2-(2-\eta)^2]^{\frac{1}{2}}}{[(2+\eta)^2-\xi^2]^{\frac{1}{2}}[\xi^2-(2-\eta)^2-8/z]^{\frac{1}{2}}}\times\int_{(\eta+1)z/w}^{\infty}dq\frac{\exp(q\mathcal{E})}{q^{\frac{1}{2}}[q-(\eta+1)z/w]^{\frac{1}{2}}}\times\left\{1-\frac{(\eta+1)z}{2qw}\left[\frac{\xi^2-(2-\eta)^2-8/z}{\xi^2-(2-\eta)^2}\right]\right\}. \quad (26)$$

The function $\mathcal{E}$, Eq. (B4), is independent of q, and it will be shown below that it is negative throughout the entire region of integration in ξ and η. Since the lower limit on the q integral goes as $1/w$, this means that as $w\to 0$, $\exp(q\mathcal{E})$ will provide strong exponential damping as q increases from its minimum value. Thus the approximations

$$\int_{(\eta+1)z/w}^{\infty}dq\frac{\exp(q\mathcal{E})}{q^{\frac{1}{2}}[q-(\eta+1)z/w]^{\frac{1}{2}}}\approx\frac{\exp[(\eta+1)z\mathcal{E}/w]}{[(\eta+1)z/w]^{\frac{1}{2}}}\int_{(\eta+1)z/w}^{\infty}dq\frac{\exp\{\mathcal{E}[q-(\eta+1)z/w]\}}{[q-(\eta+1)z/w]^{\frac{1}{2}}}\approx\pi^{\frac{1}{2}}\frac{\exp[(\eta+1)z\mathcal{E}/w]}{[-(\eta+1)z\mathcal{E}/w]^{\frac{1}{2}}}$$

and

$$\int_{(\eta+1)z/w}^{\infty}dq\frac{\exp(q\mathcal{E})}{q^{\frac{1}{2}}[q-(\eta+1)z/w]^{\frac{1}{2}}}\approx\frac{\pi^{\frac{1}{2}}}{(\eta+1)z/w}\frac{\exp[(\eta+1)z\mathcal{E}/w]}{[-(\eta+1)z\mathcal{E}/w]^{\frac{1}{2}}}$$

are valid. Equation (26) now takes the form

$$W_\perp\approx\frac{m^2z\bar{\epsilon}^2}{2(2\pi^5)^{\frac{1}{2}}}\int_{1/z}^{\infty}d\eta\times\int_{[(2-\eta)^2+8/z]^{\frac{1}{2}}}^{2+\eta}d\xi\frac{\xi^{\frac{1}{2}}[\xi^2-(2-\eta)^2]^{\frac{1}{2}}}{[(2+\eta)^2-\xi^2]^{\frac{1}{2}}[\xi^2-(2-\eta)^2-8/z]^{\frac{1}{2}}}\times\frac{\exp[(\eta+1)z\mathcal{E}/w]}{[-(\eta+1)z\mathcal{E}/w]^{\frac{1}{2}}}\left\{1-\frac{1}{2}\left[\frac{\xi^2-(2-\eta)^2-8/z}{\xi^2-(2-\eta)^2}\right]\right\}. \quad (27)$$

Equation (27) is difficult to calculate because of the unwieldy limits on the integrals. A more convenient region of integration results from a transformation to the variables r, s, where

$$\xi=r+1/z+2(s+1/z)/(r+1/z)$$
$$\eta=r+1/z+2(r-s)/(r+1/z)$$
$$r=\tfrac{1}{2}(\xi+\eta-2-2/z)$$
$$s=\tfrac{1}{8}[\xi^2-(2-\eta)^2-8/z],$$

with the Jacobian

$$\frac{\partial(\eta,\xi)}{\partial(r,s)}=\frac{4}{r+1/z}.$$

The region of integration in r and s is the wedge shaped area $0\leqslant s\leqslant r$, $0\leqslant r\leqslant\infty$. The function $\mathcal{E}$ in r, s variables is

$$\mathcal{E}=\frac{(r+1/z)+(r-s)}{(r+1/z)(1+r+1/z)+2(r-s)}\times\left[\left(1+r+\frac{1}{z}\right)^2-1\right]^{\frac{1}{2}}-\cosh^{-1}\left(1+r+\frac{1}{z}\right). \quad (28)$$

Equation (27) contains $\mathcal{E}$ in the form $(1+\eta)z\mathcal{E}/w$. The largest value of $(1+\eta)\mathcal{E}$ in the entire r, s region of integration occurs at the origin in r and s. This fact, along with the simple shape of the region of integration, is the basis for the convenience of the r, s variables. When $r=0$, $s=0$, Eq. (28) gives

$$[(1+\eta)\mathcal{E}]_{\max}=[(1+1/z)^2-1]^{\frac{1}{2}}-(1+1/z)\cosh^{-1}(1+1/z). \quad (29)$$

Equation (29) attains its largest values for the smallest values of $1/z$. This is entirely reasonable because the square of the background field strength is proportional to $z\omega^2$, and as $\omega\to 0$ the field strength (and $W_\perp$) would

vanish unless $z\to\infty$ or $1/z\to 0$. In the limit as $1/z\to 0$,

$$[(1+\eta)\mathcal{E}]_{\max}\to -\tfrac{1}{3}(2/z)^{\frac{1}{2}}.$$

This result means that the solution has strong exponential damping. In particular, since the argument of the exponential function in Eq. (27) is $(1+\eta)z\mathcal{E}/w$, the largest value this can assume is given by

$$(1+\eta)z\mathcal{E}/w\lesssim -4/(3\chi_0),$$

where by Eq. (20)

$$\chi_0=w(2z)^{\frac{1}{2}}.$$

The zero subscript refers to the fact that the amplitude of B has been used in Eq. (21).

It is to be expected that only a limited portion of the region of integration near the origin in r and s will be significant because of the exponential decay of the integrand in Eq. (27) away from the r, s origin. In the r, s variables, Eq. (27) is

$$W_\perp\approx\frac{m^2z\bar{\epsilon}^2}{2\pi^{\frac{5}{2}}}\int_0^\infty\frac{dr}{(r+1/z)(2+r+1/z)^{\frac{1}{2}}}\times\int_0^r ds\frac{(s+1/z)^{\frac{1}{2}}[(r+1/z)^2+2(s+1/z)]^{\frac{1}{2}}}{s^{\frac{1}{2}}(r-s)^{\frac{1}{2}}[-(1+\eta)z\mathcal{E}/w]^{\frac{1}{2}}}\times\exp\left[(1+\eta)\frac{z}{w}\mathcal{E}\right]\left(1-\frac{1}{2}\frac{s}{s+1/z}\right),\quad(30)$$

where the notation $(1+\eta)\mathcal{E}$ is retained for simplicity. $(1+\eta)\mathcal{E}$ can be approximated by

$$(1+\eta)\mathcal{E}\approx[(1+\eta)\mathcal{E}]_{\max}+r[(\partial/\partial r)(1+\eta)\mathcal{E}]_0+s[(\partial/\partial s)(1+\eta)\mathcal{E}]_0,\quad(31)$$

where

$$[(\partial/\partial r)(1+\eta)\mathcal{E}]_0=(1+2z)\{(2/z)[(1+1/z)^2-1]^{-\frac{1}{2}}-\cosh^{-1}(1+1/z)\}\quad(32)$$

$$[(\partial/\partial s)(1+\eta)\mathcal{E}]_0=-2z\{[(1+1/z)^2-1]^{\frac{1}{2}}-\cosh^{-1}(1+1/z)\}.\quad(33)$$

Both of these expressions are negative, as noted earlier. The largest values of r and s which will be significant in Eq. (30) are those for which the product of $r\,z/w$ with Eq. (32) and $s\,z/w$ with Eq. (33) are of the order of unity. These largest significant values of r and s can then be bounded by finding the smallest magnitudes that Eqs. (32) and (33) can assume. The largest value (or smallest magnitude) of Eq. (32) occurs as $1/z\to 0$, whence

$$[(\partial/\partial r)(1+\eta)\mathcal{E}]_0\to -\tfrac{1}{3}(2/z)^{\frac{1}{2}}.\quad(34)$$

Thus it is to be required that

$$\tfrac{1}{3}(2z)^{\frac{1}{2}}r/w<c=O(1).$$

When $r=1/z$ the left-hand side of this inequality is $2/(3\chi)$, which is very much larger than order unity. Hence only $r\ll 1/z$ will be significant in Eq. (30). Similarly, the smallest magnitude of Eq. (33) occurs as $1/z\to 0$, or when

$$[(\partial/\partial s)(1+\eta)\mathcal{E}]_0\to -\tfrac{2}{3}(2/z)^{\frac{1}{2}}.\quad(35)$$

The condition

$$\tfrac{2}{3}(2z)^{\frac{1}{2}}s/w<c=O(1)$$

then implies that only $s\ll 1/z$ need be considered in Eq. (30).

Since it has already been established that only the limit $1/z\to 0$ can contribute to $W_\perp$, it is valid to use Eqs. (34) and (35) in the expression Eq. (31). When Eq. (31) is substituted into Eq. (30) and the inequalities $r\ll 1/z$, $s\ll 1/z$ are employed, the result is

$$W_\perp\approx\frac{m^2\bar{\epsilon}^2}{2\pi^{\frac{5}{2}}}\left(\frac{3\chi_0}{4}\right)^{\frac{1}{2}}e^{-4/(3\chi_0)}\int_0^\infty dre^{-2zr/(3\chi_0)}\times\int_0^r ds\frac{e^{-4zs/(3\chi_0)}}{s^{\frac{1}{2}}(r-s)^{\frac{1}{2}}}.$$

Integration over s yields

$$\int_0^r ds s^{-\frac{1}{2}}(r-s)^{-\frac{1}{2}}e^{-4zs/(3\chi_0)}=\pi e^{-2zr/(3\chi_0)}I_0(-2zr/3\chi_0),$$

where the I_0 function which appears here is the modified Bessel function of the first kind of zero order, and is not to be confused with the I_0 of Eq. (13). The r integral is

$$\int_0^\infty dre^{-4zr/(3\chi_0)}I_0(-2zr/3\chi_0)=\sqrt{3}\chi_0/(2z).$$

The final result is then

$$W_\perp\approx\tfrac{3}{8}m^2\bar{\epsilon}^2(\chi_0/\pi)^{\frac{1}{2}}e^{-4/(3\chi_0)}$$

which is seen to correspond to Eq. (23).

For the case of parallel polarization,

$$W_{11}\approx\tfrac{3}{16}m^2\bar{\epsilon}^2(\chi_0/\pi)^{\frac{1}{2}}e^{-4/(3\chi_0)},$$

or $W_{11}=\frac{1}{2}W_\perp$ in the constant field limit, which is in agreement with Toll and Wheeler.

To complete the investigation of this limiting case, only one point remains to be cleared up, and that is that the trigonometric terms in I_0^2 were neglected. In the final parenthesis in Eq. (30), containing

$$1-\tfrac{1}{2}s/(s+1/z)$$

the first term is retained, but the other term dropped when the integrand is evaluated near $s=0$. However, the first term arises from I_{11}^2 and the second term from the nontrigonometric part of I_0^2, so it need only be shown that the trigonometric terms in I_0^2 are of the same order or less than the nonoscillatory term. When $r=0$, $s=0$, the function T, Eq. (B5), reduces to π. Hence the $\sin qT$ term vanishes and its coefficient also happens to vanish. The ratio of the coefficient of the $\cos qT$ term to the nontrigonometric term becomes unity

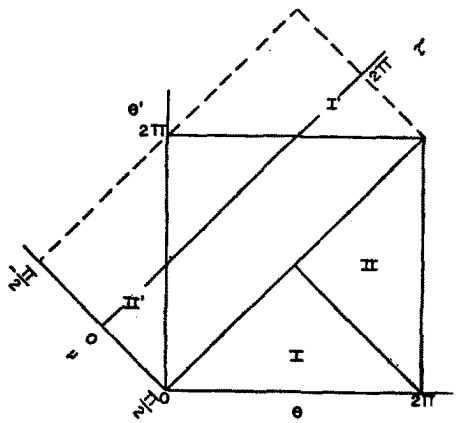


FIG. 1. Modification of Integration Region in θ, θ'. The triangular regions I and II are, respectively, equivalent to the regions I' and II' by periodicity of the integrand in θ and θ'.

when $r=0$, $s=0$, so it is justifiable to neglect the trigonometric terms in I_0^2.

APPENDIX A. SYMMETRY OF I_0^2 AND I_{11}^2 FUNCTIONS

It is desired to show that I_0^2, obtained from the square of Eq. (13), and I_{11}^2 given by Eqs. (15) and (13), are invariant with respect to a change in sign of the parameter b_1. I_0^2 is

$$I_0^2=\int_0^{2\pi}d\theta\int_0^{2\pi}d\theta'\exp\{iq[b_1(\sin\theta+\sin\theta')+\tfrac{1}{2}b_2(\sin2\theta+\sin2\theta')+\theta+\theta']\}. \quad \text{(A1)}$$

A change from θ, θ' to the variables τ, μ is to be performed, where

$$\tau=\tfrac{1}{2}(\theta+\theta'),\quad \mu=\tfrac{1}{2}(\theta-\theta'+\pi),\qquad \theta=\tau+\mu-\tfrac{1}{2}\pi,\quad \theta'=\tau-\mu+\tfrac{1}{2}\pi. \quad \text{(A2)}$$

To accomplish the transformation into a convenient region in τ, μ space, an appeal may be made to the periodicity in 2π of the integrand of I_0^2 with respect to both θ and θ'. As illustrated in Fig. 1, the region of integration in θ and θ' may be so rearranged that the transformation (A2) yields a τ, μ region of integration in which $0\leqslant\tau\leqslant2\pi$ and $-\frac{1}{2}\pi\leqslant\mu\leqslant\frac{1}{2}\pi$. With the Jacobian

$$\frac{\partial(\theta,\theta')}{\partial(\tau,\mu)}=-2,$$

then (A1) becomes

$$I_0^2=-2\int_0^{2\pi}d\tau e^{2iq\tau}\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi}d\mu\times\exp[iq(2b_1\sin\tau\sin\mu-b_2\sin2\tau\cos2\mu)]. \quad \text{(A3)}$$

Define the function $H(b_1)$ to be the μ integral in (A3). Then

$$H(b_1)-H(-b_1)=2i\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi}d\mu\times\exp(-iqb_2\sin2\tau\cos2\mu)\sin(2qb_1\sin\tau\sin\mu),$$

which vanishes because the integrand is odd in μ. Hence $H(b_1)=H(-b_1)$, and

$$I_0^2(b_1)=I_0^2(-b_1).$$

The I_{11}^2 function differs from (A1) in having a factor $(\cos\theta-\cos\theta')^2$ in the integrand. However, $(\cos\theta-\cos\theta')^2=4\sin^2\tau\cos^2\mu$, which does not affect the parity of the integrand, so

$$I_{11}^2(b_1)=I_{11}^2(-b_1).$$

APPENDIX B. I_0 AND I_1 BY THE METHOD OF STEEPEST DESCENTS

With the hypothesis that q is a very large parameter, the function I_0

$$I_0=\int_0^{2\pi}d\theta e^{qf(\theta)}$$

is to be calculated by the method of steepest descents. The function $f(\theta)$ is

$$f(\theta)=i(b_1\sin\theta+\tfrac{1}{2}b_2\sin2\theta+\theta),$$

where b_1 and b_2 are defined by Eq. (14). The condition $f'(\theta)=0$ yields

$$\cos\theta_{sp}=-b_1/4b_2\pm[b_1^2-8b_2(1-b_2)]^{\frac{1}{2}}/4b_2.$$

With the aid of Eq. (14), the argument of the square root is found to be

$$b_1^2-8b_2(1-b_2)=(16\epsilon^2/\zeta^2)(-\zeta+p_1^2+2\epsilon^2)$$

which is always negative by the definition of ζ and since $2\epsilon^2$ is real and positive. Thus, saddle points are located at

$$\cos\theta_{sp}=-b_1/4b_2\pm i[8b_2(1-b_2)-b_1^2]^{\frac{1}{2}}/4b_2, \quad \text{(B1)}$$

where the only imaginary quantity is that which is explicitly indicated. Since b_1 is non-negative and b_2 is positive, all the saddle points are in the second and third quadrants.

Application of the relations

$$\cos\theta_{rsp}\cosh\theta_{isp}=-b_1/4b_2$$
$$\sin\theta_{rsp}\sinh\theta_{isp}=\mp[8b_2(1-b_2)-b_1^2]^{\frac{1}{2}}/4b_2,$$

which follow from (B1), leads to

$$\cos\theta_{rsp}=-\tfrac{1}{2}(-\xi+\eta+2)^{\frac{1}{2}}$$
$$\cosh\theta_{isp}=\tfrac{1}{2}(\xi+\eta+2)^{\frac{1}{2}}$$
$$\sin\theta_{rsp}=\pm\tfrac{1}{2}(\xi-\eta+2)^{\frac{1}{2}}$$
$$\sinh\theta_{isp}=\pm\tfrac{1}{2}(\xi+\eta-2)^{\frac{1}{2}}$$

where ξ, η are the variables defined by Eq. (24). The ambiguous sign in $\sin\theta_{rsp}$ refers to second and third quadrant saddle points, and in $\sinh\theta_{isp}$ the ambiguous sign represents saddle points in upper and lower half-planes.

The path of integration is deformed to follow the paths of steepest descent from the saddle points above the real axis. The saddle point values required for the integration are

$$\Re f(\theta_{sp})=(\eta-\xi+4)[(\xi+\eta)^2-4]^{\frac{1}{2}}/[8(\eta+1)] -\tfrac{1}{2}\cosh^{-1}[\tfrac{1}{2}(\xi+\eta)] \quad \text{(B2)}$$
$$\Re f''(\theta_{sp})=-(\xi-\eta+2)[(\xi+\eta)^2-4]^{\frac{1}{2}}/[4(\eta+1)].$$

The imaginary parts of $f(\theta_{sp})$ and $f''(\theta_{sp})$ differ for the two saddle points, with

$$\Im f(\theta_{sp})=\pm\{(\xi+\eta+4)[4-(\xi-\eta)^2]^{\frac{1}{2}}/[8(\eta+1)] -\tfrac{1}{2}\cos^{-1}[\tfrac{1}{2}(\eta-\xi)]\}+\pi \quad \text{(B3)}$$
$$\Im f''(\theta_{sp})=\pm(\xi+\eta-2)[4-(\xi-\eta)^2]^{\frac{1}{2}}/[4(\eta+1)],$$

where in both cases the upper sign obtains in the second quadrant and the lower sign in the third quadrant. It is understood that the positive branch of the arccosh function is to be taken in Eq. (B2), and the arccos function in Eq. (B3) refers to the interval 0, π. The result of the integration of I_0 by steepest descents is

$$I_0=\int d\theta e^{qf(\theta)}\approx\left(\frac{2\pi}{q}\right)^{\frac{1}{2}}\sum_{\text{saddle pts.}}\frac{e^{qf(\theta_{sp})}}{[-f''(\theta_{sp})]^{\frac{1}{2}}}$$
$$\approx(-1)^q(2\pi/q)^{\frac{1}{2}}(\eta+1)^{\frac{1}{2}}\xi^{-\frac{1}{2}}[\xi^2-(2-\eta)^2]^{-\frac{1}{4}} \times\exp(\tfrac{1}{2}q\mathcal{E})\{\{(8\xi)^{\frac{1}{2}}+[(2+\xi)^2-\eta^2]^{\frac{1}{2}}\}^{\frac{1}{2}}\cos(\tfrac{1}{2}qT) -\{(8\xi)^{\frac{1}{2}}-[(2+\xi)^2-\eta^2]^{\frac{1}{2}}\}^{\frac{1}{2}}\sin(\tfrac{1}{2}qT)\},$$

with

$$\mathcal{E}=(\eta-\xi+4)[(\xi+\eta)^2-4]^{\frac{1}{2}}/[4(\eta+1)] -\cosh^{-1}[\tfrac{1}{2}(\xi+\eta)] \quad \text{(B4)}$$

and

$$T=(\xi+\eta+4)[4-(\xi-\eta)^2]^{\frac{1}{2}}/[4(\eta+1)] -\cos^{-1}[\tfrac{1}{2}(\eta-\xi)]. \quad \text{(B5)}$$

The calculation of I_1 proceeds in the same fashion as I_0 to yield the result

$$I_1\approx(-1)^{q+1}\tfrac{1}{4}(2\pi/q)^{\frac{1}{2}}(\eta+1)^{\frac{1}{2}}\xi^{-\frac{1}{2}}[\xi^2-(2-\eta)^2]^{-\frac{1}{4}} \times\exp(\tfrac{1}{2}q\mathcal{E})\{[\{(8\xi)^{\frac{1}{2}}+[(2+\xi)^2-\eta^2]^{\frac{1}{2}}\}^{\frac{1}{2}}[(2+\eta)^2-\xi^2]^{\frac{1}{2}} -\{(8\xi)^{\frac{1}{2}}-[(2+\xi)^2-\eta^2]^{\frac{1}{2}}\}^{\frac{1}{2}}[\xi^2-(2-\eta)^2]^{\frac{1}{2}}]\cos(\tfrac{1}{2}qT) -[\{(8\xi)^{\frac{1}{2}}-[(2+\xi)^2+\eta^2]^{\frac{1}{2}}\}^{\frac{1}{2}}[(2+\eta)^2-\xi^2]^{\frac{1}{2}} +\{(8\xi)^{\frac{1}{2}}+[(2+\xi)^2-\eta^2]^{\frac{1}{2}}\}^{\frac{1}{2}}[\xi^2-(2-\eta)^2]^{\frac{1}{2}}] \times\sin(\tfrac{1}{2}qT)\}.$$

The quadratic combinations of I_0 and I_1 are

$$I_0^2\approx(2\pi/q\xi)(\eta+1)[\xi^2-(2-\eta)^2]^{-\frac{1}{2}} \times\exp(q\mathcal{E})\{(8\xi)^{\frac{1}{2}}+[(2+\xi)^2-\eta^2]^{\frac{1}{2}}\cos(qT) -[\eta^2-(2-\xi)^2]^{\frac{1}{2}}\sin(qT)\}$$
$$I_0I_1\approx-(\pi/2q\xi)(\eta+1)[\xi^2-(2-\eta)^2]^{-\frac{1}{2}} \times\exp(q\mathcal{E})\{(8\xi)^{\frac{1}{2}}[(2+\eta)^2-\xi^2]^{\frac{1}{2}} +4[4-(\xi-\eta)^2]^{\frac{1}{2}}\cos(qT) -4[(\xi+\eta)^2-4]^{\frac{1}{2}}\sin(qT)\} \quad \text{(B6)}$$
$$I_1^2\approx(\pi/q\xi)(\eta+1)[\xi^2-(2-\eta)^2]^{-\frac{1}{2}} \times\exp(q\mathcal{E})\{\eta(8\xi)^{\frac{1}{2}}+(2-\xi) \times[(2+\xi)^2-\eta^2]^{\frac{1}{2}}\cos(qT) -(2+\xi)[\eta^2-(2-\xi)^2]^{\frac{1}{2}}\sin(qT)\}$$

with the particular quadratic form

$$b_2I_{11}^2=(1-b_2)I_0^2+b_1I_0I_1+2b_2I_1^2 =(\pi/q)(2/\xi)^{\frac{1}{2}}[\xi^2-(2-\eta)^2]^{\frac{1}{2}}\exp(q\mathcal{E}). \quad \text{(B7)}$$

The function $b_2I_{11}^2$ thus contains no trigonometric terms.

ACKNOWLEDGMENT

The author wishes to express his gratitude to Professor John S. Toll, who suggested the problem, and who directed the thesis on which this paper is based.

Causality and the R Matrix*

M. E. EBEL†
University of Wisconsin, Madison, Wisconsin
(Received August 29, 1961)

A proposed counterexample to the theorem that causality and unitarity alone will guarantee that the R matrix has no singularities off the real energy axis is considered and shown to come from an acausal interaction. The nature of the acausality in this model and in some trivial generalizations is also discussed.

A SUGGESTION has been made recently[1] that the requirements of causality and unitarity alone may not be sufficient to guarantee that the only singularities of the derivative (R) matrix lie on the real energy axis. The case in point is that of the elastic scattering of spinless uncharged particles with an R matrix given by

$$R(q)=sn(v)/[vcn(v)dn(v)], \qquad v\equiv qa, \tag{1}$$

where q is the wave number of the scattered particle and a the radius of interaction. The functions sn, cn, dn are the usual Jacobian elliptic functions, which reduce to sin, cos, and 1 as the modulus goes to zero. This may be thought of as arising from a nonlinear interaction

$$V(r)=\lambda r^2u^2(r), \quad r<a, \tag{2}$$

or an energy-dependent interaction

$$V(r,E)=k^2q^2[cn^2(qr)-sn^2(qr)], \quad r<a. \tag{3}$$

In Eq. (3), k is the modulus of the elliptic functions.

Since the R matrix may be defined independently of the specific nature of the interaction responsible for the scattering, the usual theorems[1–3] that any R matrix which satisfies the requirements of unitarity and causality should have no singularities off the real axis should also apply to nonlinear and energy-dependent interactions such as (2) and (3). Since (1) obviously has poles off the real axis unless $k=0$, if (2) and (3) are in fact causal, one would have a contradiction to these theorems. It is the purpose of this note to point out that the difficulty seems to arise from the use by Power and Saavedra[1] of an insufficiently strict definition of causality, and that the supposed counter-example in fact represents an acausal interaction for $k\neq 0$.

The proposed causality condition is that used by van Kampen,[2] which states that for a suitably normalized wave packet the probability of finding a particle outside the scattering region must be less than or equal to unity at all times. This requires that

$$-iP\int_0^\infty dk\int_0^\infty dk'\left[\frac{S_a(k)S_a^*(k')-1}{k-k'}-\frac{S_a(k)-S_a^*(k')}{k+k'}\right]A_a(k)A_a^*(k')\geq 0, \tag{4}$$

for *all* wave packets described by a square integrable momentum superposition $A_a(k)$. $S_a(k)$ is $\exp(2ika)$ times the usual scattering matrix. If this condition is expressed in terms of $R(E)$ instead of $S_a(k)$, it immediately gives the causality conditions used by Wigner and von Neumann[3,4]

$$\det|W_{ij}|\geq 0$$
$$W_{ij}\equiv(E_i-E_j)^{-1}[R(E_i)-R(E_j)], \quad i\neq j$$
$$\equiv dR(E_i)/dE_i, \quad i=j, \tag{5}$$

for all real E_i where R is not singular. This condition must be satisfied for all n, n being the rank of the determinant in (5).

The causality condition used by Power and Saavedra was obtained from (4) by considering an infinitely narrow wave packet. As such, it corresponds to the condition $dR/dE\geq 0$, which is satisfied by (1). It is easily verified, however, that (5) is not satisfied for $n=2$ (take $v_1=\frac{1}{2}K$, $v_2=\frac{1}{2}K+\epsilon$). Thus one must conclude that the interaction proposed is in fact not causal, and that the appearance of complex singularities of R should not be surprising. The importance of considering condition (5) for all n may be easily understood. In order that the causality condition used be able to place some restrictions on the interaction, it is necessary to introduce wave packets which are localized in space and time. This means that the packets must contain a superposition of many energies, corresponding to condition (5) for large n.

It is of interest to examine the nature of the acausality implied by the interaction (3). Since the interaction depends upon the wave number *outside*, it is not of the usual velocity-dependent type; and it is useful to consider it instead as a function of the energy E. If it is now used in the time-dependent Schrödinger equation, energy eigenfunctions $u_E(r)$ may be determined. It

* Supported in part by the U. S. Atomic Energy Commission.
† Alfred P. Sloan Foundation Fellow.
[1] E. A. Power and I. Saavedra, Proc. Cambridge Phil. Soc. **57**, 121 (1961).
[2] N. G. van Kampen, Phys. Rev. **91**, 1267 (1953).
[3] E. P. Wigner, Am. J. Phys. **23**, 371 (1955).
[4] E. P. Wigner and J. v. Neumann, Ann. of Math. **59**, 418 (1954).

is readily verified, however, that in general

$$\int \{u_{E'}^*(r)V(r,E)u_E(r) - [V(r,E')u_{E'}(r)]^* u_E(r)\}dr \neq 0, \quad (6)$$

so that the interaction, although real, is not Hermitian. This difficulty may be removed by employing instead the time-dependent Schrödinger equation. Then q^2 must be replaced by $(2im/\hbar)\partial/\partial t$, and the resultant equation is of infinite order in the time derivatives. This implies, however, that the interaction must be nonlocal in time. In fact, one may formally write

$$V\psi(r,t) = (2\pi\hbar)^{-1}\int_{-\infty}^{\infty} d\tau\left[\int_0^{\infty} dE V(r,E)\exp(i\tau E/\hbar)\right] \times \psi(r,t+\tau), \quad (7)$$

since ψ contains only positive energies. It is then clear that the nonlocality of (3) is in fact infinite in extent, which suggests strongly that it will be acausal and that it may not be removed by a redefinition of the nuclear radius. The acausality of (3) is thus connected with the existence of poles of V near the real axis in the first quadrant of the complex E plane, for almost all $r<a$.

A slight generalization of the interaction (3) provides an interesting illustration of these considerations. Suppose that the energy-dependent potential

$$V(r,E) = E - (\hbar^2 f^2/2m)[1+k^2 cn^2(fr) - k^2 sn^2(fr)], \quad r<a, \quad (8)$$

is introduced, with f some real function of the energy. The solution to the Schrödinger equation is

$$u = c\, sn(fr), \quad r<a \quad (9)$$

which gives for the R matrix,

$$R = sn(fa)/[(fa)cn(fa)dn(fa)]. \quad (10)$$

If now $f(E)$ is chosen, for example, to be proportional to some power of E, then R has poles off the real axis if and only if $k\neq 0$. This is just the condition that $V(r,E)$ have poles near the real axis and thus that it represent a nonlocal interaction. The case where f is proportional to $E^{\frac{1}{2}}$ is of course that considered previously. An even simpler example occurs if $f\propto E$; the nonlocality in (7) is large for infinitely large values of τ, if $k\neq 0$. As $k\to 0$, the complex poles in both R and V go off to $i\infty$, and the nonlocality vanishes. Finally, if $f(E)$ is taken as the elliptic integral $F(k,E)$, $sn(f)$ and $cn(f)$ become $\sin(E)$ and $\cos(E)$. Then there are no poles off the real axis for R and no poles in the finite plane for V.

For a precise mathematical statement of these causality and acausality regions, one must of course follow some procedure similar to that of van Kampen or Wigner. However, these simple examples may help in giving a qualitative picture of the nature of these conditions.

ACKNOWLEDGMENTS

The author is indebted to Professor H. W. Lewis and Professor E. P. Wigner for stimulating conversations on this subject.

Some Physical Solutions of Dirac-Type Equations

JACK K. HALE AND ARNOLD P. STOKES*
RIAS, 7212 Bellona Avenue, Baltimore 12, Maryland
(Received, June 12, 1961)

The Dirac equation of classical electrodynamics is a third-order differential equation. The purpose of the present paper is to give some sufficient conditions on the force field which will insure that there are solutions of Dirac's equation which approach a motion with constant velocity as time increases.

CONSIDER the equation

$$d^2z/dt^2=F(z,dz/dt,t)+\tau_0[d^3z/dt^3+f(d^2z/dt^2)z] \quad (1)$$

where z, F are n vectors, $\tau_0>0$ is a real parameter, and $f(u)$ is a Lipschitzian scalar function satisfying

$$|f(u)\leqq B\|u\|^2 \quad \text{for all} \quad u, \quad (2)$$

where $\|u\|$, u an n vector, designates any vector norm.

Equation (1) is the Dirac equation of classical electrodynamics, where $F(z,\dot{z},t)$ is the force field acting on the particle with position z, and velocity $\dot{z}$ at time t. (The dot notation indicates derivatives with respect to time.) We are interested in obtaining conditions on the field F (assumed continuous in t and Lipschitzian in z, $\dot{z}$) which, under certain boundary conditions on z, $\dot{z}$, and $\ddot{z}$, imply that "physically realizable" solutions exist. Here "physically realizable" is taken to mean that along a solution $z(t)$, the acceleration $\ddot{z}(t)\to 0$ as $t\to\infty$, and $\int_0^\infty\|\ddot{z}(u)\|du$ exists. Thus $\lim_{t\to\infty}\|\dot{z}(t)\|$ and $\lim_{t\to\infty}\|z(t)\|/t$ exist, and are finite. This is by no means the most general notion of "physically realizable" solutions,[1] or even a sufficiently general notion, as, for instance, bounded or oscillatory motions are excluded. However this notion of "physically realizable" has an immediate physical interpretation; namely, as $t\to\infty$, the solution approaches a "steady-state" motion of zero acceleration and constant velocity. Motions of this last type exist when $F\equiv 0$, as is readily seen. So, for those fields F which are "small" for $\|z\|$ large and t large, it is reasonable to suppose such "physically realizable" solutions exist.

The first theorem below is more of a description of a method of proof than a theorem, but it is of some interest in itself because it relates directly to a technique used by several physicists in discussing nonlinear equations.[2] Briefly, the device employed by the physicists is to replace some nonlinear terms by a function of time alone, reasoning that along a solution, this is the form that the nonlinear term has. The resulting equation is then discussed, and the results at least provide some guide to the intuition. Theorem 1 gives conditions under which this procedure furnishes a valid description of the nonlinear equation itself. Several lemmas show that Theorem 1 is applicable to the Dirac equation (1), and Theorems 2–4 then apply this method for different fields, each of which is of some physical interest.

Some preliminary definitions are in order. Let R^n be n-dimensional Euclidean space, and let $C_T{}^n$ be the space of all continuous functions $x(t)$ from $[T,\infty)\to R^n$, $T\geq 0$. The topology on $C_T{}^n$ is the compact open topology,[3] i.e., a sub-base of neighborhoods in $C_T{}^n$ is the collection of subsets of $C_T{}^n$ of the form $M(L,W)=\{f\epsilon C_T{}^n | f(L)\subset W\}$ where L is a compact subset of $[T,\infty)$ and W is an open set in R^n. Since R^n is a metric space, convergence in the compact open topology is equivalent to uniform convergence on compact subsets of $[T,\infty)$. Also, since R^n is a metric space and $[T,\infty)$ is the sum of a countable family of compact subsets having the property that any other compact subset of $[T,\infty)$ is covered by some finite subfamily, the space $C_T{}^n$ is a metric space and, as R^n is complete, $C_T{}^n$ is complete with respect to its metric. Furthermore, $C_T{}^n$ is a linear topological space, and to define the topology it is sufficient to give a system of neighborhoods of the identity. Now, consider the family of pseudonorms

$$\{p_n(x)\}, \quad p_n(x)=\sup_{T\leqq t\leqq T+n}\|x(t)\|,$$

where $x\epsilon C_T{}^n$ and $\|y\|$ for $y\epsilon R^n$ is any vector norm. It can be shown that the family of sets $\{\lambda V_n\}$, $\lambda\geqq 0$, $n=1, 2, \cdots$, where

$$V_n=\{x\epsilon C_T{}^n | p_n(x)\leqq 1\}, \quad (3)$$

considered as a subbase for a system of neighborhoods of $x=0$, generates a topology on $C_T{}^n$ which is equivalent to uniform convergence on compact subsets of $[T,\infty)$. Finally, the sets V_n are convex, and, summarizing the above, we have $C_T{}^n$ is a complete, locally convex, linear, topological space.[4] A set A in $C_T{}^n$ is bounded if the image of A under each pseudo-norm, $p_n(x)$, is bounded.

In $C_T{}^n$, the following fixed-point theorem, due to Tychonov[5] is valid.

* This research was partially supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract. Part of this work was completed while the second author was a National Science Foundation Fellow.

[1] See G. N. Plass, Revs. Modern Phys. 33, 37 (1961).

[2] See reference 1 and F. Rohrlich, Ann. Phys., 13, 93 (1961).

[3] J. L. Kelley, *General Topology* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1955).

[4] N. Bourbaki, *Espaces vectoriels topologiques* (Hermann & Cie, Paris, 1953), livre 5.

[5] A. Tychonov, Math. Ann. 111, 767–776 (1935). See also, J. Schauder, Studia Math. 2, 171–180 (1930).

Theorem: Let C be a complete locally convex topological vector space, and let $A \subset C$ be closed and convex. If $J: A \to C$ is continuous on A, $J(A) \subset A$, and $\mathrm{Cl}(J(A))$ is compact, then there exists $a_0 \epsilon A$ such that $J(a_0) = a_0$. For $A \subset C$, the symbol $\mathrm{Cl}(A)$ denotes the closure of A in the topology of C.

With these preliminaries, we can state:

Theorem 1. Let S and H be two subsets of $C_T{}^n$, and let J and F be two operators such that

(a) S is closed and convex;

(b) J is a continuous mapping of $S \times H \to S$; and $\mathrm{Cl}(J(S,H))$ is compact;

(c) F is a continuous mapping of $S \to H$.

Then there exists $(s_0,h_0) \epsilon S \times H$ such that

$$F(s_0) = h_0, \quad J(s_0,h_0) = s_0.$$

Proof. Define the mapping R on S by $R(s) = J[s,F(s)]$. By (b) and (c), R is defined and continuous on S, and $R(S) \subset \mathrm{Cl}(J(S,H))$. Thus $\mathrm{Cl}(R(S))$ is also compact. Then, using (a), it is clear that R and S satisfy the hypotheses of Tychonov's theorem, and so there exists $s_0 \epsilon S$ such that $R(s_0) = s_0$. Set $h_0 = F(s_0)$, and then $J(s_0,h_0) = s_0$, from the definition of R. Q. E. D.

To appreciate the relationship of this theorem to the technique referred to above, consider S as a set of functions in $C_T{}^n$ defined in such a manner that the statement "There exists a solution in S" implies that this solution has all the desired properties, whatever they may be. The operator $J(s,h)$ is an integral operator derived from the given equation, with h introduced in place of certain troublesome nonlinear terms, [e.g., $F(z,\dot{z},t)$ in 1)].

Now, as indicated, S is determined by the requirements placed on the desired solution, and J is given by the problem. H is then chosen (if possible) in such a manner that the inclusion $J(S,H) \subset S$ holds. The continuity of J is usually trivial, as is the compactness of $\mathrm{Cl}(J(S,H))$, if S, H, and J satisfy certain weak conditions.

The mapping F defined on S is given by the nonlinear terms of the equation which were replaced by h; e.g., if z, $\dot{z} \epsilon S$, then $F[z(t), \dot{z}(t), t]$, or more properly, $F(z,\dot{z},t)$ is an element of $C_T{}^n$. The continuity of F is straightforward, if the function F is continuous. Thus, the major hypothesis to be satisfied reduces to the inequality $F(S) \subset H$. This gives a condition on the type of nonlinear terms which may be introduced, in order to obtain a solution belonging to S.

The procedure outlined above seems to the authors to describe rigorously the approach used in an intuitive manner in the work of Plass and Rohrlich. In the remainder of the paper, this method will be applied to Eq. (1), and specific fields will be described for which (1) possesses "physically realizable" solutions.

Let

$$A(g) = \{a \epsilon C_T{}^n \mid \|a(t)\| \leqq g(t)\}, \tag{4}$$

where $g \epsilon C_T{}^1$, $g \geqq 0$ for $t \geqq T$, and $g(t) \to 0$ as $t \to \infty$, $\int_T{}^\infty g(t)dt < \infty$. A is clearly closed and convex. Choosing z_0, $v_0 \epsilon R^n$, let

$$v(t) = v_0 + \int_T^t a(u)du$$

and

$$z(t) = z_0 + v_0(t-T) + \int_T^t \int_T^u a(\tau)d\tau du,$$

for $a \epsilon A(g)$. Clearly v, $z \epsilon C_T{}^n$, for $a \epsilon A(g)$; $v(T) = v_0$, $z(T) = z_0$, and both limits $\lim_{t\to\infty} v(t)$, $\lim_{t\to\infty} z(t)/t$, exist, and are equal. Thus, if a solution z is such that its acceleration $\ddot{z} \epsilon A(g)$, then z is a "physically realizable" solution as described above.

Remark. There is no loss of generality in obtaining solutions in $C_T{}^n$ for $T > 0$, for it is reasonably easy to show that any solution of (1) can be continued back from T to 0, for any continuous F, and any T.

Define $H(\eta)$ by

$$H(\eta) = \{h \epsilon C_T{}^n \mid \|h(t)\| \leqq \eta(t)\}, \tag{5}$$

where $\eta \epsilon C_T{}^1$, $\eta(t) \geqq 0$ for $t \geqq T$, $0 \leqq \eta(t) \leqq N$, $t \geqq T$, for some constant N, and $\int_T{}^\infty \eta(t)dt < \infty$.

Now in (1), replace $F(z,\dot{z},t)$ by $h(t)$ and for $v_0 \epsilon R^n$, define J on $A(g) \times H(\eta)$ by

$$J(a,h)(t) = \int_t^\infty e^{(t-u)/\tau_0} \left[f[a(u)]v(u) + \frac{1}{\tau_0} h(u) \right] du, \tag{6}$$

where $v(t) = v_0 + \int_T{}^t a(u)du$, as before.

Lemma 1. Let $A(g)$, $H(\eta)$, J be defined as in (4)–(6). Then J is continuous on $A(g) \times H(\eta)$ and

$$\mathrm{Cl}(J[A(g),H(\eta)])$$

is compact (in the compact-open topology on $C_T{}^n$).

Proof. Given $\epsilon > 0$ and $m \geqq 1$, m an integer, let $V_{m,\epsilon}$ be defined by (3). We wish to show that there exists a $\delta > 0$, and $k \geqq 1$, k an integer, such that (a_1,h_1), $(a_2,h_2) \epsilon A(g) \times H(\eta)$, $(a_1 - a_2)$, $(h_1 - h_2) \epsilon V_{k,\delta}$ implies

$$[J(a_1,h_1) - J(a_2,h_2)] \epsilon V_{m,\epsilon}.$$

Now

$$\begin{aligned} &\|J(a_1,h_1)(t) - J(a_2,h_2)(t)\| \\ &\leqq \int_t^\infty e^{(t-u)/\tau_0} \Big| f[a_1(u)]v_1(u) - f[a_2(u)]v_2(u) \\ &\qquad + \frac{1}{\tau_0}[h_1(u) - h_2(u)] \Big| du \\ &\leqq \int_T^{T+k} e^{(T+k-u)/\tau_0} [c](u)du \\ &\qquad + \int_{T+k}^\infty e^{(t-u)/\tau_0} [c](u)du, \quad \text{for} \quad T \leqq t \leqq T+k, \end{aligned}$$

where $[c](u)$ denotes the integrand above, evaluated at

u. Now $\|v_0\| \leqq M$, $\int_T^\infty g \leqq G^*$ for some M, G^*. Then, for $u \geqq k$,

$$[c](u) \leqq 2B(M+G^*)g^2(u) + \frac{2}{\tau_0}\eta(u).$$

Since $[c](u)$ is integrable in $[T+k, \infty)$ there exists a k such that $\int_{T+k}^\infty e^{(t-u)/\tau_0}[c](u)du \leqq \epsilon/2$, for all $T \leqq t \leqq T+K$.

Suppose k is fixed so that the above inequality is satisfied. On the sphere $\|a\| \leqq G$, the function $f(a)$ is continuous, and therefore, uniformly continuous. So, given any $\gamma>0$ there exists a $\delta>0$ such that $\|a_1-a_2\| \leqq \delta$ implies $|f(a_1)-f(a_2)\| \leqq \gamma$. Thus, for (a_1-a_2), (h_1-h_2) $\epsilon V_{k,\delta}$, and $T \leqq u \leqq T+k$,

$$[c](u) \leqq \|f[a_1(u)][v_1(u)-v_2(u)]\| + |f[a_1(u)] - f[a_2(u)]| \cdot \|v_2(u)\| + \frac{1}{\tau_0}\|h_1(u)-h_2(u)\|$$

$$\leqq BG^2 \cdot \int_T^u \|a_1(\tau)-a_2(\tau)\|d\tau + \gamma \cdot (M+G^*) + \frac{\delta}{\tau_0}$$

$$\leqq \left(kBG^2 + \frac{1}{\tau_0}\right)\delta + (M+G^*)\gamma \equiv E(\delta,\gamma).$$

Thus,

$$\int_T^{T+k} e^{(T+k-u)/\tau_0}[c](u)du \leqq \tau_0(e^{k/\tau_0}-1)E(\delta,\gamma) \leqq \epsilon/2,$$

for suitable choices of δ, γ, and the continuity of J follows.

To show that $\mathrm{Cl}(J[A(g),H(\eta)])$ is compact, it suffices to show that every sequence in $\mathrm{Cl}(J[A(g),H(\eta)])$ contains a covergent subsequence, as $C_T{}^n$ is a metric space. To show this, it suffices to show that any sequence of functions restricted to a finite interval, contains a uniformly convergent subsequence, (for by the diagonal process, this implies convergence in the sense of the topology on $C_T{}^n$). A useful tool here is Ascoli's theorem: A sequence of equicontinuous, uniformly bounded functions mapping a compact set K into R^n contains a uniformly convergent subsequence.

We will show that if a sequence $\{a_n,h_n\} \subset A(g) \times H(\eta)$, then $\{J(a_n,h_n)\}$ is uniformly bounded and equicontinuous over a finite interval, say $[T, T+k]$.

That the sequence is uniformly bounded is evident; for if $\|v_0\| \leqq M$, $g \leqq G$, $\int_T^\infty g \leqq G^*$, and $\int_T^\infty \eta(u)du \leqq N^*$, for some constants M, G, G^*, and N^*, then

$$\|J(a_n,h_n)(t)\| \leqq \int_t^\infty e^{(t-u)/\tau_0}\left[BG^2(M+G^*) + \frac{1}{\tau_0}\eta(u)\right]$$

$$\leqq \tau_0 BG^2(M+G^*) + N^* \quad \text{for} \quad t \geqq T.$$

For equicontinuity, we must show that for any $\epsilon>0$, there exists a $\delta>0$ such that for any t_1, $t_2 \epsilon [T, T+k]$, and any $(a_n,h_n) \subset A(g) \times H(\eta)$, $|t_1-t_2| \leqq \delta$ implies $\|J(a_n,h_n)(t_1)-J(a_n,h_n)(t_2)\| \leqq \epsilon$. Now, if $t_2 \geqslant t_1$,

$$\|J(a_n,h_n)(t_1)-J(a_n,h_n)(t_2)\|$$

$$\leqq |e^{t_1/\tau_0}-e^{t_2/\tau_0}| \int_{t_1}^\infty e^{u/\tau_0}\left\| f[a_n(u)]v_n(u) + \frac{1}{\tau_0}h_n(u)\right\| du$$

$$+ e^{t_2/\tau_0}\int_{t_1}^{t_2} e^{-u/\tau_0}\left\| f[a_n(u)]v_n(u) + \frac{1}{\tau_0}h_n(u)\right\| du$$

$$\equiv I_1 + I_2.$$

Now, in the term I_1, the integral is bounded, as shown above, for any (a_n,h_n) and the function e^{t/τ_0} is uniformly continuous on the interval $[T, T+k]$. Thus $I_1 \leqq \epsilon/2$ can be satisfied by $|t_1-t_2| \leqq \delta$, for a suitable δ.

The integral I_2 can also be made $\leqq \epsilon/2$ by suitably restricting δ.

Lemma 2. Given the vector-valued function, $F(z,\dot{z},t)$, F continuous in all its arguments, the operator F: $C_T{}^n \to C_T{}^n$ defined by

$$F(a)(t) = F[z(t),v(t),t],$$

where

$$v(t) = v_0 + \int_T^t a(u)du,$$

$$z(t) = z_0 + v_0(t-T) + \int_T^t \int_T^u a(\tau)d\tau du,$$

z_0, $v_0 \epsilon R^n$, is continuous on $A(g)$ in the compact-open topology.

Proof. For any given $\epsilon>0$ define $V_{m,\epsilon}$ by 3). We wish to show there exists a $V_{k,\delta}$ such that $a_1-a_2 \epsilon V_{k,\delta}$ implies $[F(a_1)-F(a_2)] \epsilon V_{m,\epsilon}$. Here let $k=m$, and note that, for $a \epsilon A(g)$, and $t \epsilon [T, T+m]$, z and v are bounded and lie in some sphere. But $F(z,v,t)$ is uniformly continuous in z, v on this sphere. Noting further that $\|z(a_1)(t)\| \leqq m\delta^2$, and $\|v(a_1)(t)-v(a_2)(t)\| \leqq m\delta$, for $t \epsilon [T, T+m]$, and $(a_1-a_2) \epsilon V_{m,\delta}$, the result follows easily.

Lemma 3. Suppose $T \geqq 0$ and $v_0 \epsilon R^n$ are given. If $g \epsilon C_T{}^1$ is a nonnegative, nonincreasing function such that $g(t) \to 0$ as $t \to \infty$ and $\int_T^\infty g(t)dt < \infty$, then there exists a $\tau_0'>0$ such that for $0<\tau_0<\tau_0'$, $J[A(g),H(\eta)] \subset A(g)$, $\eta = (1-\tau_0/\tau_0')g$. Furthermore, if $\|v_0\| \leqq M$, $g \leqq G$, $\int_T^\infty g(t)dt \leqq G^*$, where M, G, G^* are given positive constants, then $\tau_0' = [BG(M+G^*)]^{-1}$, where B is given in (2).

Proof. Let M, G, G^*, τ_0', and $\eta(t)$, $t \geqq T$, be defined as in the lemma. We need to show that $\|a(t)\| \leqq g(t)$, $\|h(t)\| \leqq \eta(t)$, $t \geqq T$, implies $\|J(a,h)(t)\| \leqq g(t)$, $t \geqq T$. If $a \epsilon A(g)$, $h \epsilon H(\eta)$, then

$$\|J(a,h)(t)\|$$

$$\leqq \int_t^\infty e^{(t-u)/\tau_0}\left[B\|a(u)\|^2\left(\|v_0\|+\int_T^u \|a(\tau)\|d\tau\right)+\|h(u)\|/\tau_0\right]du$$

$$\leqq \int_t^\infty e^{(t-u)/\tau_0}\left[Bg^2(u)\left(M+\int_T^u g(\tau)d\tau\right)+\eta(u)/\tau_0\right]du$$

$$\leqq \int_t^\infty e^{(t-u)/\tau_0}[BG(M+G^*)+1/\tau_0-1/\tau_0']g(u)du$$

$$\leqq g(t)\int_t^\infty \tau_0^{-1}e^{(t-u)/\tau_0}du=g(t),$$

since $g(t)\geqq g(u)$, for $u\geqq t\geqq T$, and the lemma is proved.

Lemma 4. Suppose $T\geqq 0$ and $v_0\epsilon R^n$ are given. If $\eta\epsilon C_T^1$ is any function, $0\leqq\eta(t)\leqq N$, $t\geqq T$, for some constant N, $\int_T^\infty\eta(t)dt<\infty$, there exists a $\tau_1'>0$ such that for $0<\tau<\tau_1'$, there is a nonnegative function $g\epsilon C_T^1$, $g(t)\to 0$ as $t\to\infty$, $\int_T^\infty g(t)dt<\infty$ such that $J[A(g),H(\eta)]\subset A(g)$. Furthermore, if

$$\|v_0\|\leqq M,\quad 0\leqq\eta\leqq N,\quad \int_T^\infty\eta(t)dt\leqq N^*,$$

where M, N, N^* are positive constants and δ is any given number, $0<\delta<1$, then τ_1' is given by (7) below and $0\leqq g\leqq G$, $\int_T^\infty g(t)dt\leqq G^*$, where $G=N/\delta$, $G^*=N^*/\delta+\tau_0N/\delta^2$, $0<\tau_0<\tau_1'$.

Proof. Suppose M, N, N^*, G, G^*, are defined as in the lemma and let $K=BG(M+G^*)$. We first show that for any δ, $0<\delta<1$, there exists a $\tau_1'=\tau_1'(\delta)$ such that $1-\tau_0K\geqq\delta$, if $0<\tau_0<\tau_1'$. From the definition of G, G^*, $1-\tau_0K\geqq\delta$ if and only if

$$k(\tau_0,\delta)\equiv\tau_0BN(M\delta^2+N^*\delta+\tau_0N)+\delta^4-\delta^3\leqq 0.$$

But for a given δ, $0<\delta<1$, $k(\tau_0,\delta)$ is a quadratic function of τ_0 which approaches $+\infty$ as $\tau_0\to\pm\infty$ and $k(0,\delta)<0$. Therefore, $\tau_1'=\tau_1'(\delta)$ may be taken to be the positive number such that $k(\tau_1',\delta)=0$; that is,

$$\tau_1'=\frac{1}{2N}[-(M\delta^2+N^*\delta)+\{(M\delta^2+N^*\delta)^2+4(\delta^3-\delta^4)B\}^{\frac{1}{2}}]. \quad (7)$$

Of course one could find the largest τ_1' by choosing the proper value of δ, $0<\delta<1$.

Now with this choice of τ_1' and the given function $\eta(t)$, we define the function $g(t)$ by the relation

$$g(t)=\int_t^\infty \exp[(\tau_0^{-1}-K)(t-u)]\tau_0^{-1}\eta(u)du.$$

Clearly, $g(t)\leqq N/(1-\tau_0K)\leqq N/\delta=G$. Since η is integrable on $[T,\infty)$, it is evident that $g(t)\to 0$ as $t\to\infty$. Further, differentiating the relation defining g, we obtain

$$\dot g(t)=(1/\tau_0-K)g(t)-\eta(t)/\tau_0 \quad (8)$$

so that

$$\int_T^\infty g(t)dt\leqq\frac{\tau_0G}{1-\tau_0K}+\frac{1}{1-\tau_0K}\int_T^\infty\eta(t)dt$$

$$\leqq\frac{\tau_0N}{\delta^2}+\frac{N^*}{\delta}=G^*.$$

Also, from this fact, and (8), it follows that g satisfies

$$g(t)=\int_t^\infty e^{(t-u)/\tau_0}\left[Kg(u)+\frac{1}{\tau_0}\eta(u)\right]du. \quad (9)$$

Now it remains to show that $J[A(g),H(\eta)]\subset H(\eta)$.

If $a\epsilon A(g)$, $h\epsilon H(\eta)$, then, as before

$$\|J(a,h)(t)\|$$

$$\leqq\int_t^\infty e^{(t-u)/\tau_0}[BG(M+G^*)g(u)+\eta(u)/\tau_0]du$$

$$=\int_t^\infty e^{(t-u)/\tau_0}[Kg(u)+\eta(u)/\tau_0]du=g(t),$$

by (9) and, thus, $J(a,h)\epsilon A(g)$. The lemma is proved.

The previous lemmas illustrate very clearly the problem that remains to be solved in order to assert the existence of a "physically realizable" solution of (1). In fact, Lemmas 1 and 2 give us the necessary continuity properties for the application of Theorem 1 and Lemmas 3 and 4 give us relationships between $A(g)$, $H(\eta)$ which insure that $J[A(g),H(\eta)]\subset A(g)$. Consequently, the only thing that remains is to find functions $F(z,v,t)$ such that, if z, v are defined as above in terms of a, then $F(a)(t)\subset H(\eta)$ if $a\epsilon A(g)$. We illustrate these remarks in the next three theorems.

Theorem 2. Consider Eq. (1) and suppose $F(z,\dot z,t)$ is continuous in z, $\dot z$, and t, and is such that for any ρ_1, $\rho_2>0$, there exists $\gamma\epsilon C_0^1$ so that if $v\epsilon C_0^n$ has the property that $\lim_{t\to\infty}v(t)=v_\infty$ exists, and $\|v_\infty\|\leqq\rho_2$, then for $\|z_0\|\leqq\rho_1$

$$\left\|F\left(z_0+\int_0^t v(u)du, v(t), t\right)\right\|\leqq\gamma(t),\quad \gamma(t)\to 0 \text{ as } t\to\infty,\quad \int_0^\infty\gamma(u)du<\infty.$$

Then for any $\tau_0>0$, and any z_0, $v_0\epsilon R^n$ there exists a $T\geqq 0$ and a solution $z\epsilon C_T^n$ of (1) satisfying

$$z(T)=z_0,\quad \dot z(T)=v_0,\quad \|\ddot z(t)\|\to 0 \text{ as } t\to\infty,$$

and $$\int_T^\infty\|\ddot z(t)\|dt<\infty.$$

Proof. The proof uses Lemma 4. Suppose $\|v_0\| \leqq M$, $M>0$, and τ_0 is any given number, $\tau_0>0$. Choose N, N^* so that τ_1' in (7) is such that $\tau_1'>\tau_0$ and define the numbers G, G^* as in Lemma 2. Let $\rho_1=\|z_0\|$, $\rho_2=M+G^*$ and choose $T\geqq 0$ so large that $\gamma(t)\leqq N$, $t\geqq T$,

$$\int_T^\infty \gamma(t)dt \leqq N^*.$$

If we define $\eta(t)=\gamma(t)$, $t\geqq T$, then, from Lemma 4, there exists a $g\in C_T^1$ such that $g(t)\leqq G$, $t\geqq T$, $g(t)\to 0$ as $t\to\infty$, $\int_T^\infty g(t)dt<G^*$ and $J[A(g),H(\eta)]\subset A(g)$.

Recalling Lemmas 1 and 2 and Theorem 1, the proof of Theorem 2 will be complete if we can show that $F[A(g)]\subset H(\eta)$. But, $a\in A(g)$ implies that $\lim_{t\to\infty}v(t)=v_\infty$ exists and $\|v_\infty\|\leqq M+G^*$. Further, the statement $\|F(z_0+\int_0^t v(u)du, v(t), t)\|\leqq\gamma(t)$, for $v\in C_0^n$ such that $\lim_{t\to\infty}v(t)=v_\infty$, etc., implies the statement

$$\left\|F\left[z_0+\int_T^t v(u)du, v(t), t\right]\right\|_t \leqq \gamma(t),$$

for $t\geqq T$, for $v\in C_T^n$ such that $\lim_{t\to\infty}v(t)=v_\infty$, etc., because in each case the first argument of F has the form $z_0+v_\infty t+l(t)$, where $l(t)=\int_0^t[v(u)-v_\infty]du$, or $\int_T^t[v(u)-v_\infty]du$, respectively. Thus, the second statement is simply a special case of the first, given by a particular choice of the function $l(t)$. It follows then that $F[A(g)]\subset J(\eta)$ and by Theorem 1, the result follows.

Remark. An example of a field F satisfying the hypothesis of Theorem 2 is given by a bounded field acting over finite time interval, a case considered by Plass[6] in some detail.

Theorem 3. Suppose the conditions of Theorem 2 are satisfied except $\gamma(t)$ is assumed to satisfy the conditions $0\leqq\gamma(t)\leqq N$, $t\geqq 0$, $\int_0^\infty\gamma(t)dt<N^*$ for some constants N, N^*. Then for any z_0, $v_0\in R^n$, there exists a $\tau_1'>0$ (independent of z_0) such that for any τ_0, $0<\tau_0<\tau_1'$, there is a solution $z\in C_0^n$ of 1) satisfying $z(0)=z_0$, $\dot z(0)=v_0$, $\|\ddot z(t)\|\to 0$ as $t\to\infty$, and $\int_0^\infty|\ddot z(t)\|dt<\infty$.

Proof. The proof uses Lemma 4. Suppose $\|v_0\|\leqq M$, $M>0$, and define τ_1' by (7). Letting $\gamma(t)=\eta(t)$, $t\geqq 0$, the proof is completed in the same way as the proof of Theorem 2.

Theorem 4. Consider Eq. (1), and suppose $F(z,\dot z,t)$ is continuous in z, $\dot z$, and t, and satisfies

$$\|F(z,\dot z,t)\|\leqq\frac{b+c\|\dot z\|}{\|z\|^\alpha},\quad b, c>0, \alpha>1,$$

for all vectors z, $\dot z$, $\|z\|>0$, $t\geqq 0$.

Suppose $T>0$, $z_0\in R^n$, $v_0\in R^n$, are given and define $\beta=[\alpha-1)T^{\alpha-1}]^{-1}$,

$$\tau_0'=T^\alpha[B(b+c\|v_0\|)\{\|v_0\|+\beta(b+c\|v_0\|)\}]^{-1}.$$

If, for given τ_0, $0<\tau_0<\tau_0'$, there exist T, z_0, $v_0\in R^n$ such that

$$\inf_{t\geqq T}\left\|\frac{z_0}{t}+\frac{v_0(t-T)}{t}\right\|\geqq\beta(b+c\|v_0\|)+\left(\frac{1+c\beta}{1-\tau_0/\tau_0'}\right)^{1/\alpha} \quad (11)$$

then there is a solution $z\in C_T^1$ of (1) with

$$z(T)=z_0,\quad \dot z(T)=v_0,\quad \ddot z(t)\to 0$$

$$\text{as}\quad t\to\infty,\quad \int_T^\infty\|\ddot z(t)\|dt<\infty.$$

Remarks. (i) If $v_0\in R^n$ and $\tau_0>0$ are given, then there exists a $T_1>0$ such that for all $T\geqq T_1$, the constant τ_0' defined in Theorem 4 is $>\tau_0$, and there is a $z_0\in R^n$, depending on T, v_0, τ_0, such that inequality (11) is satisfied.

(ii) If $v_0\in R^n$, $\|v_0\|>1$, and $\tau_0>0$ are given then there exists a T_2 such that the constant τ_0' defined in Theorem 4 is $>\tau_0$ for all $T\geqq T_2$ and inequality (11) is satisfied if $z_0=v_0T$, $T\geqq T_2$.

(iii) In case $c=0$ in (10), the above relationships are somewhat simpler. In fact, the right-hand side of (11) is a bounded function of $\|v_0\|$.

(iv) For any z_0, v_0, the parameter $\tau_0'\to\infty$ as $T\to\infty$ and the right-hand side of (11) $\to 0$ as $T\to\infty$. However, for any fixed T, there is still a strong restriction on z_0, v_0, and it is not sufficient that the norm of z_0, v_0 be only larger than some constant. In fact, suppose every component of $v_0=\sigma>0$ and every component of $z_0=-\mu<0$. Then for any $\sigma>0$ there always exists a $t_0>T$ such that $z_0/t_0-(t_0-T)v_0/t_0=0$ and so (11) cannot hold.

(v) Inequality (11) can never be satisfied for $T=0$, since the right hand side approaches ∞ as $T\to 0$.

Proof of Theorem 4. Suppose $T>0$, $z_0\in R^n$, $v_0\in R^n$, are given and define β, τ_0' as in the theorem and let $K=b+c\|v_0\|$, $g(t)=K/t^\alpha$, $t\geqq T$. Then $g\in C_T^1$, $0\leqq g(t)\leqq G$, $t\geqq T$, $G=K/T^\alpha$, $\int_T^\infty g(t)dt=G^*=\beta K$. If τ_0' is defined as above, it follows that, for $0<\tau_0<\tau_0'$, $J[A(g),H(\eta)]\subset A(g)$, $\eta=(1-\tau_0/\tau_0')g$.

Recalling Lemmas 1 and 2 and Theorem 1, the proof of the theorem will be complete if we can show that $F[A(g)]\subset H(\eta)$; that is $\|F(a)(t)\|\leqq\eta(t)$. Now, for

$$a\in A(g),\quad \|v(t)\|\leqq\|v_0\|+G^*,$$

$$\|z(t)/t\|\geqq\|z_0/t+(t-T)v_0/t\|-G^*.$$

Using (11), the result follows immediately.

If one chose K above as $K=d(b+c\|v_0\|)$, where d is to be determined, the inequality (11) would involve d and, therefore, a wider range of z_0, v_0 would satisfy (11).

ACKNOWLEDGMENT

The authors are indebted to Professor Rohrlich for many discussions and free access to his paper prior to publication.

[6] See reference 1.

Construction of Potentials from the Phase Shifts at Fixed Energy

ROGER G. NEWTON*
Department of Physics, Indiana University, Bloomington, Indiana
(Received August 3, 1961)

The nonrelativistic potential energy between two spinless particles is deduced from a knowledge of all phase shifts at a given energy. A spherically symmetric potential is found always to exist, but it is not unique. In particular, for every energy, there exists at least one nonzero potential which causes the scattering cross section to be zero. The paper contains both the formal construction procedure and the necessary existence and uniqueness (or lack of it) proofs. Some general examples are included.

I. INTRODUCTION

THE problem of finding the nonrelativistic potential energy between two particles as a function of their relative distance, starting from information provided by scattering experiments, is of obvious physical importance as well as intrinsic interest. Since the measured scattering cross section is a function of angles as well as energy, there are several different ways of approaching this problem, depending on which of these variables are varied and which are fixed. Assuming that the potential is spherically symmetric, we may make a partial wave analysis and examine the phase shift of each angular momentum as a function of the energy. The well-known result is that the phase shift for one angular momentum, given as a function of the energy from zero to infinity, together with the binding energies of all the n-bound states of the same angular momentum, restricts the potential to an n-parameter family. These potentials were first constructed, generally and elegantly, by Jost and Kohn,[1] using an integral equation of Gel'fand and Levitan.

The deduction of the potential from a knowledge of the scattering amplitude as a function of the angle, at a fixed energy, has meanwhile received much less attention, in spite of the fact that it is much more useful from a practical point of view. After all, not only do we never know a phase shift for all energies up to infinity, but at high energies the very assumption entering into the solution of the problem, namely, nonrelativistic quantum mechanics of a single channel, is certainly incorrect.

These objections do not arise if we start from the scattering amplitude at fixed energy. Moreover, the "inverse problem" in this case also has an important theoretical aspect. As long as the potential picture is at all applicable, it is a very useful tool for the understanding of physical processes and for the prediction of others. It is therefore quite desirable to be able to construct the underlying potential even when a more fundamental theory predicts the scattering amplitude without it as an intermediary. The numerous attempts at finding a low energy nucleon-nucleon potential from field theory are a case in point; possible future potential calculations from dispersion theory would be another. The determination of a potential from a field or dispersion theoretically-calculated scattering amplitude would avoid all the ambiguities that beset the customary procedure. It would at the same time take care of the question of the *existence* of a momentum-independent potential. At each energy the scattering amplitude determines a potential; if it varies strongly with the energy then that is in the nature of the particles at hand.

To be sure, the present method is not as yet quite simple enough to make it a very practical tool for a potential calculation from field or dispersion theory. Primarily, the reason is that it relies on a phase-shift analysis. Nevertheless, it points the way in a direction to be further explored.

The problem of finding the potential from a knowledge of all the phase shifts at one energy has been treated in the past in three papers. Wheeler[2] discussed it in 1955 from the point of view of the WKB approximation. More recently Regge[3] dealt with it, partly by conjecture, by extending the value of the angular momentum into the complex plane. His approach is based on the same analogy with the Gel'fand-Levitan equation that is used in the present paper, but the introduction of complex l values proves to be an unnecessary complication. Most recently Martin and Targonski[4] treated the problem by a method applicable only to superpositions of Yukawa potentials.

The reason why the inverse problem for all phase shifts and one energy has resisted solution so much longer than that for one phase shift and all energies, is that an important tool for the solution of the latter was the completeness and orthogonality of the radial wave functions of one l value and all momenta; the wave functions of one energy and all l values have no analogous property. The present solution is based on an approach to the Gel'fand-Levitan equation which does not depend on completeness, orthogonality, or complex variable techniques. As a result, the equation could easily be transferred from a context of k integration to one of l summation.

The procedure will be as follows. In Sec. II, we do

* Supported in part by the National Science Foundation.

[1] R. Jost and W. Kohn, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **27**, No. 9 (1953). For an excellent review of this problem and a complete bibliography on it, see L. D. Faddeyev, Uspekhi Matem. Nauk **14**, 57 (1959) [Translation: New York University Research Rept. No. EM 165, Dec. 1960].

[2] J. A. Wheeler, Phys. Rev. **99**, 630 (1955).

[3] T. Regge, Nuovo cimento **14**, 951 (1959).

[4] A. Martin and Gy. Targonski, Nuovo cimento **20**, 1182 (1961).

the formal manipulations that lead us from the phase shifts to the potential and to the radial wave functions. The context is that of a Schrödinger equation in the center-of-mass coordinate system for two spinless particles. Section III provides the necessary existence and uniqueness (or lack of it) proofs for the equations of Sec. II. Section IV gives some examples. There are two appendixes. In the first, an inverse of an auxiliary infinite matrix is explicitly constructed in closed form; in the second, the procedure of Sec. II is extended to the use of an arbitrary comparison potential as a starting point.

The conclusion of the paper is this. Provided only that the phase shifts tend to zero sufficiently rapidly with increasing l value,[5] there always exists an underlying local central potential. Moreover, it can be constructed in a straightforward manner by solving an infinite set of linear algebraic equations and an integral equation of the Fredholm type. The potential, however, is not unique. In particular, there exists at least one nonzero potential which, at the given energy, causes no scattering whatever. It is not known how this potential behaves at large distances, so that it is still quite possible that within a sufficiently narrow class of potentials which asymptotically decrease rapidly enough, the correspondence to the phase shift is unique.[6]

A word should be said about the dependence on the energy at which the potential determination is made. It is physically clear that at very low energies the particles do not approach one another closely enough to be able to see the inner regions of the potential very clearly. Mathematically, this must express itself in a high degree of sensitivity to information that is experimentally inaccessible. In other words, *in principle*, even low-energy phase shifts serve to determine a potential; but then the s-phase shift dominates to such an extent that the others cannot be measured, and the potential will be sensitive to them even though they are small. In practice, therefore, one ought to work at an energy where all phase shifts that are ever going to be of significant size come into play.

II. FORMAL PROCEDURE

We start with a given function $f(r,r')$ defined by the infinite series

$$f(r,r')\equiv\sum_{l=0}^{\infty}c_l u_l(r)u_l(r'), \tag{1}$$

with real coefficients c_l and the regular spherical Riccatti-Bessel functions

$$u_l(r)\equiv r j_l(r).$$

The radial distance r is measured in units of λ, the reduced wavelength of the relative motion, which is fixed throughout the following. The differential equation satisfied by u_l is written

$$D_0(r)u_l(r)=l(l+1)u_l(r), \tag{2}$$

with the differential operator

$$D_0(r)\equiv r^2(\partial^2/\partial r^2+1). \tag{3}$$

Consequently $f(r,r')$ satisfies the partial differential equation

$$D_0(r)f(r,r')=D_0(r')f(r,r') \tag{4}$$

and the boundary condition

$$f(0,r')=f(r,0)=0. \tag{4'}$$

Let the function $K(r,r')$ be the *unique* solution of the integral equation[7]

$$K(r,r')=f(r,r')-\int_0^r dr''r''^{-2}K(r,r'')f(r'',r'); \tag{5}$$

that is, subject to proof below, we shall assume for the moment that (5) has a solution and furthermore, that that solution is unique. Next we define the auxiliary function

$$\xi(r,r')\equiv D(r)K(r,r')-D_0(r')K(r,r'),$$

with

$$D(r)\equiv D_0(r)-r^2V(r), \tag{6}$$

where[8]

$$V(r)\equiv -2r^{-1}(d/dr)[r^{-1}K(r,r)]. \tag{7}$$

It is a matter of straightforward differentiation, integration by parts, and use of the differential equations (4) as well as of the boundary condition (4'), to show that $\xi(r,r')$ satisfies the homogeneous version of (5). Since we assumed that the solution of (5) is unique, we conclude that

$$\xi(r,r')\equiv 0.$$

In other words, $K(r,r')$ satisfies the partial differential equation

$$D(r)K(r,r')=D_0(r')K(r,r'). \tag{8}$$

The integral equation (5) and (4') imply, in addition, that K fulfills the boundary condition

$$K(r,0)=K(0,r')=0. \tag{8'}$$

We now use the solution $K(r,r')$ of the integral equation (5) in order to define the function[9]

$$\varphi_l(r)\equiv u_l(r)-\int_0^r dr'r'^{-2}K(r,r')u_l(r'). \tag{9}$$

[5] This is a convenient and physically plausible sufficient condition, which we have not attempted to sharpen. It is certainly not necessary.

[6] See the remarks on the uniqueness problem on pp. 962–963 of reference 3.

[7] This is the analog of the Gel'fand-Levitan equation. The same integral equation was written down by Regge, reference 3.

[8] The actual potential energy is obtained from V by multiplication by the energy of relative motion, $E=\hbar^2k^2/2\mu$. A further implicit dependence on E comes in because r is measured in units of k^{-1}.

[9] φ_l therefore behaves at $r\to 0$ exactly as does u_l.

Application of the differential operator $D(r)$ to (9), together with two integrations by parts and use of (8) and (2), shows that $\varphi_l(r)$ satisfies the differential equation

$$D(r)\varphi_l(r)=l(l+1)\varphi_l(r). \tag{10}$$

Furthermore it follows from (8′) that φ_l is the *regular* solution of (10):

$$\varphi_l(0)=0. \tag{10′}$$

At this point we have constructed, *via* Eq. (9), the regular solutions of the radial Schrödinger equations (10) of all l values, starting with an arbitrary function $f(r,r')$ that satisfies (4) and (4′). We must now relate the input information, namely the infinite set of real numbers c_l in (1), to the phase shifts. To this end we insert (1) in (5) and infer that $K(r,r')$ can be written in the form

$$K(r,r')=\sum_{l=0}^{\infty} c_l\mathcal{K}_l(r)u_l(r'),$$

where

$$\mathcal{K}_l(r)=u_l(r)-\int_0^r dr'r'^{-2}K(r,r')u_l(r').$$

Comparison with (9) shows that $\mathcal{K}_l\equiv\varphi_l$, i.e.,

$$K(r,r')=\sum_{l=0}^{\infty} c_l\varphi_l(r)u_l(r'). \tag{11}$$

Substituting this result in (9) we obtain

$$\varphi_l(r)=u_l(r)-\textstyle\sum_{l'} L_{ll'}(r)c_{l'}\varphi_{l'}(r), \tag{12}$$

where

$$L_{ll'}(r)=\int_0^r dr'r'^{-2}u_l(r')u_{l'}(r'). \tag{13}$$

The infinite set of coupled linear-algebraic equations (12) is equivalent to the integral equation (5) (but not necessarily easier to solve).

In order to get at the phase shifts we let $r\to\infty$ in (12). Then

$$\begin{aligned}&\varphi_l(r)\sim A_l\sin(r-\tfrac{1}{2}\pi l+\delta_l)\equiv\varphi_l^{(\infty)}(r),\\ &u_l(r)\sim\sin(r-\tfrac{1}{2}\pi l)\equiv\varphi_{0l}^{(\infty)}(r),\\ &L_{ll'}(r)\sim L_{ll'}^{(\infty)},\end{aligned} \tag{14}$$

and (12) goes over into

$$\varphi_l^{(\infty)}(r)=\varphi_{0l}^{(\infty)}(r)-\textstyle\sum_{l'} L_{ll'}^{(\infty)}c_{l'}\varphi_{l'}^{(\infty)}(r). \tag{15}$$

The matrix $L_{ll'}^{(\infty)}$ is readily calculated by noticing that it is the integral of the Wronskian of u_l and $u_{l'}$. The differential equation (2) yields

$$L_{ll'}=\frac{u_lu_{l'}'-u_{l'}u_l'}{[l'(l'+1)-l(l+1)]}$$

and from this we get by inserting (14),

$$L_{ll'}^{(\infty)}=\frac{\sin\frac{1}{2}\pi(l'-l)}{(l'-l)(l'+l+1)}\quad\text{for}\quad l'\neq l. \tag{16}$$

For $l'=l$ we obtain from the recursion formulas for u_l

$$L_{ll}^{(\infty)}=\tfrac{1}{2}\pi/(2l+1). \tag{16′}$$

If we now insert the explicit functions (14) in their complex form into (15) and equate the coefficients of e^{ikr} and e^{-ikr} separately, we get the set of equations

$$e^{i\delta_l}A_l\left(1+\frac{\frac{1}{2}\pi}{2l+1}c_l\right)=1-\sum_{l'\neq l}L_{ll'}^{(\infty)}i^{l-l'}c_{l'}e^{i\delta_{l'}}A_{l'} \tag{17}$$

and another set, which by virtue of the reality of the c_l is merely the complex conjugate of (17). A glance at (16) shows that for $l'\neq l$,

$$L_{ll'}^{(\infty)}i^{l-l'}=-iM_{ll'},$$

where

$$M_{ll'}=\begin{cases}[(l'-l)(l'+l+1)]^{-1}, & \text{if}\quad l'-l\text{ is odd},\\ 0, & \text{if}\quad l'-l\text{ is even},\end{cases} \tag{18}$$

in consequence of which (17) becomes

$$A_le^{i\delta_l}=1-\tfrac{1}{2}\pi\frac{c_lA_le^{i\delta_l}}{(2l+1)}+i\sum_{l'=0}^{\infty}M_{ll'}c_{l'}A_{l'}e^{i\delta_{l'}}. \tag{19}$$

Since a knowledge of the scattering amplitude gives only the δ_l and not the A_l, we set

$$b_l\equiv c_lA_l \tag{20}$$

and multiply (19) by $e^{-i\delta_l}$,

$$A_l=e^{-i\delta_l}-\tfrac{1}{2}\pi b_l/(2l+1)+i\textstyle\sum_{l'}M_{ll'}b_{l'}e^{i(\delta_{l'}-\delta_l)}. \tag{21}$$

These equations must be solved for the infinite sets of real numbers b_l *and* A_l. We accomplish a separation of the two sets by considering separately the real and imaginary parts of (21). The imaginary part is

$$\sin\delta_l=\textstyle\sum_{l'}M_{ll'}b_{l'}\cos(\delta_{l'}-\delta_l), \tag{22}$$

whereas the real part is

$$A_l=\cos\delta_l-\tfrac{1}{2}\pi b_l/(2l+1)-\textstyle\sum_{l'}M_{ll'}b_{l'}\sin(\delta_{l'}-\delta_l). \tag{23}$$

The problem therefore is to solve the infinite set of coupled linear-algebraic equations (22) for b_l with given δ_l. Insertion of the result in (23) then explicity gives A_l and (20) yields the wanted c_l.

To finish the formal development we indicate a simplification for the practical solution of Eqs. (22). If we write $\cos(\delta_{l'}-\delta_l)$ explicitly in terms of sin and cos of $\delta_{l'}$ and δ_l, divide by $\cos\delta_l$ and set

$$a_l\equiv b_l\cos\delta_l \tag{24}$$

then we get

$$\tan\delta_l=\textstyle\sum_{l'}M_{ll'}a_{l'}(1+\tan\delta_l\tan\delta_{l'}), \tag{25}$$

or, in matrix notation,

$$\tan\Delta\hat{e}=M\hat{a}+\tan\Delta M\tan\Delta\hat{a}, \tag{25′}$$

where $\tan\Delta$ is the diagonal matrix

$$(\tan\Delta)_{ll'}=\delta_{ll'}\tan\delta_l,$$

$\hat{e}$ and $\hat{a}$ are the column matrices

$$(\hat{e})_l=1, \quad (\hat{a})_l=a_l,$$

and M is the matrix whose elements $M_{ll'}$ are given by (18) and are independent of the phase shifts. The matrix equation (25′) can then be written

$$\hat{a}=M^{-1}\tan\Delta\hat{e}-R\hat{a}, \tag{26}$$

where

$$R=M^{-1}\tan\Delta M\tan\Delta, \tag{27}$$

and (26) can be solved formally by

$$\hat{a}=(1+R)^{-1}M^{-1}\tan\Delta\hat{e}. \tag{28}$$

If M^{-1} is constructed once and for all, then for any given set of phase shifts that vanishes sufficiently rapidly as $l\to\infty$, $(1+R)^{-1}$ can be obtained, say, by the Fredholm method. Indeed, if we assume, as in any practical case we are likely to, that $\delta_l=0$ for all $l\geqslant L$, then only the first L equations in (26) need be solved, and hence only an L-dimensional matrix need be inverted; the remaining a_l for $l\geqslant L$ are explicitly given by (26) in terms of the first L. (We still need all elements of the infinite matrix M^{-1}, though.)

This ends the formal aspects of the solution of the inversion problem. The procedure, then, is this. With the given set of phase shifts, we solve Eqs. (22) for the b_l, or equivalently, (26) for the a_l, and by (20) get the c_l from them. Equation (1) uses the c_l to define the functions $f(r,r')$. We then solve the linear integral equation (5) for $K(r,r')$. Equation (7) finally gives us the potential and (9), the regular wave functions for all l values. For practical purposes, it would be quite convenient to have an explicit, closed form expression for the infinite matrix M^{-1}. Unfortunately, the author did not succeed in obtaining that, so that numerical methods must be had recourse to.

There remain the following important points to be cleared up in the next section. It must be proved that the integral equation (5) has a unique solution, it must be proved that the infinite set of linear equations (22) has a solution, and it ought to be determined whether that solution is unique. As for the last point, we shall see that the solution of (22) indeed is *not* unique. Specifically, there exists a nontrivial solution of the set of equations (22) for all $\delta_l=0$:

$$\sum_{l'} M_{ll'}b_{l'}^{(0)}=0, \quad \text{all} \quad l\geqslant 0,$$

which satisfies the inequality[10]

$$|b_l^{(0)}|\leqslant C\log(2+l).$$

It then follows from (23) and (20) that

$$|c_l^{(0)}|\leqslant C\log(2+l). \tag{29}$$

Substituting in (1) we obtain a series whose convergence is tested by (29) and the asymptotic behavior of the spherical Bessel functions[11] for fixed r and large l:

$$u_l(r)\sim C\exp[l+l\log(\tfrac{1}{2}r)-(l+1)\log l]. \tag{30}$$

It is readily seen that for every fixed R and R' the series $f(r,r')$ converges absolutely and uniformly in $0\leqslant r\leqslant R$, $0\leqslant r'\leqslant R'$. The same holds for its first two derivatives, obtained by differentiating term by term. As a result $K^{(0)}(r,r')\not\equiv 0$ exists and is twice differentiable, and $V^{(0)}(r)$ exists for every fixed $r>0$. Furthermore, Eq. (7) will not give us $V^{(0)}(r)\equiv 0$, because if it did, comparison of (1) and (11) shows that we would have to have $f^{(0)}(r,r')\equiv K^{(0)}(r,r')$; (5) would then say after insertions of (1),

$$\sum_l c_l^{(0)}L_{ll'}(r)c_{l'}^{(0)}u_l(r)u_{l'}(r')\equiv 0$$

with $L_{ll'}$ given by (13). For $r'=r$ this equation contradicts the positive definiteness of $L_{ll'}(r)$. Thus we come to the conclusion that there exists at least one nonzero potential which, at the given energy, produces no scattering. How well that potential behaves for large values of r is a difficult matter to determine. It may very well be that it decreases too slowly to be considered "acceptable." In order to clear up definitely the uniqueness question, one would have to investigate first for a given large class of potentials, how fast A_l approaches 1 as $l\to\infty$, thus restricting via (23) the asymptotic behavior of the "acceptable" sets b_l, and then show that among them none annihilates M.

We now turn to the necessary existence and uniqueness proofs.

III. EXISTENCE AND UNIQUENESS PROOFS

A. The Solution of Eq. (5)

We first want to show the existence and uniqueness of the solution to the integral equation (5). Suppose that the homogeneous version of (5) had a nontrivial solution $\chi(r,r')$:

$$\chi(r,r')=-\int_0^r dr''r''^{-2}\chi(r,r'')f(r'',r'). \tag{31}$$

Then it follows from (1) that $\chi(r,r')$ can be written

$$\chi(r,r')=\sum_l c_l\chi_l(r)u_l(r'), \tag{32}$$

where

$$\chi_l(r)=-\int_0^r dr'r'^{-2}\chi(r,r')u_l(r'). \tag{33}$$

Multiplication by $c_l\chi_l(r)$ and summation over l yields by (32),

$$\sum_l c_l\chi_l^2(r)=-\int_0^r dr'r'^{-2}\chi^2(r,r'). \tag{34}$$

We differentiate this and compare the result with that obtained by differentiating (33), multiplying by $c_l\chi_l(r)$,

[10] In the following, we use C as a general constant. It is not meant to have the same value every time it is used.

[11] See G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, New York, 1958), 2nd Edition, p. 225.

and summing over l. The two results are compatible only if

$$\chi(r,r)\equiv 0. \tag{35}$$

Next, differentiate (31) with respect to r. Because of (35), $\partial\chi(r,r')/\partial r$ is also a solution of the homogeneous equation (31). Consequently, we must also have

$$\lim_{r\to r'}(\partial/\partial r)\chi(r,r')\equiv 0,$$

and because of (35)

$$\lim_{r'\to r}(\partial/\partial r')\chi(r,r')\equiv 0.$$

This process may be repeated *ad infinitum:*

$$\lim_{r'\to r}(\partial^n/\partial r'^n)\chi(r,r')\equiv 0 \quad \text{for all} \quad n\geqslant 0. \tag{36}$$

The existence of all these derivatives and the legitimacy of differentiating under the integral sign is easily established if we make the very weak assumption that for large l, the coefficients c_l in (1) increase at most by a fixed power of l:

$$|c_l|C\leqslant l^p. \tag{37}$$

The estimate (30) of the asymptotic behavior of the u_l for large l, together with their analyticity for all finite r, shows that $f(r,r')$ is a regular analytic function of r and r' for all finite r and r'. As a result, it is clear that all solutions $\chi(r,r')$ of (31) must be analytic functions of r' regular for all r'. We may therefore differentiate (31) also with respect to r under the integral sign.

The vanishing of all r' derivatives of $\chi(r,r')$ at $r'=r$, in conjunction with its analyticity there, allows us to draw the desired conclusion,

$$\chi(r,r')\equiv 0.$$

The homogeneous version of (5) has no nontrivial solution. Hence we may conclude that if (5) has a solution it must be unique. If, furthermore, $f(r,r')$ is bounded for all finite r and r', as it surely is if (37) holds, then (5) is a Fredholm equation. It then follows from the nonexistence of a solution of the homogeneous equation that (5) indeed has a solution.

The remarks based on (37) imply at the same time that the differentiations under the integral sign performed in Sec. II are legitimate, so that the conclusions (8) and (10) are correct.

B. Existence and Nonuniqueness of Solution to Eq. (22)

We want to construct an inverse of the matrix M whose elements are given by (18). In order to do that, we introduce an auxiliary matrix N whose elements are

$$N_{ll'}=\begin{cases}[l'^2-l^2]^{-1}, & \text{for } l'-l \text{ odd},\\ 0, & \text{otherwise.}\end{cases} \tag{38}$$

An inverse of this infinite matrix can be explicitly given. In Appendix I it is proved that

$$(N^{-1})_{ll'}=\begin{cases}\dfrac{16}{\pi^2}\dfrac{l'^2}{l'^2-l^2} & \text{for } l \text{ even}, l' \text{ odd}, l\neq 0,\\ \dfrac{16}{\pi^2}\dfrac{l^2}{l'^2-l^2} & \text{for } l \text{ odd}, l' \text{ even}, l'\neq 0,\\ \dfrac{8}{\pi^2} & \text{for } l=0, l' \text{ odd},\\ -\dfrac{8}{\pi^2} & \text{for } l'=0, l \text{ odd},\\ 0 & \text{otherwise}\end{cases} \tag{39}$$

is a two-sided inverse of N. We use it to construct an inverse of the matrix

$$\mathfrak{M}_{ll'}\equiv\begin{cases}\dfrac{2l+1}{2l}M_{ll'} & \text{for } l>0,\\ M_{0l'} & \text{for } l=0,\end{cases} \tag{40}$$

by writing

$$\mathfrak{M}N^{-1}=1-(1/\pi^2)\mathfrak{M}' \tag{41}$$

where

$$\mathfrak{M}'\equiv\pi^2(N-\mathfrak{M})N^{-1}. \tag{42}$$

We then have explicitly, for example, for $l'\neq 0$, even, and $l''\neq 0$, even:

$$\mathfrak{M}_{l'l''}=\frac{8}{l'}\sum_{l \text{ odd}}\frac{l^2}{(l+l')(l+l'+1)(l''^2-l^2)}$$

and hence

$$|\mathfrak{M}_{l'l''}|\leqslant\frac{8}{l'}\sum_{l \text{ odd}}\frac{1}{(l+l')|l-l''|}.$$

By splitting the series into one part for $l<l''$, and another for $l>l''$, we find that[10]

$$|\mathfrak{M}_{l'l''}|\leqslant C\frac{\log(2+l')\log(2+l'')}{(1+l')^2}. \tag{43}$$

The same inequality is shown similarly to hold for all values of l' and l''.

We now construct

$$[1-(1/\pi^2)\mathfrak{M}']^{-1}=1+(1/\pi^2\Delta)Y \tag{44}$$

by the Fredholm method,[12]

$$Y=\sum_{n=1}^{\infty}\pi^{-2n}Y^{(n)}$$

$$\Delta=1-\sum_{n=1}^{\infty}\frac{1}{n\pi^{2n}}\operatorname{Tr}Y^{(n-1)} \tag{45}$$

[12] See, for example, R. G. Newton, J. Math. Phys. **2**, 188 (1961).

with

$$Y_{\alpha\beta}{}^{(n)}=\frac{(-)^n}{n!}\sum_{\alpha_1,\cdots,\alpha_n}\begin{vmatrix}\mathfrak{M}_{\alpha\beta}{}' & \mathfrak{M}_{\alpha\alpha_1}{}' & \cdots & \mathfrak{M}_{\alpha\alpha_n}{}'\\ \mathfrak{M}_{\alpha_1\beta}{}' & & & \vdots\\ \vdots & & & \\ \mathfrak{M}_{\alpha_n\beta}{}' & & \cdots & \mathfrak{M}_{\alpha_n\alpha_n}{}'\end{vmatrix}. \quad (46)$$

Using the inequality (43) and Hadamard's Lemma,[13] we get

$$|Y_{\alpha\beta}{}^{(n)}|\leqslant\frac{n^{\frac{1}{2}n}}{n!}\frac{\log(2+\alpha)}{(1+\alpha)^2}\log(2+\beta)\times\left\{C\sum_{\gamma=0}^{\infty}\frac{[\log(2+\gamma)]^2}{(1+\gamma)^2}\right\}^n.$$

The use of Stirling's formula for $n!$ then shows that

$$|Y_{\alpha\beta}{}^{(n)}|\leqslant\frac{C}{n^{\frac{1}{2}}}\exp[-\tfrac{1}{2}n\log n]\frac{\log(2+\alpha)\log(2+\beta)}{(1+\alpha)^2}.$$

It is therefore readily seen that both $Y_{\alpha\beta}$ and Δ are finite and furthermore

$$|Y_{\alpha\beta}|\leqslant C\frac{\log(2+\alpha)\log(2+\beta)}{(1+\alpha)^2}. \quad (47)$$

Consequently we have according to (41)

$$(\mathfrak{M}N^{-1})[1+(1/\pi^2\Delta)Y]=1.$$

Moreover, it follows from (47), (39), and (40) that the order in which the two series are summed is irrelevant, so that the product is associative:

$$\mathfrak{M}\{N^{-1}[1+(1/\pi^2\Delta)Y]\}=1. \quad (48)$$

Equation (40) finally gives us a right inverse of M:

$$(M^{-1})_{l'l}=\begin{cases}\left[N^{-1}\left(1+\frac{1}{\pi^2\Delta}Y\right)\right]_{l'l}\frac{2l+1}{2l} & \text{for } l>0,\\ \left[N^{-1}\left(1+\frac{1}{\pi^2\Delta}Y\right)\right]_{l'0} & \text{for } l=0.\end{cases} \quad (49)$$

The matrix M being antisymmetric, we obtain a left inverse from (49) by transposition and multiplication by -1. Thus we have shown that left and right inverses of M exist; but are they unique?

It is proved in Appendix I that

$$\hat{o}N=N\hat{o}=0, \quad (50)$$

where

$$(\hat{o})_l=\begin{cases}1 & \text{for } l\geqslant 2, \text{ even},\\ \frac{1}{2} & \text{for } l=0,\\ 0 & \text{for } l \text{ odd}.\end{cases} \quad (51)$$

[13] See, for example, E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (The MacMillan Company, New York, 1948), Chap. XI.

Now by (41),[14]

$$\mathfrak{M}=[1-(1/\pi^2)\mathfrak{M}']N$$

and hence

$$[1+(1/\pi^2\Delta)Y]\mathfrak{M}=N, \quad (52)$$

the necessary associativity being a simple consequence of (43), (47), and (38). Left multiplication of (52) by $\hat{o}$ yields by (50)

$$\hat{o}\{[1+(1/\pi^2\Delta)Y]\mathfrak{M}\}=0.$$

Again the product is easily seen to be associative, so that

$$M\hat{o}'-\hat{o}'M=0,$$

where

$$(\hat{o}')_l=\begin{cases}\left(1+\frac{1}{2\pi^2\Delta}Y_{0l}+\frac{1}{\pi^2\Delta}\sum_{l'>0,\text{ even}}Y_{l'l}\right)\frac{2l+1}{2l}, & \\ \qquad\qquad\qquad\qquad \text{for } l\geqslant 2, \text{ even},\\ \frac{1}{2}+\frac{1}{2\pi^2\Delta}Y_{00}+\frac{1}{\pi^2\Delta}\sum_{l'>0,\text{ even}}Y_{l'0}, & \text{for } l=0,\\ 0, & \text{otherwise}.\end{cases} \quad (53)$$

since Y connects only l values of equal parity. Insertion of the inequality (47) in (53) shows that

$$|(\hat{o}')_l|\leqslant C\log(2+l). \quad (54)$$

IV. EXAMPLES

Simple examples may be generated by setting all but one of the coefficients c_l in (1) equal to zero. Suppose that $c_l=0$ for all $l\neq L$. Then we get from (12)

$$\varphi_L(r)=\frac{u_L(r)}{1+c_LL_{LL}(r)},$$

$$\varphi_l(r)=u_l(r)-\frac{c_LL_{lL}(r)u_L(r)}{1+c_LL_{LL}(r)}, \quad \text{if } l\neq L,$$

and therefore, by (11) and (7),

$$V(r)=-\frac{2}{r}c_L\frac{d}{dr}\left\{\frac{1}{r}\frac{[u_L(r)]^2}{1+c_LL_{LL}(r)}\right\}.$$

A single coefficient c_L, however, generates infinitely many phase shifts. By (22) we have

$$\sin\delta_l=0 \quad \text{for} \quad l-L \text{ even}.$$

For $l-L$ odd we get

$$\tan\delta_l=b_L/[(L-l)(l+L+1)]$$

[14] We cannot conclude from this, by multiplying by $\hat{o}$ on the right, that $\mathfrak{M}\hat{o}=0$. The product is not associative in this case.

and by (23) and (20)

$$A_L=1-\tfrac{1}{2}\pi\frac{b_L}{2L+1}=1-\tfrac{1}{2}\pi\frac{c_LA_L}{2L+1}$$

so that

$$A_L=1\Big/\left(1+\tfrac{1}{2}\pi\frac{c_L}{2L+1}\right)$$

and consequently

$$\tan\delta_l=\frac{c_L}{1+\frac{1}{2}\pi c_L/(2L+1)}\cdot\frac{1}{(L-l)(L+l+1)}.$$

Similarly, we can solve the case in which $c_l=0$ for $l\geqslant L$. If we call $\mathfrak{L}(r)$ the $L\times L$ matrix whose elements are $L_{ll'}(r)$ for $l<L$, $l'<L, C$, the diagonal matrix whose diagonal elements are c_l for $l<L$; and $\varphi(r)$ and $u(r)$, the $L\times 1$ column matrices whose elements are $\varphi_l(r)$ and $u_l(r)$ for $l<L$, respectively, then (12) is solved by

$$\varphi(r)=[1+\mathfrak{L}(r)C]^{-1}u(r)$$

while the other radial wave functions are, for $l\geqslant L$,

$$\varphi_l(r)=u_l(r)-\sum_{l'=0}^{L-1}L_{ll'}(r)c_{l'}\varphi_{l'}(r).$$

The potential is

$$V(r)=-\frac{2}{r}\frac{d}{dr}\left\{\frac{1}{r}u(r)[1+C\mathfrak{L}(r)]^{-1}Cu(r)\right\}.$$

The phase shifts are best obtained from (19). For $l<L$ we have

$$A_le^{i\delta l}=\sum_{l'=0}^{L-1}[(1+\tfrac{1}{2}\pi C'-iMC)^{-1}]_{ll'},$$

where C' is the diagonal $L\times L$ matrix whose diagonal elements are $c_l/(2l+1)$; for $l\geqslant L$,

$$A_l\,e^{i\delta l}=\frac{1+i\sum_{l'=0}^{L-1}M_{ll'}c_{l'}A_{l'}e^{i\delta l'}}{1+\frac{1}{2}\pi c_l/(2l+1)}.$$

ACKNOWLEDGMENT

The author has enjoyed some stimulating discussions with Dr. J. Bahcall and Dr. M. Wellner.

APPENDIX I

We want to prove here that (39) is a two-sided inverse of the auxiliary matrix N given by (38). For l even, l' odd, we write $l=2r$, $l'=2n+1$ and get

$$N_{ll'}=\tfrac{1}{4}S_{rn},$$

where

$$S_{rn}=1/(n+r+\tfrac{1}{2})(n-r+\tfrac{1}{2});$$

for l odd, l' even, we write $l=2n+1$, $l'=2r$ and get

$$N_{ll'}=-\tfrac{1}{4}S_{rn}.$$

A two-sided inverse of S is given by

$$T_{nr}=\begin{cases}\dfrac{4}{\pi^2}\dfrac{(n+\frac{1}{2})^2}{(n+r+\frac{1}{2})(n-r+\frac{1}{2})}, & \text{for } r\neq 0,\\[2ex] \dfrac{2}{\pi^2}, & \text{for } r=0.\end{cases}$$

Proof: It is a matter of straightforward algebra to show that for $p>0$

$$\pi^2\sum_{n=0}^{\infty}S_{rn}T_{np}=4\frac{s_p-s_r}{p^2-r^2},$$

where

$$s_r=r^2\sum_{n=0}^{\infty}\frac{1}{(n+r+\frac{1}{2})(n-r+\frac{1}{2})}.$$

s_r is easily evaluated, since it can be written

$$s_r=\tfrac{1}{2}r\sum_{n=0}^{\infty}(x_n-x_{n+2r}),$$

where

$$x_n=1/(n-r+\tfrac{1}{2}).$$

We get

$$s_r=\tfrac{1}{2}r(x_0+\cdots+x_{2r-1})$$
$$=\tfrac{1}{2}r[(\tfrac{1}{2}-r)^{-1}+\cdots+(r-\tfrac{1}{2})^{-1}]=0.$$

Consequently,

$$\sum_{n=0}^{\infty}S_{rn}T_{np}=0,$$

unless $r=p$; it clearly holds also for $p=0$. Now take $r=p\neq 0$:

$$\pi^2\sum_{n=0}^{\infty}S_{rn}T_{np}=\sum_{n=0}^{\infty}\left(\frac{1}{n+r+\frac{1}{2}}+\frac{1}{n-r+\frac{1}{2}}\right)^2$$
$$=\sum_{n=r}^{\infty}\frac{1}{(n+\frac{1}{2})^2}+\sum_{n=-r}^{\infty}\frac{1}{(n+\frac{1}{2})^2}+\frac{2}{r^2}s_r$$
$$=2\sum_{n=0}^{\infty}\frac{1}{(n+\frac{1}{2})^2}=\pi^2,$$

since[15] $s_r=0$. Finally, for $r=p=0$,

$$\sum_{n=0}^{\infty}S_{0n}T_{n0}=\frac{2}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(n+\frac{1}{2})^2}=1.$$

This proves that T is a right inverse of S.

[15] See, for example, K. Knopp, *Theorie und Anwendung der Unendlichen Reihen* (Springer-Verlag, Berlin, Germany, 1924), pp. 239–240.

Before proceeding to the proof that it is also a left inverse, we notice that the row matrix

$$\Delta_r=\begin{cases}1 & \text{for } r\geqslant 1,\\ \frac{1}{2} & \text{for } r=0,\end{cases}$$

annihilates S on the left:

$$\sum_{r=0}^{\infty}\Delta_r S_{rn}=-\tfrac{1}{2}\frac{1}{(n+\frac{1}{2})^2}+\tfrac{1}{2}\frac{1}{(n+\frac{1}{2})^2}t_n,$$

$$t_n=\sum_{r=0}^{\infty}\frac{2(n+\frac{1}{2})^2}{(n+r+\frac{1}{2})(n-r+\frac{1}{2})}$$

$$=-(n+\tfrac{1}{2})\sum_{r=0}^{\infty}(y_r-y_{r+2n+1})$$

$$=-(n+\tfrac{1}{2})[(-n-\tfrac{1}{2})^{-1}+\cdots+(n-\tfrac{1}{2})^{-1}]=1,$$

where

$$y_r=1/(r-n-\tfrac{1}{2});$$

hence

$$\sum_{r=0}^{\infty}\Delta_r S_{rn}=0 \quad \text{for all } n.$$

Therefore, T is a left inverse of S if and only if

$$T_{nr}'=T_{nr}-\frac{4}{\pi^2}\Delta_r=\frac{4}{\pi^2}\frac{r^2}{(n+r+\frac{1}{2})(n-r+\frac{1}{2})}$$

is a left inverse.

Now straightforward algebra yields

$$\pi^2\sum_{r=0}^{\infty}T_{pr}'S_{rn}=2\frac{t_n-t_p}{(p-n)(p+n+1)}=0$$

by the above result for t_n, except for $p=n$. For $p=n$ we get

$$\pi^2\sum_{r=0}^{\infty}T_{nr}'S_{rn}=\sum_{r=0}^{\infty}\left(\frac{1}{r-n-\frac{1}{2}}+\frac{1}{n+r+\frac{1}{2}}\right)^2$$

$$=\sum_{r=-n-1}^{\infty}\frac{1}{(r+\frac{1}{2})^2}+\sum_{r=n}^{\infty}\frac{1}{(r+\frac{1}{2})^2}-\frac{t_n}{(n+\frac{1}{2})^2}$$

$$=2\sum_{r=0}^{\infty}\frac{1}{(r+\frac{1}{2})^2}=\pi^2.$$

This completes the proof that T is a two-sided inverse of S and shows at the same time that there exists a row matrix whose elements are Δ_r, which annihilates S from the left. The left inverse of S is therefore not unique.

From T we obtain an inverse of N by setting $l=2n+1$, $l'=2r$ for even l' and odd l; $l=2r$, $l'=2n+1$ for even l and odd l'; for l and l' of equal parity we take $(N^{-1})_{ll'}=0$. That gives us (39). In the same way, we use Δ_r in order to obtain a row (column) matrix that annihilates N from the left (right). The result is (51).

APPENDIX II

Everything done in Sec. II may also be done by starting with the wave functions, assumed known, of an arbitrary comparison potential V_0. In that case

$$D_0(r)\equiv r^2[(\partial^2/\partial r^2)+1-V_0(r)]$$

replaces (3), and the u_l are now regular solutions of the radial Schrödinger equations with V_0, having the same behavior at $r\to 0$ as the spherical Riccatti-Bessel functions. Of course (7) then gives the *addition* to the potential V_0. The only essential change occurs in (14), the second line of which is replaced by

$$u_l(r)\sim A_l^{(0)}\sin(r-\tfrac{1}{2}\pi l+\delta_l^{(0)})\equiv\varphi_{0l}^{(\infty)}(r).$$

Consequently the matrix $L_{ll'}^{(\infty)}$ no longer has the simple value (16), but instead

$$L_{ll'}^{(\infty)}=\frac{A_l^{(0)}A_{l'}^{(0)}\sin[\delta_l^{(0)}-\delta_{l'}^{(0)}+\frac{1}{2}\pi(l'-l)]}{(l'-l)(l'+l+1)} \quad \text{for } l'\neq l$$

$$L_{ll}^{(\infty)}=\int_0^{\infty}dr r^{-2}u_l^2(r).$$

If we write

$$L_{ll'}^{(\infty)}\equiv\mathfrak{M}_{ll'}A_l^{(0)}A_{l'}^{(0)}$$

then (22) and (23) are replaced by

$$\sin(\delta_l^{(0)}-\delta_l)=\textstyle\sum_{l'}\mathfrak{M}_{ll'}b_{l'}\sin[\delta_{l'}-\delta_l+\frac{1}{2}\pi(l-l')]$$

$$A_l/A_l^{(0)}=\cos(\delta_l^{(0)}-\delta_l)-\textstyle\sum_{l'}\mathfrak{M}_{ll'}b_{l'}\times\cos[\delta_{l'}-\delta_l+\frac{1}{2}\pi(l-l')],$$

where

$$b_l\equiv c_l A_l A_l^{(0)}$$

instead of (20). These equations are to be solved in the same manner as (22) and (23).

Multiple Scattering by Arbitrary Configurations in Three Dimensions*

VICTOR TWERSKY
Sylvania Electronic Defense Laboratories, Mountain View, California
(Received July 11, 1961)

This paper provides the extension to three dimensions of our previous treatment of two-dimensional problems. Thus we derive a system of integral equations which specifies the multiple-scattering amplitudes for many objects in terms of corresponding functions for the isolated objects. For arbitrary configurations and large spacings, the amplitudes are expanded as series of single scattered functions and their derivatives; "closed forms" involving differential operators are derived for two objects. For arbitrary spacings, the amplitudes are expanded as series of spherical harmonics to obtain algebraic sets of equations relating the multiple and single scattering coefficients. Series expansions are available for arbitrary configurations, and closed forms are given for two small objects.

1. INTRODUCTION

IN a previous paper[1] we considered multiple scattering of waves by arbitrary configurations of arbitrary scatterers in two dimensions. In the present paper, the results are extended to three dimensions

The literature of scattering by more than one object has been surveyed recently,[2] and much of the initial formalism we will use has been applied previously to related problems.[3] Because of this and because of our relatively detailed treatment of the two-dimensional case, detail will be omitted where feasible; only principal expressions will be given, auxiliary forms being obtainable by inspection of the two-dimensional analogs.[1] Relevant results of Brueckner,[4] Karp and Zitron,[5] Watson,[6] and others will be cited in their appropriate context.

2. ONE SCATTERER

The time-independent, three-dimensional scalar problem of the scattering of a plane wave by an arbitrary object is specified in the external region by a solution of

$$(\nabla^2+k^2)\psi(\mathbf{r})=0, \quad \nabla^2=\partial_x^2+\partial_y^2+\partial_z^2, \quad k=2\pi/\lambda, \quad (1)$$

subject to prescribed boundary conditions on the scatterer's surface. With increasing distance from the scatterer $r\to\infty$, the wave function ψ reduces to a plane wave

$$\psi_i=e^{i\mathbf{k}\cdot\mathbf{r}},$$

$$\mathbf{k}\cdot\mathbf{r}=(k\mathbf{i})\cdot(r\mathbf{o})=k[z\cos\theta_i+\sin\theta_i(x\cos\varphi_i+y\sin\varphi_i)] \quad (2)$$
$$=kr[\cos\theta_i\cos\theta+\sin\theta_i\sin\theta\cos(\varphi-\varphi_i)],$$

and the difference $\psi-\psi_i\equiv u\sim f(\mathbf{o},\mathbf{i})e^{ikr}/r$ is an outgoing spherical wave. Thus we write

$$\psi(\mathbf{r},\mathbf{i})=\psi_i(\mathbf{r})+u(\mathbf{r},\mathbf{i}), \quad (3)$$

where ψ_i represents the incident plane wave and u the corresponding scattered wave. The boundary conditions on ψ and its internal form (as determined by the physical constants of the scatterer) will not be considered explicitly. (For concreteness, we take the origin of coordinates as the center of the sphere circumscribing the scatterer; however, we may also use the midpoint of the axis of symmetry, or the midpoint of the longest line through the scatterer.)

Surface Integral Representation: We apply Green's theorem to the pair of functions $u(\mathbf{r}')$ and

$$ikh_0(k|\mathbf{r}-\mathbf{r}'|)/4\pi,$$

where h_0 is the spherical Hankel function of the first kind,[7] and $\mathbf{r}$ and $\mathbf{r}'$ label a field point and a point on the scatterer's surface (see Fig. 1). Integrating over a

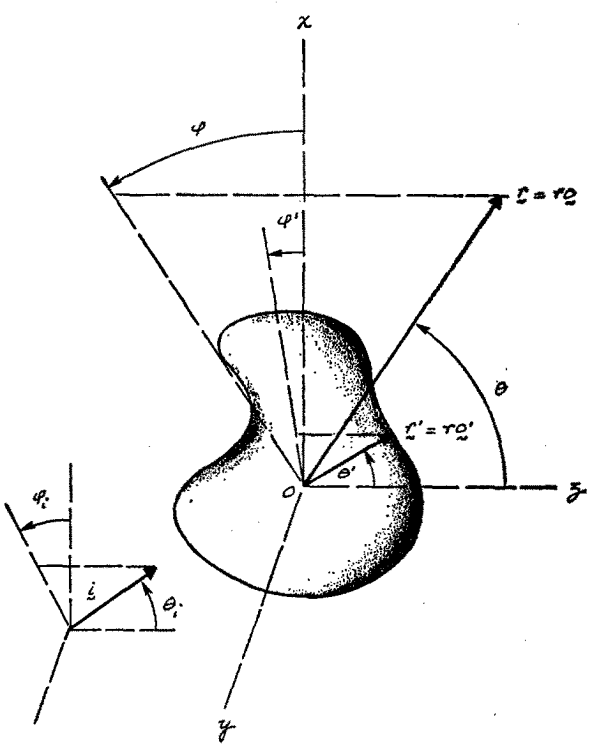


FIG. 1. Coordinates for single-body problem.

* This work was supported in part by Signal Corps contract.

[1] V. Twersky, "On scattering of waves by two objects," Report EDL-E60, Sylvania Electronic Defense Laboratories (1961) (to be published).

[2] V. Twersky, J. Research Natl. Bur. Standards **64D**, 715 (1960).

[3] V. Twersky, J. Acoust. Soc. Am. **29**, 209 (1957).

[4] K. A. Brueckner, Phys. Rev. **89**, 834 (1953).

[5] S. N. Karp and N. Zitron, "Higher Order Approximations in Multiple Scattering," Research Report EM-126, Inst. Math. Sci., New York University (1959).

[6] K. M. Watson, Phys. Rev. **87**, 575 (1953).

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

volume external to the scatterer we obtain

$$u(\mathbf{r},\mathbf{i}) = (k/4\pi i)\oint [h_0(k|\mathbf{r}-\mathbf{r}'|)\partial_n u(\mathbf{r}',\mathbf{i}) - u(\mathbf{r}',\mathbf{i})\partial_n h_0(k|\mathbf{r}-\mathbf{r}'|)]dA(\mathbf{r}') \equiv \{h_0(k|\mathbf{r}-\mathbf{r}'|), u(\mathbf{r}',\mathbf{i})\}, \quad (4)$$

where n is the outward normal. The integral may be taken over any surface inclosing the scatterer and excluding $\mathbf{r}$.

If $kr \gg 1$, $r \gg r'$, then

$$h_0(k|\mathbf{r}-\mathbf{r}'|) \sim h(kr)e^{-ik\mathbf{o}\cdot\mathbf{r}'}, \qquad h(kr) \equiv h_0(kr) = e^{ikr}/ikr, \quad (5)$$

and (4) reduces to the "far-field" form

$$u \sim (e^{ikr}/r)f(\mathbf{o},\mathbf{i}) = h(kr)g(\mathbf{o},\mathbf{i}), \quad (6)$$

where

$$g(\mathbf{o},\mathbf{i}) \equiv \{e^{-ik\mathbf{o}\cdot\mathbf{r}'}, u(\mathbf{r}',\mathbf{i})\}; \quad (7)$$

the integral is over any surface inclosing the scatterer. The "scattering amplitude" $g(\mathbf{o},\mathbf{i}) = ikf(\mathbf{o},\mathbf{i})$, which we assume to be known, indicates the far-field response in the direction of observation $\mathbf{o}$ to plane-wave excitation of direction of incidence $\mathbf{i}$. We find it convenient to use the normalized function g instead of the customary amplitude f; in terms of g, the three- and two-dimensional results have essentially the same form, and corresponding expressions may be obtained by inspection.

Plane Wave Representation: If u is known, then (7) gives g by integration. The inverse relation follows on introducing the plane-wave representation of h_0 in (4). The analog of the form of the Green's function we used for the two-dimensional case[1] may be obtained from Noether.[8] Thus

$$h_0(k|\mathbf{r}-\mathbf{r}'|) = \frac{1}{2\pi}\int \exp[ik\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')]d\Omega_p = \frac{1}{2\pi}\int_{\alpha+i\infty}^{\alpha'-i\infty} d\beta \int_{\gamma+i\infty}^{\gamma'-i\infty} \times \exp[ik\mathbf{p}(\tau,\beta)\cdot(r\mathbf{o}-r'\mathbf{o}')]\sin\tau d\tau, \quad (8)$$

where the limits and paths in the complex plane (each path analogous to one in Sommerfeld's integral for $H_0^{(1)}$ are chosen to insure $\mathrm{Im}\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')>0$; see Noether.[8] If $z-z'>0$, we may use Weyl's limits[8] $\int_{-\pi}^{\pi} d\beta \int_0^{\frac{1}{2}\pi - i\infty} d\tau$. Since the integral in (8) is spherically symmetric we may change the polar axis at will. In particular, if we measure τ from $\mathbf{o}$, and β from the plane determined by $\mathbf{o}$ and $\mathbf{o}'$, then the exponent becomes $ik(r-r'\mathbf{o}\cdot\mathbf{o}')\cos\tau - ikr'(1-(\mathbf{o}\cdot\mathbf{o}')^2)^{\frac{1}{2}}\sin\tau\cos\beta$, and we may use Weyl's limits provided that $r > r'\mathbf{o}\cdot\mathbf{o}'$. Equivalently we may retain the kernel as in (8) and introduce the corresponding dependence of the limits on θ and φ. Substituting (8) and (4), we write

$$u(\mathbf{r},\mathbf{i}) = \frac{1}{2\pi}\int e^{ik\mathbf{p}\cdot\mathbf{r}}\{e^{-ik\mathbf{p}\cdot\mathbf{r}'}, u(\mathbf{r}',\mathbf{i})\}d\Omega_p = \frac{1}{2\pi}\int e^{ik\mathbf{p}\cdot\mathbf{r}} g(\mathbf{p},\mathbf{i})d\Omega_p, \quad (9)$$

where the limits are assumed to have the appropriate dependence on θ and φ to insure convergence for all values of $\mathbf{r}'$ on the scatterer's surface, and $r > (\mathbf{r}'\cdot\mathbf{o})_{\max}$ (i.e., distance r from the "center" of the scatterer to the observation point greater than the scatterer's projection on $\mathbf{r}$). Thus the scattering amplitude g specifies the field completely, at least for all $r > r'_{\max}$.

The limiting form $u \sim hg$ given in (6) is the leading term of a series expansion of u in inverse powers of r which converges for $r > r'_{\max}$; see Sommerfeld,[9] Barrar and Kay,[10] and Wilcox.[11] We derive the series by a different method and express the series coefficients in a somewhat different form, as an incidental step in obtaining a series representation for a class of integrals which also arises in the multiple scattering problem.

Our procedure is based on the following: Starting with the usual expansion of $h_n(r)$ in inverse powers of r (e.g., as in Morse and Feshbach,[7] p. 1466), we write

$$h_n(r)i^n = h(r)\left[1 + n(n+1)\frac{i}{2r} + \frac{n(n+1)[n(n+1)-1\cdot 2]}{2!}\left(\frac{i}{2r}\right)^2 + \cdots \times \frac{n(n+1)[n(n+1)-1\cdot 2]\cdots[n(n+1)-(\nu-1)\nu]}{\nu!}\left(\frac{i}{2r}\right)^{\nu} + \cdots\right], \quad (10)$$

which terminates with $\nu = n$. Using the equation for surface harmonics (reference 7, p. 1264)

$$n(n+1)Y_n = (-1/\sin^2\theta)[\partial_\varphi^2 + \sin\theta\partial_\theta(\sin\theta\partial_\theta)]Y_n \equiv DY_n, \quad Y_n = Y_n(\theta,\varphi) = Y_n(\mathbf{o}), \quad (11)$$

[8] F. Noether, in *Theory of Functions* edited by R. Rothe, F. Ollendorff, and K. Pohlhausen (Technology Press, Cambridge, Massachusetts, 1948), p. 167, Eq. (7).

[9] A. Sommerfeld, *Partial Differential Equations in Physics* (Academic Press, Inc., New York, 1949), p. 192.

[10] R. B. Barrar and A. F. Kay, "A series development of a solution of the wave equation in powers of $1/r$," Technical Research Group, New York City, 1954.

[11] C. H. Wilcox, Proc. Am. Math. Soc. 7, 271 (1956).

we may write

$$h_n(r)i^nY_n(\mathbf{o})=h(r)\left[\sum_{\nu=0}^{\infty}\left(\frac{i}{2r}\right)^{\nu}\frac{D(D-1\cdot 2)(D-2\cdot 3)\cdots(D-[\nu-1]\cdot\nu)}{\nu!}\right]Y_n(\mathbf{o})$$

$$=h(r)\left[1+\frac{i}{2r}D+\left(\frac{i}{2r}\right)^2\frac{D(D-1\cdot 2)}{2!}+\cdots\left(\frac{i}{2r}\right)^n\frac{D(D-1\cdot 2)(D-2\cdot 3)\cdots(D-[n-1]n)}{n!}\right]Y_n\equiv h(r)\mathfrak{D}Y_n. \quad (12)$$

All terms after the last shown in the brackets contain $D-n(n+1)$ as a common factor, and are consequently [from (11)] identically zero.

We may exploit (12) in one of several ways to obtain the required representation for u. In order to further a subsequent development, we proceed by initially expanding g as a series of spherical harmonics:

$$g(\mathbf{p},\mathbf{i})=\sum_{n=0}^{\infty}Y_n=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}a_{nm}(\mathbf{i})Y_n^m(\mathbf{p}),$$
$$Y_n^m(\mathbf{p})=P_n^m(\cos\tau)e^{im\beta}, \quad (13)$$
$$Y_n^{-m}(\mathbf{p})=(-1)^m[(n-m)!/(n+m)!]P_n^m(\cos\tau)e^{-im\beta},$$

where P_n^m is an associated Legendre function. Substituting into (9), and using

$$\frac{1}{2\pi}\int e^{i\mathbf{r}\cdot\mathbf{p}}Y_n^m(\mathbf{p})d\Omega_p=i^nh_n(r)Y_n^m(\mathbf{o})$$
$$=i^nh_n(r)P_n^m(\cos\theta)e^{im\varphi}, \quad (14)$$

[essentially Eq. (15) of Friedman and Russek[12]], we obtain for $r>r'_{\max}$,

$$u=\sum_n\sum_m a_{nm}(\mathbf{i})h_n(kr)Y_n^m(\mathbf{o})i^n. \quad (15)$$

Substituting (12) into (15), and then introducing g of (13), we obtain

$$u=h(kr)\mathfrak{D}\sum_n\sum_m a_{nm}Y_n^m(\mathbf{o})=h(kr)\mathfrak{D}g(\mathbf{o},\mathbf{i})$$
$$=h(kr)\left[g+\frac{i}{2kr}Dg+\left(\frac{i}{2kr}\right)^2\frac{D(D-2)g}{2}+\cdots\right] \quad (16)$$

in terms of the scattering amplitude and its θ and φ derivatives. (This series in $r^{-\nu}$ converges absolutely and uniformly in r, θ, and φ in any region $r\geq r'_{\max}+\epsilon>r'_{\max}$.)[10,11] Thus, in addition to (9), Eq. (16) is another "inverse" of the relation for g in terms of u given in (7).

Since (13) is general in form, the steps (9) to (16) indicate that

$$(1/2\pi)\int\exp(ir\mathbf{p}\cdot\mathbf{o})F(\mathbf{p})d\Omega_p=h(r)\mathfrak{D}F(\mathbf{o}), \quad (17)$$

where the limits, as for (8) and (9), insure convergence of the integral and where F is representable as a series of surface harmonics. [In a subsequent section, we use (17) for more general scattering situations.] This same result may be obtained by a saddle point procedure. Equivalently, more or less as for the analogous two-dimensional case,[1] (17) corresponds to introducing a symbolic Taylor's series in the integral and using (14) and (10):

$$(1/2\pi)\int\exp(ir\mathbf{o}\cdot\mathbf{p})[P_n(\mathbf{o}\cdot\mathbf{p})]_DF(\mathbf{o})d\Omega$$
$$=[i^nh_n(r)]_DF(\mathbf{o})=h(r)\mathfrak{D}F(\mathbf{o}).$$

Here the first equality corresponds to (14) for $m=\theta=0$, and $[\ ]_D$ indicates the replacement of $n(n+1)$ by D in (10) and in

$$P_n(x)=1+\frac{n(n+1)}{(1!)^2}\left(\frac{x-1}{2}\right)$$
$$+\frac{n(n+1)[n(n+1)-1\cdot 2]}{(2!)^2}\left(\frac{x-1}{2}\right)^2+\cdots$$

[as follows on writing P_n as the hypergeometric function $F(n+1,-n;1;\frac{1}{2}-\frac{1}{2}x)$].

Special Function Representation: We have written g as a series of spherical harmonics (13) and used it in (9) to obtain (15) for u. These series also follow from (7) and (14) by using the appropriate series representations for $\exp(-ik\mathbf{o}\cdot\mathbf{r}')$ and $h_0(k|\mathbf{r}-\mathbf{r}'|)$. Thus since

$$e^{-i\mathbf{k}\cdot\mathbf{r}'}=\sum_{n=0}^{\infty}(2n+1)j_n(kr')P_n(-\mathbf{o}\cdot\mathbf{o}'),$$
$$P_n(-\mathbf{o}\cdot\mathbf{o}')=(-1)^nP_n(\mathbf{o}\cdot\mathbf{o}') \quad (18)$$
$$=(-1)^n\sum_{m=-n}^{n}(-1)^mY_n^m(\mathbf{o})Y_n^{-m}(\mathbf{o}'),$$

(reference 7, p. 1466) we obtain from (7),

$$g(\mathbf{o},\mathbf{i})=\sum_n\sum_m a_{nm}(\mathbf{i})Y_n^m(\mathbf{o}),$$
$$a_{nm}(\mathbf{i})=i^{-n}(2n+1)(-1)^m\{j_n(kr')Y_n^{-m}(\mathbf{o}'),u(\mathbf{r}',\mathbf{i})\}, \quad (19)$$

which provides a surface integral representation for the "scattering coefficients" a_{nm}. Similarly, if we substitute (reference 7, p. 1466)

$$h_0(k|\mathbf{r}-\mathbf{r}'|)=\sum_{n=0}^{\infty}(2n+1)j_n(kr')h_n(kr)P_n(\mathbf{o}'\cdot\mathbf{o}),$$
$$r>r' \quad (20)$$

[12] B. Friedman and J. Russek, Quart. Appl. Math. 12, 13 (1954).

into (4), we again obtain (15), which provides the "inverse" to a_{nm} in terms of u given in (19).

In general, if we know the scattering amplitude g, then we can obtain the scattering coefficients a and consequently determine u; this procedure is particularly useful when the series in (19) is rapidly convergent ("small scatterers"). Alternatively, if the series for g is not rapidly convergent ("large scatterers," in general), then we represent u in terms of g and its derivatives by means of (16).

For the special case of a homogeneous spherical scatterer (radius $r'=$constant) we have

$$a_{nm}(\mathbf{i})=a_n(2n+1)(-1)^m Y_n^{-m}(\mathbf{i}) \tag{21}$$

where the a_n are known quotients involving the radial functions, e.g., $a_n=-j_n(kr')/h_n(kr')$ if $\psi=0$ at the scatterer.

Scattering Theorems: The normalization of our scattering amplitude was chosen so that for lossless scatterers (e.g., ψ vanishing at the surface) the usual theorem for the forward amplitude reads

$$-\mathrm{Re}g(\mathbf{i},\mathbf{i})=\frac{1}{4\pi}\int|g(\mathbf{o},\mathbf{i})|^2 d\Omega_0=\frac{k^2}{4\pi}p, \tag{22}$$

where (as for the two-dimensional case[1]) the middle form is the average of $|g|^2$ over all angles of observation, and p is the total-scattering cross section. For other angles of observation, we have

$$-g(\mathbf{j},\mathbf{i})-g^*(\mathbf{i},\mathbf{j})=(1/2\pi)\int g^*(\mathbf{o},\mathbf{j})g(\mathbf{o},\mathbf{i})d\Omega_0. \tag{23}$$

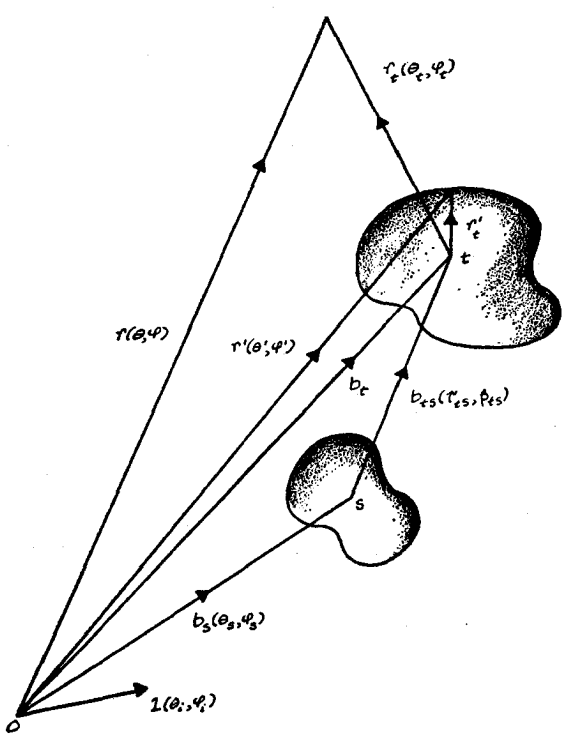


FIG. 2. Coordinates for many-body problem.

Applying this result to spheres [(19) with a_{nm} as in (21)], yields

$$-\mathrm{Re}a_n=|a_n|^2. \tag{24}$$

More generally, if loss is present, the total energy cross section $p_t=p+p_a$ (radiation plus absorption losses) is given by

$$-\mathrm{Re}g(\mathbf{i},\mathbf{i})=\frac{k^2}{4\pi}p_t=\frac{k^2}{4\pi}p_a+\frac{1}{4\pi}\int|g(\mathbf{o},\mathbf{i})|^2 d\Omega. \tag{25}$$

In addition, for the general case, the reciprocity principle gives

$$g(\mathbf{j},\mathbf{i})=g(-\mathbf{i},-\mathbf{j}). \tag{26}$$

3. MANY SCATTERERS

For many scatterers in the geometry of Fig. 2, we write the field as

$$\Psi=\psi_i+\mathfrak{U},\quad \mathfrak{U}\sim h(kr)\mathfrak{G}(\mathbf{o},\mathbf{i}), \tag{27}$$

where $\mathfrak{U}$ and $\mathfrak{G}$ have the forms (4) and (7) with u replaced by $\mathfrak{U}$. The "compound amplitude" $\mathfrak{G}$ fulfills the same theorems as g [i.e., (25), (26), etc. hold with g replaced by $\mathfrak{G}$].

Proceeding as in reference 1, we express the total scattered field as

$$\mathfrak{U}=\sum U_s(\mathbf{r}-\mathbf{b}_s)\exp(i\mathbf{k}\cdot\mathbf{b}_s),$$

$$U_s=\{h_0(k|\mathbf{r}_s-\mathbf{r}_s'|),\ U_s(\mathbf{r}')\}, \tag{28}$$

$$U_s\sim h(kr_s)\{e^{-ik\mathbf{o}_s\cdot\mathbf{r}_s'},\ U_s(\mathbf{r}_s')\}\equiv h(kr_s)G_s(\mathbf{o},\mathbf{i}), \tag{29}$$

where G_s, the "multiple scattered amplitude" of scatterer s, reduces to the single scattered function g_s as the others recede to infinity. In terms of G_s, the compound amplitude equals

$$\mathfrak{G}(\mathbf{o},\mathbf{i})=\sum_s\exp[ik(\mathbf{i}-\mathbf{o})\cdot\mathbf{b}_s]G_s(\mathbf{o},\mathbf{i}). \tag{30}$$

Integral Equations: Substituting (8) into U_s of (28) and using G_s defined in (29) yields

$$U_s=\int\exp(ik\mathbf{p}\cdot\mathbf{r}_s)G_s(\mathbf{p},\mathbf{i})d\Omega_p/2\pi, \tag{31}$$

$$\mathfrak{U}=\sum_s\exp(i\mathbf{k}\cdot\mathbf{b}_s)\int\exp[ik\mathbf{p}\cdot(\mathbf{r}-\mathbf{b}_s)]\times G_s(\mathbf{p},\mathbf{i})d\Omega_p/2\pi. \tag{32}$$

Proceeding as in reference 1, we express ψ_i and $\mathfrak{U}$ in the local coordinates of scatterer t, and write the total field as a set of plane waves plus one outgoing wave U_t. Then using the superposition principle, we obtain

$$U_t=u_t(\mathbf{i})+\sum_s{}'\int\exp[ik(\mathbf{b}_t-\mathbf{b}_s)\cdot(\mathbf{p}-\mathbf{i})]\times u_t(\mathbf{p})G_s(\mathbf{p},\mathbf{i})d\Omega_p/2\pi, \tag{33}$$

where u_t is the single scattered value. (The prime on the sum means $s\neq t$.)

Similarly, we obtain the "self-consistent" system of

integral equations

$$G_t(\mathbf{o},\mathbf{i})=g_t(\mathbf{o},\mathbf{i})+\sum_s{}'\int \exp[ik(\mathbf{b}_t-\mathbf{b}_s)\cdot(\mathbf{p}-\mathbf{i})]\times g_t(\mathbf{o},\mathbf{p})G_s(\mathbf{p},\mathbf{i})d\Omega_p/2\pi. \quad (34)$$

From (26) for g_t, and (7) for $g_t(-\mathbf{p},-\mathbf{o})$ and $G_s(\mathbf{p},\mathbf{i})$, the integral converges if

$$\mathrm{Im}\mathbf{p}\cdot(\mathbf{b}_t-\mathbf{b}_s+\mathbf{r}_t'-\mathbf{r}_s')_{\max}>0.$$

Thus in terms of $\mathbf{b}_{ts}=\mathbf{b}_t-\mathbf{b}_s=b_{ts}\hat{b}_{ts}$ we require

$$b_{ts}>[(\mathbf{r}_t'+\mathbf{r}_s')\cdot\hat{b}_{ts}]_{\max},$$

i.e., that the separation of scatterer centers (b_{ts}) be greater than the sum of the scatterer's projections on $\mathbf{b}_{ts}$.

Large spacings: We obtain forms of (34) convenient for large $k|\mathbf{b}_t-\mathbf{b}_s|=kb_{ts}$ by applying (17). Thus

$$G_t(\mathbf{o},\mathbf{i})=g_t(\mathbf{o},\mathbf{i})+\sum_s{}'\mathfrak{F}_{ts}g_t(\mathbf{o},\hat{b}_{ts})G_s(\hat{b}_{ts},\mathbf{i}),$$
$$\mathfrak{F}_{ts}=h(kb_{ts})\exp[-i\mathbf{k}\cdot\mathbf{b}_{ts}]\mathfrak{D}_{ts}$$
$$\equiv\mathfrak{K}_{ts}(b^{-1})+\mathfrak{M}_{ts}(b^{-2})+\mathfrak{N}_{ts}(b^{-3})+\cdots, \quad (35)$$

where the subscripts on $\mathfrak{D}_{ts}$ indicate that the differentiations [as in (11) and (12)] are to be performed with respect to the angles associated with the unit vector $\hat{b}_{ts}$. [If we keep only the leading term of $\mathfrak{D}$ (i.e., unity) then (35) reduces to a system of inhomogeneous algebraic equations for the amplitudes for the case of "far-field multiple scattering" (i.e., each scatterer in the far field of all others); see Karp,[13] and Twersky.[14]]

The leading term of (35) (which is independent of kb) is the single scattered value, or equivalently the "first order" of scattering:

$$g_t(\mathbf{o},\mathbf{i}). \quad (36)$$

Iterating (35) starting with (36) yields a series in inverse powers of kb_{ts} which involves g and its derivatives. Thus the $(kb)^{-1}$ term is the far-field multiple scattering form of the second order of scattering:

$$\sum_s{}'\mathfrak{K}_{ts}g_t(\mathbf{o},\hat{b}_{ts})g_s(\hat{b}_{ts},\mathbf{i}),$$
$$\mathfrak{K}_{ts}\equiv h(kb_{ts})\exp[-i\mathbf{k}\cdot\mathbf{b}_{ts}]=\frac{\exp[ikb_{ts}-i\mathbf{k}\cdot\mathbf{b}_{ts}]}{ikb_{ts}}. \quad (37)$$

Terms to order $(kb)^{-2}$ are given by

$$\sum_s{}'\mathfrak{K}_{ts}g_t(\mathbf{o},\hat{b}_{ts})\sum_p{}'\mathfrak{K}_{sp}g_s(\hat{b}_{ts},\hat{b}_{sp})g_p(\hat{b}_{sp},\mathbf{i})$$
$$+\sum_s\mathfrak{M}_{ts}g_t(\mathbf{o},\hat{b}_{ts})g_s(\hat{b}_{ts},\mathbf{i}),$$
$$\mathfrak{M}_{ts}\equiv\mathfrak{K}_{ts}(i/2kb_{ts})D_{ts}, \quad (38)$$
$$D_{ts}=\frac{-1}{\sin^2\tau_{ts}}\left[\frac{\partial^2}{\partial\beta_{ts}^2}+\sin\tau_{ts}\frac{\partial}{\partial\tau_{ts}}\left(\sin\tau_{ts}\frac{\partial}{\partial\tau_{ts}}\right)\right],$$

where the double sum corresponds to the third far-field order, and the single sum is the first "mid-field" correction to the second far-field order. (Note that no derivatives of the exponentials in $\mathfrak{K}$ are involved: these would introduce positive powers of kb.) The next terms in the expansion of G, the terms of order $(kb)^{-3}$, are given by

$$\sum_s{}'\mathfrak{K}_{ts}g_t(\mathbf{o},\hat{b}_{ts})\sum_p{}'\mathfrak{K}_{sp}g_s(\hat{b}_{ts},\hat{b}_{sp})\times\sum_q{}'\mathfrak{K}_{pq}g_p(\hat{b}_{sp},\mathbf{b}_{pq})g_q(\hat{b}_{pq},\mathbf{i})$$
$$+\sum_s{}'\begin{bmatrix}\mathfrak{K}\\ \mathfrak{M}\end{bmatrix}_{ts}g_t(\mathbf{o},\hat{b}_{ts})\sum_p{}'\begin{bmatrix}\mathfrak{M}\\ \mathfrak{K}\end{bmatrix}_{sp}g_s(\hat{b}_{ts},\hat{b}_{sp})g_p(\hat{b}_{sp},\mathbf{i})$$
$$+\sum_s{}'\mathfrak{N}_{ts}g_t(\mathbf{o},\hat{b}_{ts})g_s(\hat{b}_{ts},\mathbf{i}),$$
$$\mathfrak{N}_{ts}\equiv\mathfrak{K}_{ts}(i/2kb_{ts})^2\tfrac{1}{2}D_{ts}(D_{ts}-2), \quad (39)$$

where the triple sum is the fourth far-field order, the two double sums (i.e., terms containing either $\mathfrak{K}_{ts}\mathfrak{M}_{sp}$ or $\mathfrak{M}_{ts}\mathfrak{K}_{sp}$) are the first corrections to the third far-field order, and the single sum is the second correction to the second far-field order. Thus we have

$$G_t(\mathbf{o},\mathbf{i})=(36)+(37)+(38)+(39)+\mathcal{O}[(kb)^{-4}]. \quad (40)$$

Terms to order $(kb)^{-2}$ for two scatterers were obtained originally by Karp and Zitron.[5]

Algebraic Equations: If we substitute spherical harmonic representations analogous to (13) for g and G in (34), i.e.,

$$g_t(\mathbf{o},\mathbf{i})=\sum_n\sum_m a^t{}_{nm}(\mathbf{i})Y_n{}^m(\mathbf{o}),$$

and

$$G_t(\mathbf{o},\mathbf{i})=\sum\sum A^t{}_{nm}(\mathbf{i})Y_n{}^m(\mathbf{o}),$$

and also expand the scattering coefficients for the isolated body as a series of spherical harmonics in the angles of $\mathbf{i}$,

$$a^t{}_{nm}(\mathbf{i})=\sum_\nu\sum_\mu b^t{}_{nm\nu\mu}Y_\nu{}^{-\mu}(\mathbf{i}), \quad (41)$$

then

$$A^t{}_{nm}(\mathbf{i})=a^t{}_{nm}(\mathbf{i})+\sum_s{}'\sum_\nu\sum_\mu\sum_r\sum_q b^t{}_{nm\nu\mu}A^s{}_{rq}E(st;\nu\mu,rq),$$
$$E(st;\nu\mu,rq)=e^{-i\mathbf{k}\cdot\mathbf{b}_{ts}}\int e^{-ik\mathbf{p}\cdot\mathbf{b}_{ts}}Y_\nu{}^{-\mu}(\mathbf{p})Y_r{}^q(\mathbf{p})d\Omega/2\pi. \quad (42)$$

To reduce E, we use

$$Y_\nu{}^{-\mu}Y_r{}^q=\sum_{m=|\nu-r|}^{\nu+r}d_m(\mu\nu;qr)Y_m{}^{q-\mu},$$

and (14). Thus

$$E(st;\nu\mu,rq)=\exp[-i\mathbf{k}\cdot\mathbf{b}_{ts}]\times\sum_{m=|\nu-r|}^{\nu+r}d_m(\mu\nu;qr)i^m h_m(kb_{ts})Y_m{}^{q-\mu}(\hat{b}_{ts}); \quad (43)$$

[13] S. N. Karp, "Diffraction by combinations of obstacles," Proc. McGill Symp. Microwave Optics (1953), p. 198.

[14] V. Twersky, J. Appl. Phys. 23, 407 (1952); additional results are given in Research Report EM-34, Institute of Mathematical Sciences, New York University, 1951.

see Friedman and Russek,[12] and others, for the numbers d_m.

In particular, for spheres, (42) reduces to

$$A^t{}_{nm}=(-1)^m(2n+1)a_n{}^t \times[Y_n{}^{-m}(\mathbf{i})+\sum_s{}'\sum_r\sum_q A^s{}_{rq}E(st;nm,rq]. \quad (44)$$

The special case of (44) for a periodic lattice is considered by Morse.[15]

In general, for arbitrary configurations and arbitrary scatterers, we expand (42) and (44) by Neumann iteration to obtain orders of scattering series for the coefficients A in terms of powers of a; see analogous series in reference 1. For two small scatterers, we obtain closed form approximations in the next section.

4. TWO SCATTERERS

For two objects, we take the primary origin ($r=0$) as the midpoint of the line joining the scatterer's origins (say the centers of their circumscribed spheres). The centers are located at

$$\mathbf{b}_1=\mathbf{b}=b\mathbf{b}_+(\tau_+,\beta_+),$$
$$\mathbf{b}_2=-\mathbf{b}=b\mathbf{b}_-(\tau_-,\beta_-)=b\mathbf{b}_-(\pi+\tau_+,\pi+\beta_+),$$

and the local coordinates with respect to these centers are $\mathbf{r}_1=\mathbf{r}_+(r_+,\theta_+,\varphi_+)$, $\mathbf{r}_2=\mathbf{r}_-(r_-,\theta_-,\varphi_-)$. For this case the total scattered field of (28) reduces to

$$\mathfrak{U}(\mathbf{r})=e^{i\delta}U_+(\mathbf{r}_+)+e^{-i\delta}U_-(\mathbf{r}_-),$$
$$\pm\delta\equiv b\mathbf{k}\cdot\mathbf{b}_\pm=kb\mathbf{i}\cdot\mathbf{b}_\pm=\pm\mathbf{k}\cdot\mathbf{b}, \quad (45)$$

and the compound scattering amplitude equals

$$\mathcal{G}(\mathbf{o},\mathbf{i})=e^{i(\delta-\Delta)}G_+(\mathbf{o},\mathbf{i})+e^{-i(\delta-\Delta)}G_-(\mathbf{o},\mathbf{i}),$$
$$\pm\Delta\equiv kb\mathbf{o}\cdot\mathbf{b}_\pm=\pm k\mathbf{o}\cdot\mathbf{b}. \quad (46)$$

The plane wave representation (31) yields

$$U_\pm=\int \exp[ik\mathbf{p}\cdot\mathbf{r}_\pm]G_\pm(\mathbf{p},\mathbf{i})d\Omega_p/2\pi, \quad (47)$$

where, from (34), the G's are determined by the integral equations

$$G_\pm(\mathbf{o},\mathbf{i})=g_\pm(\mathbf{o},\mathbf{i})+\exp[-i2kb\mathbf{i}\cdot\mathbf{b}_\pm] \times\int \exp[i2kb\mathbf{p}\cdot\mathbf{b}_\pm]g_\pm(\mathbf{o},\mathbf{p})G_\mp(\mathbf{p},\mathbf{i})d\Omega_p/2\pi. \quad (48)$$

4.1 Large Spacings

For two scatterers, (35) reduces to

$$G_\pm(\mathbf{o},\mathbf{i})=g_\pm(\mathbf{o},\mathbf{i})+\mathcal{F}_\pm g_\pm(\mathbf{o},\mathbf{b}_\pm)G_\mp(\mathbf{b}_\pm,\mathbf{i})$$
$$\mathcal{F}_\pm=h(2kb)\exp[-i2kb\mathbf{i}\cdot\mathbf{b}_\pm]\mathfrak{D}_\pm \equiv\mathcal{K}_\pm(b^{-1})+\mathfrak{M}_\pm(b^{-2})+\mathfrak{N}_\pm(b^{-3})+\cdots, \quad (49)$$

where the subscripts on $\mathfrak{D}_\pm$, etc., indicate that the differentiations [as in (11) and (12)] are to be performed with respect to the angles associated with the unit vectors $\mathbf{b}_\pm$. Proceeding essentially as in reference 1, we replace $\mathbf{o}$ by $\mathbf{b}_\mp$ in (49) to obtain two equations relating $G_\mp(\mathbf{b}_\pm,i)$; we "solve" these symbolically in terms of $g_\pm$, and substitute the results into (49) to obtain the closed operator form

$$G_\pm(\mathbf{o},\mathbf{i})=g_\pm(\mathbf{o},\mathbf{i})+\mathcal{F}_\pm g_\pm(\mathbf{o},\mathbf{b}_\pm) \times\left[\frac{g_\mp(\mathbf{b}_\pm,\mathbf{i})+\mathcal{F}_\mp g_\mp(\mathbf{b}_\pm,\mathbf{b}_\mp)g_\pm(\mathbf{b}_\mp,\mathbf{i})}{1-\mathcal{F}_\mp g_\mp(\mathbf{b}_\pm,\mathbf{b}_\mp)\mathcal{F}_\pm g_\pm(\mathbf{b}_\mp,\mathbf{b}_\pm)}\right]. \quad (50)$$

Here it is understood that within the brackets the expanded denominator operates on the numerator from the left, and that we work from right to left in performing the $\mathcal{F}$ operations in the generated chains $\cdots\mathcal{F}g\mathcal{F}g\mathcal{F}gg$. (If we replace $\mathcal{F}$ by $\mathcal{K}$ we obtain the far-field multiple scattering closed form, which includes the leading term of each order of scattering for large separation; see Karp,[13] and Twersky.[14])

Expanding the closed form (50) in powers of b^{-1} yields the analogs of the results (36) to (40) obtained by iterating the system of equations (35). Thus terms to order b^{-1} are given by

$$g_\pm(\mathbf{o},\mathbf{i})+\mathcal{K}_\pm g_\pm(\mathbf{o},\mathbf{b}_\pm)g_\mp(\mathbf{b}_\pm,\mathbf{i}). \quad (51)$$

The b^{-2} terms equal

$$\mathcal{K}_\pm g_\pm(\mathbf{o},\mathbf{b}_\pm)\mathcal{K}_\mp g_\mp(\mathbf{b}_\pm,\mathbf{b}_\mp)g_\pm(\mathbf{b}_\mp,\mathbf{i}) +\mathfrak{M}_\pm g_\pm(\mathbf{o},\mathbf{b}_\pm)g_\mp(\mathbf{b}_\pm,\mathbf{i}). \quad (52)$$

Similarly the b^{-3} terms equal

$$\mathcal{K}_\pm g_\pm(\mathbf{o},\mathbf{b}_\pm)\mathcal{K}_\mp g_\mp(\mathbf{b}_\pm,\mathbf{b}_\mp)\mathcal{K}_\pm g_\pm(\mathbf{b}_\mp,\mathbf{b}_\pm)g_\mp(\mathbf{b}_\pm,\mathbf{i}) +\begin{bmatrix}\mathcal{K}\\ \mathfrak{M}\end{bmatrix}_\pm g_\pm(\mathbf{o},\mathbf{b}_\pm)\begin{bmatrix}\mathfrak{M}\\ \mathcal{K}\end{bmatrix}_\mp g_\mp(\mathbf{b}_\pm,\mathbf{b}_\mp)g_\pm(\mathbf{b}_\mp,\mathbf{i}) +\mathfrak{N}_\pm g_\pm(\mathbf{o},\mathbf{b}_\pm)g_\mp(\mathbf{b}_\pm,\mathbf{i}). \quad (53)$$

Thus

$$G_\pm(\mathbf{o},\mathbf{b}_\pm)=(51)+(52)+(53)+\mathcal{O}(b^{-4}). \quad (54)$$

Terms to order $(kb)^{-2}$ were obtained originally by Karp and Zitron[5] by a different procedure. Starting with U_t correct to order b^{-2} [essentially the first two terms of (16) with g replaced by G_t], they expand in plane waves around scatterer s to order b^{-2}, and then carry out a process of successive scatterings to appropriate order to obtain essentially $G\approx(51)+(52)$.

4.2 Small Scatterers

For two spheres, we use

$$G_\pm(\mathbf{o},\mathbf{i})=\sum_n\sum_m A_{nm}{}^\pm(\mathbf{i})Y_n{}^m(\mathbf{o}),$$
$$Y_n{}^m(\mathbf{o})=P_n{}^m(\cos\theta)e^{im\varphi}, \quad (55)$$
$$g_\pm(\mathbf{o},\mathbf{i})=\sum\sum a_n{}^\pm(2n+1)Y_n{}^{-m}(\mathbf{i})Y_n{}^m(\mathbf{o})(-1)^m,$$

[15] P. M. Morse, Proc. Natl. Acad. Sci. U. S. 42, 276 (1956).

where from (44), we have

$$A_{nm}^{\pm}=(2n+1)a_n^{\pm}(-1)^m[Y_n^{-m}(\mathbf{i}) +e^{\mp i2\delta}\sum_\nu\sum_\mu A_{\nu\mu}{}^{\mp}H_{\pm}(mn;\nu\mu)],$$

$$H_{\pm}(mn;\nu\mu)=\int \exp[i2kb\mathbf{p}\cdot\mathbf{b}_{\pm}]Y_n^{-m}(\mathbf{p})Y_\nu^{\mu}(\mathbf{p})d\Omega/2\pi$$

$$=\sum_q d_q(mn;\mu\nu)i^q h_q(2kb)Y_q^{\mu-m}(\mathbf{b}_{\pm}). \quad (56)$$

Equivalently, in terms of

$$B_n^{\pm}=A_n^{\pm}e^{\pm i2\delta},\ b_n^{\pm}=a_n^{\pm}e^{\pm i2\delta}(2n+1),$$

we have

$$B_{nm}^{\pm}=b_n^{\pm}(-1)^m[Y_n^{-m}(\mathbf{i})+\sum_\nu\sum_\mu B_{\nu\mu}{}^{\mp}H_{\pm}(mn;\nu\mu)]. \quad (57)$$

If we keep only the monopole ($n=\nu=0$) and dipole ($n=\nu=1$) terms in (57), and suppress the arguments $2kb$ in h_n, and $\mathbf{b}_{\pm}$ in Y_n, we obtain

$$\begin{aligned}
B_{00}^{\pm}&=b_0^{\pm}[1+B_{00}{}^{\mp}h_0+ih_1(B_{10}{}^{\mp}Y_1^0+B_{11}{}^{\mp}Y_1^1 +B_{1-1}{}^{\mp}Y_1^{-1})],\\
B_{10}^{\pm}&=b_1^{\pm}[Y_1^0(\mathbf{i})+B_{00}{}^{\mp}ih_1Y_1^0 +B_{10}{}^{\mp}(h_0-2h_2Y_2^0)/3 -B_{11}Y_2^1h_2/3-B_{1-1}{}^{\mp}Y_2^{-1}h_2],\\
B_{11}^{\pm}&=-b_1^{\pm}[Y_1^{-1}(\mathbf{i})+B_{00}{}^{\mp}ih_1Y_1^{-1} -B_{10}{}^{\mp}Y_2^{-1}h_2-B_{11}(Y_2^0h_2+h_0)/3 -B_{1-1}{}^{\mp}Y_2^{-2}h_2],\\
B_{1-1}^{\pm}&=-b_1^{\pm}[Y_1^1(\mathbf{i})+B_{00}{}^{\mp}ih_1Y_1^1-B_{10}{}^{\mp}Y_2^1h_2/3 -B_{11}{}^{\mp}Y_2^2h_2/3-B_{1-1}{}^{\mp}(Y_2^0h_2+h_0)/3].
\end{aligned} \quad (58)$$

Introducing $\tau_{\pm}=\pi/2$, and $\beta_{\pm}=\pm\pi/2$ (i.e., scatterers located at $y_{\pm}=\pm b$, $x=z=0$) we simplify (58) to

$$\begin{aligned}
B_{00}^{\pm}&=b_0^{\pm}[1+B_{00}{}^{\mp}h_0\mp(B_{11}{}^{\mp}+B_{1-1}{}^{\mp}/2)h_1],\\
B_{10}^{\pm}&=b_1^{\pm}[\cos\theta_i+B_{10}^{\pm}(h_0+h_2)/3],\\
B_{11}^{\pm}&=\frac{b_1^{\pm}}{2}[\sin\theta_i e^{-i\varphi_i}\pm B_{00}{}^{\mp}h_1+B_{11}(2h_0-h_2)/3 -B_{11}h_2/2],\\
B_{1-1}^{\pm}&=-b_1^{\pm}[\sin\theta_i e^{i\varphi_i}\mp B_{00}{}^{\mp}h_1+B_{11}{}^{\mp}h_2 -B_{1-1}{}^{\mp}(2h_0-h_2)/6].
\end{aligned} \quad (59)$$

It is convenient to write

$$e^{\pm i2\delta}G_{\pm}=B_0^{\pm}+B_1^{\pm}\cos\theta+C_1^{\pm}\sin\theta\cos\varphi +D_1^{\pm}i\sin\theta\sin\varphi, \quad (60)$$

$$C_1=B_{11}-\tfrac{1}{2}B_{1-1},\quad D_1=B_{11}+\tfrac{1}{2}B_{1-1}.$$

Introducing these new coefficients into (59) yields

$$\begin{aligned}
B_0^{\pm}&=b_0^{\pm}[1+B_0{}^{\mp}h_0\mp D_1{}^{\mp}h_1],\\
B_1^{\pm}&=b_1^{\pm}[\cos\theta_i+B_1{}^{\mp}(h_0+h_2)/3],\\
C_1^{\pm}&+b_1^{\pm}[\sin\theta_i\cos\varphi_i+C_1{}^{\mp}(h_0+h_2)/3],\\
D_1^{\pm}&=b_1^{\pm}[-i\sin\theta_i\sin\varphi_i\pm B_0{}^{\mp}h_1+D_1{}^{\mp}(h_0-2h_2)/3],
\end{aligned} \quad (61)$$

where $(h_0+h_2)/3=h_1(2kb)/2kb$, and

$$(h_0-2h_2)/3=h_1'=\partial h_1(2kb)/\partial 2kb.$$

Monopoles: If the scatterers are monopoles, then (61) reduces to

$$B_0^{\pm}=b_0^{\pm}(1+B_0{}^{\mp}h_0)=\frac{b_0^{\pm}(1+b_0{}^{\mp}h_0)}{1-b_0{}^{+}b_0{}^{-}h_0{}^2}. \quad (62)$$

Equivalently, the multiple scattering amplitudes

$$G_{\pm}=a_0^{\pm}\frac{[1+a_0^{\pm}e^{\mp i2\delta}h_0]}{1-a_0{}^{+}a_0{}^{-}h_0{}^2} \quad (63)$$

are closed forms of the geometrical progression of "orders of scattering"; here a_0 is the monopole coefficient of the appropriate isolated scatterer. (See Twersky[14] for detailed discussion of the analogous form for two cylinders.) Applying the cross-section theorem to the compound object yields

$$k^2P/4\pi=-\mathrm{Re}\mathcal{G}(\mathbf{i},\mathbf{i})=-\mathrm{Re}(G_{+}+G_{-})$$

$$=-\mathrm{Re}\left[\frac{a_0{}^{+}+a_0{}^{-}+a_0{}^{+}a_0{}^{-}h_0(2kb)2\cos(2\mathbf{k}\cdot\mathbf{b})}{1-a_0{}^{+}a_0{}^{-}h_0{}^2}\right] \quad (64)$$

where P is the total energy cross section. Analogous results are given by Brueckner[4] and by Watson.[6]

For the special symmetrical case of incidence normal to the line joining identical scatterers, we obtain

$$\begin{aligned}
G_{\pm}&=G=a_0/[1-a_0h_0(2kb)],\\
\mathcal{G}&=2a_0\cos\Delta/(1-a_0h_0).
\end{aligned} \quad (65)$$

For lossless scatterers, the analog of (22) for the compound object yields

$$\frac{k^2P}{4\pi}=-2\,\mathrm{Re}G=\frac{|G|^2}{4\pi}\int|e^{-i\Delta}+e^{i\Delta}|^2d\Omega$$

$$=|G|^2 2[1+j_0(2kb)]. \quad (66)$$

It is simple to demonstrate that the explicit forms (65) fulfills (66). Thus since $h=j_0+in_0$, we obtain

$$\mathrm{Re}G=\frac{\mathrm{Re}[a_0(1-a_0h_0)^*]}{|1-a_0h_0|^2}=\frac{\mathrm{Re}a_0-|a_0|^2j_0}{|1-a_0h_0|^2};$$

and since [from (24)] $-\mathrm{Re}a_0=|a_0|^2=k^2p/4\pi$ holds

for the isolated lossless scatterer, we see that

$$\mathrm{Re}G=\frac{\mathrm{Re}a_0(1+j_0)}{|1-a_0h_0|^2}=-\left|\frac{a_0}{1-a_0h_0}\right|^2(1+j_0)$$
$$=-|G|^2(1+j_0), \quad (67)$$

where the final form followed from the identity $|G|^2=|a_0/(1-a_0h_0)|^2$ obtained from (65). Thus the explicit form G in terms of a_0 given in (65) is consistent within the framework of the scattering theorems: it yields a compound amplitude $\mathcal{G}$ which, as we have in effect proved, fulfills the required scattering theorem (22).

Dipoles: If only the dipole terms are significant, then (61) reduces to

$$B_1^{\pm}=b_1^{\pm}[\cos\theta_i+B_1^{\mp}h_1/\rho]$$
$$=\frac{b_1^{\pm}\cos\theta_i(1+b_1^{\mp}h_1/\rho)}{1-b_1^{+}b_1^{-}(h_1/\rho)^2}, \quad \rho=2kb;$$
$$C_1^{\pm}=b_1^{\pm}[\sin\theta_i\cos\varphi_i+C_1^{\mp}h_1/\rho]$$
$$=\frac{b_1^{\pm}\sin\theta_i\cos\varphi_i(1+b_1^{\mp}h_1/\rho)}{1-b_1^{+}b_1^{-}(h_1/\rho)^2}, \quad (68)$$
$$D_1^{\pm}=b_1^{\pm}[-i\sin\theta_i\sin\varphi_i+D_1^{\mp}h_1']$$
$$=\frac{-ib_1^{\pm}\sin\theta_i\sin\varphi_1(1+b_1^{\mp}h_1')}{1-b_1^{+}b_1^{-}(h_1')^2}.$$

Consequently the multiple scattering amplitudes equal

$$G_{\pm}=3a_1^{\pm}\frac{(1+3a_1^{\mp}e^{\mp i2b}h_1/\rho)}{1-9a_1^{+}a_1^{-}(h_1/\rho)^2}$$
$$\times[\cos\theta\cos\theta_i+\sin\theta\sin\theta_i\cos\varphi\cos\varphi_i], \quad (69)$$
$$+3a_1^{\pm}\frac{(1+3a_1^{\mp}e^{\mp i2b}h_1')}{1-9a_1^{+}a_1^{-}(h_1')^2}\sin\theta\sin\theta_i\sin\varphi\sin\varphi_i.$$

Note that while each isolated scattering amplitude consists of a single dipole in the direction of incidence [i.e., $g_{\pm}=a_1^{\pm}\cos(\mathbf{o}\cdot\mathbf{i})$], the multiple scattering amplitudes consist of two uncoupled "compound dipoles" (in the sense that each is the closed form of the corresponding geometrical progression of orders of scattering)—one in the plane perpendicular to the axis of the "dumbbell" (involving $o_xi_x+o_zi_z$), and one parallel to the axis (involving o_yi_y). Analogous results are given by Brueckner,[4] and Watson,[6] and essentially the same forms were obtained for circular and elliptic cylinders.[1]

For incidence normal to the axis of the scatterers, the coefficients C_1 and D_1 of (68) vanish. In particular, for the symmetrical case of identical scatterers, we obtain simply

$$B_1^{\pm}=B_1=\frac{3a_1}{1-3a_1h_1/\rho},$$
$$\mathcal{G}=\frac{6a_1\cos\theta\cos\Delta}{1-3a_1h_1/\rho}=2B_1\cos\theta\cos\Delta. \quad (70)$$

The scattering theorem for lossless scatterers gives

$$k^2P/4\pi=-2\,\mathrm{Re}B_1=\tfrac{2}{3}|B_1|^2[1+3j_1(2kb)/2kb]. \quad (71)$$

Since

$$\mathrm{Re}2B_1=\frac{\mathrm{Re}6a_1(1-3a_1h_1/\rho)^*}{|1-3a_1h_1/\rho|^2}=\frac{6(\mathrm{Re}a_1-3|a_1|^2j_1/\rho)}{|1-3a_1h_1/\rho|^2},$$

and $-\mathrm{Re}a_1=|a_1|^2$ [from (24)], we see essentially as for (67) that the explicit form (70) fulfills $-\mathrm{Re}\mathcal{G}=(1/4\pi)\int|\mathcal{G}|^2d\Omega$.

Monopoles Plus Dipoles: If both monopole and dipole terms are significant, then we keep all terms of (61). The equations for $B_1^{\pm}$ and $C_1^{\pm}$ are identical with those in (68), i.e.,

$$B_1^{\pm}=\beta^{\pm}i_z, \quad C_1^{\pm}=\beta^{\pm}i_x, \quad \beta^{\pm}=\frac{b_1^{\pm}(1+b_1^{\mp}h_1/\rho)}{1-b_1^{+}b_1^{-}(h_1/\rho)^2}. \quad (72)$$

On the other hand, $B_0^{\pm}$ and $D_1^{\pm}$ are coupled through

$$B_0^{\pm}=b_0^{\pm}[1+B_0^{\mp}h_0\mp D_1^{\mp}h_1],$$
$$D_1^{\pm}=b_1^{\pm}[S\pm B_0^{\mp}h_1+D_1^{\mp}h_1'], \quad (73)$$
$$S\equiv-i\sin\theta_i\sin\varphi_i,$$

which constitutes a system of four equations for the four unknowns:

$$\begin{bmatrix}1 & -b_0^{+}h_0 & 0 & b_0^{+}h_1\\ -b_0^{-}h_0 & 1 & -b_0^{-}h_1 & 0\\ 0 & -b_1^{+}h_1 & 1 & -b_1^{+}h_1'\\ b_1^{-}h_1 & 0 & -b_1^{-}h_1' & 1\end{bmatrix}\begin{bmatrix}B_0^{+}\\ B_0^{-}\\ D_1^{+}\\ D_1^{-}\end{bmatrix}=\begin{bmatrix}b_0^{+}\\ b_0^{-}\\ b_1^{+}S\\ b_1^{-}S\end{bmatrix}. \quad (74)$$

Explicitly, we have

$$\Delta B_0^{\pm}=b_0^{\pm}[(1+b_0^{\mp}h_0)(1-b_1^{+}b_1^{-}h_1'^2)$$
$$-b_0^{\mp}b_1^{\pm}h_1^2(1+b_1^{\mp}h_1')]$$
$$-b_0^{\pm}h_1S\{b_1^{\mp}(1+b_1^{\pm}h_1')$$
$$-b_0^{\mp}b_1^{\pm}[h_0(1+b_1^{\mp}h_1')+b_1^{\mp}h_1^2]\},$$
$$\Delta D_1^{\pm}=b_1^{\pm}S[(1+b_1^{\mp}h_1')(1-b_0^{+}b_1^{-}h_0^2)$$
$$-b_0^{\pm}b_1^{\mp}h_1^2(1+b_0^{\mp}h_0)]\pm b_1^{\pm}h_1\{b_0^{\mp}(1+b_0^{\pm}h_0)$$
$$-b_0^{\pm}b_1^{\mp}[h_1'(1+b_0^{\mp}h_0)+b_0^{\mp}h_1^2]\}, \quad (75)$$
$$\Delta=1-b_0^{+}b_0^{-}h_0^2-b_1^{+}b_1^{-}h_1'^2-b_0^{-}b_1^{+}h_1^2-b_0^{+}b_1^{-}h_1^2$$
$$+b_0^{+}b_0^{-}b_1^{+}b_1^{-}(h_1^4+h_0^2h_1'^2+2h_0h_1^2h_1').$$

Equations (72), (75), (60), and (46) provide the explicit closed form for multiple scattering by two arbitrarily separated objects such that each scatterer is fully specified by its appropriate monopole and dipole

terms when isolated. This case of course includes the previous two illustrations. With reference to applications, the previous result for monopoles (62) is of interest (for example) as an approximation for the scattering of sound by two bubbles in water; for such problems the present results (which include all orders of monopole-dipole coupling, etc.) provide explicit corrections. Similarly, the result for dipoles (68) is of interest in acoustics for scatterers having the same compressibility as the medium, but different densities. The present results (monopole plus dipole) apply for two sets of compressibilities and densities differing from those of the medium, as well as for two different sized rigid scatterers. For two very small rigid scatterers, we have $a_0^{\pm}\approx -id^{\pm}-|d^{\pm}|^2$, $a_1^{\pm}\approx i2d^{\pm}-|2d^{\pm}|^2$ [where $d^{\pm}=\frac{1}{3}(kr_{\pm}')^3$ in terms of the radii $r_{\pm}'$] and the expressions simplify considerably. Other simple forms follow for $kb\gg 1$ ("far-field" multiple scattering, for which case $h_1'\sim h_0$, and $h'/kb\sim 0$), and also for $kb\sim 0$ ("near-field" multiple scattering, for which case only the largest inverse powers of kb would be retained in the h's); see corresponding expressions for two cylinders given previously.[14]

Other applications are based on appropriate symmetry components of the above. Thus we may now treat one scatterer near a perfectly reflecting infinite plane, two protuberances on such a plane, and one protuberance on the wall of a perfectly reflecting quadrant; see reference 14. Quantum mechanical applications of analogous forms are given by Brueckner[4] and Watson.[6]

For normal incidence and identical scatterers, (75) simplifies to

$$D_1^{\pm}=\pm D_1=\pm b_1h_1/(1+b_1h_1')B_0\equiv\pm\epsilon h_1B_0,$$
$$B_0=B_0^{\pm}=b_0/(1-b_0h_0-\epsilon b_0h_1^2). \tag{76}$$

Including the appropriate form of (72), we have

$$G_{\pm}=[B_0+B_1\cos\theta\pm D_1 i\sin\theta\sin\varphi]$$
$$=\frac{b_0(1\pm i\epsilon h_1\sin\theta\sin\varphi)}{1-b_0h_0-\epsilon b_0h_1^2}+\frac{b_1\cos\theta}{1-b_1h_1/\rho}, \tag{77}$$

and

$$\mathcal{G}=2[B_0+B_1\cos\theta]\cos\Delta+2D_1\sin\theta\sin\varphi\sin\Delta. \tag{78}$$

If we neglect the terms containing ϵ, then the explicit forms for lossless scatterers fulfill

$$k^2P/4\pi=-\mathrm{Re}2(B_0+B_1)$$
$$=2|B_0|^2(1+j_0)+\tfrac{2}{3}|B_1|^2[1+j_0+j_2]. \tag{79}$$

The above are merely illustrations. Using (57), it is but a question of algebra to obtain explicit results when higher multipoles are retained.

Similarly there are other special configurations for which the multiple scattered functions may be obtained explicitly in terms of single scattered analogs and functions depending on the geometry of the array. In particular, all uniform distributions (periodic, or random) of identical scatterers with centers lying either on an infinite line or on an infinite plane reduce essentially to an equivalent "one-body" problem, i.e., for such symmetrical cases the multiple scattering functions (except for the phase factor of the incident plane wave) are essentially independent of a scatterer's position.[2] This holds similarly for "slab region" volume distributions for which G_s depends essentially only on the distance of s from the slab face.[2] The same applies for other essentially symmetrical cases, e.g, for a "ring distribution" of identical scatterers (periodically or randomly distributed on a circle) each scatterer is identically excited by all others. Other configurations have the required symmetry for limiting values of a parameter, e.g., for very long wavelengths (so that phase effects can be neglected) each scatterer of a spherical shell of monopoles is specified by the same scattering function, etc.

Small Angle X-Ray Scattering from Randomly Oriented Cylinders of Arbitrary Cross Section*

AKELEY MILLER† AND PAUL W. SCHMIDT
Department of Physics, University of Missouri, Columbia, Missouri
(Received May 5, 1961)

At the larger angles of the small-angle region, asymptotic expansions are the appropriate means of calculating the intensity of the small angle x-ray scattering from a sample of identical particles with uniform electron density and random position and orientation. Since other workers have previously shown that this intensity can be expressed as a Fourier integral transform involving a function characteristic of the size and shape of the particles, standard techniques for asymptotic expansion of Fourier integrals can be employed. In order to make a more general study than would be possible by consideration of particles with a single shape, right cylinders of arbitrary cross section have been chosen for this investigation. These generalized cylinders permit a fairly wide range of choice of particle shape and are well adapted to the study of both elongated and flattened particles. An integral has been developed which gives the characteristic function for the generalized cylinder in terms of the characteristic function of the two-dimensional cross section of the cylinder. As the asymptotic expressions for the scattered intensity depend quite critically on the discontinuities in derivatives of the characteristic function, a study has been made of these derivatives. An expression for the scattered intensity from rectangular parallepipeds has been calculated. A treatment of the limiting case of highly elongated cylinders is given.

INTRODUCTION

SCATTERING curves calculated for ideal particles with selected geometric forms are useful in interpreting small angle x-ray scattering data in order to obtain information about the size and shape of the colloidal particles in the sample which is being studied. As an exact calculation of these theoretical scattering curves can be carried out in closed form for only a very few particle shapes, recourse must be made to approximate methods.

Different approximation methods are best suited for different angular regions of the scattering curve. At the smaller angles of the small-angle region, a series in powers of the scattering angle is convenient. Even though the power series converges at all angles, it is not convenient to use at larger angles, because of the large number of terms which must be retained. At these relatively large angles of the small angle region, other methods of evaluating the intensity are more appropriate. In particular, asymptotic expansion techniques can be applied.[1]

In order to investigate the relation between the scattered intensity and the size and shape of the scattering particle, this paper considers the scattering from the class of particles which may be described as generalized right cylinders of arbitrary convex cross section and uniform electron density. The consideration of such a class of particle shapes yields more general results than can be obtained from the calculations for single particle shapes that have been given in previous work. A completely general treatment for an arbitrary particle shape is not possible at present.

The analytic expressions which come out of the present treatment of a class of shapes demonstrate some general results that would not be apparent from numerical calculations. Particular attention is paid to the scattered intensity at relatively large angles, where asymptotic expansions are convenient.

Prisms are an important subclass of the class of generalized cylinders. The expansion developed by Stokes[2] for the light scattering from elongated prisms applies equally well to the small angle x-ray scattering. We will discuss the relation of Stokes' result to the expressions developed here.

DISCONTINUITIES IN THE DERIVATIVES OF THE CHARACTERISTIC FUNCTION

For a sample of identical cylindrical particles which are random in position and in orientation, the small-angle x-ray scattering intensity $i(h)$ can be expressed by[3]

$$i(h)=\int_0^{D_3} dr \frac{4\pi r^2}{V}\gamma_0(r)\frac{\sin hr}{hr}, \tag{1}$$

where V is the particle volume, D_3 is the maximum diameter of the particle (i.e., the length of the longest line which will fit inside the particle), ϕ is the scattering angle, λ is the x-ray wavelength, $h=4\pi\lambda^{-1}\sin(\phi/2)$, and $\gamma_0(r)$ is a function called the characteristic function, the form of which is determined by the size and shape of the scattering particle. The characteristic function $\gamma_0(r)$ represents the probability that from a point picked at random in the particle and in a direction picked at random with respect to the particle, a displacement

* The material in this paper is discussed in more detail in a thesis presented by Akeley Miller in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Missouri. Work supported by the National Science Foundation.

† Now at the Department of Physics, Utah State University, Logan, Utah.

[1] P. W. Schmidt and R. Hight, Jr., J. Appl. Phys. 30, 867 (1959).

[2] A. R. Stokes, Proc. Phys. Soc. (London) B70, 384 (1957).

[3] A. Guinier, G. Fournet, and C. B. Walker, *Small Angle Scattering of X-rays* (John Wiley & Sons, Inc., New York, 1955), pp. 12–13.

of magnitude r will terminate in another point which also lies in particle. Alternatively, one can say that $4\pi V^{-1}r^2\gamma_0(r)$ is the probability density function for the distance r between any pair of points in the particle.

One can also express $i(h)$ by

$$i(h)=\frac{1}{V^2}\int_{V_1}dV_1\int_{V_2}dV_2\frac{\sin hr_{12}}{hr_{12}},$$

where both volume integrations are over the volume of the particle and r_{12} is the distance between the volume elements. For right cylinders the volume integrations may each be separated into one integration over the cross section and another over the axial dimension. In this separation note that the limits of integration over the cross section are independent of the limits of integration along the axis. One can therefore represent the double integration over the cross section in terms of a single integral over a characteristic function $\beta_0(q)$ for the cross section. The two axial integrations, which are independent of the integrals over the cross section, then can be expressed in terms of an axial characteristic function $\alpha_0(p)$, to give

$$i(h)=\int_0^{D_1}dp\frac{2}{D_1}\alpha_0(p)\int_0^{D_2}dq\frac{2\pi q}{A}\beta_0(q)\frac{\sin h(p^2+q^2)^{\frac{1}{2}}}{h(p^2+q^2)^{\frac{1}{2}}},\quad(2)$$

where D_1 is the axial length of the particle and D_2 is the maximum diameter of the cross section. Thus, for a cylinder, $D_3=(D_1^2+D_2^2)^{\frac{1}{2}}$. As can be shown,[3]

$$\alpha_0(p)=1-p/D_1.$$

Then $2D_1^{-1}\alpha_0(p)$ is the probability density function for the axial component p, while the probability density for a distance q in the cross section, the area of which is A, will be given by $2\pi A^{-1}\beta_0(q)$.

If the variables of integration are changed from p, q to θ, r by the transformation,

$$p=r\sin\theta,$$
$$q=r\cos\theta,$$

then, when the elongation $v\equiv D_1/D_2\geq 1$, Eq. (2) becomes

$$\begin{aligned}i(h)=\frac{4\pi}{V}\Bigg\{&\int_0^{D_2}drr^2\frac{\sin hr}{hr}\int_0^{\pi/2}d\theta\cos\theta\alpha_0(r\sin\theta)\beta_0(r\cos\theta)\\&+\int_{D_2}^{D_1}drr^2\frac{\sin hr}{hr}\int_{\cos^{-1}D_2/r}^{\pi/2}d\theta\cos\theta\alpha_0(r\sin\theta)\beta_0(r\cos\theta)\\&+\int_{D_1}^{D_3}drr^2\frac{\sin hr}{hr}\\&\times\int_{\cos^{-1}D_2/r}^{\sin^{-1}D_1/r}d\theta\cos\theta\alpha_0(r\sin\theta)\beta_0(r\cos\theta)\Bigg\}.\quad(3)\end{aligned}$$

It is convenient to define

$$H(r)\equiv(4\pi r/V)\gamma_0(r)\quad(4)$$

so that (1) becomes

$$i(h)=\frac{1}{h}\int_0^{D_3}drH(r)\sin hr.\quad(5)$$

From (3) and (5), $H(r)$ is seen to have the form

$$H(r)=\frac{4\pi}{V}r\int_0^{\pi/2}d\theta\cos\theta\alpha_0(r\sin\theta)\beta_0(r\cos\theta)$$

for $0\leq r\leq D_2$. Three pieces of the function $H(r)$ are evident in (3) for contiguous ranges in r. Three similar pieces exist for the case $v\leq 1$, although for simplicity we will consider only the case $v\geq 1$.

If β_0 for the cross section also has more than one piece, say $k+1$ pieces, then analysis shows that H will have $2k+3$ pieces. For example, β_0 for a rectangular cross section has three pieces, so that $k=2$, and H for the rectangular parallelepiped has seven pieces. Let the boundary points for the $k+1$ pieces of β_0 be $a_i=b_iD_2=b_i(v^2+1)^{-\frac{1}{2}}D_3$ for $i=0,\ 1,\ \cdots,\ k+1$, and let $\beta_{0i}(r\cos\theta)$ denote the piece for $a_i\leq r\cos\theta\leq a_{i+1}$. Note that $0=b_0<b_1<\cdots<b_k<b_{k+1}=1$. When (3) is rewritten to indicate these pieces of β_0 with appropriate limits for the r and θ integrations, the $2k+3$ pieces of H may be identified. Thus

$$H(r)\equiv H_i(r)\quad\text{for}\quad a_i\leq r\leq a_{i+1}$$

where $i=0,\ \cdots,\ 2k+2$ with

$$a_i=b_iD_2=b_i(v^2+1)^{-\frac{1}{2}}D_3\quad\text{for}\quad i=0,1,\cdots,k+1$$

$$=(D_1^2+b^2{}_{i-k-2}D_2^2)^{\frac{1}{2}}=\left(\frac{v^2+b^2{}_{i-k-2}}{v^2+1}\right)^{\frac{1}{2}}D_3$$

$$\text{for}\quad i=k+2,\ k+3,\ \cdots,\ 2k+3.$$

Just as was done for $\gamma_0(r)$ by Guinier *et al.*,[3] one can express $\beta_0(q)$ as an integral over a line distribution function involving a probability density function $g_2(M)$ for the length M of a line segment which terminates on the boundary of the convex cross section, which passes through a point selected at random in the cross section, and which lies in a direction selected at random in the cross section. The function $g_2(M)$ is analogous to the function $g(M)$ used by Guinier *et al.*

The fact that $\beta_0(r)$ can be written as an integral involving $g_2(M)$ leads to the results that β_0 and β_0' are continuous, just as γ_0 and $\gamma_0'(r)$ were found to be continuous because $\gamma_0(r)$ could be expressed as an integral over $g(M)$.[1]

In $\beta_0''(q)$, however, infinite but integrable discontinuities may be found. In all known cases, these infinite discontinuities occur at the boundary points of the pieces of $\beta_0(q)$.

As γ_0 and γ_0' are continuous, $H(r)$ and $H'(r)$ will also be continuous. In $H''(r)$, however, both finite and infinite discontinuities sometimes may be found at the boundary points of the pieces. For a given cylindrical

particle, all but one of these discontinuities are related to the infinite discontinuities in $\beta_0''(q)$.

We will now assume the existence of H'' and the integrability of H'''. Then after two partial integrations of (5), there results

$$i(h)=-\frac{1}{h^4}\sum_{i=0}^{2k+2}\left[\mathfrak{D}_i^{(2)}\cos ha_i+\int_{a_i}^{a_{i+1}}dr H_i'''(r)\cos hr\right] \quad (6)$$

where

$$\begin{aligned}\mathfrak{D}_i^{(2)}&\equiv H_0''(0) && \text{for } i=0\\ &\equiv H_i''(a_i)-H_{i-1}''(a_i) && \text{for } i=1, 2, \cdots, 2k+2.\end{aligned}$$

The meaning of (6) must be discussed in terms of the existence of finite non-vanishing $\mathfrak{D}_i^{(2)}$ and of the integrability of H'''. For $\mathfrak{D}_0^{(2)}\equiv H_0''(0)$,[1,4]

$$\begin{aligned}H_0''(0)&=(4\pi/V)[2\gamma_0'(r)+r\gamma_0''(r)]_{r=0}\\ &=(4\pi/V)2\gamma_0'(0)=-2\pi s/V^2.\end{aligned} \quad (7)$$

For $i=1, 2, \cdots, k$,

$$\mathfrak{D}_i^{(2)}=\frac{4\pi}{V}\lim_{r\to a_i^+}\int_0^{\cos^{-1}a_i/r}d\theta\cos^3\theta r \times\alpha_0(r\sin\theta)\beta_{0i}''(r\cos\theta), \quad (8)$$

where $r\to a_i^+$ indicates that the limit is approached from above. Since the upper limit in (8) goes to zero in the limit as $r\to a_i^+$, the right-hand side of (8) can be nonzero only if β_{0i}'' is infinitely discontinuous at $r=a_i$.

Analysis of (8) shows that the right-hand side vanishes unless $\beta_{0i}''(r\cos\theta)$ contains $(r^2\cos^2\theta-a_i^2)^{-\sigma}$ as a factor, that convergence of $\mathfrak{D}_i^{(2)}$ requires $\sigma\leq\frac{1}{2}$, and that the convergent $\mathfrak{D}_i^{(2)}$ is nonvanishing only when $\sigma=\frac{1}{2}$. Such a factor $(r^2\cos^2\theta-a_i^2)^{-\frac{1}{2}}$ occurs in each of the second and third pieces of β_0'' for a rectangular cross section with a_i equal to the width of the rectangle in the second piece and to the length of the rectangle in the third piece. A geometrical argument, which makes use of the function $g_2(M)$ mentioned above, strongly suggests that this type of infinite discontinuity in β_0'' and the corresponding finite discontinuity $\mathfrak{D}_i^{(2)}$ in H'' are associated with polygonal portions of the periphery of the cross section which are *parallel* and *opposite*. In the case of the rectangular cross section of width a_1, (8) becomes

$$\mathfrak{D}_1^{(2)}=4\pi/Va_1. \quad (9)$$

When $i=k+1$, for all generalized cylinders,

$$\mathfrak{D}_{k+1}^{(2)}=\frac{4\pi}{V}\left[\lim_{r\to D_2^+}\int_{\cos^{-1}D_2/r}^{\cos^{-1}a_k/r}d\theta\cos^3\theta r\alpha_0(r\sin\theta)\beta_{0k}''(r\cos\theta) - \lim_{r\to D_2^-}\int_0^{\cos^{-1}a_k/r}d\theta\cos^3\theta r\alpha_0(r\sin\theta)\beta_{0k}''(r\cos\theta)\right]. \quad (10)$$

In the case of a circular cross section ($k=0$), $\beta_{00}''(r\cos\theta)$ contains a factor $(D_2^2-r^2\cos^2\theta)^{-\frac{1}{2}}$, and this factor makes both limits indicated on the right-hand side of (10) infinite. The quantities responsible for this discontinuous behavior are complete elliptic integrals of the first kind whose arguments approach unity as r approaches D_2^+ and D_2^- in the first and second terms of (10), respectively. Not only does $\mathfrak{D}_{k+1}^{(2)}$ not exist in any well defined sense for a circular cross section, but H''' for such a cross section is not integrable at $a_1=D_2$. Geometric considerations which again make use of the function $g_2(M)$ lead us to believe that (6) is thus invalid not only for the circular cross section but also for any convex cross section for which the periphery contains portions which are diametrically opposite arcs of a circle which has a diameter equal to the maximum diameter of the cross section. We can describe these last as "partially circular." It is believed that $\mathfrak{D}_{k+1}^{(2)}=0$ for all cylinders except those with circular or "partially circular" cross sections.

For $i=k+2$ we have

$$\begin{aligned}\mathfrak{D}_{k+2}^{(2)}&=(4\pi/V)(D_1/r^2)\beta_{00}[(r^2-D_1^2)^{\frac{1}{2}}]|_{r=D_1}\\ &=4\pi/VD_1.\end{aligned} \quad (11)$$

This finite discontinuity at the axial dimension exists for all cylindrical particles. It is very closely related to the type of discontinuity (9). Both (9) and (11) are associated with portions of the surface of the particle which are *parallel* and *opposite planes*.

For $i=k+3, k+4, \cdots, 2k+2$

$$\mathfrak{D}_i^{(2)}=-\frac{4\pi}{V}\lim_{r\to a_i^-}\int_{\cos^{-1}b_{i-k-1}D_2/r}^{\sin^{-1}D_1/r}d\theta\cos^3\theta r \times\alpha_0(r\sin\theta)\beta''_{0(i-k-3)}(r\cos\theta). \quad (12)$$

These $\mathfrak{D}_i^{(2)}$ are zero for all known β_0''.

These results extend those reported earlier by Schmidt and Hight concerning the dependence of $i(h)$ on the inverse fourth power of h and on the surface area of the particle.[1,5,6] In (6) the existence of the finite discontinuities $\mathfrak{D}_i^{(2)}$ in H'' in (9) and (11) associated with portions of the particle surface which are parallel and opposite planes establishes that oscillatory terms of the order of h^{-4} exist in $i(h)$ for high enough h. (Calculations now in progress suggest that an h^{-4} dependence may hold at large h even for cylinders with circular or "partially-circular" cross sections, in spite of the infinite discontinuities in H''.) The nature of these oscillatory terms will now be further explored.

THE RECTANGULAR PARALLELEPIPED

The connection between finite discontinuities and parallel and opposite plane portions of the particle surface lends some interest to an expansion of the form (6) for a rectangular parallelepiped. The characteristic function $\gamma_0(r)$ and its first three derivatives have been

[4] Reference 3, pp. 14–15.

[5] G. Porod, Acta Phys. Austriaca 2 (1949), pp. 267–268.

[6] Reference 3, p. 17.

evaluated for a rectangular parallelepiped of dimensions $2l \leq 2m \leq 2n$, for both the case $n \geq (l^2+m^2)^{\frac{1}{2}}$ and the case $n \leq (l^2+m^2)^{\frac{1}{2}}$. The use of these in (6) results, for the integral portion, in improper integrals which may be evaluated by a theorem due to Erdelyi.[1,7] When this is done and when the notation

$$p \equiv 2(l^2+m^2+n^2)^{\frac{1}{2}}, \quad \Delta_1 \equiv 2l/p, \quad \Delta_2 \equiv 2m/p, \quad \Delta_3 \equiv 2n/p$$
$$k \equiv 8/\Delta_1\Delta_2\Delta_3, \quad z \equiv hp$$

is adopted, the first few terms of the resulting asymptotic expansion for $i(z)$ for the rectangular parallelepiped are

$$\begin{aligned} i(z) = \frac{k^2}{8}\frac{1}{z^4}\Big\{ & \frac{4\pi}{k}\Big[\frac{1-\cos\Delta_1 z}{\Delta_1}+\frac{1-\cos\Delta_2 z}{\Delta_2}+\frac{1-\cos\Delta_3 z}{\Delta_3}\Big] \\ & +(1/z)[\Delta_1 \sin\Delta_1 z+\Delta_2 \sin\Delta_2 z+\Delta_3 \sin\Delta_3 z] \\ & +(\pi/2z)^{\frac{1}{2}}[(\Delta_1)^{\frac{1}{2}}(\Delta_2+\Delta_3)\cos(\Delta_1 z+\pi/4) \\ & +(\Delta_2)^{\frac{1}{2}}(\Delta_1+\Delta_3)\cos(\Delta_2 z+\pi/4) \\ & +(\Delta_3)^{\frac{1}{2}}(\Delta_1+\Delta_2)\cos(\Delta_3 z+\pi/4)]\Big\}+O(z^{-11/2}). \end{aligned} \tag{13}$$

Note that the dominant terms of order z^{-4} involve oscillations of period $4\pi/\Delta_1$, $4\pi/\Delta_2$, and $4\pi/\Delta_3$ in z. In the case of the cube, $\Delta_1=\Delta_2=\Delta_3$ and the terms of order z^{-4} vanish periodically. Such zeros in the intensity have previously been known only for the sphere.[8] Here we have established that such zeros exist at the largest angles whenever the ratios of Δ_1, Δ_2, and Δ_3 to one another are all rational numbers. Thus there exist large undamped oscillations in the limiting form of the scattered intensity at the largest angles of the small angle scattering region, but these oscillations occur in terms of order h^{-4} not h^{-3} as some work has indicated.[6] The oscillations would complicate the determination of the surface area of the particle from the scattering data at the largest angles of the small angle region.

The nature of the convergence of the asymptotic expansions of the improper integrals evaluated by the Erdelyi theorem indicates that this asymptotic expansion (13) for the intensity of scattering from a rectangular parallelepiped is best for particles of nearly equal dimensions (particles with maximum and minimum diameters that are of the order of magnitude of the mean diameter), and is not useful for greatly elongated or greatly flattened particles. In these extreme cases ($v \gg 1$ or $v \ll 1$) it is always possible to take h large enough to obtain asymptotic convergence, but these very large angles are rarely of practical interest.

We will now consider a different type of asymptotic expansion which is adapted to highly elongated particles.

[7] A. Erdelyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956), p. 49.
[8] Reference 3, p. 55.

ASYMPTOTIC EXPANSION FOR THE ELONGATED CYLINDER

When $D_2=L$, $D_1=vL$, $D_3=(v^2+1)^{\frac{1}{2}}L$, and when the variables $s=q/D_2$ and $t=p/D_1$ replace the variables q and p, respectively, then with appropriate change in the functional form of α_0 and β_0 to take account of the change in the argument, (2) becomes

$$i(h)=\int_0^1 ds\frac{2\pi s}{A_1}\beta_0(s)\int_0^1 dt 2\alpha_0(t)\frac{\sin hL(s^2+v^2t^2)^{\frac{1}{2}}}{hL(s^2+v^2t^2)^{\frac{1}{2}}}, \tag{14}$$

where $A_1=A/D_2^2$. Since both β_0 and β_0' are continuous, a notation which took account of the pieces in β_0 would be superfluous in this expression. Denote

$$g(h;s)\equiv 2\int_0^1 dt(1-t)\frac{\sin hL(s^2+v^2t^2)^{\frac{1}{2}}}{hL(s^2+v^2t^2)^{\frac{1}{2}}} \tag{15}$$

noting that $\alpha_0(t)=1-t$. With the transformation,

$$y=(s^2+v^2t^2)^{\frac{1}{2}}$$

(15) becomes

$$g(h;s)=\frac{2}{hvL}\Big[Z(s)-\frac{1}{v}\int_s^{(v^2+s^2)^{\frac{1}{2}}}\sin hLy\,dy\Big] \tag{16}$$

where

$$Z(s)\equiv \tfrac{1}{2}\pi J_0(hLs)-Y(s) \tag{17}$$

and

$$Y(s)\equiv\int_{(v^2+s^2)^{\frac{1}{2}}}^{\infty}\frac{\sin hLy}{(y^2+s^2)^{\frac{1}{2}}}dy. \tag{18}$$

When (18) is partially integrated, (14) may be written in the form

$$i(h)=i_1(h)+i_2(h)+i_3(h)+i_4(h)+i_{4R}(h) \tag{19}$$

where

$$\beta_2(s)\equiv\frac{2\pi s}{A_1}\beta_0(s),$$

$$i_1(h)\equiv\frac{\pi}{hvL}\int_0^1 ds\beta_2(s)J_0(hLs)$$

$$i_2(h)\equiv-\frac{2}{(hvL)^2}\int_0^1 ds\beta_2(s)\cos hLs$$

$$i_3(h)\equiv-\frac{2}{(hvL)^3}\int_0^1 ds\beta_2(s)\frac{(v^2+s^2)^{\frac{1}{2}}}{v}\sin hL(v^2+s^2)^{\frac{1}{2}}$$

$$i_4(h)\equiv\frac{2}{(hvL)^4}\int_0^1 ds\beta_2(s)\frac{2v^2+3s^2}{v^2}\cos hL(v^2+s^2)^{\frac{1}{2}}$$

$$i_{4R}(h)\equiv-\frac{6}{(hvL)^4}\int_0^1 ds\beta_2(s)\times\int_{(v^2+s^2)^{\frac{1}{2}}}^{\infty}dy\frac{v^3y(2y^2+3s^2)}{(y^2-s^2)^{\frac{1}{2}}}\cos hLy.$$

Preliminary analysis of $i_1(h)$ indicates it is of order $(hLv)^{-1}$ for $hL \ll 1 \ll hvL$, and that it is of order $v^{-1}(hL)^{-4}$ for $1 \ll hL \ll hvL$. Since all terms in $i(h)$ other than $i_1(h)$ are of inverse power in v equal to or greater than 2, it follows that $i_1(h)$ is the dominant term of $i(h)$ for highly elongated particles. For such elongated particles, then, the scattering at angles for which $hL \ll 1$ is like that from a rod of length vL with negligible cross section; that is, the intensity at these relatively small angles is proportional to $(hL)^{-1}$ or h^{-1}. For $hL \gg 1$, the intensity of scattering from the elongated particle varies with $(hL)^{-4}$ or h^{-4}, as is expected in the small angle scattering from all three-dimensional particles of uniform electron density when the smallest particle dimensions L is such that $hL \gg 1$.

When one takes account of the infinite discontinuities which may occur in $\beta_0''(s)$, partial integration shows that $i_2(h)$ and $i_3(h)$ are each of order h^{-4} when $hL \gg 1$. Thus, $i(h)$ is of order h^{-4} for large h and (19) is seen to be an asymptotic expansion which is valid for h as large as may be desired. It may be noted that for large hL and large v the $\cos hLs$ in the integrand of $i_2(h)$ oscillates rapidly, while the $\sin hL(v^2+s^2)^{\frac{1}{2}}$ remains nearly constant as s varies from 0 to 1. For this reason $i_3(h)$ seems to behave as though it were of order v^{-2} instead of v^{-3}, and $i_2(h)$ and $i_3(h)$ seem to be of comparable magnitude for large enough h.

By series evaluation of $i_3(h)$ and $i_4(h)$, making use of the trigometric identity

$$\sin hL(v^2+s^2)^{\frac{1}{2}} = \sin hvL \cos hL[(v^2+s^2)^{\frac{1}{2}}-v] + \cos hvL \sin hL[(v^2+s^2)^{\frac{1}{2}}-v]$$

and of the corresponding identity for $\cos hL(v^2+s^2)^{\frac{1}{2}}$, one can obtain the expression derived by Stokes[2] for the light scattering from elongated prisms. Stokes' work extends an earlier calculation by Porod.[5] Among the terms of their expression, both workers obtained $i_1(h)$ and $i_2(h)$. Their equations are derived under the assumption that $v > hL > 1$.

In the light-scattering problem with which Stokes was concerned, there was no need to consider the behavior of the intensity for $hL \gg 1$, since such a large h was experimentally unattainable. For numerical evaluation at intermediate values of h, Stokes' equation may be the most convenient expression to use. In small-angle x-ray scattering, on the other hand, since hL can be experimentally large relative to 1, an investigation of the asymptotic behavior of $i(h)$ is of interest. Because in Stokes' work the assumption was made that $v > hL > 1$, his expression cannot be an asymptotic expansion.

In addition to giving an asymptotic expression for $i(h)$, Eq. (19) provides a means of readily estimating the error involved in dropping i_2, i_3, or i_4. The order of dependence of these terms on v and H can also be determined. Thus (19) can provide a way of knowing when Stokes' equation will be a good approximation to the scattered intensity.

An analog of (19) for use with very flattened platelets is being developed.

Use of the Displacement Vector in Electromagnetic Theory*

K. G. DEDRICK AND ROBERT N. WILSON†
Microwave Laboratory, W. W. Hansen Laboratories of Physics, Stanford University, Stanford, California
(Received May 26, 1961)

Problems in the electrodynamics of charged fluids often require calculation of the charge and current densities, given prior knowledge of the initial distributions and the displacement vector field. In this paper, solutions are obtained for these new distributions, with the property that if the initial distributions have a sharply defined boundary, terms arise which can be interpreted as distributions of electric charge and current multipoles located on the initial boundary surface. These solutions cannot be considered valid near the initial boundary, but prove to be useful in that many of their properties are given correctly. The electromagnetic potentials due to the charge and current densities in the displaced configuration are calculated, and form the basis for a discussion of the complicated boundary value problem encountered in the description of the electrodynamics near the boundary surface. The results are directly applicable to the theory of high frequency electron tubes, and certain formulas are of use in hydrodynamics and elasticity.

I. INTRODUCTION

HYDRODYNAMICS provides a conceptual framework and many mathematical methods which can be directly applied to a variety of physical situations. Topics in the electrodynamics of charged fluids are often treated in terms of a hydrodynamic model. Such treatments may require the solution of an easily stated problem. Consider an initial distribution of electric charge described by a density function $\rho_0(\mathbf{r},t)$. Let each element of charge be given a displacement ζ which is a function of the initial position of the element $\mathbf{r}$ and the time t so that the new position of the charge element is $\mathbf{r}+\zeta(\mathbf{r},t)$. We ask for the new charge density $\rho(\mathbf{r},t)$ in terms of ρ_0 and the displacement vector ζ. Similarly, given an initial current density $\mathbf{J}_0(\mathbf{r},t)$, we ask for the new current density in the displaced configuration.

In the theory of high frequency electron tubes, this problem has received considerable attention. Here, we require a description of the electrodynamics of an electron beam moving under the influence of externally applied electromagnetic fields in addition to the fields due to the beam itself. The boundary value problem for the electromagnetic field is complicated by the motion of the beam boundary under the influence of these fields. It has been treated by solving a related problem in which the boundary is held stationary, and the effects of the actual boundary motion are replaced by locating surface charge and current distributions on this stationary boundary which is taken to be the boundary in the absence of excitation.[1] A rigorous justification for this convenient method seems to be lacking.

Many different procedures[2] are available for calculation of the surface charge and current distributions when carried to first-order terms in the displacement vector, but it is not always clear how to proceed to higher orders. Chu[1] and Sturrock[3] have developed two different mathematical procedures leading to the surface distributions, and they give results through second order in the displacement. Chu calculates the charge density in the displaced configuration by application of the Taylor expansion and the Jacobian of the transformation connecting the initial and displaced positions of the charge elements. Toward the edge of the electron beam, the Taylor expansion is not valid. In order to circumvent this difficulty, a *modified problem* is introduced which is defined by two requirements. These are: (a) that the electric charge must be conserved, and (b) that the electromagnetic power in the modified problem is required to be equal to the power in the actual system under study. As a result, surface charge and current densities are required to support discontinuities of the electromagnetic field vectors in the modified problem. Sturrock's method is based on the Fourier transform of the charge density and is employed here in Sec. III. In his work, as well as in ours, the surface distributions appear without hypothesis.

We obtain expressions for the surface distributions in a way which relates directly to the central problem outlined in the first paragraph. In the course of analysis of this central problem, terms arise which may be interpreted physically as surface distributions of charge and current located on the boundary of the initial charge configuration. The appearance of these terms can be accounted for only by understanding that mathematical rigor has been sacrificed in the analysis. We believe that the central problem has not been solved completely and choose to examine the solutions from the point of view of their usefulness.

Our developments are based primarily on an integral theorem which is the subject of Sec. II. In Sec. III, the electric charge density in the displaced configuration is calculated and in Sec. IV is applied to a discussion of the electrostatic potential. The current density in the displaced configuration is the subject of Sec. V. In Sec. VI, the retarded potentials due to the charge and current densities are calculated. These potentials form the basis

* This work was supported jointly by the U. S. Army Signal Corps, the U. S. Air Force, and the Office of Naval Research.
† Now at Kane Engineering Laboratories, Palo Alto, California.
[1] An historical account of this method is given by E. L. Chu, J. Appl. Phys. 31, 381 (1960).
[2] See for example, items 1–5, 7, and 8 in the list of references given by Chu.[1]
[3] P. A. Sturrock, J. Math. Phys. 1, 405 (1960).

of a demonstration that the electromagnetic fields generated by the charge and current densities in the displaced configuration can be calculated using the expressions for the charge and current densities obtained in Secs. III and V, even though these densities are not always correct everywhere. Section VII includes a discussion of the complicated boundary value problem encountered in the description of the electrodynamics of charges moving under the influence of their collective fields. This problem is particularly troublesome in the region near the boundary surface of the charge distribution and is conveniently discussed in terms of the electromagnetic fields generated by the volume and surface charge and current distributions.

II. AN INTEGRAL THEORM

Let $\Psi(\mathbf{r})$ and $\Phi(\mathbf{r})$ be analytic scalar functions of the position vector $\mathbf{r}$ in the volume V enclosed by the surface S. We are interested in evaluating volume integrals of the form

$$\int_V d\tau \Psi(\mathbf{r})\Phi[\mathbf{r}+\boldsymbol{\zeta}(\mathbf{r})], \tag{1}$$

where $\boldsymbol{\zeta}(\mathbf{r})$ is a vector function of $\mathbf{r}$ analytic within V and on the surface S. The quantity $\Phi[\mathbf{r}+\boldsymbol{\zeta}(\mathbf{r})]$ can be expanded by Taylor's theorem

$$\Phi[\mathbf{r}+\boldsymbol{\zeta}(\mathbf{r})]=\sum_{i,j,k=0}^{\infty}\frac{1}{i!j!k!}\zeta_x{}^i\zeta_y{}^j\zeta_z{}^k\frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial y^j}\frac{\partial^k}{\partial z^k}\Phi(\mathbf{r}). \tag{2}$$

By collecting terms in this triple summation, the Taylor operator may be rewritten as

$$\sum_{i,j,k=0}^{\infty}\frac{1}{i!j!k!}\zeta_x{}^i\zeta_y{}^j\zeta_z{}^k\frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial y^j}\frac{\partial^k}{\partial z^k}=\sum_{\nu=0}^{\infty}\frac{1}{\nu!}\boldsymbol{\zeta}^\nu(\cdot\boldsymbol{\nabla})^\nu=\exp(\boldsymbol{\zeta}\cdot\boldsymbol{\nabla}), \tag{3}$$

where the symbolic form[4] $\exp(\boldsymbol{\zeta}\cdot\boldsymbol{\nabla})$ is *defined* by (3). The notation used in the terms of the second expression for the Taylor operator means that the νth-rank tensor whose elements are the elements of $\boldsymbol{\zeta}^\nu$ is completely contracted with the νth-rank tensor whose elements are the elements of $\boldsymbol{\nabla}^\nu$. This is an extension of the familiar dot product notation of vector analysis, and is more compact and convenient for the problems under discussion than the usual notation for the contraction of tensors as expressed by the summation convention. In all cases, this dot product notation is defined by the summation convention.[5]

[4] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics* (Cambridge University Press, New York, 1946), p. 242.

[5] For example, by $\mathbf{ABC}(\cdot\mathbf{D})(\cdot\mathbf{E})^2$, we mean $A_\gamma B_\beta C_\alpha D_\alpha E_\beta E_\gamma$ where repeating indices are summed from 1 to 3. In the notation of Gibbs, this quantity would be written $\mathbf{ABC}\dot{:}\mathbf{EED}$. See J. W. Gibbs and E. B. Wilson, *Vector Analysis* (Dover Publications Inc., New York), p. 306.

If we use the notation of (3) and suppress the functional dependence of Ψ, Φ, and $\boldsymbol{\zeta}$, the integral (1) becomes

$$\int_V d\tau\Psi\exp(\boldsymbol{\zeta}\cdot\boldsymbol{\nabla})\Phi=\sum_{\nu=0}^{\infty}\frac{1}{\nu!}\int_V d\tau\Psi\boldsymbol{\zeta}^\nu(\cdot\boldsymbol{\nabla})^\nu\Phi. \tag{4}$$

We are interested mainly in the values of the integrals on the right-hand side of (4) when the differentiations have all been transformed to operate on Ψ and $\boldsymbol{\zeta}^\nu$ rather than on Φ. To accomplish this, we integrate by parts ν times. It is necessary to use a generalization of the divergence theorem. Let $\mathbf{A}$ be a vector differential operator such that the result of operating upon f with $\mathbf{A}$ may be written as $\mathbf{A}f$, which in turn is a vector. The divergence theorem yields

$$\int_V d\tau(\boldsymbol{\nabla}\cdot\mathbf{A}f)=\int_S da(\mathbf{n}\cdot\mathbf{A}f), \tag{5}$$

where $\mathbf{n}$ is the unit outward normal vector on S. We may readily write

$$\boldsymbol{\nabla}\cdot\mathbf{A}f=(\mathbf{A}\cdot\boldsymbol{\nabla})f+[{}^\dagger\boldsymbol{\nabla}\cdot\mathbf{A}]f, \tag{6}$$

where $\boldsymbol{\nabla}^\dagger$ operates only on $\mathbf{A}$ in the second term and not on f. Combining Eqs. (6) and (5) gives the desired expression

$$\int_V d\tau(\mathbf{A}\cdot\boldsymbol{\nabla})f=\int_S da(\mathbf{n}\cdot\mathbf{A}f)-\int_V d\tau[\boldsymbol{\nabla}^\dagger\cdot\mathbf{A}]f. \tag{7}$$

Equation (7) is now applied to any of the integrals in the series on the right hand side of (4) as follows[6]:

$$\begin{aligned}\int_V d\tau\Psi\boldsymbol{\zeta}^\nu(\cdot\boldsymbol{\nabla})^\nu\Phi&=\int_V d\tau\{\Psi\boldsymbol{\zeta}^\nu(\cdot\boldsymbol{\nabla})^{\nu-1}\}(\cdot\boldsymbol{\nabla}\Phi)\\&=\int_S da\mathbf{n}\cdot\{\Psi\boldsymbol{\zeta}^\nu(\cdot\boldsymbol{\nabla})^{\nu-1}\}\Phi\\&\quad-\int_V d\tau[\boldsymbol{\nabla}^\dagger\cdot\Psi\boldsymbol{\zeta}^\nu](\cdot\boldsymbol{\nabla})^{\nu-1}\Phi.\end{aligned} \tag{8}$$

The last volume integral in (8) is treated by the method of Eq. (7) also. This integral becomes

$$\begin{aligned}&\int_V d\tau\{[\boldsymbol{\nabla}^\dagger\cdot\Psi\boldsymbol{\zeta}^\nu](\cdot\boldsymbol{\nabla})^{\nu-2}\}(\cdot\boldsymbol{\nabla}\Phi)\\&\quad=\int_S da\mathbf{n}\cdot[\boldsymbol{\nabla}^\dagger\cdot\Psi\boldsymbol{\zeta}^\nu](\cdot\boldsymbol{\nabla})^{\nu-2}\Phi\\&\quad\quad-\int_V d\tau[(\boldsymbol{\nabla}^\dagger\cdot)^2\Psi\boldsymbol{\zeta}^\nu](\cdot\boldsymbol{\nabla})^{\nu-2}\Phi.\end{aligned} \tag{9}$$

[6] Throughout this work, the operator $\boldsymbol{\nabla}$ differentiates quantities only on the right-hand side of it.

Equation (7) is applied a total of ν times to the integral in question. The result is

$$\int_V d\tau\Psi\zeta^\nu(\cdot\nabla)^\nu\Phi=(-)^\nu \times\int_V d\tau\Phi[(\nabla^\dagger\cdot)^\nu\Psi\zeta^\nu]+(-)^{\nu-1}\sum_{\sigma=0}^{\nu-1}(-)^\sigma \times\int_S da\mathbf{n}\cdot[(\nabla^\dagger\cdot)^{\nu-1-\sigma}\Psi\zeta^\nu](\cdot\nabla)^\sigma\Phi. \quad (10)$$

This expression is now used in (4). The sums on ν and σ in the surface terms are best rearranged so that all terms involving $(\cdot\nabla)^\sigma\Phi$ are grouped. The desired theorem is

$$\int_V d\tau\Psi\exp(\zeta\cdot\nabla)\Phi=\int_V d\tau\Phi\{\exp(-\nabla\cdot\zeta)\}_{op}\Psi +\int_S da\mathbf{n}\cdot\sum_{\sigma=0}^{\infty}\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!} \times[(\nabla^\dagger\cdot)^\alpha\Psi\zeta^{\alpha+\sigma+1}](\cdot\nabla)^\alpha\Phi, \quad (11)$$

where the following symbolic notation is employed:

$$\{\exp(-\nabla\cdot\zeta)\}_{op}\Psi\equiv\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\nabla\cdot)^\nu\zeta^\nu\Psi. \quad (12)$$

In the surface integrals of Eq. (11), $\nabla^\dagger$ operates on all terms to the right of it which are within the square brackets.[7]

Another useful integral theorem is obtained by writing Eq. (11) for the three scalar components of a vector $\boldsymbol{\Psi}$ and adding so as to form a vector equivalent of Eq. (11). Thus we write

$$\boldsymbol{\Psi}(\mathbf{r})=\mathbf{i}\Psi_x(\mathbf{r})+\mathbf{j}\Psi_y(\mathbf{r})+\mathbf{k}\Psi_z(\mathbf{r}), \quad (13)$$

and form the integrals

$$\int_V d\tau\Psi_a\exp(\zeta\cdot\nabla)\Phi; \quad (a=x,y,z).$$

We now add the results of these integrations so as to form a true vector equation, namely

$$\int_V d\tau\boldsymbol{\Psi}\exp(\zeta\cdot\nabla)\Phi=\int_V d\tau\Phi\{\exp(-\nabla\cdot\zeta)\}_{op}\boldsymbol{\Psi} +\int_S da\mathbf{n}\cdot\sum_{\sigma=0}^{\infty}\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!} \times[(\nabla^\dagger\cdot)^\alpha\zeta^{\alpha+1}\boldsymbol{\Psi}\zeta^\sigma](\cdot\nabla)^\sigma\Phi. \quad (14)$$

[7] For example, $\mathbf{n}\cdot[\nabla^\dagger\cdot\Psi\zeta^3](\cdot\nabla)\Phi$ is written

$$n_\alpha[\partial(\Psi\zeta_\beta\zeta_\alpha\zeta_\gamma)/\partial x_\beta](\partial\Phi/\partial x_\gamma)$$

in the usual notation.

The important point is the placement of $\boldsymbol{\Psi}$ among all the vectors ζ in the square brackets of the surface terms. Equation (14) indicates that we contract $\mathbf{n}$ and $(\nabla^\dagger\cdot)^\alpha$ with $\zeta^{\alpha+1}$. The remaining vectors ζ^σ contract with $(\cdot\nabla)^\sigma$ to the right of the square brackets.

III. ELECTRIC CHARGE DENSITY

Both Chu[1] and Sturrock[3] have derived expressions for the electric charge density in terms of the displacement vector. Their expressions are carried through second-order terms in the displacement vector, but it is clear in both papers how to proceed to any desired order. This extension is easier using Sturrock's method, and in the following discussion we shall begin by using his Fourier analysis scheme.[8]

Let $\rho_0(\mathbf{r})$ be the initial charge density at $\mathbf{r}$. The charges are now displaced so that those which were at $\mathbf{r}$ in the initial configuration are now at $\mathbf{r}+\zeta(\mathbf{r})$ where ζ is the displacement vector. It is required that ζ be an analytic function of $\mathbf{r}$ and may be an explicit function of the time t as well. We now ask for the charge density $\rho(\mathbf{r})$ in the displaced configuration. Following Sturrock,[3] $\rho(\mathbf{r})$ is calculated in terms of its Fourier transform:

$$\rho(\mathbf{r})=\int d\mathbf{k}\rho(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{r}}. \quad (15)$$

The Fourier transform $\rho(\mathbf{k})$ is given by

$$\rho(\mathbf{k})=\frac{1}{(2\pi)^3}\int d\mathbf{r}'\rho(\mathbf{r}')e^{i\mathbf{k}\cdot\mathbf{r}'} =\frac{1}{(2\pi)^3}\int d[\mathbf{r}'+\zeta(\mathbf{r}')]\,\rho[\mathbf{r}'+\zeta(\mathbf{r}')] \times\exp\{i\mathbf{k}\cdot[\mathbf{r}'+\zeta(\mathbf{r}')]\}, \quad (16)$$

where the second form is obtained from the first by a change of variables. The charge in the volume element $d[\mathbf{r}'+\zeta(\mathbf{r}')]$ is $\rho[\mathbf{r}'+\zeta(\mathbf{r}')]d[\mathbf{r}'+\zeta(\mathbf{r}')]$, which according to charge conservation is equal to $\rho_0(\mathbf{r}')d\mathbf{r}'$, viz.: the charge in the corresponding volume element in the initial configuration.[9] In this way, Eq. (16) becomes

$$\rho(\mathbf{k})=\frac{1}{(2\pi)^3}\int d\mathbf{r}'\rho_0(\mathbf{r}')\exp\{i\mathbf{k}\cdot[\mathbf{r}'+\zeta(\mathbf{r}')]\} =\frac{1}{(2\pi)^3}\int d\mathbf{r}'\rho_0\exp(\zeta\cdot\nabla')e^{i\mathbf{k}\cdot\mathbf{r}'}. \quad (17)$$

The latter form is obtained through use of the Taylor operator of Eq. (3), and for simplicity the functional dependences of ρ_0 and ζ have been suppressed. This integral is of the type treated in the previous section,

[8] See Sec. 3 of reference 3.

[9] This charge conservation statement is correct if one and only one point $\mathbf{r}'$ corresponds to the point $\mathbf{r}'+\zeta(\mathbf{r}')$.

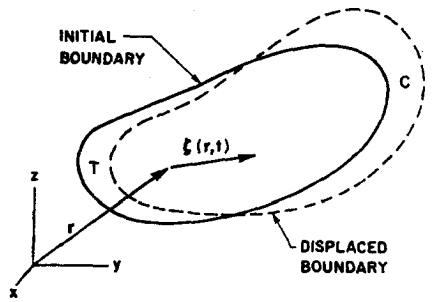


FIG. 1. The initial and displaced boundaries of a charge distribution.

and we shall apply Eq. (11). The result is then used in (15) to give

$$\frac{1}{(2\pi)^3}\int d\mathbf{k}e^{-i\mathbf{k}\cdot\mathbf{r}}\int_V d\mathbf{r}'e^{i\mathbf{k}\cdot\mathbf{r}'}\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\nabla'\cdot)^\nu\rho_0\zeta^\nu$$
$$+\frac{1}{(2\pi)^3}\int d\mathbf{k}e^{-i\mathbf{k}\cdot\mathbf{r}}\int_S da'\mathbf{n}\cdot\sum_{\sigma=0}^{\infty}\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!}$$
$$\times[(\nabla^{\dagger\prime}\cdot)^\alpha\rho_0\zeta^{\alpha+\sigma+1}](\cdot\nabla')^\sigma e^{i\mathbf{k}\cdot\mathbf{r}'}. \quad (18)$$

A further integration over $\mathbf{k}$ yields the delta function $\delta(\mathbf{r}-\mathbf{r}')$. The result is denoted by $\hat{\rho}(\mathbf{r})$ and it will be seen below that $\hat{\rho}(\mathbf{r})$ is not, in general, equal to $\rho(\mathbf{r})$ everywhere. For this reason we write

$$\rho(\mathbf{r})\sim\hat{\rho}(\mathbf{r}), \quad (19)$$

where

$$\hat{\rho}(\mathbf{r})=\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\nabla\cdot)^\nu\rho_0\zeta^\nu$$
$$+\int_S da'\mathbf{n}\cdot\sum_{\sigma=0}^{\infty}\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!}$$
$$\times[(\nabla^{\dagger\prime}\cdot)^\alpha\rho_0\zeta^{\alpha+\sigma+1}](\cdot\nabla')^\sigma\delta(\mathbf{r}-\mathbf{r}'). \quad (20)$$

The terms in the sum over ν represent $\hat{\rho}(\mathbf{r})$ when $\mathbf{r}$ is in V and the surface integrals give $\hat{\rho}(\mathbf{r})$ when $\mathbf{r}$ is on S. If $\mathbf{r}$ is outside V and not on S, then $\hat{\rho}(\mathbf{r})$ is zero.[10]

The choice of V and S is important in the discussion to follow. If ρ_0 and all its derivatives are continuous functions everywhere and vanish at infinity, the surface S is taken at infinity and V is the whole space. In many problems, we are interested in dealing with initial charge density functions which are well behaved within some volume V, and such that ρ_0 falls abruptly to zero just outside V. In these problems, we take S to be just inside the surface of discontinuity of ρ_0 so that ρ_0, ζ and all their derivatives are defined. Then V is chosen as the volume within S. This means that all terms in the series in (20) are well behaved to the extent that no effects of the discontinuity in ρ_0 are in evidence.

[10] For example, let $g(\mathbf{r})=\int_V d\mathbf{r}'f(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')$. Then $g(\mathbf{r})=f(\mathbf{r})$ if $\mathbf{r}$ is in V, and g is zero otherwise.

If ρ_0 and all its derivatives are continuous and vanish at infinity, the surface integrals in (20) are all zero, and $\hat{\rho}(\mathbf{r})$ is given by the terms in the sum over ν. These terms define a contribution to $\hat{\rho}(\mathbf{r})$ which is specifically a volume charge density, and we shall denote them by $\hat{\rho}_V(\mathbf{r})$. From (20) it is clear that $\hat{\rho}_V(\mathbf{r})$ persists if we allow the discontinuity in ρ_0, and hence it is useful to consider $\hat{\rho}_V(\mathbf{r})$ separately in all cases. From (20) then, for $\mathbf{r}$ in V, we obtain

$$\hat{\rho}_V(\mathbf{r})=\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\nabla\cdot)^\nu\rho_0\zeta^\nu=\rho_0(\mathbf{r})-\nabla\cdot\mathbf{P}, \quad (21a)$$

$$\mathbf{P}(\mathbf{r})=\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{(\nu+1)!}(\nabla\cdot)^\nu\rho_0\zeta^{\nu+1}, \quad (21b)$$

and for $\mathbf{r}$ outside V, we say that $\hat{\rho}_V$ is zero. The second form of (21a) is useful in demonstrating that charge conservation is not destroyed through the operations leading to its development. We are also led to a possible physical interpretation of $\mathbf{P}$ in that its role in (21a) is that of the electric polarization vector.[11] Equation (21a) is in agreement with the extension of Sturrock's[3] Eq. (3.6) and through second order in ζ, is equal to Chu's[1] Eq. (9).

When expressed as a function of $\mathbf{r}$, the surface integrals in (20) are strongly localized on the surface S. The $\sigma=0$ part may be viewed as a multiple of a delta function located on the surface of the initial charge distribution, and we are led to consider that this part may be interpreted as a surface charge distribution. Similarly, for $\sigma=1, 2, \cdots$, derivatives of the delta function occur and suggest interpretation in terms of surface dipole, quadrupole, $\cdots$, distributions. This tentative interpretation is strengthened in the next section where the electrostatic potential generated by these terms is calculated.

A typical example of initial and displaced charge distributions is shown in Fig. 1. The *correct* expression for the charge density in the displaced configuration would show that ρ is zero in T and nonzero in C. The result of our development given by (20) shows, on the contrary, that $\hat{\rho}$ is zero in C and generally nonzero in T. The volume density $\hat{\rho}_V$ of (21) is generally nonzero inside the initial boundary S of Fig. 1, and the surface terms of (20) are located on this boundary. Thus the operations leading to (20) give a result which, in general, is not correct everywhere. We may say however, that $\hat{\rho}(\mathbf{r})$ is in a sense *equivalent* to the correct charge density $\rho(\mathbf{r})$ in that the results of calculations of certain *integral properties* of $\rho(\mathbf{r})$ are given correctly when $\hat{\rho}(\mathbf{r})$ of (20) is used.

Integral properties of the charge density may be calculated in two different ways. One way is to use $\hat{\rho}(\mathbf{r})$ given by Eq. (20) and the other is to start afresh and

[11] See, for example, J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 12.

calculate the integral property using (11).[12] Let $f(\mathbf{r})$ be an arbitrary analytic function. The integral property G, of the charge density ρ, is then defined by

$$G=\int_\infty d\mathbf{r}\rho(\mathbf{r})f(\mathbf{r})=\int_V d\mathbf{r}\rho_0(\mathbf{r})f[\mathbf{r}+\boldsymbol{\zeta}(\mathbf{r})]. \quad (22)$$

The second form on the right is obtained from the first on use of the charge-conservation argument employed in obtaining (17) from (16). This second expression for G is of the form of (1) and is readily evaluated using (11). The result is written immediately as

$$G=\int_V d\mathbf{r}f(\mathbf{r})\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\boldsymbol{\nabla}\cdot)^\nu\rho_0\boldsymbol{\zeta}^\nu + \int_S da\mathbf{n}\cdot\sum_{\sigma=0}^{\infty}\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!} \times[(\boldsymbol{\nabla}^\dagger\cdot)^\alpha\rho_0\boldsymbol{\zeta}^{\alpha+\sigma+1}](\cdot\boldsymbol{\nabla})^\sigma f(\mathbf{r}). \quad (23)$$

Inspection shows that calculation of G using $\hat{\rho}(\mathbf{r})$ of Eq. (20) for $\rho(\mathbf{r})$ in the first form of (22) again gives (23).[13]

The total charge in the displaced configuration is calculated by choosing $f(\mathbf{r})$ equal to unity in (23). On using (21a) and (23), the volume integral in this case is

$$\int_V \hat{\rho}_V(\mathbf{r})d\mathbf{r}=\int_V \rho_0(\mathbf{r})d\mathbf{r}-\int_V(\boldsymbol{\nabla}\cdot\mathbf{P})d\mathbf{r} =Q_0-\int_S da(\mathbf{n}\cdot\mathbf{P}), \quad (24)$$

where Q_0 is the total charge in the initial configuration. Only the $\sigma=0$ term in the surface integrals of (23) contributes in this example, and the contribution is

$$\int_S da\mathbf{n}\cdot\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+1)!}[(\boldsymbol{\nabla}^\dagger\cdot)^\alpha\rho_0\boldsymbol{\zeta}^{\alpha+1}]=\int_S da(\mathbf{n}\cdot\mathbf{P}). \quad (25)$$

On adding (25) to (24), we find the total charge in the displaced configuration to be Q_0 as required by charge conservation.

The fact that *only* the $\sigma=0$ term of (20) contributes to total charge permits interpretation of this term as a surface charge density. This surface charge density may be written $(\mathbf{n}\cdot\mathbf{P})$ where $\mathbf{P}$ is evaluated on the surface S of the initial charge distribution. Through second order, this expression for the surface charge density is in agreement with Chu's[1] Eq. (27).

[12] We recall that $\hat{\rho}(\mathbf{r})$ is calculated through the use of Fourier analysis, while in the latter procedure, only charge conservation and the integral relation (11) are required.

[13] The volume V and the surface S are chosen as described following (20).

IV. ELECTROSTATIC POTENTIAL

The electrostatic potential $\phi(\mathbf{r})$ can be calculated in two different ways. One way is to evaluate the potential due to the charge distribution $\hat{\rho}(\mathbf{r})$ of Eq. (20), and the other is based on employment of (11) directly. This later procedure gives the potential correctly only if $\mathbf{r}$ lies in certain regions of the coordinate system. The potential also provides a framework for the physical interpretation of the surface terms of (20). This interpretation is identical to that suggested in the discussion following Eq. (21).

The potential of a displaced charge distribution is

$$\phi(\mathbf{r})=\frac{1}{4\pi\epsilon}\int\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'=\frac{1}{4\pi\epsilon}\int_V\frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'-\boldsymbol{\zeta}(\mathbf{r}')|}d\mathbf{r}', \quad (26)$$

where the second form is obtained from the first on application of the charge conservation argument employed in obtaining (17) from (16). The potential is an integral property of the charge density and might be treated by applying Eq. (23). The difficulty is that $1/|\mathbf{r}-\mathbf{r}'-\boldsymbol{\zeta}(\mathbf{r}')|$ cannot be represented everywhere by a Taylor series. The Taylor series is valid for all $\mathbf{r}'$ such that $|\boldsymbol{\zeta}(\mathbf{r}')|<|\mathbf{r}-\mathbf{r}'|$, and we write[14]

$$\frac{1}{|\mathbf{r}-\mathbf{r}'-\boldsymbol{\zeta}(\mathbf{r}')|}=\exp(\boldsymbol{\zeta}\cdot\boldsymbol{\nabla}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}; \quad |\boldsymbol{\zeta}(\mathbf{r}')|<|\mathbf{r}-\mathbf{r}'|. \quad (27)$$

If $\mathbf{r}$ is chosen far enough outside the boundary of the initial charge configuration so that the above condition is satisfied, we use (27) in (26) and apply (11). The result is

$$\phi(\mathbf{r})=\frac{1}{4\pi\epsilon}\int_V\frac{d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\boldsymbol{\nabla}'\cdot)^\nu\rho_0\boldsymbol{\zeta}^\nu + \frac{1}{4\pi\epsilon}\int_S da'\mathbf{n}\cdot\sum_{\sigma=0}^{\infty}\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!} \times[(\boldsymbol{\nabla}^{\dagger\prime}\cdot)^\alpha\rho_0\boldsymbol{\zeta}^{\alpha+\sigma+1}](\cdot\boldsymbol{\nabla}')^\sigma\frac{1}{|\mathbf{r}-\mathbf{r}'|}. \quad (28)$$

The potential at points $\mathbf{r}$ within the charge distribution is calculated in a similar manner except that the integration indicated in (26) must be done separately over two regions. A surface S_δ enclosing $\mathbf{r}$ is constructed on which $|\boldsymbol{\zeta}(\mathbf{r}')|=|\mathbf{r}-\mathbf{r}'|$ so that (26) becomes

$$\phi(\mathbf{r})=\frac{1}{4\pi\epsilon}\int_{V-V_\delta}\frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'-\boldsymbol{\zeta}(\mathbf{r}')|}d\mathbf{r}' + \frac{1}{4\pi\epsilon}\int_{V_\delta}\frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'-\boldsymbol{\zeta}(\mathbf{r}')|}d\mathbf{r}'. \quad (29)$$

[14] See, for example, E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Chelsea Publishing Company, New York, 1955), p. 15.

In the first of these integrals, the Taylor series of (27) is valid and the contribution to $\phi(\mathbf{r})$ is given by (28) where V is replaced by $V-V_\delta$ and S by $S-S_\delta$. The second integral in (29) is properly treated using a different expansion for $1/|\mathbf{r}-\mathbf{r}'-\boldsymbol{\zeta}(\mathbf{r}')|$ than the one given by (27). It is believed that this integral may be calculated using (27). The contribution is then of the form of (28) and is expected to be valid, providing that the resulting expression is absolutely convergent.[15] A little consideration[16] shows that (28) cannot be relied upon to give $\phi(\mathbf{r})$ correctly if $\mathbf{r}$ is within the distance ζ of the surface S of the initial charge configuration.

The potential calculated using $\hat{\rho}(\mathbf{r})$ of (20) in the integral on the left of (26) immediately gives (28) also. Since $\hat{\rho}(\mathbf{r})$ is usually incorrect within the distance $|\boldsymbol{\zeta}|$ of S, the potential cannot be reliably calculated if $\mathbf{r}$ is within this region. It is clear that $\hat{\rho}(\mathbf{r})$ may be used to calculate the potential correctly at any point beyond a distance $|\boldsymbol{\zeta}|$ from S, and it is believed to give $\phi(\mathbf{r})$ correctly for $\mathbf{r}$ within S by the distance $|\boldsymbol{\zeta}|$.

The volume integrals in (28) give the potential due to the volume charge density $\hat{\rho}_V(\mathbf{r}')$ of Eq. (21). The surface integrals in Eq. (28) can be written

$$\frac{1}{4\pi\epsilon}\sum_{\sigma=0}^{\infty}(-)^\sigma\int_S da'\mathfrak{D}^{(\sigma)}(\cdot\nabla)^\sigma\frac{1}{|\mathbf{r}-\mathbf{r}'|} \tag{30a}$$

where

$$\mathfrak{D}^{(\sigma)}=\mathbf{n}\cdot\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!}[(\nabla^{\dagger\prime}\cdot)^\alpha\rho_0\zeta^{\alpha+\sigma+1}]. \tag{30b}$$

The terms in (30a) are surface integrals of the quantities

$$\frac{(-)^\sigma}{4\pi\epsilon}\left\{da'\mathfrak{D}^{(\sigma)}(\cdot\nabla)^\sigma\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right\}, \tag{31}$$

which in turn are the potentials at $\mathbf{r}$ of 2^σ multipoles[17] located at $\mathbf{r}'$, the moments being of strength $da'\sigma!\mathfrak{D}^{(\sigma)}$. The quantities $\sigma!\mathfrak{D}^{(\sigma)}$ are interpreted as multipole moment surface densities. The surface charge density is the scalar $\mathfrak{D}^{(0)}=\mathbf{n}\cdot\mathbf{P}$ where $\mathbf{P}$ is given by (21b), the surface dipole moment density is the vector $\mathfrak{D}^{(1)}$, the surface quadrupole moment density is the dyadic $2!\mathfrak{D}^{(2)}$, etc.

V. CURRENT DENSITY

An expression for the current density is obtained by following the same general procedure used in Sec. III above in arriving at $\hat{\rho}(\mathbf{r})$ of Eq. (20), the difference being that we are concerned here with a vector problem requiring the use of (14) rather than (11).

Corresponding to (15), we write

$$\mathbf{J}(\mathbf{r})=\int d\mathbf{k}\mathbf{J}(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{32}$$

where

$$\mathbf{J}(\mathbf{k})=\frac{1}{(2\pi)^3}\int d[\mathbf{r}'+\boldsymbol{\zeta}(\mathbf{r}',t)]\,\mathbf{J}[\mathbf{r}'+\boldsymbol{\zeta}(\mathbf{r}',t)] \times\exp\{i\mathbf{k}\cdot[\mathbf{r}'+\boldsymbol{\zeta}(\mathbf{r}',t)]\}, \tag{33}$$

and where the explicit dependence of $\mathbf{J}$ on the time t has been suppressed. We shall write $\mathbf{J}=\rho\mathbf{v}$ as in the Eulerian approach to hydrodynamics[18] and evaluate $\rho\mathbf{v}$ at $\mathbf{r}'+\boldsymbol{\zeta}(\mathbf{r}',t)$. The velocity $\mathbf{v}$ of charges at the displaced position $\mathbf{r}'+\boldsymbol{\zeta}(\mathbf{r}',t)$ is

$$\mathbf{v}=(d/dt)[\mathbf{r}'+\boldsymbol{\zeta}(\mathbf{r}',t)]=\mathbf{v}_0(\mathbf{r}',t) +(\partial/\partial t)\boldsymbol{\zeta}(\mathbf{r}',t)+(\mathbf{v}_0\cdot\nabla')\boldsymbol{\zeta}(\mathbf{r}',t), \tag{34}$$

where $\mathbf{v}_0(\mathbf{r}',t)$ is the velocity in the initial configuration. This expression is now used in (33) and we employ the charge conservation argument used in obtaining (17). The result is then treated according to (14) and used in (32) to give

$$\mathbf{J}(\mathbf{r},t)\sim\hat{\mathbf{J}}(\mathbf{r},t), \tag{35}$$

where

$$\hat{\mathbf{J}}(\mathbf{r},t)=\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\nabla\cdot)^\nu\boldsymbol{\zeta}^\nu\left\{\mathbf{J}_0+\rho_0\frac{\partial\boldsymbol{\zeta}}{\partial t}+(\mathbf{J}_0\cdot\nabla)\boldsymbol{\zeta}\right\} +\int_S da'\mathbf{n}\cdot\sum_{\sigma=0}^{\infty}\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!} \times\left[(\nabla^{\dagger\prime}\cdot)^\alpha\boldsymbol{\zeta}^{\alpha+1}\left\{\mathbf{J}_0+\rho_0\frac{\partial\boldsymbol{\zeta}}{\partial t}+(\mathbf{J}_0\cdot\nabla')\boldsymbol{\zeta}\right\}\boldsymbol{\zeta}^\sigma\right] \times(\cdot\nabla')^\sigma\delta(\mathbf{r}-\mathbf{r}'). \tag{36}$$

In this expression, the functional dependences of $\mathbf{J}_0$, $\boldsymbol{\zeta}$, and ρ_0 have been dropped, and we write $\mathbf{J}_0=\rho_0\mathbf{v}_0$ which is the current density in the initial configuration. As in Eq. (20), the terms in the sum over ν in (36) represent $\hat{\mathbf{J}}(\mathbf{r},t)$ when $\mathbf{r}$ is in V, and the surface integrals give $\hat{\mathbf{J}}(\mathbf{r},t)$ when $\mathbf{r}$ is on S. For $\mathbf{r}$ outside V and not on S, we take $\hat{\mathbf{J}}(\mathbf{r},t)$ to be zero.

The terms in the sum over ν in (36) give the space or volume current density and are considered separately. Thus we write (for $\mathbf{r}$ in V)

$$\hat{\mathbf{J}}_V=\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\nabla\cdot)^\nu\boldsymbol{\zeta}^\nu\left\{\mathbf{J}_0+\rho_0\frac{\partial\boldsymbol{\zeta}}{\partial t}+(\mathbf{J}_0\cdot\nabla)\boldsymbol{\zeta}\right\}. \tag{37}$$

For some purposes, it is useful to rewrite this expression in another form. The desired form should allow demon-

[15] See, for example, E. C. Titchmarch, *The Theory of Functions* (Oxford University Press, 1932), Sec. 1.77.

[16] A convenient example is the uniformly charged sphere displaced by a constant distance $\boldsymbol{\zeta}$.

[17] See, for example, reference 11, Sec. 3.12.

[18] This expression for $\mathbf{J}$ restricts us to consideration of single stream flow models, i.e., at any time t, the velocity $\mathbf{v}$ shall be a single valued function of the position.

stration of charge conservation as expressed by the continuity equation

$$\nabla\cdot\hat{\mathbf{J}}_V+\partial\hat{\rho}_V/\partial t=0. \tag{38}$$

On addition and subtraction of $(\partial\mathbf{P}/\partial t)$ to (37), we have

$$\hat{\mathbf{J}}_V=\mathbf{J}_0+\partial\mathbf{P}/\partial t+\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{(\nu+1)!}\{-(\nabla\cdot)^{\nu+1}\zeta^{\nu+1}\mathbf{J}_0$$
$$+(\nu+1)(\nabla\cdot)^\nu\zeta^\nu\rho_0(\partial\zeta/\partial t)+(\nu+1)(\nabla\cdot)^\nu\zeta^\nu(\mathbf{J}_0\cdot\nabla)\zeta$$
$$-(\partial/\partial t)[(\nabla\cdot)^\nu\zeta^{\nu+1}\rho_0]\} \tag{39}$$

where $\mathbf{P}$ is given by (21b). It is clear that Eq. (39) and $\hat{\rho}_V$ of (21a) will satisfy (38) if the terms in the sum over ν in (39) can be written as the curl of a vector, or equivalently, as the divergence of an antisymmetric dyadic.[19]

To reduce the expression in the curly brackets of Eq. (39) to the divergence of an antisymmetric dyadic, the differentiation with respect to time is performed and charge conservation is postulated in the initial configuration so that

$$\nabla\cdot\mathbf{J}_0+(\partial\rho_0/\partial t)=0. \tag{40}$$

We then add $(\nabla\cdot)^{\nu+1}\zeta^\nu\mathbf{J}_0\zeta$ and subtract the same quantity in which one of the differentiations indicated by $(\nabla\cdot)$ has been carried out. The curly bracket in (39) becomes

$$\nabla\cdot[(\nabla\cdot)^\nu\zeta^\nu(\mathbf{J}_0\zeta-\zeta\mathbf{J}_0)+\nu(\nabla\cdot)^{\nu-1}\zeta^{\nu-1}(\zeta\boldsymbol{\Gamma}-\boldsymbol{\Gamma}\zeta)], \tag{41a}$$

where

$$\boldsymbol{\Gamma}=\rho_0(\partial\zeta/\partial t)+(\mathbf{J}_0\cdot\nabla)\zeta. \tag{41b}$$

Expression (41a) is thus the curl of a vector, and is written

$$\nabla\times[(\nabla\cdot)^\nu\zeta^\nu(\zeta\times\mathbf{J}_0)+\nu(\nabla\cdot)^{\nu-1}\zeta^{\nu-1}(\boldsymbol{\Gamma}\times\zeta)], \tag{42}$$

so that (39) becomes

$$\hat{\mathbf{J}}_V=\mathbf{J}_0+(\partial\mathbf{P}/\partial t)+\nabla\times\mathbf{F} \tag{43}$$

where

$$\mathbf{F}=\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\nabla\cdot)^\nu\zeta^\nu\left[\frac{1}{\nu+1}(\zeta\times\mathbf{J}_0)+\frac{1}{\nu+2}(\zeta\times\boldsymbol{\Gamma})\right]. \tag{44}$$

Equation (43) together with (21a) satisfy the equation of continuity (38) identically.

The last term in (43) represents a contribution to $\hat{\mathbf{J}}_V$ which is solenoidal, and is of the form taken by magnetization currents[20,21] so that $\mathbf{F}$ may be interpreted as the equivalent density of magnetization.

The surface terms in (36) represent the various multipole currents or the equivalent surface densities of magnetic multipoles. This interpretation follows the procedure developed in Sec. IV. The surface current density $\mathbf{j}_s$ is the $\sigma=0$ part of these terms:

$$\mathbf{j}_s=\mathbf{n}\cdot\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+1)!}(\nabla\cdot)^\alpha\zeta^{\alpha+1}$$
$$\times[\mathbf{J}_0+\rho_0(\partial\zeta/\partial t)+(\mathbf{J}_0\cdot\nabla)\zeta] \tag{45a}$$
$$=\partial\mathfrak{D}^{(1)}/\partial t+\mathbf{F}\times\mathbf{n}+\mathbf{n}\cdot\sum_{\alpha=1}^{\infty}\frac{(-)^\alpha}{(\alpha+2)!}(\nabla\cdot)^\alpha\zeta^{\alpha-1}$$
$$\times[\alpha(\zeta\boldsymbol{\Gamma}-\boldsymbol{\Gamma}\zeta)\zeta+(\alpha+2)(\zeta\mathbf{J}_0-\mathbf{J}_0\zeta)\zeta]. \tag{45b}$$

The surface current is not generally directed along the surface S but may have a component in the normal direction as well. The term $\mathbf{F}\times\mathbf{n}$ is a tangential contribution to $\mathbf{j}_s$ which is related to the density of magnetization,[22] and $\partial\mathfrak{D}^{(1)}/\partial t$ is the rate of growth of the surface dipole moment.[23]

VI. RETARDED POTENTIALS

Questions of radiation from charge and current distributions are conveniently discussed in terms of the scalar potential $\phi(\mathbf{r},t)$ and the vector potential $\mathbf{A}(\mathbf{r},t)$. These potentials may be calculated in mks units using the familiar expressions for the retarded potentials[24]

$$\phi(\mathbf{r},t)=\frac{1}{4\pi\epsilon}\int\frac{\rho(\mathbf{r}',t-|\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' \tag{46}$$

$$\mathbf{A}(\mathbf{r},t)=\frac{\mu}{4\pi}\int\frac{\mathbf{J}(\mathbf{r}',t-|\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'. \tag{47}$$

The electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ are then given by

$$\mathbf{E}=-\nabla\phi-(\partial\mathbf{A}/\partial t) \tag{48}$$

$$\mathbf{B}=\nabla\times\mathbf{A}. \tag{49}$$

The potentials are first calculated by transforming Eqs. (46) and (47) according to the methods used in proceeding from (26) to (28). The result is then shown to be equal to that obtained by using $\hat{\rho}$ and $\hat{\mathbf{J}}$ in Eqs. (46) and (47), provided that $\mathbf{r}$ is not within the distance $|\zeta|$ of the boundary S.

The charge density $\rho(\mathbf{r}',t)$, at any time t, is written in

---

[19] Let $\mathfrak{C}$ be an arbitrary antisymmetric dyadic. Then $\mathfrak{C}$ can be written: $\mathfrak{C}=\mathbf{AB}-\mathbf{BA}$ where $\mathbf{A}$ and $\mathbf{B}$ are vectors. By forming $\nabla\cdot\mathfrak{C}$, it is readily seen that $\nabla\cdot\mathfrak{C}=\nabla\times\mathbf{R}$ where $\mathbf{R}=\mathbf{B}\times\mathbf{A}$.

[20] See reference 11, Sec. 4.10.

[21] W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism* (Addison-Wesley Publishing Co., Reading, Massachusetts, 1955), Secs. 7.9 to 7.12.

[22] See reference 11, Sec. 4.10.

[23] We do not expect (45a) and the surface charge density $\mathbf{n}\cdot\mathbf{P}$ to satisfy Chu's[1] surface continuity equation [Eq. (28)] since the latter is based on surface currents having only tangential components. Therefore, it cannot be expected that our Eq. (45a) for the surface current will agree with Chu's Eq. (29) since the latter expression is obtained from his Eq. (28).

[24] See, for example, reference 21, p. 214.

terms of its Fourier transform $\tilde{\rho}(\mathbf{r}',\omega)$

$$\rho(\mathbf{r}',t)=\frac{1}{2\pi}\int\tilde{\rho}(\mathbf{r}',\omega)e^{-i\omega t}d\omega, \tag{50}$$

where

$$\tilde{\rho}(\mathbf{r}',\omega)=\int\rho(\mathbf{r}',t')e^{i\omega t'}dt'. \tag{51}$$

At the retarded time $t^*=t-|\mathbf{r}-\mathbf{r}'|/c$, Eq. (50) is

$$\rho(\mathbf{r}',t-|\mathbf{r}-\mathbf{r}'|/c)=\frac{1}{2\pi}\int\tilde{\rho}(\mathbf{r}',\omega)e^{-i\omega(t-|\mathbf{r}-\mathbf{r}'|/c)}d\omega. \tag{52}$$

Equations (51) and (52) are combined and the result used in (46):

$$\phi(\mathbf{r},t)=\frac{1}{8\pi^2\epsilon}\int dt'\int d\omega e^{i\omega(t'-t)}\times\int d\mathbf{r}'\rho(\mathbf{r}',t')\frac{e^{i\omega|\mathbf{r}-\mathbf{r}'|/c}}{|\mathbf{r}-\mathbf{r}'|}. \tag{53}$$

The integral over $\mathbf{r}'$ is treated using the arguments required to obtain (28) from (26). The range of validity of the resulting expression is discussed below Eq. (28). Hence the integral over $\mathbf{r}'$ in (53) is

$$\int d\mathbf{r}'\frac{e^{i\omega|\mathbf{r}-\mathbf{r}'|/c}}{|\mathbf{r}-\mathbf{r}'|}\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\nabla'\cdot)^\nu\rho_0(\mathbf{r}',t')\{\zeta(\mathbf{r}',t')\}^\nu + \int da'\mathbf{n}\cdot\sum_{\sigma=0}^{\infty}\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!}\times[(\nabla^{t'}\cdot)^\alpha\rho_0(\mathbf{r}',t')\{\zeta(\mathbf{r}',t')\}^{\alpha+\sigma+1}]\times(\cdot\nabla')^\sigma\frac{e^{i\omega|\mathbf{r}-\mathbf{r}'|/c}}{|\mathbf{r}-\mathbf{r}'|}. \tag{54}$$

The integral over ω in (53) then yields 2π times the delta function $\delta(t'-t+|\mathbf{r}-\mathbf{r}'|/c)$ and the integral over t' requires that the result be evaluated at the retarded time $t^*=t-|\mathbf{r}-\mathbf{r}'|/c$, so that the potential is given by

$$\phi(\mathbf{r},t)=\frac{1}{4\pi\epsilon}\int d\mathbf{r}'\frac{1}{|\mathbf{r}-\mathbf{r}'|}\left\{\sum_{\nu=0}^{\infty}\frac{(-)^\nu}{\nu!}(\nabla'\cdot)^\nu\rho_0\zeta^\nu\right\}_{t'=t^*} + \int da'\left\{\mathbf{n}\cdot\sum_{\sigma=0}^{\infty}\sum_{\alpha=0}^{\infty}\frac{(-)^\alpha}{(\alpha+\sigma+1)!}\times[(\nabla^{t'}\cdot)^\alpha\rho_0\zeta^{\alpha+\sigma+1}](\cdot\nabla')^\sigma\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right\}_{t'=t^*} \tag{55}$$

where $\mathbf{r}$ is not within the distance $|\zeta|$ of the initial boundary. This expression is also immediately obtained on using $\hat{\rho}$ of Eq. (20) in (46). Equation (55) is then more compactly written as

$$\phi(\mathbf{r},t)=\frac{1}{4\pi\epsilon}\int\frac{d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}[\hat{\rho}(\mathbf{r}',t')]_{t'=t^*}. \tag{56}$$

The vector potential is calculated using either of these methods. The result can be written

$$\mathbf{A}(\mathbf{r},t)=\frac{\mu}{4\pi}\int\frac{d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}[\hat{\mathbf{J}}(\mathbf{r}',t')]_{t'=t^*}, \tag{57}$$

where $\hat{\mathbf{J}}$ is given[25] by (36).

The fact that ϕ and $\mathbf{A}$ can be calculated correctly by using $\hat{\rho}$ and $\hat{\mathbf{J}}$ justifies the use of the equivalent surface charge and current densities in calculating the electromagnetic fields. More specifically, if ρ_0, $\mathbf{J}_0$, and ζ are given, the *true* charge density ρ and the current density $\mathbf{J}$ are uniquely determined. The scalar potential is correctly calculated using (55), provided the series in (55) converges and $\mathbf{r}$ is chosen within the domain where (55) is relevant. The vector potential is treated in a like manner. The fields are then calculated on application of (48) and (49), and are denoted by $\mathbf{E}_1$ and $\mathbf{B}_1$. An entirely distinct problem is defined by asking for the scalar and vector potentials generated by $\hat{\rho}$ and $\hat{\mathbf{J}}$ of (20) and (36), respectively. This is accomplished by the use of (56) and (57) and gives rise to the fields $\mathbf{E}_2$ and $\mathbf{B}_2$ on use of (48) and (49). These fields are then the exact solution of this problem. As suggested previously, $\mathbf{E}_1$ equals $\mathbf{E}_2$ and $\mathbf{B}_1$ equals $\mathbf{B}_2$ if $\mathbf{r}$ is chosen in the region where $\mathbf{E}_1$ and $\mathbf{B}_1$ are accurately known, viz., at points removed from the boundary of ρ_0 by the distance $|\zeta|$. It is often possible to calculate $\mathbf{E}_2$ and $\mathbf{B}_2$ by some method other than through the use of (56) and (57).[26] Since the fields $\mathbf{E}_2$ and $\mathbf{B}_2$ are unique,[27] $\mathbf{E}_2$ and $\mathbf{B}_2$ then represent $\mathbf{E}_1$ and $\mathbf{B}_1$ for $\mathbf{r}$ removed from the boundary of ρ_0 by the distance $|\zeta|$.

VII. DISCUSSION

Many dynamical quantities can be calculated by using the method presented. The charge conservation argument is clearly an important part of the procedure and consequently we are usually restricted to calculating quantities which are proportional to the charge density $\rho(\mathbf{r},t)$ or the current density $\mathbf{J}(\mathbf{r},t)$. The kinetic energy density $\frac{1}{2}\rho v^2/\eta$ and the kinetic energy flux $\frac{1}{2}(\rho v^2/\eta)\mathbf{v}$ are

[25] These potentials are readily seen to satisfy the Lorentz condition, provided $\hat{\rho}$ and $\hat{\mathbf{J}}$ satisfy the continuity equation. See, for example, reference 11, Sec. 8.2.

[26] An expanison in eigenfunctions appropriate to the geometry of the boundary surface of ρ_0 may be useful in many cases. The boundary conditions required of the fields at the boundary of ρ_0 are then properly written in terms of the discontinuities in $\mathbf{E}$ and $\mathbf{B}$ generated by the surface charge and current complexes of (20) and (36).

[27] Reference 11, Sec. 9.2.

of this nature and are readily calculated to yield volume and surface terms.[28] In these examples, the velocity $\mathbf{v}$ is involved. As in Sec. V, we first apply Fourier analysis and the charge conservation argument, and then use Eq. (34) for the velocity. Integral properties of the charge and current densities such as the total angular momentum $\int \rho(\mathbf{r}\times\mathbf{v})/\eta d\tau$ and moment of inertia $\frac{1}{2}\int(\rho r^2/\eta)d\tau$ can be handled similarly except that the Fourier analysis is not necessary.

In applications to the theory of electron tubes and plasmas, the displacement vector $\boldsymbol{\zeta}$ is the solution to an equation of motion where the driving term is the Lorentz force $e(\mathbf{E}+\mathbf{v}\times\mathbf{B})$. The fields $\mathbf{E}$ and $\mathbf{B}$ may be due to the charge and current densities and are therefore dependent on $\boldsymbol{\zeta}$. We can calculate $\mathbf{E}$ and $\mathbf{B}$ according to Eqs. (56), (57), (48), and (49) reliably only if the point $\mathbf{r}$ at which the fields are required is not within the distance $|\boldsymbol{\zeta}|$ of the boundary of the initial charge distribution. However, solution of the equation of motion requires knowledge of the fields *everywhere* within the displaced boundary shown in Fig. 1. If there are no discontinuities in the charge and current densities in the displaced configuration other than at the boundary, the actual fields are continuous and have continuous derivatives within the displaced boundary. Hence it may be postulated that our interior solutions for the fields can be analytically continued into the crest-like region C of Fig. 1. Similarly, the fields in the region T of Fig. 1 may be evaluated by continuation of the exterior solutions. If this procedure is valid, the entire electrodynamic problem can in principle be solved without encountering the usual difficulties associated with the motion of the boundary.

The electromagnetic power $\int(\mathbf{E}\cdot\mathbf{J})d\tau$ in Poynting's theorem can be treated according to the procedure discussed above. In this case we write

$$\int(\mathbf{E}\cdot\mathbf{J})d\mathbf{r}=\int_V d\mathbf{r}'\{\mathbf{J}_0+\rho_0(\partial\boldsymbol{\zeta}/\partial t)+(\mathbf{J}_0\cdot\nabla')\boldsymbol{\zeta}\}\cdot\mathbf{E}[\mathbf{r}'+\boldsymbol{\zeta}(\mathbf{r}',t)]. \quad (58)$$

For values of $\mathbf{r}$ near the initial boundary S, we again require $\mathbf{E}$ in the troublesome region. If analytic continuation of the interior fields is permissible, the power given by (58) may be calculated entirely in terms of ρ_0, $\mathbf{J}_0$, and $\boldsymbol{\zeta}$, or alternatively, in terms of the surface and volume distributions of charge and current.

Use of the equivalent surface charge and current complex proves to be a powerful technique in solving certain difficult boundary value problems. It is clear that these surface quantities are purely the result of mathematical operations, and consequently their employment must be justified in all cases. Other equivalent surface and volume distributions of charge and current exist which are equally useful. The ambiguity is generated by transforming surface terms into volume terms and vice versa. As usual, terms may be added or subtracted which produce no change in the calculated value of physical quantities. The representation chosen here is associated with the method of integration developed in Sec. II. We feel that this representation is consistent in itself and of a *special nature*, since it gives the volume charge and current densities in a form which in no way depends on the existence of a sharp boundary for the initial charge and current distributions.

ACKNOWLEDGMENTS

It is a pleasure to extend our acknowledgment to Dr. E. L. Chu for many helpful discussions and criticisms. During the early stages of this work, we had several profitable discussions with Professor P. A. Sturrock for which we are grateful. We would also like to thank Professor M. Chodorow for his interest and encouragement. Work with the manuscript was greatly facilitated by the efforts of Mrs. Geo. Wada, Miss Sue Zentell, and A. S. Braun.

[28] η is equal to the ratio of charge to mass of the charged particles, hence ρ/η is the mass density.

Eigenvalues and Eigenvectors of a Symmetric Matrix of 6j Symbols*

MORRIS E. ROSE† AND C. N. YANG‡
Brookhaven National Laboratory, Upton, New York
(Received July 27, 1961)

A real orthogonal symmetrical matrix M is defined. It represents the transformation between two coupling schemes for the addition of the angular momenta b, a, b to form a'. The eigenvectors and eigenvalues of M are found. A physical application is briefly discussed.

CONSIDER the matrix M with elements defined by

$$M_{ll'}=[(2l+1)(2l'+1)]^{\frac{1}{2}}\begin{Bmatrix} a & b & l \\ a' & b' & l' \end{Bmatrix}$$

for arbitrary integral or half integral a, b, a', b' for which the $6j$ symbols[1] do not identically vanish. It is well known that M is real and orthogonal. Therefore its eigenvalues have modulus unity. We specialize to the case $b=b'$. For such a case M is symmetric. Hence the eigenvalues of M are ± 1. To find the eigenvectors of M we use a well-known relation[2]

$$\sum_{l_3}(-1)^{j+j_3+l_3}(2l_3+1)\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}\begin{Bmatrix} j_1 & l_1 & j \\ j_2 & l_2 & l_3 \end{Bmatrix} = \begin{Bmatrix} j_1 & j_2 & j_3 \\ l_2 & l_1 & j \end{Bmatrix}. \quad (1)$$

Putting $j_1=a$, $j_2=a'$, $j_3=k$, $l_1=l_2=b$, $l_3=l'$, $j=l$ one obtains

$$\sum_{l'}(-1)^{l+k+l'}(2l'+1)\begin{Bmatrix} a & a' & k \\ b & b & l' \end{Bmatrix}\begin{Bmatrix} a & b & l \\ a' & b & l' \end{Bmatrix} = \begin{Bmatrix} a & a' & k \\ b & b & l \end{Bmatrix}. \quad (2)$$

Multiplying (2) by $(-1)^l(2l+1)^{\frac{1}{2}}$ one obtains

$$\sum_{l'}(-1)^{2l+k+l'}(2l'+1)^{\frac{1}{2}}\begin{Bmatrix} a & a' & k \\ b & b & l' \end{Bmatrix}M_{ll'} = (-1)^l(2l+1)^{\frac{1}{2}}\begin{Bmatrix} a & a' & k \\ b & b & l \end{Bmatrix}. \quad (3)$$

Now $a+b+l=$integer for the $6j$ symbols to be non-vanishing. Hence one may put $(-1)^{2l}=(-1)^{-2a-2b}$. Thus (3) shows that

$$\sum_{l'} M_{ll'}N_{l'k}=(-1)^{+2a+2b-k}N_{lk}, \quad (4)$$

where N_{lk} is a real orthogonal matrix defined by

$$N_{lk}\equiv(-1)^l(2l+1)^{\frac{1}{2}}(2k+1)^{\frac{1}{2}}\begin{Bmatrix} a & a' & k \\ b & b & l \end{Bmatrix} \quad 2(a+b)=\text{even},$$
$$N_{lk}\equiv(-1)^{l+\frac{1}{2}}(2l+1)^{\frac{1}{2}}(2k+1)^{\frac{1}{2}}\begin{Bmatrix} a & a' & k \\ b & b & l \end{Bmatrix} \quad 2(a+b)=\text{odd}. \quad (5)$$

Equation (4) allows for an easy diagonalization of M:

$$\sum_{l,l'} N_{k'l}M_{ll'}N_{l'k}=(-1)^{2a+2b-k}\delta_{kk'}, \quad (6)$$

giving all eigenvectors of M.

One physical application of this result is as follows. Consider the implication of Pomeranchuk's theorem[3] for the high energy limit of cross sections where isotopic spin invariance is valid. For a collision between a particle A (of isotopic spin a) and a particle B (of isotopic spin b), if B and its antiparticle belong to the same isotopic spin multiplet (hence $b=$integer), it can be shown[4] that the number n of linearly independent cross sections at infinite energy is equal to the number of eigenvalues $(-1)^{2a}$ for the matrix M for the case $a=a'$, $b=b'$. Equation (6) then implies that

for $a\geqq b$ (=integer), $n=b+1$,
for $a\leqq b$ (=integer), $n=a+\frac{1}{2}$ for $2a=$odd,
$n=a+1$ for $2a=$even.

ACKNOWLEDGMENT

We wish to thank the members of the Physics Department of the Brookhaven National Laboratory for the hospitality extended to us during our visit.

* Work performed under the auspices of the U. S. Atomic Energy Commission.
† Summer visitor from Oak Ridge National Laboratory. Permanent address after September, 1961: Department of Physics, University of Virginia, Charlottesville, Virginia.
‡ Permanent address: Institute for Advanced Study, Princeton, New Jersey.
[1] See for example, M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957). The relationship between the $6j$ symbols and the W coefficients is

$$\begin{Bmatrix} a & b & c \\ a' & b' & c' \end{Bmatrix}=(-1)^{a+b+b'+a'}W(abb'a',cc').$$

[2] See, e.g., M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Worten, Jr., *The 3-j and 6-j Symbols* (MIT, Cambridge, Massachusetts, 1959), p. 15.
[3] I. Ia. Pomeranchuk, Soviet Phys.—JETP **34**, 499 (1958).
[4] C. N. Yang (to be published).

Inverse Overlap Matrix for Periodic Arrays of Atoms*

T. L. GILBERT
Argonne National Laboratory, Argonne, Illinois
(Received July 18, 1961)

The calculation of the inverse overlap matrix for an infinite chain of single orbital atoms is reduced to the problem of calculating the roots of a polynomial of degree n constructed from the overlap integrals between n neighbors. The familiar method of diagonalizing the overlap matrix, inverting it, and then transforming back to the original representation is used. The final transformation leads to contour integrals which can be evaluated by the method of residues. This diagonalization method is shown to be useful also for finite chains of atoms with Born von Kármán boundary conditions, for chains of atoms with more than one orbital per atom, and for calculating the inverse root of the overlap matrix of single-orbital atoms when the only overlap is between nearest neighbors. Application of the method to two- and three-dimensional arrays of atoms leads to contour integrals containing branch points which cannot be reduced to simple analytic expressions. However, it is shown that an iterative procedure can be devised which permits the calculation of 2^n terms of the Löwdin expansion for the inverse with only $2n$ matrix multiplications.

1. INTRODUCTION

THE inverse overlap matrix for periodic arrays of atoms plays an important role in the calculation of cohesive energies and band structures of ionic and inert gas crystals.[1-3] It is also required for calculating the distortion which occurs in the orbitals of an atom when the atom is in a crystal.[4] The most generally applicable method for calculating the inverse is the expansion method used extensively by Löwdin.[3] It has the advantage of being applicable to any system, but it also has the serious disadvantage that it does not converge very rapidly unless all of the overlap integrals are very small. In a recent publication by Löwdin, Pauncz, and de Heer, three alternative methods for calculating the inverse overlap matrix of cyclic systems were presented.[5] These methods circumvent the convergence problem, but they have a rather restricted range of usefulness. They become very tedious and complicated when there are many neighbors for which the overlap integrals are non-negligible and they are not easily extended to systems of atoms with several orbitals per atom.

In the following we shall present an alternative method which reduces the calculation of the inverse overlap matrix for an infinite one-dimensional array of single-orbital atoms to the trivial problem of finding the roots of a polynomial of degree n, where n is the number of neighbors for which the overlap integral is non-negligible. The method is a direct application of the well-known diagonalization method, i.e., the overlap matrix is transformed to diagonal form, inverted, and then transformed back to the original representation. A useful byproduct of this method is the information it gives us on the constraints which the overlap integrals must satisfy as a consequence of the requirement that the eigenvalues of the overlap matrix be positive definite. These constraints are discussed in Sec. 3. The simplicity of the method is demonstrated in Sec. 4 where the simple examples of an infinite chain with first- and second-neighbor overlap are treated explicitly. In Sec. 5 it is shown how the inverse overlap matrix for an infinite chain may be used to calculate the inverse overlap matrix for a finite chain of arbitrary length with cyclic boundary conditions. The extension of the method to linear arrays of atoms with several orbitals per atom is presented in Sec. 6. The application of the diagonalization method to the calculation of the inverse root of the overlap matrix for an infinite linear chain is considered in Sec. 7. The contour integrals for the matrix elements of the inverse root contain branch points so that a reduction to a simple algebraic form is not possible. In the case of nearest neighbor overlap only, the integrals reduce to Legendre functions of the second kind.

When one attempts to apply the diagonalization method to two- and three-dimensional arrays, one also obtains contour integrals which contain branch points. The form of the integrand is not very simple, so that it does not appear that the diagonalization method is a practical one for such systems. As an alternative, one may still use the Löwdin expansion method for two- and three-dimensional arrays. In Sec. 9 it is shown how Löwdin's expansion may be obtained by an iteration procedure which makes possible the calculation of 2^n terms of the expansion with only $2n$ matrix multiplications. With this modification, the expansion method becomes a practical one even when the convergence of the direct expansion is rather poor.

2. INVERSE OVERLAP MATRIX FOR AN INFINITE CHAIN OF SINGLE-ORBITAL ATOMS

Let $\varphi_l = \varphi(x_1 - la, x_2, x_3)$ be the atomic orbital on the lth atom of an infinite chain of single-orbital atoms separated from each other by a distance a. The overlap

* Based on work performed under the auspices of the U. S. Atomic Energy Commission.

[1] P. O. Löwdin, *A Theoretical Investigation into Some Properties of Ionic Crystals* (Almqvist and Wiksells, Uppsala, Sweden, 1948).
[2] P. O. Löwdin, J. Chem. Phys. **18**, 365 (1950); **19**, 1579 (1951).
[3] P. L. Löwdin, Advances in Phys. **5**, 1 (1956).
[4] T. L. Gilbert (to be published).
[5] P. O. Löwdin, R. Pauncz, and J. de Heer, J. Math. Phys. **1**, 461 (1960). See also, K. Ruedenberg, J. Chem. Phys. **34**, 1878 (1961).

matrix for this system will be $\mathbf{S}=[S_{l,l'}]$, where $S_{l,l'}=\int\varphi^*_l\varphi_{l'}dv=S_{l-l'}=S^*_{l'-l}$ is a function of the distance between the atoms only, so that the overlap matrix is completely characterized by the overlap integrals, $S_l=S^*_{-l}=\int\varphi^*_l\varphi_0 dv$. We shall, temporarily, assume that the range of the overlap is finite, i.e., $S_l=0$ when $|l|>n$, where n is a finite integer which may, however, be as large as we please.

The overlap matrix may be diagonalized by the unitary transformation, $\mathbf{U}=[(k|U|l)]=[(2\pi)^{-\frac{1}{2}}e^{ikl},]$ where k is a continuous variable in the interval $-\pi<k\leqslant\pi$. The diagonal form is

$$\mathbf{S}'=\mathbf{USU}^\dagger=[(2\pi)^{-1}\textstyle\sum_{ll'}e^{ikl}S_{l-l'}e^{-ik'l'}]$$
$$=[\delta(k-k')s(e^{ik})],$$

where $s(z)=\sum_{-\infty}^{\infty}S_lz^l$ is a function of a complex parameter z which will be referred to as the "overlap function." We may invert the diagonal form obtaining $\mathbf{S}'^{-1}=[\delta(k-k')/s(e^{ik})]$, and then transform from the k representation back to the l representation. The result is

$$\mathbf{S}^{-1}=\mathbf{U}^\dagger\mathbf{S}'^{-1}\mathbf{U}=[S_{l-l'}{}^{-1}], \tag{1}$$

where[6]

$$S^{-1}{}_{l-l'}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-ik(l-l')}dk}{s(e^{ik})}=\frac{1}{2\pi i}\oint\frac{z^{l'-l-1}dz}{s(z)}$$
$$=\frac{1}{2\pi i}\oint\frac{z^{l-l'-1}dz}{s^*(z)}. \tag{2}$$

The first contour integral is the most convenient to use when $l'\geqslant l$; the second when $l\geqslant l'$. The contour is the unit circle in the z plane, traversed counterclockwise, for both integrals.

A direct proof that the contour integral defines the elements of an inverse overlap matrix may be given as follows. We have

$$\textstyle\sum_{l''}S_{l-l''}S^{-1}{}_{l''-l'}$$
$$=\textstyle\sum_{l''}S_{l''}S^{-1}{}_{l-l'-l''}$$
$$=(2\pi i)^{-1}\oint dz\{z^{l'-l-1}(\textstyle\sum_{l''}S_{l''}z^{l''})/s(z)\}$$
$$=(2\pi i)^{-1}\oint dzz^{l'-l-1}=\delta_{l-l'}.$$

This proof is valid as long as $\oint dzz^{l-1}/s(z)$ is bounded for all values of l and the sums are convergent.[7]

The contour integral may be evaluated in a straightforward manner. When all of the roots of $s(z)=0$ are distinct, the result is

$$S^{-1}{}_{l-l'}=\begin{cases}\displaystyle\sum_{i=1}^{n}\frac{w_i{}^{l'-l}}{w_ids(w_i)/dz} & l'\geqslant l\\ \displaystyle\sum_{i=1}^{n}\frac{w_i{}^{l-l'}}{w_ids^*(w_i)/dz} & l\geqslant l'\end{cases} \tag{3a}$$

where $w_1, w_2, \cdots, w_n$ are the roots of $s(z)=0$ (the poles of the integrand) which lie within the unit circle. There will be a total of $2n$ roots of $s(z)=0$ when $S_n\neq 0$ and $S_l=0$ for $|l|>n$, but there will be only n roots within the unit circle because $w_i{}^{-1*}$ must be a root if w_i is a root. If the overlap integrals are all real, Eq. (3a) reduces to

$$S^{-1}{}_{l-l'}=\sum_{i=1}^{n}\frac{w_i{}^{|l-l'|}}{w_is'(w_i)}, \tag{3b}$$

where $s'(z)=ds/dz$. For the sake of simplicity we will henceforth assume that the overlap integrals are real. Modifications for complex overlap integrals (which are unavoidable only if there are imaginary terms in the Hamiltonian, and are of practical importance only when magnetic fields are too strong to be treated as perturbations) are straightforward and need not be considered explicitly.

The restriction that the roots be distinct is not essential. If two or more roots come into confluence within the unit circle, then $s'(z)$ will vanish at this point and the corresponding terms in Eqs. (3) will become infinite, but in such a manner that their sum remains finite. For example, if $w_q\to w_r=w$, then $s'(w_q)\to 0$ and $s'(w_r)\to 0$, but

$$\frac{w_q{}^{|l-l'|}}{w_qs'(w_q)}+\frac{w_r{}^{|l-l'|}}{w_rs'(w_r)}\to w^{|l-l'|}\left\{\frac{2(|l-l'|-1)}{w^2s^{(2)}(w)}-\frac{2}{3}\frac{w^3s^{(3)}(w)}{[w^2s^{(2)}(w)]^2}\right\},$$

where $s^{(n)}(z)=d^ns/dz^n$.

When the overlap integrals are real, the calculation of the inverse overlap matrix elements may be reduced to the problem of calculating the roots of a polynomial of nth degree. This is done by means of the substitution $z=u[1-(1-u^{-2})^{\frac{1}{2}}]$, which gives

$$r(u)\equiv s[z(u)]=1+\sum_{l=1}^{n}\sum_{m=0}^{[\frac{1}{2}l]}\binom{l}{2m}u^{l-2m}(u^2-1)^m$$
$$=1+2uS_1+(4u^2-2)S_2+(8u^3-6u)S_3$$
$$+(16u^4-16u^2+2)S_4+\cdots, \tag{4}$$

where $[\frac{1}{2}l]$ denotes the largest integer in $\frac{1}{2}l$. The roots of the polynomial $s(z)$, of degree $2n$, may be obtained directly from the roots of the polynomial $r(u)$, of degree n.

[6] The convention $s^*(z)=[s(z^*)]^*$ is used, i.e., $s^*(z)=\Sigma_{-\infty}{}^{\infty}S_l{}^*z^l$.

[7] This proof is somewhat deceptive because it remains valid for any contour which includes the origin, thereby suggesting that the inverse is not unique. The inverse is unique, however, if we demand that $\mathbf{S}^{-1}\to\mathbf{1}$ as $\mathbf{S}\to\mathbf{1}$.

The restriction to a finite range of overlap may be removed, provided that the infinite sum,

$$S^{-1}{}_{l-l'}=\sum_{i=1}^{\infty}\frac{w_i^{|l-l'|-1}}{s'(w_i)} \tag{5}$$

converges, where w_i are the (infinitely many) zeros of the overlap function, $s(z)=\sum_{-\infty}^{\infty}S_l z^l$. The restriction on the range of the overlap cannot be removed before evaluating the contour integral because $s(z)$ has an essential singularity at the origin when the range of the overlap is infinite. But once Eq. (2) has been evaluated, we obtain an expression for the inverse overlap, [Eqs. (3)], which is well-defined for an arbitrarily large range of overlap and which may also have a well-defined limit as $n\rightarrow\infty$. If this limit exists, then Eq. (5) must define the elements of the inverse overlap matrix.

We have not been able to determine the necessary and sufficient conditions which must be imposed on the overlap matrix elements S_l in order to make the sum converge; hence we will bypass this mathematical problem by assuming that the sum does converge for all overlap matrices which may be constructed from a set of linearly independent orbitals. (The proof of this statement for an infinite chain with an infinite range of overlap does not appear to be a trivial problem.) This assumption covers all overlap matrices of interest.

3. CONSTRAINTS ON THE OVERLAP INTEGRALS

The eigenvalues of an overlap matrix constructed from a finite set of orbitals are always non-negative and are positive definite if the orbitals are linearly independent.[8] This same statement also applies to the overlap matrix for an infinite linear chain of simple atoms.[9] We can use this fact to derive certain inequalities which the overlap integrals must satisfy.

The eigenvalues of the overlap matrix are the values of the overlap function on the unit circle. Hence, if the basis orbitals φ_l are linearly independent, the overlap function must be positive definite on the unit circle.

$$s(e^{ik})=\sum_{-\infty}^{\infty}S_n e^{ink}>0. \tag{6}$$

Equation (6) may be rewritten as

$$\sum_1^{\infty}S_n\cos nk>-\tfrac{1}{2}. \tag{7}$$

The extrema of the left-hand side of Eq. (7) will occur at the zeros of the function $s'(z)$ which lie on the unit circle. When the overlap integrals are all real, there will always be zeros at $z=\pm1$, because $zs'(z)+z^{-1}s'(z^{-1})=0$. These points will, in many cases of interest, correspond to the maximum and minimum values of the left-hand side of Eq. (7). At these points, Eq. (7) reduces to:

$$-\sum_1^{\infty}S_n<\tfrac{1}{2} \qquad (z=1,\ k=0), \tag{8}$$

$$\sum_1^{\infty}S_{2n-1}-\sum_1^{\infty}S_{2n}<\tfrac{1}{2} \qquad (z=-1,\ k=\pi). \tag{9}$$

Equation (8) is the relevant restriction when the nearest-neighbor overlap integrals are dominant and negative. Equation (9), which is the relevant inequality when the nearest-neighbor overlap integrals are dominant and positive, is of particular interest. The nearest-neighbor overlap, when positive, can exceed the value $\frac{1}{2}$. However, the inequality must always be satisfied; hence the sum of the more distant neighbor overlaps must be greater than the amount by which the nearest neighbor overlap exceeds $\frac{1}{2}$. This result indicates that the neglect of more distant neighbor overlap when the nearest-neighbor overlap is large may lead to serious errors. The error may be particularly serious for three-dimensional arrays of atoms.

As the orbitals of the chain approach linear dependence, the zeros of $s(z)$ will move out toward the unit circle. In the limiting case of linear dependence at least one of the zeros inside the unit circle will become confluent on the unit circle with the corresponding zero outside the unit circle. The contour integral which defines the inverse matrix elements will then become infinite because the contour lies between the confluent poles of the integrand. However, if (by artificial construction of a matrix which cannot be constructed from overlap integrals) the overlap matrix elements are increased further, the zeros will move apart along the unit circle and the inverse matrix can again be defined by choosing an appropriate contour which is deformed into "hooks" about the isolated poles on the unit circle. The inverse matrix elements will then be periodic and nondecreasing instead of exponentially decreasing [see Eqs. (3)]. The inverse of an infinite periodic matrix will, therefore, be defined under very general conditions and will be a rather complicated function of the elements of the original matrix, with many poles and branch points. A pole of the inverse matrix elements (regarded as functions of the direct matrix elements) will occur whenever the zeros of $s(z)$ become confluent *on the unit circle*. These poles are also branch points, with the various branches corresponding to the various ways in which the hook integrals can be chosen to avoid the separated zeros of $s(z)$ on the unit circle. This extended domain of the matrix elements is of interest in connection with the Green's operator for a linear chain, which is useful for calculating the orbitals and energy levels associated with localized impurities.[10,11] The Green's operator is $\mathbf{G}(E)=(\mathbf{H}-E)^{-1}$, where $\mathbf{H}$ is the

[8] The proof proceeds as follows. Let $\mathbf{a}^{(k)}$ and $s^{(k)}$ be the eigenvectors and eigenvalues of $\mathbf{S}=[\int\varphi_i^*\varphi_j d\mathbf{r}]$; i.e., $\Sigma_j S_{ij}a_j^{(k)}=s^{(k)}a_i^{(k)}$. Let $\psi_k(\mathbf{r})=\Sigma_i\varphi_i(\mathbf{r})a_i^{(k)}$ be the (unnormalized) "canonical orbitals" constructed from these eigenvectors. It can readily be shown that $\int|\psi_k|^2d\mathbf{r}=s^{(k)}$. Hence, $s^{(k)}$ must be non-negative. If s^k vanishes, then $\Sigma_i\varphi_i(\mathbf{r})a_i^{(k)}=0$ and the set $\{\varphi_i\}$ is linearly dependent. This proof is valid as long as k is a discrete variable. It breaks down for an infinite linear chain because k is then a continuous variable in the interval $-\pi\leq k<\pi$ and the ψ_k are not normalizable.

[9] See Appendix.

[10] M. Lax, Phys. Rev. **94**, 1391 (1954).

[11] G. F. Koster and J. C. Slater, Phys. Rev. **94**, 1392 (1954); **95**, 1167 (1955).

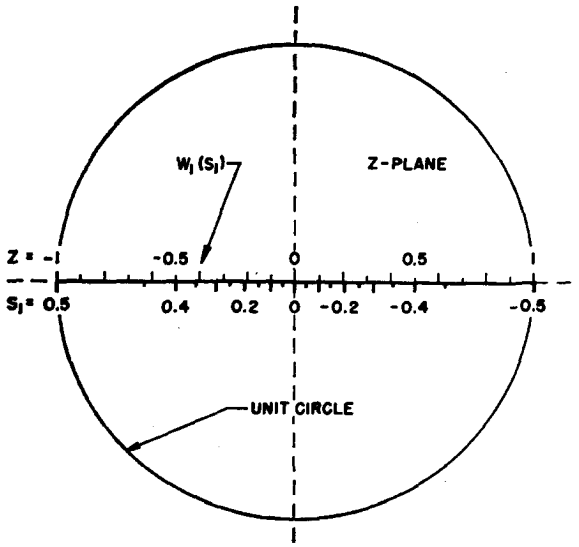


FIG. 1. Locus of the root of the overlap function for nearest-neighbor overlap. S_1=overlap integral, w_1=root of the overlap function, $s(z)$=overlap function.

Hamiltonian of the linear chain in the LCAO approximation. If $\mathbf{H}$ is expressed in a discrete representation, $\mathbf{H}=[H_{l-l'}]$, where $H_{l-l'}=\int\psi^*_l H\psi_{l'}dv$ are the matrix elements of the one-electron Hamiltonian H with respect to the localized orthogonal orbitals $\psi_l=\sum_{l'}\varphi_{l'}S^{-\frac{1}{2}}_{l'-l}$, then the calculation of $\mathbf{G}(E)$ proceeds in exactly the same manner as the calculation of $\mathbf{S}^{-1}$. The only essential difference is that the overlap function $s(z)=\sum_{-\infty}^{\infty}S_l z^l$ is replaced by the characteristic function $h(z)=\sum_{-\infty}^{\infty}H_l z^l-E$, which will have zeros lying on the unit circle when E lies within the energy band. When E lies in a band gap, then $\mathbf{G}(E)$ has the same properties as $\mathbf{S}^{-1}$, and will be a single-valued and analytic function of E (and of H_l). The poles of $\mathbf{G}(E)$ will occur at the band edges. When E lies within the band, there will be various branches of $G(E)$ which correspond to the various types of scattered waves. We shall not pursue these interesting ramifications any further in this paper. All overlap integrals lie within the domain within which the inverse matrix elements are single valued and analytic functions of the original matrix elements.

4. TWO EXAMPLES

The first example is a system with nearest-neighbor overlap only; i.e., $S_l=0$ for $l>1$. Equation (3) then reduces to

$$S^{-1}{}_{l-l'}=w^{|l-l'|}/S_1(w-w^{-1}), \tag{10}$$

where

$$w=-[1-(1-4S_1^2)^{\frac{1}{2}}]/2S_1.$$

The constraints imposed by linear independence are $|w|<1$ and $|S_1|<\frac{1}{2}$. If $S_1=\pm\frac{1}{2}$, so that $w=\mp1$, then $\sum_{-\infty}^{\infty}(\mp1)^l\varphi_l=0$. We observe that, if $S_1>\frac{1}{2}$ (which is not possible for the nearest-neighbor overlap integral when more distant overlap integrals vanish), the inverse matrix is still given by Eq. (10) if we set $w=-[1-i(4S_1^2-1)^{\frac{1}{2}}]/2S_1\equiv e^{i\theta}$. The inverse matrix elements would then be periodic and nondecreasing instead of exponentially decreasing.

The locus of the zeros of $s(z)$ is shown in Fig. 1 for $0\leqslant|S_1|<\frac{1}{2}$, and the inverse overlap matrix elements for various values of S_1 are shown in Fig. 2 [note that $S_n^{-1}=(-1)^n|S_n^{-1}|$].

The second example is that of a system with first- and second-neighbor overlap only. Substituting

$$z=u[1-(1-u^{-2})^{\frac{1}{2}}]$$

into the overlap function we obtain the polynomial $r(u)=4S_2u^2+2S_1u+1-2S_2$. The roots of this polynomial are

$$\begin{aligned}v_1&=[-S_1+(S_1^2+8S_2^2-4S_2)^{\frac{1}{2}}]/4S_2,\\ v_2&=[-S_1-(S_1^2+8S_2^2-4S_2)^{\frac{1}{2}}]/4S_2.\end{aligned} \tag{11}$$

Both roots are real when $0\leqslant S_2\leqslant A\equiv\frac{1}{4}[1-(1-2S_1^2)^{\frac{1}{2}}]$. The roots become confluent when $S_2=A$ and are complex conjugate when $A_a<S_2<B=\frac{1}{4}[1+(1-2S_1^2)^{\frac{1}{2}}]$. When $S_2=B$ the zeros of $s(z)$ reach the unit circle and the inverse becomes infinite. The finite zeros of $s(z)$ are

$$\left.\begin{aligned}w_1&=v_1[1-(1-v_1^{-2})^{\frac{1}{2}}]\\ w_2&=v_2[1-(1-v_2^{-2})^{\frac{1}{2}}]\end{aligned}\right\}0\leqslant S_2\leqslant A,$$
$$\left.\begin{aligned}w_1&=w_r+iw_i\\ w_2&=w_r-iw_i\end{aligned}\right\}A<S_2<B, \tag{12}$$

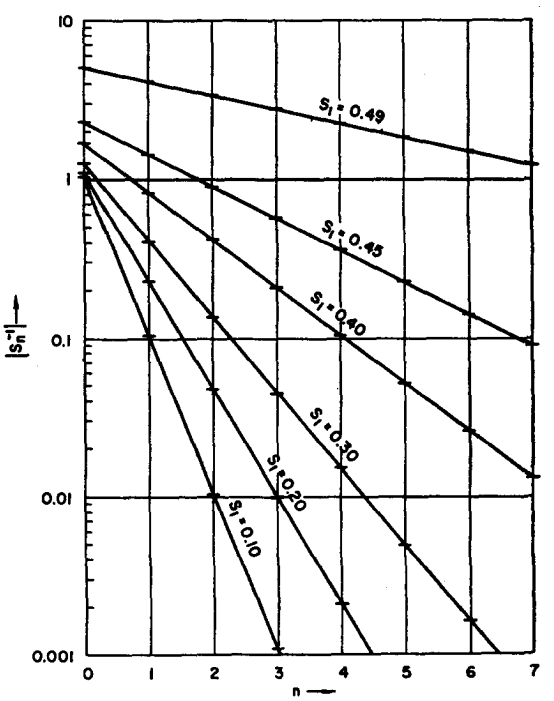


FIG. 2. Magnitude of the elements of the inverse overlap matrix for nearest-neighbor overlap only. S_1=overlap integral, $S_n^{-1}=(-1)^n|S_n^{-1}|$=inverse overlap matrix elements.

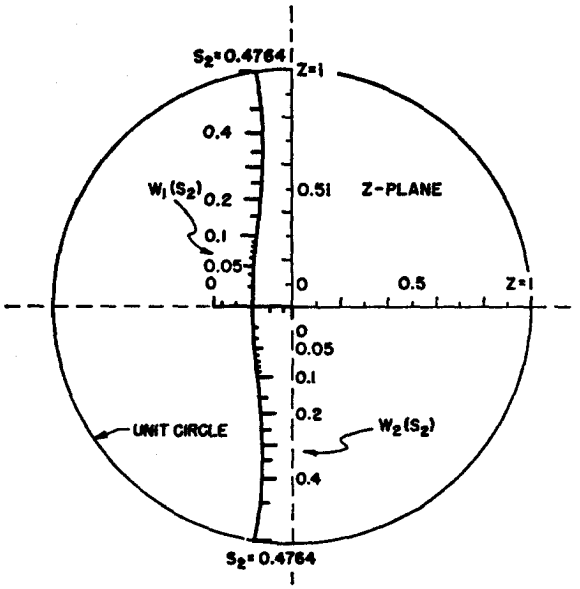


FIG. 3. Locus of the roots of the overlap function when second-neighbor overlap is included. S_1 and S_2 are the first- and second-neighbor overlap integrals, and w_1 and w_2 are the roots of the overlap function. The locus is shown for $S_1=0.3$ and $0\leqslant S_2\leqslant 0.4764$. The two roots are confluent at $z=-0.1616$ when $S_2=0.02362$.

where

$$w_r=v_r[1-\{\tfrac{1}{2}[(f^2+g^2)^{\frac{1}{2}}+f]v_r^{-2}\}^{\frac{1}{2}}],$$
$$w_i=v_i[1-\{\tfrac{1}{2}[(f^2+g^2)^{\frac{1}{2}}-f]v_i^{-2}\}^{\frac{1}{2}}], \tag{13}$$

with $v_r=-S_1/2S_2$, $v_i=(4S_2-S_1^2-8S_2^2)^{\frac{1}{2}}/4S_2$, $f=v_r^2-v_i^2-1$, and $g=2v_iv_r$. The inverse overlap matrix elements are, explicitly,

$$S^{-1}{}_{l-l'}=\frac{w_1{}^{|l-l'|}}{S_1(w_1-w_1^{-1})+2S_2(w_1^2-w_1^{-2})}+\frac{w_2{}^{|l-l'|}}{S_1(w_2-w_2^{-1})+2S_2(w_2-w_2^{-2})}$$
$$=\begin{cases}\dfrac{1}{2S_1[1+(8S_2^2-4S_2)/S_1^2]^{\frac{1}{2}}}\times\left[\dfrac{w_1{}^{|l-l'|}}{(v_1^2-1)^{\frac{1}{2}}}-\dfrac{w_2{}^{|l-l'|}}{(v_2^2-1)^{\frac{1}{2}}}\right] & 0\leqslant S_2\leqslant A\\ |w|^{|l-l'|}\cos(n\theta+\varphi)/R & A<S_2<B,\end{cases} \tag{14}$$

where $|w|=(w_r^2+w_i^2)^{\frac{1}{2}}$, $\theta=\tan^{-1}(w_i/w_r)$,

$$\varphi=(v_rv_i/|v_rv_i|)\tan^{-1}[(f^2+g^2)^{\frac{1}{2}}+f]^{\frac{1}{2}}/[(f^2+g^2)^{\frac{1}{2}}-f]^{\frac{1}{2}},$$

where n must be chosen so that $\cos\varphi>0$, and $R=(f^2+g^2)^{\frac{1}{4}}(4S_2-S_1^2-8S_2^2)^{\frac{1}{4}}$. The locus of the zeros of $s(z)$ for $S_1=0.3$ and $0\leqslant S_2\leqslant 0.476$ is shown in Fig. 3. The dependence of the inverse overlap matrix elements on S_2 is shown in Fig. 4.

5. INVERSE OVERLAP MATRIX FOR A FINITE CHAIN OF SINGLE-ORBITAL ATOMS WITH PERIODIC BOUNDARY CONDITIONS

The inverse overlap matrix for a finite chain of l_0 single-orbital atoms with periodic boundary conditions, i.e., $S_{l-l'}=S_{l-l'\pm l_0}$, may be calculated by the same general procedure of first diagonalizing the overlap matrix, then inverting it, and finally transforming it back to the original representation. In order to avoid irrelevant complications, we assume that the overlap vanishes between neighbors which are half a period or more apart, i.e., $S_{l-l'}=0$ when $|l-l'|\geqslant[\frac{1}{2}l_0]$ where $[\frac{1}{2}l_0]$ denotes the largest integer in $\frac{1}{2}l_0$. (This restriction imposes the requirement that $l_0\geqslant 3$ for nearest-neighbor overlap, $l_0\geqslant 5$ for next-nearest-neighbor overlap, etc.)

Let $S^{-1}{}_{(l_0),l-l'}$ be an inverse overlap matrix element for a finite chain with period l_0 and let $S^{-1}{}_{l-l'}=S^{-1}{}_{(\infty),l-l'}$ be an inverse overlap matrix element for an infinite chain. The results of the calculation for the finite chain can be expressed in the form,

$$S^{-1}{}_{(l_0),l-l'}=\sum_{n=-\infty}^{\infty}S^{-1}{}_{l-l'+nl_0} \tag{15}$$

so that the inverse overlap matrix for a finite chain with periodic boundary conditions can be calculated immediately from the inverse overlap matrix for an infinite chain.

The constraints on the overlap integrals for an infinite chain will be different from the constraints on

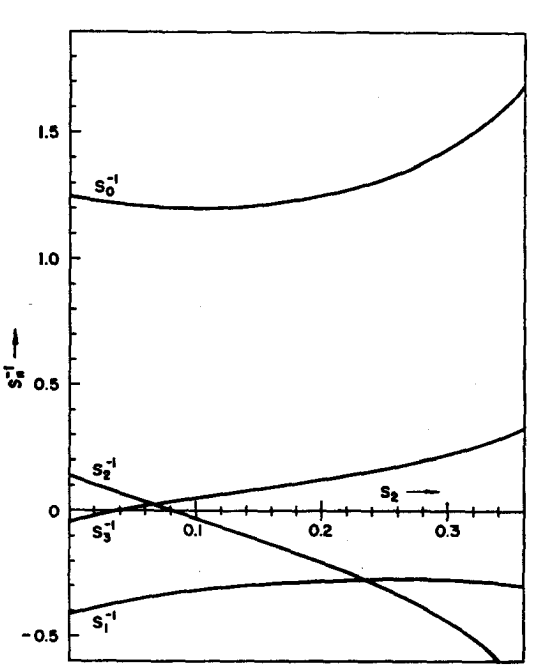


FIG. 4. Dependence of the inverse overlap matrix elements on the second-neighbor overlap for a fixed value of the first-neighbor overlap. $S_n{}^{-1}$=inverse overlap matrix elements, $S_1=0.3$=first-neighbor overlap, S_2=second-neighbor overlap.

the overlap integrals for a finite chain because the eigenvalues of the overlap matrix for a finite chain are the values of the overlap function at the discrete set of points, $z=e^{2\pi ik/l_0}$, where $k=0, 1, 2, \cdots, l_0-1$. The overlap function can vanish on the unit circle as long as it does not vanish at any one of these discrete points. In particular, if the number of atoms is odd, a zero of $s(z)$ can occur at the point $z=-1$ without violating the requirement that the eigenvalues of the overlap matrix be positive. The constraints given by Eq. (9) are therefore inapplicable to a finite chain with an odd number of atoms. Equation (8) remains applicable to both finite and infinite chains.

The significance of the modification in the constraints may be seen by noting that a triatomic homonuclear molecule with the nuclei located at the vertices of an equilateral triangle (i.e., a finite chain with $l_0=3$), is the one unique cyclic system in which second and more distant neighbor overlap integrals can vanish when the nearest-neighbor overlap exceeds $+\frac{1}{2}$. The limiting value of the nearest-neighbor overlap for this system is $S_1=1$. As S_1 increases from 0 to $\frac{1}{2}$, the smallest zero of $s(z)$ moves along the real axis from $z=0$ to $z=-1$. The individual terms in Eq. (15) diverge when $S_1=\frac{1}{2}$, but the sum does not. As S_1 increases from $\frac{1}{2}$ to 1, the zeros move apart along the unit circle. As $S_1\rightarrow 1$, the zeros of $s(z)$ approach the limiting values $w=e^{\pm 2\pi i/3}$ and the orbitals become linearly dependent. When the overlap is negative, the limiting value is $S_1=-\frac{1}{2}$, $(w=+1)$, and the orbitals become linearly dependent when this limit is attained.

This behavior may be seen more clearly from the general form of the inverse overlap matrix for a finite chain with nearest-neighbor overlap only, which is

$$S^{-1}{}_{(l_0),l-l'}=\frac{w^{|l-l'|}+w^{l_0-|l-l'|}}{S_1(w-w^{-1})(1-w^{l_0})}, \tag{16}$$

where w is given by Eq. (10). It is clear that this expression becomes infinite when $w\rightarrow+1$. If l_0 is even, it also becomes infinite when $w\rightarrow-1$. However, when l_0 is odd ($l_0=2N+1$), it can be written in the form

$$S^{-1}{}_{(2N+1),l-l'}\quad\frac{w^{|l-l'|+1}\sum_{r=0}^{2N-2|l-l'|}(-w)^r}{S_1(w-1)(1-w^{2N+1})} \tag{17}$$

which is obviously bounded when $w=-1$.

6. INVERSE OVERLAP MATRIX FOR A CHAIN OF MANY-ORBITAL ATOMS

In Sec. 2 we reduced the problem of calculating the elements of the inverse overlap matrix for an infinite chain of single-orbital atoms to the problem of calculating the poles of the inverse of the overlap function. When the range of the overlap is finite, the problem reduces to the simple algebraic problem of finding the roots of a polynomial.

The calculation of the inverse overlap matrix for a chain of atoms with more than one orbital per atom can also be reduced to the problem of finding the poles of certain functions; however, the functions are obtained by inverting matrices rather than functions, so that the problem of determining the poles is not so trivial. In spite of this complication, the method should be feasible when all but one or two orbitals have a rather small overlap so that perturbation techniques can be used for determining the positions of the poles.

Let φ_j, $j=1, 2, \cdots, r$, be a set of localized orbitals centered on a given atom. The basis orbitals for the chain will be $\varphi_{jl}(x_1,x_2,x_3)=\varphi_j(x_1-la, x_2, x_3)$. The overlap matrix will be $\mathbf{S}=[S_{jl,j'l'}]$ where $S_{jl,j'l'}=\int\varphi^*_{jl}\varphi_{j'l'}dv=S_{l-l',jj'}=S^*_{l'-l,j'j}$. The overlap matrix may be partially diagonalized by using the unitary transformation $\mathbf{U}=[(jk|U|j'l)]=[\delta_{jj'}(2\pi)^{-\frac{1}{2}}e^{ikl}]$. We obtain $\mathbf{S'}=\mathbf{USU}^\dagger=[\delta(k-k')\mathbf{s}(e^{ik})]$. The quantity, $\mathbf{s}(z)=[s_{jj'}(z)]$, is the "overlap function" which is now a matrix with elements given by

$$s_{jj'}(z)=\sum_{l=-\infty}^{\infty}S_{l,jj'}z^l. \tag{18}$$

The inverse overlap function can be calculated by inverting this matrix for each value of z. Let $\mathbf{s}^{-1}(z)=[s^{-1}{}_{jj'}(z)]$ be this inverse. In the limiting case that $s_{jj'}(z)=\delta_{jj'}s_j(z)$ then $s^{-1}{}_{jj'}=\delta_{jj'}/s_j(z)$. This case will not occur in practice, so that the elements of $\mathbf{s}^{-1}(z)$ will be somewhat more complicated functions which must be calculated by solving the equations

$$\sum_{j''}s_{jj''}(z)s^{-1}{}_{j''j'}(z)=\delta_{jj'}.$$

If the off-diagonal elements are small, it is likely that the elements of the inverse overlap function and, in particular, the location of the poles of these functions, can be determined by applying perturbation techniques to the zero order solution obtained by neglecting the off-diagonal elements. The case where a single off-diagonal element is large should not be too difficult for practical calculations because the inversion of a 2×2 matrix can be carried out explicitly. The difficult cases will be those for which the off diagonal elements of the overlap function are large for several orbital pairs.

Once $\mathbf{s}^{-1}(z)$ has been obtained, the inverse overlap matrix elements can immediately be expressed as the integral

$$S^{-1}{}_{l-l',jj'}=\frac{1}{2\pi i}\oint s^{-1}{}_{jj'}(z)z^{|l-l'|-1}dz. \tag{19}$$

If, as we shall assume, the poles are all distinct and do not lie on the unit circle, and there are no branch points within the unit circle, then the contour integral

can be evaluated by the method of residues. The result is

$$S^{-1}{}_{l-l',jj'}=\sum_{i=1}^{\infty} w_{jj',i}{}^{|l-l'|}/w_{jj',i}s'_{jj'}(w_{jj',i}), \quad (20)$$

where $w_{jj',i}$ are the poles of $s_{jj'}(z)$ which lie within the unit circle and

$$s'_{jj'}(z)=(d/dz)(1/s^{-1}{}_{jj'}(z). \quad (21)$$

7. INVERSE ROOT OF THE OVERLAP MATRIX FOR AN INFINITE CHAIN OF SIMPLE ATOMS

The diagonalization method can also be used to reduce the calculation of the inverse root of the overlap matrix for a simple infinite chain to the evaluation of an integral. Proceeding in the usual manner, we first reduce the elements of the inverse root to the following contour integral around the unit circle.

$$S^{-\frac{1}{2}}{}_{l-l'}=\frac{1}{2\pi i}\oint \frac{z^{|l-l'|-1}}{[s(z)]^{\frac{1}{2}}}dz. \quad (22)$$

The reduction of this integral to an algebraic expression cannot be carried out directly because each pole in the integrand is now a branch point. We may, however, proceed in the following alternative manner. We first "cut" the plane by introducing open curves consisting of straight line segments connecting the branch points of $[s(z)]^{-\frac{1}{2}}$, [i.e., the origin and the zeros of $s(z)$] in pairs. The contour in Eq. (22) may then be "shrunk" to a contour enclosing zero area which traverses each side of the cuts. The integrands on opposite sides of a cut differ only in sign; hence the contour integral may be replaced by a line integral along the cut. The sign of the line integral will depend on the direction of the path of integration and must be determined in such a manner that the diagonal elements of the inverse root are positive.

In the particular case of nearest-neighbor overlap only, the line integral which one obtains is

$$S^{-\frac{1}{2}}{}_{l-l'}=\pm\frac{1}{i\pi}\int_0^w \frac{z^{|l-l'|-1}dz}{[1+S_1(z+z^{-1})]^{\frac{1}{2}}} \quad (23)$$

where $w=-[1-(1-4S_1{}^2)^{\frac{1}{2}}]/2S_1$ is the solitary zero of $s(z)$ which lies within the unit circle. In this particular case we may obtain a closed analytic form by making use of the following integral formula for a Legendre function of the second kind[12]

$$Q_n(z')=\int_0^{z'-(z'^2-1)} \frac{z^n dz}{(z^2-2zz'+1)^{\frac{1}{2}}} \quad \mathcal{R}l(n+1)>0. \quad (24)$$

[12] Whittaker and Watson, *Modern Analysis* (Cambridge University Press, New York, 1943), problem 32, p. 334. In this reference $Q_n(z)$ is defined by the Laplacian integral

$$Q_n(z)=\int_0^{\infty} [z+(z^2-1)^{\frac{1}{2}}\cosh\theta]^{-(n+1)}d\theta$$

in the cut plane. If n is not an integer, Q_n is single valued in the z-plane cut from 1 to $-\infty$ along the real axis.

The analytic expression for $-\frac{1}{2}<S_1\leqslant 0$ is

$$S^{-\frac{1}{2}}{}_{l-l'}=Q_{|l-l'|-\frac{1}{2}}(-1/2S_1)/\pi(-S_1)^{\frac{1}{2}}. \quad (25)$$

This expression can also be used for $0\leqslant S_1<\frac{1}{2}$, provided that the limiting value

$$Q_n(-x)=\lim_{\delta\to 0} Q_n(-x+i\delta)$$

is used. $Q_n(-x)$ will be pure imaginary for $x>0$ so that $S^{-\frac{1}{2}}{}_{l-l'}$ is always real.

The reduction of the integrals for the matrix elements of the inverse root to a simple analytic form in the case of nearest-neighbor interaction appears to be a fortunate happenstance which is unlikely to occur when second- and more-distant-neighbor overlap integrals are included.

8. INVERSE OVERLAP MATRIX FOR TWO- AND THREE-DIMENSIONAL ARRAYS OF ATOMS

Although the diagonalization method is also applicable, in principle, to two- and three-dimensional arrays of atoms, its practical usefulness appears to be limited to one-dimensional arrays. The reason is that the contour integrals which occur in the two- and three-dimensional cases cannot be evaluated in simple closed form. In order to demonstrate the difficulties which arise, let us consider the case of a two dimensional array of atoms.

Let $\varphi_{l_1l_2}(x_1,x_2x_3)=\varphi(x_1-l_1a_1,\ x_2-l_2a_2,\ x_3)$ be the basis orbitals for a two-dimensional array of single-orbital atoms, where $\varphi(x_1,x_2,x_3)$ is the atomic orbital of a single atom expressed in terms of the coordinates x_1, x_2, measured along the directions of the primitive lattice vectors, and the coordinate x_3 measured perpendicular to the plane of the array. The a_i are primitive lattice distances and l_1, $l_2=0$, ±1, ±2, $\cdots$, $\pm\infty$. The overlap matrix will be $\mathbf{S}=[S_{l_1l_2,l_1'l_2'}]$ where $S_{l_1l_2,l_1'l_2'}=\int\varphi^*_{l_1l_2}\varphi_{l_1'l_2'}dv=S_{l_1-l_1',l_2-l_2'}$.

Applying the diagonalization method, we can easily express the inverse overlap matrix elements as a double contour integral,

$$S^{-1}{}_{l_1-l_1',l_2-l_2'}=\frac{1}{(2\pi i)^2}\oint dz_1\oint dz_2\frac{z_1{}^{|l_1-l_1'|}z_2{}^{|l_2-l_2'|}}{z_1z_2s(z_1,z_2)} \quad (26)$$

where $s(z_1,z_2)=\sum_{l_1=-\infty}^{\infty}\sum_{l_2=-\infty}^{\infty}S_{l_1l_2}z_1{}^{l_1}z_2{}^{l_2}$ is the overlap function for a two-dimensional array and the contours are unit circles in the complex z-planes. The first contour integral may be evaluated immediately to give

$$S^{-1}{}_{l_1-l_1',l_2-l_2'}=\frac{1}{2\pi i}\sum_i\oint dz_1\frac{z_1{}^{|l_1-l_1'|}[w_i(z_1)]^{|l_2-l_2'|}}{z_1w_i(z_1)s_i(z_1)} \quad (27)$$

where $w_i(z_1)$ are functions of z_1 defined implicitly by the constraint $s(z_1,w)=0$ and

$$s_i(z_1)=\partial s(z_1,z_2)/\partial z_2|_{z_2=w_i(z_1)}.$$

The limits $\lim_{z_1\to 0} z_1 w_i(z_1) s_i(z_1)$ are finite and the w_i are bounded functions of z_1 which can vanish only when $z_1=0$; hence the poles of the integrands of Eq. (27) will be the zeros of the functions $s_i(z)$. The difficulty comes from the fact that the poles of $s_i(z)$ are also branch points; hence, the best that can be done is to reduce Eq. (27) to line integrals along the paths which connect the branch points. These integrals cannot, in general, be reduced to a closed analytic form.

As an example, consider the case of nearest-neighbor overlap for which the only nonvanishing overlap integrals are $S_{10}=S_{-10}\equiv A$ and $S_{01}=S_{0-1}=B$, so that the overlap function is

$s(z_1,z_2)=1+A(z_1+z_1^{-1})+B(z_2+z_2^{-1})$. We obtain

$$w=-\tfrac{1}{2}f(z_1)\{1-[1-4f^{-2}(z_1)]^{\frac{1}{2}}\}, \tag{28}$$

where $f(z_1)=[1+A(z_1+z_1^{-1})]/B$ and

$$w\partial s(z_1,w)/\partial z_2=B(w-w^{-1})=[f^2(z_1)-4]^{\frac{1}{2}}, \tag{29}$$

so that

$$S^{-1}{}_{l_1-l_1',l_2-l_2'}=\frac{1}{2\pi Bi}\oint \frac{z_1^{|l-l'|-1}[w(z_1)]^{|l-l'|}}{[f^2(z_1)-4]^{\frac{1}{2}}}dz_1. \tag{30}$$

The poles of the integrand which lie within the unit circle are $z_a=\{-1-2B+[(1+2B)^2-A^2]^{\frac{1}{2}}\}/2A$ and $z_b=\{-1+2B+[(1-2B)^2-A^2]^{\frac{1}{2}}\}/2A$. [We observe that $w(z_a)=1$ and $w(z_b)=-1$.] These poles are also branch points; hence, the best that can be done is to reduce Eq. (30) to a line integral along the path connecting the branch points.

$$\begin{aligned} S^{-1}{}_{l_1-l_1',l_2-l_2'} &= \pm\frac{1}{i\pi B}\int_{z_a}^{z_b}\frac{z_1^{|l_1-l_1'|-1}[w(z_1)]^{|l_2-l_2'|}}{[f^2(z_1)-4]^{\frac{1}{2}}}dz_1 \\ &= \pm\frac{1}{i\pi A}\int_{-1}^{+1}\frac{[z_1(w)]^{|l_1-l_1'|}w^{|l_2-l_2'|-1}}{z_1(w)-z^{-1}(w)}dw \end{aligned} \tag{31}$$

where the sign must be chosen so that $S^{-1}{}_{00}>0$. This line integral cannot, in general, be reduced to a simple analytic form. One might attempt to use numerical or other approximate techniques to evaluate the integral, but this approach does not appear to be very promising.

Even though we cannot calculate the inverse overlap matrix elements of two- and three-dimensional arrays in a simple manner, we can still obtain some qualitative information on the range of the inverse overlap, (i.e., the magnitude of $[(l_1-l_1')^2+(l_2-l_2')^2]^{\frac{1}{2}}$ beyond which $S^{-1}{}_{l_1-l_1',l_2-l_2'}$ is smaller than some preassigned value), rather easily. We know that the eigenvalues of the overlap matrix must be positive; hence the overlap integrals must satisfy the constraint

$$s(e^{ik_1},e^{ik_2})=\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}S_{n_1n_2}e^{i(n_1k_1+n_2k_2)}>0 \tag{32}$$

for all $-\pi<k_i\leqslant\pi$. In the one-dimensional case the range of the inverse overlap was directly related to the distance of the zeros of the overlap function from the unit circle. As a zero approached the unit circle, the range of the inverse overlap became very large. Furthermore, the magnitude of the overlap function on the unit circle (i.e., the eigenvalues of the overlap matrix) must become very small in the vicinity of a zero which is approaching the unit circle. It follows that, at least in the limiting case of large range, the range will increase uniformly as the minimum eigenvalue of the overlap matrix decreases. It is reasonable to assume that this last statement is valid for two- and three-dimensional arrays as well as one-dimensional arrays.

As an example of the application of this prescription, let us examine the conditions under which the range of the inverse overlap matrix may be expected to be small for three-dimensional arrays of atoms with S orbitals. The eigenvalues for this system are

$$s(e^{ik_1},e^{ik_2},e^{ik_3}) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}S_{n_1n_2n_3}\cdot e^{i(k_1n_1+k_2n_2+k_3n_3)}$$

where $S_{n_1n_2n_3}$ are the overlap integrals. In a simple cubic system, the nearest-neighbor overlap integrals will be $S_1=S_{\pm 100}=S_{0\pm 10}=S_{00\pm 1}$. If nearest neighbors only are considered, the minimum eigenvalue will be $s(e^{i\pi},e^{i\pi},e^{i\pi})=1-6S_1$. We infer that the range will be small only if $S_1\ll\frac{1}{6}$. This is a much more stringent requirement than the one-dimensional case where a nearest-neighbor overlap integral which was small compared to $\frac{1}{2}$ would give a short range. In the one-dimensional case an overlap integral of 0.1 may be regarded as small (i.e., the inverse overlap matrix elements will be negligible beyond second or third neighbors). An overlap integral of this magnitude cannot be regarded as small in the three-dimensional case. These considerations also serve to emphasize the importance of more-distant-neighbor overlap. If we include the overlap integral S_2 for the twelve next-nearest neighbors, the minimum eigenvalue becomes $1-6S_1+12S_2$. We may infer that the inclusion of second-neighbor overlap integrals may lead to a non-negligible decrease in the range of the inverse overlap matrix. In a face-centered cubic system, the minimum eigenvalue for nearest-neighbor overlap is $s(1,e^{i\pi},e^{i\pi})=1-4S_1$. The inclusion of second- and third-neighbor overlap increases this minimum to $1-4S_1+6S_2+8S_3$.

9. REMARKS OF THE EXPANSION METHOD FOR CALCULATING THE INVERSE OVERLAP MATRIX FOR ARBITRARY SYSTEMS

We may sum up the major results which have been obtained so far by the following statements. The diagonalization method appears to be a simple and practical method for calculating the inverse overlap matrix of one-dimensional arrays of atoms. It also provides us with some insight into the nature of the problem by enabling us to relate the range of the inverse overlap matrix to the zeros of a simple function constructed from the overlap integrals, and permits us

to study the effect of distant-neighbor overlap in simple systems. However, it does not have much practical use for calculating the inverse overlap matrix of two- and three-dimensional arrays of atoms.

Under these circumstances it is worthwhile to re-examine the Löwdin expansion method for calculating the inverse overlap matrix for two- and three-dimensional arrays.[3] Löwdin's expansion is

$$\mathbf{S}^{-1}=S_{\max}^{-1}\sum_{i=0}^{\infty}\mathbf{d}^i, \tag{33}$$

where $S_{\max}$ is the largest eigenvalue of $\mathbf{S}$ and $\mathbf{d}=\mathbf{1}-\mathbf{S}/S_{\max}$. This expansion is convergent for all overlap matrices constructed from linearly independent orbitals; however, the convergence may be very slow. It is for this reason that one seeks to find alternate methods for calculating the inverse overlap matrix. Because it now appears that the expansion method may still be needed for most physical systems of interest, it becomes urgent to find means of circumventing the convergence problem. The following simple iteration-expansion scheme, which does not appear to have been considered heretofore for this particular problem, is one possible means of improving convergence.

We start with the Löwdin expansion, retaining only a small number of terms.

$$\mathbf{d}_1=\mathbf{1}-\mathbf{S}/S_{\max},\quad \mathbf{S}_1^{-1}=S_{\max}^{-1}\sum_{i=0}^{n-1}\mathbf{d}_1{}^i. \tag{34}$$

The matrix $\mathbf{S}_1^{-1}$ gives a first approximation to the inverse, albeit a rather poor one when n is small. We calculate a second approximation using the relations

$$\mathbf{d}_2=\mathbf{1}-\mathbf{S}\mathbf{S}_1^{-1},\quad \mathbf{S}_2^{-1}=\mathbf{S}_1^{-1}\sum_{i=0}^{n-1}\mathbf{d}_2{}^i. \tag{35}$$

Let

$$\mathbf{d}_m=\mathbf{1}-\mathbf{S}\mathbf{S}_{m-1}^{-1},\quad \mathbf{S}_m^{-1}=\mathbf{S}_{m-1}^{-1}\sum_{i=0}^{n-1}\mathbf{d}_m{}^i \tag{36}$$

be the final iteration required to obtain the inverse to the desired accuracy. We note that each iteration involves n matrix multiplications; hence, the calculation of the final approximate inverse $\mathbf{S}_m^{-1}$ involves a total of mn matrix multiplications. This final result, expressed as a power series in $\mathbf{d}_1$, is

$$\mathbf{S}_m^{-1}=S_{\max}^{-1}\sum_{i=0}^{n^m-1}\mathbf{d}_1{}^i. \tag{37}$$

A direct calculation of this expansion using Eq. (33) would have involved n^m-1 matrix multiplications.

One can readily show that the number of matrix multiplications required to achieve a given accuracy is minimized by choosing $n=2$. This choice reduces the problem to iterating the equation,

$$\mathbf{S}_r^{-1}=2\mathbf{S}_{r-1}^{-1}-\mathbf{S}_{r-1}^{-1}\mathbf{S}\mathbf{S}_{r-1}^{-1} \tag{38}$$

m times, starting from the zero order approximation, $\mathbf{S}_0=\mathbf{1}/S_{\max}$.

As an example of the efficiency of this procedure, suppose that 1000 terms in Eq. (33) are required to obtain an inverse overlap matrix of the desired accuracy. By using the above-described iteration procedure, this same accuracy can be obtained by means of 20 matrix multiplications. Thus, the use of an iteration procedure reduces the time of the calculation by nearly two orders of magnitude. If the number of non-negligible overlap integrals is of the order of 10^4 or less, then the total number of multiplications required to calculate the inverse would be of the order of 2×10^5. Such a computation would require a time of the order of 1 minute on an electronic computer with a speed comparable to the IBM 704. A calculation of the same accuracy using Eq. (33) directly would require approximately 1 hour.

APPENDIX: THE SIGNATURE AND ZEROS OF THE OVERLAP FUNCTION FOR AN INFINITE CHAIN

Consider a single atomic orbital which vanishes outside of an interval of $n+1$ contiguous lattice periods; i.e., $\varphi(x_1,x_2,x_3)=0$ when $|x_1|>\frac{1}{2}(n+1)a$. This condition is imposed in order to limit the range of overlap to n-nearest neighbors. We may then decompose the localized orbital into the sum,

$$\varphi(x_1,x_2,x_3)=\sum_{m=1}^{n+1}\psi_m[x_1+\tfrac{1}{2}(n+3-2m)a,\,x_2,\,x_3], \tag{A1}$$

where $\psi_m(x_1,x_2,x_3)=\varphi[x_1+\frac{1}{2}(2m-n-3)a,\,x_2,\,x_3]$ when $0\leqslant x_1<a$ and $\psi_m(x_1,x_2,x_3)=0$ otherwise. Substituting this decomposition of the atomic orbitals into the overlap integrals $S_l=\int\varphi_0^*\varphi_l dv$ and simplifying, we obtain the following expression for the overlap function.

$$s(z)=\sum_{-\infty}^{\infty}S_l z^l=\sum_{m=1}^{n+1}\sum_{m'=1}^{n+1}z^{m-m'}\int\psi_m^*\psi_{m'}dv. \tag{A2}$$

We now minimize the overlap function with respect to arbitrary variations in the orbital components ψ_m, subject to the normalization condition

$$\int|\varphi|^2dv=\sum_{1}^{n+1}\int|\psi_m|^2dv=1.$$

This is a well-defined variational problem as long as $s(z)$ is real for all values of the varied quantities, i.e., as long as z lies on the unit circle.[13] The variational equations are

$$\sum_{m'}z^{m-m'}\psi_{m'}(x_1,x_2,x_3)=s(z)\psi_m(x_1,x_2,x_3). \tag{A3}$$

Equation (A3) implies a relation between the values of an orbital φ at points along the x_1 axis which are separated by a distance a. Orbitals which satisfy this relation will give extremal values of the overlap function. There is no restriction on the functional form of any one orbital component we choose to single out.

[13] When $s(z)=s'(z)+is''(z)$ is complex, then $\delta s(z)=0$ implies two equations: $\delta s'(z)=0$ and $\delta s''(z)=0$. There will, in general, be no set of values of the varied quantities which satisfies both of these equations. For example, the variational equations for $\delta s(z)=0$ when $|z|\neq1$ are Eq. (A3) together with the equation $\sum_{l'}\psi_{l'}{}^*z^{l'-l}=s(z)\psi_l{}^*$. This equation is not compatible with Eq. (A3) unless $|z|=1$, in which case $s(z)$ is real.

However, once the functional form of a single component ψ_m has been established, [thereby establishing the functional form of the orbital over an interval $\frac{1}{2}(2m-n-3)\leqslant x_1<\frac{1}{2}(2m-n-1)$], Eq. (A3) determines the functional form of the remaining orbital components [thereby determining the functional form of the orbital throughout the remainder of the interval $-\frac{1}{2}(n+1)\leqslant x_1<\frac{1}{2}(n+1)$].

Equation (A3) tells us that the extremal values of the overlap function are the eigenvalues of the $(n+1)\times(n+1)$ matrix $\mathbf{Z}=[z^{m-m'}]$. This matrix, which is Hermitian when $|z|=1$, has two eigenvalues; a nondegenerate eigenvalue $s(z)=n+1$ with the eigenvector $\{1,z,z^2,\cdots,z^n\}$, and an n-fold degenerate eigenvalue $s(z)=0$ with the eigenvectors $\{1,-z,0,\cdots,0,0\}$, $\{0,1,-z,\cdots,0,0\},\cdots,\{0,0,0,\cdots,1,-z\}$. The eigenvalues of the overlap matrix are bounded because $s(z)$ is bounded on the unit circle. Hence, the eigenvalues of the overlap matrix for an infinite chain with a finite range of overlap must lie between the limits 0 and $n+1$. Passing to the limit $n\rightarrow\infty$, we conclude that the eigenvalues of the overlap matrix for an infinite chain are nonnegative. We now prove that the orbital set $\{\varphi_l\}$, where $\varphi_l(x_1,x_2,x_3)=\varphi(x_1-la,x_2,x_3)$, is linearly dependent when $s(e^{ik})=0$ for any real value of k.

The orbital components

$$\{\psi_m{}^{(q)}\}=\{\psi_1{}^{(q)},\psi_2{}^{(q)},\cdots,\psi_{n+1}{}^{(q)}\},$$

$q=1,2,\cdots,n+1$, which satisfy Eq. (A3) are as follows:

$$\left.\begin{array}{l}\{\psi_m{}^{(1)}\}=\{f_1,-zf_1,0,\cdots,0,0\}\\ \{\psi_m{}^{(2)}\}=\{0,f_2,-zf_2,\cdots,0,0\}\\ \cdots\cdots\cdots\cdots\\ \{\psi_m{}^{(n)}\}=\{0,0,0,\cdots,f_n,-zf_n\}\end{array}\right\}\quad s(z)=0 \tag{A4}$$

$$\{\psi_m{}^{(n+1)}\}=\{f_{n+1},zf_{n+1},z^2f_{n+1},\cdots,z^nf_{n+1}\},\quad s(z)=n+1$$

where $f_1(x_1,x_2,x_3)$, $f_2(x_1,x_2,x_3)$, $\cdots$, $f_{n+1}(x_1,x_2,x_3)$ are arbitrary functions which vanish when x_1 lies outside the interval $0\leqslant x_1<a$. Consider an arbitrary orbital φ with orbital components ψ_m. If the functions f_q are appropriately chosen, we may always express the orbital components as linear combinations of the components of the extremal orbitals [i.e., the orbitals with components given by Eq. (A4)]. The relation is

$$\psi_m=\sum_{q=1}^{n+1}\psi_m{}^{(q)}=\sum_{q=1}^{n+1}A_{mq}f_q, \tag{A5}$$

where

$$\mathbf{A}(z)=[A_{mq}]=\begin{bmatrix}1&0&0&\cdots&0&1\\-z&1&0&\cdots&0&z\\0&-z&1&\cdots&0&z^2\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\0&0&0&\cdots&1&z^{n-1}\\0&0&0&\cdots&-z&z^n\end{bmatrix}. \tag{A6}$$

The functions f_q are uniquely determined by the relation

$$f_q=\sum_{m=1}^{n+1}A_{qm}{}^{-1}\psi_m \tag{A7}$$

where

$$\mathbf{A}^{-1}(z)=[A_{qm}{}^{-1}]$$

$$=\frac{1}{(n+1)z^n}\begin{bmatrix}nz^n&-z^{n-1}&-z^{n-2}&-z^{n-3}&\cdots&-z&-1\\(n-1)z^{n+1}&(n-1)z^n&-2z^{n-1}&-2z^{n-2}&\cdots&-2z^2&-2z\\(n-2)z^{n+2}&(n-2)z^{n+1}&(n-2)z^n&-3z^{n-1}&\cdots&-3z^3&-3z^2\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\z^{2n-1}&z^{2n-2}&z^{2n-3}&z^{2n-2}&\cdots&z^n&-nz^{n-1}\\z^n&z^{n-1}&z^{n-2}&z^{n-3}&\cdots&z&1\end{bmatrix}. \tag{A8}$$

The functional form of f_q for a given orbital φ will, of course, depend upon the choice of the parameter z. In order to be explicit we should write $f_q=f_q(x_1,x_2,x_3;z)$.

It is convenient to rewrite Eq. (A5) as

$$\psi_m=\psi_m{}^{(0)}+\psi_m{}^{(n+1)} \tag{A9}$$

where $\psi_m{}^{(0)}=\sum_{q=1}^{n}\psi_m{}^{(q)}$ and $\psi_m{}^{(n+1)}$ is as previously defined. We may then write

$$\varphi=\varphi^{(0)}+\varphi^{(n+1)}, \tag{A10}$$

where

$$\varphi^{(0)}(x_1,x_2,x_3;z)=\sum_{m=1}^{n+1}\psi_m{}^{(0)}[x_1+\tfrac{1}{2}(n+3-2m)a,x_2,x_3;z]$$

and

$$\varphi^{(n+1)}(x_1,x_2,x_3;z)$$

$$=\sum_{m=1}^{n+1}\psi_m{}^{(n+1)}(x_1+\tfrac{1}{2}(n+3-2m)a,x_2,x_3;z).$$

(These unnormalized orbitals depend upon z in such a manner that their sum is independent of z.) Using Eq. (A4), we can easily establish the fact that

$$\int\varphi_l{}^{(n+1)*}\varphi_0{}^{(0)}dv=0 \tag{A11}$$

for all values of l; hence

$$S_l=\int \varphi_l^*\varphi_0 dv=N^{(0)}S_l^{(0)}+N^{(n+1)}S_l^{(n+1)} \quad \text{(A12)}$$

where

$$N^{(0)}=\int |\varphi^{(0)}|^2 dv\leqslant 1, \quad N^{(n+1)}=\int |\varphi^{(n+1)}|^2 dv\leqslant 1$$

and

$$S_l^{(0)}=\int \varphi_l^{(0)*}\varphi_0^{(0)} dv/N^{(0)}$$

and

$$S_l^{(n+1)}=\int \varphi_l^{(n+1)*}\varphi_l^{(n+1)} dv/N^{(n+1)}.$$

The quantities $N^{(0)}$, $N^{(n+1)}$, $S^{(0)}$, and $S^{(n+1)}$ are all functions of the parameter z. We may now write

$$s(z)=N^{(0)}s^{(0)}(z)+N^{(n+1)}s^{(n+1)}(z)=(n+1)N^{(n+1)} \quad \text{(A13)}$$

where $s^{(0)}(z)=\sum_{-n}^{n} S_l^{(0)}z^l=0$ and

$$s^{(n+1)}(z)=\sum_{-n}^{n} S_l^{(n+1)}z^l=n+1.$$

[We have used the fact that $\varphi^{(0)}$ and $\varphi^{(n+1)}$ were constructed from solutions of Eq. (A3).]

We conclude from Eq. (A13) that $s(z)=0$ if and only if $\varphi^{(n+1)}(x_1,x_2,x_3;z)=0$, i.e., if and only if

$$\varphi(x_1,x_2,x_3)=\varphi^{(0)}(x_1,x_2,x_3;z) \quad \text{(A14)}$$

for the particular value of z at which $s(z)$ vanishes. But we can easily show that, for any orbital constructed entirely from the components $\psi_m^{(q)}$, $q=1, 2, \cdots, n$, the relation

$$\sum_{-\infty}^{\infty} z^l\varphi_l^{(0)}=0 \quad \text{(A15)}$$

must hold [use Eqs. (A1) and (A3)]. We thereby establish the result that the basis orbitals $\{\varphi_l\}$ of an infinite linear chain must be linearly dependent if the eigenvalues of the overlap matrix vanish at any point in the spectrum. The proof, as given, is applicable for an overlap of arbitrarily large, but finite, range. If Eq. (A15) holds for an overlap of any finite range, it must also hold in the limit as the range of overlap becomes infinite. We therefore conclude that this result is valid for any infinite chain, provided that the sum converges for almost all values of x_1, x_2, and x_3.

Traces of Products of Angular Momentum Matrices

E. AMBLER, J. C. EISENSTEIN, AND J. F. SCHOOLEY
National Bureau of Standards, Washington, D. C.

(Received May 11, 1961)

Closed formulas are given for evaluating Tr $J_a{}^pJ_b{}^qJ_c{}^r\cdots$ where a, b, c, $\cdots$ are equal to x, y, or z and p, q, r, $\cdots$ are non-negative integers for which $p+q+r+\cdots\leq 10$. All possible combinations of the angular momentum components for $p+q+r+\cdots\leq 9$ are included. Numerical values of the traces are given for $J=\frac{1}{2}$, 1, $\cdots 10$. The procedures used in evaluating the traces are described.

1. INTRODUCTION

A METHOD originally due to Van Vleck[1] is often used for the computation of thermal and magnetic properties of paramagnetic salts at low temperatures. The method is applicable to a salt in thermodynamic equilibrium and yields the quantities of interest as power series in $1/T$ where T is the absolute temperature. Although the method was given originally with explicit reference to the calculation of entropy and susceptibility, it can be extended in an obvious way to the calculation of any property of the system. Suppose such a property is represented quantum mechanically by an operator O. Then the average value of O is given by

$$\bar{O}=(\mathrm{Tr}\,O\rho)/(\mathrm{Tr}\,\rho), \tag{1}$$

where ρ is the density matrix for the system. For thermodynamic equilibrium,

$$\rho=\exp\{-\mathfrak{H}/kT\}\equiv 1-\frac{\mathfrak{H}}{kT}+\frac{\mathfrak{H}^2}{2!k^2T^2}-\frac{\mathfrak{H}^3}{3!k^3T^3}+\cdots, \tag{2}$$

where $\mathfrak{H}$ is the Hamiltonian. Therefore

$$\bar{O}=\left[\langle O\rangle-\frac{\langle O\mathfrak{H}\rangle}{kT}+\frac{\langle O\mathfrak{H}^2\rangle}{2!k^2T^2}-\cdots\right]\times\left[1-\frac{\langle\mathfrak{H}\rangle}{kT}+\frac{\langle\mathfrak{H}^2\rangle}{2!k^2T^2}-\cdots\right]^{-1}, \tag{3}$$

where, for example,

$$\langle O\mathfrak{H}^n\rangle=(\mathrm{Tr}\,O\mathfrak{H}^n)/(\mathrm{Tr}\,1). \tag{4}$$

We have, further,

$$\bar{O}=\left[\langle O\rangle-\frac{\langle O\mathfrak{H}\rangle}{kT}+\frac{\langle O\mathfrak{H}^2\rangle}{2!k^2T^2}-\cdots\right]\times\left[1+\frac{\langle\mathfrak{H}\rangle}{kT}+\left\{-\frac{\langle\mathfrak{H}^2\rangle}{2!k^2T^2}+\frac{\langle\mathfrak{H}\rangle^2}{k^2T^2}\right\}+\left\{\frac{\langle\mathfrak{H}^3\rangle}{3!k^3T^3}-\frac{2\langle\mathfrak{H}\rangle\langle\mathfrak{H}^2\rangle}{2!k^3T^3}+\frac{\langle\mathfrak{H}\rangle^3}{k^3T^3}\right\}+\cdots\right]. \tag{5}$$

The final expansion in $1/T$ can be obtained from (5).

The essential advantage of the method lies in the fact that the result is given as a sum of traces, which, on account of the invariance of a trace under canonical transformation, may be evaluated in any representation. A convenient one is chosen, therefore, and it is not necessary to solve an eigenvalue problem. The advantage of this fact is greatest when couplings between different ions in a crystal must be considered, i.e., for many-body problems, although the method is still useful in other cases.

With the introduction of the spin Hamiltonian,[2] the method became more straightforward to apply since the traces that need to be evaluated in this case are those of operator products of the form $J_a{}^pJ_b{}^qJ_c{}^r$, where a, b and c are equal to x, y or z, J is an angular momentum, and p, q and r are integers. It is essential to note, however, that on account of the non-commutation of the terms both in O and in $\mathfrak{H}$, quantities such as $O\mathfrak{H}^n$ may contain the sum of products of angular momentum operators permuted in different ways. It is this fact that adds complication to the method, principally because the terms in the expansion become very unwieldy. It seemed to us that individual problems could be simplified to some extent if there were available a table of the traces of products such as $J_x{}^pJ_y{}^qJ_z{}^r$ and all the permutations. The evaluation and tabulation of these traces has been the aim of this paper although for the sake of completeness we have given, in addition, a few of the formulas for calculating observable quantities most frequently encountered in practice.

2. EVALUATION OF THE TRACES

We used purely algebraic methods to obtain formulas for the traces. In obtaining these formulas it is necessary to use elementary properties of the trace; i.e., that $\mathrm{Tr}(AB)=\mathrm{Tr}(BA)$, and that the trace is invariant under canonical transformation. It is also necessary to use the fact that the rotations which belong to the cubic group are canonical transformations and to use the commutation relations for the angular momentum operators. All the traces can be derived[3] from those of the form Tr $J_z{}^{2n}$ provided one starts with the lowest values of $(p+q+r)$ and works up.

[1] J. H. Van Vleck, J. Chem. Phys. 5, 320 (1937).

[2] A. Abragam and M. H. L. Pryce, Proc. Roy. Soc. (London) **A205**, 135 (1951).

[3] i.e., all the traces for which we have obtained formulas. We believe the statement is true in general, but we have not proved it.

Tr J_z^{2n} can itself be evaluated from the following well-known formula.[4]

$$\sum_{s=\alpha}^{J} s^{2n}=\frac{1}{2n+1}[B_{2n+1}(J+1)-B_{2n+1}(\alpha)]. \quad (6)$$

Here s increases by integral steps from α to J, B_{2n+1} is the Bernoulli polynomial of order unity, and n is a non-negative integer. For J integral or half-integral (that is, for $\alpha=0$ or $\frac{1}{2}$) and $n\geqslant 1$, $B_{2n+1}(\alpha)=0$.

It is fortunate that a great many of the traces vanish. To enumerate the ones that are zero, we can use the fact that the trace is invariant under rotations through an angle π about any one of the coordinate axes. This operation changes many of the operators O into $-O$. One has $\mathrm{Tr}\,O=\mathrm{Tr}(-O)$ and therefore Tr O equals zero. In this way we can deduce that the traces of all operators vanish except the ones listed below:

J_a^p with p even, (7a)

$J_a^p J_b^q$ with p and q both even, (7b)

$J_a^p J_b^q J_c^r$ with *either* p, q and r all even,
or p, q and r all odd. (7c)

It is to be noted that by writing terms such as $J_a^p J_b^q J_c^r$ in (7), we mean to include any permutation of the individual J_a, J_b and J_c.

By using the other rotations of the cubic group, that is, those through $\pi/2$ about a fourfold axis and $2\pi/3$ about a threefold axis, we were able to reduce considerably the number of traces which had to be evaluated. For example, a counterclockwise rotation through $\pi/2$ about the z axis sends J_x into J_y and J_y into $-J_x$. Therefore, one has Tr $J_x^4 J_y^2 J_z^2=$ Tr $J_y^4 J_x^2 J_z^2$. The trace is also invariant under cyclic permutation of the operators so that, for example, Tr $J_x^4 J_y^2 J_z^2$ $=$Tr $J_x^3 J_y^2 J_z^2 J_x=\cdots=$Tr $J_y^2 J_z^2 J_x^4=$Tr $J_z^2 J_x^4 J_y^2$.

A further reduction in the number of traces that must be evaluated independently can be achieved by making use of the Hermitian property of the angular momentum operators. If A, B, C, $\cdots$ are Hermitian we have

$$(ABC\cdots)^\dagger=\cdots C^\dagger B^\dagger A^\dagger=\cdots CBA$$

and, therefore, since transposition does not affect the trace,

$$\mathrm{Tr}(ABC\cdots)^\dagger=\mathrm{Tr}(ABC\cdots)^*=(\mathrm{Tr}\,ABC\cdots)^*$$
$$=\mathrm{Tr}\cdots CBA=\mathrm{Tr}\,A\cdots CB. \quad (8)$$

Here the dagger denotes Hermitian conjugate and the star complex conjugate. In the case of the angular momentum operators

$$(\mathrm{Tr}\,ABC\cdots)^*=(-)^n\,\mathrm{Tr}\,ABC\cdots$$

when n is the total number of operators.

[4] See, for example, J. M. Milne-Thomson, *The Calculus of Finite Differences* (MacMillan and Company, Ltd., London, 1951), p. 137.

It follows, for example, that

$$\mathrm{Tr}\,J_x^3 J_y J_z J_y J_x^2 J_y=-\mathrm{Tr}\,J_x^3 J_y J_x^2 J_y J_z J_y$$

and that

$$\mathrm{Tr}\,J_x^2 J_y J_x^2 J_y J_z J_x J_y=-\mathrm{Tr}\,J_x^2 J_y J_x^2 J_y J_x J_z J_y$$

since in the first case J_x^3 can be used as the matrix A, and in the second case $J_x^2 J_y J_x^2$ can be used for A.

In order to demonstrate the methods we used for actual evaluations we give below a few examples.

Example 1. Tr $J_x^3 J_y J_z$. A counterclockwise rotation about the x axis through 90° sends J_y into J_z, J_z into $-J_y$ and $J_x^3 J_y J_z$ into $-J_x^3 J_z J_y$. Therefore, we have

$$2\,\mathrm{Tr}\,J_x^3 J_y J_z=\mathrm{Tr}\{J_x^3 J_y J_z-J_x^3 J_z J_y\}.$$

Because of the commutation relations the expression on the right can be written as i Tr J_x^4, which can be readily evaluated by using (6).

Example 2. Tr $J_x^4 J_y^2$.

$$\mathrm{Tr}\,J_x^4 J_y^2=\mathrm{Tr}\,J_x^4(\mathbf{J}^2-J_x^2-J_z^2)$$
$$=J(J+1)\,\mathrm{Tr}\,J_x^4-\mathrm{Tr}\,J_x^6-\mathrm{Tr}\,J_x^4 J_z^2.$$

A counterclockwise rotation about the x axis through 90° sends J_y^2 into J_z^2. Therefore Tr $J_x^4 J_z^2=$Tr $J_x^4 J_y^2$ and

$$2\,\mathrm{Tr}\,J_x^4 J_y^2=J(J+1)\,\mathrm{Tr}\,J_x^4-\mathrm{Tr}\,J_x^6.$$

The traces on the right can be obtained again from (6).

The number of distinct traces is given in Table I for each order up to and including ten; by an order we mean the value of $(p+q+r)$. In the second column we include all nonzero traces but count only one for all those operators, O, for which $O\rightarrow\pm O$ under the rotations of the cubic group (which includes, of course, cyclic permutations of x, y and z). In the third column we again impose this restriction in counting but, in addition, count only one for those operators connected by Eq. (8). In listing the traces explicitly, however, in Tables II and III, which we shall describe more fully

TABLE I. Numbers of independent traces for products of the angular momentum matrices. The order is the sum of the exponents in $J_a^p J_b^q J_c^r\cdots$. The number in column A is obtained by using only the invariance of the trace under the canonical transformations which belong to the cubic group. The number in column B is obtained by using also the Hermitian property of the matrices.

Order	Number (A)	Number (B)
0	1	1
1	0	0
2	1	1
3	1	1
4	3	3
5	2	2
6	9	9
7	13	10
8	41	34
9	95	57
10	261	156

later, we list the number according to column two, but indicate by means of an equation at the appropriate part of the table, the connection implied in column three.

We performed the algebraic evaluation in a systematic way in the manner illustrated in the examples given above. We met no unusual difficulty until we reached the eighth order and beyond. We give below those operators that caused some difficulty, and also sketch the method we used to evaluate them.

TABLE II. Analytical expressions for the traces of products of the angular momentum matrices.

$\mathrm{Tr}\, J_x^2 = (1/3)J(J+1)(2J+1)$

$\mathrm{Tr}\, J_xJ_yJ_z = (i/6)J(J+1)(2J+1)$

$\mathrm{Tr}\, J_x^4 = (1/15)J(J+1)(2J+1)(3J^2+3J-1)$
$\mathrm{Tr}\, J_x^2J_y^2 = (1/30)J(J+1)(2J+1)(2J^2+2J+1)$
$\mathrm{Tr}\, J_xJ_yJ_xJ_y = (1/15)J(J+1)(2J+1)(J-1)(J+2)$

$\mathrm{Tr}\, J_x^3J_yJ_z = (i/30)J(J+1)(2J+1)(3J^2+3J-1)$
$\mathrm{Tr}\, J_x^2J_yJ_xJ_z = (i/30)J(J+1)(2J+1)(J-1)(J+2)$

$\mathrm{Tr}\, J_x^6 = (1/21)J(J+1)(2J+1)(3J^4+6J^3-3J+1)$
$\mathrm{Tr}\, J_x^4J_y^2 = (1/210)J(J+1)(2J+1)(6J^4+12J^3+14J^2+8J-5)$
$\mathrm{Tr}\, J_x^3J_yJ_xJ_y = (1/210)J(J+1)(2J+1)(J-1)(J+2)(6J^2+6J-1)$
$\mathrm{Tr}\, J_x^2J_y^2J_z^2 = (1/210)J(J+1)(2J+1)(J-1)(J+2)(2J^2+2J-5)$
$\mathrm{Tr}\, J_x^2J_yJ_x^2J_y = (1/105)J(J+1)(2J+1)(J-1)(J+2)(3J^2+3J-4)$
$\mathrm{Tr}\, J_x^2J_yJ_z^2J_y = (1/210)J(J+1)(2J+1)(2J^4+4J^3+21J^2+19J-11)$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_yJ_z = (1/105)J(J+1)(2J+1)(J-1)(J+2)(J^2+J+1)$
$\mathrm{Tr}\, J_xJ_yJ_zJ_xJ_yJ_z = (1/210)J(J+1)(2J+1)(2J^4+4J^3-28J^2-30J+17)$
$\mathrm{Tr}\, J_xJ_yJ_xJ_zJ_yJ_z = (1/210)J(J+1)(2J+1)(J-1)(J+2)(2J^2+2J-5)$

$\mathrm{Tr}\, J_x^5J_yJ_z = (i/42)J(J+1)(2J+1)(3J^4+6J^3-3J+1)$
$\mathrm{Tr}\, J_x^4J_yJ_xJ_z = (i/210)J(J+1)(2J+1)(J-1)(J+2)(9J^2+9J-5)$
$\mathrm{Tr}\, J_x^3J_yJ_x^2J_z = (i/210)J(J+1)(2J+1)(J-1)(J+2)(3J^2+3J-4)$
$\mathrm{Tr}\, J_x^3J_y^3J_z = (i/420)J(J+1)(2J+1)(18J^4+36J^3+21J^2+3J-8)$
$\mathrm{Tr}\, J_x^3J_y^2J_zJ_y = (i/420)J(J+1)(2J+1)(J-1)(J+2)(6J^2+6J-1)$
$\mathrm{Tr}\, J_x^3J_yJ_zJ_y^2 = -\mathrm{Tr}\, J_x^3J_y^2J_zJ_y$
$\mathrm{Tr}\, J_x^3J_y^2J_xJ_yJ_z = (i/420)J(J+1)(2J+1)(J-1)(J+2)(10J^2+10J-11)$
$\mathrm{Tr}\, J_x^2J_y^2J_zJ_xJ_y = \mathrm{Tr}\, J_x^2J_y^2J_xJ_yJ_z$
$\mathrm{Tr}\, J_x^2J_y^2J_xJ_zJ_y = (-i/420)J(J+1)(2J+1)(2J^4+4J^3+49J^2+47J-32)$
$\mathrm{Tr}\, J_x^2J_y^2J_xJ_y^2J_z = (i/60)J(J+1)(2J+1)(J-1)(J+2)(2J^2+2J-1)$
$\mathrm{Tr}\, J_x^2J_yJ_xJ_yJ_zJ_y = (i/420)J(J+1)(2J+1)(J-1)(J+2)(2J^2+2J-5)$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_yJ_zJ_y = -\mathrm{Tr}\, J_x^2J_yJ_xJ_yJ_zJ_y$
$\mathrm{Tr}\, J_xJ_yJ_xJ_yJ_xJ_yJ_z = (i/420)J(J+1)(2J+1)(J-1)(J+2)(6J^2+6J-1)$

$\mathrm{Tr}\, J_x^8 = (1/45)J(J+1)(2J+1)(5J^6+15J^5+5J^4-15J^3-J^2+9J-3)$
$\mathrm{Tr}\, J_x^6J_y^2 = (1/630)J(J+1)(2J+1)(10J^6+30J^5+55J^4+60J^3-23J^2-48J+21)$
$\mathrm{Tr}\, J_x^5J_yJ_xJ_y = (1/630)J(J+1)(2J+1)(J-1)(J+2)(10J^4+20J^3+10J^2-3)$
$\mathrm{Tr}\, J_x^4J_yJ_x^2J_y = (1/630)J(J+1)(2J+1)(J-1)(J+2)(10J^4+20J^3-17J^2-27J+12)$
$\mathrm{Tr}\, J_x^3J_yJ_x^3J_y = (1/315)J(J+1)(2J+1)(J-1)(J+2)(5J^4+10J^3-13J^2-18J+12)$
$\mathrm{Tr}\, J_x^4J_y^4 = (1/210)J(J+1)(2J+1)(2J^6+6J^5+14J^4+18J^3+J^2-7J+1)$
$\mathrm{Tr}\, J_x^3J_y^3J_xJ_y = (1/420)J(J+1)(2J+1)(J-1)(J+2)(4J^4+8J^3+10J^2+6J-5)$
$\mathrm{Tr}\, J_x^3J_y^2J_xJ_y^2 = (1/105)J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3+J^2-1)$
$\mathrm{Tr}\, J_x^2J_y^2J_x^2J_y^2 = (1/210)J(J+1)(2J+1)(2J^6+6J^5-2J^4-14J^3+29J^2+37J-23)$
$\mathrm{Tr}\, J_x^2J_y^2J_xJ_yJ_xJ_y = (1/420)J(J+1)(2J+1)(J-1)(J+2)(4J^4+8J^3-6J^2-10J+7)$
$\mathrm{Tr}\, J_x^2J_yJ_xJ_y^2J_xJ_y = (1/210)J(J+1)(2J+1)(J-1)(J+1)(2J^4+4J^3-2J^2-4J+1)$
$\mathrm{Tr}\, J_xJ_yJ_xJ_yJ_xJ_yJ_xJ_y = (1/105)J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3-3J^2-4J+2)$
$\mathrm{Tr}\, J_x^4J_y^2J_z^2 = (1/630)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-19J^2-21J+12)$
$\mathrm{Tr}\, J_x^4J_yJ_z^2J_y = (1/630)J(J+1)(2J+1)(2J^6+6J^5+53J^4+96J^3-13J^2-60J+21)$
$\mathrm{Tr}\, J_x^4J_yJ_zJ_yJ_z = (1/630)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+8J^2+6J-3)$
$\mathrm{Tr}\, J_x^3J_y^2J_xJ_z^2 = (1/630)J(J+1)(2J+1)(2J^6+6J^5+17J^4+24J^3+50J^2+39J-33)$
$\mathrm{Tr}\, J_x^3J_y^2J_zJ_xJ_z = (1/1260)J(J+1)(2J+1)(J-1)(J+2)(4J^4+8J^3-20J^2-24J+21)$
$\mathrm{Tr}\, J_x^3J_yJ_xJ_yJ_z^2 = \mathrm{Tr}\, J_x^3J_y^2J_zJ_xJ_z$
$\mathrm{Tr}\, J_x^3J_yJ_z^2J_xJ_y = (1/1260)J(J+1)(2J+1)(J-1)(J+2)(4J^4+8J^3+52J^2+48J-33)$
$\mathrm{Tr}\, J_x^3J_yJ_xJ_z^2J_y = \mathrm{Tr}\, J_x^3J_yJ_z^2J_xJ_y$
$\mathrm{Tr}\, J_x^3J_yJ_xJ_zJ_yJ_z = (1/1260)J(J+1)(2J+1)(J-1)(J+2)(4J^4+8J^3-2J^2-6J-3)$

TABLE II (*continued*).

$\mathrm{Tr}\, J_x^3 J_y J_z J_y J_x J_z = \mathrm{Tr}\, J_x^3 J_y J_x J_z J_y J_z$

$\mathrm{Tr}\, J_x^3 J_y J_z J_x J_y J_z = (1/630) J(J+1)(2J+1)(2J^6+6J^5-28J^4-66J^3-34J^2+15)$

$\mathrm{Tr}\, J_x^3 J_y J_z J_x J_z J_y = (1/630) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+17J^2+15J-15)$

$\mathrm{Tr}\, J_x^2 J_y^2 J_x^2 J_z^2 = (1/630) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+29J^2+27J-24)$

$\mathrm{Tr}\, J_x^2 J_y^2 J_z J_x^2 J_z = (1/630) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+5J^2+3J-6)$

$\mathrm{Tr}\, J_x^2 J_y J_z^2 J_x^2 J_y = \mathrm{Tr}\, J_x^2 J_y^2 J_z J_x^2 J_z$

$\mathrm{Tr}\, J_x^2 J_y J_x^2 J_z J_y J_z = (1/315) J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3-2J^2-3J+3)$

$\mathrm{Tr}\, J_x^2 J_y J_z J_x^2 J_y J_z = (1/630) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-40J^2-42J+33)$

$\mathrm{Tr}\, J_x^2 J_y J_z J_x^2 J_z J_y = (1/630) J(J+1)(2J+1)(2J^6+6J^5+5J^4+71J^2+72J-51)$

$\mathrm{Tr}\, J_x^2 J_y^2 J_z J_z J_x J_z = (1/1260) J(J+1)(2J+1)(J-1)(J+2)(4J^4+8J^3+28J^2+24J-15)$

$\mathrm{Tr}\, J_x^2 J_y J_x J_y J_x J_z^2 = \mathrm{Tr}\, J_x^2 J_y^2 J_x J_z J_x$

$\mathrm{Tr}\, J_x^2 J_y J_x J_z^2 J_x J_y = (1/630) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+17J^2+15J-15)$

$\mathrm{Tr}\, J_x^2 J_y J_x J_z J_y J_x J_z = (1/1260) J(J+1)(2J+1)(J-1)(J+2)(4J^4+8J^3-50J^2-54J+33)$

$\mathrm{Tr}\, J_x^2 J_y J_z J_x J_y J_z J_z = \mathrm{Tr}\, J_x^2 J_y J_x J_z J_y J_z$

$\mathrm{Tr}\, J_x^2 J_y J_x J_z J_z J_x J_y = (1/1260) J(J+1)(2J+1)(J-1)(J+2)(4J^4+8J^3+4J^2+3)$

$\mathrm{Tr}\, J_x^2 J_y J_z J_x J_z J_x J_y = \mathrm{Tr}\, J_x^2 J_y J_x J_z J_z J_y$

$\mathrm{Tr}\, J_x^2 J_y J_x J_y J_z J_x J_z = (1/630) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-7J^2-9J+3)$

$\mathrm{Tr}\, J_x^2 J_y J_x J_z J_y J_x J_z = (1/630) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-16J^2-18J+15)$

$\mathrm{Tr}\, J_x J_y J_z J_y J_z J_z J_x = (1/630) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+5J^2+3J-6)$

$\mathrm{Tr}\, J_x J_y J_x J_z J_x J_y J_z J_x = (1/315) J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3-14J^2-15J+12)$

$\mathrm{Tr}\, J_x^7 J_y J_z = (i/90) J(J+1)(2J+1)(5J^6+15J^5+5J^4-15J^3-J^2+9J-3)$

$\mathrm{Tr}\, J_x^6 J_y J_x J_z = (i/630) J(J+1)(2J+1)(J-1)(J+2)(25J^4+50J^3-20J^2-45J+21)$

$\mathrm{Tr}\, J_x^6 J_y J_z^2 J_z = (i/210) J(J+1)(2J+1)(J-1)(J+2)(5J^4+10J^3-10J^2-15J+8)$

$\mathrm{Tr}\, J_x^4 J_y J_z^3 J_z = (i/630) J(J+1)(2J+1)(J-1)(J+2)(5J^4+10J^3-13J^2-18J+12)$

$\mathrm{Tr}\, J_x^5 J_y^3 J_z = (i/420) J(J+1)(2J+1)(10J^6+30J^5+40J^4+30J^3-23J^2-33J+16)$

$\mathrm{Tr}\, J_x^5 J_z J_y^3 = -\mathrm{Tr}\, J_x^5 J_y^3 J_z$

$\mathrm{Tr}\, J_x^5 J_y^2 J_z J_y = (i/1260) J(J+1)(2J+1)(J-1)(J+2)(10J^4+20J^3+10J^2-3)$

$\mathrm{Tr}\, J_x^5 J_y J_z J_y^2 = -\mathrm{Tr}\, J_x^5 J_y^2 J_z J_y$

$\mathrm{Tr}\, J_x^4 J_y^3 J_x J_z = (i/420) J(J+1)(2J+1)(J-1)(J+2)(6J^4+12J^3+12J^2+6J-7)$

$\mathrm{Tr}\, J_x^4 J_z J_x J_y^3 = -\mathrm{Tr}\, J_x^4 J_y^3 J_x J_z$

$\mathrm{Tr}\, J_x^4 J_y^2 J_x J_y J_z = (i/1260) J(J+1)(2J+1)(J-1)(J+2)(22J^4+44J^3-2J^2-24J+3)$

$\mathrm{Tr}\, J_x^4 J_z J_y J_x J_y^2 = -\mathrm{Tr}\, J_x^4 J_y^2 J_x J_y J_z$

$\mathrm{Tr}\, J_x^4 J_y^2 J_x J_z J_y = (i/1260) J(J+1)(2J+1)(2J^6+6J^5-112J^4-234J^3+29J^2+147J-48)$

$\mathrm{Tr}\, J_x^4 J_y J_z J_x J_y^2 = -\mathrm{Tr}\, J_x^4 J_y^2 J_x J_z J_y$

$\mathrm{Tr}\, J_x^4 J_y^2 J_z J_x J_y = (i/180) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-4J^2-6J+3)$

$\mathrm{Tr}\, J_x^4 J_y J_x J_z J_y^2 = -\mathrm{Tr}\, J_x^4 J_y^2 J_z J_x J_y$

$\mathrm{Tr}\, J_x^4 J_z J_y^2 J_x J_y = (-i/1260) J(J+1)(2J+1)(J-1)(J+2)(26J^4+52J^3+14J^2-12J-3)$

$\mathrm{Tr}\, J_x^4 J_y J_x J_y^2 J_z = -\mathrm{Tr}\, J_x^4 J_z J_y^2 J_x J_y$

$\mathrm{Tr}\, J_x^4 J_y J_x J_y J_z J_y = (i/420) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-2J^2-4J+1)$

$\mathrm{Tr}\, J_x^4 J_y J_x J_y J_x J_y = -\mathrm{Tr}\, J_x^4 J_y J_x J_y J_z J_y$

$\mathrm{Tr}\, J_x^3 J_y^3 J_x^2 J_z = (i/210) J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3+J^2-1)$

$\mathrm{Tr}\, J_x^3 J_z J_x^2 J_y^3 = -\mathrm{Tr}\, J_x^3 J_y^3 J_x^2 J_z$

$\mathrm{Tr}\, J_x^3 J_y^2 J_x^2 J_y J_z = (i/1260) J(J+1)(2J+1)(14J^6+42J^5+20J^4-30J^3+119J^2+141J-96)$

$\mathrm{Tr}\, J_x^3 J_z J_y J_x^2 J_y^2 = -\mathrm{Tr}\, J_x^3 J_y^2 J_x^2 J_y J_z$

$\mathrm{Tr}\, J_x^3 J_y^2 J_x^2 J_z J_y = (-i/420) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+30J^2+28J-23)$

$\mathrm{Tr}\, J_x^3 J_y J_z J_x^2 J_y^2 = -\mathrm{Tr}\, J_x^3 J_y^2 J_x^2 J_z J_y$

$\mathrm{Tr}\, J_x^3 J_y^2 J_z J_x^2 J_y = (i/210) J(J+1)(2J+1)(J-1)(J+2)(3J^4+6J^3-8J^2-11J+7)$

$\mathrm{Tr}\, J_x^3 J_y J_x^2 J_z J_y^2 = -\mathrm{Tr}\, J_x^3 J_y^2 J_z J_x^2 J_y$

$\mathrm{Tr}\, J_x^3 J_y J_z^2 J_y^2 J_z = (i/630) J(J+1)(2J+1)(J-1)(J+2)(11J^4+22J^3-19J^2-30J+15)$

$\mathrm{Tr}\, J_x^3 J_z J_y^2 J_x^2 J_y = -\mathrm{Tr}\, J_x^3 J_y J_x^2 J_y^2 J_z$

$\mathrm{Tr}\, J_x^3 J_y J_x^2 J_y J_z J_y = (i/630) J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3-2J^2-3J+3)$

$\mathrm{Tr}\, J_x^3 J_y J_z J_y J_x^2 J_y = -\mathrm{Tr}\, J_x^3 J_y J_x^2 J_y J_z J_y$

$\mathrm{Tr}\, J_x^3 J_y^2 J_z J_y J_x J_z = (i/1260) J(J+1)(2J+1)(J-1)(J+2)(10J^4+20J^3-14J^2-24J+15)$

$\mathrm{Tr}\, J_x^3 J_z J_x J_y J_z J_y^2 = -\mathrm{Tr}\, J_x^3 J_y^2 J_x J_y J_z J_z$

$\mathrm{Tr}\, J_x^3 J_y^2 J_z J_z J_x J_y = (i/420) J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-26J^2-28J+19)$

$\mathrm{Tr}\, J_x^3 J_y J_z J_z J_x J_y^2 = -\mathrm{Tr}\, J_x^3 J_y^2 J_z J_z J_x J_y$

$\mathrm{Tr}\, J_x^3 J_y J_z J_y^2 J_z J_z = (i/630) J(J+1)(2J+1)(J-1)(J+2)(7J^4+14J^3-8J^2-15J+6)$

$\mathrm{Tr}\, J_x^3 J_z J_z J_y^2 J_z J_y = -\mathrm{Tr}\, J_x^3 J_y J_z J_y^2 J_z J_z$

$\mathrm{Tr}\, J_x^3 J_y J_z J_y J_z J_y J_z = (i/420) J(J+1)(2J+1)(J-1)(J+2)(6J^4+12J^3-12J^2-18J+11)$

$\mathrm{Tr}\, J_x^3 J_z J_y J_z J_y J_z J_y = -\mathrm{Tr}\, J_x^3 J_y J_z J_y J_z J_y J_z$

TABLE II (*continued*).

$\mathrm{Tr}\, J_x^3J_yJ_xJ_yJ_xJ_y = (-i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+56J^2+54J-39)$
$\mathrm{Tr}\, J_x^3J_yJ_xJ_xJ_yJ_y = -\mathrm{Tr}\, J_x^3J_yJ_xJ_yJ_xJ_y$
$\mathrm{Tr}\, J_x^3J_yJ_xJ_yJ_xJ_y = (i/630)J(J+1)(2J+1)(J-1)(J+2)(5J^4+10J^3-13J^2-18J+12)$
$\mathrm{Tr}\, J_x^3J_yJ_xJ_xJ_yJ_y = -\mathrm{Tr}\, J_x^3J_yJ_xJ_yJ_xJ_y$
$\mathrm{Tr}\, J_x^2J_y^2J_x^2J_yJ_xJ_z = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+32J^2+30J-21)$
$\mathrm{Tr}\, J_x^2J_y^2J_x^2J_xJ_yJ_z = -\mathrm{Tr}\, J_x^2J_y^2J_x^2J_yJ_xJ_z$
$\mathrm{Tr}\, J_x^2J_y^2J_xJ_yJ_x^2J_z = (-i/630)J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3-2J^2-3J+3)$
$\mathrm{Tr}\, J_x^2J_yJ_x^2J_yJ_xJ_y^2 = -\mathrm{Tr}\, J_x^2J_y^2J_xJ_yJ_x^2J_z$
$\mathrm{Tr}\, J_x^2J_y^2J_xJ_xJ_x^2J_y = (i/630)J(J+1)(2J+1)(J-1)(J+2)(5J^4+10J^3-25J^2-30J+21)$
$\mathrm{Tr}\, J_x^2J_yJ_x^2J_xJ_yJ_y^2 = -\mathrm{Tr}\, J_x^2J_y^2J_xJ_xJ_x^2J_y$
$\mathrm{Tr}\, J_x^2J_yJ_x^2J_yJ_xJ_yJ_z = (i/630)J(J+1)(2J+1)(J-1)(J+2)(7J^4+14J^3-20J^2-27J+15)$
$\mathrm{Tr}\, J_x^2J_yJ_x^2J_xJ_yJ_xJ_y = -\mathrm{Tr}\, J_x^2J_yJ_x^2J_yJ_xJ_yJ_z$
$\mathrm{Tr}\, J_x^2J_yJ_x^2J_yJ_xJ_zJ_y = (-i/210)J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3+J^2-1)$
$\mathrm{Tr}\, J_x^2J_yJ_x^2J_yJ_xJ_yJ_z = -\mathrm{Tr}\, J_x^2J_yJ_x^2J_yJ_xJ_zJ_y$
$\mathrm{Tr}\, J_x^2J_yJ_xJ_yJ_x^2J_yJ_z = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(10J^4+20J^3+10J^2-3)$
$\mathrm{Tr}\, J_x^2J_yJ_xJ_yJ_x^2J_zJ_y = -\mathrm{Tr}\, J_x^2J_yJ_xJ_yJ_x^2J_yJ_z$
$\mathrm{Tr}\, J_x^2J_yJ_xJ_yJ_xJ_yJ_xJ_z = (i/210)J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3-3J^2-4J+2)$
$\mathrm{Tr}\, J_x^2J_zJ_xJ_yJ_xJ_yJ_xJ_y = -\mathrm{Tr}\, J_x^2J_yJ_xJ_yJ_xJ_yJ_xJ_z$
$\mathrm{Tr}\, J_x^2J_yJ_xJ_yJ_xJ_zJ_xJ_y = (i/630)J(J+1)(2J+1)(J-1)(J+2)(J^4+2J^3-14J^2-15J+12)$
$\mathrm{Tr}\, J_x^2J_yJ_xJ_zJ_xJ_yJ_xJ_y = -\mathrm{Tr}\, J_x^2J_yJ_xJ_yJ_xJ_zJ_xJ_y$
$\mathrm{Tr}\, J_x^3J_y^3J_z^3 = (i/420)J(J+1)(2J+1)(6J^6+18J^5+15J^4+24J^2+27J-20)$
$\mathrm{Tr}\, J_x^3J_y^2J_z^3J_y = (i/420)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+23J^2+21J-16)$
$\mathrm{Tr}\, J_x^3J_yJ_z^3J_y^2 = -\mathrm{Tr}\, J_x^3J_y^2J_z^3J_y$
$\mathrm{Tr}\, J_x^3J_y^2J_z^2J_yJ_z = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(10J^4+20J^3+31J^2+21J-24)$
$\mathrm{Tr}\, J_x^3J_yJ_zJ_y^2J_z^2 = \mathrm{Tr}\, J_x^3J_y^2J_z^2J_yJ_z$
$\mathrm{Tr}\, J_x^3J_y^2J_zJ_yJ_z^2 = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(14J^4+28J^3+11J^2-3J-3)$
$\mathrm{Tr}\, J_x^3J_yJ_z^3J_y^2J_z = (i/1260)J(J+1)(2J+1)(2J^6+6J^5+89J^4+168J^3+8J^2-75J+12)$
$\mathrm{Tr}\, J_x^3J_yJ_z^3J_yJ_zJ_y = (-i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+17J^2+15J-15)$
$\mathrm{Tr}\, J_x^3J_yJ_zJ_yJ_z^2J_y = -\mathrm{Tr}\, J_x^3J_yJ_z^2J_yJ_zJ_y$
$\mathrm{Tr}\, J_x^3J_yJ_zJ_yJ_zJ_yJ_z = (i/420)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+11J^2+9J-7)$
$\mathrm{Tr}\, J_x^2J_y^2J_z^2J_zJ_yJ_z = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(10J^4+20J^3-65J^2-75J+48)$
$\mathrm{Tr}\, J_x^2J_y^2J_z^2J_zJ_zJ_y = (i/420)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-9J^2-11J+8)$
$\mathrm{Tr}\, J_x^2J_y^2J_z^2J_yJ_zJ_z = \mathrm{Tr}\, J_x^2J_y^2J_z^2J_zJ_zJ_y$
$\mathrm{Tr}\, J_x^2J_y^2J_zJ_z^2J_yJ_z = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-55J^2-57J+39)$
$\mathrm{Tr}\, J_x^2J_y^2J_zJ_zJ_z^2J_y = \mathrm{Tr}\, J_x^2J_y^2J_zJ_z^2J_yJ_z$
$\mathrm{Tr}\, J_x^2J_yJ_z^2J_zJ_y^2J_z = (i/420)J(J+1)(2J+1)(-1)(J+2)(2J^4+4J^3+39J^2+37J-28)$
$\mathrm{Tr}\, J_x^2J_yJ_z^2J_zJ_zJ_yJ_z = (i/420)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-25J^2-27J+20)$
$\mathrm{Tr}\, J_x^2J_y^2J_zJ_yJ_zJ_z = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(10J^4+20J^3-17J^2-27J+12)$
$\mathrm{Tr}\, J_x^2J_yJ_z^2J_zJ_yJ_zJ_y = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+65J^2+63J-51)$
$\mathrm{Tr}\, J_x^2J_yJ_z^2J_yJ_zJ_yJ_z = \mathrm{Tr}\, J_x^2J_yJ_z^2J_zJ_yJ_zJ_y$
$\mathrm{Tr}\, J_x^2J_yJ_z^2J_yJ_zJ_zJ_y = (-i/1260)J(J+1)(2J+1)(2J^6+6J^5+41J^4+72J^3+92J^2+57J-60)$
$\mathrm{Tr}\, J_x^2J_yJ_z^2J_yJ_zJ_zJ_y = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-7J^2-9J+3)$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_z^2J_yJ_zJ_y = (-i/420)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-J^2-3J+2)$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_y^2J_zJ_zJ_z = -\mathrm{Tr}\, J_x^2J_yJ_zJ_z^2J_yJ_zJ_y$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_zJ_yJ_zJ_yJ_z = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+5J^2+3J-6)$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_yJ_zJ_yJ_zJ_z = \mathrm{Tr}\, J_x^2J_yJ_zJ_zJ_yJ_zJ_yJ_z$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_zJ_yJ_zJ_yJ_z = (i/420)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-5J^2-7J+5)$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_yJ_zJ_zJ_yJ_z = \mathrm{Tr}\, J_x^2J_yJ_zJ_zJ_yJ_zJ_yJ_z$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_zJ_zJ_yJ_yJ_y = (-i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-31J^2-33J+21)$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_yJ_zJ_zJ_zJ_y = -\mathrm{Tr}\, J_x^2J_yJ_zJ_zJ_zJ_yJ_zJ_y$
$\mathrm{Tr}\, J_x^2J_yJ_zJ_yJ_zJ_zJ_yJ_z = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-67J^2-69J+48)$
$\mathrm{Tr}\, J_xJ_yJ_zJ_zJ_zJ_yJ_yJ_z = (i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-19J^2-21J+12)$
$\mathrm{Tr}\, J_xJ_yJ_zJ_zJ_zJ_zJ_yJ_y = (-i/420)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3-9J^2-11J+8)$
$\mathrm{Tr}\, J_xJ_yJ_zJ_zJ_yJ_zJ_yJ_z = (-i/1260)J(J+1)(2J+1)(J-1)(J+2)(2J^4+4J^3+17J^2+15J-15)$
$\mathrm{Tr}\, J_xJ_yJ_zJ_zJ_yJ_zJ_zJ_y = (i/420)J(J+1)(2J+1)(2J^6+6J^5-13J^4-36J^3-34J^2-15J+20)$

$\mathrm{Tr}\, J_x^{10} = (1/33)J(J+1)(2J+1)(3J^8+12J^7+8J^6-18J^5-10J^4+24J^3+2J^2-15J+5)$
$\mathrm{Tr}\, J_x^8J_y^2 = (1/990)J(J+1)(2J+1)(10J^8+40J^7+100J^6+160J^5-26J^4-272J^3+36J^2+192J-75)$
$\mathrm{Tr}\, J_x^6J_y^4 = (1/6930)J(J+1)(2J+1)(30J^8+120J^7+410J^6+810J^5+428J^4-354J^3-343J^2-51J+105)$
$\mathrm{Tr}\, J_x^6J_y^2J_z^2 = (1/6930)J(J+1)(2J+1)(J-1)(J+2)(10J^6+30J^5-185J^4-420J^3+211J^2+426J-210)$
$\mathrm{Tr}\, J_x^4J_y^4J_z^2 = (1/6930)J(J+1)(2J+1)(J-1)(J+2)(6J^6+18J^5-166J^4-362J^3-199J^2-15J+105)$

TABLE III. Numerical values for the traces of products of the angular momentum matrices.

Operator	$J=1/2$	1	3/2	2	5/2	3	7/2	4	9/2	5
J_z^2	1/2	2	5	10	35/2	28	42	60	165/2	110
$J_zJ_yJ_z$	$i/4$	i	$5i/2$	$5i$	$35i/4$	$14i$	$21i$	$30i$	$165i/4$	$55i$
J_z^4	1/8	2	41/4	34	707/8	196	777/2	708	9669/8	1958
$J_z^2J_y^2$	1/8	1	17/4	13	259/8	70	273/2	246	3333/8	671
$J_zJ_yJ_zJ_y$	−1/8	0	7/4	8	189/8	56	231/2	216	3003/8	616
$J_z^3J_yJ_z$	$i/16$	i	$41i/8$	$17i$	$707i/16$	$98i$	$777i/4$	$354i$	$9669i/16$	$979i$
$J_z^3J_yJ_zJ_z$	$-i/16$	0	$7i/8$	$4i$	$189i/16$	$28i$	$231i/4$	$108i$	$3003i/16$	$308i$
J_z^6	1/32	2	365/16	130	16 355/32	1588	33 501/8	9780	665 445/32	41 030
$J_z^4J_y^2$	1/32	1	125/16	37	4195/32	382	7725/8	2190	145 893/32	8855
$J_z^3J_yJ_zJ_y$	−1/32	0	43/16	20	2781/32	284	6171/8	1836	126 555/32	7876
$J_z^2J_y^2J_z^2$	1/32	0	5/16	4	675/32	76	1749/8	540	38 181/32	2420
$J_z^2J_yJ_z^2J_y$	1/32	0	29/16	16	2403/32	256	5709/8	1728	120 549/32	7568
$J_z^2J_yJ_z^2J_y$	1/32	1	101/16	25	2467/32	202	3765/8	1002	63 525/32	3707
$J_z^2J_yJ_zJ_yJ_z$	−1/32	0	19/16	8	1053/32	104	2211/8	648	44 187/32	2728
$J_zJ_yJ_zJ_zJ_yJ_z$	−1/32	−1	−77/16	−13	−739/32	−22	195/8	186	18 843/32	1441
$J_zJ_yJ_zJ_zJ_yJ_z$	1/32	0	5/16	4	675/32	76	1749/8	540	38 181/32	2420
$J_z^5J_yJ_z$	$i/64$	i	$365i/32$	$65i$	$16\,355i/64$	$794i$	$33\,501i/16$	$4890i$	$665\,645i/64$	$20\,515i$
$J_z^4J_yJ_zJ_z$	$-i/64$	0	$115i/32$	$28i$	$7965i/64$	$412i$	$18\,051i/16$	$2700i$	$373\,659i/64$	$11\,660i$
$J_z^3J_yJ_z^2J_z$	$i/64$	0	$29i/32$	$8i$	$2403i/64$	$128i$	$5709i/16$	$864i$	$120\,549i/64$	$3784i$
$J_z^3J_y^3J_z$	$i/64$	i	$293i/32$	$47i$	$11\,171i/64$	$524i$	$21\,621i/16$	$3108i$	$418\,341i/64$	$12\,793i$
$J_z^3J_y^2J_zJ_y$	$-i/64$	0	$43i/32$	$10i$	$2781i/64$	$142i$	$6171i/16$	$918i$	$126\,555i/64$	$3938i$
Operator	$J=11/2$	6	13/2	7	15/2	8	17/2	9	19/2	10
J_z^2	143	182	455/2	280	340	408	969/2	570	665	770
$J_zJ_yJ_z$	$143i/2$	$91i$	$455i/4$	$140i$	$170i$	$204i$	$969i/4$	$285i$	$665i/2$	$385i$
J_z^4	12 155/4	4550	52 871/8	9352	12 937	17 544	187 017/8	30 666	158 669/4	50 666
$J_z^2J_y^2$	4147/4	1547	17 927/8	3164	4369	5916	62 985/8	10 317	53 333/4	17 017
$J_zJ_yJ_zJ_y$	3861/4	1456	17 017/8	3024	4199	5712	61 047/8	10 032	52 003/4	16 632
$J_z^3J_yJ_z$	$12\,155i/8$	$2275i$	$52\,871i/16$	$4676i$	$12\,937i/2$	$8772i$	$187\,017i/16$	$15\,333i$	$158\,669i/8$	$25\,333i$
$J_z^2J_yJ_zJ_z$	$3861i/8$	$728i$	$17\,017i/16$	$1512i$	$4199i/2$	$2856i$	$61\,047i/16$	$5016i$	$52\,003i/8$	$8316i$
J_z^6	1 218 503/16	134 342	7 263 815/32	369 640	2 331 805/4	893 928	42 792 009/32	1 956 810	44 918 945/16	3 956 810
$J_z^4J_y^2$	259 831/16	28 379	1 523 015/32	77 036	483 565/4	184 620	8 807 241/32	401 565	9 194 993/16	808 225
$J_z^3J_yJ_zJ_y$	235 521/16	26 104	1 417 273/32	72 360	457 691/4	175 848	8 433 207/32	386 232	8 877 655/16	782 892
$J_z^2J_y^2J_z^2$	73 359/16	8216	449 735/32	23 112	146 965/4	56 712	2 729 673/32	125 400	2 889 881/16	255 420
$J_z^2J_yJ_z^2J_y$	227 799/16	25 376	1 383 239/32	70 848	449 293/4	172 992	8 311 113/32	381 216	8 773 649/16	774 576
$J_z^2J_yJ_z^2J_y$	105 391/16	11 219	589 511/32	29 300	181 237/4	68 340	3 225 801/32	145 749	3 311 225/16	289 069
$J_z^2J_yJ_zJ_yJ_z$	81 081/16	8944	483 769/32	24 624	155 363/4	59 568	2 851 767/32	130 416	2 993 887/16	263 736
$J_zJ_yJ_zJ_zJ_yJ_z$	49 049/16	5941	343 993/32	18 436	121 091/4	47 940	2 355 639/32	110 067	2 572 543/16	230 087
$J_zJ_yJ_zJ_zJ_yJ_z$	73 359/16	8216	449 735/32	23 112	146 965/4	56 712	2 729 673/32	125 400	2 889 881/16	255 420
$J_z^5J_yJ_z$	$1\,218\,503i/32$	$67\,171i$	$7\,263\,815i/64$	$184\,820i$	$2\,331\,805i/8$	$446\,964i$	$42\,792\,009i/64$	$978\,405i$	$44\,918\,945i/32$	$1\,978\,405i$
$J_z^4J_yJ_zJ_z$	$698\,841i/32$	$38\,792i$	$4\,217\,785i/64$	$107\,784i$	$1\,364\,675i/8$	$262\,344i$	$25\,177\,527i/64$	$576\,840i$	$26\,528\,959i/32$	$1\,170\,180i$
$J_z^3J_yJ_z^2J_z$	$227\,799i/32$	$12\,688i$	$1\,383\,239i/64$	$35\,424i$	$449\,293i/8$	$86\,496i$	$8\,311\,113i/64$	$190\,608i$	$8\,773\,649i/32$	$387\,288i$
$J_z^3J_y^3J_z$	$755\,183i/32$	$41\,431i$	$4\,463\,303i/64$	$113\,216i$	$1\,424\,821i/8$	$272\,544i$	$26\,047\,689i/64$	$594\,681i$	$27\,267\,641i/32$	$1\,199\,671i$
$J_z^3J_y^2J_zJ_y$	$235\,521i/32$	$13\,052i$	$1\,417\,273i/64$	$36\,180i$	$457\,691i/8$	$87\,924i$	$8\,433\,207i/64$	$193\,116i$	$8\,877\,655i/32$	$391\,446i$

TABLE III (*continued*).

Operator	$J=1/2$	1	3/2	2	5/2	3	7/2	4	9/2	5
$J_x^2J_y^2J_xJ_yJ_z$	$i/64$	0	$53i/32$	$14i$	$4131i/64$	$218i$	$9669i/16$	$1458i$	$202\,917i/64$	$6358i$
$J_x^2J_y^2J_xJ_zJ_y$	$-i/64$	$-i$	$-197i/32$	$-23i$	$-4259i/64$	$-164i$	$-5781i/16$	$-732i$	$-88\,869i/64$	$-2497i$
$J_x^2J_yJ_zJ_y^2J_z$	$-i/64$	0	$91i/32$	$22i$	$6237i/64$	$322i$	$14\,091i/16$	$2106i$	$291\,291i/64$	$9086i$
$J_x^2J_yJ_zJ_yJ_xJ_y$	$i/64$	0	$5i/32$	$2i$	$675i/64$	$38i$	$1749i/16$	$270i$	$38\,181i/64$	$1210i$
$J_xJ_yJ_zJ_yJ_xJ_yJ_z$	$-i/64$	0	$43i/32$	$10i$	$2781i/64$	$142i$	$6171i/16$	$918i$	$126\,555i/64$	$3938i$
J_x^8	1/128	2	3281/64	514	397 187/128	13 636	1 540 497/32	144 708	49 208 707/128	925 958
$J_x^6J_y^2$	1/128	1	1097/64	133	87 619/128	2710	285 033/32	25 446	8 335 173/128	152 471
$J_x^6J_yJ_zJ_y$	$-1/128$	0	367/64	68	54 909/128	1916	218 031/32	20 556	7 004 283/128	131 956
$J_x^4J_y^2J_z^2J_y$	1/128	0	137/64	40	38 979/128	1504	181 929/32	17 856	6 256 965/128	120 296
$J_x^2J_yJ_z^2J_y^2$	$-1/128$	0	79/64	32	34 173/128	1376	170 511/32	16 992	6 015 867/128	116 512
$J_x^4J_y^4$	1/128	1	881/64	97	60 547/128	1810	186 033/32	16 338	5 287 557/128	95 843
$J_x^4J_y^2J_zJ_y$	$-1/128$	0	295/64	50	38 205/128	1286	142 791/32	13 230	4 450 875/128	83 050
$J_x^2J_y^2J_zJ_y^2$	1/128	0	209/64	40	32 643/128	1144	130 449/32	12 312	4 197 765/128	79 112
$J_x^2J_y^2J_z^2J_y^2$	1/128	1	497/64	49	32 899/128	1090	122 673/32	11 586	3 969 669/128	75 251
$J_x^2J_y^2J_xJ_yJ_zJ_y$	$-1/128$	0	103/64	26	24 381/128	926	111 111/32	10 854	3 791 931/128	72 754
$J_x^2J_yJ_zJ_y^2J_zJ_y$	1/128	0	113/64	28	25 731/128	964	114 609/32	11 124	3 868 293/128	73 964
$J_xJ_yJ_zJ_yJ_zJ_yJ_zJ_y$	1/128	0	17/64	16	18 819/128	784	98 769/32	9936	3 538 821/128	68 816
$J_x^4J_y^2J_z^2$	1/128	0	$-103/64$	-8	$-1341/128$	64	15 609/32	2016	820 677/128	17 336
$J_x^4J_yJ_z^2J_y$	1/128	1	857/64	85	47 299/128	1270	118 713/32	9606	2 898 885/128	49 511
$J_x^4J_yJ_zJ_yJ_z$	$-1/128$	0	127/64	20	14 589/128	476	51 711/32	4716	1 567 995/128	28 996
$J_x^2J_y^2J_zJ_z^2$	1/128	1	569/64	49	26 563/128	730	71 193/32	6042	1 910 469/128	34 067

Operator	$J=11/2$	6	13/2	7	15/2	8	17/2	9	19/2	10
$J_x^2J_y^2J_zJ_yJ_z$	$382\,239i/32$	$21\,268i$	$2\,316\,743i/64$	$59\,292i$	$751\,621i/8$	$144\,636i$	$13\,892\,553i/64$	$318\,516i$	$14\,657\,417i/32$	$646\,866i$
$J_x^2J_y^2J_zJ_zJ_y$	$-137\,423i/32$	$-7111i$	$-729\,287i/64$	$-17\,744i$	$-215\,509i/8$	$-39\,984i$	$-3\,721\,929i/64$	$-83\,049i$	$-3\,732\,569i/32$	$-161\,359i$
$J_x^2J_yJ_zJ_y^2J_z$	$544\,401i/32$	$30\,212i$	$3\,284\,281i/64$	$83\,916i$	$1\,062\,347i/8$	$204\,204i$	$19\,596\,087i/64$	$448\,932i$	$20\,645\,191i/32$	$910\,602i$
$J_x^2J_yJ_zJ_yJ_xJ_y$	$73\,359i/32$	$4108i$	$449\,735i/64$	$11\,556i$	$146\,965i/8$	$28\,356i$	$2\,729\,673i/64$	$62\,700i$	$2\,889\,881i/32$	$127\,710i$
$J_xJ_yJ_zJ_yJ_zJ_yJ_z$	$235\,521i/32$	$13\,052i$	$1\,417\,273i/64$	$36\,180i$	$457\,691i/8$	$87\,924i$	$8\,433\,207i/64$	$193\,116i$	$8\,877\,655i/32$	$391\,446i$
J_x^8	131 783 795/64	4 285 190	8.432018×10^6	1.581479×10^7	2.845460×10^7	4.936922×10^7	8.295270×10^7	1.354627×10^8	2.156368×10^8	3.354626×10^8
$J_x^6J_y^2$	21 231 067/64	678 587	1.316975×10^6	2 442 524	4.354270×10^6	7 496 796	1.251513×10^7	2.032512×10^7	3.220238×10^7	4.989322×10^7
$J_x^6J_yJ_zJ_y$	18 794 061/64	611 416	$1\,203478\times10^6$	2 257 704	4.062795×10^6	7 049 832	1.184650×10^7	1.934671×10^7	3.079866×10^7	4.791481×10^7
$J_x^4J_yJ_z^2J_y$	17 396 379/64	572 624	1.137575×10^6	2 149 920	3.892210×10^6	6 787 488	1.145310×10^7	1.876987×10^7	2.996963×10^7	4.674463×10^7
$J_x^2J_yJ_z^2J_y^2$	16 940 781/64	559 936	1.115962×10^6	2 114 496	3.836049×10^6	6 700 992	1.132324×10^7	1.857926×10^7	2.969545×10^7	4.635734×10^7
$J_x^4J_y^4$	13 251 667/64	421 187	8.137580×10^5	1 503 716	2.672571×10^6	4 589 796	7.645902×10^6	1.239482×10^7	1.960744×10^7	3.033834×10^7
$J_x^4J_y^2J_zJ_y$	11 741 301/64	379 756	7.440189×10^5	1 390 500	2.494468×10^6	4 317 252	7.238907×10^6	1.180014×10^7	1.875532×10^7	2.913867×10^7
$J_x^2J_y^2J_zJ_y^2$	11 270 259/64	366 704	7.218740×10^5	1 354 320	2.437257×10^6	4 229 328	7.107138×10^6	1.160702×10^7	1.847790×10^7	2.874722×10^7
$J_x^2J_y^2J_z^2J_y^2$	10 780 627/64	352 547	6.970700×10^5	1 312 772	2.370243×10^6	4 124 676	6.948222×10^6	1.137156×10^7	1.813649×10^7	2.826172×10^7
$J_x^2J_y^2J_zJ_yJ_zJ_y$	10 505 781/64	345 436	6.856749×10^5	1 295 028	2.343304×10^6	4 084 692	6.890067×10^6	1.128851×10^7	1.801985×10^7	2.810036×10^7
$J_x^2J_yJ_zJ_y^2J_zJ_y$	10 652 499/64	349 544	6.927020×10^5	1 306 584	2.361675×10^6	4 113 048	6.932718×10^6	1.135121×10^7	1.811016×10^7	2.822807×10^7
$J_xJ_yJ_zJ_yJ_zJ_yJ_zJ_y$	10 034 739/64	332 384	6.635301×10^5	1 258 848	2.286093×10^6	3 996 768	6.758298×10^6	1.109539×10^7	1.774242×10^7	2.770891×10^7
$J_x^4J_y^2J_z^2$	2 673 099/64	92 144	1.894850×10^5	367 776	6.799755×10^5	1 206 048	2.063493×10^6	3 420 912	5.515220×10^6	8 673 192
$J_x^4J_yJ_z^2J_y$	6 507 787/64	198 107	3.688850×10^5	660 380	1.142035×10^6	1 915 356	3.125517×10^6	4 976 157	7.747967×10^6	1.182178×10^7
$J_x^4J_yJ_zJ_yJ_z$	4 070 781/64	130 936	2.553879×10^5	475 560	8.505599×10^5	1 468 392	2.456892×10^6	3 997 752	6.344250×10^6	9 843 372
$J_x^2J_y^2J_zJ_z^2$	4 654 507/64	146 627	2.813691×10^5	517 172	9.152895×10^5	1 566 516	2.602257×10^6	4 208 709	6.644760×10^6	1.026431×10^7

TABLE III (*continued*).

Operator	$J=1/2$	1	3/2	2	5/2	3	7/2	4	9/2	5
$J_x^2J_y^2J_zJ_xJ_z$	−1/128	0	−17/64	2	4221/128	206	27 951/32	2934	1 073 787/128	21 274
$J_x^2J_yJ_z^2J_xJ_y$	−1/128	0	271/64	38	24 957/128	746	75 471/32	6498	2 062 203/128	36 718
$J_x^2J_yJ_zJ_xJ_yJ_z$	1/128	0	41/64	10	9027/128	334	39 369/32	3798	1 314 885/128	25 058
$J_x^2J_yJ_xJ_zJ_yJ_z$	−1/128	−1	−545/64	−37	−13 315/128	−190	−3873/32	690	478 203/128	12 265
$J_x^2J_yJ_xJ_zJ_zJ_y$	1/128	0	185/64	28	19 395/128	604	63 129/32	5580	1 809 093/128	32 780
$J_x^2J_y^2J_z^2J_x^2$	1/128	0	281/64	40	26 307/128	784	78 969/32	6768	2 138 565/128	37 928
$J_x^2J_y^2J_xJ_z^2J_z$	1/128	0	89/64	16	12 483/128	424	47 289/32	4392	1 479 621/128	27 632
$J_x^2J_yJ_z^2J_xJ_yJ_z$	−1/128	0	31/64	8	7677/128	296	35 871/32	3528	1 238 523/128	23 848
$J_x^2J_yJ_xJ_z^2J_yJ_z$	−1/128	0	−257/64	−28	−13 059/128	−244	−11 649/32	−36	250 107/128	8404
$J_x^2J_yJ_xJ_z^2J_xJ_y$	1/128	1	473/64	37	19 651/128	550	55 353/32	4854	1 580 997/128	28 919
$J_x^2J_y^2J_zJ_xJ_zJ_z$	−1/128	0	175/64	26	18 045/128	566	59 631/32	5310	1 732 731/128	31 570
$J_x^2J_yJ_zJ_x^2J_zJ_y$	1/128	0	185/64	28	19 395/128	604	63 129/32	5580	1 809 093/128	32 780
$J_x^2J_yJ_zJ_xJ_z^2J_yJ_z$	1/128	0	−151/64	−14	−4797/128	−26	7689/32	1422	655 941/128	14 762
$J_x^2J_yJ_zJ_xJ_zJ_xJ_y$	−1/128	0	79/64	14	11 133/128	386	43 791/32	4122	1 403 259/128	26 422
$J_x^2J_yJ_xJ_yJ_xJ_zJ_z$	1/128	0	−7/64	4	5571/128	244	31 449/32	3204	1 150 149/128	22 484
$J_x^2J_yJ_zJ_xJ_yJ_zJ_x$	−1/128	0	−65/64	−4	765/128	116	20 031/32	2340	909 051/128	18 700
$J_xJ_yJ_xJ_yJ_zJ_xJ_zJ_z$	1/128	0	89/64	16	12 483/128	424	47 289/32	4392	1 479 621/128	27 632
$J_xJ_yJ_zJ_xJ_zJ_yJ_zJ_x$	−1/128	0	−161/64	−16	−6147/128	−64	4191/32	1152	579 579/128	13 552
$J_x^7J_yJ_z$	i/256	i	3281i/128	257i	397 187i/256	6818i	1 540 497i/64	72 354i	$1.922215\times10^{6}i$	462 979i
$J_x^6J_yJ_zJ_z$	−i/256	0	1087i/128	124i	221 949i/256	4108i	970 431i/64	46 908i	$1.271030\times10^{6}i$	310 508i
$J_x^6J_yJ_z^2J_z$	i/256	0	353i/128	56i	112 131i/256	2192i	534 369i/64	26 352i	18 529 797i/256	178 552i

Operator	$J=11/2$	6	13/2	7	15/2	8	17/2	9	19/2	10
$J_x^2J_y^2J_zJ_xJ_z$	3 144 141/64	105 196	2.116299×10^{6}	403 956	7.371869×10^{6}	1 293 972	2.195262×10^{6}	3 614 028	5.792646×10^{6}	9 064 638
$J_x^2J_yJ_z^2J_xJ_y$	4 997 421/64	156 676	2.991459×10^{6}	547 164	9.639329×10^{6}	1 642 812	2.718522×10^{6}	4 381 476	6.895853×10^{6}	1.062211×10^{7}
$J_x^2J_yJ_zJ_xJ_yJ_z$	3 599 739/64	117 884	2.332431×10^{6}	439 380	7.933485×10^{6}	1 380 468	2.325123×10^{6}	3 804 636	6.066823×10^{6}	9 451 926
$J_x^2J_yJ_xJ_zJ_yJ_z$	2 089 373/64	76 453	1.635039×10^{6}	326 164	6.152459×10^{6}	1 107 924	1.918128×10^{6}	3 209 955	5.214709×10^{6}	8 252 255
$J_x^2J_yJ_xJ_zJ_xJ_y$	4 526 379/64	143 624	2.770010×10^{6}	510 984	9.067215×10^{6}	1 554 888	2.586753×10^{6}	4 188 360	6.618426×10^{6}	1.023066×10^{7}
$J_x^2J_y^2J_z^2J_x^2$	5 144 139/64	160 784	3.061730×10^{6}	558 720	9.823035×10^{6}	1 671 168	2.761173×10^{6}	4 444 476	6.986162×10^{6}	1.074982×10^{7}
$J_x^2J_y^2J_xJ_z^2J_z$	3 908 619/64	126 464	2.478290×10^{6}	463 248	8.311395×10^{6}	1 438 608	2.412333×10^{6}	3 932 544	6.250691×10^{6}	9 711 504
$J_x^2J_yJ_z^2J_xJ_yJ_z$	3 453 021/64	113 776	2.262159×10^{6}	427 824	7.749779×10^{6}	1 352 112	2.282472×10^{6}	3 741 936	5.976514×10^{6}	9 324 216
$J_x^2J_yJ_xJ_z^2J_yJ_z$	1 599 741/64	62 296	1.386999×10^{6}	284 616	5.482319×10^{6}	1 003 272	1.759212×10^{6}	2 974 488	4.873308×10^{6}	7 766 748
$J_x^2J_yJ_xJ_z^2J_xJ_y$	4 036 747/64	129 467	2.521970×10^{6}	469 436	8.397075×10^{6}	1 450 236	2.427837×10^{6}	3 952 893	6.277025×10^{6}	9 745 153
$J_x^2J_y^2J_zJ_xJ_zJ_z$	4 379 661/64	139 516	2.699739×10^{6}	499 428	8.883509×10^{6}	1 526 532	2.544102×10^{6}	4 125 660	6.528117×10^{6}	1.010295×10^{7}
$J_x^2J_yJ_zJ_x^2J_zJ_y$	4 526 379/64	143 624	2.770010×10^{6}	510 984	9.067215×10^{6}	1 554 888	2.586753×10^{6}	4 188 360	6.618426×10^{6}	1.023066×10^{7}
$J_x^2J_yJ_zJ_xJ_yJ_xJ_z$	2 364 219/64	83 564	1.748990×10^{6}	343 908	6.421845×10^{6}	1 147 908	1.976283×10^{6}	3 293 004	5.331352×10^{6}	8 413 614
$J_x^2J_yJ_zJ_xJ_zJ_xJ_y$	3 761 901/64	122 356	2.408019×10^{6}	451 692	8.127689×10^{6}	1 410 252	2.369682×10^{6}	3 869 844	6.160382×10^{6}	9 583 794
$J_x^2J_yJ_zJ_yJ_xJ_zJ_z$	3 290 859/64	109 304	2.186570×10^{6}	415 512	7.555575×10^{6}	1 322 328	2.237913×10^{6}	3 676 728	5.882955×10^{6}	9 192 348
$J_x^2J_yJ_zJ_xJ_yJ_zJ_x$	2 835 261/64	96 616	1.970439×10^{6}	380 088	6.993959×10^{6}	1 235 832	2.108052×10^{6}	3 486 120	5.608779×10^{6}	8 805 060
$J_xJ_yJ_xJ_yJ_zJ_xJ_zJ_z$	3 908 619/64	126 464	2.478290×10^{6}	463 248	8.311395×10^{6}	1 438 608	2.412333×10^{6}	3 932 544	6.250691×10^{6}	9 711 504
$J_xJ_yJ_zJ_xJ_zJ_yJ_zJ_x$	2 217 501/64	79 456	1.678719×10^{6}	332 352	6.238139×10^{6}	1 119 552	1.933632×10^{6}	3 230 304	5.241043×10^{6}	8 285 904
$J_x^7J_yJ_z$	$1.029561\times10^{6}i$	2 142 595i	$4.216009\times10^{6}i$	7 907 396i	$1.422730\times10^{7}i$	$2.468461\times10^{7}i$	$4.147635\times10^{7}i$	$6.773133\times10^{7}i$	$1.078184\times10^{8}i$	$1.677313\times10^{8}i$
$J_x^6J_yJ_zJ_z$	89 321 659i/128	1 464 008i	$2.899034\times10^{6}i$	5 464 872i	$9.873030\times10^{6}i$	$1.718782\times10^{7}i$	$2.896122\times10^{7}i$	$4.740621\times10^{7}i$	$7.561602\times10^{7}i$	$1.178381\times10^{8}i$
$J_x^6J_yJ_z^2J_z$	51 733 537i/128	852 592i	$1.695556\times10^{6}i$	3 207 168i	$5.810235\times10^{6}i$	$1.013798\times10^{7}i$	$1.711472\times10^{7}i$	$2.805950\times10^{7}i$	$4.481736\times10^{7}i$	$6.992330\times10^{7}i$

TABLE III (*Continued*).

Operator	$J=1/2$	1	3/2	2	5/2	3	7/2	4	9/2	5
$J_x^4J_yJ_x^3J_z$	$-i/256$	0	$79i/128$	$16i$	$34\,173i/256$	$688i$	$170\,511i/64$	$8496i$	$6\,015\,867i/256$	$58\,256i$
$J_x^5J_y^3J_z$	$i/256$	i	$2561i/128$	$167i$	$230\,147i/256$	$3668i$	$788\,097i/64$	$35\,724i$	$23\,674\,629i/256$	$218\,449i$
$J_x^5J_y^3J_zJ_y$	$-i/256$	0	$367i/128$	$34i$	$54\,909i/256$	$958i$	$218\,031i/64$	$10\,278i$	$7\,004\,283i/256$	$65\,978i$
$J_x^4J_y^3J_zJ_z$	$-i/256$	0	$799i/128$	$70i$	$109\,053i/256$	$1858i$	$416\,031i/64$	$19\,386i$	$13\,099\,515i/256$	$122\,606i$
$J_x^3J_y^3J_x^2J_yJ_z$	$i/256$	0	$593i/128$	$62i$	$106\,371i/256$	$1922i$	$447\,249i/64$	$21\,402i$	$14\,740\,869i/256$	$139\,942i$
$J_x^4J_y^2J_zJ_zJ_y$	$-i/256$	$-i$	$-1601i/128$	$-71i$	$-68\,867i/256$	$-788i$	$-122\,817i/64$	$-4044i$	$-1\,929\,477i/256$	$-12\,529i$
$J_x^4J_y^3J_xJ_zJ_y$	$i/256$	0	$161i/128$	$26i$	$52\,227i/256$	$1022i$	$249\,249i/64$	$12\,294i$	$8\,645\,637i/256$	$83\,314i$
$J_x^4J_xJ_y^2J_zJ_y$	$i/256$	0	$-847i/128$	$-82i$	$-135\,549i/256$	$-2398i$	$-550\,671i/64$	$-26\,118i$	$-17\,876\,859i/256$	$-168\,938i$
$J_x^4J_yJ_zJ_yJ_zJ_y$	$i/256$	0	$113i/128$	$14i$	$25\,731i/256$	$482i$	$114\,609i/64$	$5562i$	$3\,868\,293i/256$	$36\,982i$
$J_x^3J_y^3J_z^2J_z$	$i/256$	0	$209i/128$	$20i$	$32\,643i/256$	$572i$	$130\,449i/64$	$6156i$	$4\,197\,765i/256$	$39\,556i$
$J_x^3J_y^3J_z^2J_yJ_z$	$i/256$	i	$1313i/128$	$71i$	$94\,211i/256$	$1508i$	$328\,737i/64$	$15\,132i$	$10\,166\,277i/256$	$94\,897i$
$J_x^3J_y^3J_z^2J_xJ_y$	$-i/256$	0	$-881i/128$	$-62i$	$-81\,027i/256$	$-1202i$	$-241\,329i/64$	$-10\,314i$	$-6\,504\,069i/256$	$-57\,574i$
$J_x^3J_y^3J_xJ_z^2J_y$	$-i/256$	0	$127i/128$	$28i$	$60\,669i/256$	$1228i$	$305\,151i/64$	$15\,228i$	$10\,793\,211i/256$	$104\,588i$
$J_x^3J_yJ_z^2J_y^2J_z$	$i/256$	0	$305i/128$	$44i$	$85\,635i/256$	$1652i$	$399\,729i/64$	$19\,620i$	$13\,752\,453i/256$	$132\,220i$
$J_x^3J_yJ_z^2J_yJ_zJ_y$	$-i/256$	0	$31i/128$	$4i$	$7677i/256$	$148i$	$35\,871i/64$	$1764i$	$1\,238\,523i/256$	$11\,924i$
$J_x^3J_y^3J_zJ_yJ_z^2$	$-i/256$	0	$175i/128$	$22i$	$41\,085i/256$	$778i$	$186\,351i/64$	$9090i$	$6\,345\,339i/256$	$60\,830i$
$J_x^3J_y^3J_z^2J_zJ_y$	$i/256$	0	$-463i/128$	$-22i$	$-15\,741i/256$	$-58i$	$19\,569i/64$	$1998i$	$1\,891\,461i/256$	$21\,538i$
$J_x^3J_yJ_zJ_y^2J_z^2$	$i/256$	0	$257i/128$	$32i$	$59\,139i/256$	$1112i$	$265\,089i/64$	$12\,888i$	$8\,975\,109i/256$	$85\,888i$
$J_x^3J_yJ_zJ_yJ_zJ_yJ_z$	$-i/256$	0	$223i/128$	$34i$	$67\,581i/256$	$1318i$	$320\,991i/64$	$15\,822i$	$11\,122\,683i/256$	$107\,162i$
$J_x^3J_yJ_zJ_yJ_zJ_xJ_y$	$i/256$	0	$-511i/128$	$-34i$	$-42\,237i/256$	$-598i$	$-115\,071i/64$	$-4734i$	$-2\,885\,883i/456$	$-24\,794i$
$J_x^3J_yJ_xJ_yJ_zJ_zJ_y$	$-i/256$	0	$79i/128$	$16i$	$34\,173i/256$	$688i$	$170\,511i/64$	$8496i$	$6\,015\,867i/256$	$58\,256i$

Operator	$J=11/2$	6	13/2	7	15/2	8	17/2	9	19/2	10
$J_x^4J_yJ_x^3J_z$	1.323498×10^6i	$279\,968i$	$142\,843\,127i/256$	$1\,057\,248i$	1.9180294×10^6i	$3\,350\,496i$	5.661621×10^6i	9.289632×10^6i	1.484773×10^7i	2.317867×10^7i
$J_x^5J_y^3J_z$	4.785640×10^6i	$984\,295i$	1.918714×10^6i	$3\,571\,376i$	6.385668×10^6i	1.102171×10^7i	1.843838×10^7i	2.999847×10^7i	4.760171×10^7i	7.385062×10^7i
$J_x^5J_y^3J_zJ_y$	1.468286×10^6i	$305\,708i$	$154\,045\,175i/256$	$1\,128\,852i$	2.031397×10^6i	$3\,524\,916i$	5.923251×10^6i	9.673356×10^6i	1.539933×10^7i	2.395741×10^7i
$J_x^4J_y^3J_zJ_z$	2.715067×10^6i	$563\,108i$	1.104956×10^6i	$2\,067\,660i$	3.713097×10^6i	6.431916×10^6i	1.079248×10^7i	1.760365×10^7i	2.799427×10^7i	4.351228×10^7i
$J_x^4J_y^3J_zJ_yJ_z$	3.132739×10^6i	$655\,252i$	1.294441×10^6i	$2\,435\,436i$	4.393072×10^6i	7.637964×10^6i	1.285597×10^7i	2.102456×10^7i	3.350949×10^7i	5.218547×10^7i
$J_x^4J_y^3J_zJ_xJ_y$	$-2\,363\,075i/128$	$-23\,335i$	$-5\,768\,711i/256$	$-7088i$	$1\,241\,663i/32$	$141\,168i$	3.408420×10^6i	$699\,447i$	1.307113×10^6i	$2\,292\,257i$
$J_x^4J_y^3J_xJ_zJ_y$	1.885958×10^6i	$397\,852i$	$202\,553\,344i/256$	$1\,496\,628i$	2.711373×10^6i	$4\,730\,964i$	7.986744×10^6i	1.309427×10^7i	2.091455×10^7i	3.263060×10^7i
$J_x^4J_xJ_y^3J_zJ_y$	-3.768798×10^6i	$-786\,188i$	-1.549829×10^6i	-2.910996×10^6i	-5.243632×10^6i	$-9.106\,356\times10^6i$	-1.531286×10^7i	-2.502232×10^7i	-3.985374×10^7i	-6.202885×10^7i
$J_x^4J_yJ_xJ_yJ_zJ_y$	$10\,652\,499i/128$	$174\,772i$	$88\,665\,863i/256$	$653\,292i$	1.180838×10^6i	$2\,056\,524i$	3.466359×10^6i	$5\,675\,604i$	9.055080×10^6i	1.411403×10^7i
$J_x^3J_y^3J_z^2J_z$	$11\,270\,259i/128$	$183\,352i$	3.609370×10^6i	$677\,160i$	1.218629×10^6i	$2\,114\,664i$	3.553569×10^6i	$5\,803\,512i$	9.238948×10^6i	1.437361×10^7i
$J_x^3J_y^3J_z^2J_yJ_z$	2.099028×10^6i	$435\,175i$	8.539360×10^6i	$1\,598\,288i$	2.871105×10^6i	$4\,975\,152i$	8.351088×10^6i	1.362625×10^7i	2.167635×10^7i	3.370256×10^7i
$J_x^3J_y^3J_z^2J_zJ_y$	-1.218326×10^6i	$-243\,412i$	$-118\,537\,991i/256$	$-844\,236i$	-1.483166×10^6i	$-2\,521\,644i$	-4.164039×10^6i	$-6\,698\,868i$	-1.052602×10^7i	-1.619066×10^7i
$J_x^3J_y^3J_zJ_z^2J_y$	2.377230×10^6i	$503\,048i$	1.002854×10^6i	$1\,900\,584i$	3.448560×10^6i	6.024936×10^6i	1.0182006×10^7i	1.670830×10^7i	2.670720×10^7i	4.169524×10^7i
$J_x^3J_yJ_z^2J_z^2J_z$	2.987951×10^6i	$629\,512i$	1.250683×10^6i	$2\,363\,832i$	4.279699×10^6i	7.463544×10^6i	1.259434×10^7i	2.064084×10^7i	3.295789×10^7i	5.140674×10^7i
$J_x^3J_yJ_z^2J_yJ_zJ_y$	$3\,453\,021i/128$	$56\,888i$	1.131080×10^6i	$213\,912i$	$12\,399\,647i/32$	$676\,056i$	1.141236×10^6i	$1\,870\,968i$	2.988257×10^6i	$4\,662\,108i$
$J_x^3J_y^3J_zJ_yJ_zJ_z$	$17\,558\,541i/128$	$288\,548i$	5.725670×10^6i	$1\,081\,116i$	1.955815×10^6i	$3\,408\,636i$	5.748831×10^6i	9.417540×10^6i	1.503159×10^7i	2.343825×10^7i
$J_x^3J_y^3J_zJ_zJ_xJ_y$	$6\,945\,939i/128$	$123\,292i$	2.588350×10^6i	$510\,084i$	$1\,908\,183i/2$	$1\,707\,684i$	2.943099×10^6i	$4\,908\,156i$	7.951874×10^6i	1.255657×10^7i
$J_x^3J_yJ_xJ_y^2J_zJ_z$	$24\,758\,019i/128$	$406\,432i$	$805\,810i$	$1\,520\,496i$	2.749164×10^6i	$4\,789\,104i$	8.073954×10^6i	1.322218×10^7i	2.109842×10^7i	3.289018×10^7i
$J_x^3J_yJ_zJ_yJ_zJ_yJ_z$	$31\,046\,301i/128$	$511\,628i$	1.017440×10^6i	$1\,924\,452i$	3.486351×10^6i	6.083076×10^6i	1.026922×10^7i	1.683620×10^7i	2.689106×10^7i	4.195481×10^7i
$J_x^3J_yJ_zJ_yJ_zJ_xJ_y$	$-6\,541\,821i/128$	$-99\,788i$	-1.860380×10^6i	$-333\,252i$	-5.764440×10^6i	$-966\,756i$	-1.577286×10^6i	$-2\,510\,508i$	-3.907596×10^6i	-5.959998×10^6i
$J_x^3J_yJ_zJ_yJ_zJ_zJ_y$	$16\,940\,781i/128$	$279\,968i$	5.579810×10^6i	$1\,057\,248i$	1.918024×10^6i	$3\,350\,496i$	5.661621×10^6i	$9.289\,632\times10^6i$	1.484773×10^7i	2.317867×10^7i

TABLE III (*Continued*).

Operator	$J=1/2$	1	3/2	2	5/2	3	7/2	4	9/2	5
$J_x^2J_y^2J_x^2J_yJ_xJ_z$	$-i/256$	0	$319i128$	$22i$	$28\,413i/256$	$418i$	$83\,391i/64$	$3546i$	$2\,226\,939i/256$	$19\,646i$
$J_x^2J_y^2J_xJ_yJ_x^2J_z$	$i/256$	0	$-31i/128$	$-4i$	$-7677i/256$	$-148i$	$-35\,871i/64$	$-1764i$	$-1\,238\,523i/256$	$-11\,924i$
$J_x^2J_y^2J_xJ_zJ_x^2J_y$	$-i/256$	0	$-113i/128$	$4i$	$20\,349i/256$	$508i$	$138\,831i/64$	$7308i$	$5\,356\,923i/256$	$53\,108i$
$J_x^2J_yJ_x^2J_yJ_xJ_yJ_z$	$i/256$	0	$65i/128$	$20i$	$45\,315i/256$	$932i$	$233\,409i/64$	$11\,700i$	$8\,316\,165i/256$	$80\,740i$
$J_x^2J_yJ_x^2J_yJ_xJ_zJ_y$	$-i/256$	0	$-209i/128$	$-20i$	$-32\,643i/256$	$-572i$	$-130\,449i/64$	$-6156i$	$-4\,197\,765i/256$	$-39\,556i$
$J_x^2J_yJ_xJ_yJ_x^2J_yJ_z$	$-i/256$	0	$367i/128$	$34i$	$54\,909i/256$	$958i$	$218\,031i/64$	$10\,278i$	$7\,004\,283i/256$	$65\,978i$
$J_x^2J_yJ_xJ_yJ_xJ_yJ_xJ_z$	$i/256$	0	$17i/128$	$8i$	$18\,819i/256$	$392i$	$98\,769i/64$	$4968i$	$3\,538\,821i/256$	$34\,408i$
$J_x^2J_yJ_xJ_yJ_xJ_zJ_xJ_y$	$-i/256$	0	$-161i/128$	$-8i$	$-6147i/256$	$-32i$	$4191i/64$	$576i$	$5\,795\,79i/256$	$6776i$
$J_x^2J_y^2J_x^2$	$i/256$	i	$1625i/128$	$95i$	$128\,195i/256$	$2048i$	$443\,577i/64$	$20\,280i$	$13\,543\,365i/256$	$125\,785i$
$J_x^2J_y^2J_xJ_y$	$-i/256$	0	$727i/128$	$52i$	$69\,309i/256$	$1048i$	$214\,071i/64$	$9288i$	$5\,933\,499i/256$	$53\,108i$
$J_x^2J_y^2J_xJ_yJ_x$	$i/256$	0	$521i/128$	$44i$	$66\,627i/256$	$1112i$	$245\,289i/64$	$11\,304i$	$7\,574\,853i/256$	$70\,444i$
$J_x^2J_y^2J_xJ_yJ_x^2$	$-i/256$	0	$487i/128$	$46i$	$75\,069i/256$	$1318i$	$301\,191i/64$	$14\,238i$	$9\,722\,427i/256$	$91\,718i$
$J_x^2J_yJ_x^2J_y^2J_x$	$i/256$	i	$1529i/128$	$71i$	$75\,203i/256$	$968i$	$174\,297i/64$	$6816i$	$3\,988\,677i/256$	$33\,121i$
$J_x^2J_yJ_x^2J_yJ_xJ_y$	$-i/256$	0	$-185i/128$	$-14i$	$-19\,395i/256$	$-302i$	$-63\,129i/64$	$-2790i$	$-1\,809\,093i/256$	$-16\,390i$
$J_x^2J_yJ_xJ_yJ_xJ_yJ_z$	$-i/256$	0	$439i/128$	$34i$	$48\,573i/256$	$778i$	$166\,551i/64$	$7506i$	$4\,945\,083i/256$	$45\,386i$
$J_x^2J_y^2J_x^2J_xJ_yJ_z$	$i/256$	0	$-247i/128$	$-4i$	$11\,331i/256$	$392i$	$118\,569i/64$	$6552i$	$4\,939\,077i/256$	$49\,852i$
$J_x^2J_y^2J_x^2J_xJ_zJ_y$	$-i/256$	0	$-41i/128$	$4i$	$14\,013i/256$	$328i$	$87\,351i/64$	$4536i$	$3\,297\,723i/256$	$32\,516i$
$J_x^2J_y^2J_xJ_z^2J_yJ_z$	$i/256$	0	$-391i/128$	$-22i$	$-22\,077i/256$	$-238i$	$-31\,911i/64$	$-774i$	$-167\,739i/256$	$946i$
$J_x^2J_yJ_x^2J_zJ_y^2J_z$	$-i/256$	0	$1111i/128$	$76i$	$96\,957i/256$	$1408i$	$277\,431i/64$	$11\,664i$	$7\,251\,387i/256$	$63\,404i$
$J_x^2J_y^2J_xJ_zJ_xJ_yJ_z$	$-i/256$	0	$-425i/128$	$-20i$	$-13\,635i/256$	$-32i$	$23\,991i/64$	$2160i$	$1\,979\,835i/256$	$22\,220i$
$J_x^2J_y^2J_xJ_yJ_xJ_zJ_x$	$i/256$	0	$137i/128$	$20i$	$38\,979i/256$	$752i$	$181\,929i/64$	$8928i$	$6\,256\,965i/256$	$60\,148i$

Operator	$J=11/2$	6	13/2	7	15/2	8	17/2	9	19/2	10
$J_x^2J_y^2J_x^2J_yJ_xJ_z$	$5\,306\,301i/128$	$82\,628i$	$40\,157\,689i/256$	$285\,516i$	$16\,027\,583i/32$	$850\,476i$	$1.402866\times10^{6}i$	$2\,254\,692i$	$3.539860\times10^{6}i$	$5\,440\,842i$
$J_x^2J_y^2J_xJ_yJ_x^2J_z$	$-3\,453\,021i/128$	$-56\,888i$	$-28\,955\,641i/256$	$-213\,912i$	$-12\,399\,647i/32$	$-676\,056i$	$-1.141236\times10^{6}i$	$-1\,870\,968i$	$-2.988257\times10^{6}i$	$-4\,662\,108i$
$J_x^2J_y^2J_xJ_zJ_x^2J_y$	$15\,705\,261i/128$	$262\,808i$	$135\,375\,095i/256$	$1\,009\,512i$	$1.842442\times10^{6}i$	$3\,234\,216i$	$5.487201\times10^{6}i$	$9\,033\,816i$	$1.447999\times10^{7}i$	$2.265952\times10^{7}i$
$J_x^2J_yJ_x^2J_yJ_xJ_yJ_z$	$23\,522\,499i/128$	$389\,272i$	$7.766380\times10^{6}i$	$1\,472\,760i$	$2.673582\times10^{6}i$	$4\,672\,824i$	$7.899534\times10^{6}i$	$1.296636\times10^{7}i$	$2.073068\times10^{7}i$	$3.237102\times10^{7}i$
$J_x^2J_yJ_x^2J_yJ_xJ_zJ_y$	$-11\,270\,259i/128$	$-183\,352i$	$-3.609370\times10^{6}i$	$-677\,160i$	$-1.218629\times10^{6}i$	$-2\,114\,664i$	$-3.553569\times10^{6}i$	$-5\,803\,512i$	$-9.238948\times10^{6}i$	$-1.437361\times10^{7}i$
$J_x^2J_yJ_xJ_yJ_x^2J_yJ_z$	$18\,794\,061i/128$	$305\,708i$	$6.017390\times10^{6}i$	$1\,128\,852i$	$2.031397\times10^{6}i$	$3\,524\,916i$	$5.923251\times10^{6}i$	$9.673356\times10^{6}i$	$1.539933\times10^{7}i$	$2.395741\times10^{7}i$
$J_x^2J_yJ_xJ_yJ_xJ_yJ_xJ_z$	$10\,034\,739i/128$	$166\,192i$	$3.317650\times10^{6}i$	$629\,424i$	$1.143047\times10^{6}i$	$1\,998\,384i$	$3.379149\times10^{6}i$	$5.547696\times10^{6}i$	$8.871212\times10^{6}i$	$1.385446\times10^{7}i$
$J_x^2J_yJ_xJ_yJ_xJ_zJ_xJ_y$	$2\,217\,501i/128$	$39\,728i$	$21\,487\,609i/256$	$166\,176i$	$9\,981\,023i/32$	$559\,776i$	$9.668159\times10^{5}i$	$1\,615\,152i$	$2.620522\times10^{6}i$	$4\,142\,952i$
$J_x^2J_y^2J_x^3$	$35\,464\,715i/128$	$572\,455i$	$1.120\,131\times10^{6}i$	$2\,091\,560i$	$3.749746\times10^{6}i$	$6.486792\times10^{6}i$	$1.087291\times10^{7}i$	$1.771930\times10^{7}i$	$2.815769\times10^{7}i$	$4.373957\times10^{7}i$
$J_x^2J_y^2J_x^2J_y$	$14\,521\,221i/128$	$228\,488i$	$112\,037\,495i/256$	$802\,656i$	$1.417294\times10^{6}i$	$2\,420\,256i$	$4.011898\times10^{6}i$	$6\,475\,656i$	$1.020507\times10^{7}i$	$1.573744\times10^{7}i$
$J_x^2J_y^2J_x^2J_yJ_x$	$19\,867\,419i/128$	$320\,632i$	$6.271315\times10^{6}i$	$1\,170\,432i$	$2.097269\times10^{6}i$	$3\,626\,304i$	$6.075391\times10^{6}i$	$9.896568\times10^{6}i$	$1.572029\times10^{7}i$	$2.441063\times10^{7}i$
$J_x^2J_y^2J_xJ_yJ_x^2$	$26\,155\,701i/128$	$425\,828i$	$8.387615\times10^{6}i$	$1\,574\,388i$	$2.834456\times10^{6}i$	$4\,920\,276i$	$8.270653\times10^{6}i$	$1.351060\times10^{7}i$	$2.151293\times10^{7}i$	$3.347527\times10^{7}i$
$J_x^2J_yJ_x^2J_y^2J_x$	$8\,489\,195i/128$	$126\,295i$	$2.303845\times10^{6}i$	$404\,888i$	$6.886748\times10^{6}i$	$1\,137\,912i$	$1.832141\times10^{6}i$	$2\,881\,977i$	$4.438754\times10^{6}i$	$6\,706\,447i$
$J_x^2J_yJ_x^2J_yJ_xJ_y$	$-4\,526\,379i/128$	$-71\,812i$	$-1.385005\times10^{6}i$	$-255\,492i$	$-14\,507\,545i/32$	$-777\,444i$	$-1.293377\times10^{6}i$	$-2\,094\,180i$	$-3.309213\times10^{6}i$	$-5\,115\,330i$
$J_x^2J_yJ_xJ_yJ_xJ_yJ_x$	$12\,667\,941i/128$	$202\,748i$	$3.938885\times10^{6}i$	$731\,052i$	$1.303921\times10^{6}i$	$2\,245\,836i$	$3.750268\times10^{6}i$	$6.091932\times10^{6}i$	$9.653463\times10^{6}i$	$1.495870\times10^{7}i$
$J_x^2J_y^2J_x^2J_xJ_yJ_z$	$14\,925\,339i/128$	$251\,992i$	$5.104435\times10^{6}i$	$979\,488i$	$1.794941\times10^{6}i$	$3\,161\,184i$	$5.377711\times10^{6}i$	$8.873304\times10^{6}i$	$1.424934\times10^{7}i$	$2.233400\times10^{7}i$
$J_x^2J_y^2J_x^2J_zJ_xJ_y$	$9\,579\,141i/128$	$159\,848i$	$3.209585\times10^{6}i$	$611\,712i$	$1.114966\times10^{6}i$	$1\,955\,136i$	$3.314218\times10^{6}i$	$5\,452\,392i$	$8.734124\times10^{6}i$	$1.366081\times10^{7}i$
$J_x^2J_y^2J_xJ_z^2J_yJ_z$	$819\,819i/128$	$20\,332i$	$13\,052\,039i/256$	$112\,284i$	$7\,251\,673i/32$	$428\,604i$	$7.701165\times10^{5}i$	$1\,326\,732i$	$2.206007\times10^{6}i$	$3\,557\,862i$
$J_x^2J_yJ_x^2J_zJ_y^2J_z$	$16\,992\,261i/128$	$262\,808i$	$4.959905\times10^{6}i$	$898\,128i$	$1.568458\times10^{6}i$	$2\,652\,816i$	$4.360738\times10^{6}i$	$6.987288\times10^{6}i$	$1.094054\times10^{7}i$	$1.677575\times10^{7}i$
$J_x^2J_y^2J_xJ_zJ_xJ_yJ_z$	$7\,108\,101i/128$	$125\,528i$	$2.626145\times10^{6}i$	$516\,240i$	$9.638017\times10^{5}i$	$1\,722\,576i$	$2.965378\times10^{6}i$	$4.940760\times10^{6}i$	$7.998653\times10^{6}i$	$1.262250\times10^{7}i$
$J_x^2J_y^2J_xJ_yJ_xJ_zJ_x$	$17\,396\,379i/128$	$286\,312i$	$5.687875\times10^{6}i$	$1\,074\,960i$	$1.946105\times10^{6}i$	$3\,393\,744i$	$5.726551\times10^{6}i$	$9\,384\,936i$	$1.498482\times10^{7}i$	$2.337232\times10^{7}i$

TABLE III (*continued*).

Operator	$J=1/2$	1	3/2	2	5/2	3	7/2	4	9/2	5
$J_x^2J_y^2J_z^2J_xJ_yJ_zJ_y$	$i/256$	0	$569i/128$	$38i$	$47\,043i/256$	$662i$	$126\,489i/64$	$5166i$	$3\,126\,981i/256$	$26\,686i$
$J_x^2J_y^2J_z^2J_yJ_xJ_zJ_y$	$-i/256$	$-i$	$-1145i/128$	$-47i$	$-47\,555i/256$	$-608i$	$-110\,937i/64$	$-4440i$	$-2\,670\,789i/256$	$-22\,825i$
$J_x^2J_y^2J_z^2J_yJ_zJ_xJ_y$	$i/256$	0	$-7i/128$	$2i$	$5571i/256$	$122i$	$31\,449i/64$	$1602i$	$1\,150\,149i/256$	$11\,242i$
$J_x^2J_y^2J_z^2J_yJ_xJ_yJ_z$	$i/256$	0	$-151i/128$	$-16i$	$-27\,837i/256$	$-508i$	$-119\,031i/64$	$-5724i$	$-3\,956\,667i/256$	$-37\,664i$
$J_x^2J_y^2J_z^2J_yJ_zJ_yJ_z$	$i/256$	0	$89i/128$	$8i$	$12\,483i/256$	$212i$	$47\,289i/64$	$2196i$	$1\,479\,621i/256$	$13\,816i$
$J_x^2J_y^2J_z^2J_zJ_yJ_xJ_yJ_z$	$-i/256$	0	$55i/128$	$10i$	$20\,925i/256$	$418i$	$103\,191i/64$	$5130i$	$3\,627\,195i/256$	$35\,090i$
$J_x^2J_y^2J_z^2J_zJ_yJ_zJ_y$	$-i/256$	0	$199i/128$	$10i$	$8253i/256$	$58i$	$231i/64$	$-414i$	$-491\,205i/256$	$-6094i$
$J_x^2J_y^2J_z^2J_yJ_zJ_xJ_yJ_z$	$i/256$	0	$-487i/128$	$-28i$	$-28\,989i/256$	$-328i$	$-47\,751i/64$	$-1368i$	$-497\,211i/256$	$-1628i$
$J_xJ_yJ_zJ_xJ_zJ_yJ_xJ_yJ_z$	$i/256$	0	$-103i/128$	$-4i$	$-1341i/256$	$32i$	$15\,609i/64$	$1008i$	$820\,677i/256$	$8668i$
$J_xJ_yJ_zJ_xJ_zJ_zJ_yJ_xJ_y$	$i/256$	0	$41i/128$	$-4i$	$-14\,013i/256$	$-328i$	$-87\,351i/64$	$-4536i$	$-3\,297\,723i/256$	$-32\,516i$
$J_xJ_yJ_zJ_xJ_yJ_zJ_xJ_yJ_z$	$-i/256$	0	$-185i/128$	$-14i$	$-19\,395i/256$	$-302i$	$-63\,129i/64$	$-2790i$	$-1\,809\,093i/256$	$-16\,390i$
$J_xJ_yJ_zJ_xJ_yJ_zJ_zJ_yJ_z$	$-i/256$	$-i$	$-905i/128$	$-23i$	$-7235i/256$	$112i$	$55\,383i/64$	$3480i$	$2\,765\,499i/256$	$28\,655i$
J_x^{10}	1/512	2	29 525/256	2050	9 824 675/512	120 148	5.708 983×10⁵	2 217 300	7.381 024×10⁶	2.174 855×10⁷
$J_x^8J_y^2$	1/512	1	9845/256	517	2 038 435/512	21 742	11 988 165/128	338 430	1.066971×10⁶	3 015 095
$J_x^6J_y^4$	1/512	1	7685/256	337	1 191 715/512	11 842	6 217 365/128	169 410	5.201418×10⁵	1 440 275
$J_x^6J_y^2J_z^2$	1/512	0	−1075/256	−56	−163 485/512	−1064	−248 451/128	1080	12 580 581/512	118 760
$J_x^4J_y^4J_z^2$	1/512	0	−2155/256	−92	−264 285/512	−1964	−714 651/128	−12 060	−9 157 083/512	−5260

Operator	$J=11/2$	6	13/2	7	15/2	8	17/2	9	19/2	10
$J_x^2J_y^2J_z^2J_xJ_yJ_zJ_y$	$6\,997\,419i/128$	$106\,132i$	$1.968445\times10^{6}i$	$350\,964i$	$19\,344\,793i/32$	$1\,010\,004i$	$1.642217\times10^{6}i$	$2\,605\,812i$	$4.044684\times10^{6}i$	$6\,153\,642i$
$J_x^2J_y^2J_z^2J_yJ_xJ_zJ_y$	$-6\,018\,155i/128$	$-91\,975i$	$-1.720405\times10^{6}i$	$-309\,416i$	$-17\,200\,345i/32$	$-905\,352i$	$-1.483301\times10^{6}i$	$-2\,370\,345i$	$-3.703283\times10^{6}i$	$-5\,668\,135i$
$J_x^2J_y^2J_z^2J_yJ_zJ_xJ_y$	$3\,290\,859i/128$	$54\,652i$	$1.093285\times10^{6}i$	$207\,756i$	$12\,088\,921i/32$	$661\,164i$	$1.118957\times10^{6}i$	$1\,838\,364i$	$2.941478\times10^{6}i$	$4\,596\,174i$
$J_x^2J_y^2J_z^2J_yJ_zJ_yJ_z$	$-10\,814\,661i/128$	$-177\,008i$	$-3.501305\times10^{6}i$	$-659\,448i$	$-1.190548\times10^{6}i$	$-2\,071\,416i$	$-3.488638\times10^{6}i$	$-5.708208\times10^{6}i$	$-9.101860\times10^{6}i$	$-1.417997\times10^{7}i$
$J_x^2J_y^2J_z^2J_zJ_yJ_xJ_yJ_z$	$3\,908\,619i/128$	$63\,232i$	$1.239145\times10^{6}i$	$231\,624i$	$13\,298\,233i/32$	$719\,304i$	$1.206167\times10^{6}i$	$1\,966\,272i$	$3.125345\times10^{6}i$	$4\,855\,752i$
$J_x^2J_y^2J_z^2J_zJ_yJ_zJ_yJ_z$	$10\,196\,901i/128$	$168\,428i$	$3.355445\times10^{6}i$	$635\,580i$	$1.152757\times10^{6}i$	$2\,013\,276i$	$3.401428\times10^{6}i$	$5\,580\,300i$	$8.917992\times10^{6}i$	$1.392039\times10^{7}i$
$J_x^2J_y^2J_z^2J_zJ_yJ_zJ_yJ_z$	$-2\,055\,339i/128$	$-37\,492i$	$-20\,520\,071i/256$	$-160\,020i$	$-9\,670\,297i/32$	$-544\,884i$	$-9.445365\times10^{5}i$	$-1\,582\,548i$	$-2.573742\times10^{6}i$	$-4\,077\,018i$
$J_x^2J_y^2J_z^2J_yJ_zJ_zJ_yJ_z$	$202\,059i/128$	$11\,752i$	$9\,318\,023i/256$	$88\,416i$	$6\,042\,361i/32$	$370\,464i$	$6.829065\times10^{5}i$	$1\,198\,824i$	$2.022139\times10^{6}i$	$3\,298\,284i$
$J_xJ_yJ_zJ_xJ_zJ_yJ_zJ_yJ_z$	$2\,673\,099i/128$	$46\,072i$	$24\,254\,087i/256$	$183\,888i$	$10\,879\,609i/32$	$603\,024i$	$1.031747\times10^{6}i$	$1\,710\,456i$	$2.757610\times10^{6}i$	$4\,336\,596i$
$J_xJ_yJ_zJ_xJ_zJ_zJ_yJ_xJ_y$	$-9\,579\,141i/128$	$-159\,848i$	$-3.209585\times10^{6}i$	$-611\,712i$	$-1.114966\times10^{6}i$	$-1\,955\,136i$	$-3.314218\times10^{6}i$	$-5\,452\,392i$	$-8.734124\times10^{6}i$	$-1.366081\times10^{7}i$
$J_xJ_yJ_zJ_xJ_yJ_zJ_xJ_yJ_z$	$-4\,526\,379i/128$	$-71\,812i$	$-1.385005\times10^{6}i$	$-255\,492i$	$-14\,507\,545i/32$	$-777\,444i$	$-1.293377\times10^{6}i$	$-2\,094\,180i$	$-3.309213\times10^{6}i$	$-5\,115\,330i$
$J_xJ_yJ_zJ_xJ_yJ_zJ_zJ_yJ_z$	$8\,705\,125i/128$	$148\,265i$	$3.020045\times10^{6}i$	$581\,656i$	$1.068607\times10^{6}i$	$1\,885\,368i$	$3.211504\times10^{6}i$	$5.304135\times10^{6}i$	$8.523922\times10^{6}i$	$1.336759\times10^{7}i$
J_x^{10}	5.804005×10^{7}	1.426809×10^{8}	3.272949×10^{8}	7.076314×10^{8}	1.453565×10^{9}	2.855115×10^{9}	5.391053×10^{9}	9.828684×10^{9}	1.736579×10^{10}	2.982868×10^{10}
$J_x^8J_y^2$	7.786774×10^{6}	1.864854×10^{7}	4.188298×10^{7}	8.899847×10^{7}	1.802078×10^{8}	3.497345×10^{8}	6.536888×10^{8}	1.181478×10^{9}	2.071989×10^{9}	3.536105×10^{9}
$J_x^6J_y^4$	3.660878×10^{6}	8.656739×10^{6}	1.924319×10^{7}	4.054644×10^{7}	8.152605×10^{7}	1.572911×10^{8}	2.925304×10^{8}	5.264728×10^{8}	9.199112×10^{8}	1.564964×10^{9}
$J_x^6J_y^2J_z^2$	4.118889×10^{5}	1 195 376	3.076370×10^{6}	7.236432×10^{6}	1.585090×10^{7}	3.274363×10^{7}	6.437721×10^{7}	1.213097×10^{8}	2.202869×10^{8}	3.871854×10^{8}
$J_x^4J_y^4J_z^2$	20 618 559/256	376 376	1.184333×10^{6}	3 115 224	7.324302×10^{6}	1.588303×10^{7}	3.234571×10^{7}	6.258828×10^{7}	1.160193×10^{8}	2.072900×10^{8}

$$\begin{aligned}\mathrm{Tr}\, J_x^4J_y^2J_z^2&=\mathrm{Tr}\, J_x^2(\mathbf{J}^2-J_y^2-J_z^2)J_y^2J_z^2\\&=J(J+1)\,\mathrm{Tr}\, J_x^2J_y^2J_z^2-\mathrm{Tr}\, J_x^2J_y^4J_z^2\\&\quad-\mathrm{Tr}\, J_x^2J_z^2J_y^2J_z^2.\end{aligned}$$

By using the commutation relations one can bring the last term on the right-hand side to the form

$$\begin{aligned}&-\mathrm{Tr}\, J_x^2J_z^2J_y^2J_z^2\\&=-\mathrm{Tr}\, J_x^2J_y^2J_z^4+\mathrm{Tr}\{2iJ_x^2J_yJ_xJ_z^3-2J_x^2J_y^2J_z^2\\&\quad+2iJ_x^3J_yJ_z^3+2J_x^4J_z^2\}.\end{aligned}$$

Since $\mathrm{Tr}\, J_x^2J_y^2J_z^4=\mathrm{Tr}\, J_x^4J_y^2J_z^2$ one has

$$\begin{aligned}&3\,\mathrm{Tr}\, J_x^4J_y^2J_z^2\\&=J(J+1)\,\mathrm{Tr}\, J_x^2J_y^2J_z^2+2i\,\mathrm{Tr}\, J_x^2J_yJ_xJ_z^3\\&\quad+2i\,\mathrm{Tr}\, J_x^3J_yJ_z^3+2\,\mathrm{Tr}\, J_x^4J_z^2-2\,\mathrm{Tr}\, J_x^2J_y^2J_z^2.\end{aligned}$$

The terms on the right are of lower order and were previously evaluated.

$$\begin{aligned}\mathrm{Tr}\, J_x^5J_y^3J_z&=\mathrm{Tr}\, J_x^5J_y(\mathbf{J}^2-J_x^2-J_z^2)J_z\\&=\mathrm{Tr}\{\mathbf{J}^2J_x^5J_yJ_z-J_x^5J_yJ_x^2J_z-J_x^5J_yJ_z^3\}.\end{aligned}$$

One can show, by using property (8) and then rotating through $\pi/2$ about the x axis, that

$$\mathrm{Tr}\, J_x^5J_yJ_z^3=\mathrm{Tr}\, J_x^5J_y^3J_z.$$

Hence,

$$2\,\mathrm{Tr}\, J_x^5J_y^3J_z=J(J+1)\,\mathrm{Tr}\, J_x^5J_yJ_z-\mathrm{Tr}\, J_x^5J_yJ_x^2J_z.$$

The terms on the right can be easily evaluated.

$$\begin{aligned}\mathrm{Tr}\, J_x^6J_y^4=\mathrm{Tr}\{&(\mathbf{J}^2-J_y^2-J_z^2)(\mathbf{J}^2-J_y^2-J_z^2)\\&\times(\mathbf{J}^2-J_y^2-J_z^2)J_y^4\}\end{aligned}$$

The terms on the right are multiplied out. After some manipulation in which the commutation relations are used, one obtains

$$\begin{aligned}\mathrm{Tr}\, J_x^6J_y^4=\tfrac{1}{5}\,\mathrm{Tr}\{&\mathbf{J}^6J_x^4-3\mathbf{J}^4J_x^6-3\mathbf{J}^4J_x^4J_y^2+3\mathbf{J}^2J_x^8\\&+6\mathbf{J}^2J_x^6J_y^2+3\mathbf{J}^2J_x^4J_y^4-J_x^{10}-3J_x^8J_y^2\\&-iJ_x^5J_yJ_z^3+iJ_x^5J_yJ_zJ_y^2+iJ_x^4J_zJ_yJ_xJ_y^2\\&+iJ_x^4J_yJ_zJ_xJ_y^2\}.\end{aligned}$$

All terms on the right-hand side except J_x^{10} and $J_x^8J_y^2$ will have been evaluated previously, because they are of lower than tenth order; the two tenth order terms can be evaluated without difficulty.

The complete results are given in Tables II and III at the end of the paper. The tabulation is complete up to and including order nine. The computation of higher orders would have been too lengthy, and we decided to end the list by giving only those traces of order ten which seem to be the key ones from which the rest can be deduced. In Table II we give the algebraic forms for the traces, while in Table III we have listed the actual numerical values for angular momentum $J=\frac{1}{2}, 1, \frac{3}{2}\cdots 10$. The numerical values were obtained by using an electronic computer.

3. SPECIFIC FORMULAS

It may be useful to have the following formulas written out explicitly.

The partition function Z.

$$\begin{aligned}Z=\mathrm{Tr}\{\exp(-\mathfrak{IC}/kT)\}=\mathrm{Tr}\{&1-(\mathfrak{IC}/kT)\\&+(\mathfrak{IC}^2/2!k^2T^2)-\cdots\}.\end{aligned}\quad(9)$$

We shall assume henceforth that $\mathfrak{IC}$ is always constructed so that $\mathrm{Tr}\,\mathfrak{IC}\equiv 0$, i.e., $\langle\mathfrak{IC}\rangle\equiv 0$.

The Entropy S.

$$S=-\frac{\partial}{\partial T}(-kT\ln Z)=k\ln Z+\frac{kT}{Z}\frac{\partial Z}{\partial T}$$

Assuming we may use the formula

$$\ln(1+x)=x-\tfrac{1}{2}x^2+\tfrac{1}{3}x^3-\tfrac{1}{4}x^4+\cdots$$

to expand $\ln Z$ (i.e., $-1<x\leqslant 1$) we obtain

$$\begin{aligned}S=k[&\ln(\mathrm{Tr}\,1)-(1/2!k^2T^2)\{\langle\mathfrak{IC}^2\rangle\}\\&+(1/3!k^3T^3)\{2\langle\mathfrak{IC}^3\rangle\}-(1/4!k^4T^4)\\&\times\{3\langle\mathfrak{IC}^4\rangle-9\langle\mathfrak{IC}^2\rangle^2\}+\cdots].\end{aligned}\quad(10)$$

The specific heat, C, can then be obtained under any given conditions, since

$$C=T(\partial S/\partial T).$$

The magnetic susceptibility is given by

$$\chi=\lim_{H\to 0}\left(\frac{M}{H}\right),$$

where M is the magnetic moment and H the applied magnetic field. The explicit form can be obtained by splitting the Hamiltonian into field independent and field dependent parts. For H small we have[4a]:

$$\mathfrak{IC}=\mathfrak{IC}_0+H\mathfrak{IC}_M.$$

We then obtain, assuming $M=0$ when $H=0$, i.e., no spontaneous magnetization,

$$\begin{aligned}\chi=&(1/kT)\{\langle\mathfrak{IC}_m^2\rangle\}+(1/k^2T^2)\{-\tfrac{1}{2}\langle\mathfrak{IC}_M^2\mathfrak{IC}_0\rangle\\&-\tfrac{1}{2}\langle\mathfrak{IC}_M\mathfrak{IC}_0\mathfrak{IC}_M\rangle\}+(1/k^3T^3)\{-\tfrac{1}{2}\langle\mathfrak{IC}_M^2\rangle\langle\mathfrak{IC}_0^2\rangle\\&+\tfrac{1}{6}\langle\mathfrak{IC}_M^2\mathfrak{IC}_0^2\rangle+\tfrac{1}{6}\langle\mathfrak{IC}_M\mathfrak{IC}_0\mathfrak{IC}_M\mathfrak{IC}_0\rangle+\tfrac{1}{6}\langle\mathfrak{IC}_M\mathfrak{IC}_0^2\mathfrak{IC}_M\rangle\}\\&+(1/k^4T^4)\{\tfrac{1}{6}\langle\mathfrak{IC}_M^2\rangle\langle\mathfrak{IC}_0^3\rangle+\tfrac{1}{4}\langle\mathfrak{IC}_0^2\rangle\langle\mathfrak{IC}_M^2\mathfrak{IC}_0\rangle\\&+\tfrac{1}{4}\langle\mathfrak{IC}_0^2\rangle\langle\mathfrak{IC}_M\mathfrak{IC}_0\mathfrak{IC}_M\rangle-(1/4!)\langle\mathfrak{IC}_M^2\mathfrak{IC}_0^3\rangle\\&-(1/4!)\langle\mathfrak{IC}_M\mathfrak{IC}_0\mathfrak{IC}_M\mathfrak{IC}_0^2\rangle-(1/4!)\langle\mathfrak{IC}_M\mathfrak{IC}_0^2\mathfrak{IC}_M\mathfrak{IC}_0\rangle\\&-(1/4!)\langle\mathfrak{IC}_M\mathfrak{IC}_0^3\mathfrak{IC}_M\rangle\}+\cdots.\end{aligned}\quad(11)$$

The statistical tensors[5] are also of interest, particularly with reference to nuclear orientation experiments. They are defined by the relation:

$$\begin{aligned}[\langle J'|\rho_\alpha|J\rangle]_q^k=\sum_{MM'}(-)^{J-M}&\langle J'M'J-M|kq\rangle\\&\times\langle J'M'|\rho_\alpha|JM\rangle\end{aligned}\quad(12)$$

[4a] *Note added in proof.* The usual term in H^2 which gives rise to diamagnetic effects is omitted from this Hamiltonian.

When $\mathfrak{IC}$ is taken to be the spin Hamiltonian one should remember that a term in $\mathfrak{IC}$ of the form $\mathbf{H}\cdot\mathbf{\Lambda}\cdot\mathbf{H}$ is seldom given explicitly since it is not usually relevant. This term gives rise to the Van Vleck temperature-independent paramagnetism.

[5] U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic Press, Inc., New York, 1959).

TABLE IV. The explicit forms of the statistical operators T_q^k, and their relation to the harmonic polynomials, y_{kq}. The normalization constants are given, and are defined such that $\langle J'\|T^k\|J\rangle=(2k+1)^{\frac{1}{2}}$.

Harmonic polynomial, $y_{kq}=r^kY_{kq}$	Statistical operator, T_q^k	Normalization, a_k
$y_{00}=-1/(2\pi)^{\frac{1}{2}}$	$a_0 1$	$a_0=\langle JJJ-J\|00\rangle=(2J+1)^{-\frac{1}{2}}$
$y_{10}=\frac{1}{2}(3/\pi)^{\frac{1}{2}}z$	a_1J_z	$a_1=2\langle JJJ-J\|10\rangle(2J)^{-1}$
$y_{1\pm1}=\mp\frac{1}{2}(3/2\pi)^{\frac{1}{2}}(x\pm iy)$	$\mp a_1(1/\sqrt{2})J_\pm$	$=2\sqrt{3}[(2J+2)(2J+1)(2J)]^{-\frac{1}{2}}$
$y_{20}=\frac{1}{4}(5/\pi)^{\frac{1}{2}}(3z^2-r^2)$	$a_2\{3J_z^2-J(J+1)\}$	$a_2=2\langle JJJ-J\|20\rangle[(2J)(2J-1)]^{-1}$
$y_{2\pm1}=\mp\frac{1}{2}(15/2\pi)^{\frac{1}{2}}(z)(x\pm iy)$	$\mp(3/2)^{\frac{1}{2}}a_2\{J_zJ_\pm+J_\pm J_z\}$	$=2(5)^{\frac{1}{2}}[(2J+3)(2J+2)(2J+1)(2J)(2J-1)]^{-\frac{1}{2}}$
$y_{2\pm2}=\frac{1}{4}(15/2\pi)^{\frac{1}{2}}(x\pm iy)^2$	$(3/2)^{\frac{1}{2}}a_2\{J_\pm^2\}$	
$y_{30}=\frac{1}{4}(7/\pi)^{\frac{1}{2}}(5z^3-3zr^2)$	$a_3\{5J_z^3-3J_zJ(J+1)+J_z\}$	$a_3=4\langle JJJ-J\|30\rangle[(2J)(2J-1)(2J-2)]^{-1}$
$y_{3\pm1}=\mp\frac{1}{8}(21/\pi)^{\frac{1}{2}}(5z^2-r^2)(x\pm iy)$	$\mp(1/12)^{\frac{1}{2}}a_3\{4(J_z^2J_\pm+J_zJ_\pm J_z+J_\pm J_z^2)$ $-(J_\pm^2J_\mp+J_\pm J_\mp J_\pm+J_\mp J_\pm^2)\}$	$=4(7)^{\frac{1}{2}}[(2J+4)(2J+3)(2J+2)(2J+1)(2J)$ $\times(2J-1)(2J-2)]^{-\frac{1}{2}}$
$y_{3\pm2}=\frac{1}{4}(105/2\pi)^{\frac{1}{2}}(z)(x\pm iy)^2$	$(5/6)^{\frac{1}{2}}a_3\{J_zJ_\pm^2+J_\pm J_zJ_\pm+J_\pm^2J_z\}$	
$y_{3\pm3}=\mp\frac{1}{8}(35/\pi)^{\frac{1}{2}}(x\pm iy)^3$	$\mp(5/4)^{\frac{1}{2}}a_3\{J_\pm^3\}$	
$y_{40}=\frac{1}{16}(9/\pi)^{\frac{1}{2}}(35z^4-30z^2r^2+3r^4)$	$a_4\{35J_z^4-30J_z^2J(J+1)+25J_z^2-6J(J+1)$ $+3J^2(J+1)^2\}$	
$y_{4\pm1}=\mp\frac{1}{8}(45/\pi)^{\frac{1}{2}}(7z^2-3r^2)(z)(x\pm iy)$	$\mp(20)^{\frac{1}{2}}a_4\{(J_z^3J_\pm+J_z^2J_\pm J_z+J_zJ_\pm J_z^2$ $+J_\pm J_z^3)-\frac{1}{4}(J_zJ_\pm^2J_\mp+J_zJ_\pm J_\mp J_\pm$ $+J_zJ_\mp J_\pm^2+J_\pm J_zJ_\pm J_\mp+J_\pm J_zJ_\mp J_\pm$ $+J_\mp J_zJ_\pm^2+J_\pm^2J_zJ_\mp+J_\pm J_\mp J_zJ_\pm$ $+J_\mp J_\pm J_zJ_\pm+J_\pm^2J_\mp J_z+J_\pm J_\mp J_\pm J_z$ $+J_\mp J_\pm^2J_z)\}$	$a_4=2\langle JJJ-J\|40\rangle[(2J)(2J-1)(2J-2)(2J-3)]^{-1}$ $=2(9)^{\frac{1}{2}}[(2J+5)(2J+4)(2J+3)(2J+2)(2J+1)$ $\times(2J)(2J-1)(2J-2)(2J-3)]^{-\frac{1}{2}}$
$y_{4\pm2}=\frac{1}{8}(45/\pi)^{\frac{1}{2}}(7z^2-r^2)(x\pm iy)^2$	$(10)^{\frac{1}{2}}a_4\{(J_z^2J_\pm^2+J_\pm^2J_z^2+J_zJ_\pm^2J_z+J_\pm J_z^2J_\pm$ $+J_zJ_\pm J_zJ_\pm+J_\pm J_zJ_\pm J_z)-\frac{1}{4}(J_\pm^3J_\mp$ $+J_\pm^2J_\mp J_\pm+J_\pm J_\mp J_\pm^2+J_\mp J_\pm^3)\}$	
$y_{4\pm3}=\mp\frac{1}{8}(315/\pi)^{\frac{1}{2}}(z)(x\pm iy)^3$	$(35/4)^{\frac{1}{2}}a_4\{J_zJ_\pm^3+J_\pm J_zJ_\pm^2+J_\pm^2J_zJ_\pm$ $+J_\pm^3J_z\}$	
$y_{4\pm4}=\frac{1}{16}(315/\pi)^{\frac{1}{2}}(x\pm iy)^4$	$(35/2)^{\frac{1}{2}}a_4\{J_\pm^4\}$	

The density matrix of the state, $|a\rangle$, of the system under consideration is ρ_a, and $\langle J'M'J-M|kq\rangle$ is a Clebsch-Gordan coefficient. The statistical tensors or, rather, their complex conjugates, can be related to the expectation values, $\langle a|T_q^k|a\rangle$, of a set of tensor operators T_q^k, as follows. We have

$$\langle a|T_q^k|a\rangle=\sum_{JJ'MM'}\langle a|J'M'\rangle\langle J'M'|T_q^k|JM\rangle\langle JM|a\rangle \tag{13}$$

and, from the definition of the reduced matrix element,[5,6]

$$\langle J'M'|T_q^k|JM\rangle=(-)^{J-M}(2k+1)^{-\frac{1}{2}}\times\langle J'M'J-M|kq\rangle\langle J'\|T^k\|J\rangle, \tag{14}$$

one obtains

$$\langle a|T_q^k|a\rangle=\sum_{JJ'MM'}\frac{\langle J'\|T^k\|J\rangle}{(2k+1)^{\frac{1}{2}}}(-)^{J-M}\times\langle J'M'J-M|kq\rangle\langle a|J'M'\rangle\langle JM|a\rangle. \tag{15}$$

Since ρ_a is Hermitian

$$\langle J'M'|\rho_a|JM\rangle^*\equiv\langle J'M'|a\rangle^*\langle a|JM\rangle^*=\langle a|J'M'\rangle\langle JM|a\rangle.$$

Hence (15) becomes

$$\langle a|T_q^k|a\rangle=\sum_{JJ'}\frac{\langle J'\|T^k\|J\rangle}{(2k+1)^{\frac{1}{2}}}\sum_{MM'}(-)^{J-M}\times\langle J'M'J-M|kq\rangle\langle J'M'|\rho_a|JM\rangle^*.$$

[6] M. E. Rose, Phys. Rev. **108**, 362 (1957).

Since the Clebsch-Gordan coefficients are real,

$$\langle a|T_q^k|a\rangle=\sum_{JJ'}\frac{\langle J'\|T^k\|J\rangle}{(2k+1)^{\frac{1}{2}}}[\langle J'|\rho_a|J\rangle]_q^{k*}. \tag{16}$$

If we now normalize the operators T_q^k so that

$$\langle J'\|T^k\|J\rangle=(2k+1)^{\frac{1}{2}}, \tag{17}$$

we have, for a state of sharp angular momentum J,

$$\langle a|T_q^k|a\rangle=[\langle J|\rho_a|J\rangle]_q^{k*}. \tag{18}$$

We give the explicit forms of the statistical operators in Table IV up to and including the set for which $k=4$. They were constructed from the harmonic polynomials by using the procedures given by Edmonds[7] and Stevens[8] and were normalized according to Eq. (17). Koster and Statz[9] have given a similar table in which the explicit forms of the operators have a rather different appearance.

ACKNOWLEDGMENT

We wish to thank J. D. Waggoner who coded the formulas for the traces for machine calculation.

[7] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1957), p. 71.
[8] K. W. H. Stevens, Proc. Phys. Soc. (London) **A65**, 209 (1952).
[9] G. F. Koster and H. Statz, Phys. Rev. **113**, 445 (1959).

Variational Method for the Quantum Statistics of Interacting Particles

M. GIRARDEAU*
Boeing Scientific Research Laboratories, Seattle, Washington
(Received July 27, 1961)

A variational method is developed for calculating the thermodynamic potential of quantum-mechanical many-body systems with pair-wise interactions. The method is based on Peierls' theorem and yields an upper bound to the thermodynamic potential density in the limit of an infinite system. Evaluation and minimization of the bound involves solution of a set of coupled nonlinear integral equations for the distribution function of elementary excitations and for functions defining a unitary transformation from bare particles to elementary excitations. Application of the theory to the BCS model of superconductivity reproduces the BCS results, and application to a degenerate imperfect Bose gas gives equations which are shown to be equivalent to those of Tolmachev and Wentzel.

I. INTRODUCTION

IN 1938 Peierls published a variational theorem[1] for the quantum-statistical partition function which is a natural generalization to nonzero temperature of the familiar variational theorem for the ground state. Peierls' theorem implies that one obtains a rigorous lower bound to the exact partition function and hence an upper bound to the thermodynamic potential (free energy), if in evaluating the partition function one replaces the Hamiltonian by its diagonal part in any representation; the theorem holds equally well for the grand partition function if one replaces the Hamiltonian H by $H-\mu N$.

By choosing a representation in terms of states of the form $U\phi_\alpha^{(0)}$ where the $\phi_\alpha^{(0)}$ are independent-particle states and the unitary "model operator" U induces a linear canonical transformation of the single-particle annihilation and creation operators depending upon variational parameters (actually arbitrary functions), one obtains a workable variational method for calculating the thermodynamic properties of quantum-mechanical many-body systems; the variational parameters are simply related to the properties of the approximate "elementary excitations" described by the states $U\phi_\alpha^{(0)}$. Provided that the elementary excitation energies are all positive, the quantum-statistical variational treatment reduces at zero temperature to a quantum-mechanical variational treatment using a trial ground state of the form $U|0\rangle$, where $|0\rangle$ is the vacuum state and hence $U|0\rangle$ is the state containing no elementary excitations. The method can therefore be regarded as a certain generalization of Gross' theory[2] to nonzero temperature in the case of a Bose system, and as the corresponding generalization of Bogolubov's theory[3] in the case of a Fermi system.

The general theory is developed in Sec. II; it is noted that when applied to the BCS model[4] of superconductivity it reduces to a variational principle obtained previously by Koppe and Mühlschlegel,[5] which reproduces the BCS results. The general theory is applied to a different example, the degenerate imperfect Bose gas, in Sec. III; the resulting equations are shown to be equivalent to those of Tolmachev[6] and Wentzel[7] for the cases of a grand ensemble and a canonical ensemble respectively.

II. GENERAL THEORY

A. Formulation

The Hamiltonian of a nonrelativistic system of identical particles interacting by two-body interactions can be written in second-quantized form as

$$H=\sum_{ij}\epsilon_{ij}a_i^\dagger a_j+\tfrac{1}{2}\sum_{ijkl}v_{ijkl}a_i^\dagger a_j^\dagger a_l a_k, \qquad (1)$$

where the ϵ_{ij} are single-particle matrix elements of the kinetic energy and any external field (if present), the v_{ijkl} are interparticle interaction matrix elements, and the a_i and $a_i^\dagger$ are single-particle annihilation and creation operators satisfying the usual relations

$$[a_i,a_j^\dagger]_\pm=\delta_{ij}, \quad [a_i,a_j]_\pm=[a_i^\dagger,a_j^\dagger]_\pm=0, \qquad (2)$$

where anticommutators $[\ ]_+$ are to be taken for the case of Fermi statistics and commutators $[\ ]_-$ for Bose statistics. The indices $i, j, \cdots$ are most commonly taken to refer to momentum in the case of Bose particles, and momentum and spin in the case of Fermi particles, but we do not restrict ourselves to this particular choice of single-particle states. The system possesses a total particle number operator N which commutes with the Hamiltonian and has the general form

$$N=\sum_{ij}n_{ij}a_i^\dagger a_j; \qquad (3)$$

the matrix n_{ij} is only diagonal in the free-particle representation. In addition, if the system is translationally invariant and periodic boundary conditions are adopted, there will be a total linear momentum operator $\mathbf{P}$ which

* Now at Enrico Fermi Institute for Nuclear Studies, The University of Chicago, Chicago, Illinois.

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

[2] E. P. Gross, Ann. Phys. **9**, 292 (1960).

[3] N. N. Bogolubov, Doklady Akad. Nauk. S.S.S.R. **119**, 244 (1958), translated in Soviet Phys.—Doklady **3**, 292 (1958).

[4] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).

[5] H. Koppe and B. Mühlschlegel, Z. Phys. **151**, 613 (1958), Appendix 2 and Sec. IV.

[6] V. V. Tolmachev, Doklady Akad. Nauk. S.S.S.R. **134**, 1324 (1960), translated in Soviet Phys.—Doklady **5**, 984 (1961).

[7] G. Wentzel, Phys. Rev. **120**, 1572 (1960).

commutes with the Hamiltonian and has the general form

$$\mathbf{P}=\sum_{ij}\mathbf{p}_{ij}a_i^\dagger a_j. \tag{4}$$

In such a case the controllable[8] constants of the motion are H, N, and $\mathbf{P}$, so that the thermal equilibrium properties of the system are described by a density operator

$$\rho=Z^{-1}e^{-\beta(H-\mu N-\mathbf{u}\cdot\mathbf{P})} \tag{5}$$

with

$$Z=\mathrm{Tr}e^{-\beta(H-\mu N-\mathbf{u}\cdot\mathbf{P})}. \tag{6}$$

The parameter β is, as usual, $(kT)^{-1}$ with k Boltzmann's constant and T the absolute temperature, μ is the chemical potential, and $\mathbf{u}$ is a parameter related to the average total linear momentum. For systems which are not capable of superfluidity, or for systems which are not translationally invariant, one can set $\mathbf{u}=0$ so that Z reduces to the usual grand partition function; the treatment of superfluids is, however, facilitated[9] by use of the more general ensemble (5). We can define an appropriate thermodynamic potential (free energy) W by

$$Z=e^{-\beta W}. \tag{7}$$

Then the mean particle number n, mean total linear momentum $\mathbf{p}$, and internal energy E are

$$\begin{aligned} n&=\mathrm{Tr}(N\rho)=-(\partial W/\partial\mu)_{\beta,\mathbf{u}},\\ \mathbf{p}&=\mathrm{Tr}(\mathbf{P}\rho)=-(\partial W/\partial\mathbf{u})_{\beta,\mu},\\ E&=\mathrm{Tr}(H\rho)=[\partial(\beta W)/\partial\beta]_{\beta\mu,\beta\mathbf{u}}, \end{aligned} \tag{8}$$

where in the last equation the quantities held constant in the differentiation are $\beta\mu$ and $\beta\mathbf{u}$ rather than μ and $\mathbf{u}$.

In order to obtain a variational principle for the thermodynamic potential W we employ a theorem due to Peierls[1] which implies in our case that

$$\begin{aligned} Z&=\sum_\alpha\langle\alpha|\exp[-\beta(H-\mu N-\mathbf{u}\cdot\mathbf{P})]|\alpha\rangle\\ &\geqslant\sum_\alpha\exp[-\beta\langle\alpha|H-\mu N-\mathbf{u}\cdot\mathbf{P}|\alpha\rangle]\equiv Z_{\mathrm{var}} \end{aligned} \tag{9}$$

and hence that

$$W\leqslant W_{\mathrm{var}} \tag{10}$$

where the variational approximation W_{var} to the thermodynamic potential is defined by

$$Z_{\mathrm{var}}=e^{-\beta W_{\mathrm{var}}}. \tag{11}$$

The summations in (9) are carried out over any complete orthonormal set of states $|\alpha\rangle$; the equality holds only if the $|\alpha\rangle$ are the eigenstates of H. One employs (9) in a variational calculation by letting the states $|\alpha\rangle$, chosen by a compromise between physical reality and mathematical tractability, depend upon undetermined parameters or functions and choosing these so as to minimize W_{var}, thereby bringing it as close as possible to the exact thermodynamic potential W. It is important to realize that (10) does *not* necessarily imply that $(W/n)\leqslant(W_{\mathrm{var}}/n_{\mathrm{var}})$ with n given by the first Eq. (8) and n_{var} by its analog with W replaced by W_{var}. What *is* implied is that

$$(W/\Omega)\leqslant(W_{\mathrm{var}}/\Omega)\quad\text{for }\textit{given}\ \mu\text{ and }\mathbf{u}, \tag{12}$$

where Ω is the volume of the system. For the case $\mathbf{u}=0$, Z reduces to the usual grand partition function. In this case W is related to the pressure p by[10]

$$W=-p\Omega\quad(\mathbf{u}=0), \tag{13}$$

and the minimum principle (12) for (W/Ω) reduces to a maximum principle for the pressure.

In order to obtain a complete set $\{|\alpha\rangle\}$ sufficiently simple that the summation on the right side of (9) can actually be performed, we shall choose the states $|\alpha\rangle$ to be of the general form

$$|\{\eta_i\}\rangle=U[\textstyle\prod_i(\eta_i!)^{-\frac{1}{2}}(a_i^\dagger)^{\eta_i}]|0\rangle, \tag{14}$$

where $|0\rangle$ is the vacuum state and U is a unitary operator which induces a *linear* canonical transformation of the annihilation and creation operators[11]:

$$U^{-1}a_iU=s_i+\sum_j(f_{ij}a_j+g_{ij}a_j^\dagger). \tag{15}$$

The coefficients s_i, f_{ij}, and g_{ij} are variational parameters, while the eigenvalues η_i in (14) take on all integral values 0, 1, 2, $\cdots$ for the case of a Bose system and the values 0, 1 for the case of a Fermi system. The "model operator" U transforms the independent-particle states $[\prod_i(\eta_i!)^{-\frac{1}{2}}(a_i^\dagger)^{\eta_i}]|0\rangle$ into the states $|\{\eta_i\}\rangle$; the quantum numbers[12] η_i specify the numbers of "elementary excitations" with properties labeled by the subscripts i. In most cases simplifying assumptions will be made which reduce the number of variational parameters appearing in (15).

In terms of the states (14), the variational approximation Z_{var} [Eq. (9)] to the partition function is

$$Z_{\mathrm{var}}=\sum_{\{\eta_i\}}\exp[-\beta\langle\{\eta_i\}|(H-\mu N-\mathbf{u}\cdot\mathbf{P})|\{\eta_i\}\rangle]. \tag{16}$$

It is clear from (14) that only the part of

$$U^{-1}(H-\mu N-\mathbf{u}\cdot\mathbf{P})U$$

which is diagonal in the independent-particle basis contributes to (16). Since the canonical transformation (15) is linear, this diagonal part can be written in the form

$$\begin{aligned} &[U^{-1}(H-\mu N-\mathbf{u}\cdot\mathbf{P})U]_{\mathrm{diag}}\\ &\qquad=W_0+\sum_i W_iN_i+\tfrac{1}{2}\sum_{ij}W_{ij}N_iN_j, \end{aligned} \tag{17}$$

where $N_i\equiv a_i^\dagger a_i$, the occupation number operator for the single-particle state with label i, and where the

[8] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover Publications, New York, 1949), p. 51.

[9] N. N. Bogolubov, J. Phys. (U.S.S.R.) 11, 23 (1947).

[10] See, e.g., A. Münster, *Statistische Thermodynamik* (Springer-Verlag, Berlin, Germany, 1956), p. 187.

[11] It is not necessary here to write out the subsidiary conditions which must be imposed upon the coefficients f_{ij} and g_{ij} in order that (15) actually be canonical.

[12] α in (9) summarizes the whole set $\{\eta_i\ (\text{all } i)\}$.

coefficients W_0, W_i, W_{ij} depend upon the variational parameters in (15). The function W_{ij} represents the effects of the diagonal interactions between the elementary excitations. In terms of (17), Eq. (16) can be rewritten in the form

$$Z_{\rm var}=e^{-\beta W_0}\sum_{\{\eta_i\}}\exp[-\beta(\sum_i W_i\eta_i+\tfrac{1}{2}\sum_{ij}W_{ij}\eta_i\eta_j)]. \tag{18}$$

B. Thermodynamic Potential: Variational Equations

In order to proceed further we have to evaluate (18). An exact evaluation is exceedingly difficult; however, the thermodynamic functions are determined by the asymptotic behavior as the system volume $\Omega\to\infty$, and such an asymptotic evaluation is easy provided that the following conditions are satisfied:

$$W_i=O(1),\quad W_{ij}=O(\Omega^{-1}),\quad \sum_i=O(\Omega). \tag{19}$$

The meaning of the first two requirements is clear; they will in general be satisfied if the single-particle states labeled by i are spacially nonlocalized and if the coefficients defining the unitary transformation (15) are properly chosen. The third requirement (19) means that each summation over i contributes a factor Ω. This will be the case, e.g., if $\sum_i$ means $\sum_{\mathbf{k}}$ for bosons and $\sum_{\mathbf{k}\sigma}$ for fermions, where $\mathbf{k}$ is an allowed single-particle momentum; then $\sum_{\mathbf{k}}\to(2\pi)^{-3}\Omega\int d^3k$ as $\Omega\to\infty$; more generally, the same qualitative behavior occurs for any spacially nonlocalized single-particle states. We shall assume from now on that the conditions (19) are indeed satisfied; it is important to verify that this is the case in each particular application of the general theory.

There are many ways of evaluating (18) asymptotically for large Ω. The simplest derivation is based on the inequality[13–15]

$$W_{\rm var}=-\beta^{-1}\ln{\rm Tr}e^{-\beta(\mathcal{H}_0+\mathcal{H}_1)}\leqslant-\beta^{-1}\ln{\rm Tr}e^{-\beta\mathcal{H}_0}+\frac{{\rm Tr}(\mathcal{H}_1e^{-\beta\mathcal{H}_0})}{{\rm Tr}e^{-\beta\mathcal{H}_0}}, \tag{20}$$

where $W_{\rm var}=-\beta^{-1}\ln Z_{\rm var}$ [cf. (11)], and $\mathcal{H}_0$ and $\mathcal{H}_1$ are defined by an appropriate decomposition of (17):

$$\begin{aligned}&[U^{-1}(H-\mu N-\mathbf{u}\cdot\mathbf{P})U]_{\rm diag}=\mathcal{H}_0+\mathcal{H}_1,\\ &\mathcal{H}_0\equiv W_0-\tfrac{1}{2}\sum_{ij}W_{ij}f_if_j+\sum_i\epsilon_iN_i,\\ &\mathcal{H}_1\equiv\tfrac{1}{2}\sum_{ij}W_{ij}(N_i-f_i)(N_j-f_j),\\ &\epsilon_i\equiv W_i+\sum_jW_{ij}f_j.\end{aligned} \tag{21}$$

The physical meaning of this decomposition is that the interaction involves only products of fluctuations (N_i-f_i) instead of the products of the occupation numbers themselves, provided that f_i is the thermal average of N_i; the relationship to the work of Bogolubov, Zubarev, and Tserkovnikov[16] and Wentzel[7] is clear. The evaluation of $W^{(0)}\equiv-\beta^{-1}\ln{\rm Tr}e^{-\beta\mathcal{H}_0}$ is trivial, while the second term $\langle\mathcal{H}_1\rangle_0$ in (20) can be evaluated with the aid of the relationships

$$\begin{aligned}&\partial W^{(0)}/\partial\epsilon_i=\langle N_i\rangle_0,\\ &\partial^2W^{(0)}/\partial\epsilon_i\partial\epsilon_j=-\beta\langle N_iN_j\rangle_0+\beta\langle N_i\rangle_0\langle N_j\rangle_0,\end{aligned} \tag{22}$$

where $\langle\cdots\rangle_0\equiv{\rm Tr}(\cdots e^{-\beta\mathcal{H}_0})/{\rm Tr}e^{-\beta\mathcal{H}_0}$. The resulting expression for the right side of (20) is[17]

$$W_0\mp\beta^{-1}\sum_i\ln(1\pm e^{-\beta\epsilon_i})-\sum_{ij}W_{ij}(e^{\beta\epsilon_i}\pm1)^{-1}f_j+\tfrac{1}{2}\sum_{ij}W_{ij}(e^{\beta\epsilon_i}\pm1)^{-1}(e^{\beta\epsilon_j}\pm1)^{-1}, \tag{23}$$

where the upper signs are to be taken for fermions and the lower for bosons. Minimization of (23) with respect to f_i, taking account of the definition (21) of ϵ_i, leads to the following nonlinear integral equation for f_i:

$$f_i=\exp[\beta(W_i+\sum_jW_{ij}f_j)\pm1]^{-1}. \tag{24}$$

Substitution of (24) and the expression (21) for ϵ_i into (23) leads to the following expression for the upper bound to the thermodynamic potential:

$$W_{\rm var}=W_0\pm\beta^{-1}\sum_i\ln(1\mp f_i)-\tfrac{1}{2}\sum_{ij}W_{ij}f_if_j. \tag{25}$$

In writing down (25) we have anticipated the fact that the inequality (20) is in fact an equality to $O(\Omega)$ provided that (19) and (24) are satisfied; this requires

---

[13] This inequality and its use in a variational principle seem to be due originally to Bogolubov. The author is not aware of any published account available in English; an indirect reference is given in footnote 4 of a paper by Tolmachev.[6] We sketch here a proof of (20) due to Mühlschlegel,[14] to whom the author is indebted for pointing out the error in a previous derivation[15] and for informing him of this method. One has

$${\rm Tr}\exp[-\beta(\mathcal{H}_0+\mathcal{H}_1)]\geqslant\Sigma_\alpha\exp\{-\beta[E_\alpha+\langle\alpha|\mathcal{H}_1|\alpha\rangle]\}\geqslant\exp(-\beta\langle\mathcal{H}_1\rangle_0)\,{\rm Tr}\exp(-\beta\mathcal{H}_0)$$

where $\{|\alpha\rangle\}$ is a complete orthonormal set of eigenstates of $\mathcal{H}_0$ with eigenvalues E_α, and

$$\langle\mathcal{H}_1\rangle_0={\rm Tr}[\mathcal{H}_1\exp(-\beta\mathcal{H}_0)]/{\rm Tr}\exp(-\beta\mathcal{H}_0).$$

The first inequality follows immediately from Peierls' theorem, while the second inequality involves a second application of the convexity property

$$\Sigma_\alpha w_\alpha f(x_\alpha)\geqslant f(\Sigma_\alpha w_\alpha x_\alpha);\quad w_\alpha\geqslant0,\quad \Sigma_\alpha w_\alpha=1$$

valid for any function f which is convex downward. The convexity condition was already applied once in obtaining Peierls' theorem; it is applied again with $w_\alpha=e^{-\beta E_\alpha}/{\rm Tr}\exp(-\beta\mathcal{H}_0)$, $x_\alpha=\langle\alpha|\mathcal{H}_1|\alpha\rangle$, $f(x)=e^{-\beta x}$ in order to obtain (20).

[14] B. Mühlschlegel, Sitzber. bayer. Akad. Wiss. München, Math.-naturw. Kl. **10** (1960), Sec. 3.

[15] M. Girardeau, Bull. Am. Phys. Soc. **6**, 16 (1961).

[16] Bogolubov, Zubarev, and Tserkovnikov, Doklady Akad. Nauk. S.S.S.R. **117**, 788 (1957), translated in Soviet Phys.—Doklady **2**, 535 (1957).

[17] An extra term

$$\tfrac{1}{2}\Sigma_iW_{ii}e^{\beta\epsilon_i}(e^{\beta\epsilon_i}\pm1)^{-2}$$

has been dropped, since it is only $O(1)$ [instead of $O(\Omega)$] because of (19). A possible exception can occur for Bose statistics, in case the single term corresponding to $\mathbf{k}=0$ should contribute $O(\Omega)$. This difficulty can be avoided by proper choice of s_0 in (15), as we shall see in Sec. III.

separate proof.[18–20] For the special case of Fermi statistics and separable interaction $W_{ij}=\lambda v_i v_j$, Eqs. (24) and (25) reduce to results obtained previously by Koppe and Mühlschlegel,[5] which were shown to reproduce the BCS[4] results for the thermodynamics of the BCS model Hamiltonian.

It is readily verified that

$$\langle N_i\rangle_0=\partial W^{(0)}/\partial\epsilon_i=f_i \tag{26}$$

and that

$$\langle\mathcal{H}_1\rangle_0=O(1). \tag{27}$$

This justifies the interpretation of f_i as the distribution function of elementary excitations; the elementary excitation energy ϵ_i [Eq. (21)] includes both the bare excitation energy W_i and the self-energy $\sum_j W_{ij}f_j$ due to interaction of the bare excitations.

The variational method consists of minimizing the upper bound (25) with respect to the parameters defining the canonical transformation (15). Renaming the *independent* parameters ϕ_k, one can write the variational equations in the form

$$\frac{\partial W_{\rm var}}{\partial\phi_k}=\frac{\partial W_0}{\partial\phi_k}-\beta^{-1}\sum_i\frac{\partial f_i/\partial\phi_k}{1\mp f_i}$$
$$-\sum_{ij}f_i\left[\frac{1}{2}\left(\frac{\partial W_{ij}}{\partial\phi_k}\right)f_j+W_{ij}\left(\frac{\partial f_j}{\partial\phi_k}\right)\right]=0. \tag{28}$$

It follows from (24) that

$$\frac{\partial f_i}{\partial\phi_k}=-\beta f_i(1\mp f_i)\left\{\frac{\partial W_i}{\partial\phi_k}+\sum_j\left[\left(\frac{\partial W_{ij}}{\partial\phi_k}\right)f_j+W_{ij}\left(\frac{\partial f_j}{\partial\phi_k}\right)\right]\right\}. \tag{29}$$

Using (29) to eliminate $\partial f_i/\partial\phi_k$ from (28), one finds

$$\frac{\partial W_0}{\partial\phi_k}+\sum_i f_i\left(\frac{\partial W_i}{\partial\phi_k}\right)+\tfrac{1}{2}\sum_{ij}f_if_j\left(\frac{\partial W_{ij}}{\partial\phi_k}\right)=0. \tag{30}$$

When the explicit forms of W_0, W_i, and W_{ij} are substituted into (30) and the derivatives evaluated, Eqs. (24) and (30) become a set of coupled nonlinear integral equations for f_i and ϕ_k.

C. Generalization to Multicomponent Systems

One is often interested in the quantum statistical mechanics of mixtures of two or more species of particles each of which would alone be described by a Hamiltonian of the general type (1); two examples are He^3–He^4 mixtures and multicomponent plasmas. The Hamiltonian of such a mixture is of the form

$$H=\sum_S H_S+\sum_{(S,S')}H_{SS'} \tag{31}$$

where S and S' are species (component) labels; $\sum_S$ is a sum over all different species and $\sum_{(S,S')}$ is a sum over all pairs (S,S') of different species. Each H_S is of the general form (1), and $H_{SS'}$ is the interaction between systems S and S', of the general form

$$H_{SS'}=\sum_{ijkl}c_{ijklSS'}a_{iS}{}^\dagger a_{jS}a_{kS'}{}^\dagger a_{lS'}, \tag{32}$$

where a_{iS} and $a_{iS}{}^\dagger$ are annihilation and creation operators for particles of species S in single-particle state i. The controllable[8] constants of the motion are the total energy H, the number operator N_S for particles of species S, and the total linear momentum operator $\mathbf{P}$. The number operators and the total linear momentum operator have the general forms

$$N_S=\sum_{ij}n_{ijS}a_{iS}{}^\dagger a_{jS} \tag{33}$$

and

$$\mathbf{P}=\sum_S\sum_{ij}\mathbf{p}_{ijS}a_{iS}{}^\dagger a_{jS}, \tag{34}$$

where the quantities n_{ijS} and $\mathbf{p}_{ijS}$ are matrix elements whose precise forms depend upon the choice of representation [cf. (3) and (4)]. The thermodynamic properties of the system are determined by the generalized partition function

$$Z=\mathrm{Tr}\exp[-\beta(H-\textstyle\sum_S\mu_S N_S-\mathbf{u}\cdot\mathbf{P})]\equiv e^{-\beta W} \tag{35}$$

---

[18] To see how this comes about one can adopt an alternative approach due to Bogolubov, Zubarev, and Tserkovnikov.[16] Instead of writing (20) as an inequality, one expands the exponential in powers of $\mathcal{H}_1$, obtaining (note that $\mathcal{H}_0$ and $\mathcal{H}_1$ commute)

$$W_{\rm var}=-\beta^{-1}\ln\mathrm{Tr}\exp(-\beta\mathcal{H}_0)-\beta^{-1}\ln[1+\sum_{l=1}^{\infty}(l!)^{-1}(-\beta)^l\langle\mathcal{H}_1{}^l\rangle_0].$$

With the choice (24) of f_i, the $l=1$ term is only $O(1)$ [Eq. (27)]. The BZT argument,[16] which we sketch here, shows that this is also true of the terms with $l\geq 2$.

One writes

$$\langle\mathcal{H}_1{}^l\rangle_0=2^{-l}\sum_{i_1\cdots i_l}\sum_{j_1\cdots j_l}W_{i_1j_1}\cdots W_{i_lj_l}\langle(N_{i_1}-f_{i_1})(N_{j_1}-f_{j_1})\cdots$$
$$\times(N_{i_l}-f_{i_l})(N_{j_l}-f_{j_l})\rangle_0.$$

The BZT argument shows that as a consequence of (26), the only nonvanishing terms are those in which not more than l indices are different, and hence, by (19), $\langle\mathcal{H}_1{}^l\rangle_0=O(1)$. Under the assumption of convergence of the sum over l it follows that (25), with f_i given by (24), is an exact expression for $W_{\rm var}$ to $O(\Omega)$. The BZT proof has been questioned because of this convergence assumption, especially because (24) and the associated equations for the variational parameters in (15) may admit more than one solution, as in the example of the BCS model of superconductivity, where there is a "normal-state" solution as well as the "superconducting" solution. It is therefore reassuring to know that (18) can be evaluated by an entirely different method,[19] based on the familiar procedure of picking out the largest term in the partition sum; the single-particle states must first be collected into groups of states "adjacent in i-space" in order to avoid problems associated with large fluctuations of individual occupation numbers due to the interactions. This alternative procedure leads to precisely the same Eqs. (24) and (25), provided that Eqs. (19) are satisfied; trivial solutions of (24) are to be rejected, since only one solution maximizes the largest term in the partition sum. It should also be mentioned that there exists an alternative version of the BZT proof, due to Bogolubov,[20] which avoids the use of perturbation theory, but is mathematically much more complex. Finally, it should be noted that

$$W\leq W_{\rm var}\leq W_{\rm var}'$$

rigorously, where $W_{\rm var}'$ is the right side of (20), and that $W_{\rm var}'$ is *rigorously* equal to the right side of (25) to $O(\Omega)$ provided that (19) and (24) are satisfied; the fact that $W_{\rm var}'=W_{\rm var}$ to $O(\Omega)$ is additional information which is not really necessary for the variational argument.

[19] B. A. Jacobsohn (private communication).

[20] N. N. Bogolubov, Physica **26**, 1 (1960).

where μ_S is the chemical potential of species S. Peierls' theorem (9) still holds after replacement of the statistical operator in (9) by the more general operator in (35), and can be used to obtain an upper bound $W_{\rm var}$ to the thermodynamic potential W. We choose the states $|\alpha\rangle$ in Peierls' theorem to be of the form

$$|\{\eta_{iS}\}\rangle = U[\prod_S \prod_i (\eta_{iS}!)^{-\frac{1}{2}}(a_{iS}^\dagger)^{\eta_{iS}}]|0\rangle \quad (36)$$

where

$$U^{-1}a_{iS}U = s_{iS} + \sum_j (f_{ijS}a_{jS} + g_{ijS}a_{jS}^\dagger). \quad (37)$$

The nonnegative integers η_{iS} specify the numbers of elementary excitations with properties labeled by the subscripts i and S.[21] The lower bound $Z_{\rm var}$ to Z is

$$Z_{\rm var} = {\rm Tr}\ \exp\{-\beta[U^{-1}(H - \sum_S \mu_S N_S - \mathbf{u}\cdot\mathbf{P})U]_{\rm diag}\} \quad (38)$$

where the diagonal part of the transformed operator has the general form[22]

$$[U^{-1}(H - \sum_S \mu_S N_S - \mathbf{u}\cdot\mathbf{P})U]_{\rm diag} = W_0 + \sum_S \sum_i W_{iS}N_{iS} + \tfrac{1}{2}\sum_S \sum_{ij} W_{ijS}N_{iS}N_{jS} = \sum_{(S,S')} \sum_{ij} C_{ijSS'}N_{iS}N_{jS'} \quad (39)$$

with $N_{iS} \equiv a_{iS}^\dagger a_{iS}$. Equations (36)–(39) are obvious generalizations of (14), (15), (18), and (17).

The problem of evaluating $W_{\rm var}$ can now be reduced to the corresponding problem for a one-component system by introduction of the appropriate notation. We merely have to notice that the labels i, j, $\cdots$ already include internal variables (such as spin). But there is nothing to prevent us from also considering the species label S as an internal variable; indeed, this interpretation corresponds closely with the physics, since, e.g., an He^3 atom with a given momentum and an He^4 atom with the same momentum differ only in their internal structure. We shall therefore represent the pair (i,S) by a single boldface label $\mathbf{i}$; it is then to be understood that the nonboldface label i includes the appropriate external variable (usually the momentum) and all internal variables *except* the species label S. Hence $\sum_{\mathbf{i}}$ is to be interpreted as $\sum_S \sum_i$. Equation (39) then becomes

$$[\ ]_{\rm diag} = W_0 + \sum_{\mathbf{i}} W_{\mathbf{i}}N_{\mathbf{i}} + \tfrac{1}{2}\sum_{\mathbf{ij}} W_{\mathbf{ij}}N_{\mathbf{i}}N_{\mathbf{j}} \quad (40)$$

where

$$W_{\mathbf{ij}} = \begin{cases} W_{ijS} \text{ if } \mathbf{i} \text{ and } \mathbf{j} \text{ belong to the same species } S \\ C_{ijSS'} \text{ if } \mathbf{i} \text{ belongs to species } S \text{ and } \mathbf{j} \text{ to species } S' \ (\neq S). \end{cases} \quad (41)$$

The derivation in Sec. II B then goes through essentially unchanged; one finds that

$$W_{\rm var} = W_0 + \beta^{-1}\sum_i [\pm \ln(1 \mp f_i)] - \tfrac{1}{2}\sum_{ij} W_{ij}f_i f_j, \quad (42)$$

where f_i satisfies the integral equation

$$f_i = \exp[\beta(W_i + \textstyle\sum_j W_{ij}f_j) \pm 1]^{-1}. \quad (43)$$

The variational equation (30) is

$$\frac{\partial W_0}{\partial \phi_k} + \sum_i f_i\left(\frac{\partial W_i}{\partial \phi_k}\right) + \tfrac{1}{2}\sum_{ij} f_i f_j\left(\frac{\partial W_{ij}}{\partial \phi_k}\right) = 0 \quad (44)$$

where the ϕ_k are the *independent* parameters defining the canonical transformation (37). Upon transforming (42)–(44) back into the notation employed in (31)–(39), one finds

$$W_{\rm var} = W_0 + \beta^{-1}\sum_S [\pm \sum_i \ln(1 \mp f_{iS})] - \tfrac{1}{2}\sum_S \sum_{ij} W_{ijS}f_{iS}f_{jS} - \sum_{(S,S')} \sum_{ij} C_{ijSS'}f_{iS}f_{jS'}, \quad (45)$$

where the elementary excitation distribution functions f_{iS} satisfy the integral equations

$$f_{iS} = \{\exp[\beta(W_{iS} + \textstyle\sum_j W_{ijS}f_{jS} + \sum_{S'(\neq S)} \sum_j C_{ijSS'}f_{jS'})] \pm 1\}^{-1}. \quad (46)$$

The variational equations are

$$\frac{\partial W_0}{\partial \phi_k} + \sum_S \sum_i f_{iS}\left(\frac{\partial W_{iS}}{\partial \phi_k}\right) + \tfrac{1}{2}\sum_S \sum_{ij} f_{iS}f_{jS}\left(\frac{\partial W_{ijS}}{\partial \phi_k}\right) + \sum_{(S,S')} \sum_{ij} f_{iS}f_{jS'}\left(\frac{\partial C_{ijSS'}}{\partial \phi_k}\right) = 0. \quad (47)$$

III. EXAMPLE: THE DEGENERATE IMPERFECT BOSE GAS

A. Grand Ensemble

We consider a system of bosons with spherically symmetric, pairwise additive interactions. We put $u = 0$ in (5) and subsequent equations, since the greater generality afforded by taking $u \neq 0$ does not add anything to the illustration of the formalism of Sec. II. One has

$$H - \mu N = \sum_k (\tfrac{1}{2}k^2 - \mu)a_k^\dagger a_k + \tfrac{1}{2}\Omega^{-1}\sum_{qkk'} v(q)a_{k+q}^\dagger a_{k'-q}^\dagger a_{k'}a_k \quad (48)$$

where the indices label single-particle momenta and $v(q)$ is the interaction in momentum space, a function only of $|q|$; we assume that $v(0) > 0$, i.e., that the interaction is mainly repulsive. We restrict the general linear canonical transformation (15) as follows:

$$U^{-1}a_k U = s_0\delta_{k0} + (1 - \phi_k^2)^{-\frac{1}{2}}(a_k - \phi_k a_{-k}^\dagger), \quad s_0 = s_0^*, \quad \phi_k = \phi_{-k} = \phi_k^*, \quad |\phi_k| < 1. \quad (49)$$

The part of the transformation mixing a_k and $a_{-k}^\dagger$ is a Bogolubov transformation,[9] while the term $s_0\delta_{k0}$ is introduced so as to permit a rigorous treatment of the effects of Bose-Einstein condensation into the zero-momentum single-particle state.[23] The diagonal part of the transform of (48) then has the form (17) with

[21] It should not, however, be assumed that the elementary excitations with species label S propagate through system S alone; they may, e.g., represent sound waves propagating through the mixture of all species. Our labeling is motivated by the idea of an "adiabatic switching on" of the interactions between different species.

[22] We assume without loss of generality that $W_{ijS} = W_{jiS}$ and $C_{ijSS'} = C_{jiS'S}$.

[23] This method of treating the effects of macroscopic occupation of the $k = 0$ state is due to Gross [Eq. (2.7) of reference 2].

$$W_0=-\Omega\mu\rho_0+\tfrac{1}{2}\Omega\nu(0)\rho_0^2+\sum_k\{\tfrac{1}{2}k^2-\mu+\rho_0[\nu(0)+\nu(k)]\}\left(\frac{\phi_k^2}{1-\phi_k^2}\right)-\sum_k\rho_0\nu(k)\left(\frac{\phi_k}{1-\phi_k^2}\right)$$
$$+\tfrac{1}{2}\Omega^{-1}\sum_{kk'}\nu(k-k')\left(\frac{\phi_k}{1-\phi_k^2}\right)\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)+\tfrac{1}{2}\Omega^{-1}\sum_{kk'}[\nu(0)+\nu(k-k')]\left(\frac{\phi_k^2}{1-\phi_k^2}\right)\left(\frac{\phi_{k'}^2}{1-\phi_{k'}^2}\right),$$
$$W_k=\left(\frac{1+\phi_k^2}{1-\phi_k^2}\right)\left\{\tfrac{1}{2}k^2-\mu+\rho_0[\nu(0)+\nu(k)]+\Omega^{-1}\sum_{k'}[\nu(0)+\nu(k-k')]\left(\frac{\phi_{k'}^2}{1-\phi_{k'}^2}\right)\right\}$$
$$-2\left(\frac{\phi_k}{1-\phi_k^2}\right)\left[\rho_0\nu(k)-\Omega^{-1}\sum_{k'}\nu(k-k')\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\right], \tag{50}$$
$$W_{kk'}=2\Omega^{-1}[\nu(k-k')+\nu(k+k')]\left(\frac{\phi_k}{1-\phi_k^2}\right)\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)+\Omega^{-1}[\nu(0)+\nu(k-k')]\left[\frac{1+\phi_k^2\phi_{k'}^2}{(1-\phi_k^2)(1-\phi_{k'}^2)}\right]$$
$$+\Omega^{-1}[\nu(0)+\nu(k+k')]\left[\frac{\phi_k^2+\phi_{k'}^2}{(1-\phi_k^2)(1-\phi_{k'}^2)}\right]$$

with

$$\rho_0\equiv s_0^2/\Omega. \tag{51}$$

In obtaining (50) we have omitted terms in W_0 of order unity, in W_k of order Ω^{-1}, and in $W_{kk'}$ of order Ω^{-2}, since these cannot affect the thermodynamic potential to $O(\Omega)$.

The integral equation (24) for the elementary excitation distribution function f_k can be written[24]

$$f_k=(e^{\beta\epsilon_k}-1)^{-1},$$
$$\epsilon_k=\left(\frac{1+\phi_k^2}{1-\phi_k^2}\right)\left\{\tfrac{1}{2}k^2-\mu+\rho_0[\nu(0)+\nu(k)]+\Omega^{-1}\sum_{k'}[\nu(0)+\nu(k-k')]\left[\frac{\phi_{k'}^2}{1-\phi_{k'}^2}+f_{k'}\left(\frac{1+\phi_{k'}^2}{1-\phi_{k'}^2}\right)\right]\right\}$$
$$-2\left(\frac{\phi_k}{1-\phi_k^2}\right)\left[\rho_0\nu(k)-\Omega^{-1}\sum_{k'}\nu(k-k')(1+2f_{k'})\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\right] \tag{52}$$

using the definition (21) of ϵ_k and the expressions (50) for W_k and $W_{kk'}$. This is completely equivalent to Tolmachev's[6] first Eq. (7), as can be seen by making the following changes of notation in Tolmachev's equations:

$$(N_0/V)\rightarrow\rho_0,\quad E(k)\rightarrow\tfrac{1}{2}k^2,\quad n_k\rightarrow f_k,\quad u_k\rightarrow(1-\phi_k^2)^{-\frac{1}{2}},\quad v_k\rightarrow-\phi_k(1-\phi_k^2)^{-\frac{1}{2}}. \tag{53}$$

The variational equation (30) for ϕ_k is found with the aid of (50) to be[24]

$$2\phi_k\left\{\tfrac{1}{2}k^2-\mu+\rho_0[\nu(0)+\nu(k)]-\Omega^{-1}\sum_{k'}[\nu(0)+\nu(k-k')]\left[\frac{\phi_{k'}^2}{1-\phi_{k'}^2}+f_{k'}\left(\frac{1+\phi_{k'}^2}{1-\phi_{k'}^2}\right)\right]\right\}$$
$$-(1+\phi_k^2)\left[\rho_0\nu(k)-\Omega^{-1}\sum_{k'}\nu(k-k')(1+2f_{k'})\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\right]=0, \tag{54}$$

which is equivalent to Tolmachev's second Eq. (7). Finally, the variational equation for ρ_0, obtained by replacing ϕ_k by ρ_0 in (30), is

$$-\mu+\rho_0\nu(0)+\Omega^{-1}\sum_k[\nu(0)+\nu(k)]$$
$$\times\left[\frac{\phi_k^2}{1-\phi_k^2}+f_k\left(\frac{1+\phi_k^2}{1-\phi_k^2}\right)\right]$$
$$-\Omega^{-1}\sum_k\nu(k)(1+2f_k)\left(\frac{\phi_k}{1-\phi_k^2}\right)=0 \tag{55}$$

which is equivalent to Tolmachev's first Eq. (8) after correction of a minor error.[25] However, we do not obtain any analog of Tolmachev's second Eq. (8), since the variational principle (12) is only rigorously valid for *given* μ (*not* given n). The agreement of (52), (54), and (55) with Tolmachev's results was to be expected, since his work was also based on the variational principle (20). However, Tolmachev's treatment of the effects of Bose-Einstein condensation does differ somewhat from ours. Tolmachev followed Bogolubov's

[24] We assume that f_k, like ϕ_k, is an even function of k.

[25] The coefficient of the first summation in Tolmachev's Eq. (8) should be $(1/V)$ instead of $(1/2V)$.

procedure[9] of simply replacing the annihilation and creation operators a_0 and $a_0^\dagger$ by a c-number $n_0^{\frac{1}{2}}$ determined by a self-consistency condition, whereas we introduced an additional variational parameter $\rho_0=(s_0^2/\Omega)$ into the quasi-particle transformation (49) and determined ρ_0 by the variational equation (55). The agreement between our results and those of Tolmachev shows that the simple Bogolubov replacement is in fact rigorously justified, at least in a variational sense.

We have restricted ourselves to the special form (49) of the general linear canonical transformation (15) in order to simplify the analysis and permit ready comparison with the work of Tolmachev. Gross has shown[2] that a more general transformation may be necessary in order to treat an interaction with a sufficiently strong attractive tail. Our method would provide a generalization of Gross' work to nonzero temperature.

B. Canonical Ensemble

The results of Sec. III(A) suffer from the disadvantage that they depend upon the unknown *exact* chemical potential μ, which cannot be determined from knowledge of the *approximate* thermodynamic potential. It is possible to take advantage of the macroscopic occupation of the zero-momentum state so as to permit evaluation of a lower bound to the *canonical* partition function, hence an upper bound to the *Helmholtz* free energy. Then the known number of particles, rather than the unknown chemical potential, appears in the equations. We shall carry out such a treatment in this section. This treatment is suggested by and in fact equivalent to Wentzel's treatment[7] of the imperfect Bose gas, and suggests a more general equivalence between Wentzel's method of the "thermodynamically equivalent Hamiltonian" [7] and our variational method with appropriate choices of variational parameters.

The Hamiltonian is given by (48) with the term μN omitted. We shall restrict the general linear canonical transformation (15) to a Bogolubov transformation mixing a_k and $a_{-k}^\dagger$. Then the only part of H having diagonal matrix elements after the transformation, and hence contributing to W_{var}, is the "pair Hamiltonian," [26] a function only of $N_k\equiv a_k^\dagger a_k$, $\alpha_k\equiv a_{-k}a_k$, and $\alpha_k^\dagger=a_k^\dagger a_{-k}^\dagger$:

$$H_P=\tfrac{1}{2}(n-1)\rho v(0)+\textstyle\sum_k{}'\,[\tfrac{1}{2}k^2+(N_0/\Omega)v(k)]N_k +\tfrac{1}{2}\Omega^{-1}\textstyle\sum_k{}'\,v(k)(\alpha_k^\dagger\alpha_0+\alpha_0^\dagger\alpha_k) +\tfrac{1}{2}\Omega^{-1}\textstyle\sum_{kk'}{}'\,v(k-k')N_kN_{k'} +\tfrac{1}{2}\Omega^{-1}\textstyle\sum_{kk'}{}'\,v(k-k')\alpha_k^\dagger\alpha_{k'} \quad (56)$$

where the primes on the summation forbid any momentum index not explicitly zero from assuming the value zero; in the next to the last summation $k=k'$ is also to be excluded, and in the last summation $k=\pm k'$. Since we intend to evaluate the thermodynamic potential in a canonical ensemble rather than a grand ensemble, all states contributing to the trace will be eigenstates of the total particle number operator N belonging to eigenvalue n; we have accordingly separated out[27] explicitly in (56) a part of H depending only on N and replaced N by its eigenvalue n; ρ is the particle number density (n/Ω).

In order to simplify the mathematics it is desirable to eliminate N_0, α_0, and $\alpha_0^\dagger$ from (56). Following Wentzel,[7] N_0 can be eliminated by the substitution (the primed summation excludes $k=0$)

$$N_0=N-\textstyle\sum_k{}'\,N_k\rightarrow n-\sum_k{}'\,N_k \quad (57)$$

which is rigorously justified for a canonical ensemble; however, elimination of α_0 and $\alpha_0^\dagger$ is not quite so trivial. We first introduce the familiar phase-amplitude decomposition of the annihilation and creation operators a_0 and $a_0^\dagger$:

$$a_0=e^{i\Theta_0}N_0^{\frac{1}{2}}\equiv\beta_0^{\frac{1}{2}}N_0^{\frac{1}{2}},\quad a_0^\dagger=N_0^{\frac{1}{2}}e^{-i\Theta_0}=N_0^{\frac{1}{2}}\beta_0^{-\frac{1}{2}} \quad (58)$$

(we recall that $N_0\equiv a_0^\dagger a_0$). Then[28]

$$\alpha_0=a_0^2=\beta_0^{\frac{1}{2}}N_0^{\frac{1}{2}}\beta_0^{\frac{1}{2}}N_0^{\frac{1}{2}}=\beta_0[N_0(N_0-1)]^{\frac{1}{2}} =\beta_0(N_0-\tfrac{1}{2}-\tfrac{1}{8}N_0^{-1}+\cdots), \quad (59)$$

$$\alpha_0^\dagger=(N_0-\tfrac{1}{2}-\tfrac{1}{8}N_0^{-1}+\cdots)\beta_0^{-1}.$$

The indicated expansion of the square root is justified in the limit $\Omega\rightarrow\infty$ provided that the states contributing to the thermodynamic potential to $O(\Omega)$ all have $N_0=O(\Omega)$ [or equivalently $N_0=O(n)$], i.e., provided that there is Bose-Einstein condensation into the single-particle state with zero momentum.[29] Then all terms in the expansion except the leading term N_0 may be dropped without affecting the thermodynamic potential to $O(\Omega)$, and with (57) one obtains

$$\alpha_0\rightarrow\beta_0(n-\textstyle\sum_k{}'\,N_k),\quad \alpha_0^\dagger\rightarrow\beta_0^{-1}(n-\sum_k{}'\,N_k). \quad (60)$$

Substitution of (57) and (60) into (56) then gives

$$H_P=\tfrac{1}{2}n\rho v(0)+\textstyle\sum_k{}'\,[f_k^{\text{II}}N_k+\tfrac{1}{2}h_k^{\text{II}}(\beta_0^{-1}\alpha_k+\beta_0\alpha_k^\dagger)] +\tfrac{1}{2}\textstyle\sum_{kk'}{}'\,[i_{kk'}N_kN_{k'}+j_{kk'}\alpha_k^\dagger\alpha_{k'} +l_{kk'}(\beta_0^{-1}\alpha_kN_{k'}+\beta_0N_{k'}\alpha_k^\dagger)]+O(1) \quad (61)$$

where the quantities f_k^{II}, h_k^{II}, $i_{kk'}$, $j_{kk'}$, and $l_{kk'}$ are defined by Wentzel's Eq. (24) with appropriate changes of notation; here and henceforth, all primed summations exclude zero momentum indices. The term "$O(1)$" does not contribute to the thermodynamic potential to $O(\Omega)$ at temperatures below the Bose-Einstein con-

[26] M. Girardeau and R. Arnowitt, Phys. Rev. **113**, 755 (1959).

[27] This is done by picking out the $q=0$ terms in (48) and applying the Bose commutation relations.

[28] Compare with Eq. (15) of reference 26.

[29] This heuristic argument can be given a more rigorous justification by noting that

$$\text{Tr}_n e^{-\beta H}>\text{Tr}_n' e^{-\beta H}$$

where Tr_n is a trace over a complete set of n-particle states and Tr_n' is a trace over an incomplete set from which states failing to satisfy the condition $N_0=O(n)$ are excluded. Thus inclusion only of states with $N_0=O(n)$ leads to a rigorous upper bound to the free energy $F=-\beta^{-1}\ln\text{Tr}_n e^{-\beta H}$, which is sufficient for a variational treatment.

densation temperature. Equation (61) differs from Wentzel's Eq. (23) only in the occurence of the unitary zero-momentum-pair annihilation and creation operators β_0 and β_0^{-1}; our main reason for carrying out the derivation of (61) here is to show that the usual procedure[9] of treating a_0 and $a_0^\dagger$ as c numbers can be justified by an equivalent but more rigorous treatment.

We now subject the annihilation and creation operators with $k\neq 0$ to a modified Bogolubov transformation[9,26]:

$$\begin{aligned} U^{-1}a_kU&=(1-\phi_k^2)^{-\frac{1}{2}}(a_k-\phi_k\beta_0a_{-k}^\dagger),\\ U^{-1}a_k^\dagger U&=(1-\phi_k^2)^{-\frac{1}{2}}(a_k^\dagger-\phi_k\beta_0^{-1}a_{-k}),\\ \phi_k&=\phi_{-k}=\phi_k^*,\quad |\phi_k|<1; \end{aligned} \tag{62}$$

it should be noted that, in contradistinction to (49), the transformation (62) conserves the total number of particles, since β_0 and $\beta_0^{-1}(=\beta_0^\dagger)$, respectively, annihilate and create two particles with momentum zero. Since we wish to work with a canonical ensemble, we write (14) in the modified form

$$\begin{aligned} |\{\eta_k\}\rangle&=U[\textstyle\prod_k{}'(\eta_k!)^{-\frac{1}{2}}\beta_0^{\eta_k}(a_k^\dagger)^{\eta_k}]\,|n\rangle,\\ |n\rangle&\equiv(n!)^{-\frac{1}{2}}(a_0^\dagger)^n\,|0\rangle, \end{aligned} \tag{63}$$

and accordingly write (9) and (10) in the form

$$F\leqslant F_{\text{var}}\equiv-\beta^{-1}\ln \text{Tr}_n \exp[-\beta(U^{-1}HU)_{\text{diag}}] \tag{64}$$

where F is the *Helmholtz* free energy and Tr_n denotes a trace over states of n particles. It follow from (61) and (62) that

$$\begin{aligned} (U^{-1}HU)_{\text{diag}}&=(U^{-1}H_PU)_{\text{diag}}\\ &=W_0+\textstyle\sum_k{}'W_kN_k+\frac{1}{2}\sum_{kk'}{}'W_{kk'}N_kN_{k'} \end{aligned} \tag{65}$$

with

$$\begin{aligned} W_0&=\tfrac{1}{2}n\rho v(0)+\textstyle\sum_k{}'\left[f_k^{\text{II}}\left(\frac{\phi_k^2}{1-\phi_k^2}\right)-h_k^{\text{II}}\left(\frac{\phi_k}{1-\phi_k^2}\right)\right]+\frac{1}{2}\sum_{kk'}{}'\left[i_{kk'}\left(\frac{\phi_k^2}{1-\phi_k^2}\right)\left(\frac{\phi_{k'}^2}{1-\phi_{k'}^2}\right)\right.\\ &\qquad\left.+j_{kk'}\left(\frac{\phi_k}{1-\phi_k^2}\right)\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)-2l_{kk'}\left(\frac{\phi_k}{1-\phi_k^2}\right)\left(\frac{\phi_{k'}^2}{1-\phi_{k'}^2}\right)\right],\\ W_k&=\left(\frac{1+\phi_k^2}{1-\phi_k^2}\right)\left\{f_k^{\text{II}}+\textstyle\sum_{k'}{}'\left[i_{kk'}\left(\frac{\phi_{k'}^2}{1-\phi_{k'}^2}\right)-l_{k'k}\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\right]\right\}\\ &\qquad-2\left(\frac{\phi_k}{1-\phi_k^2}\right)\left\{h_k^{\text{II}}+\textstyle\sum_{k'}{}'\left[l_{kk'}\left(\frac{\phi_{k'}^2}{1-\phi_{k'}^2}\right)-j_{kk'}\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\right]\right\},\\ W_{kk'}&=i_{kk'}\left[\frac{1+\phi_k^2\phi_{k'}^2}{(1-\phi_k^2)(1-\phi_{k'}^2)}\right]+i_{k,-k'}\left[\frac{\phi_k^2+\phi_{k'}^2}{(1-\phi_k^2)(1-\phi_{k'}^2)}\right]+2(j_{kk'}+j_{k,-k'})\left(\frac{\phi_k}{1-\phi_k^2}\right)\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\\ &\qquad-2l_{kk'}\left(\frac{\phi_k}{1-\phi_k^2}\right)\left(\frac{1+\phi_{k'}^2}{1-\phi_{k'}^2}\right)-2l_{k'k}\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\left(\frac{1+\phi_k^2}{1-\phi_k^2}\right). \end{aligned} \tag{66}$$

The derivation of the expression for F_{var} goes through as in Sec. II(B) with the differences that F_{var}, the *Helmholtz* free energy, replaces W_{var} and that summation exclude $k=0$, and subject to the restriction that the result obtained is only valid at temperatures low enough that $n-\sum_{k\neq 0}\langle N_k\rangle=O(n)$ where $\langle N_k\rangle$ denotes the canonical average number of particles with momentum k. F_{var} is given by (25):

$$F_{\text{var}}=W_0-\beta^{-1}\textstyle\sum_k{}'\ln(1+f_k)-\frac{1}{2}\sum_{kk'}{}'W_{kk'}f_kf_{k'} \tag{67}$$

where the distribution function f_k of elementary excitations is determined by the nonlinear integral equation (24):

$$f_k=\{\exp[\beta(W_k+\textstyle\sum_{k'}{}'W_{kk'}f_{k'})]-1\}^{-1}. \tag{68}$$

The variational integral equation (30) for ϕ_k becomes, after substitution from (66) and evaluation of the derivatives,[24]

$$\begin{aligned} &2\phi_k\left\{f_k^{\text{II}}+\textstyle\sum_{k'}{}'\left[i_{kk'}\left(\frac{\phi_{k'}^2}{1-\phi_{k'}^2}\right)-l_{k'k}\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\right]+\sum_{k'}{}'f_{k'}\left[i_{kk'}\left(\frac{1+\phi_{k'}^2}{1-\phi_{k'}^2}\right)-2l_{k'k}\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\right]\right\}\\ &-(1+\phi_k^2)\left\{h_k^{\text{II}}+\textstyle\sum_{k'}{}'\left[l_{kk'}\left(\frac{\phi_{k'}^2}{1-\phi_{k'}^2}\right)-j_{kk'}\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\right]+\sum_{k'}{}'f_{k'}\left[l_{kk'}\left(\frac{1+\phi_{k'}^2}{1-\phi_{k'}^2}\right)-2j_{kk'}\left(\frac{\phi_{k'}}{1-\phi_{k'}^2}\right)\right]\right\}=0. \end{aligned} \tag{69}$$

At temperature $T=0$ ($\beta=\infty$), f_k vanishes and (67) and (69) reduce to previously known results.[30] For arbitrary temperature Eqs. (68) and (69) are completely equivalent to Wentzel's Eqs. (29), (32), and (40), and our expression (67) for the free energy is equivalent to Wentzel's expression (35). This can be seen by making the substitutions[31]

$$\begin{aligned} f_k &\rightarrow 1/(e^{\beta\epsilon_k}-1), \\ \phi_k/(1-\phi_k^2) &\rightarrow h_k/2\epsilon_k, \\ \phi_k^2/(1-\phi_k^2) &\rightarrow \tfrac{1}{2}(f_k/\epsilon_k-1), \\ (1+\phi_k^2)/(1-\phi_k^2) &\rightarrow f_k/\epsilon_k \end{aligned} \tag{70}$$

in our equations; the algebraic reduction is straightforward. Wentzel has shown that his method gives results which are exact to $O(\Omega)$ for Hamiltonians of a certain special class [Eq. (2) of reference 7] to which the pair Hamiltonian (56) belongs. Because of the equivalence between our results and his, we conclude that our variational method also gives the thermodynamic potential of the pair Hamiltonian (56) exactly to $O(\Omega)$.[32] This result is perhaps somewhat superfluous, since our variational method is designed to give an *upper bound* to the thermodynamic potential of the *full* Hamiltonian [(48) with μN omitted], rather than to give the thermodynamic potential of a *portion* of the Hamiltonian *exactly*. Nevertheless, the equivalence between our method and Wentzel's in this case shows that Wentzel's method also gives an upper bound to the thermodynamic potential of the full Hamiltonian. This conclusion also holds for Fermi statistics [BCS model of superconductivity[4,5,16]], and it no doubt holds for Wentzel's more general partial Hamiltonian [Eq. (2) of reference 7] as well as for the pair Hamiltonian.

The treatments of the imperfect Bose gas carried out in Sec. III(A) and in this section represent two different variational approximations. That of Sec. III(A), based on the grand ensemble, is better suited to illustrate the general theory of Sec. II; the treatment in this section, based on the canonical ensemble, is more convenient for actual calculations because the unknown chemical potential does not occur, but the fact that we were able to treat the canonical ensemble was a result of the macroscopic occupation of the $k=0$ state, and does not carry over to Fermi systems or to Bose systems in which Bose-Einstein condensation is absent.

[30] Equations (21)–(23) of reference 26.
[31] Do not confuse our f_k with Wentzel's f_k.
[32] This result was already known for the special case of zero temperature; cf. Appendix B of reference 26.

ACKNOWLEDGMENTS

It is a pleasure for me to thank A. Isihara, G. Wentzel, Elliott Lieb, R. V. Hanks, B. Mühlschlegel, and B. A. Jacobsohn for valuable discussions, suggestions, and criticisms. I am particularly indebted to Professor Mühlschlegel and Professor Jacobsohn for pointing out an error in a previous evaluation[15] of (18), and for suggesting correct methods of evaluation; the method based on Eqs. (20) and (21) is due to Mühlschlegel, and a different method due to Jacobsohn[19] leads to the same Eqs. (24) and (25).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 1 JANUARY-FEBRUARY, 1962

Statistical Theory of the Energy Levels of Complex Systems. I

FREEMAN J. DYSON
Institute for Advanced Study, Princeton, New Jersey
(Received September 15, 1961)

New kinds of statistical ensemble are defined, representing a mathematical idealization of the notion of "all physical systems with equal probability." Three such ensembles are studied in detail, based mathematically upon the orthogonal, unitary, and symplectic groups. The orthogonal ensemble is relevant in most practical circumstances, the unitary ensemble applies only when time-reversal invariance is violated, and the symplectic ensemble applies only to odd-spin systems without rotational symmetry. The probability-distributions for the energy levels are calculated in the three cases. Repulsion between neighboring levels is strongest in the symplectic ensemble and weakest in the orthogonal ensemble. An exact mathematical correspondence is found between these eigenvalue distributions and the statistical mechanics of a one-dimensional classical Coulomb gas at three different temperatures. An unproved conjecture is put forward, expressing the thermodynamic variables of the Coulomb gas in closed analytic form as functions of temperature. By means of general group-theoretical arguments, the conjecture is proved for the three temperatures which are directly relevant to the eigenvalue distribution problem. The electrostatic analog is exploited in order to deduce precise statements concerning the entropy, or degree of irregularity, of the eigenvalue distributions. Comparison of the theory with experimental data will be made in a subsequent paper.

I. INTRODUCTION

RECENT theoretical analyses[1] have had impressive success in interpreting the detailed structure of the low-lying excited states of complex nuclei. Still, there must come a point beyond which such analyses of individual levels cannot usefully go. For example, observations of levels of heavy nuclei in the neutron-capture region[2] give precise information concerning a stretch of levels from number N to number $(N+n)$, where n is an integer of the order of 100 while N is of the order of 10^6. It is improbable that level assignments based on shell structure and collective or individual-particle quantum numbers can ever be pushed as far as the millionth level. It is therefore reasonable to inquire whether the highly excited states may be understood from the diametrically opposite point of view, assuming as a working hypothesis that all shell structure is washed out and that no quantum numbers other than spin and parity remain good. The result of such an inquiry will be a statistical theory of energy levels. The statistical theory will not predict the detailed sequence of levels in any one nucleus, but it will describe the general appearance and the degree of irregularity of the level structure that is expected to occur in any nucleus which is too complicated to be understood in detail.

In ordinary statistical mechanics a comparable renunciation of exact knowledge is made. By assuming all states of a very large ensemble to be equally probable, one obtains useful information about the over-all behavior of a complex system, when the observation of the state of the system in all its detail is impossible. This type of statistical mechanics is clearly inadequate for the discussion of nuclear energy levels. We wish to make statements about the fine detail of the level structure, and such statements cannot be made in terms of an ensemble of states. What is here required is a new kind of statistical mechanics, in which we renounce exact knowledge not of the state of a system but of the nature of the system itself. We picture a complex nucleus as a "black box" in which a large number of particles are interacting according to unknown laws. The problem then is to define in a mathematically precise way an ensemble of systems in which all possible laws of interaction are equally probable.

The idea of a statistical mechanics of nuclei based on an ensemble of systems is due to Wigner.[3] Wigner's program has been energetically pursued by Porter and Rosenzweig,[4] by Gaudin and Mehta,[5] and by others.[6] The results of this work are encouraging, but progress has been held back by the extreme difficulty of calculating the ensemble averages in Wigner's model. The difficulties seem to be of a purely mathematical nature, and they are as severe as those which arise in more orthodox statistical analyses of many-body systems with strong interactions. Only during the last year have Gaudin and Mehta[5] shown, by a beautiful exercise of analytical skill, that these difficulties are not insuperable. The way now lies open to develop the new statistical mechanics on a broad front and to use it for quantitative interpretation of experiments.

The present series of papers will explore the new statistical mechanics in its various ramifications. This, the first paper of the series, is mainly mathematical in content. Its purpose is to introduce a new type of

[1] See, for example, L. S. Kisslinger and R. A. Sorensen, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. 32, No. 9 (1960); M. Baranger, Phys. Rev. 120, 957 (1960).

[2] J. L. Rosen, J. S. Desjardins, J. Rainwater, and W. W. Havens, Jr., Phys. Rev. 118, 687 (1960); 120, 2214 (1960).

[3] E. P. Wigner, Ann. Math. 53, 36 (1951); 62, 548 (1955); 65, 203 (1957); 67, 325 (1958).

[4] C. E. Porter and N. Rosenzweig, Suomalaisen Tiedeakat. Toimituksia, AVI, No. 44 (1960), and Phys. Rev. 120, 1698 (1960).

[5] M. L. Mehta, Nuclear Phys. 18, 395 (1960); M. L. Mehta and M. Gaudin, *ibid.* 18, 420 (1960); M. Gaudin, *ibid.* 25, 447 (1961).

[6] R. G. Thomas and C. E. Porter, Phys. Rev. 104, 483 (1956); I. I. Gurevich and M. I. Pevsner, Nuclear Phys. 2, 575 (1957).

ensemble, different from those studied by Wigner and Mehta, and in this way to translate the whole subject into the language of abstract group theory. Powerful group-theoretical methods can then be applied to obtain results which would otherwise require heavy calculation. The new ensembles, while mathematically more elegant than Wigner's, are based on the same physical assumptions and imply the same consequences wherever a comparison has been made.

The question, whether either Wigner's ensembles or ours correspond well to the actual behavior of a heavy nucleus, can be answered only by experiment. In ordinary statistical mechanics, there is a rather strong logical expectation (though no rigorous mathematical proof) that an ensemble average will correctly describe the behavior of one particular system which is under observation. The expectation is strong, because the system "might be" in a huge variety of states, and very few of these states will deviate much from a properly chosen ensemble average. In the new statistical mechanics we have an ensemble of systems, and this ensemble is supposed to describe the behavior of a unique object, for example the nucleus U_{239}. The logical presumption that U_{239} is really a good sample of the ensemble cannot be compelling. It will not be a disaster if it turns out that U_{239} in fact deviates quite strongly from the ensemble average. On the contrary, deviations from the ensemble average will reveal important physical information concerning the extent to which hidden quantum numbers (shell structure or other unknown integrals of the motion) may persist into the domain of neutron capture resonances.

II. GAUSSIAN ENSEMBLE AND ORTHOGONAL ENSEMBLE

The Gaussian ensemble E_G, the most convenient in practice of the class of ensembles introduced by Wigner,[3] may be defined as follows. A system is characterized by a Hamiltonian which is a real symmetric matrix H_{ij}, $i, j=1, \cdots, N$. The integer N is fixed, and the H_{ij} for $i \leq j$ are $\frac{1}{2}(N^2+N)$ independent Gaussian random variables with the joint distribution function

$$D(H_{ij})=A \exp[-(\sum_i H_{ii}^2+2\sum_{i<j} H_{ij}^2)/4a^2]. \quad (1)$$

Here A and a are constants. The meaning of (1) is that each system having N quantum states occurs in the ensemble E_G with the statistical weight $D(H_{ij})$. The Hamiltonian is taken real rather than merely Hermitian in order to restrict attention to systems invariant under time reversal.

It was shown by Porter and Rosenzweig[4] that the special form of Eq. (1) is implied by two apparently more general requirements: (i) the various components H_{ij} to be statistically independent, and (ii) the function $D(H_{ij})$ to be invariant under all transformations $H \rightarrow R^{-1}HR$, where R is a real orthogonal matrix. The requirement (ii) is a natural one in any ensemble that attempts to give equal weight to all kinds of interactions. However, requirement (i) is artificial and without clear physical motivation. To picture the H_{ij} as resulting from some "random process" of a conventional kind does not seem reasonable. Therefore the definition of E_G remains somewhat arbitrary.

The basic reason for the unsatisfactory features of Eq. (1) is that one cannot define a uniform probability distribution on an infinite range. Thus some arbitrary restriction of the magnitudes of the H_{ij} is inevitable. It is impossible to define an ensemble in terms of the H_{ij} in which all interactions are equally probable.

By a rather slight formal change we can define a new ensemble E_1, which is free from the arbitrary features of E_G and which also turns out to be mathematically easier to handle. The ensemble E_1 will be called "the orthogonal ensemble" because its structure is closely connected with that of the N-dimensional orthogonal group. A system is represented in E_1 not by its Hamiltonian H but by an $(N \times N)$ unitary matrix S. The eigenvalues of S are N complex numbers $[\exp(i\theta_j)]$, $j=1, \cdots, N$, distributed around the unit circle. The precise connection between S and H need not be specified. We assume only that S is a function of H, so that the angles θ_i are a function of the energy levels ϵ_j of the system. Over a small range of angles, the relation between θ_j and ϵ_j will be approximately linear. Our basic statistical hypothesis is then the following: *The behavior of n consecutive levels of an actual system, where n is small compared with the total number of levels, is statistically equivalent to the behavior in the ensemble E_1 of n consecutive angles θ_i on the unit circle, where n is small compared with N.*

It may be helpful for the reader to imagine a definite relation between S and H, for example,

$$S=\exp[-iH\tau], \quad S=[1-i\tau H]/[1+i\tau H]. \quad (2)$$

However, such a definite relation will never correspond to reality except over a limited range of energy. Both the Gaussian ensemble and the orthogonal ensemble which we shall shortly define are restricted to $(N \times N)$ matrices. Both ensembles are gross mutilations of an actual nucleus, which has an infinite number of energy levels. The most one can ask of any such ensemble is that it correctly reproduces level distributions over an energy range small compared with the total energy of excitation. The relation between S and the "true Hamiltonian" is bound to be wrong, considered in the large. It is therefore better to leave the connection between S and H vague. The connection between the ensemble E_1 and physical reality is then *only* the connection which we have stated above as the basic statistical hypothesis.

The over-all distributions of energy levels predicted by the Gaussian ensemble and by the orthogonal

ensemble are both unrealistic. The Gaussian ensemble gives for the distribution in the large the famous[7] "semi-circle distribution"

$$p(\epsilon)=[2\pi Na^2]^{-1}[4Na^2-\epsilon^2]^{\frac{1}{2}}, \quad \epsilon^2<4Na^2,$$
$$p(\epsilon)=0, \quad \epsilon^2>4Na^2, \tag{3}$$

which is totally unlike the level distribution of a nucleus. The orthogonal ensemble gives for the distribution in the large a uniform distribution around the unit circle. Both distributions are unphysical, but the orthogonal distribution has the advantage of simplicity and absence of spurious end effects.

After these preliminary remarks, we now state the precise definition of the orthogonal ensemble E_1. A system is characterized by a symmetric unitary matrix S having N rows and columns. Since the space T_1 of all S is compact, it makes sense to require that the ensemble E_1 contain all possible S with equal probability. However, to give a meaning to equal probability, we require a measure μ in the space T_1. Since the S do not form a group, the definition of μ is not entirely trivial. We choose the following definition. Every S can be represented in the form

$$S=U^TU, \tag{4}$$

where U is a unitary matrix, U^T its transposed matrix. An infinitesimal neighborhood of S in T_1 is given by

$$S+dS=U^T[1+idM]U, \tag{5}$$

where dM is a real symmetric infinitesimal matrix with elements dM_{ij}, and the elements dM_{ij} for $i\leq j$ vary independently through some small intervals of lengths $d\mu_{ij}$. The measure of this neighborhood is then defined to be

$$\mu(dS)=\prod_{i\leq j} d\mu_{ij}. \tag{6}$$

The ensemble E_1 is defined by the statement: *The probability that a system of E_1 belongs to the volume-element dS is*

$$P(dS)=(V_1)^{-1}\mu(dS), \tag{7}$$

where

$$V_1=\int \mu(dS) \tag{8}$$

is the total volume of the space T_1.

To make this definition unique, it remains to be proved that $\mu(dS)$ is independent of the particular U which was chosen in Eq. (4). Suppose then

$$S=U^TU=V^TV, \tag{9}$$

where both U and V are unitary. The operator

$$R=VU^{-1} \tag{10}$$

is unitary, and also satisfies

$$R^TR=(U^T)^{-1}V^TVU^{-1}=(U^T)^{-1}U^TUU^{-1}=1. \tag{11}$$

Therefore, R is real and orthogonal. Let

$$\mu'(dS)=\prod_{i\leq j} d\mu_{ij}' \tag{12}$$

be the measure derived from V as $\mu(dS)$ was derived from U. We have

$$S+dS=V^T[1+idM']V \tag{13}$$

with

$$dM'=RdMR^{-1}. \tag{14}$$

To prove $\mu(dS)=\mu'(dS)$, we need to show that the Jacobian

$$J=\det|\partial dM_{ij}'/\partial dM_{kl}| \tag{15}$$

has absolute value unity, when dM, dM' are real symmetric matrices related by Eq. (14). To prove $|J|=1$ in general, it is sufficient to consider only two special forms of R, (i) R is a reflection

$$R_{ij}=\eta_j\delta_{ij}, \quad \eta_j=\pm 1, \tag{16}$$

and (ii) R is an infinitesimal rotation

$$R_{ij}=\delta_{ij}+a_{ij}, \quad a_{ij}=-a_{ji}. \tag{17}$$

In case (i) the result $|J|=1$ is trivial. In case (ii) we have to first order in the a_{ij}:

$$(\partial dM_{ij}'/\partial dM_{kl})=\delta_{ik}\delta_{jl}+\delta_{ik}a_{jl}+a_{ik}\delta_{jl}+(i\rightleftarrows j), \tag{18}$$

$$J=1+2\,\mathrm{spur}(a_{ij})=1. \tag{19}$$

This proves that the measure $d\mu(S)$ is unique. Incidentally, we have established that for fixed S the unitary matrix U in Eq. (4) is undetermined precisely to the extent of a transformation

$$U\rightarrow RU, \tag{20}$$

where R is an arbitrary real orthogonal matrix.

The motivation for the choice of the ensemble E_1 will become clearer in view of the following theorem.

Theorem 1. The orthogonal ensemble E_1 is uniquely defined, in the space T_1 of symmetric unitary matrices, by the property of being invariant under every automorphism

$$S\rightarrow W^TSW \tag{21}$$

of T_1 into itself, where W is any unitary matrix.

Theorem 1 comprises two statements, (i) that E_1 is invariant under the automorphisms (21), and (ii) that no other ensemble is invariant. To prove (i), we suppose that a neighborhood $S+dS$ of S is transformed into a neighborhood $S'+dS'$ of S' by the automorphism (21). Equations (4) and (5) then hold, and therefore

$$S'=V^TV, \quad V=UW, \tag{22}$$

[7] E. P. Wigner, Ann. Math. 65, 203 (1957).

$$S'+dS'=V^T(1+idM)V. \tag{23}$$

The measures $\mu(dS)$ and $\mu(dS')$ are then identical by definition. This proof of (i) becomes trivial because we made a convenient choice of the unitary operator V associated with S' by Eq. (22); it was shown before that the value of $\mu(dS')$ is independent of the choice of V. To prove (ii), let E_1' be any ensemble invariant under Eq. (21). The probability distribution of E_1' will define a certain measure $\mu'(dS)$ of neighborhoods in T_1. The ratio

$$\varphi(S)=\mu'(dS)/\mu(dS) \tag{24}$$

is a function of S defined on T_1 and invariant under Eq. (21). But Eq. (4) shows that every S may be transformed into the identity operator by Eq. (21). Therefore $\varphi(S)=\varphi(I)=$constant, and E_1' is identical with E_1.

Theorem 1 states in mathematical language the precise meaning of the vague statement "all systems occur in E_1 with equal probability." The point here is that the automorphism (21) is not a mere change in the representation of states; it is a physical alteration of the system S into a different system. Intuitively speaking, we may visualize S as representing an unknown system enclosed in a "black box," S being the transformation matrix of the system from some initial state φ to some final state ψ. The transformation $S\rightarrow W^TSW$ then means that we subject the initial state to some further interaction W, and the final state to the same interaction W^T in a time-symmetric manner. If we are totally ignorant of the interactions occurring inside the black box, the additional interaction W cannot increase or decrease our ignorance. If all systems S were equally probable to start with, the application of W must leave them equally probable. Invariance of the ensemble E_1 under the transformations (21) is a reasonable mathematical idealization of the hypothetical "state of total ignorance."

It remains only to justify on physical grounds the choice of the basic space T_1 of symmetric unitary matrices. Here, alternative choices are possible, and will be discussed in the next section. The choice of T_1 has the same motivation as Wigner's choice of real symmetric matrices for his ensemble E_G. Symmetric unitary matrices are physically appropriate under two alternative conditions, (i) if the systems are invariant under time inversion and under space rotations, or (ii) if the systems are invariant under time inversion and contain an even number of half-integer spin particles. The symmetry of the S matrix for systems satisfying condition (i) has been proved in a particularly simple way by Coester.[8] In applying the theory to neutron capture resonances, conditions (i) will always hold, and so the ensemble E_1 is the one to use.

[8] F. Coester, Phys. Rev. **89**, 619 (1953).

III. TIME-REVERSAL SYMMETRY. SYMPLECTIC ENSEMBLE

To find out whether the orthogonal ensemble is a reasonable one to use under all circumstances, a more careful analysis must be made of the consequences of time-reversal invariance. It will turn out that under some (perhaps not very realistic) circumstances a quite different ensemble should be used. The new ensemble will be called symplectic, because it bears the same relation to the symplectic group as E_1 bears to the orthogonal group.

We begin by recapitulating the basic notions of time-reversal invariance.[9] The operation of time reversal applied to a state ψ is defined by

$$T\psi=K\psi^c, \tag{25}$$

where ψ^c is the complex conjugate of ψ, and K is a constant unitary matrix. The operation of time reversal applied to a matrix A is defined by

$$A^R=KA^TK^{-1}. \tag{26}$$

A is called self-dual if $A^R=A$. A physical system is invariant under time reversal if the Hamiltonian is self-dual, i.e., if

$$H^R=H. \tag{27}$$

When Eq. (27) is satisfied, any unitary matrix S which is a function of H, for example the S given by Eq. (2), will also be self-dual,

$$S^R=S. \tag{28}$$

When the representation of states is transformed by a unitary transformation $\psi\rightarrow U\psi$, the K matrix transforms according to

$$K\rightarrow UKU^T. \tag{29}$$

So far the operation of time reversal has been purely formal, and the matrix K is quite arbitrary. Physical definiteness is given to the operation by requiring

$$J^R=-J, \tag{30}$$

where J is any component of the total angular momentum operator. It is not assumed that angular momentum is necessarily conserved. However, it is always true that the system has either integer spin or half-odd-integer spin. That is to say, the eigenvalues of components of J are either all integers or all half-odd integers, and the two possibilities do not mix. For brevity we call these two possibilities the even-spin case and the odd-spin case, respectively. The consequence of Eq. (30) is that in the even-spin case

$$T^2=KK^c=1, \tag{31}$$

and K is a symmetric unitary matrix, while in the odd-

[9] E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (English translated edition, Academic Press, Inc., New York, 1959), Chap. 26.

spin case

$$T^2=KK^c=-1, \tag{32}$$

and K is antisymmetric and unitary. Strictly speaking, Eqs. (31) and (32) need hold only for systems not possessing super-selection rules.[10] It will be assumed that this condition is satisfied by all the systems which we discuss.

Suppose now that the even-spin case holds and Eq. (31) is valid. Then a unitary operator U exists such that

$$K=UU^T. \tag{33}$$

By Eq. (29), the transformation $\psi\rightarrow U^{-1}\psi$ performed on the states ψ brings K to unity. Thus in the even-spin case the representation of states can always be chosen so that

$$K=1. \tag{34}$$

After one such representation is found, further transformations $\psi\rightarrow R\psi$ are allowed only with R a real orthogonal matrix, so that Eq. (34) remains valid. The consequence of Eq. (34) is that self-dual matrices are symmetric. In the even-spin case, every system invariant under time reversal will be associated, if the representation of states is suitably chosen, with a symmetric unitary matrix S. For even-spin systems with time-reversal invariance, the orthogonal ensemble E_1 is always appropriate.

Suppose next that we are dealing with a system invariant under space rotations. The spin may now be either even or odd. The matrix S representing the system commutes with every component of J. If we use the standard representation of the J matrices with J_1 and J_3 real and J_2 imaginary, the conditions (30) may be satisfied by the usual choice

$$K=\exp[i\pi J_2] \tag{35}$$

for K. With this choice of K, S and K commute and S^R reduces to S^T. Thus a rotation-invariant system is represented by a symmetric unitary S. The ensemble E_1 is in this case again appropriate.

In the case of rotational symmetry, the matrices S do not couple together states of different total angular momentum. A separate ensemble E_1 must be introduced for each value of J. Levels belonging to different J values belong to different ensembles and are statistically uncorrelated. A similar remark applies if there are other conserved quantities in the problem, for example parity or isotopic spin. In such cases the known integrals of the motion must first be eliminated, and the ensemble E_1 applied separately to each of the resulting uncorrelated series of levels.

For the remainder of this section we shall discuss the situation to which the ensemble E_1 does not apply, a system having odd spin, invariance under time reversal, but no rotational symmetry. In this case Eq. (32) holds, K cannot be made diagonal by any transformation of the form (29), and there is no integral of the motion by means of which the double-valuedness of the time-reversal operation can be trivially eliminated.

Every antisymmetric unitary operator can be reduced by a transformation (29) to the standard form

$$Z=\begin{bmatrix} 0 & -1 & 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 & & & & \\ 0 & 0 & 0 & -1 & & & & \\ 0 & 0 & 1 & 0 & & & & \\ \cdot & & & & \cdot & & & \\ \cdot & & & & & \cdot & & \\ \cdot & & & & & & \cdot & \\ \cdot & & & & & & & \cdot \end{bmatrix}, \tag{36}$$

consisting of (2×2) blocks

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

along the leading diagonal, all other elements of Z being zero. We suppose the representation of states chosen so that K is reduced to this form. The number of rows and columns of all matrices must now be even, and it is convenient to denote this number by $2N$ instead of by N. After one representation of states is chosen for which $K=Z$, further transformations $\psi\rightarrow B\psi$ are allowed only with B a unitary $(2N\times2N)$ matrix for which

$$Z=BZB^T. \tag{37}$$

Such matrices B form precisely the N-dimensional symplectic group,[11] usually denoted by $Sp(N)$.

It is well known[12] that the algebra of the symplectic group can be expressed most naturally in terms of quaternions. We therefore introduce the standard quaternion notation for (2×2) matrices

$$\tau^1=\begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix},\quad \tau^2=\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},\quad \tau^3=\begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, \tag{38}$$

with the multiplication table

$$(\tau^1)^2=(\tau^2)^2=(\tau^3)^2=-1, \tag{39}$$

$$\tau^1\tau^2=-\tau^2\tau^1=\tau^3,\quad \tau^2\tau^3=-\tau^3\tau^2=\tau^1,\quad \tau^3\tau^1=-\tau^1\tau^3=\tau^2. \tag{40}$$

Note that here and always i will be the ordinary imaginary unit, not a quaternion unit. All the $(2N\times2N)$ matrices will be considered as cut into N^2 blocks of (2×2), and each (2×2) block regarded as a quaternion.

[10] G. C. Wick, A. S. Wightman, and E. P. Wigner, Phys. Rev. 88, 101 (1952).

[11] H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, New Jersey, 1946), 2nd Ed., Chap. 6.

[12] C. Chevalley, *Theory of Lie Groups* (Princeton University Press, Princeton, New Jersey, 1946), pp. 18–24. J. Dieudonné, Ergeb. d. Math. 5, (1955).

In general a $(2N\times 2N)$ matrix with complex elements becomes an $(N\times N)$ matrix with complex quaternion elements. In particular, the matrix Z is now

$$Z=\tau_2 I, \tag{41}$$

where I is the $(N\times N)$ unit matrix. It is easy to verify that the rules of matrix multiplication are not changed by this transcription.

We call a quaternion "real" if it is of the form

$$q=q^0+(q\cdot\tau), \tag{42}$$

with real coefficients q^0, q^1, q^2, q^3. Thus a real quaternion does not correspond to a (2×2) matrix with real elements. Any complex quaternion q has a "conjugate quaternion"

$$\bar{q}=q^0-(q\cdot\tau), \tag{43}$$

which is distinct from its "complex conjugate,"

$$q^c=q^{0c}+(q^c\cdot\tau). \tag{44}$$

A quaternion with $q=q^c$ is real; one with $q=\bar{q}$ is a scalar. Applying both types of conjugation together, we obtain the "Hermitian conjugate"

$$q^+=\bar{q}^c=q^{0c}-(q^c\cdot\tau). \tag{45}$$

Now consider a general $(2N\times 2N)$ matrix A which is to be written as an $(N\times N)$ matrix Q with quaternion elements q_{ij}; $i, j=1, \cdots, N$. The standard matrix operations on A then reflect themselves on Q in the following way: Transposition,

$$(Q^T)_{ij}=-\tau^2\bar{q}_{ji}\tau^2, \tag{46}$$

Hermitian conjugation,

$$(Q^+)_{ij}=q_{ji}{}^+, \tag{47}$$

Time reversal,

$$(Q^R)_{ij}=\bar{q}_{ji}. \tag{48}$$

The usefulness of the quaternion algebra is a consequence of the simplicity of the relations (47) and (48). In particular it is noteworthy that the time-reversal operator K does not appear explicitly in Eq. (48) as it did in Eq. (26). By Eqs. (47) and (48), the condition

$$Q^R=Q^+ \tag{49}$$

is necessary and sufficient for the elements of Q to be real quaternions. When Eq. (49) holds we call Q "quaternion real."

A unitary matrix B satisfying Eq. (37) is automatically quaternion real. In fact it satisfies the conditions

$$B^R=B^+=B^{-1}, \tag{50}$$

which define the symplectic group. The matrices S representing physical systems are not quaternion real. They are unitary and self-dual, that is

$$S^R=S, \quad S^+=S^{-1}. \tag{51}$$

We now require a theorem of quaternion algebra.[13]

Theorem 2. Let H be any Hermitian quaternion-real $(N\times N)$ matrix. Then there exists a symplectic matrix B such that

$$H=B^{-1}DB, \tag{52}$$

where D is diagonal, real, and scalar.

The fact that D is scalar means that it consists of N blocks of the form

$$\begin{bmatrix} D_j & 0 \\ 0 & D_j \end{bmatrix}.$$

Thus the eigenvalues of H consist of N equal pairs. The Hamiltonian of any system which is invariant under time reversal and has odd spin satisfies the conditions of Theorem 2. All energy levels of such a system must be doubly degenerate. This is the Kramers degeneracy,[14] and Theorem 2 shows how it appears naturally in the quaternion language.

An immediate extension of Theorem 2 states that if S_1 and S_2 are two commuting Hermitian quaternion-real matrices, there exists a symplectic matrix B such that

$$S_1=B^{-1}D_1B, \quad S_2=B^{-1}D_2B, \tag{53}$$

with D_1 and D_2 both diagonal, real, and scalar. From this extension we can deduce

Theorem 3. Let S be any unitary self-dual $(N\times N)$ quaternion matrix. Then there exists a symplectic matrix B such that

$$S=B^{-1}EB, \tag{54}$$

where E is diagonal and scalar. The diagonal elements of E are N complex numbers $[exp(i\theta_j)]$ on the unit circle, each repeated twice.

To prove Theorem 3, we write

$$S=S_1+iS_2, \tag{55}$$

where S_1 and S_2 are quaternion real. The operation of time reversal applied to a matrix does not involve complex conjugation. Therefore, when S is self-dual, each of S_1 and S_2 must be separately self-dual. Being self-dual and quaternion real, S_1 and S_2 are also Hermitian. Moreover, since S is unitary,

$$S^+S=(S_1-iS_2)(S_1+iS_2)=1. \tag{56}$$

Separating quaternion real and quaternion imaginary parts in Eq. (56), we find

$$S_1{}^2+S_2{}^2=1, \quad S_1S_2-S_2S_1=0. \tag{57}$$

The S_1 and S_2 commute, and the extension of Theorem 2

[13] This theorem is presumably well known to the experts, but we are unable to find a reference to it in the mathematical literature. A nonrigorous "physicist's proof" of it is given in Appendix A of this paper.

[14] H. A. Kramers, Proc. Acad. Sci. Amsterdam 33, 959 (1930).

applies. Let then B be chosen to satisfy Eq. (53). Equation (54) will hold, with

$$E=D_1+iD_2$$

diagonal and scalar. If d_j, d_j' are corresponding eigenvalues of D_1 and D_2, Eq. (57) gives

$$d_j^2+(d_j')^2=1. \tag{58}$$

Hence we may write

$$d_j=\cos\theta_j, \quad d_j'=\sin\theta_j, \tag{59}$$

and the diagonal elements of E become

$$e_j=d_j+id_j'=\exp(i\theta_j), \tag{60}$$

each repeated twice.

It is convenient to state at this point the analog of Theorem 3 for the even-spin case. This is

Theorem 4. Let S be any unitary symmetric $(N\times N)$ matrix. Then there exists a real orthogonal matrix R such that

$$S=R^{-1}ER, \tag{61}$$

where E is diagonal. The diagonal elements of E are N complex numbers $[\exp(i\theta_j)]$ on the unit circle.

The proof of Theorem 4 is word for word the same as that of Theorem 3, only substituting "symmetric" for "self-dual," "real" for "quaternion real," "orthogonal" for "symplectic." This parallelism between the odd-spin and even-spin cases will always be maintained.

Now we return to the odd-spin case and define the symplectic ensemble E_4, the odd-spin analog of the orthogonal ensemble E_1. We work in the space T_4 of unitary self-dual quaternion matrices. The problem is again to define an invariant measure in T_4, in spite of the fact that the matrices of T_4 do not form a group.

Every matrix S in T_4 can be written in the form

$$S=U^RU, \tag{62}$$

with U unitary. To see that this is possible, observe that in the old prequaternion notations (SZ) is an antisymmetric unitary matrix and can be reduced to the canonical form

$$SZ=VZV^T \tag{63}$$

with V unitary; substituting $(UZ)^T$ for V then gives Eq. (62). For given S, the unitary matrix U in Eq. (62) is undetermined precisely to the extent of a transformation

$$U\rightarrow BU, \tag{64}$$

where B is an arbitrary symplectic matrix; the proof of this statement is word for word the same as that of Eq. (20). An infinitesimal neighborhood of S in T_4 is given by

$$S+dS=U^R[1+idM]U, \tag{65}$$

where dM is a quaternion-real self-dual infinitesimal matrix with elements

$$dM_{ij}=dM_{ij}^0+(dM_{ij}\cdot\tau). \tag{66}$$

The real coefficients dM_{ij}^α satisfy

$$dM_{ij}^0=dM_{ji}^0, \quad dM_{ij}^\alpha=-dM_{ji}^\alpha \quad \text{for} \quad \alpha=1,\,2,\,3. \tag{67}$$

There are $(2N^2-N)$ independent real variables dM_{ij}^α, and they are supposed to vary through some small intervals of lengths $d\mu_{ij}^\alpha$. The neighborhood of S thus defined has the measure

$$\mu(dS)=\prod_{\alpha,i,j} d\mu_{ij}^\alpha. \tag{68}$$

In terms of this measure, the symplectic ensemble E_4 is defined exactly like E_1. *The statistical weight of the neighborhood dS in E_4 is*

$$P(dS)=(V_4)^{-1}\mu(dS), \tag{69}$$

where V_4 is the total volume of the space T_4.

We can now repeat almost without change the arguments of Sec. II. We must first prove that the measure $\mu(dS)$ is independent of the choice of U in Eq. (62). This involves showing that the Jacobian

$$J=\det|\partial dM_{ij}'^\alpha/\partial dM_{kl}^\beta| \tag{70}$$

has absolute value unity, where

$$dM'=BdMB^{-1} \tag{71}$$

and B is symplectic. As before, it is enough to consider the case

$$B=I+A, \tag{72}$$

where A is infinitesimal, quaternion real, and anti-self-dual. Let the quaternion coefficients of A be a_{ij}^α, with $\alpha=0,\,1,\,2,\,3$. The diagonal elements of the matrix $(\partial dM_{ij}'^\alpha/\partial M_{kl}^\beta)$ can only involve the coefficients a_{ij}^0, and these occur just like the a_{ij} in Eq. (18). The conclusion that $J=1$ to first order in A follows as before.

The analog of Theorem 1 is

Theorem 5. The symplectic ensemble E_4 is uniquely defined, in the space T_4 of self-dual unitary quaternion matrices, by the property of being invariant under every automorphism

$$S\rightarrow W^RSW \tag{73}$$

of T_4 into itself, where W is any unitary matrix.

Theorem 5 can be proved by following word for word the proof of Theorem 1. Theorem 5 shows that the symplectic ensemble uniquely represents the notion of "uniform *a priori* probability" in the space T_4.

IV. SYSTEMS WITHOUT TIME-REVERSAL SYMMETRY. UNITARY ENSEMBLE

For completeness we briefly discuss a much simpler ensemble, the unitary ensemble E_2, which would apply to systems without invariance under time reversal.

Such systems are in principle easily created, for example by putting an ordinary atom or nucleus into an externally generated magnetic field. However, for the unitary ensemble to be applicable, the splitting of the levels by the magnetic field must be at least as large as the average level spacing in the absence of the field. The magnetic interaction must in fact be so strong that it completely "mixes up" the level structure which would exist in zero field. Such a state of affairs could never occur in nuclear physics; in atomic or molecular physics a practical application of the unitary ensemble may perhaps be possible.

A system without invariance under time reversal has a Hamiltonian which may be an arbitrary Hermitian matrix, not restricted to be symmetric or self-dual. We represent the system by an $(N\times N)$ unitary matrix S belonging to the space T_2 of all unitary matrices. It is now a trivial matter to define a uniform ensemble E_2 in T_2, because the space T_2 is simply the unitary group $U(N)$, and an invariant group measure in $U(N)$ is already provided.[15]

The formal definition of E_2 is as follows. A neighborhood of S in T_2 is given by

$$S+dS=U(1+idH)V, \tag{74}$$

where U, V are any two unitary matrices such that $S=UV$, while dH is an infinitesimal Hermitian matrix with elements $dH_{ij}=dH_{ij}^1+idH_{ij}^2$. The components dH_{ij}^1, dH_{ij}^2, in number N^2, vary independently through small intervals of length $d\mu_{ij}^1$, $d\mu_{ij}^2$. The invariant group measure $\mu(dS)$ is defined by

$$\mu(dS)=\prod_{i,j} d\mu_{ij}^1 d\mu_{ij}^2, \tag{75}$$

and is independent of the choice of U and V. The ensemble E_2 gives to each neighborhood dS the statistical weight

$$P(dS)=(V_2)^{-1}\mu(dS), \tag{76}$$

where V_2 is the volume of the space T_2.

The invariance property of E_2 analogous to Theorems 1 and 5 is stated in

Theorem 6. The unitary ensemble E_2 is uniquely defined, in the space T_2 of unitary matrices, by the property of being invariant under every automorphism

$$S\rightarrow USW \tag{77}$$

of T_2 into itself, where U, W are any two matrices of T_2.

This theorem merely expresses the fact that $\mu(dS)$ is the invariant group-measure on $U(N)$.

V. CALCULATION OF THE JOINT EIGENVALUE DISTRIBUTIONS

The joint distribution of the eigenvalues $(\epsilon_1,\cdots,\epsilon_N)$ of the Hamiltonian in the Gaussian ensemble was derived by Porter and Rosenzweig.[4] Their result is the following

Theorem 7. In the Gaussian ensemble defined by Eq. (1), the probability for finding an eigenvalue in each of the intervals $[\epsilon_j, \epsilon_j+d\epsilon_j]$, $j=1, \cdots, N$, is given by $P_N(\epsilon_1,\cdots,\epsilon_N)d\epsilon_1\cdots d\epsilon_N$, where

$$P_N(\epsilon_1,\cdots,\epsilon_N)=K_N\{\prod_{i<j}|\epsilon_i-\epsilon_j|\}\exp(-\sum_j \epsilon_j^2/4a^2), \tag{78}$$

and K_N is a constant.

Following the same method of proof, we shall obtain the corresponding formulas for the joint distribution function of the eigenvalues $[\exp(i\theta_j)]$ in the orthogonal, unitary and symplectic ensembles.

Theorem 8. In the ensemble E_β, the probability for finding eigenvalues $[exp(i\varphi_j)]$ of S with an angle φ_j in each of the intervals $[\theta_j, \theta_j+d\theta_j]$, $j=1, \cdots, N$, is given by $Q_{N\beta}(\theta_1,\cdots,\theta_N)d\theta_1\cdots d\theta_N$, where

$$Q_{N\beta}(\theta_1,\cdots,\theta_N)=C_{N\beta}\prod_{i<j}|e^{i\theta_i}-e^{i\theta_j}|^\beta. \tag{79}$$

Here $\beta=1$ for the orthogonal, $\beta=2$ for the unitary, and $\beta=4$ for the symplectic ensemble.

This theorem explains the choice of the notation E_β and T_β for the three ensembles and their corresponding spaces.

The unitary case ($\beta=2$) of Theorem 8 is a well-known result in the theory of the unitary group.[16] It will be enough for us to prove the theorem in detail for $\beta=1$ and to indicate the necessary modifications in the proof for $\beta=2$, 4.

Let $\beta=1$. By Theorem 4, every S in T_1 may be diagonalized into the form

$$S=R^{-1}ER, \tag{80}$$

with R orthogonal. We wish now to express the measure $\mu(dS)$ in terms of measures $\mu(dE)$, $\mu(dR)$ defined on the matrices E, R separately. Small neighborhoods of E and R are given by

$$dE=iEd\theta, \tag{81}$$

$$dR=dA.R. \tag{82}$$

Here $d\theta$ means the diagonal matrix with elements $[d\theta_1,\cdots,d\theta_N]$, and dA is a real antisymmetric infinitesimal matrix with elements dA_{ij}. We define

$$\mu(dE)=\prod_j d\theta_j, \tag{83}$$

$$\mu(dR)=\prod_{i<j} dA_{ij}, \tag{84}$$

the latter being the invariant group measure in the orthogonal group $O(N)$.

The measure $\mu(dS)$ is defined by Eq. (6), where dM is

[15] H. Weyl, reference 11, p. 188.

[16] H. Weyl, reference 11, p. 197, Theorem 7.4C.

given by Eq. (5) and U is any unitary matrix satisfying Eq. (4). The relation between dM, $d\theta$, and dA is given by Eqs. (5), (80), (81), and (82) and is

$$iRU^{T}dMUR^{-1}=iEd\theta+EdA-dAE. \tag{85}$$

Since E is a diagonal unitary matrix it has a square root F with elements $[\exp(\frac{1}{2}i\theta_j)]$. There is an ambiguity of a factor ± 1 in each element of F; it does not matter how these signs are chosen. A convenient choice for U satisfying Eq. (4) is then

$$U=FR, \tag{86}$$

by virtue of Eq. (80). With this choice of U, Eq. (85) reduces to

$$dM=d\theta-i[FdAF^{-1}-F^{-1}dAF]. \tag{87}$$

The last equation is a separate equation for each component dM_{ij}, namely,

$$dM_{jj}=d\theta_j, \tag{88}$$

$$dM_{ij}=2\sin[\tfrac{1}{2}(\theta_i-\theta_j)]dA_{ij}, \quad i\neq j. \tag{89}$$

Assembling the definitions (6), (83), and (84), we deduce from Eqs. (88) and (89)

$$\begin{aligned}\mu(dS)&=\{\prod_{i<j}|2\sin\tfrac{1}{2}(\theta_i-\theta_j)|\}\mu(dE)\mu(dR)\\&=\{\prod_{i<j}|e^{i\theta_i}-e^{i\theta_j}|\}\mu(dE)\mu(dR).\end{aligned} \tag{90}$$

If now the angles $[\theta_1,\cdots,\theta_N]$ are held fixed, and Eq. (90) is integrated with respect to dR over the whole orthogonal group $O(N)$, the result is Eq. (79) with $\beta=1$. Theorem 1 is thus proved in the orthogonal case.

Next let $\beta=2$. In this case R in Eq. (80) is unitary, and dA in Eq. (82) is anti-Hermitian, while dM is Hermitian. The equations (87) and (88) hold as before, but now Eq. (89) holds separately for the real and imaginary parts of each nondiagonal dM_{ij}, these being independent variables. The diagonal elements dA_{jj} are pure imaginary and do not appear in Eq. (89). Their absence reflects the fact that S is unchanged by the substitution

$$R\rightarrow GR \tag{91}$$

in Eq. (80), if G is any diagonal unitary matrix. The relation between measures analogous to Eq. (90) then becomes

$$\mu(dS)\mu(dG)=\{\prod_{i<j}|e^{i\theta_i}-e^{i\theta_j}|^{2}\}\mu(dE)\mu(dR). \tag{92}$$

Here $\mu(dS)$ and $\mu(dR)$ are invariant measures on $U(N)$, $\mu(dE)$ is defined by Eq. (83), and

$$\mu(dG)=\textstyle\prod_j d\eta_j, \tag{93}$$

where G is the diagonal matrix with elements $[\exp(i\eta_j)]$. The step from Eq. (92) to the theorem goes as before.

Lastly let $\beta=4$. In this case, by Theorem 3, Eq. (80) holds with R symplectic. A neighborhood of R in the symplectic group $Sp(N)$ is given by Eq. (82), where dA is now an anti-Hermitian quaternion-real infinitesimal matrix. The components of dA are $dA_{ij}{}^{\alpha}$, which are independent real variables, antisymmetric in (i,j) for $\alpha=0$, symmetric in (i,j) for $\alpha=1, 2, 3$. The total number of the $dA_{ij}{}^{\alpha}$ is $(2N^2+N)$. The invariant measure on the symplectic group is

$$\mu(dR)=\prod_{i,j,\alpha} dA_{ij}{}^{\alpha}. \tag{94}$$

The measure $\mu(dS)$ is given by Eq. (68), with dM given by Eqs. (62) and (65). The matrix dM is Hermitian and quaternion real, and has $(2N^2-N)$ independent components according to Eq. (67). The algebra leading up to Eq. (87) goes exactly as before. Equation (88) still holds, the diagonal elements dM_{jj} being real scalar quaternions with only one independent component. Also Eq. (89) holds, for the nondiagonal elements, separately in each of the four quaternion components $\alpha=0, 1, 2, 3$.

There are now $3N$ diagonal components $dA_{jj}{}^{\alpha}$, $\alpha=1, 2, 3$, which do not appear in Eq. (89). Their absence reflects the fact that S is unchanged by the substitution (91) in Eq. (80), where G may now be an arbitrary diagonal matrix whose elements are real unimodular quaternions. Let g_j be the diagonal elements of G, satisfying

$$\bar{g}_j g_j=1. \tag{95}$$

A neighborhood of G is defined by writing

$$dg_j=(d\eta_j)g_j, \tag{96}$$

where $d\eta_j$ is a pure vector quaternion, quaternion real and anti-Hermitian. There are $3N$ independent components $d\eta_j{}^{\alpha}$, and a measure in the space of G is defined by writing

$$\mu(dG)=\prod_{\alpha,j} d\eta_j{}^{\alpha}. \tag{97}$$

A comparison of Eq. (82) with Eq. (96) gives

$$d\eta_j{}^{\alpha}=dA_{jj}{}^{\alpha}, \quad \alpha=1, 2, 3. \tag{98}$$

Multiplying together the N equations (88), the $(2N^2-2N)$ equations (89), and the $3N$ equations (98), we obtain the relation

$$\mu(dS)\mu(dG)=\{\prod_{i<j}|e^{i\theta_i}-e^{i\theta_j}|^{4}\}\mu(dE)\mu(dR), \tag{99}$$

which establishes Theorem 8 for $\beta=4$.

The most obvious physical consequence of Theorem 8 is the so-called "repulsion of energy levels." The probability of finding an unusually small separation $\Delta=(\theta_i-\theta_j)$ between two levels tends to zero with Δ like Δ^{β}. According to Theorem 7 this phenomenon occurs also in the Gaussian model (where effectively $\beta=1$), a fact which was well known to Wigner and others. What

is new and unexpected in Theorem 8 is that the level repulsion is so drastically stronger in the symplectic ensemble, going with Δ^4 instead of Δ. Qualitatively speaking, one may say that the presence of the Kramers degeneracy makes any additional accidental degeneracy enormously more unlikely.

It is possible to understand the different powers β which appear in Eq. (79) by a simple mathematical argument based on counting dimensions. The dimension of the space T_1 is $\frac{1}{2}(N^2+N)$, while the dimension of the subspace T_1' composed of matrices in T_1 with two equal eigenvalues is $[\frac{1}{2}(N^2+N)-2]$. The difference in dimension, being 2 instead of 1, accounts for a factor in Eq. (79) linear in Δ. Similarly, when $\beta=2$, the dimension of T_2 is N^2 while that of T_2' is $[N^2-3]$. When $\beta=4$, the dimension of T_4 is $[2N^2-N]$ while that of T_4' is $[2N^2-N-5]$.

VI. ELECTROSTATIC ANALOG

Consider an infinitely thin circular conducting wire of radius 1. Let N unit charges be free to move on the wire, the positions of the charges being identified by angular variables $[\theta_1,\cdots,\theta_N]$. The universe is supposed to be two-dimensional, being merely the plane in which the circle lies. The charges repel each other, according to the Coulomb law of two-dimensional electrostatics, with a potential energy

$$W=-\sum_{i<j}\ln|e^{i\theta_i}-e^{i\theta_j}|. \tag{100}$$

We shall study the statistical mechanics of this Coulomb gas, considered as a classical system.

In classical statistical mechanics the velocity distribution of the charges is trivial and can be separated from the position distribution. We shall simply discard the velocity-dependent factors and their (easily calculable) contributions to the thermodynamics of the problem. The probability distribution of the angles $[\theta_1,\cdots,\theta_N]$, when the Coulomb gas is in thermal equilibrium at temperature T, is then given by

$$Q_{N\beta}(\theta_1,\cdots,\theta_N)=C_{N\beta}\exp[-\beta W], \tag{101}$$

with

$$\beta=1/T. \tag{102}$$

The nontrivial contributions to the thermodynamic variables are to be calculated from the positional partition-function

$$\Psi_N(\beta)=(2\pi)^{-N}\int\cdots\int_0^{2\pi}\exp[-\beta W]d\theta_1\cdots d\theta_N. \tag{103}$$

The reader will probably have observed that the distributions (79) and (101) are identical. We have therefore established

Theorem 9. There is a precise mathematical identity between the distribution of eigenvalues of a random matrix S and the distribution of positions of charges in a finite Coulomb gas at a finite temperature T. When S is taken from the orthogonal ensemble, the unitary ensemble, or the symplectic ensemble, the corresponding temperature of the gas is $T=1$, $T=\frac{1}{2}$, or $T=\frac{1}{4}$, respectively.

The beauty of Theorem 9 is that it shows the expression "repulsion of energy levels" to be more than an empty phrase. Energy levels do indeed behave exactly as if they were like charges, repelling each other with a force varying inversely with the first power of the distance. Another consequence of Theorem 9 is that the thermodynamic notions of entropy, specific heat, etc., can be transferred from the Coulomb gas to the eigenvalue series. This will prove very useful, as it gives us a precise and well-understood language in which to describe the statistical properties of the eigenvalue series.

It may seem strange that the temperature in Theorem 9 is a dimensionless quantity. The reason for this is that we have chosen the magnitude of the charges to be unity. In two-dimensional electrostatics, the dimensions of charge are [energy]$^{\frac{1}{2}}$. If the charges had been taken to be equal to e, then the temperatures in Theorem 9 would be $T=e^2$, $T=\frac{1}{2}e^2$, $T=\frac{1}{4}e^2$, giving T the dimensions of energy.

VII. CALCULATION OF THE PARTITION FUNCTION

In this section we use a simple group-theoretical argument to evaluate the partition function

$$\Psi_N(\beta)=(2\pi)^{-N}\int\cdots\int_0^{2\pi}\{\prod_{i<j}|e^{i\theta_i}-e^{i\theta_j}|^\beta\}\times d\theta_1\cdots d\theta_N, \tag{104}$$

for the physical values 1, 2, and 4 of β. The result can be checked easily by a direct calculation in the case $\beta=2$; for $\beta=1$ a direct calculation is very difficult but possible, using the methods of Mehta[5]; for $\beta=4$ no method of direct calculation has yet been found.

The method of procedure is to integrate over all the variables dS, dG, dE, and dR in Eqs. (90), (92), and (99). The relation (80) shows that S is fixed when E and R are given. However, when S is given, E and R are still subject to freedom of choice in two respects. (i) The order of the N angles θ_j may be permuted in any of $(N!)$ ways, provided that the rows of R are simultaneously permuted in the same way. (ii) R may be multiplied on the left by the diagonal matrix G, each of whose N diagonal elements may be chosen independently. The elements of G belong to a space $\sum_\beta$ which is geometrically the surface of a sphere of unit radius in a space of β dimensions. When $\beta=1$, $\sum_\beta$ consists of the two points ±1 only. When $\beta=2$, $\sum_\beta$ consists of the complex numbers with unit modulus. When $\beta=4$, $\sum_\beta$ consists of the unimodular real quaternions.

The integration of Eqs. (90), (91), and (99) then gives

$$N!V_\beta[S_\beta]^N=(2\pi)^N\Psi_N(\beta)\Omega_\beta. \tag{105}$$

Here V_β is the volume of the ensemble space T_β, defined by the measure $\int\mu(dS)$. S_β is the surface area of the β-dimensional sphere $\sum_\beta$, namely,

$$S_\beta=[2\cdot\pi^{\beta/2}/\Gamma(\tfrac{1}{2}\beta)]. \tag{106}$$

In particular, when $\beta=1$, $S_\beta=2$, which is correct since $\int\mu(dG)$ then reduces to a summation over N independent choices of ±1. Finally,

$$\Omega_\beta=\int\mu(dR) \tag{107}$$

is the volume of the orthogonal group $O(N)$ for $\beta=1$, of the unitary group $U(N)$ for $\beta=2$, and of the symplectic group $Sp(N)$ for $\beta=4$. The value of Ω_β is

$$\Omega_\beta=S_{N\beta}S_{(N-1)\beta}\cdots S_\beta, \tag{108}$$

with S_b given by Eq. (106). To prove Eq. (108), consider for example Ω_4. The general matrix in $Sp(N)$ consists of N vectors each having N quaternion components; these vectors are of unit length and pair-wise orthogonal in the quaternion sense. We can choose the first vector to be any set of N quaternions such that the sum of the squares of their $4N$ real coefficients is unity. The first vector then is free to move on a $4N$-dimensional sphere of measure S_{4N}. The remaining $(N-1)$ vectors are all perpendicular to the first and form a symplectic matrix in $Sp(N-1)$. Therefore

$$\Omega_4(N)=S_{4N}\Omega_4(N-1), \tag{109}$$

and this proves Eq. (108) for $\beta=4$. The proof for $\beta=1, 2$ is the same. Note that here $O(N)$ is the full orthogonal group, including reflections, so that the last factor $S_1=2$ occurs correctly in Eq. (108) when $\beta=1$.

It remains only to determine the V_β. Take first $\beta=1$ and go back to the definition of $\mu(dS)$ by Eqs. (4)–(6). Let us define a neighborhood of U in the unitary group $U(N)$ by

$$dU=idHU, \tag{110}$$

with dH Hermitian. Writing $dH_{ij}=dH_{ij}^1+idH_{ij}^2$, the measure $\mu(dU)$ is given by Eq. (75). Assembling Eqs. (4), (5), and (110), we find

$$dM_{ij}=dH_{ij}+dH_{ji}=2dH_{ij}^1. \tag{111}$$

The antisymmetric components dH_{ij}^2 of dH do not appear in Eq. (111). In fact, when S is fixed, the matrix U is undetermined by a transformation (20) with R orthogonal, and the measure of a neighborhood of R is given precisely by

$$dR=-dH^2R, \tag{112}$$

$$\mu(dR)=\prod_{i<j}dH_{ij}^2. \tag{113}$$

Assembling Eqs. (6), (75), and (113), we find

$$\mu(dS)\mu(dR)=2^{\frac{1}{2}N(N+1)}\mu(dU). \tag{114}$$

Integrating this over all the variables gives the desired result,

$$V_1=2^{\frac{1}{2}N(N+1)}[\Omega_2/\Omega_1]. \tag{115}$$

In the case $\beta=2$ the evaluation of V_β is trivial, since T_2 is $U(N)$ and therefore

$$V_2=\Omega_2. \tag{116}$$

Lastly let $\beta=4$. The measure $\mu(dS)$ in T_4 is defined by Eqs. (62), (65), (66), and (68), where U now belongs to the group $U_q(N)$ of unitary $(N\times N)$ matrices with complex quaternion elements. A neighborhood of U in $U_q(N)$ is given by Eq. (110), with dH a Hermitian quaternion matrix. An element of dH has the form

$$dH_{ij}=dH_{ij}^{10}+idH_{ij}^{20}+\sum_{\alpha=1}^{3}(dH_{ij}^{1\alpha}+idH_{ij}^{2\alpha})\tau^\alpha, \tag{117}$$

with eight independent real coefficients $dH_{ij}^{k\alpha}$. The measure in $U_q(N)$ is given by

$$\mu(dU)=\prod_{i,j,k,\alpha}dH_{ij}^{k\alpha}. \tag{118}$$

Assembling Eqs. (62), (65), and (110) gives

$$dM_{ij}^\alpha=2dH_{ij}^{1\alpha},\quad \alpha=0, 1, 2, 3, \tag{119}$$

the components $dH_{ij}^{2\alpha}$ again not appearing. When S is fixed, U is undetermined by a symplectic transformation (64), and the measure of a neighborhood of B in the symplectic group $Sp(N)$ is given by

$$dB=-dH^2B, \tag{120}$$

$$\mu(dB)=\prod_{i,j,\alpha}dH_{ij}^{2\alpha}. \tag{121}$$

Since the number of equations (119) is $(2N^2-N)$, Eqs. (68), (118), and (121) give

$$\mu(dS)\mu(dB)=2^{N(2N-1)}\mu(dU). \tag{122}$$

Integrating this over all variables gives

$$V_4=2^{N(2N-1)}[\Omega_8/\Omega_4], \tag{123}$$

where Ω_8 is the volume of $U_q(N)$, and Ω_4 that of $Sp(N)$.

Although Eq. (108) holds for $\beta=1, 2, 4$, it does not hold for $\beta=8$, since the complex quaternions do not form a division algebra. Instead, $U_q(N)$ is merely a different parametrization of the ordinary unitary group $U(2N)$. The group measure in $U(2N)$ is given by

$$dU=idKU,\quad dK=dK^1+idK^2, \tag{124}$$

$$\mu(dU)=\prod_{i,j}dK_{ij}^1dK_{ij}^2, \tag{125}$$

where dK^1 and dK^2 are real $(2N\times 2N)$ matrices, symmetric and antisymmetric, respectively. The matrices dH and dK must be identical, only dH is expressed in quaternion components by Eq. (117) while dK is expressed in ordinary matrix components by Eq. (124). The relation between the $dH_{ij}{}^{k\alpha}$ and the $dK_{ij}{}^{k}$ is then the following. A complex quaternion (117), which may be written for brevity

$$a^1+ia^2+(b^1+ib^2)\tau^1+(c^1+ic^2)\tau^2+(d^1+id^2)\tau^3, \quad (126)$$

appears in the dK matrix as a (2×2) block

$$\begin{bmatrix} a^1+d^2+i(a^2-d^1) & b^2-c^1+i(-b^1-c^2) \\ b^2+c^1+i(c^2-b^1) & a^1-d^2+i(a^2+d^1) \end{bmatrix}. \quad (127)$$

For each nondiagonal dH_{ij}, all eight components are present in Eqs. (126) and (127). For the diagonal elements dH_{jj}, $a^2=b^1=c^1=d^1=0$ and only a^1, b^2, c^2, and d^2 survive.

The Jacobian of the transformation from the components a^1, b^1, c^1, d^1, a^2, b^2, c^2, d^2 to the linear combinations appearing in Eq. (127) is 2^4 for each of $\frac{1}{2}(N^2-N)$ nondiagonal dH_{ij}, and 2 for each of N diagonal dH_{jj}. Hence the Jacobian of the total transformation from the $dH_{ij}{}^{k\alpha}$ to the $dK_{ij}{}^{k}$ is

$$J=\det\left|\partial dK_{mn}{}^{p}/\partial dH_{ij}{}^{k\alpha}\right|=2^{N(2N-1)}. \quad (128)$$

According to Eqs. (118) and (125), the volume of $U(2N)$ is just J times the volume of $U_q(N)$. Therefore, Eq. (123) reduces to

$$V_4=[\Omega_2(2N)/\Omega_4(N)]. \quad (129)$$

It is now only a matter of simple arithmetic to compute $\Psi_N(\beta)$ from Eqs. (105), (106), (108), (115), (116), and (129). We find

$$\Psi_N(1)=N!\pi^{-N}2^{\frac{1}{2}N(N+1)}[\Omega_2/\Omega_1{}^2]$$
$$=\Gamma(1+\tfrac{1}{2}N)/[\Gamma(\tfrac{3}{2})]^N, \quad (130)$$

$$\Psi_N(2)=N!, \quad (131)$$

$$\Psi_N(4)=N!\pi^N\{\Omega_2(2N)/[\Omega_4(N)]^2\}=2^{-N}(2N)!. \quad (132)$$

These results lead to the following general statement:

Conjecture A. For every integer N and real or complex β, we have identically

$$\Psi_N(\beta)=\Gamma(1+\tfrac{1}{2}N\beta)/[\Gamma(1+\tfrac{1}{2}\beta)]^N. \quad (133)$$

The evidence for the truth of this conjecture is overwhelmingly strong. We have proved it for $\beta=1$, 2, or 4, and for many other special cases to be described in the following section. But a general proof is still lacking. The failure of our strenuous efforts to find a proof has led us to surmise that some novel and interesting mathematics is lurking behind this innocent-looking identity.

VIII. MATHEMATICAL STATUS OF THE CONJECTURE

In this section we marshal the mathematical evidence in favor of conjecture A.

We first examine the analytic behavior of $\Psi_N(\beta)$ as a function of the complex variable β. It is easy to verify (and physically intuitive) that the maximum value of the quantity

$$y=\prod_{i<j}|e^{i\theta_i}-e^{i\theta_j}| \quad (134)$$

is attained when the points $e^{i\theta_j}$ are arranged at the vertices of a regular N-sided polygon, and that the maximum is equal to

$$Y=N^{\frac{1}{2}N}. \quad (135)$$

Therefore

$$\Psi_N(\beta)=\int_0^Y P(y)y^\beta dy, \quad (136)$$

where $P(y)$ is a positive weight function. In other words, $\Psi_N(\beta)$ is a moment-function defined on a finite interval.[17] Such a function must possess very special analytic properties. It must be analytic in the half-plane $(\mathrm{Re}\beta>0)$, and it must satisfy there the inequality

$$|\Psi_N(\beta)|<C|Y^\beta|. \quad (137)$$

Now the function

$$\psi_N(\beta)=\Gamma(1+\tfrac{1}{2}N\beta)/[\Gamma(1+\tfrac{1}{2}\beta)]^N \quad (138)$$

certainly satisfies these conditions. It has singularities only on the negative real axis, and its asymptotic behavior for large $|\beta|$ is

$$\psi_N(\beta)\sim N^{\frac{1}{2}}(\pi\beta)^{-\frac{1}{2}(N-1)}Y^\beta. \quad (139)$$

The function

$$\Delta_N(\beta)=Y^{-\beta}[\Psi_N(\beta)-\psi_N(\beta)] \quad (140)$$

is thus regular and bounded in $\mathrm{Re}\beta>0$.

Now a theorem of Carlson[18] states:

Carlson's Theorem. If a function of β is regular and bounded in $\mathrm{Re}\beta>0$, and if it is zero for $\beta=2, 4, 6, \cdots$, then it is identically zero.

Applying this theorem to the function $\Delta_N(\beta)$, we deduce that conjecture A, for any fixed value of N, must hold identically in β if it holds for every even integer $\beta=2k$. For $\beta=2k$, the integrand in Eq. (104) reduces to a finite polynomial in the variables $z_j=\exp(i\theta_j)$. We have thus proved that conjecture A is equivalent to the following purely algebraic statement:

---

[17] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (The American Mathematical Society, Providence, Rhode Island, 1943), p. 8.

[18] See E. C. Titchmarsh, *Theory of Functions* (Oxford University Press, Oxford, England, 1939), 2nd Ed., p. 186.

Conjecture B. For positive integer values of N and k, the coefficient of

$$[z_1 z_2 \cdots z_N]^{(N-1)k} \tag{141}$$

in the polynomial

$$\prod_{i \neq j} [z_i - z_j]^k \tag{142}$$

is equal to

$$(Nk)!/[k!]^N. \tag{143}$$

Note that each pair (i,j) in the product (142) is counted twice. This way of writing the product eliminates an unaesthetic minus sign from Eq. (143).

This algebraic form of the conjecture looks so simple that it ought to be provable by elementary combinatorial methods. However, the illusion of simplicity is quickly dispelled if one looks at the previous history of the problem.

When $N=1$ or 2, the conjecture is indeed trivial. So far as we have been able to discover, the only nontrivial case of the conjecture that has been previously known is the case $N=3$. The case $N=3$ appears, in heavily disguised form, as Eq. VI,(3) in the first letter[19] written by Ramanujan to Hardy in 1913, the letter which resulted in Ramanujan being discovered as a mathematical genius. Like all Ramanujan's statements, this one is very far from being trivial; however, it was not new in 1913. An equivalent form of conjecture B for $N=3$ is the identity

$$\sum_{j=-k}^{k} (-1)^j \binom{2k}{k+j}^3 = \frac{(3k)!}{(k!)^3}, \tag{144}$$

where

$$\binom{2k}{k+j}$$

is a binomial coefficient. In this form the statement was first proposed as a conjecture by Morley,[20] and was proved by Dixon[21] in 1891. Subsequently, Dixon[22] found and proved a natural generalization of Eq. (144), namely

$$\sum_j (-1)^j \binom{a+b}{a+j}\binom{b+c}{b+j}\binom{c+a}{c+j} = \frac{(a+b+c)!}{a!b!c!}, \tag{145}$$

valid for any positive integers a, b, and c. As is often the case in such problems, an inductive proof of Eq. (145) is easier than a direct proof of the special case (144).

In trying to deal with the case of general N by algebraic methods, one is led to the following generalized conjecture:

Conjecture C. Let $(a_1,a_2,\cdots,a_N)$ be any set of N positive integers. Then the constant term in the expansion of the polynomial

$$\prod_{i \neq j} \left[1 - \frac{z_j}{z_i}\right]^{a_i} \tag{146}$$

in powers of $(z_1,\cdots,z_N)$ is

$$(a_1+\cdots+a_N)!/[a_1!a_2!\cdots a_N!]. \tag{147}$$

In the case $N=3$, conjecture C reduces to Eq. (145) and is thus known to be correct. For general N, conjecture C reduces to conjecture B when $a_1=a_2=\cdots=a_N=k$.

The evidence in favor of conjecture C is again overwhelming. We have succeeded in proving it (and hence also conjectures A and B) for the cases $N=4$ and 5. The proof for $N=4$ is given in Appendix B of this paper. It is based on the "evergreen proof," a combination of the principles of induction and symmetry, first invented in a different connection by Dougall.[23] The same method works, with greater complications of detail, when $N=5$. Beyond $N=5$, these algebraic devices seem to fail utterly.

To summarize the evidence for conjecture A, it is known to be true for all β if it is true for $\beta=2k$; it is known to be true for all N with $\beta=1$, 2, or 4; and it is known to be true for all β with $N=1$, 2, 3, 4, or 5. The simplest case in which it could conceivably be false is $N=\beta=6$, and even to test it in this special case would require a prohibitive amount of numerical computation.

In the next section we shall find some additional independent evidence. Namely, conjecture A makes definite predictions concerning the thermodynamic behavior of the Coulomb gas in both the high-temperature limit $\beta \to 0$ and the low-temperature limit $\beta \to \infty$. These predictions can be checked against perturbation-theory expansions in powers of β and β^{-1}, respectively. In all cases which have been examined, the agreement is exact.

Note added in proof. Conjecture C has now been proved by Dr. Kenneth Wilson of Harvard University and by Dr. J. Gunson of the University of Birmingham, England. Wilson's proof will shortly be published in this journal. Gunson's proof is essentially the same as Wilson's, but was found independently.

IX. PHYSICAL CONSEQUENCES OF THE CONJECTURE

Conjecture A specifies precisely the statistical properties of a finite Coulomb gas of N particles. For physical applications, and in particular for describing the statistical properties of eigenvalues of complex systems, we are only interested in the limit $N \to \infty$. We study in

[19] S. Ramanujan, *Collected Papers* (Cambridge University Press, Cambridge, England, 1927), p. 26 of the Introductory Notice. The same equation appears as Eq. (1.1) in G. H. Hardy, *Ramanujan* (Cambridge University Press, Cambridge, England, 1940), p. 7.

[20] By private communication. See, also, F. Morley, Proc. London Math. Soc. 34, 397 (1902).

[21] A. C. Dixon, Messenger of Math. 20, 79 (1891).

[22] A. C. Dixon, Proc. London Math. Soc. 35, 285 (1903).

[23] J. Dougall, Proc. Edinburgh Math. Soc. 25, 114 (1907).

TABLE I. Values of the thermodynamic quantities F, U, S, and C as functions of $\beta=T^{-1}$. γ is Euler's constant.

	$\beta\to 0$	$\beta=1$	$\beta=2$	$\beta=4$	$\beta\to\infty$
F	$\frac{1}{2}(1-\gamma-\ln(\frac{1}{2}\beta))+(\pi^2/48)\beta$	$\frac{1}{2}(1+\ln\frac{1}{2}\pi)=0.726$	$\frac{1}{2}=0.500$	$\frac{1}{2}-\frac{1}{4}\ln 2=0.327$	$(1/2\beta)\ln(\pi\beta)+(1/6\beta^2)$
U	$\frac{1}{2}(-\gamma-\ln(\frac{1}{2}\beta))+(\pi^2/24)\beta$	$1-\frac{1}{2}(\gamma+\ln 2)=0.365$	$\frac{1}{2}(1-\gamma)=0.211$	$\frac{3}{4}-\frac{1}{2}(\gamma+\ln 2)=0.115$	$1/2\beta-1/6\beta^2$
S	$-\frac{1}{2}\beta+(\pi^2/48)\beta^2$	$\frac{1}{2}(1-\gamma-\ln\pi)=-0.361$	$-\gamma=-0.577$	$1-2\gamma-\ln 2=-0.848$	$\frac{1}{2}(1-\ln(\pi\beta))-1/3\beta$
C	$\frac{1}{2}\beta-(\pi^2/24)\beta^2$	$\frac{3}{2}-\frac{1}{8}\pi^2=0.266$	$2-\frac{1}{6}\pi^2=0.355$	$7-\frac{2}{3}\pi^2=0.420$	$\frac{1}{2}-1/3\beta$

this section the thermodynamics of the infinite Coulomb gas, or equivalently the statistics of an infinitely long series of eigenvalues.

The partition function (104) is normalized so that the potential energy of the gas is zero at infinite temperature ($\beta=0$). The potential energy at zero temperature ($\beta=\infty$) is then the ground-state energy

$$W_0=-\ln Y=-\tfrac{1}{2}N\ln N. \tag{148}$$

In order to obtain well-defined limits for the thermodynamic variables as $N\to\infty$, we must first change the zero of energy to the position W_0. The gas has then by definition zero energy at zero temperature, and a positive energy at any positive temperature. The partition function defined on the new energy-scale is

$$\Phi_N(\beta)=Y^{-\beta}\Psi_N(\beta). \tag{149}$$

The free energy per particle $F_N(\beta)$ is

$$F_N(\beta)=-[\beta N]^{-1}\ln\Phi_N(\beta). \tag{150}$$

Taking the limit $N\to\infty$ in conjecture A, we deduce

Conjecture D. As $N\to\infty$, the free energy per particle of the Coulomb gas at temperature $T=\beta^{-1}$ tends to the limiting value

$$F(\beta)=\beta^{-1}L(\tfrac{1}{2}\beta)+\tfrac{1}{2}[1-\ln(\tfrac{1}{2}\beta)], \tag{151}$$

$$L(z)=\ln\Gamma(1+z). \tag{152}$$

In what follows we shall always assume that conjecture D is correct.

From Eq. (151) the values of the other thermodynamic quantities follow. These are:

Energy per particle:

$$U=F+\beta(\partial F/\partial\beta)=\tfrac{1}{2}[L'(\tfrac{1}{2}\beta)-\ln(\tfrac{1}{2}\beta)]. \tag{153}$$

Entropy per particle:

$$S=\beta^2(\partial F/\partial\beta)=\tfrac{1}{2}\beta[L'(\tfrac{1}{2}\beta)-1]-L(\tfrac{1}{2}\beta). \tag{154}$$

Specific heat per particle:

$$C=-\beta^2(\partial U/\partial\beta)=-\tfrac{1}{4}\beta^2L''(\tfrac{1}{2}\beta)+\tfrac{1}{2}\beta. \tag{155}$$

Note that although Eq. (151) has been rigorously proved for $\beta=1, 2, 4$, the same is not true of Eqs. (153), (154), and (155). These last equations depend on the validity of conjecture D for general β. However, in the case $\beta=2$, the first two derivatives of $F(\beta)$ can be directly computed. This has been done, and the results agree with Eqs. (153), (154), and (155) at $\beta=2$. So we have yet another independent check on conjecture D. The values of the thermodynamic functions for physically interesting values of β are summarized in Table I.

There follow some miscellaneous remarks concerning the interpretation of these results.

A. Physical Nature of the Coulomb Gas

The thermodynamic functions are analytic over the whole range from $\beta=0$ to $\beta=\infty$. The Coulomb gas is a single-phase system with no thermodynamic transition at any finite temperature. In a later paper we will prove that the system possesses a long-range order of crystalline type at all temperatures. Thus it might be appropriate to call it a "crystal" rather than a gas. In a one-dimensional system the distinction between crystal and gas is somewhat arbitrary.

At low temperatures ($\beta\to\infty$) the charges are regularly spaced in a crystalline lattice arrangement, and the thermal excitations are compressional waves of small amplitude (phonons) running through the lattice. As the temperature is raised, the local disorder becomes greater, although some degree of long-range lattice structure always persists. At high temperatures ($\beta\to 0$) we can define a Debye length Λ, with the property that all charge fluctuations are neutralized by correlated motions of other charges within a distance of the order of Λ. The system then behaves approximately like a gas of independent particles, each particle carrying with it a neutralizing "charge cloud" of size Λ. The energy U is the electrostatic energy of a particle interacting with its induced charge cloud. Since U is normalized to be zero at low temperatures, when Λ is equal to the level spacing

$$\Delta=2\pi/N, \tag{156}$$

we may define Λ in general by the equation

$$U=\tfrac{1}{2}\ln(\Lambda/\Delta). \tag{157}$$

The factor $\frac{1}{2}$ appears because the interaction includes the self-energy of the induced charge itself. At high temperatures we have then from column 1 of the table

$$\Lambda=2e^{-\gamma}[\Delta/\beta]\sim\Delta T. \tag{158}$$

In other words, the induced charge cloud is spread out over about $(2T)$ neighboring particles.

For $\beta=1$, the value of chief interest in applications to the eigenvalue problem, the Debye length is only of the order of one level spacing. In this case the notion of

a Debye length hardly applies, and all violations of charge neutrality involving more than one or two particles are highly improbable. The long-range regularity of the eigenvalues is extremely rigid, and the eigenvalue series looks qualitatively more like a "wobbly crystal" than a classical plasma. The same remark applies even more strongly to the cases $\beta=2$ and 4.

B. Entropy as a Measure of Information Content

The entropy S provides us with a quantitative and exact notion of the "degree of irregularity" of an arrangement of atoms. Because of the existence of the analogy between Coulomb gas and eigenvalue series, the same quantity S gives a precise measure of the degree of irregularity of a long sequence of eigenvalues. It is appropriate here to use the language of information theory.[24]

A perfectly random sequence of N numbers, with mean spacing Δ and with values determined within some observational limit of accuracy δ, can carry a quantity of information

$$I_0=[1+\ln(\Delta/\delta)][N/\ln 2], \tag{159}$$

measured in the practical unit of binary digit or bit. A series of N eigenvalues taken from the statistical ensemble E_β can carry only a reduced amount of information

$$I(\beta)=I_0+S(\beta)[N/\ln 2]. \tag{160}$$

This loss of information content is a direct measure of the statistical regularity of the eigenvalue series. According to the numbers in the table, the loss of information is

0.521 bit per level in the even-spin case ($\beta=1$),
0.833 bit per level in the case without time-reversal symmetry ($\beta=2$),
1.223 bits per level in the odd-spin case ($\beta=4$).

It would be quite feasible to compute the entropy of an observed sequence of levels and see whether the result agreed with these numbers. However, for a practical test of the statistical model the quantities U and C would undoubtedly be more convenient.

C. Statistical Interpretation of U and C

Denoting by $\langle\ \rangle$ an ensemble-average, we have by Eqs. (103) and (149)

$$NU=\langle W-W_0\rangle, \tag{161}$$

$$NC=\langle(W-\langle W\rangle)^2\rangle, \tag{162}$$

where W is the electrostatic energy given by Eq. (100), and $W_0=-\frac{1}{2}N\ln N$ is the value of W for a uniformly spaced series. Thus U is, apart from normalization, the ensemble average of the logarithm of the geometric mean of all distances between pairs of eigenvalues. And C is the statistical mean-square fluctuation of the same quantity.

As a statistic for analyzing the properties of observed eigenvalue series, W seems to be the best expression to use. It has two great advantages over other possible statistics such as F and S; (i) W can be computed from the eigenvalue pair-correlation function alone, without analyzing higher order correlations, and (ii) the statistical uncertainty of W is known from the value of C.

We summarize the situation in the following

Theorem 10. Let $(z_1,\cdots,z_N)$ be the eigenvalues of a random matrix taken from one of the ensembles E_1, E_2, E_4. The statistic

$$W-W_0=\tfrac{1}{2}N\ln N-\sum_{i<j}\ln|z_i-z_j| \tag{163}$$

has the expectation value NU and the root-mean-square deviation $(NC)^{\frac{1}{2}}$, with the values of U and C given in the table above.

This theorem is unfortunately not yet adapated to practical use. In practice we never have a complete series of N eigenvalues all round the unit circle. We have a comparatively small number n of observed levels, which are supposed to be a small section of a complete eigenvalue series of order $N\gg n$. In order to analyze the statistics of the observed levels, it is necessary to work out in detail the predicted behavior of a section of n levels chosen at random from a matrix of one of the basic ensembles. We shall find that a statement substantially identical with Theorem 10 can be proved, with the summation in W restricted to the observed levels, and with NU, $(NC)^{\frac{1}{2}}$ replaced by nU, $(nC)^{\frac{1}{2}}$. A full discussion of this and other properties of partial level series will be given in later papers of the series.

ACKNOWLEDGMENTS

The author is grateful to E. P. Wigner for conversations which stimulated him to pursue this study. He is indebted to Wigner and to Dr. M. L. Mehta for copies of unpublished work, and to Dr. J. L. Rosen and Dr. H. Palevsky for discussions of the experimental material. He wishes also to thank the Brookhaven National Laboratory for its hospitality during the summer of 1961 when this paper was being written.

APPENDIX A

Proof That a Hermitian Self-Dual Matrix Can Be Diagonalized in Quaternion Algebra

Let H be a Hermitian self-dual $(N\times N)$ quaternion matrix. Let $(q_1,\cdots,q_N)$ be an N-component vector

[24] C. E. Shannon, Bell System Tech. J. **27**, 379 and 623 (1948). Reprinted in book form, C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois, Urbana, Illinois, 1949).

whose components are real quaternions. The expression

$$Q=\sum_{ij}\bar{q}_iH_{ij}q_j \tag{A1}$$

is a scalar (because H is self-dual) and is real (because H is Hermitian). The ratio

$$R=(Q/P),\quad P=\sum_i \bar{q}_iq_i, \tag{A2}$$

is a bounded real scalar function of the q_j. For some set of real quaternions

$$q_j=A_{j1}, \tag{A3}$$

R attains its maximum value D_1. The fact that R is stationary at the point (A3) implies

$$\sum_j H_{ij}A_{j1}=D_1A_{i1}. \tag{A4}$$

Proceeding in the same way through the successive stationary values of R, we find N real scalar quantities D_i and N real quaternion vectors A_{jk} such that

$$\sum_j H_{ij}A_{jk}=D_kA_{ik}. \tag{A5}$$

Writing B for the matrix A^{-1}, we have

$$H=B^{-1}DB, \tag{A6}$$

with D real, scalar, and diagonal, while B is quaternion real.

From Eq. (A6) we deduce

$$DBB^R=BB^RD. \tag{A7}$$

Thus BB^R either is diagonal, or (if several of the D_j are equal) can be chosen to be diagonal. Since the normalization of each vector B_{jk} is arbitrary, we can choose them all to have unit length. Then

$$BB^R=1, \tag{A8}$$

i.e., B is symplectic.

APPENDIX B

Proof of Conjecture C for $N=4$

Let the function $F(a_1,\cdots,a_N)$ of the non-negative integer variables $(a_1,\cdots,a_N)$ be defined by

$$F=\sum_{\lambda_{jk}}\prod_{j<k}\left\{(-1)^{\lambda_{jk}}\binom{a_j+a_k}{a_j+\lambda_{jk}}\right\}. \tag{B1}$$

The summation variables λ_{jk} are integers subject to the conditions

$$\lambda_{jk}=-\lambda_{kj},\quad j,k=1,\cdots,N, \tag{B2}$$

$$-a_j\leq\lambda_{jk}\leq a_k, \tag{B3}$$

$$\sum_k\lambda_{jk}=0,\quad j=1,\cdots,N. \tag{B4}$$

We also define $f(a_1,\cdots,a_N)$ by

$$f=[a_1+\cdots+a_N]!/[a_1!\cdots a_N!]. \tag{B5}$$

When all the factors in Eq. (146) are expanded by the binomial theorem, conjecture C reduces to the statement

$$F=f. \tag{B6}$$

Suppose next that $(a_1,\cdots,a_{N-1})$ are non-negative integers with

$$a_1+\cdots+a_{N-1}=m, \tag{B7}$$

while $x=a_N$ is free to be non-integral. A function $F(x)=F(a_1,\cdots,a_{N-1},x)$ can then be defined by Eq. (B1), the conditions (B3) with $j=N$ or $k=N$ being omitted. This $F(x)$ is a polynomial in x of degree m, and is therefore well defined for all positive or negative x. Moreover, when x is a non-negative integer the new definition reduces to the old one. Similarly,

$$f(x)=(m+x)!/[a_1!\cdots a_{N-1}!x!] \tag{B8}$$

has an obvious meaning as a polynomial in x of degree m. Conjecture C is thus equivalent to the statement that Eq. (B6) holds as an identity in x for non-negative integer values of $(a_1,\cdots,a_{N-1})$.

The essential step in the proof is the following Lemma which holds for all N.

Lemma 1. $F(x)$ has the symmetry property

$$F(x)=(-1)^mF(-m-1-x). \tag{B9}$$

Note: It is trivial that $f(x)$ has the same symmetry. To prove the lemma, we go back to Eq. (146). We can represent F by the contour integral

$$F=\frac{1}{(2\pi i)^N}\int\cdots\int dz_1\cdots dz_N\{\prod_{i\neq j}(z_i-z_j)^{a_j}\}\times\prod_j z_j^{-m-1-a_N+a_j}. \tag{B10}$$

So long as all the a_j are non-negative integers, the paths of integration can be chosen in any way provided that each variable z_j circles the origin once in the positive direction. When $a_N=x$ is allowed to be nonintegral, the contours for $(z_1,\cdots,z_{N-1})$ are still arbitrary, but the contour for z_N must be chosen to circle the origin inside all the other contours. This gives the correct value for F, since for $|z_N|\ll|z_j|$ all the nonterminating binomial series $(z_j-z_N)^{a_j+x}$ can be expanded in ascending powers of z_N as required by Eq. (B1).

Now make in Eq. (B10) the transformation of variables

$$z_N=-y_N, \tag{B11}$$

$$z_j=y_j-y_N,\quad j=1,\cdots,N-1. \tag{B12}$$

The rule for choosing the contours of integration is the same for the y_j as it was for the z_j. In terms of the variables y_j, the expression for F is identical with Eq. (B10) except for an over-all sign $(-1)^m$ and the replacement of x by $(-m-1-x)$. This proves Lemma 1.

The second Lemma also holds for all N.

Lemma 2. Let $0 \leq a_1 \leq a_2 < \cdots \leq a_{N-1}$. *Then* $F(x) = f(x) = 0$ *when* x *is any negative integer in the range*

$$-a_1 \leq x < 0. \tag{B13}$$

To prove Lemma 2, observe that $F(x)$ contains, in every term of the expansion (B1), a product of $(N-1)$ factors

$$\binom{a_j+x}{a_j+\lambda_{jN}}. \tag{B14}$$

A factor (B14) will vanish whenever x satisfies Eq. (B13) and

$$\lambda_{jN} > x. \tag{B15}$$

But in view of Eq. (B4), not all λ_{jN} can be negative, and so at least one factor (B14) will always vanish. This proves Lemma 2.

We now complete the proof of the conjecture for $N=4$, following the method of Dougall.[23] We suppose a_1, a_2, and a_3 to be non-negative integers with

$$0 \leq a_1 \leq a_2 \leq a_3, \quad a_1+a_2+a_3 = m. \tag{B16}$$

Let it be assumed as an inductive hypothesis that $F(x) = f(x)$ holds as an identity in x whenever $a_1+a_2+a_3 < m$. Because $F(a_1,a_2,a_3,a_4)$ is formally symmetric between a_3 and a_4, the inductive hypothesis implies

$$F(a_1,a_2,a_3,x) = f(a_1,a_2,a_3,x) \tag{B17}$$

for the integer values of x

$$x = 0, 1, \cdots, a_3-1. \tag{B18}$$

Lemma 2 states that Eq. (B17) holds for

$$x = -a_1, -a_1+1, \cdots, -1. \tag{B19}$$

Lemma 1 in combination with Eqs. (B18) and (B19) implies that Eq. (B17) holds for

$$x = -a_1-a_2-2a_3, \cdots, -a_2-a_3-1. \tag{B20}$$

The three ranges (B18), (B19), and (B20) do not overlap, and they contain altogether

$$2(a_1+a_3) \geq m+a_1 \tag{B21}$$

distinct values of x. But the two sides of Eq. (B17) are polynomials in x of degree m. If $a_1 > 0$, Eq. (B17) must hold as an identity in x. If $a_1 = 0$, Eq. (B17) reduces to the statement

$$F(a_2,a_3,x) = f(a_2,a_3,x), \tag{B22}$$

which is true since conjecture C has already been proved for $N=3$. Therefore Eq. (B17) always holds, and the conjecture is proved for $N=4$.

The same proof also applies in the case $N=3$, when instead of Eq. (B21) we obtain

$$2(a_1+a_2) = 2m > m \tag{B23}$$

distinct values of x.

When we try to extend the argument to $N=5$, we have $2(a_1+a_4)$ values of x for which the analog of Eq. (B17) is proved, and this is not necessarily greater than $m = a_1+a_2+a_3+a_4$. Two further steps are then required. (i) By using all the inequalities (B3) and (B4), we can strengthen Lemma 2 so that it holds for negative integers x down to $(-a_2)$ instead of $(-a_1)$. We then have instead of Eq. (B21)

$$2(a_2+a_4) \geq m, \tag{B24}$$

which is better but still not quite good enough. (ii) Equation (B17) can be directly verified for one more value of x, namely, $x = -a_1-1$. The argument for $N=5$ then just squeezes through. There is clearly no hope of obtaining by such piecemeal methods enough x values to deal with $N \geq 6$.

Statistical Theory of the Energy Levels of Complex Systems. II

FREEMAN J. DYSON
Institute for Advanced Study, Princeton, New Jersey
(Received September 15, 1961)

The distribution function of spacings S between nearest neighbors, in a long series of energy levels with average spacing D, is studied. The statistical properties of S are defined in terms of an ensemble of systems described in a previous paper. For large values of $t=(\pi S/2D)$, it is shown that the distribution of S can be deduced from the thermodynamical properties of a certain model. The model, which replaces the eigenvalue distribution by a continuous fluid, can be studied by the methods of classical electrostatics, potential theory, and thermodynamics. In this way the distribution function of spacings S is found to be asymptotically

$$Q(t)=At^{17/8}\exp[-\tfrac{1}{4}t^2-\tfrac{1}{2}t]$$

for large t. The numerical constant A can in principle not be determined from such a continuum model. Reasons are given for considering the remaining factors in the formula for $Q(t)$ to be reliable.

I. INTRODUCTION

IN the first paper of the present series,[1] a general theory was developed with the purpose of describing the statistical behavior of energy levels in complex systems. In this paper the theory will be applied to one of the classical problems of energy-level analysis, the study of the theoretical distribution law of spacings between nearest-neighbor levels.

The theory of level spacings was begun by Wigner[2] with a famous conjecture—that in a long series of levels with average spacing D, the proportion of spacings which lie between S and $(S+dS)$ is given by

$$Q_W(t)dt, \quad t=(\pi S/2D). \tag{1}$$

$$Q_W(t)=(2t/\pi)\exp(-t^2/\pi). \tag{2}$$

This formula was supposed to apply to a series of levels having the same values of all identifiable quantum numbers such as angular momentum and parity. It is very well supported by experimental data,[3] and by numerical tests[4] with random matrices of high order. However, it is now known to be false. Mehta and Gaudin[5] have obtained an analytic expression for the correct distribution function $Q(t)$ and have computed $Q(t)$ numerically. They find that $Q_W(t)$ is not identical with $Q(t)$ but is astonishingly close to it, the difference $|Q_W-Q|$ being less than 0.0162 over the whole range of t. For practical purposes, Wigner's intuition has been abundantly justified.

The analytic expressions of Mehta and Gaudin provide in principle a complete solution to the level-spacing distribution problem. Unfortunately, their formula for $Q(t)$ is very complicated, being obtained as the Fredholm determinant of a certain integral equation in which t occurs as a parameter. The formula can be used for numerical computation when S is of the order of D (Gaudin[5] has computed $Q(t)$ for $t\leq 5$), and it provides a series expansion of $Q(t)$ in ascending powers of t which is useful for $t\leq 1$. However, it gives no precise information about the behavior of $Q(t)$ for large t. Further analytic study might well yield an asymptotic formula for $Q(t)$ valid in the limit $t\to\infty$, but this remains to be demonstrated.

The purpose of this paper is to attack the problem of the large level spacings, for which the Gaudin-Mehta analysis has not yet proved useful, by an entirely different approach. The behavior of $Q(t)$ for large t will be deduced from arguments of a mathematically non-rigorous kind, based upon thermodynamical considerations. The results are undoubtedly correct in their main features and fill the existing gap in our knowledge of the spacing distribution. If it should later turn out that a rigorous and more exact calculation of $Q(t)$ for large t can be extracted from the Mehta-Gaudin analysis, then the results of this paper would be interesting for a different reason. A comparison of the two calculations would then show what are the limits within which thermodynamical arguments may be trusted and beyond which such arguments may be misleading.

II. CONTINUUM MODEL

According to Sec. V of (I), the joint distribution function of the eigenvalues $[\exp(i\theta_j)]$, $j=1, \cdots, N$, of a random unitary $(N\times N)$ matrix is

$$Q_{N\beta}(\theta_1, \cdots, \theta_N)=C_{N\beta}\exp[-\beta W], \tag{3}$$

$$W=-\sum_{i<j}\ln|e^{i\theta_i}-e^{i\theta_j}|. \tag{4}$$

Here $C_{N\beta}$ is a normalization constant; β is a parameter which under normal circumstances takes the value 1, but may also be equal to 2 or 4 under certain conditions described in (I). Equation (3) is identical with the distribution function of N point charges on the unit

[1] Freeman J. Dyson, J. Math. Phys. 3, 140 (1962). Quoted in what follows as (I).
[2] E. P. Wigner, Gatlinberg Conference on Neutron Physics, Oak Ridge National Laboratory Report ORNL 2309 (1957), p. 59.
[3] N. Rosenzweig and C. E. Porter, Phys. Rev. 120, 1698 (1960). See, also, reference 4.
[4] C. E. Porter and N. Rosenzweig, Suomalaisen Tiedeakat. Toimituksia A VI, No. 44 (1960); see especially Figs. 19 and 20.
[5] M. L. Mehta, Nuclear Phys. 18, 395 (1960); M. L. Mehta and M. Gaudin, *ibid.* 18, 420 (1960); M. Gaudin, *ibid.* 25, 447 (1961).

circle, repelling each other according to the laws of classical 2-dimensional electrostatics with the potential energy W, in thermodynamic equilibrium at temperature $T=\beta^{-1}$.

The main objective of this paper is to calculate the quantity $R_\beta(\alpha)$, defined as the probability that the angle $[-\alpha\leq\theta\leq\alpha]$ contains none of the θ_j. From Eq. (3) it follows that

$$R_\beta(\alpha)=[\Psi_{N\beta}(\alpha)/\Psi_{N\beta}(0)], \tag{5}$$

$$\Psi_{N\beta}(\alpha)=\int\cdots\int_{\alpha}^{2\pi-\alpha}\exp[-\beta W]d\theta_1\cdots d\theta_N. \tag{6}$$

Let now N become very large. Then $\Psi_{N\beta}(0)$ is the partition function of the Coulomb gas on the whole unit circle, while $\Psi_{N\beta}(\alpha)$ is the partition function of the same gas compressed into a circular arc of length $2(\pi-\alpha)$. In other words

$$R_\beta(\alpha)=\exp[\beta\{F_{N\beta}(0)-F_{N\beta}(\alpha)\}], \tag{7}$$

where $F_{N\beta}(\alpha)$ is the free energy of the Coulomb gas on the arc $2(\pi-\alpha)$.

We now make three assumptions for which no rigorous mathematical justification exists.

(i) There is a macroscopic density function $\sigma_\alpha(\theta)$, a continuous function of θ on the arc $[\alpha<\theta<2\pi-\alpha]$, such that $\sigma_\alpha(\theta)d\theta$ is the number of θ_j in the range $[\theta<\theta_j<\theta+d\theta]$.

(ii) For a given density function $\sigma_\alpha(\theta)$, the free energy of the gas is composed of two parts

$$F=V+F_1, \tag{8}$$

where V is the macroscopic Coulomb energy

$$V=-\tfrac{1}{2}\int\int_{\alpha}^{2\pi-\alpha}\sigma_\alpha(\theta)\sigma_\alpha(\varphi)\ln|e^{i\theta}-e^{i\varphi}|d\theta d\varphi, \tag{9}$$

and F_1 is a sum of contributions from the individual arcs $[\theta,\theta+d\theta]$ of the gas,

$$F_1=\int_{\alpha}^{2\pi-\alpha}\sigma_\alpha(\theta)f_\beta[\sigma_\alpha(\theta)]d\theta, \tag{10}$$

where $f_\beta(\sigma)$ is the free energy per particle in a Coulomb gas having *uniform* density σ on the whole unit circle.

(iii) The overwhelmingly dominant contribution to the integral (6) comes from configurations not deviating significantly from a particular macroscopic density-distribution $\sigma_\alpha(\theta)$, namely that function $\sigma_\alpha(\theta)$ which makes F given by Eqs. (8)–(10) a minimum subject to

$$\int_{\alpha}^{2\pi-\alpha}\sigma_\alpha(\theta)d\theta=N. \tag{11}$$

These assumptions (i)–(iii) can be summarized in the single statement that *for large N the Coulomb gas obeys the laws of classical thermodynamics*. The assumption (10) means that the free energy (apart from the macroscopic Coulomb energy) is an extensive property of the system, the free-energy density at any point being a function of the local temperature and density alone. To a physicist these assumptions are so hallowed by custom that they hardly require justification. Every application of thermodynamics to systems of strongly interacting atoms or molecules rests on assumptions of this kind. We make no effort here to explore more deeply the mathematics of the problem.

The "continuum model" of the Coulomb gas on the arc $[\alpha<\theta<2\pi-\alpha]$ is defined to be a classical compressible fluid of density $\sigma_\alpha(\theta)$ per unit angle, obeying the laws of classical thermodynamics. The total free energy $F_{N\beta}(\alpha)$ of the continuum model is defined to be the minimum value of F given by Eqs. (8)–(10), and the function $\sigma_\alpha(\theta)$ is determined by requiring that F be a minimum subject to Eq. (11).

It remains to specify the function $f_\beta(\sigma)$. Let

$$f_\beta(\sigma)=U_\beta(\sigma)-\beta^{-1}S_\beta(\sigma), \tag{12}$$

where U_β is the energy and S_β the entropy per particle in a uniform gas of

$$N'=2\pi\sigma \tag{13}$$

charges on the whole unit circle. According to Eq. (I, 163), the energy per particle is

$$U_\beta(\sigma)=-\tfrac{1}{2}\ln N'+U(\beta), \tag{14}$$

since we are now talking about the total Coulomb energy including the ground-state energy $[-\frac{1}{2}N'\ln N']$. The physical meaning of Eq. (14) is made clearer by remembering that

$$U_\beta(\sigma)=+\tfrac{1}{2}\ln\Lambda, \tag{15}$$

where Λ is a Debye length representing the size of the neutralizing charge cloud around each particle. Since the Debye length must vary inversely with N', the dependence of $U_\beta(\sigma)$ on σ can only have the simple form (14).

According to the calculations of Sec. (IX) of I, the entropy $S_\beta(\sigma)$ should be independent of N' for large N'. However, we here run into an interesting example of Gibbs' paradox.[6] The entropy has been calculated in I for a classical gas of N distinguishable particles. Gibbs' Paradox lies in the fact that entropy so defined is not an extensive quantity. To obtain an extensive quantity, one must subtract $[\ln N!]$ from the classical entropy, which is equivalent to treating the particles as undistinguishable. This means that in Eq. (12) one should use

$$S_\beta(\sigma)=\ln(N/N')+S(\beta), \tag{16}$$

with $S(\beta)$ given by Eq. (154) of I.

[6] E. Schrödinger, *Statistical Thermodynamics* (Cambridge University Press, New York, 1952), pp. 58–62.

Assembling Eqs. (8)–(10), (12), (14), (16), we find

$$\beta F = G_2 + G_1 + G_0, \tag{17}$$

$$G_2 = -\tfrac{1}{2}\beta(N/2\pi)^2 \int\int_{\alpha}^{2\pi-\alpha} \rho_\alpha(\theta)\rho_\alpha(\varphi) \times \ln|e^{i\theta}-e^{i\varphi}|\,d\theta d\varphi, \tag{18}$$

$$G_1 = (1-\tfrac{1}{2}\beta)(N/2\pi)\int_{\alpha}^{2\pi-\alpha} \rho_\alpha(\theta)\,\ln(\rho_\alpha(\theta))d\theta, \tag{19}$$

$$G_0 = \beta N[F(\beta) - \tfrac{1}{2}\ln N], \tag{20}$$

with

$$\rho_\alpha(\theta) = (2\pi/N)\sigma_\alpha(\theta), \tag{21}$$

$$\int_{\alpha}^{2\pi-\alpha} \rho_\alpha(\theta)d\theta = 2\pi, \tag{22}$$

and $F(\beta)$ given by Eq. (151) of I. The fact that the energy and entropy both contribute to G_1 a term in $(\rho \ln\rho)$ is due to the special form of the Coulomb potential. When $\beta=2$ (the case of the unitary ensemble) the term G_1 is absent and the model becomes particularly simple.

When $\alpha=0$, the equilibrium density is $\rho_0(\theta)=1$, which makes $G_2=G_1=0$. Therefore

$$\beta F_{N\beta}(0) = G_0 = \beta N[F(\beta) - \tfrac{1}{2}\ln N]. \tag{23}$$

Hence Eq. (7) gives

$$\ln R_\beta(\alpha) = -\min_\rho[G_2+G_1]. \tag{24}$$

The term G_0 is the only one in Eqs. (17)–(20) which depended on the detailed microstructure of the Coulomb gas, and it has disappeared from Eq. (24). Since our purpose in this paper is to compute $R_\beta(\alpha)$, we simply drop the constant G_0 and write

$$\beta F = G_2 + G_1, \tag{25}$$

both terms G_2 and G_1 being purely macroscopic in form.

The variational problem (24) is equivalent to the following set of equations.

$$\rho_\alpha(\theta) = A\,\exp[-\gamma V_\alpha(\theta)], \quad \alpha<\theta<2\pi-\alpha, \tag{26}$$

$$\gamma = [N\beta/\{\pi(2-\beta)\}], \tag{27}$$

$$V_\alpha(\theta) = -\int_{\alpha}^{2\pi-\alpha} \rho_\alpha(\varphi)\,\ln|e^{i\theta}-e^{i\varphi}|\,d\varphi. \tag{28}$$

This V_α is the electrostatic potential produced by the the charge-distribution ρ_α. Equation (26) is a "self-consistent field" type of equation, expressing the fact that the charge ρ_α is in thermal equilibrium in the potential V_α which it itself generates.

The only unexpected feature of these equations is that the "effective temperature" appearing in the exponent in Eq. (26) is not $T=\beta^{-1}$ but

$$T_e = \beta^{-1} - \tfrac{1}{2}. \tag{29}$$

When $\beta>2$, the effective temperature is negative, that is to say the charge has a statistical preference for regions of higher potential. The simple case of zero effective temperature occurs not at $\beta=\infty$ but at $\beta=2$. When $\beta=2$, we must replace Eq. (26) by

$$V_\alpha(\theta) = V_\alpha = \text{const}, \quad \alpha<\theta<2\pi-\alpha, \tag{30}$$

and the minimum problem reduces to a problem of classical electrostatics without any thermodynamics. In this case Eq. (24) gives

$$\ln R_2(\alpha) = (N^2/2\pi)V_\alpha. \tag{31}$$

III. SOLUTION FOR $\beta=2$

The continuum model for $\beta=2$ is defined by Eqs. (28) and (30). It consists of a charge-density $\rho_\alpha(\theta)$ distributed on a conducting wire which forms a circular arc of angle $2(\pi-\alpha)$. Since this is a standard problem of 2-dimensional potential theory, it can be immediately solved by the method of conformal mapping. Although the solution is well known, we reproduce the details of it here. The details will be needed in Sec. IV, when we go on to the more difficult case $\beta\neq 2$.

Let z be a complex variable representing points in the physical plane. The conducting wire consists of the curve

$$z = e^{i\theta}, \quad \alpha\leq\theta\leq 2\pi-\alpha, \tag{32}$$

lying in the z plane. The function

$$W_\alpha(z) = -\int_{\alpha}^{2\pi-\alpha} \rho_\alpha(\varphi)\,\ln(z-e^{i\varphi})d\varphi, \tag{33}$$

is analytic and many-valued in the whole z plane outside the wire. Its real part is one-valued, and by Eqs. (28), (30)

$$\mathrm{Re}W_\alpha(z) = V_\alpha, \quad z=e^{i\theta}, \quad \alpha\leq\theta\leq 2\pi-\alpha. \tag{34}$$

By Eq. (22),

$$W_\alpha(z) \sim 2\pi \ln z \quad \text{as} \quad |z| \to \infty. \tag{35}$$

The potential is completely determined by the statement that $W_\alpha(z)$ is analytic and satisfies Eqs. (34) and (35).

The charge density is related to W_α by

$$\rho_\alpha{}^{\pm}(\theta) = \frac{1}{2\pi}[\lim_{z\to e^{i\theta}} |\partial W_\alpha/\partial z|]. \tag{36}$$

The limit $z\to e^{i\theta}$ may be taken from the outside of the unit circle, giving $\rho_\alpha{}^{+}(\theta)$, or from the inside, giving

$\rho_\alpha^-(\theta)$. The charge ρ_α^+ may be thought of as localized on the outer surface of the wire and ρ_α^- as localized on the inner surface. The total charge density is given by

$$\rho_\alpha(\theta)=\rho_\alpha^+(\theta)+\rho_\alpha^-(\theta), \tag{37}$$

$$\rho_\alpha^\pm(\theta)=\tfrac{1}{2}[\rho_\alpha(\theta)\pm 1]. \tag{38}$$

A convenient series of mappings is the following

$$\zeta=\frac{1-z}{1+z}, \quad z=\frac{1-\zeta}{1+\zeta}. \tag{39}$$

$$w=\frac{\zeta}{(\zeta^2+\delta^2)^{\frac{1}{2}}+\delta}, \quad \zeta=\frac{2w\delta}{1-w^2} \tag{40}$$

$$u=\frac{1-ww_0}{w-w_0}, \quad w=\frac{1+uw_0}{u+w_0}. \tag{41}$$

Equation (39) maps the z plane onto the ζ plane with cuts along the imaginary axis from $(\pm i\delta)$ to $(\pm i\infty)$, where

$$\delta=\tan\tfrac{1}{2}\alpha. \tag{42}$$

Equation (40) maps the ζ plane onto the inside of the unit circle in the w plane, the point $z=\infty$, $\zeta=-1$ mapping onto $w=w_0$, with

$$w_0=-\tan\epsilon, \quad \epsilon=\tfrac{1}{4}(\pi-\alpha). \tag{43}$$

The end-points $z=\exp(\pm i\alpha)$ of the wire in the z-plane map onto the points $w=\pm i$. Equation (41) maps the inside of the unit circle in the w-plane onto the outside of the unit circle in the u plane. The point $z=\infty$ maps onto $u=\infty$ with the proportionality factor

$$u\sim z\sec(\tfrac{1}{2}\alpha), \quad |z|\to\infty. \tag{44}$$

The solution of the potential problem is simply

$$W_\alpha(z)=2\pi\ln[u\cos(\tfrac{1}{2}\alpha)]. \tag{45}$$

This satisfies Eqs. (34), (35) with

$$V_\alpha=2\pi\ln\cos(\tfrac{1}{2}\alpha). \tag{46}$$

The corresponding charge densities $\rho_\alpha^\pm(\theta)$ are given by Eq. (38) with

$$\rho_\alpha(\theta)=\sin(\tfrac{1}{2}\theta)[\sin^2(\tfrac{1}{2}\theta)-\sin^2(\tfrac{1}{2}\alpha)]^{-\frac{1}{2}}. \tag{47}$$

These results give immediately the asymptotic form of the spacing distribution in a series of energy levels with $\beta=2$. Let D be the average level-spacing, and

$$t=(\pi S/2D) \tag{48}$$

large compared with unity. The proportion of level spacings of size between t and $(t+dt)$ is $Q(t)dt$, where

$$Q(t)=\tfrac{1}{2}\pi d^2P/dt^2, \tag{49}$$

and $P(t)$ is the probability that a randomly chosen interval of length $(2tD/\pi)$ will be free of energy levels. By Eq. (5) we have

$$P(t)=R_2(\alpha), \quad t=\tfrac{1}{2}\alpha N. \tag{50}$$

Equations (31) and (46) then give in the limit as $N\to\infty$,

$$P(t)=\exp[N^2\ln\cos(t/N)]=\exp[-\tfrac{1}{2}t^2], \tag{51}$$

$$Q(t)\sim(\pi/2)t^2\exp[-\tfrac{1}{2}t^2], \quad \beta=2. \tag{52}$$

The continuum model predicts that the spacing distribution for $t\gg 1$ has the form (52). Unfortunately it is impossible to estimate by continuum-model calculations what the inherent errors of the model are likely to be. Therefore the degree of accuracy with which Eq. (52) holds is unknown. Clearly the discreteness of charge would make the distribution (47) wrong for angles θ within a range of about N^{-2} from the end point $\theta=\alpha$. Taking an optimistic view, one may conjecture that *the free energies of the continuum model and of a real Coulomb gas with discrete charges differ by an amount which remains bounded as* $t\to\infty$. One could hardly expect, in view of the unavoidable effect of discreteness at the end-points, that "remains bounded" could be replaced by "tends to zero" in this statement. The consequence of this conjecture is that, in all asymptotic formulas such as Eq. (52), the exponential factor and the power of t standing outside the exponential are probably correct, but the numerical coefficient is not to be taken seriously.

It is of some interest to compare the formula (52) with the result one would deduce from simple arguments of the kind which Wigner[2] used in making his conjecture Eq. (2). To derive Eq. (2), Wigner assumed

$$Q_W(t)=At\exp(-Bt^2), \tag{53}$$

and determined the constants A and B from the conditions

$$\int_0^\infty Q(t)dt=1, \tag{54}$$

$$\int_0^\infty Q(t)tdt=(\pi/2), \tag{55}$$

which must hold exactly for the correct distribution function $Q(t)$. The motivation for Eq. (53) came from three arguments: (i) "the repulsion of levels" is known to make the distribution linear in t for small t, (ii) the level repulsion should make the distribution approximately Gaussian for large t, and (iii) the formula should be as simple as possible. The resulting Eq. (2) was applicable to a level series with $\beta=1$, the case which normally occurs in experiments.

In the case $\beta=2$, which applies when time-reversal invariance is abandoned, the level-repulsion will make

$Q(t)$ quadratic in t for small t. The "Wigner conjecture" for this case would be

$$Q_W^2(t)=At^2\exp(-Bt^2), \tag{56}$$

with A and B still determined by Eqs. (54) and (55). Putting in the numerical values, Eq. (56) becomes

$$Q_W^2(t)=[256t^2/\pi^5]\exp[-16t^2/\pi^3]. \tag{57}$$

Comparing Eqs. (52) and (57), one sees that the Wigner conjecture is incorrect, but that the conjectured exponential factor differs from the true one only by the ratio $[32/\pi^3]$. It is probable that, as in the case $\beta=1$, the Wigner conjecture (57) lies numerically very close to the true distribution function over the whole range of t. An additional check on this point comes from the known form of the exact distribution function at small t. In Paper III of this series, we shall prove that

$$Q(t)\sim(8/3\pi)t^2,\quad t\ll 1,\quad \beta=2. \tag{58}$$

The Wigner conjecture (57) then differs from the true $Q(t)$ at small t by the ratio $[96/\pi^4]$. It is remarkable that an incorrect formula can come as near as this to the truth.

IV. SOLUTION FOR $\beta\neq 2$

When $\beta\neq 2$, the continuum model leads to the nonlinear equations (26), (28), and an analytic solution is not to be expected. The problem becomes tractable only after introducing some kind of perturbation-theory approximations. Since the objective is an asymptotic formula valid for large values of $t=\frac{1}{2}\alpha N$, the perturbation theory should if possible represent an expansion in inverse powers of (αN). Fortunately, the formulation of the problem by the minimum principle Eq. (24) makes such an expansion possible. According to Eqs. (18) and (19), G_2 is of order $(\alpha N)^2$ while G_1 is of order (αN). The "unperturbed system" can be taken to be given by G_2 alone, the addition of G_1 being the "perturbation."

The unperturbed system is, apart from the constant factor β, identical with the case $\beta=2$ considered in Sec. III. So the unperturbed free energy is given by Eqs. (25), (31), and (46) and is

$$\beta F_0=-\tfrac{1}{2}\beta N^2\ln\cos(\tfrac{1}{2}\alpha). \tag{59}$$

The unperturbed charge density will now be denoted by $\bar\rho_\alpha(\theta)$ and is given by Eq. (47).

Since Eq. (24) is a variation principle for the free energy, the first-order perturbation of βF is merely the value which G_1 takes with the unperturbed charge-distribution, namely

$$\beta F_1=(1-\tfrac{1}{2}\beta)(N/2\pi)\int_\alpha^{2\pi-\alpha}\bar\rho_\alpha(\theta)\,\ln\bar\rho_\alpha(\theta)d\theta. \tag{60}$$

To evaluate Eq. (60) it is convenient to transform the integral into the u plane. Since charge is invariant in a conformal mapping,

$$\int\bar\rho_\alpha(\theta)d\theta=\int|du|. \tag{61}$$

Hence Eqs. (47) and (60) give

$$\beta F_1=(1-\tfrac{1}{2}\beta)(N/2\pi)\times\int\ln\left|\frac{(u-\tan\epsilon)(u-\cot\epsilon)}{[u-\exp(\frac{1}{2}i\alpha)][u-\exp(-\frac{1}{2}i\alpha)]}\right||du|, \tag{62}$$

the integral being taken around the unit circle. Now

$$\int(\ln|u-a|)\,|du|=2\pi\max[\ln|a|,0], \tag{63}$$

this being the potential at the position a of a uniformly charged circle of radius 1. Therefore Eq. (62) gives

$$\beta F_1=(1-\tfrac{1}{2}\beta)N\ln\cot\epsilon=(1-\tfrac{1}{2}\beta)N\times\ln[\sec\tfrac{1}{2}\alpha+\tan\tfrac{1}{2}\alpha]. \tag{64}$$

In the limit $N\to\infty$, Eqs. (59) and (64) give

$$\beta F_0+\beta F_1=\tfrac{1}{4}\beta t^2+(1-\tfrac{1}{2}\beta)t, \tag{65}$$

as the first two terms of the desired expansion of the free energy in powers of t^{-1}.

The next term in the expansion will require second-order perturbation theory. The calculation becomes necessarily more complicated, but much of the pain can be avoided by working in the w plane defined by Eqs. (39), (40). On the unit circle of the w plane,

$$\bar\rho_\alpha(\theta)=\sec\tfrac{1}{2}\alpha\,|\sec\psi|,\quad w=\exp(i\psi), \tag{66}$$

$$\int\bar\rho_\alpha(\theta)d\theta=\int m(\psi)d\psi, \tag{67}$$

$$m(\psi)=|du/dw|=\sin\tfrac{1}{2}\alpha[1+\cos\tfrac{1}{2}\alpha\cos\psi]^{-1}. \tag{68}$$

The perturbed charge density is written in the form

$$\rho_\alpha(\theta)=\bar\rho_\alpha(\theta)+q(\psi)h(\psi), \tag{69}$$

$$q(\psi)=|dw/dz|=[\sec^2\tfrac{1}{2}\alpha-\cos^2\psi]/[2\tan\tfrac{1}{2}\alpha|\cos\psi|], \tag{70}$$

where $h(\psi)$ is the unknown perturbation and is supposed to be a small quantity. According to Eq. (40), each point $z=\exp(i\theta)$ is mapped onto two points $w=\exp(i\psi)$, $w=\exp[i(\pi-\psi)]$. The outside surface of the arc $[\alpha<\theta<2\pi-\alpha]$ in the z plane is mapped onto the left half of the unit circle $[\cos\psi<0]$ in the w plane, while the inside surface of the arc in the z plane is mapped onto the right half of the unit circle $[\cos\psi>0]$ in the w plane. Therefore Eq. (69) may be written

$$\rho_\alpha(\theta)=q(\psi)[m(\psi)+m(\pi-\psi)+h(\psi)], \tag{71}$$

and $h(\psi)$ must be regarded as an even function of $(\cos\psi)$. Since the total charge is not changed by the perturbation,

$$\int_0^{2\pi} h(\psi)d\psi=0. \tag{72}$$

It is now necessary to express the free energy (G_2+G_1) in terms of w-plane integrals. For G_1, Eqs. (19), (70), and (71) give

$$G_1=\tfrac{1}{2}(1-\tfrac{1}{2}\beta)(N/2\pi)\int_0^{2\pi}[M(\psi)+h(\psi)] \times\ln\{q(\psi)[M(\psi)+h(\psi)]\}d\psi, \tag{73}$$

$$M(\psi)=m(\psi)+m(\pi-\psi) =2\sin\tfrac{1}{2}\alpha[1-\cos^2(\tfrac{1}{2}\alpha)\cos^2\psi]^{-1}. \tag{74}$$

Expanding Eq. (73) to second order in $h(\psi)$, we find

$$G_1=\tfrac{1}{2}(1-\tfrac{1}{2}\beta)(N/2\pi)\int_0^{2\pi}[(M\ln\bar{\rho}_\alpha)+(h\ln\bar{\rho}_\alpha) +\tfrac{1}{2}M^{-1}h^2]d\psi. \tag{75}$$

The term independent of h is the first-order free energy given by Eqs. (60), (64). Thus to second order in $h(\psi)$

$$G_1=\beta F_1+\tfrac{1}{2}(1-\tfrac{1}{2}\beta)(N/2\pi)\int_0^{2\pi}[h(\psi)\ln|\sec\psi| +\tfrac{1}{4}\csc\tfrac{1}{2}\alpha(1-\cos^2\tfrac{1}{2}\alpha\cos^2\psi)h^2(\psi)]d\psi. \tag{76}$$

The transformation of G_2 into the w plane can be made without any approximation. Equation (18) may be written

$$G_2=+\tfrac{1}{2}\beta(N/2\pi)^2\int_\alpha^{2\pi-\alpha}\rho_\alpha(\theta)\,\mathrm{Re}[W_\alpha(e^{i\theta})]d\theta, \tag{77}$$

where $W_\alpha(z)$ is given by Eq. (33). If $\rho_\alpha(\varphi)$ in Eq. (33) is taken to be the unperturbed charge density $\bar{\rho}_\alpha(\varphi)$, then $W_\alpha(z)$ has the value given by Eq. (45). However, $\rho_\alpha(\varphi)$ is now defined by Eq. (69), and therefore

$$W_\alpha(z)=2\pi\ln[u\cos(\tfrac{1}{2}\alpha)]+Y(z), \tag{78}$$

$$Y(z)=-\int_\alpha^{2\pi-\alpha}q(\psi)h(\psi)\ln(z-e^{i\varphi})d\varphi \tag{79}$$

$$=-\tfrac{1}{2}\int_0^{2\pi}h(\psi)\ln(z-e^{i\varphi})d\psi. \tag{80}$$

The factor $(\frac{1}{2})$ in Eq. (80) appears because the ψ integration corresponds to the arc $[\alpha<\varphi<2\pi-\alpha]$ taken twice. Now this function $Y(z)$ is analytic in z and tends to zero as $z\to\infty$ by virtue of Eq. (72). It is, therefore, analytic in w inside the unit circle. According to Eqs. (36), (70), and (79),

$$\frac{1}{2\pi}\lim_{w\to e^{i\psi}}[\partial Y/\partial w]=\frac{1}{2\pi q(\psi)}\lim_{z\to e^{i\varphi}}[\partial Y/\partial z]=\tfrac{1}{2}h(\psi), \tag{81}$$

the derivatives being taken in the radial direction. Therefore the function

$$-\int_0^{2\pi}h(\psi)\ln(w-e^{i\psi})d\psi \tag{82}$$

has the same normal derivative as $Y(z)$ at every point of the unit circle in the w plane, and can differ from $Y(z)$ only by a constant. By virtue of Eq. (72) this implies

$$\int_\alpha^{2\pi-\alpha}q(\psi)h(\psi)\,\mathrm{Re}[Y(e^{i\theta})]d\theta =-\tfrac{1}{2}\int\int_0^{2\pi}h(\psi)h(w)\ln|e^{i\psi}-e^{iw}|d\psi dw. \tag{83}$$

This expresses in a simple form in the w plane the part of G_2 which is quadratic in $h(\psi)$. The term linear in $h(\psi)$ vanishes since the unperturbed charge distribution $\bar{\rho}_\alpha(\theta)$ was chosen so as to make G_2 stationary. The term independent of $h(\psi)$ is just βF_0 given by Eq. (59). Therefore Eq. (77) reduces to

$$G_2=\beta F_0-\tfrac{1}{4}\beta(N/2\pi)^2\int\int_0^{2\pi}h(\psi)h(w) \times\ln|e^{i\psi}-e^{iw}|d\psi dw. \tag{84}$$

The total free energy to order $h^2(\psi)$ is

$$\beta F=\beta F_0+\beta F_1+\beta F_2, \tag{85}$$

where βF_2 is the sum of the terms involving $h(\psi)$ in Eqs. (76) and (84).

To determine $h(\psi)$, the quadratic form βF_2 must be minimized. It is convenient to expand $h(\psi)$ in a Fourier series

$$h(\psi)=\sum_{n=1}^{\infty}h_n\cos(2n\psi). \tag{86}$$

The constant term is zero by Eq. (72), and the odd terms are zero since $h(\psi)$ is even in $(\cos\psi)$. Substituting Eq. (86) into Eqs. (76) and (84), and taking the limit $N\to\infty$, we find

$$\beta F_2=(1-\tfrac{1}{2}\beta)[\tfrac{1}{4}\textstyle\sum n^{-1}u_n+(32t)^{-1}\sum(u_n^2+u_nu_{n+1})] +(\beta/32)\textstyle\sum n^{-1}u_n^2, \tag{87}$$

with

$$u_n=(-1)^nNh_n. \tag{88}$$

For large values of t, the u_n which makes Eq. (87) a minimum will be a slowly varying function of n. With

negligible error we may replace the term $(u_n u_{n+1})$ by (u_n^2), and the minimization then becomes trivial. The result is

$$\beta F_2 = -\tfrac{1}{2}\beta^{-1}(1-\tfrac{1}{2}\beta)^2 \sum_{n=1}^{\infty} n^{-1}[1+(2n/\beta t) \times(1-\tfrac{1}{2}\beta)]^{-1}. \quad (89)$$

The series is convergent and gives for large t the asymptotic expression

$$\beta F_2 = -\tfrac{1}{2}\beta^{-1}(1-\tfrac{1}{2}\beta)^2\{\ln[\beta t/2(1-\tfrac{1}{2}\beta)]+\gamma\}, \quad (90)$$

where γ is Euler's constant.

As we observed in Sec. III, the continuum model cannot be expected to give the constant term in the free energy correctly as $t\to\infty$. The constant term in Eq. (90) is probably meaningless. Therefore we drop the constant term and obtain the final expression for the free energy

$$\beta F = \tfrac{1}{4}\beta t^2 + (1-\tfrac{1}{2}\beta)t - \tfrac{1}{2}\beta^{-1}(1-\tfrac{1}{2}\beta)^2 \ln t, \quad (91)$$

with an error which should be bounded as $t\to\infty$. The term in $(\ln t)$ is probably reliable. At the very worst, the second-order perturbation calculation, being based on a variation principle, shows that the error in Eq. (91) cannot be of greater order than $(\ln t)$.

Equations (5) and (24) give the result

$$P_\beta(t)\sim A t^{f(\beta)} \exp[-\tfrac{1}{4}\beta t^2-(1-\tfrac{1}{2}\beta)t], \quad f(\beta)=(1-\tfrac{1}{2}\beta)^2/2\beta \quad (92)$$

for the probability that a randomly-chosen interval of length $(2tD/\pi)$ be empty of levels in a series with mean spacing D. According to Eq. (49), this gives for the distribution-function of large spacings

$$Q_\beta(t)\sim A t^{2+f(\beta)} \exp[-\tfrac{1}{4}\beta t^2-(1-\tfrac{1}{2}\beta)t]. \quad (93)$$

In the case $\beta=2$, these expressions reduce to Eqs. (51) and (52). In applying the theory to nuclear level distributions, in which there is invariance under rotations and under time reversal, the case $\beta=1$ is relevant, and we find for $t\to\infty$

$$Q_1(t)\sim A t^{17/8} \exp[-\tfrac{1}{4}t^2-\tfrac{1}{2}t]. \quad (94)$$

This is the case to which Wigner's conjecture Eq. (2) applied. We conclude that Wigner's conjecture underestimates the frequency of large spacings by a factor which tends to infinity as $t\to\infty$. Needless to say, the range of t for which Wigner's conjecture is seriously in error includes so few level-spacings that it is for practical purposes completely unimportant.

For systems with odd spin, invariance under time reversal, and no rotational symmetry, we showed in Sec. III of I that the case $\beta=4$ applies. The level-spacing distribution is then asymptotically

$$Q_4(t)=A t^{17/8} \exp[-t^2+t]. \quad (95)$$

V. ADDITIONAL REMARKS

a. Accuracy of the Perturbation-Theory Calculation

The calculations of this paper are subject to two kinds of errors, (i) the inherent inaccuracy of the continuum model, and (ii) the inaccuracy of the treatment of the continuum model by perturbation theory in Sec. IV. We believe that the magnitude of type (i) errors in the free energy is bounded as $t\to\infty$; this belief is based only on physical intuition and cannot be checked by calculations within the continuum model. The magnitude of type (ii) errors can in principle be checked by pushing the perturbation calculations further.

We here examine the magnitude of the perturbation term $h(\psi)$ in Eq. (71) in comparison with the unperturbed term $M(\psi)$ given by Eq. (74). The explicit form of u_n obtained by minimizing Eq. (87) is

$$u_n=-2t[n+(\beta/(2-\beta))t]^{-1}. \quad (96)$$

With Eqs. (86) and (88), this gives

$$h(\psi)=-\alpha\sum_1^{\infty}\{n+[\beta/(2-\beta)]t\}^{-1}(-1)^n\cos 2n\psi. \quad (97)$$

The order of magnitude of $h(\psi)$ is

$$h(\psi)\sim\alpha\ln|t\cos\psi|, \quad |\cos\psi|<t^{-1}, \quad (98)$$

$$h(\psi)\sim\alpha|t\cos\psi|^{-2}, \quad t^{-1}<|\cos\psi|<t^{-\frac{1}{2}}, \quad (99)$$

$$h(\psi)\sim\alpha t^{-1}, \quad t^{-\frac{1}{2}}<|\cos\psi|. \quad (100)$$

The comparison term $M(\psi)$ is by Eq. (74) always at least of the order α. We have then

$$h(\psi)\ll M(\psi) \quad (101)$$

except in the range (98). So the perturbation theory is reliable except in the range of angles ψ within t^{-1} of $(\pm\frac{1}{2}\pi)$. The excluded range is mapped in the z plane onto the range

$$\alpha<|\theta|<\alpha+t^{-2}\alpha, \quad (102)$$

at the extreme tips of the arc $[\alpha<\theta<2\pi-\alpha]$. Equation (47) then shows that the total amount of the unperturbed charge in the excluded region is approximately one unit. Therefore the perturbation theory breaks down just in the space occupied by a single charge at the tips of the arc, where the continuum model is anyhow meaningless.

The foregoing argument indicates that the series (86) for $h(\psi)$ has a meaning up to frequencies n of the order of t, while the terms with $n>t$ are meaningless. The same conclusion holds for the series (89) giving the second-order contribution βF_2 to the free energy. The terms in βF_2 up to $n\sim t$ give the part of Eq. (90)

proportional to $(\ln t)$, while the terms with $n>t$ affect only the constant in Eq. (90). We thus arrive at the following general conclusions concerning the accuracy of the calculations.

(a) Errors of type (ii) are of the order of unity in the free energy and do not affect the logarithmic term in Eq. (91).

(b) Errors of type (i) probably appear only where the perturbation theory breaks down, and therefore the perturbation theory makes physical sense as far as we have carried it.

(c) It would make no sense to carry the perturbation theory to higher orders, since any higher-order terms would be of the same order of magnitude as the type (i) errors.

b. Gaussian Model

The calculations of this paper were based on the circular ensembles defined in (I). The same thermodynamic methods could just as well have been applied directly to the Gaussian ensemble[7] which has been the starting-point for the other workers[2,4,5] in this field. In the Gaussian ensemble, the angles $[\theta_1,\cdots,\theta_N]$ are replaced by real numbers $[\epsilon_1,\cdots,\epsilon_N]$ free to vary from $(-\infty)$ to $(+\infty)$. The potential energy W is given by

$$W=-\sum_{i<j}\ln|\epsilon_i-\epsilon_j|+(4a^2)^{-1}\sum_j\epsilon_j^2, \qquad (103)$$

instead of by Eq. (4). The continuum model is then a classical charged fluid, confined to a straight conducting wire and attracted to a fixed point 0 on the wire by a harmonic potential.

The analysis of the Gaussian continuum model proceeds almost as easily as for the circular model. There is only one essential complication. The conducting wire cannot be allowed to be infinite, because the attractive potential would then bring in charge from large distances in indefinite amounts. Negative charge-density is allowed by classical electrostatics but not by the conditions of this model. The appropriate model is a conducting wire of finite length, the length being chosen so that the charge-density shall be positive everywhere on the wire but zero at the end-points. When the model with a gap is introduced, the length of the wire must be adjusted so that the condition of zero charge density at the ends is maintained. Once this is done, the calculations proceed as before, and the final results are identical with those we have obtained in Secs. III and IV.

[7] E. P. Wigner, Proc. 4th Can. Math. Congress, p. 174 (Toronto, 1959), has in fact used this method to determine the over-all eigenvalue distribution of the Gaussian ensemble. For the over-all distribution, in contrast to the distribution of level spacings, a zero-temperature approximation is sufficient. Wigner was therefore able to derive the "semi-circle law" for the eigenvalue distribution, using a purely electrostatic model without any thermodynamics.

c. Case of Negative t

The partition function $\Psi_{N\beta}$ given by Eq. (6) has a well-defined meaning when the angle α is replaced by $(-\alpha)$. The integration with respect to each variable θ_j is then to be taken from 0 to 2π, with the interval from $(-\alpha)$ to $(+\alpha)$ counted twice. The ratio $R_\beta(-\alpha)$ is the expectation value of 2^k, where k is the number of the θ_j lying in the range $|\theta|<\alpha$. The function $P_\beta(-t)$ is the expectation value of 2^k, where k is the number of energy levels, in a series with mean spacing D, which happen to lie in a randomly chosen interval of length $(2tD/\pi)$. The expectation values are to be taken from the usual ensemble at temperature β^{-1}.

At infinite temperature ($\beta=0$) the value of $P_\beta(-t)$ is

$$P_0(-t)=[1+(\alpha/\pi)]^N=\exp[2t/\pi]. \qquad (104)$$

At any temperature we have

$$P_\beta(-t)=\langle 2^k\rangle\geq 2^{\langle k\rangle}=\exp[2t\ln 2/\pi]. \qquad (105)$$

In fact $P_\beta(-t)$ is a decreasing function of β and always lies between the limits (104) and (105).

The behavior of $P_\beta(-t)$ for large t can be determined from a continuum model. Instead of having a gap from $\theta=-\alpha$ to $\theta=+\alpha$, the model is now a complete circular wire with a potential

$$U=-\beta^{-1}\ln 2 \qquad (106)$$

applied to the interval $(|\theta|<\alpha)$. This adds a term

$$G_1'=-(N/2\pi)\ln 2\int_{-\alpha}^{\alpha}\rho_\alpha(\theta)d\theta \qquad (107)$$

to the free energy given by Eqs. (18), (19), the other integrals now all running from 0 to 2π.

We can calculate the free energy as before by perturbation theory, using G_2 alone for the unperturbed system. The calculation is much simpler than for positive t. The unperturbed charge density is $\bar\rho_\alpha(\theta)=1$, and the unperturbed free energy is zero. The first-order perturbation produced by Eq. (107) is

$$\beta F_1=-(2t/\pi)\ln 2. \qquad (108)$$

Second-order perturbation theory adds to this a contribution

$$\beta F_2=-\beta^{-1}[(\ln 2)/\pi]^2\ln t. \qquad (109)$$

The asymptotic behavior of $P_\beta(-t)$ at large t is thus

$$P_\beta(-t)\sim At^{g(\beta)}\exp[(2\ln 2/\pi)t],$$
$$g(\beta)=(\ln 2)^2/\pi^2\beta \qquad (110)$$

and the asymptotic behavior of $Q_\beta(-t)$ is the same.

These results for negative t are not of any practical importance. Their chief interest is that they impose necessary conditions which any exact analytic formula

for $Q(t)$ must satisfy. In particular, even the elementary inequality (105) is not satisfied by Wigner's conjectured Eq. (2), and this provides the shortest proof that Wigner's conjecture cannot be exactly correct.

d. Magnitude of Fluctuations in Level Density

One further consequence emerges from Eq. (110). Let an interval of length (MD) be chosen at random in a long series of energy levels with average spacing D. Let k be the number of levels lying in the interval. Then Eq. (110) may be written

$$\langle 2^k \rangle = A M^{g(\beta)} 2^M. \tag{111}$$

This implies that the variable k is distributed about its mean value $\langle k \rangle = M$ with a mean-square fluctuation

$$\langle (k-M)^2 \rangle = [(2/\pi^2\beta) \ln M] + R, \tag{112}$$

the remainder term R being bounded for large M. Equation (112) shows that the fluctuations in k are enormously less than they would be for an uncorrelated series of levels, which would give

$$\langle (k-M)^2 \rangle = M. \tag{113}$$

The difference between Eqs. (112) and (113) is a measure of the power of the long-range level repulsion in suppressing large fluctuations of level density. For the case $\beta = 1$ which applies to observed level-series, a more precise result than Eq. (112) will be proved in Paper IV.

Statistical Theory of the Energy Levels of Complex Systems. III

FREEMAN J. DYSON
Institute for Advanced Study, Princeton, New Jersey
(Received September 15, 1961)

A systematic method is developed for calculating the n-level correlation-function $R_n(x_1,\cdots,x_n)$, defined as the probability for finding n levels at positions $(x_1,\cdots,x_n)$, regardless of the positions of other levels. It is supposed that the levels of a complex system are statistically equivalent to the eigenvalues of a random symmetric unitary matrix of order $N\gg n$, according to the general theory described in an earlier paper. The 2-level correlation-function is found to be

$$R_2(x_1,x_2)=1-\{s(r)\}^2-\left\{\int_r^\infty s(t)dt\right\}\{ds(r)/dr\},$$
$$s(r)=[\sin(\pi r)/\pi r],\quad r=|x_1-x_2|,$$

the scale of energy being chosen so that the mean level-spacing is unity. It is shown how this result could in principle be used in order to determine the proportions of levels in two uncorrelated and superimposed series. An analytic expression for the distribution of nearest-neighbor level-spacings, discovered by Gaudin and Mehta, is rederived, and a similar expression is found for the distribution of spacings between next-nearest neighbors. An unexplained identity relates the nearest and next-nearest neighbor spacing distributions of a system invariant under time-reversal to the level-spacing distribution of a system without time-reversal invariance.

INTRODUCTION

THIS paper will be concerned with the study of the statistical properties of N points $[\exp(i\theta_j)]$, $j=1,\cdots,N$, distributed around the unit circle with the probability distribution function

$$P_N(\theta_1,\cdots,\theta_N)=C_N\prod_{i<j}|e^{i\theta_i}-e^{i\theta_j}|. \tag{1}$$

In paper I of the present series[1] it was shown that the distribution (1) holds for the eigenvalues of a symmetric unitary $(N\times N)$ matrix chosen at random o t of a certain ensemble, called the orthogonal ensemble. It was suggested that the series of angles $[\theta_1,\cdots,\theta_N]$ derived from this particular ensemble should provide a good model for the statistical behavior of the energy levels of a sufficiently complicated system. According to Eq. (I, 130), the probability distribution (1) is correctly normalized if the constant C_N has the value

$$C_N=[2^{2N}\pi^N\Gamma(1+\tfrac{1}{2}N)]^{-1}. \tag{2}$$

The main objective of the analysis is to calculate the *n-level correlation function*

$$R_n(\theta_1,\cdots,\theta_n)=[N!/(N-n)!]\int\cdots\int_0^{2\pi}P_N(\theta_1,\cdots,\theta_N)d\theta_{n+1}d\theta_{n+2}\cdots d\theta_N, \tag{3}$$

which measures the probability of finding a level (regardless of labeling) in each of the small intervals $[\theta_1,\ \theta_1+d\theta_1]$, $\cdots$, $[\theta_n,\ \theta_n+d\theta_n]$, the positions of the remaining $(N-n)$ levels being unobserved. In particular

$$R_1(\theta)=(N/2\pi) \tag{4}$$

is the constant over-all level density. Each function R_n for $n>1$ contains terms of various kinds, describing the grouping of n levels into various subgroups or clusters. For practical purposes it is convenient to work with the *n-level cluster function* defined by

$$T_n(\theta_1,\cdots,\theta_n)=\sum_G(-1)^{n-m}(m-1)!\times\prod_{j=1}^m R_{G_j}[\theta_k \text{ with } k \text{ in } G_j]. \tag{5}$$

Here G stands for any division of the indices $[1, 2, \cdots, n]$ into subgroups $[G_1, G_2, \cdots, G_m]$. Eqation (5) is a finite sum of products of R-functions, the first term in the sum being $[(-1)^{n-1}R_n(\theta_1,\cdots,\theta_n)]$, and the last being the constant

$$(n-1)!(N/2\pi)^n. \tag{6}$$

The inverse of Eq. (5) is

$$R_n(\theta_1,\cdots,\theta_n)=\sum_G(-1)^{n-m}\prod_{j=1}^m T_{G_j}[\theta_k \text{ with } k \text{ in } G_j]. \tag{7}$$

Thus each set of functions R_n and T_n is easily determined in terms of the other. The advantage of the cluster functions is that they have the property of vanishing when any one (or several) of the separations $|\theta_i-\theta_j|$ becomes large in comparison with the mean level spacing $(2\pi/N)$. The function T_n describes the correlation properties of a single cluster of n levels, isolated from the more trivial effects of lower-order correlations.

Of special interest for comparison with experiment are those features of the statistical model which tend to definite limits as $N\to\infty$. The cluster functions are convenient also from this point of view. In the limit $N\to\infty$, the angles θ_j must be replaced by real numbers

$$x_j=(N/2\pi)\theta_j, \tag{8}$$

[1] F. J. Dyson, J. Math. Phys. 3, 140 (1962); 3, 157 (1962), these two papers will be quoted as I and II.

each x_j being free to vary from $(-\infty)$ to $(+\infty)$, and the index j running from $(-\infty)$ to $(+\infty)$ also. The x_j then form a statistical model for an infinite series of energy-levels with mean spacing $D=1$. The cluster functions,

$$Y_n(x_1,\cdots,x_n)=\lim_{N\to\infty}(N/2\pi)^{-n}T_n(\theta_1,\cdots,\theta_n), \quad (9)$$

are well defined and finite everywhere. In particular

$$Y_1(x)=1, \quad (10)$$

while

$$Y_2(x_1,x_2)=Y_2(|x_1-x_2|) \quad (11)$$

defines the shape of the neutralizing charge-cloud induced by each particle around itself, when the model is interpreted as a classical Coulomb gas [see Sec. VI of I].

The cluster functions satisfy the identity

$$\int_{-\infty}^{\infty}Y_n(x_1,\cdots,x_n)dx_n=(n-1)Y_{n-1}(x_1,\cdots,x_{n-1}), \quad (12)$$

for $n>1$. This means that each Y_n is an integrable function of the $(n-1)$ variables $(x_1-x_n, x_2-x_n, \cdots, x_{n-1}-x_n)$, and has a Fourier transform

$$y_n(k_1,\cdots,k_{n-1}) = \int\cdots\int_{-\infty}^{\infty}Y_n(x_1,\cdots,x_n)\exp\{2\pi i[k_1(x_1-x_n)+\cdots +k_{n-1}(x_{n-1}-x_n)]\}dx_1\cdots dx_{n-1}. \quad (13)$$

The limiting forms of the correlation functions R_n are not integrable, and their Fourier transforms involve products of δ functions.

Many important properties of the level-distribution depend only on the *two-level form-factor* defined by

$$b(k)=y_2(k)=\int_{-\infty}^{\infty}Y_2(r)e^{2\pi ikr}dr. \quad (14)$$

The normalization is chosen to make $b(0)=1$, by Eqs. (10) and (12).

In this paper it will be shown that all the cluster-functions Y_n are in principle calculable. The Y_n will be exhibited as coefficients in the expansion of a certain determinant. However, the elementary algebra that is required for the extraction of the higher Y_n is very tedious. Explicit evaluations will be made only for the two-level functions $Y_2(r)$ and $b(k)$.

The method of calculation is essentially copied from the work of Gaudin and Mehta,[2] who first discovered how to deal with integrals of the form of Eq. (3). All the serious difficulties are overcome by the Gaudin-Mehta method. The analysis differs from th t of Gaudin-Mehta in two respects. (i) We deal with distributions around a circle, while Gaudin and Mehta used a Gaussian distribution on a straight line. (ii) We are interested in a precise evaluation of the two-level correlation function, while they considered only the more difficult problem of the distribution of level-spacings.

II. GAUDIN-MEHTA METHOD ON A CIRCLE

To avoid minor complications, let $N=2m$ be even. Let $U(\theta)$ and $V(\theta)$ be any two functions defined on the unit circle $(-\pi\leq\theta\leq\pi)$. Consider the quantity H, defined as the expectation-value

$$H=\langle[\prod\nolimits_{\text{alt}}U(\theta_j)][\prod\nolimits'_{\text{alt}}V(\theta_j)]\rangle, \quad (15)$$

taken with respect to the probability-distribution (1). Here $\prod_{\text{alt}}$ means a product taken over a set of m alternate points θ_j as they lie on the unit circle, and $\prod'_{\text{alt}}$ means a product over the remaining m points θ_j. The alternating series may start anywhere on the circle, no special end-point being singled out.

By Eqs. (1) and (2), an explicit formula for H is

$$H=\frac{(2m)!}{2^{4m+1}\pi^m m!}\int_{-\pi}^{\pi}d\theta_1\int_{\theta_1}^{\pi}d\theta_2\cdots\int_{\theta_{2m-1}}^{\pi}d\theta_{2m} \times\prod_{i<j}[2\sin\tfrac{1}{2}(\theta_j-\theta_i)]\{\prod_{k=1}^{m}U(\theta_{2k-1})V(\theta_{2k}) +\prod_{k=1}^{m}V(\theta_{2k-1})U(\theta_{2k})\}. \quad (16)$$

This may be transformed into

$$H=\frac{(2m)!}{2^{5m}\pi^m(m!)^2}\int\cdots\int_{-\pi}^{\pi}d\theta_1\cdots d\theta_{2m} \times\prod_{i<j}[2\sin\tfrac{1}{2}(\theta_j-\theta_i)]\prod_{k=1}^{m}\{V(\theta_{2k-1})U(\theta_{2k}) \times\epsilon(\theta_{2k}-\theta_{2k-1})\}, \quad (17)$$

where $\epsilon(x)=(x/|x|)$. To deduce Eq. (17) from Eq. (16), let J be the integral on the right of Eq. (17). If any two of the θ_{2j} lie consecutively on the circle, the integrand of J is odd under interchange of these two variables, and that part of J vanishes. Similarly the part of J with any two of the θ_{2j-1} lying consecutively vanishes. The surviving part of J has the θ_{2j} and the θ_{2j-1} forming two alternate series, but not necessarily arranged in order as they appear in Eq. (16). Suppose that the θ_{2j} in J appear in a permutation P of the natural order, while the θ_{2j-1} appear in a permutation Q. The entire integral J then reduces to the integral

[2] M. L. Mehta, Nuclear Phys. 18, 395 (1960); M. L. Mehta and M. Gaudin, Nuclear Phys. 18, 420 (1960); M. Gaudin, Nuclear Phys. 25, 447 (1960).

appearing in Eq. (16), multiplied by the constant

$$C=\sum_{P,Q}\epsilon_P\epsilon_Q\prod_{k=1}^{m}\epsilon(2P_k-2Q_k+1). \tag{18}$$

But C is just $(m!)$ multiplied by an $(m\times m)$ determinant whose element number (i,j) is $(+1)$ when $i\geq j$, (-1) when $i<j$. The value of this determinant is 2^{m-1}, hence

$$C=2^{m-1}m!, \tag{19}$$

and Eq. (17) is proved.

The next step is to write

$$\prod_{i<j}[2\sin\tfrac{1}{2}(\theta_j-\theta_i)]=i^{-m}\det|\exp(ip\theta_j)|, \tag{20}$$

where the determinant is $(2m\times 2m)$, the column index j taking the values $(1,\cdots,2m)$ while the row-index p takes the values

$$p=-m+\tfrac{1}{2},\quad -m+\tfrac{3}{2},\quad \cdots,\quad m-\tfrac{1}{2}. \tag{21}$$

Squaring Eq. (17), using Eq. (20) and expanding the determinants, we obtain

$$H^2=\frac{[(2m)!]^2(-1)^m}{2^{12m}\pi^{2m}(m!)^4}\sum_{P,Q}\epsilon_P\epsilon_Q \times\prod_{k=1}^{m}\{g(P_{2k},P_{2k-1})g(Q_{2k},Q_{2k-1})\}, \tag{22}$$

with

$$g(p,q)=\int\int_{-\pi}^{\pi}d\theta d\varphi U(\theta)V(\varphi)\epsilon(\theta-\varphi) \times[\exp(ip\theta+iq\varphi)-\exp(iq\theta+ip\varphi)]. \tag{23}$$

P_j and Q_j are any two permutations taking values in the range (21) for $j=1,\cdots,2m$.

Let (p,q) be indices in the range (21). We say that (p,q) are "partners in P" if for some k we have $p=P_{2k}$, $q=P_{2k-1}$ or $p=P_{2k-1}$, $q=P_{2k}$. Similarly we define partners in Q. We can then construct a permutation R on $(-m+\frac{1}{2},\cdots,m-\frac{1}{2})$ such that $R_p=q$ if and only if (p,q) are partners in P or in Q. Such a permutation R must consist of a number L of cycles each of even order. When P and Q are given, the composition of the cycles in R is fixed, and only the sense of those cycles which have order greater than 2 is undetermined. Therefore, the number of distinct R associated with a given P and Q is

$$2^{L-h}, \tag{24}$$

where h is the number of cycles in R of order 2. This h is also the number of pairs (p,q) which are partners in both P and Q.

Conversely, if a permutation R containing only even cycles is given, this determines the pairing of partners in P and in Q with precisely the same degree of ambiguity 2^{L-h}. Given the pairings, the complete specification of P can be made in $2^m(m!)$ ways, and similarly for Q. Therefore every R can arise from

$$2^{L-h+2m}(m!)^2 \tag{25}$$

distinct pairs (P,Q).

The parity of the permutation R is

$$\epsilon_R=(-1)^L. \tag{26}$$

The combined parity $\epsilon_P\epsilon_Q$ is the parity of the permutation

$$S=\begin{bmatrix}P_1P_2\cdots P_{2m}\\ Q_1Q_2\cdots Q_{2m}\end{bmatrix}.$$

Now ϵ_S is unchanged if we interchange the pairs (Q_{2j-1},Q_{2j}) in blocks so as to make $P_1=Q_1$, $P_3=Q_3$, etc. The resulting permutation S is

$$\begin{bmatrix}P_2P_4\cdots P_{2m}\\ Q_2Q_4\cdots Q_{2m}\end{bmatrix}, \tag{27}$$

and this is obtained from R by taking just half the indices in each cycle. The parity of S is therefore

$$\epsilon_P\epsilon_Q=\epsilon_S=(-1)^\lambda, \tag{28}$$

and λ is the number of cycles in R whose length is divisible by 4. Since the sum of lengths of all the cycles in R is $2m$, we have

$$L-\lambda\equiv m\ (\text{mod. }2). \tag{29}$$

Equations (26), (28), and (29) give

$$\epsilon_P\epsilon_Q=(-1)^m\epsilon_R. \tag{30}$$

The sum (22) is now expressed in terms of the permutations R alone. Combining Eq. (22) with Eqs. (24), (25), (30), we find

$$H^2=\left[\frac{(2m)!}{2^{5m}\pi^m m!}\right]^2(-1)^m\sum_R\epsilon_R\prod_{p=-m+\frac{1}{2}}^{m-\frac{1}{2}}g(p,R_p). \tag{31}$$

The factor $(-1)^m$ reappears in Eq. (31) because we used the relation

$$g(p,q)=-g(q,p) \tag{32}$$

which follows from Eq. (23). The sum in Eq. (31) is over permutations R consisting of even cycles only. However, by virtue of Eq. (32), any permutation including an odd cycle would cancel in Eq. (31) against the same permutation with the odd cycle taken in the opposite sense. Therefore the sum over R may be extended to all permutations, and H^2 reduces to a determinant. The factor $(-1)^m$ may be again absorbed by changing $(q\rightarrow -q)$ in $g(p,q)$, which is equivalent to reversing the order of the $(2m)$ columns of $\det|g(p,q)|$. The constant factor in Eq. (31) may be absorbed by multiplying the pth row of the determinant by $[p/8\pi i]$.

In this way Eq. (31) reduces to

$$H^2=\det|f_{pq}|, \tag{33}$$

with

$$f_{pq}=\frac{p}{8\pi i}\int\int_{-\pi}^{\pi}d\theta d\varphi U(\theta)V(\varphi)\epsilon(\theta-\varphi)\times[\exp(ip\theta-iq\varphi)-\exp(ip\varphi-iq\theta)], \tag{34}$$

both p and q taking the values $[-m+\frac{1}{2},\cdots,m-\frac{1}{2}]$.

The result (33) corresponds to Eq. (10) in Mehta's paper.[3] Mehta does not make any use of his Eq. (10), and instead concentrates his attention on a determinantal expression for the first power of the integral corresponding to our H, namely his Eq. (14). The analogous formula for the first power of H is

$$H=\det|F_{pq}|\,;\ p,q=\tfrac{1}{2},\tfrac{3}{2},\cdots,m-\tfrac{1}{2}, \tag{35}$$

with

$$F_{pq}=\frac{p}{4\pi}\int\int_{-\pi}^{\pi}d\theta d\varphi U(\theta)V(\varphi)\epsilon(\theta-\varphi)\times[\cos p\varphi\,\sin q\theta-\cos p\theta\,\sin q\varphi]. \tag{36}$$

Equation (35) follows immediately from Eq. (33), provided that

$$U(\theta)V(\varphi)=U(-\theta)V(-\varphi), \tag{37}$$

which means in practice that U and V must either be both even functions or both odd functions on $[-\pi,+\pi]$.

Superficially, Eq. (35) appears simpler and more elegant than Eq. (33). However, the restriction (37) is highly inconvenient and makes it difficult to obtain directly from Eq. (35) any information about the cluster functions T_n. For our purposes Eq. (33), which holds without restriction on the functions U and V, is much more useful.

Another advantage of Eq. (33) is that it is independent of the arbitrary choice of the end points $[-\pi,\pi]$ on the circle. Since p and q are both half-odd-integers, one may write

$$f_{pq}=\frac{p}{8\pi i}\int\int d\theta d\varphi U(\theta)V(\varphi)\times[\exp(ip\theta-iq\varphi)-\exp(ip\varphi-iq\theta)], \tag{38}$$

the range of integration being limited only by

$$0<\theta-\varphi<2\pi. \tag{39}$$

A similar independence of the end point does not hold for Eq. (36).

When $U(\theta)=V(\theta)=1$, Eq. (34) gives

$$f_{pq}=\delta_{pq} \tag{40}$$

and so $H^2=1$, as it should be according to Eq. (15). This is a confirmation by direct calculation that the normalization of probability given by Eq. (2) is correct. In paper I this normalization was deduced independently, by a group-theoretical argument.

[3] The first paper in reference 2.

III. TWO-LEVEL CORRELATION FUNCTIONS

Write $U(\theta)=V(\theta)=1+A(\theta)$ in Eq. (15), where $A(\theta)$ is still an arbitrary function. Then

$$H=\langle\prod[1+A(\theta_j)]\rangle, \tag{41}$$

the product now extending over all the $(2m)$ levels θ_j. Equations (33) and (34) now become

$$H^2=\det|\delta_{pq}+r_{pq}|, \tag{42}$$

$$r_{pq}=\frac{1}{2\pi}\left(1+\frac{p}{q}\right)\int_{-\pi}^{\pi}A(\theta)\exp[i(p-q)\theta]d\theta+\left(\frac{p}{4\pi i}\right)\int\int_{-\pi}^{\pi}A(\theta)A(\varphi)\epsilon(\theta-\varphi)\times\exp[i(p\theta-q\varphi)]d\theta d\varphi. \tag{43}$$

All the cluster-functions $T_n(\theta_1,\cdots,\theta_n)$ can, in principle, be determined by expanding the two sides of the identity (42) in powers of $A(\theta)$. According to Eqs. (3), (41), and (7),

$$H=\sum_{n=0}^{\infty}\frac{1}{n!}\int\cdots\int_0^{2\pi}R_n(\theta_1,\cdots,\theta_n)\times A(\theta_1)\cdots A(\theta_n)d\theta_1\cdots d\theta_n \tag{44}$$

$$=\exp\left\{\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n!}\int\cdots\int_0^{2\pi}T_n(\theta_1,\cdots,\theta_n)\times A(\theta_1)\cdots A(\theta_n)d\theta_1\cdots d\theta_n\right\}. \tag{45}$$

The determinant for H^2 can be expanded along its leading diagonal. The result is a series beginning with the terms

$$H^2=1+(2m/\pi)\int A(\theta)d\theta+\frac{m(2m-1)}{\pi^2}\left[\int A(\theta)d\theta\right]^2+\sum_p\left(\frac{p}{4\pi i}\right)\int\int A(\theta)A(\varphi)\epsilon(\theta-\varphi)\times\exp[ip(\theta-\varphi)]d\theta d\varphi-\frac{1}{8\pi^2}\sum_{p\neq q}\left(1+\frac{p}{q}\right)\left(1+\frac{q}{p}\right)\times\int\int A(\theta)A(\varphi)\exp[i(p-q)(\theta-\varphi)]d\theta d\varphi, \tag{46}$$

the remaining terms being of order A^3 and higher. Since the function $A(\theta)$ is arbitrary (and this is here essential), each T_n can be picked out as the coefficient

of $[A(\theta_1)\cdots A(\theta_n)]$ in the logarithm of the series (46).

In the case $n=2$, this procedure gives

$$T_2(\theta,\varphi)=-\sum_p (p/4\pi i)\epsilon(\theta-\varphi)\exp[ip(\theta-\varphi)]$$
$$+\frac{1}{8\pi^2}\sum_{p,q}\left(2+\frac{p}{q}+\frac{q}{p}\right)\exp[i(p-q)(\theta-\varphi)]$$
$$=+\tfrac{1}{2}\epsilon(\theta-\varphi)Ds_N(\theta-\varphi)-\{Is_N(\theta-\varphi)\}$$
$$\times\{Ds_N(\theta-\varphi)\}+\{s_N(\theta-\varphi)\}^2, \quad (47)$$

where we have written

$$s_N(\alpha)=\frac{1}{2\pi}\sum_p e^{ip\alpha}=\frac{\sin(m\alpha)}{2\pi\sin(\frac{1}{2}\alpha)}, \quad (48)$$

$$Df(\alpha)=(\partial/\partial\alpha)f(\alpha), \quad (49)$$

$$If(\alpha)=\int_0^\alpha f(\alpha')d\alpha'. \quad (50)$$

In the limit $N\to\infty$, Eq. (47) becomes

$$Y_2(x_1,x_2)=\{\tfrac{1}{2}-Is(r)\}\{Ds(r)\}+\{s(r)\}^2, \quad (51)$$
$$r=|x_1-x_2|,$$

$$s(r)=\lim_{N\to\infty}\{(2\pi/N)s_N(2\pi r/N)\}=\frac{\sin(\pi r)}{\pi r}. \quad (52)$$

Since

$$\int_0^\infty s(r)dr=\tfrac{1}{2}, \quad (53)$$

Eq. (51) is equivalent to the formula for R_2 stated in the abstract.

The behavior of Y_2 for small and large values of r is given by

$$Y_2(r)=1-\frac{1}{6}\pi^2r+\frac{1}{60}\pi^4r^3-\frac{1}{135}\pi^4r^4+\cdots, \quad (54)$$

$$Y_2(r)=\frac{1}{\pi^2r^2}-\frac{1+\cos^2(\pi r)}{\pi^4r^4}+\cdots. \quad (55)$$

The Fourier transform of Y_2 gives the two-level form factor according to Eq. (14),

$$b(k)=1-2|k|+|k|\ln(1+2|k|), \quad (|k|\le 1),$$
$$=-1+|k|\ln[(2|k|+1)/(2|k|-1)], \quad (|k|\ge 1). \quad (56)$$

This has the behavior

$$b(k)=1-2|k|+2k^2+\cdots \quad (57)$$

$$b(k)=\frac{1}{12k^2}+\frac{1}{80k^4}+\cdots \quad (58)$$

for small and large k. At the points $(k=\pm 1)$ where the analytic form of $b(k)$ changes, not only $b(k)$ but also its first two derivatives are continuous. There is a discontinuity only in the third derivative. This is connected with the fact that the oscillating term in $Y_2(r)$ according to Eq. (55) is of order r^{-4} for large r.

The oscillating term in Eq. (55) is of considerable interest, as it indicates the presence of an incipient crystal-lattice structure or long-range order in a series of eigenvalues. Even at large separations, two eigenvalues feel the natural periodicity of the lattice, and have a slight preference for separations which are an integer multiple of the mean level spacing. Unfortunately, the r^{-4} dependence of this effect makes it unobservable in practice. To see the second maximum (at $r=1$) of the oscillatory term standing out from statistical fluctuations, one would need a well-observed series of more than 10 000 levels.

The Gaudin-Mehta method gives information not only about the total eigenvalue distribution but also about the separate distributions of odd-numbered or even-numbered levels. For example, one may take in Eq. (15) $U(\theta)=1$, $V(\theta)=1+A(\theta)$. Then

$$H=\langle\textstyle\prod_{\text{alt}}[1+A(\theta_j)]\rangle, \quad (59)$$

the product extending over m levels lying alternately on the unit circle. Equation (42) now holds with

$$r_{pq}=\frac{1}{4\pi}\left(1+\frac{p}{q}\right)\int_{-\pi}^{\pi}A(\theta)\exp[i(p-q)\theta]d\theta. \quad (60)$$

The expression of H involves correlation functions of the alternate eigenvalue series. The analysis proceeds as before, only the term in $\frac{1}{2}\epsilon(\theta-\varphi)$ is now missing from Eq. (47). The results are the following.

In an infinite eigenvalue series with mean spacing $D=1$, let

$$\tfrac{1}{2}[1-Y_2^e(x_1,x_2)]dx_1dx_2 \quad (61)$$

be the probability for finding two eigenvalues in the intervals $[x_1, x_1+dx_1]$, $[x_2, x_2+dx_2]$, both belonging to the same alternate series. Then

$$Y_2^e(x_1,x_2)=\{s(r)\}^2-\{Is(r)\}\{Ds(r)\}, \quad (62)$$

with the same notations as in Eq. (51). For large and small r we find

$$Y_2^e(r)=1-\frac{1}{135}\pi^4r^4+\cdots, \quad (63)$$

$$Y_2^e(r)=-\frac{\cos(\pi r)}{2r}+\frac{1+\frac{1}{2}\pi\sin(\pi r)}{\pi^2r^2}-\frac{1+\cos^2(\pi r)}{\pi^4r^4}$$
$$+\cdots. \quad (64)$$

The corresponding two-level form factor is

$$b^e(k)=2-2|k|+|k|\ln(|2|k|-1|), \quad (|k|<1),$$
$$=0, \quad (|k|>1). \quad (65)$$

The long-range order of the eigenvalue series appears much more strongly in Eq. (64) than in Eq. (55), and it shows clearest of all in the singularity of the Fourier transform $b^e(k)$ at $k=\frac{1}{2}$. This behavior of the alternate eigenvalues proves that the long-range crystalline structure of the level-series is real. In a one-dimensional gas, the operation of merely picking out alternate atoms for examination could not create long-range order if long-range order had not been present to start with.

In an observed series of levels of practical length, say a few hundred levels, the first few oscillations of Eq. (64) should be distinguishable. But for this test of the theory to be meaningful, it is necessary to be sure that all the observed levels belong to a single series and that none have been missed.

A very intriguing possibility suggests itself in the situation, very frequent in practice, in which an observed set of levels is a superposition of two uncorrelated series. This situation arises, for example, when slow neutrons are captured by an odd-A nucleus into levels with two possible spin values. In general it is difficult to say which levels belong to which series, and even the proportion of levels in the two series is a matter of conjecture.[4]

Let a level-series be a mixture of two uncorrelated series 1 and 2, in the proportions f, $(1-f)$. The 2-level cluster function $\bar{Y}_2$ of the combined series is then

$$\bar{Y}_2(r)=f^2Y_2(fr)+(1-f)^2Y_2[(1-f)r], \tag{66}$$

and the form factor is

$$\bar{b}(k)=fb(k/f)+(1-f)b[k/(1-f)]. \tag{67}$$

The function $\bar{b}(k)$ can be measured by Fourier analysis of the observed 2-level correlations. In principle, if the series were long enough, one could find the discontinuities in the third derivative of $\bar{b}(k)$ and so determine f directly. In practice this will not be possible, because the function $b(k)$ is too smooth. The stronger discontinuity of $b^e(k)$ is also of no help, since there is no way to pick out alternate levels from two superimposed series. Practical methods for determining f in such cases will be discussed in paper IV.

IV. REGULARITY OF EIGENVALUES AROUND THE CIRCLE

As a simple application of the theory of Sec. III, we calculate the mean-square deviation of the eigenvalues $[\exp(i\theta_j)]$ from a regular arrangement of N points on the unit circle. This mean-square deviation is

$$\Delta^2=\left\langle \frac{1}{N}\min_{\alpha}\sum_{j=1}^{N}\left|\exp(i\theta_j)-\exp\left(\frac{2\pi ij}{N}+i\alpha\right)\right|^2\right\rangle. \tag{68}$$

The minimization over the angle α is elementary, and gives

$$\Delta^2=1-(\zeta/N^2), \tag{69}$$

$$\zeta=\langle|\sum_{j=1}^{N}\exp[i\theta_j-(2\pi ij/N)]|^2\rangle. \tag{70}$$

The ensemble average can be expressed in terms of two-level correlations only, and Eq. (70) becomes

$$\zeta=\pi^2\csc^2(\pi/N)\left[1-(\pi/N^2)\times\int_{-\pi}^{\pi}T_2(\varphi)\{2\pi|\varphi|-\varphi^2\}d\varphi\right]. \tag{71}$$

By use of the Fourier expansion

$$2\pi|\varphi|-\varphi^2=\frac{8}{3}\pi^2-4\sum_{1}^{\infty}k^{-2}\cos(k\varphi), \tag{72}$$

Eq. (71) can be reduced to the form

$$\zeta=\frac{\pi^2}{\sin^2(\pi/N)}\left\{1+\frac{2}{N}\sum_{1}^{\infty}k^{-2}[b(k/N)-1]\right\}. \tag{73}$$

Now let $N\to\infty$. The sum in (73) may be split into two parts, $1\leq k<\eta N$ and $k>\eta N$, where η is small compared with unity. In the first sum we approximate $b(k/N)$ by $[1-(2k/N)]$ according to Eq. (56). The second sum reduces to the integral

$$\int_{\eta}^{\infty}[b(x)-1]x^{-2}dx=2[(\ln\eta)-1]+\tfrac{1}{4}\pi^2. \tag{74}$$

Equations (69) and (73) then give

$$\Delta^2=\frac{4}{N^2}\left[(\ln N)+\gamma+1-\frac{5\pi^2}{24}\right]. \tag{75}$$

Since the mean level spacing is $D=(2\pi/N)$, we have for large N

$$(\Delta^2/D^2)\sim(1/\pi^2)\ln N. \tag{76}$$

Thus the deviations of the θ_j from a regular polygonal arrangement are very small, on the average, even when N is as large as 10^3 or 10^6.

V. COMPARISON WITH THE UNITARY ENSEMBLE

It is of some interest to compare the results hitherto obtained with the corresponding results for an eigen-

[4] The general belief among nuclear theorists is that, when a nucleus of spin J captures a slow neutron, the compound states of spin $(J-\frac{1}{2})$ and $(J+\frac{1}{2})$ will occur roughly in the proportion J to $(J+1)$. E.g., T. D. Newton, Can. J. Phys. **34**, 804 (1956). The experimental evidence for or against this belief is quite meager. See for example J. A. Harvey, D. J. Hughes, R. S. Carter, and V. E. Pilcher, Phys. Rev. **99**, 10 (1955).

value series with the probability distribution

$$P_{Nu}(\theta_1,\cdots,\theta_N)=(1/N!)(2\pi)^{-N}\prod_{i<j}|e^{i\theta_i}-e^{i\theta_j}|^2, \quad (77)$$

instead of Eq. (1). The distribution P_{Nu} would be correct for the eigenvalues of a random unitary matrix in the unitary ensemble defined in paper I. This would be a model for the energy levels of a complex system without invariance under time reversal.

In the unitary ensemble the integrals (3) can be performed without difficulty, giving the n-level correlation functions

$$R_{nu}(\theta_1,\cdots,\theta_n)=\det|s_N(\theta_j-\theta_k)|, \quad (78)$$

the determinant being $(n\times n)$ and having every diagonal element equal to $(N/2\pi)$. It is easy then to verify from Eq. (7) that the cluster functions are

$$T_{nu}(\theta_1,\cdots,\theta_n)=\sum\prod[s_N(\theta_j-\theta_k)]. \quad (79)$$

Here $s_N(\alpha)$ is given by Eq. (48), and $\sum\prod$ means a sum over $[(n-1)!]$ products, of which one is

$$s_N(\theta_1-\theta_2)s_N(\theta_2-\theta_3)\cdots s_N(\theta_n-\theta_1), \quad (80)$$

the others being obtained from it by cyclic permutations of $[1,2,\cdots,n]$. In particular, when $n=2$,

$$T_{2u}(\theta_1,\theta_2)=\{s_N(\theta_1-\theta_2)\}^2, \quad (81)$$

and therefore

$$\begin{aligned} b_u(k)&=1-|k|, \quad (|k|<1),\\ &=0, \quad (|k|>1). \end{aligned} \quad (82)$$

The analogs to Eqs. (54), (55) are

$$Y_{2u}(r)=1-\frac{1}{3}\pi^2r^2+\frac{2}{45}\pi^4r^4+\cdots, \quad (83)$$

$$Y_{2u}(r)=\frac{\sin^2(\pi r)}{\pi^2r^2}. \quad (84)$$

Equation (84) shows that the long-range order is much more marked in the unitary case than it was in Eq. (55). This was to be expected, since the unitary case is equivalent to a Coulomb gas at temperature $T=\frac{1}{2}$ instead of $T=1$.

The analog of Eq. (75) is

$$\Delta_u^2=(2/N^2)[(\ln N)+\gamma-(\pi^2/6)]. \quad (85)$$

Thus the mean-square displacement of the eigenvalues from a regular lattice is only half as great as before.

It is an interesting problem, which we have not been able to solve, to calculate the correlation-functions for the symplectic ensemble of paper I. In that case the square in Eq. (77) is replaced by a fourth power. The effects of long-range order must then be even stronger than in the unitary case.

VI. ENERGY AND SPECIFIC HEAT OF THE EQUIVALENT COULOMB GAS

We now return to the study of the probability distribution (1) of the eigenvalues of symmetric unitary matrices. In Sec. VI of I it was shown that this is also the distribution function of a classical Coulomb gas at temperature $T=1$. A measure of the energy of the gas is provided by the quantity

$$W=-\sum_{i<j}\ln|e^{i\theta_i}-e^{i\theta_j}|+\tfrac{1}{2}N\ln N. \quad (86)$$

This W is a convenient statistic by which to compare observed eigenvalue distributions with the theory, since we saw in Sec. IX (C) of I that the theoretical expectation-value and variance of W are both calculable. In fact, for large N,

$$\langle W\rangle=NU, \quad U=1-\tfrac{1}{2}\gamma-\tfrac{1}{2}\ln 2=0.365, \quad (87)$$

$$\langle(W-\langle W\rangle)^2\rangle=NC, \quad C=\frac{3}{2}-\frac{\pi^2}{8}=0.266. \quad (88)$$

Here U is the mean energy, and C the specific heat per particle, in the Coulomb gas at temperature 1. The values given in Eqs. (87), (88) were obtained from the conjectured analytic form of the partition function of the gas.

The value of $\langle W\rangle$ can also be expressed in terms of the 2-level cluster functions. A little manipulation of Eq. (86) gives

$$\begin{aligned} \langle W\rangle&=\tfrac{1}{2}\int\int T_2(\theta,\varphi)\ln|e^{i\theta}-e^{i\varphi}|\,d\theta\,d\varphi+\tfrac{1}{2}N\ln N\\ &=\tfrac{1}{2}N\int_{-\infty}^{\infty}Y_2(x)\ln|2\pi x|\,dx\\ &=-\tfrac{1}{2}N\lim_{\eta\to 0}\left\{\int_\eta^\infty\frac{b(k)}{k}dk+(\ln\eta)+\gamma\right\}. \end{aligned} \quad (89)$$

The integral (89) is elementary, and we obtain thereby a direct check of Eq. (87).

Suppose next that the observed levels are a mixture of two uncorrelated series in the proportions f, $(1-f)$. Equation (89) still holds, with the form factor $\bar{b}(k)$ given by Eq. (67). The expectation-value of W for the mixed series is therefore

$$\langle W\rangle=NU-\tfrac{1}{2}N\{f\ln f+(1-f)\ln(1-f)\}. \quad (90)$$

The expectation value of W^2 will bring in cluster functions of 2, 3, and 4 levels. After some algebra we find

$$\langle(W-\langle W\rangle)^2\rangle$$

$$=-\tfrac{1}{4}N\int\int\int Y_4(x,y,z,w)\ln(2\pi|x-y|)$$

$$\times\ln(2\pi|z-w|)dydzdw+N\int\int Y_3(x,y,z)$$

$$\times\ln(2\pi|x-y|)\ln(2\pi|x-z|)dydz-\tfrac{1}{2}N\int Y_2(x,y)$$

$$\times\{\ln(2\pi|x-y|)\}^2dy+\tfrac{1}{2}N\int\int\int[Y_2(x,z)$$

$$-\delta(x-z)][Y_2(y,w)-\delta(y-w)]\ln(2\pi|x-y|)$$

$$\times\ln(2\pi|z-w|)dydzdw. \quad (91)$$

The last term of Eq. (91) may be simply expressed in terms of the 2-level formfactor $b(k)$; it is in fact

$$\tfrac{1}{4}N\int_0^\infty[b(k)-1]^2k^{-2}dk. \quad (92)$$

The integrals in Eq. (91) could all in principle be evaluated, using the methods of Sec. III to determine the functions Y_3 and Y_4. However, this would be a tremendous labor; even to write down the explicit form of Y_3 takes many lines of print. It is very fortunate that the sum of all the integrals is known independently, being given by Eq. (88).

In the case of the unitary probability-distribution, the Y_n are simple functions given by Eq. (79), and the integrals (89) and (91) can be performed without too much trouble. In this case we obtain Eqs. (87) and (88) with

$$U=\tfrac{1}{2}(1-\gamma), \quad C=2-(\pi^2/6), \quad (93)$$

in agreement with the results of Sec. IX of I. This serves as a double check, verifying both the conjectured partition-function of paper I and the algebra leading to Eq. (91).

When the variance of W is computed for a mixture of two level-series, it turns out that only the last term of Eq. (91) is affected. After much cancellation of terms, we find

$$\langle(W-\langle W\rangle)^2\rangle=NC+\tfrac{1}{2}N\int_0^\infty[b(fk)-1]$$

$$\times[b((1-f)k)-1]k^{-2}dk. \quad (94)$$

The integral in Eq. (94) does not seem to be expressible in terms of elementary functions. In the simplest case $f=\frac{1}{2}$, the added term (94) is just equal to the term (92) already appearing in Eq. (91), but even in this case no analytic integration seems possible.

If $b(k)$ given by Eq. (56) is plotted numerically, it is seen that

$$b(k)=\exp(-2|k|) \quad (95)$$

is quite an accurate approximation except for large $|k|$. Large values of $|k|$ are unimportant in Eq. (94), and so Eq. (95) should give a useful approximation. Substitution of Eq. (95) into (94) gives the result

$$\langle(W-\langle W\rangle)^2\rangle=NC-N\{f\ln f+(1-f)\ln(1-f)\}, \quad (96)$$

which should be accurate within 10% for all f. Note that Eq. (90), although of similar appearance to Eq. (96), is an exact formula not resting upon the approximation (95).

Suppose that a series of N eigenvalues around the unit circle is known to be a mixture of two series having separate probability distributions of the form (1), only the proportions f, $(1-f)$ of the two series being unknown. Then a single measurement of W for the combined series will yield a value for f by Eq. (90). The variance of this measurement of f is by Eq. (96)

$$\langle(\Delta f)^2\rangle=\frac{4}{N}\frac{C-f\ln f-(1-f)\ln(1-f)}{\ln f-\ln(1-f)}. \quad (97)$$

Unless f happens to be very close to $\frac{1}{2}$, the expected error in the measurement of f is of the order of $N^{-\frac{1}{2}}$.

This method of measuring f is obviously far more precise than the method discussed in Sec. III. However, it is not yet a practical method, since it requires observation of the eigenvalues round an entire circle. In paper (IV) we shall show how the method can be adapted to the practical situation in which we observe eigenvalues only on a small part of the circle.

VII. LEVEL-SPACING DISTRIBUTION

Until now we have studied only the probability-distributions $R_n(\theta_1,\cdots,\theta_n)$ for finding n levels at a given set of positions, irrespective of the remaining levels. Gaudin and Mehta,[2] following Wigner, were mainly interested in the level-spacing distribution function $S(\theta)$. This is defined by the statement that $S(\theta_2-\theta_1)d\theta_1d\theta_2$ is the probability for finding 2 levels in the intervals $[\theta_1,\theta_1+d\theta_1]$ and $[\theta_2,\theta_2+d\theta_2]$, and no levels in the interval $[\theta_1+d\theta_1,\theta_2]$. Also

$$S(\theta)=[d^2R/d\theta^2], \quad (98)$$

where $R(\theta)$ is the probability that a randomly chosen interval of length θ is empty of eigenvalues.

The connection between $R(\theta)$ and the cluster-functions T_n is

$$R(2\alpha)=\exp\Big\{-\sum_{n=1}^{\infty}\frac{1}{n!}\int\cdots\int_{-\alpha}^{\alpha}T_n(\theta_1,\cdots,\theta_n)$$

$$\times d\theta_1\cdots d\theta_n\Big\}. \quad (99)$$

In fact $R(2\alpha)$ is precisely the expression H given by

Eqs. (41) and (45), if we choose for $U(\theta)=V(\theta)=1+A(\theta)$ the function

$$U(\theta)=1(-\pi+\alpha<\theta<\pi-\alpha)$$
$$=0(\pi-\alpha<\theta<\pi+\alpha). \quad (100)$$

It is convenient to choose the center of the excluded interval to be at $\theta=\pi$, so that $U(\theta)$ is an even function in $(-\pi, +\pi)$. Then Eqs. (35), (36) can be used, and we deduce

$$R(2\alpha)=\det\left|\delta_{pq}-\frac{\sin[(p-q)\alpha]}{\pi(p-q)}-\frac{\sin[(p+q)\alpha]}{\pi(p+q)}\right|, \quad (101)$$

p and q taking the values $(\frac{1}{2}, \cdots, m-\frac{1}{2})$. This is the analog, for the circle, of Gaudin's Eq. (12),[5] which gives the corresponding expression for the Gaussian model with finite N. The analogy becomes even closer if we write Eq. (101) as

$$R(2\alpha)=\det\left|\delta_{pq}-\frac{1}{\pi}\int_{-\alpha}^{\alpha}\cos(p\theta)\cos(q\theta)d\theta\right|. \quad (102)$$

When $N\to\infty$, the determinant (102) becomes the Fredholm determinant of an integral equation, and our results coincide with those of Gaudin. We find that, for an infinite eigenvalue series with mean spacing $D=1$, the probability $E(x)$ that a random interval of length x be empty of eigenvalues is

$$E(x)=\prod_{j=1}^{\infty}(1-x\lambda_j^2), \quad (103)$$

where the λ_j are the eigenvalues of the integral equation

$$\lambda F(y)=\int_0^1\cos(\tfrac{1}{2}\pi xyz)F(z)dz. \quad (104)$$

Gaudin[3] has shown how to use Eq. (103) for the practical computation of $E(x)$.

A different application of the Gaudin-Mehta method is made by choosing, instead of Eq. (100),

$$U(\theta)=V(\theta)=1(-\pi+\alpha<\theta<\pi-\alpha)$$
$$U(\theta)=0, \quad V(\theta)=2(\pi-\alpha<\theta<\pi+\alpha). \quad (105)$$

Let $R'(2\alpha)$ denote the expression H resulting from this choice. $R'(\theta)$ is then the probability that a randomly chosen interval of length θ will contain *not more than one eigenvalue*. The function

$$S'(\theta)=[d^2(R+R')/d\theta^2] \quad (106)$$

is the probability distribution for spacings between pairs of *next-nearest neighbors*. Equations (35) and (36) now give

$$R'(2\alpha)=\det\left|\delta_{pq}-\frac{\sin(p-q)\alpha}{\pi(p-q)}+\frac{\sin(p+q)\alpha}{\pi(p+q)}\right| \quad (107)$$

$$=\det\left|\delta_{pq}-\frac{1}{\pi}\int_{-\alpha}^{\alpha}\sin(p\theta)\sin(q\theta)d\theta\right|. \quad (108)$$

The symmetry between Eqs. (102) and (108) is remarkable, and we do not understand why it exists.

When $N\to\infty$, the limit of $R'(2\alpha)$ is $E'(x)$, the probability for an interval x to contain not more than one eigenvalue in a series with mean spacing $D=1$. Equation (108) gives the result

$$E'(x)=\prod_{j=1}^{\infty}(1-x\mu_j^2), \quad (109)$$

where the μ_j are the eigenvalues of the integral equation

$$\mu F(y)=\int_0^1\sin(\tfrac{1}{2}\pi xyz)F(z)dz. \quad (110)$$

Gaudin's method would allow one to compute $E'(x)$, and hence the next-nearest neighbor spacing distribution, numerically.

Now comes a still more peculiar coincidence. Let $R_u(2\alpha)$ be the probability for an interval (2α) to be empty, in an eigenvalue series taken from the unitary probability-distribution (77). Then Eq. (78) gives

$$R_u(2\alpha)-\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\int\cdots\int_{-\alpha}^{\alpha}\det|s_N(\theta_j-\theta_k)|\times d\theta_1\cdots d\theta_n. \quad (111)$$

Using Eq. (48) and standard theorems from the algebra of determinants, Eq. (111) becomes

$$R_u(2\alpha)=\det\left|\delta_{pq}-\frac{\sin(p-q)\alpha}{\pi(p-q)}\right|$$
$$=\det\left|\delta_{pq}-\frac{1}{2\pi}\int_{-\alpha}^{\alpha}e^{i(p-q)\theta}d\theta\right|. \quad (112)$$

The determinant here is $(N\times N)$, the indices p and q taking the values $[-m+\frac{1}{2}, -m+\frac{3}{2}, \cdots, m-\frac{1}{2}]$. The similarity to Eqs. (102) and (108) is again striking. If we now add and subtract the rows and columns of the determinant (112) with indices $(\pm p)$, $(\pm q)$, we find the identity

$$R_u(2\alpha)=R(2\alpha)R'(2\alpha). \quad (113)$$

When $N\to\infty$, the limit of $R_u(2\alpha)$ is $E_u(x)$, the probability for an interval x to be empty in an infinite eigenvalue series with mean spacing $D=1$ in the unitary ensemble. Equation (112) then reduces to the

[5] Page 450 of the third paper in reference 2.

Fredholm determinant

$$E_u(x)=\prod_{j=1}^{\infty}(1-\epsilon_j x\nu_j^2), \tag{114}$$

where the ν_j are the eigenvalues of the integral equation

$$\nu F(y)=\tfrac{1}{2}\int_{-1}^{1}\exp(\tfrac{1}{2}\pi ixyz)F(z)dz, \tag{115}$$

and the ϵ_j are $(+1)$ or (-1) according as the corresponding eigenfunction is even or odd. The even ν_j are identical with the λ_j satisfying Eq. (104), while the odd ν_j are equal to $(i\mu_j)$ with μ_j satisfying Eq. (110). Therefore Eqs. (103), (109), and (114) satisfy the identity

$$E_u(x)=E(x)E'(x), \tag{116}$$

which is just the limit of Eq. (113) as $N\to\infty$.

The meaning of Eq. (113) can be illustrated in a concrete way as follows. Let $\sum_u$ be an eigenvalue series of order $N=2m$, taken from the unitary probability distribution (77). Let $\sum_M$ be another eigenvalue series of order N, constructed according to the following recipe: take two independent eigenvalue series $\sum_1$ and $\sum_2$, each of order N and belonging to the usual probability-distribution (1), superimpose the two series, and then pick out alternate eigenvalues from the mixed series. In view of a certain biological analogy, the suffix M may here be considered to stand for the word "meiosis." Now the product $R(2\alpha)R'(2\alpha)$ is just the probability that a random interval of length (2α) contains no term of the series $\sum_M$. Therefore Eq. (113) has the following meaning: *the distributions of level-spacings in the series $\sum_u$ and $\sum_M$ are identical.*

This property of the series $\sum_M$ suggests that we also examine its 2-level cluster function $T_{2M}(\alpha)$, which can be derived quite easily from the results of Sec. III. The construction of $\sum_M$ gives the formula

$$T_{2M}=\left(\frac{N}{2\pi}\right)^2-(\sum R_{2j})(\sum P_{2k+1})$$
$$-(\sum R_{2j+1})(\sum P_{2k})-2(\sum Q_{2j})(\sum Q_{2k+1}). \tag{117}$$

Here $P_j(\alpha)$ is defined to be the probability, for the separate eigenvalue series $\sum_1$ or $\sum_2$, that an interval α shall contain precisely j eigenvalues. $Q_j(\alpha)$ is the probability that an interval α, with one end point at an eigenvalue of $\sum_1$, shall contain j additional eigenvalues. $R_j(\alpha)$ is the probability that an interval α has eigenvalues of $\sum_1$ at both end points and j additional eigenvalues in its interior. From Eqs. (47) and (62) it is easy to compute

$$\sum_0^{m-1}R_{2j}=\tfrac{1}{2}(k^2-s_N{}^2+Is_NDs_N-Ds_N),$$
$$k=(N/2\pi), \tag{118}$$

$$\sum_0^{m-2}R_{2j+1}=\tfrac{1}{2}(k^2-s_N{}^2+Is_NDs_N), \tag{119}$$

$$\sum_0^{m-1}Q_{2j}=\tfrac{1}{2}(k+s_N), \tag{120}$$

$$\sum_0^{m-1}Q_{2j+1}=\tfrac{1}{2}(k-s_N), \tag{121}$$

$$\sum_0^{m}P_{2j}=1-Is_N, \tag{122}$$

$$\sum_0^{m-1}P_{2j+1}=Is_N, \tag{123}$$

the notations being defined by Eqs. (48)–(50). Substituting these expressions into Eq. (117) and using (81), we find

$$T_{2M}(\alpha)=s_N{}^2=T_{2u}(\alpha). \tag{124}$$

So the series $\sum_M$ and $\sum_u$ have identical 2-level correlations. This fact is additional and distinct from the identity of their spacing distributions.

In view of the foregoing, we make the general conjecture that *all statistical properties of the eigenvalue series $\sum_M$ and $\sum_u$ are identical.* We can find no general argument to explain why this conjecture should be true; but, if it were false, the identities (113) and (124) would become even more mysterious than they already are.

Note added in proof: This conjecture has subsequently been proved by Dr. J. Gunson of the University of Birmingham, England.

Cluster Sums for the Ising Model

G. S. RUSHBROOKE AND H. I. SCOINS
University of Durham, King's College, Newcastle upon Tyne, England

(Received July 5, 1961)

Expressions are derived for the first seven irreducible cluster sums, $\beta_1, \cdots, \beta_7$, for plane-square, simple-cubic, and body-centered cubic Ising models. The connection between these expressions and high-temperature expansions of the zero-field partition function and zero-field inverse susceptibility is indicated. Comparison is made with the work of other authors.

1. INTRODUCTION

THE calculations reported here were in fact completed some six years ago but have not hitherto been published. The delay has been due not only to the pressure of other work but also, perhaps primarily, to the forbidding demands of any adequate presentation. Simply to give the final answers, without any indication of how they were obtained, would seem both presumptuous and uninteresting. On the other hand it is certainly not feasible, and probably not helpful, to transcribe the computations in detail. We shall therefore attempt a compromise, less adequate than we would wish, by indicating the main route followed while not elaborating every small complication or sophistication. The essential techniques used are, we believe, not without value in casting light on the structure of the Ising problem.

In language appropriate to a lattice gas, we are primarily concerned with calculating successive virial coefficients (as functions of temperature). We prefer, however, to use the language of either regular solution theory or of ferromagnetism. Starting with the definition of $\Lambda(\alpha,\eta)$, namely,

$$\sum_{N_A}\sum_{N_{AB}} g(N;N_B,N_{AB})\alpha^{N_B}\eta^{N_{AB}}=[\Lambda(\alpha,\eta)]^N, \quad (1)$$

where $g(N;N_B,N_{AB})$ is the number of arrangements of N_A A systems and N_B B systems on a lattice of $N_A+N_B(=N)$ sites, for which there are N_{AB} $A-B$ contacts, we know that the Ursell-Mayer formalism (see, for example, Rushbrooke and Scoins[1]) leads first to the equation

$$\ln\Lambda(\alpha,\eta)=\sum_{l\geq 1}(\alpha\eta^z)^l b_l, \quad (2)$$

where

$$Nb_l=\frac{1}{l!}\,\mathrm{S}_{(l)}\sum\prod f_{ij}(1)(2)\cdots(l), \quad (3)$$

and thence, defining x by

$$x=\alpha(\partial/\partial\alpha)\ln\Lambda(\alpha,\eta), \quad (4)$$

to the equations

$$\ln\Lambda(\alpha,\eta)=x\left(1-\sum_{k\geq 1}\frac{k}{k+1}\beta_k x^k\right) \quad (5)$$

and

$$\ln\alpha=\ln x-z\ln\eta-\sum\beta_k x^k, \quad (6)$$

where

$$N\beta_k=\frac{1}{k!}\,\mathrm{S}_{(k+1)}\sum{}^*\prod f_{ij}(1)(2)\cdots(k+1). \quad (7)$$

In (2) and (6), z is the lattice coordination number. The sums in (3) and (7) refer to sums over all conceivable positions on the lattice, including multiple occupancy of sites, of l and $k+1$ B systems, respectively. $\sum\prod f_{ij}$ denotes the sum of all products of f_{ij} factors whereby the l systems are at least singly connected, while $\sum^*\prod f_{ij}$ denotes the sum of all such products whereby the $k+1$ systems are always multiply connected (except for $k=1$). The basic factor in the integrands f_{ij} is defined by

$$f_{ij}=\begin{cases}-1 \text{ if systems } i \text{ and } j \text{ are on the same lattice site,}\\ \eta^{-2}-1 \text{ if they are on adjacent lattice sites,}\\ 0 \text{ otherwise.}\end{cases}$$

We shall denote $\eta^{-2}-1$ by f. The factor η, in (1), is of course the Boltzmann factor usually denoted by $\exp(-w/kT)$, where w is the energy gained by creation of an extra $A-B$ contact for the same values of N_A and N_B. In the regular solution problem x [defined by (4)], measures the equilibrium concentration $\bar{N}_B/(\bar{N}_A+\bar{N}_B)$; for ferromagnetism, the magnetization I is given by $I=I_{\max}(1-2x)$, where $I_{\max}$ is the saturation magnetization.

Our aim has been to compute as many of the cluster sums, b_l or β_k, as practicable without inordinate labor. We have found it possible for open lattices (i.e., lattices not involving triangles of nearest neighbors) to proceed as far as β_7 and b_8. Fuchs,[2] who first applied the Ursell-Mayer formalism to the regular solution problem, calculated β_1, β_2, β_3, and β_4 for the body-centered cubic lattice. Wakefield,[3] in his extensive calculations on the simple-cubic lattice, gave low-temperature results equivalent to a knowledge of $\beta_1\cdots\beta_6$; and Domb and Sykes[4] have since published additional low-temperature terms equivalent to a knowledge of β_7. Katsura[5] has recently published $\beta_1\cdots\beta_7$, and $b_1\cdots b_8$, for the plane-

[1] G. S. Rushbrooke and H. I. Scoins, Proc. Roy. Soc. (London) **A230**, 74 (1955).

[2] K. Fuchs, Proc. Roy. Soc. (London) **A179**, 340 (1942).

[3] A. J. Wakefield, Proc. Camb. Phil. Soc. **47**, 419 (1951).

[4] C. Domb and M. F. Sykes, Proc. Roy. Soc. (London) **A235**, 247 (1956).

[5] S. Katsura, Progr. Theoret. Phys. (Kyoto) **20**, 192 (1958).

square lattice. Apart from trivial misprints (noted below), our results confirm all these earlier calculations. It may justly be held that it is only for the body-centered cubic lattice that anything much has been gained, but our approach has been very different from that of either Wakefield or Katsura. In particular, it is such that all three lattices are covered by the same basic calculations.

2. BASIC PROCEDURE

The starting point of the calculations is the theorem, a proof of which we have already published,[1] that $\sum^* \prod f_{ij}$ on the right-hand side of (7) vanishes identically unless the lattice sites occupied by the systems $1, 2, \cdots, k+1$, referred to subsequently as the occupied sites, are themselves topologically multiply connected by "bonds" on the lattice; a "bond" between two sites signifies that they are nearest neighbors. This enables us to split up the calculation of any particular β into the sum of contributions from different sets of occupied sites, and at the same time confine attention to such multiply-connected sets of sites on the lattice.

But we can go further than this. Suppose that we construct models, as we in fact did for the body-centered cubic lattice, of multiply-connected sets of lattice sites occurring on the lattice. These are to be three-dimensional models, reproducing lattice sites with bonds joining those which are nearest neighbors. For β_7 we require all models up to and including those having eight sites (and also the special cases of a single site and a pair of neighboring sites). Let us label the different models by a subscript i, reserving $i=1, 2$ for the two special cases just noted. Let Nm_i be the number of ways in which the ith model, with its vertices regarded as labeled, can be placed, conceptually, on the lattice, (e.g., $m_1=1$, $m_2=z$). Now consider the computation of (7). We are concerned with $k+1$ labeled systems; first take that part of (7) in which they occupy, in an unspecified way, p specific lattice sites. Since multiple occupancy of sites is allowed, $p \leqslant k+1$. Then, by the above theorem, these p sites lie on a model, say the ith, in one of its possible positions on the lattice. Now suppose we specify the way in which the labeled systems occupy the p specific lattice sites; let the subscript a specify such a definite assignment. Then for this assignment of definite systems to definite sites $\sum^* \prod f_{ij}$ will be a polynomial in f, say $P_{i,k,a}(f)$. As the notation indicates, this polynomial depends only on the model i and the particular assignment of $k+1$ labeled systems to the p vertices of this model themselves regarded as distinguishable. Thus

$$N\beta_k = \frac{1}{k!} \sum_i \frac{Nm_i}{s_i} \sum_a P_{i,k,a}(f), \tag{8}$$

where s_i is the number of ways in which model i with labeled vertices can be superimposed on a model i whose vertices are not labeled (i.e., the number of elements in the symmetry group of an unlabeled model i). We have to divide by s_i on the right-hand side of (8), since m_i was defined for a model with labeled vertices and this labeling is supplied only by the $k+1$ systems, over all arrangements of which we are summing.

TABLE I. Multiply connected maps of up to eight points for open lattices with their appropriate weights.

t	Map	χ(ps)	χ(sc)	χ(bcc)	t	Map	χ(ps)	χ(bc)	χ(bcc)
1	•	1	1	1	17			12	8
2	⟷	2	3	4	18				24
3		1	3	12	19				24
4				12	20				24
5				3	21				24
6			4	4	22				120
7		2	18	96	23				288
8				24	24				6
9			8	8	25				48
10				168	26				24
11				48	27				48
12				24	28				72
13		1	27	216	29				4
14			48	168	30			1	
15		2	42	408					
16		4	84	744					

But the polynomial $P_{i,k,a}(f)$ depends only on the topological properties of the model i, not its spatial geometrical structure. The number of distinct (i.e., not conceptually superimposible) unlabeled models having a given number of vertices can be very large. A full account of these calculations would, of course, have to detail them and would involve drawing them out in perspective. Both brevity and generality, however, are achieved by grouping together models of the same topological type. To do this we represent each model by a two-dimensional point-line map and preserve only those maps which are topologically distinct. Any particular map may represent numerous models, but for all these models the polynomial $P_{i,k,a}(f)$ is the same. We number the distinct maps by the subscript t, where again 1 and 2 have their special usage. Thus (8) is

written, more appropriately,

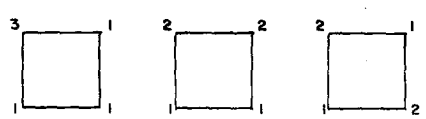
$$\beta_k=\frac{1}{k!}\sum_t \nu_t \sum_a P_{t,k,a}(f), \tag{9}$$

where $\nu_t=\sum m_i/s_i$ over all models i corresponding to map t. The suffix a in (9) refers now, of course, to assignments of $k+1$ labeled systems to the p vertices of map t, themselves regarded as distinguishable.

The maps t are, except for $t=1$ and 2, multiply-connected point-line graphs (which could be drawn without any initial construction of models). The numbers ν_t, on the other hand, are specific lattice parameters; in a vague sense, made precise above, they tell how often a given map occurs on the lattice. Table I lists all the maps involving 8 or fewer points occurring on the plane-square, simple-cubic, or body-centered cubic lattices. This restriction to open lattices is not essential, but implies confining attention to graphs not involving triangles. And against each map is given the appropriate value, if nonzero, of ν_t for these three lattices, signified by ps, sc, and bcc, respectively.

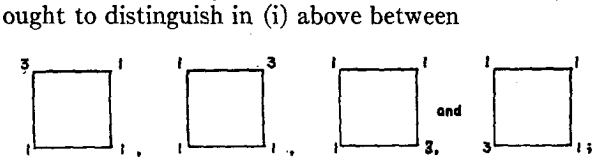
It is perhaps worth commenting that for the body-centered cubic lattice we have 17 maps involving 8 sites ($t=13$–29); there were 117 equivalent models.

So much for the factors ν_t in Eq. (9). It remains to consider further the polynomials $P_{t,k,a}(f)$, which are tied to maps and have, in themselves, nothing to do with any specific lattice. A partial specification of the assignment a can be made by distribution numbers. Labeling the vertices of the tth map 1, 2, $\cdots$, p, let n_v systems occupy vertex v, where, of course, $n_v\geqslant 1$, $v=1, 2, \cdots, p$. Then there are $(k+1)!\prod_v n_v!$ possible assignments of the $k+1$ systems which accord with the specification $n_1, n_2, \cdots, n_p$, and for all these assignments $\sum^*\prod f_{ij}$ has the same value, since the names of the systems involved in a do not affect $P_{t,k,a}(f)$. An illustration may now be helpful. Consider the case $t=3$ (map represented by a square) and $k=5$ (6 systems); then three basically different allocations of distribution numbers are as shown in (i), (ii), and (iii). But since,



(i) (ii) (iii)

so far, we have regarded the maps as having labeled vertices (ultimately, as localized on the lattice) we ought to distinguish in (i) above between



and, similarly, to distinguish 4 cases corresponding to (ii) and 2 corresponding to (iii). It is, however, convenient to regard only (i), (ii), and (iii) as basically different allocations, i.e., to lump together in one class all distributions (n) which differ only in the reallocation of the given distribution numbers to equivalent points in the map t. Distinguishing essentially different distributions by the subscript α, we must then introduce the weight $w(\alpha)$ to allow for this grouping. In the examples (i), (ii), and (iii) above, $w(\alpha)$ has, of course, the values 4, 4, and 2.

Returning to Eq. (9), this may now be written

$$\beta_k=\sum_t \nu_t F_{t,k}(f), \tag{10}$$

where

$$F_{t,k}(f)=(k+1)\sum_\alpha \frac{w(\alpha)}{\prod_v n_v!}P_{t,k,\alpha}(f). \tag{11}$$

$P_{t,k,\alpha}(f)$ is a polynomial in f, there being one such polynomial for each essentially different allocation (α) of distribution numbers (n) to the vertices (v) of the map t.

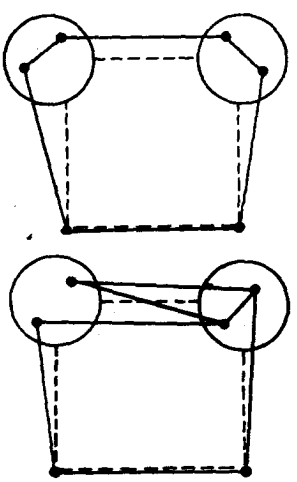


FIG. 1.

FIG. 2.

Of course we still have to calculate the polynomials $P_{t,k,\alpha}(f)$. It is not feasible to describe this work in detail, and the briefest illustration must suffice. Let us take the case (ii) above, so that we are concerned with $P_{3,5,(ii)}(f)$. It is convenient to represent the map by a dotted diagram, multiply-occupied vertices of which are replaced by circles. Within these circles we put points representing systems occupying these vertices. These points, within a circle, must be regarded as labeled, i.e., distinguishable. Then we require to join up all the points of the figure by full lines which (i) either lie along dotted lines or are within circles and (ii) multiply connect all the points of the figure. Figures 1 and 2 illustrate two of the many ways of doing this. We now, from the definition of f_{ij}, associate a factor -1 with each full line lying within a circle, and a factor f with each other full line. For Figs. 1 and 2 above we have f^4 and $-f^6$, respectively. To avoid drawing topologically identical figures we give numerical weights, 4 in Fig. 1 and 8 in Fig. 2, on account of the (suppressed)

labeling of their points. And we take the sum of all such possibilities.

Although tedious, this work has not seemed to us particularly liable to error (we discuss checking procedures below). It is much simplified by a number of elementary observations, of which we shall give only one instance. Any figure containing, *inter alia*, the line graph shown in Fig. 3 can be disregarded. An extra line within the circle produces the factor -1 without otherwise affecting matters.

In principle, the problem of computing β_k from Eqs. (10) and (11) is now solved. In practice, however, the computations would hardly be feasible were it not for drastic simplifications produced by the symmetry of the Ising problem. We turn, therefore, to these symmetry considerations before tabulating the basic information concerning the polynomials $F_{t,k}(f)$ obtained by the methods just outlined.

3. SYMMETRY CONSIDERATIONS, AND BASIC $F_{t,k}(f)$ RESULTS

It is convenient first to dispose of the contributions $F_{1,k}(f)$ and $F_{2,k}(f)$. These do not need to be calculated by the methods of Sec. 2 above, for we know them on other grounds already. We have shown earlier[1] that retaining only the terms $t=1$ and 2, but preserving all values of k, is equivalent to the well-known quasi-chemical approximation. In fact $F_{1,k}(f)=-1/k$ and $F_{2,k}(f)=2\beta_k^{(2)}$, in our earlier notation,[1] where[6]

$$\sum_{k\geq 1}\frac{k}{k+1}\beta_k^{(2)}x^{k+1}=-\frac{1}{2}\ln\left[\frac{2(1-x)+\eta^2(2x-1+\xi)}{2}\right] \quad (12)$$

with $\xi=[1+4x(1-x)f]^{\frac{1}{2}}$. Differentiating (12) with respect to x, we find

$$x(1-x)\sum_{k\geq 1}k\beta_k^{(2)}x^{k-1}=\tfrac{1}{2}(1-\xi^{-1}),$$

whence

$$\sum_{k\geq 1}k\beta_k^{(2)}x^{k-1}=\sum_{n\geq 1}\frac{(-1)^{n+1}(2n-1)!}{n!(n-1)!}f^n x^{n-1}(1-x)^{n-1}. \quad (13)$$

Equation (13) enables us to draw up the table

$$\beta_1^{(2)}=f$$
$$2\beta_2^{(2)}=-3f^2$$
$$3\beta_3^{(2)}=3f^2+10f^3$$
$$4\beta_4^{(2)}=-20f^3-35f^4$$
$$5\beta_5^{(2)}=10f^3+105f^4+126f^5$$
$$6\beta_6^{(2)}=-105f^4-504f^5-462f^6$$
$$7\beta_7^{(2)}=35f^4+756f^5+2310f^6+1716f^7$$
$$8\beta_8^{(2)}=-504f^5-4620f^6-10\,296f^7-6435f^8,$$

which effectively disposes of the polynomials $F_{2,k}(f)$.

[6] This is Eq. (37) of Rushbrooke and Scoins.[1] Unhappily a misprint gives the power of x on the left-hand side as k rather than $k+1$; the other equations are, however, printed there correctly.

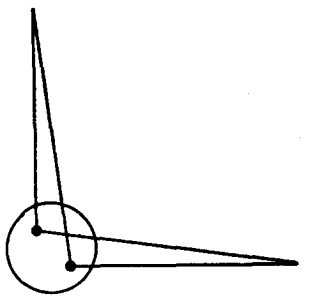


FIG. 3.

We turn now to $F_{t,k}(f)$ when $t\geq 3$. It is convenient to write

$$\sum_{t\geq 3}\nu_t F_{t,k}(f)=\delta_k=\sum_n \delta_{k,n}f^n.$$

By no means all the coefficients $\delta_{k,n}$ need to be computed by the methods described above. For in the language of regular solution theory, the free energy of mixing, ΔA, is given by

$$\Delta A/NkT=x\ln\alpha-\ln\Lambda(\alpha,\eta)$$

which, by (5) and (6), reads

$$\frac{\Delta A}{NkT}=x\ln x-xz\ln\eta-x-\sum_{t\geq 1}\sum_k\frac{\nu_t F_{t,k}(f)}{k+1}x^{k+1}$$

while on the quasi-chemical approximation

$$\frac{\Delta A^{QC}}{NkT}=x\ln x-xz\ln\eta-x-\sum_{t=1,2}\sum_k\frac{\nu_t F_{t,k}(f)}{k+1}x^{k+1}$$

so that

$$\frac{\Delta A-\Delta A^{QC}}{NkT}=-\sum_{t\geq 3}\frac{\nu_t F_{t,k}(f)}{k+1}x^{k+1}=-\sum_k\frac{\delta_k}{k+1}x^{k+1}. \quad (14)$$

But the left-hand side of (14) is symmetrical in x and $1-x$. We may therefore write

$$\sum_k\frac{\delta_k}{k+1}x^{k+1}=\sum_n d_n y^{n+1}, \quad (15)$$

where $y=x(1-x)$. It is easy to prove, by the methods of Sec. 2, that $\delta_{k,n}=0$ if $k>2n-1$. Allowing the possibility $k=2$, in order to give the argument in a form equally applicable to close-packed lattices, (15) may therefore be written

$$\sum_{k,n}\frac{\delta_{k,n}}{k+1}f^n x^{k+1}=\sum_{n,s}d_{n,s}f^n y^{n-s} \quad (0\leq s<n-2) \quad (16)$$

or

$$\delta_{k,n}=\sum_{s=0}^{n-3}d_{n,s}\frac{(k+1)(n-s)!(-1)^{k+s+1-n}}{(k+s+1-n)!(2n-2s-k-1)!}, \quad k\geq 2. \quad (17)$$

TABLE II. Structure of $\delta_{k,n}$ coefficients.

$k \backslash n$	4	5	6	7	8	9
3	A					
4	$-5A$	B	C			
5	$9A$	$-6B$	D	E	G	
6	$-7A$	$14B$	$-28C-7D$	F	H	L
7	$2A$	$-16B$	$104C+20D$	$-36E-8F$	K	M
8		$9B$	$-171C-30D$	$159E+27F$	$-165G-45H-9K$	N
9		$-2B$	$148C+25D$	$-325E-50F$	...	...
10			$-66C-11D$	$374E+55F$	...	...
11			$12C+2D$	$-250E-36F$	...	...
12				$91E+13F$	...	...
13				$-14E-2F$	...	...

Equation (17) shows that of the $2n-2$ values of $\delta_{k,n}$, $k=2, 3, \cdots 2n-1$, only $n-2$ (corresponding to $s=0, 1, \cdots n-3$) are independent quantities. We may, indeed, regard $\delta_{k,n}$, $k=n, n+1, \cdots 2n-1$ as determinate combinations of the values $\delta_{k,n}$, $k=2, 3, \cdots n-1$. This at once more than halves the computational labor involved in finding the irreducible cluster sums β_k. But the saving is in fact much greater than this since many of the coefficients $\delta_{k,n}$, $k=2, 3, \cdots n-1$, are zero. The situation for our open lattices is revealed by Table II. Below the lower stepped line $\delta_{k,n}=0$. And the values of $\delta_{k,n}$ corresponding to entries between the two stepped lines are linear combinations, determined from (17), of the values $A, B, \cdots, N$ which, up to $n=9$, are the only nonvanishing $\delta_{k,n}$ occurring above the upper stepped line.

Although presented only for the complete lattice, the above theory is found in fact to hold also for the contributions to $\delta_{k,n}$ from individual maps. A proper proof of this would require consideration of somewhat sophisticated trees and will not be given here. But the point may be illustrated by the particular case $t=3$ (for which a proof follows from simple pseudolattice arguments[1]). Writing

$$F_{t,k}(f)=\sum_t c_{t,k,n} f^n$$

so that

$$\delta_{k,n}=\sum_t \nu_t c_{t,k,n},$$

Table III lists those values of $c_{3,k,n}$ which were obtained by direct counting as described in Sec. 2. Taken in conjunction with Table II, this illustrates one form of checking which we have used on our results. Other checks are provided by the considerations of Sec. 4 below, and indeed the coefficients of all powers of f up to and including f^8 have been fully checked in one way or another. For higher powers of f it is necessary to rely on agreement between the two authors working independently.

TABLE III. Values of $c_{3,k,n}$.

$k \backslash n$	4	5	6	7	8	9
3	4					
4	-20	-40				
5	36	240	264			
6				-1456	-14	
7					K	288
8						N
9						
10						
11			528			

It will be observed that there are two gaps not yet filled in Table III, namely, the values of K and N. These values (7452 and $-38\,088$, respectively), together with a check on K, are provided by considerations based on the arguments of the next section. We would add here, however, that for all maps $t>3$ it is only the value of N which has been found by these other methods. And N is not necessary for the calculation of $\beta_1, \cdots, \beta_7$. It is, however, necessary to know N in order to have results which are complete up to and including terms in f^9; which we have felt to be desirable.

For maps $t>3$, A and B are zero. We conclude this section with values of the remaining coefficients, C, $D, \cdots, N$, together with any remaining terms, for all such maps. First, the basic coefficients for maps 4 to 12: Secondly, the remaining terms, not included above, for

TABLE IV. Basic coefficients for maps 4–12.

t	C	D	E	F	G	H	K	L	M	N
4	5	−54	−72	840	0	567	−7152	0	−3104	42 336
5	0	0	0	0	6	−126	1032	−112	2432	−21 168
6	0	6	0	−84	0	0	720	0	0	−4896
7	0	6	6	−154	0	−112	2120	0	1224	−21 276
8	0	12	24	−448	6	−546	7736	−112	6784	−89 424
9	0	0	0	0	0	21	−264	7	−504	6093
10	0	0	0	0	0	7	−88	7	−248	2637
11	0	0	0	0	0	35	−440	42	−1360	14 094
12	0	0	0	0	0	14	−176	28	−736	7092

these maps:

$t=4$, $c_{7,10}=-96$

$t=5$, $c_{7,10}=928$

$t=8$, $c_{7,10}=1120$

$t=9$, $c_{7,10}=-144$

$t=10$, $c_{7,10}=-160$

$t=11$, $c_{6,10}=7$, $c_{7,10}=-1072$, $c_{7,11}=-160$,

$t=12$, $c_{6,10}=7$, $c_{7,10}=-792$, $c_{7,11}=-160$.

Finally, for maps 13 to 30, we give their full contributions to $F_{t,7}(f)$:

$t=13$ $8f^8$

14 $8f^8+8f^9$

15 $8f^8+16f^9+8f^{10}$

16 $8f^8+16f^9+8f^{10}$

17 $8f^9+8f^{10}$

18 $16f^8+56f^9+40f^{10}+8f^{11}$

19 $8f^{10}+8f^{11}$

20 $8f^{10}+8f^{11}$

21 $8f^{10}+8f^{11}$

22 $8f^8+32f^9+32f^{10}+8f^{11}$

23 $16f^8+48f^9+40f^{10}+8f^{11}$

24 $16f^{10}+32f^{11}+8f^{12}$

25 $16f^8+80f^9+104f^{10}+48f^{11}+8f^{12}$

26 $32f^8+128f^9+160f^{10}+64f^{11}+8f^{12}$

27 $40f^{10}+48f^{11}+8f^{12}$

28 $24f^8+112f^9+144f^{10}+64f^{11}+8f^{12}$

29 $96f^8+568f^9+880f^{10}+480f^{11}+104f^{12}+8f^{13}$

30 $48f^8+256f^9+336f^{10}+96f^{11}+8f^{12}$.

For these maps, $t=13-30$, appropriately weighted, the total contributions to N for the plane-square, simple-cubic, and body-centered cubic lattices, respectively, are −2304, −66 960 and −1 151 280.

4. HIGH-TEMPERATURE EXPANSIONS AND CRITICAL CONDITIONS

Returning to Eqs. (14) and (16), and putting $x=\frac{1}{2}$, i.e., $y=\frac{1}{4}$, we have

$$\left(\frac{\Delta A^{QC}-\Delta A}{NkT}\right)_{x=\frac{1}{2}}=\sum_k \frac{\delta_k}{k+1}\left(\frac{1}{2}\right)^{k+1}=\sum_{n,s} d_{n,s} f^n\left(\frac{1}{4}\right)^{n-s}. \tag{18}$$

Introducing the high-temperature variable

$$u=(1-\eta)/(1+\eta)=\tanh\kappa,$$

where $\kappa=w/(2kT)$, so that $f=4u/(1-u)^2$, Eq. (18) becomes

$$\left(\frac{\Delta A^{QC}-\Delta A}{NkT}\right)_{x=\frac{1}{2}}=\sum_{n,s} 4^s d_{n,s}\frac{u^n}{(1-u)^{2n}}. \tag{19}$$

But, by (17), in terms of the basic symbols of Table II,

$d_{4,0}=\frac{1}{4}A$, $d_{5,0}=\frac{1}{5}B$, $d_{6,1}=\frac{1}{5}C$, $d_{6,0}=C+\frac{1}{6}D$, $d_{7,1}=\frac{1}{6}E$, $d_{7,0}=E+(1/7)F$, $d_{8,2}=\frac{1}{6}G$, $d_{8,1}=G+(1/7)H$, $d_{8,0}=\frac{9}{2}G+H+\frac{1}{8}K$, $d_{9,2}=(1/7)L$, $d_{9,1}=L+\frac{1}{8}M$, $d_{9,0}=5L+M+\frac{1}{9}N$,

and therefore, expanded in powers of u, (19) reads

$$\begin{aligned}\left(\frac{\Delta A^{QC}-\Delta A}{NkT}\right)_{x=\frac{1}{2}}&=\frac{1}{4}Au^4+\left(2A+\frac{1}{5}B\right)u^5+\left(9A+2B+\frac{9}{5}C+\frac{1}{6}D\right)u^6\\&+\left(30A+11B+\frac{108}{5}C+2D+\frac{10}{6}E+\frac{1}{7}F\right)u^7\\&+\left(\frac{165}{2}A+44B+\frac{702}{5}C+13D+\frac{70}{3}E\right.\\&\left.+2F+\frac{67}{6}G+\frac{11}{7}H+\frac{1}{8}K\right)u^8\\&+\left(198A+143B+\frac{3276}{5}C+\frac{182}{3}D+175E\right.\\&+15F+\frac{536}{3}G+\frac{176}{7}H+2K+\frac{79}{7}L\\&\left.+\frac{3}{2}M+\frac{1}{9}N\right)u^9+\cdots.\end{aligned} \tag{20}$$

But

$$\left(\frac{\Delta A^{QC}-\Delta A}{NkT}\right)_{x=\frac{1}{2}}=\ln\Lambda(1,\eta)-\ln\Lambda^{QC}(1,\eta),$$

since, in general and in the quasi-chemical approximation, $x=\frac{1}{2}$ corresponds to $\alpha=1$. We know, however, that

$$[\Lambda(1,\eta)]^N=\left(\frac{1+\eta}{2}\right)^{zN/2}2^N(1+\textstyle\sum_n a_n u^n), \quad (21)$$

where a_n=number of n-link lattice graphs with only even vertices (see Wannier[7]). And the quasi-chemical approximation corresponds to taking all $a_n=0$. Thus

$$\left(\frac{\Delta A^{QC}-\Delta A}{NkT}\right)_{x=\frac{1}{2}}=\frac{1}{N}\ln(1+\textstyle\sum_n a_n u^n)$$
$$=\textstyle\sum c_n u^n, \text{ say,} \quad (22)$$

But for an open lattice $a_n=0$ when n is odd; consequently for open lattices $c_n=0$ when n is odd, and the right-hand side of (20) is necessarily an even function of u.

Although the above theory has been presented for the full lattice coefficients A, B, $\cdots$, N of Table II, it does in fact hold also for the map coefficients of Tables III and IV, though a rather more sophisticated proof is required. It is easy to verify, in fact, that the data of Tables III and IV lead to zero contributions from u^5 and u^7 in (20). The values of N were fixed from the requirement that the coefficient of u^9 should also vanish.

Moreover, we can check the even powers of u as well as the odd. Equation (20) and the data of Tables III and IV lead to contributions to the right-hand side of (22) from individual maps. These contributions can in fact be computed directly in closed form from Eq. (21). We shall not go into details nor give these closed forms here, but simply present the terms up to u^{12} in contributions to $\ln\Lambda(1,\eta)$ from maps 1 to 12:

TABLE V. Contributions to $\ln\Lambda(1,\eta)$ from unweighted maps, $t=1-12$.

t	Contribution to $\ln\Lambda(1,\eta)$
1	$\ln 2$
2	$-\ln(1+u)$
3	$u^4-\frac{1}{2}u^8+\frac{1}{3}u^{12}$
4	$-3u^8+8u^{12}$
5	$-2u^8+32u^{12}$
6	$u^6-\frac{1}{2}u^{12}$
7	$u^6-u^8-2u^{10}+\frac{3}{2}u^{12}$
8	$2u^6-4u^8-10u^{10}+22u^{12}$
9	$-6u^{10}-4u^{12}$
10	$-4u^{10}+3u^{12}$
11	$-18u^{10}+9u^{12}$
12	$-12u^{10}+20u^{12}$

The remaining maps, 13–30, contribute 7, 207, and 2736 to the coefficient of u^8 in $\ln\Lambda(1,\eta)$ for the plane-square, simple-cubic, and body-centered cubic lattices, respectively.

We turn at this stage to the critical equation. It follows from (4) and (6) that

$$\left[\alpha\frac{\partial}{\partial\alpha}\alpha\frac{\partial}{\partial\alpha}\ln\Lambda(\alpha,\eta)\right]^{-1}=\frac{\partial\ln\alpha(x,\eta)}{\partial x}=\frac{1}{x}(1-\textstyle\sum k\beta_k x^k) \quad (23)$$

and, as previously,[1] we take the vanishing of $\partial\ln\alpha/\partial x$ at $\alpha=1$, i.e., $x=\frac{1}{2}$, as the equation defining a critical temperature, η_c. In the ferromagnetic application of the Ising model $\alpha=\exp(2\mu H/kT)$, where μ is the magnetic moment of a system and H the external magnetic field. In this case the left-hand side of (23) measures $4N\mu^2/\chi kT$, where χ is the susceptibility, and our condition $\partial\ln\alpha/\partial x=0$ at $\alpha=1$ is the condition for the vanishing of the zero-field reciprocal susceptibility, i.e., the normal condition for a Curie temperature.

From (10)

$$\frac{\partial\ln\alpha}{\partial x}=\frac{1}{x}+\frac{1}{1-x}-z\textstyle\sum_k k\beta_k^{(2)}x^{k-1}-\sum_k k\delta_k x^{k-1},$$

and we have shown (Rushbrooke and Scoins,[1] Sec. 6) that

$$\textstyle\sum_k k\beta_k^{(2)}(\frac{1}{2})^{k-1}=2(1-\eta).$$

Thus our critical condition becomes

$$4-2z(1-\eta)-\textstyle\sum k\delta_k(\frac{1}{2})^{k-1}=0. \quad (24)$$

As observed by Fuchs,[2] this equation gives poor convergence to η_c, on adding terms corresponding to increasing k. We therefore apply the transformations of Sec. 3 above (which discussion owes something in spirit, though not in detail, to Fuchs's work) and use (i) the symmetrical form in y rather than x, and (ii) the high-temperature variable u.

Differentiating (15) twice, we find

$$-\textstyle\sum k\delta_k(\frac{1}{2})^{k-1}=2\sum(n-s)d_{n,s}4^{s+1}\left(\frac{f}{4}\right)^n;$$

consequently,

$$\left(\frac{\partial\ln\alpha}{\partial x}\right)_{\alpha=1}=4-2z(1-\eta)+2\sum_{n,s}(n-s)d_{n,s}4^{s+1}\left(\frac{f}{4}\right)^n,$$

which in terms of u reads

$$\left(\frac{\partial\ln\alpha}{\partial x}\right)_{\alpha=1}=4-\frac{4zu}{1+u}+2\sum_{n,s}(n-s)d_{n,s}4^{s+1}\frac{u^n}{(1-u)^{2n}}. \quad (25)$$

Expressed in terms of the basic symbols of Table II, and using the language of ferromagnetism, (25) reads

[7] G. H. Wannier, Revs. Modern Phys. 17, 50 (1945).

$$N\mu^2/\chi kT=1-zu+zu^2-zu^3+(2A+z)u^4$$
$$+(16A+2B-z)u^5+(72A+20B$$
$$+20C+2D+z)u^6+(240A+110B$$
$$+240C+24D+22E+2F-z)u^7$$
$$+(660A+440B+1560C+156D$$
$$+308E+28F+160G+24H+2K+z)u^8$$
$$+(1584A+1430B+7280C+728D$$
$$+2310E+210F+2560G+384H+32K$$
$$+186L+26M+2N-z)u^9+\cdots. \quad (26)$$

The critical equation is simply the vanishing of this expression for $N\mu^2/\chi kT$.

Since this critical equation is linear in the coefficients ν_l, we are at liberty to speak about contributions to the critical equation from individual maps; but we shall not list these since, in general, we have not checked them by independent, alternative, calculations. We have done so, using the methods of our earlier paper,[1] for maps 3 and 6 (thereby obtaining a check on K for map 3); we believe similar calculations could be made for the other maps, though they would be laborious.

5. CLUSTER SUMS AND RELATED EXPRESSIONS

We list our results under the headings of the three lattices for which we have found them. There is no need to give explicitly the basic coefficients $A, B, \cdots, N$, since they can be read off from the appropriate expressions; the quasi-chemical approximation does not affect these coefficients, and K, for example, is simply the coefficient of f^8 in β_7. Besides giving the irreducible cluster sums β_k, in terms of f, we also give $\eta^{zl}b_l$ (which, following Katsura,[5] we denote by $\bar{b}_l$) in terms of η. These are computed algebraically from the well-known formula

$$b_l=\frac{1}{l!}\sum_{[\Sigma kn_k=l-1]}\prod\frac{(l\beta_k)^{n_k}}{n_k!}.$$

And we give the expressions for $\ln\Lambda(1,\eta)$ and $N\mu^2/\chi kT$ which, by (20) and (26), are the direct consequences of our irreducible cluster sums.

A. Plane-Square Lattice

Our expressions for $\beta_1, \cdots, \beta_7$, and the terms up to f^9 in β_8, agree precisely with those obtained by Katsura[5]; so do our expressions for $b_1, \cdots, b_8$ in terms of f. Regarding $\bar{b}_1, \cdots, \bar{b}_8$ in terms of η, we agree with Katsura except that

the coefficient of η^{18} in $\bar{b}_7$ is -5664 and not -5644,
the coefficient of η^{16} in $\bar{b}_8$ is $389/2$ and not 389,
the coefficient of η^{26} in $\bar{b}_8$ is 867 670 and not 87 670.

The latter figures are clearly due to printing errors in Katsura's paper.

Equation (20) gives

$$\ln\Lambda(1,\eta)=2\ln(1+\eta)-\ln 2+u^4+2u^6+\tfrac{9}{2}u^8+\cdots, \quad (27)$$

while (26) yields

$$N\mu^2/\chi kT=1-4u+4u^2-4u^3+12u^4-20u^5$$
$$+44u^6-84u^7+188u^8-372u,+\cdots. \quad (28)$$

B. Simple-Cubic Lattice

We find

$$\beta_1=-1+6f$$
$$\beta_2=-\tfrac{1}{2}-9f^2$$
$$\beta_3=-\tfrac{1}{3}+6f^2+20f^3+12f^4$$
$$\beta_4=-\tfrac{1}{4}-30f^3-112\tfrac{1}{2}f^4-120f^5$$
$$\beta_5=-\tfrac{1}{5}+12f^3+234f^4+871\tfrac{1}{2}f^5+924f^6+108f^7$$
$$\beta_6=-\tfrac{1}{6}-189f^4-2184f^5-6930f^6-7476f^7$$
$$-1890f^8+56f^9,$$
$$\beta_7=-(1/7)+54f^4+2568f^5+20\,460f^6+(57\,390+6/7)f^7$$
$$+62\,940f^8+21\,616f^9+288f^{10}+96f^{11}+8f^{12}$$
$$\beta_8=-\tfrac{1}{8}-1458f^5-31\,185f^6-192\,402f^7$$
$$-486\,261\tfrac{1}{4}f^8-535\,032f^9\cdots$$

whence

$$\bar{b}_1=\eta^6$$
$$\bar{b}_2=3\eta^{10}-3\tfrac{1}{2}\eta^{12}$$
$$\bar{b}_3=15\eta^{14}-36\eta^{16}+21\tfrac{1}{3}\eta^{18}$$
$$\bar{b}_4=3\eta^{16}+83\eta^{18}-328\tfrac{1}{2}\eta^{20}+405\eta^{22}-162\tfrac{3}{4}\eta^{24}$$
$$\bar{b}_5=48\eta^{20}+426\eta^{22}-2804\eta^{24}+5532\eta^{26}-4608\eta^{28}+1406\tfrac{1}{5}\eta^{30}$$
$$\bar{b}_6=18\eta^{22}+496\eta^{24}+1575\eta^{26}-22\,144\tfrac{1}{2}\eta^{28}+64\,574\eta^{30}$$
$$-84\,738\eta^{32}+53\,370\eta^{34}-13\,150\tfrac{2}{3}\eta^{36}$$
$$\bar{b}_7=8\eta^{24}+378\eta^{26}+3888\eta^{28}-1360\eta^{30}-157\,380\eta^{32}$$
$$+674\,652\eta^{34}-1\,261\,904\eta^{36}+1\,240\,035\eta^{38}$$
$$-628\,236\eta^{40}+(129\,919+1/7)\eta^{42}$$
$$\bar{b}_8=\eta^{24}+306\eta^{28}+4622\eta^{30}+22\,396\tfrac{1}{2}\eta^{32}-106\,113\eta^{34}$$
$$-947\,582\tfrac{1}{2}\eta^{36}+6\,392\,769\eta^{38}-16\,362\,155\tfrac{1}{4}\eta^{40}$$
$$+22\,521\,935\eta^{42}-17\,686\,675\tfrac{1}{2}\eta^{44}+7\,496\,787\eta^{46}$$
$$-1\,336\,290\tfrac{3}{8}\eta^{48}.$$

On taking the exponential of each side of Eq. (2), the above expressions for $\bar{b}_1, \cdots \bar{b}_7$ lead to terms up to those involving α^8 in $\Lambda(\alpha,\eta)$ identical with those found (up to α^7) by Wakefield[3] and (up to α^8) by Domb and Sykes.[4]

Equation (20) gives

$$\ln\Lambda(1,\eta)=3\ln(1+\eta)-2\ln 2+1+3u^4+22u^6$$
$$+192u^8+\cdots, \quad (29)$$

while (26) yields

$$N\mu^2/\chi kT=1-6u+6u^2-6u^3+30u^4-54u^5+318u^6$$
$$-726u^7+3726u^8-9718u^9+\cdots. \quad (30)$$

C. Body-Centered Cubic Lattice

We find

$$\beta_1=-1+8f$$

$$\beta_2=-\tfrac{1}{2}-12f^2$$

$$\beta_3=-\tfrac{1}{3}+8f^2+26\tfrac{2}{3}f^3+48f^4$$

$$\beta_4=-\tfrac{1}{4}-40f^3-310f^4-480f^5+60f^6$$

$$\beta_5=-\tfrac{1}{5}+16f^2+600f^4+3081\tfrac{3}{5}f^5+3408f^6+288f^7+162f^8$$

$$\beta_6=-\tfrac{1}{6}-476f^4-7392f^5-26\,152f^6-33\,264f^7-14\,238f^8+896f^9+504f^{10}$$

$$\beta_7=-(1/7)+136f^4+8544f^5+77\,040f^6+(257\,705+1/7)f^7+378\,408f^8+181\,280f^9-19\,072f^{10}+5376f^{11}+2000f^{12}+32f^{13}$$

$$\beta_8=-\tfrac{1}{8}-4824f^5-117\,120f^6-862\,632f^7-2\,798\,127f^8-4\,033\,584f^9 \text{ plus powers of } f \text{ up to } f^{15},$$

whence

$$\bar{b}_1=\eta^8$$

$$\bar{b}_2=4\eta^{14}-4\tfrac{1}{2}\eta^{16}$$

$$\bar{b}_3=28\eta^{20}-64\eta^{22}+36\tfrac{1}{3}\eta^{24}$$

$$\bar{b}_4=12\eta^{24}+204\eta^{26}-798\eta^{28}+948\eta^{30}-366\tfrac{1}{4}\eta^{32}$$

$$\bar{b}_5=12\eta^{28}+216\eta^{30}+1262\eta^{32}+9072\eta^{34}+17\,592\eta^{36}-14\,184\eta^{38}+4174\tfrac{1}{5}\eta^{40}$$

$$\bar{b}_6=27\eta^{32}+312\eta^{34}+2368\eta^{36}+4312\eta^{38}-92\,992\eta^{40}+275\,021\tfrac{1}{3}\eta^{42}-353\,640\eta^{44}+216\,036\eta^{46}-51\,444\tfrac{1}{2}\eta^{48}$$

$$\bar{b}_7=72\eta^{36}+704\eta^{38}+4404\eta^{40}+17\,616\eta^{42}-36\,348\eta^{44}-833\,064\eta^{46}+3\,795\,726\eta^{48}-7\,072\,736\eta^{50}+6\,798\,900\eta^{52}-3\,344\,712\eta^{54}+(669\,438+1/7)\eta^{56}$$

$$\bar{b}_8=4\eta^{38}+198\eta^{40}+2016\eta^{42}+10\,300\eta^{44}+41\,352\eta^{46}+55\,536\eta^{48}-989\,076\eta^{50}-6\,007\,194\eta^{52}+46\,866\,408\eta^{54}-122\,039\,509\eta^{56}+166\,096\,620\eta^{58}-127\,471\,458\eta^{60}+52\,501\,716\eta^{62}-9\,066\,913\tfrac{1}{8}\eta^{64}.$$

Our expressions for $\beta_1, \cdots \beta_4$ accord with those given by Fuchs except that the coefficient of f^2 in β_2 is -12 and not 12 as printed.[2]

Equation (20) gives

$$\ln\Lambda(1,\eta)=4\ln(1+\eta)-3\ln 2+12u^4+148u^6+2496u^8+\cdots, \quad (31)$$

in agreement with Trefftz,[8] while (26) yields

$$N\mu^2/\chi kT=1-8u+8u^2-8u^3+104u^4-200u^5+1880u^6-5288u^7+44\,072u^8-143\,560u^9+\cdots. \quad (32)$$

We end with a few comments on these results.

Equations (27), (29), and (31) are not particularly interesting, except as showing the extent to which our low-temperature cluster sums reproduce the high-temperature expansions. In every case more terms of these high-temperature series are in fact known (see Domb,[9] Sec. 4-5.3); indeed Rushbrooke and Eve[10] have recently computed terms in (29) up to that in u^{14}.

Equations (28), (30), and (32) are much more interesting. If we take the reciprocals of these equations, we find expressions for $\chi kT/N\mu^2$ in powers of u up to u^9. And such high-temperature expansions of χ in powers of u can be obtained directly by Oguchi's method.[11] In fact in 1957 Domb and Sykes[12] gave expressions equivalent to such expansions up to u^9 for our three lattices. We have exact agreement with these expansions. Very recently the same authors[13] have added two terms to the simple-cubic lattice series, and Sykes[14] has added six more terms to the series for the plane-square lattice. The matter of interest to which we would draw attention is that cluster sum work gives directly the expansion for χ^{-1}; Oguchi's high-temperature method gives the expansion for χ.

Finally, we would stress that there is a lot more information in the cluster sums β_k than in the corresponding high-temperature expansions for $\ln\Lambda(1,\eta)$ or χ^{-1}. In particular, we can use the cluster sum expressions $\bar{b}_l$ to discuss successive approximations to the specific-heat anomaly on the low-temperature side of the critical temperature. We can also derive, from the β_k's, the phase boundary. These numerical aspects of this work will be published separately in a joint paper with Dr. J. Eve.

[8] E. Trefftz, Z. Physik **127**, 371 (1950).
[9] C. Domb, Phil. Mag. suppl. **9**, 149 (1960).
[10] G. S. Rushbrooke and J. Eve, J. Math. Phys. **3**, 185 (1962).
[11] T. Oguchi, J. Phys. Soc. Japan **6**, 31 (1951).
[12] C. Domb and M. F. Sykes, Proc. Roy. Soc. (London) **A240**, 214 (1957).
[13] C. Domb, and M. F. Sykes, J. Math. Phys. **2**, 63 (1961).
[14] M. F. Sykes, J. Math. Phys. **3**, 52 (1961).

High-Temperature Ising Partition Function and Related Noncrossing Polygons for the Simple Cubic Lattice

G. S. RUSHBROOKE AND J. EVE
Physics Department, King's College, Newcastle upon Tyne, England, and
University of Durham Computing Laboratory, Newcastle upon Tyne, England
(Received August 29, 1961)

We have found the high-temperature expansion of the partition function for the simple-cubic lattice Ising problem up to the term involving u^{14}, where u is the high-temperature variable $\tanh w/2kT$. Comment is passed on an odd feature of the coefficients in this expansion and the corresponding expressions for the specific heat in terms of both u and $w/2kT$ are presented. The calculations involve machine counts of the numbers p_n of non-self-crossing lattice polygons; the method of obtaining these is described, and the value of p_{16} reported. The paper ends with a brief discussion of the trend of the numbers p_n and of the closely related noncrossing chain numbers c_n (using the data of Sykes).

1. INTRODUCTION

FOR the Ising problem, we may define the partition function per site, $\Lambda(\alpha,\eta)$, by the equation

$$\sum_{N_B}\sum_{N_{AB}} g(N;N_B,N_{AB})\alpha^{N_B}\eta^{N_{AB}}=[\Lambda(\alpha,\eta)]^N. \quad (1)$$

Here $g(N;N_B,N_{AB})$ is the number of arrangements of N_A A systems and N_B B systems on a lattice of $N_A+N_B(=N)$ sites, for which there are N_{AB} $A-B$ contacts; η is the differential Boltzmann factor often denoted by $e^{-w/kT}$, and for the zero-field ferromagnetic application of the Ising problem, $\alpha=1$. We are concerned here with this zero-field case, when it is well known[1] that for high temperatures (above the transition point) Eq. (1) can be written as

$$[\Lambda(1,\eta)]^N=\left(\frac{1+\eta}{2}\right)^{zN/2}2^N[1+\sum_n a_n u^n], \quad (2)$$

where $u=(1-\eta)/(1+\eta)=\tanh(w/2kT)$ and z is the coordination number of the lattice. In (2), a_n denotes the number of n-link lattice graphs with only even vertices, i.e., polygonal graphs whose lines join neighboring lattice sites, an even number of lines meeting at each vertex. Equation (2) gives an expansion of the zero-field partition function in powers of u and thence, if required, in powers of w/kT. It provides the simplest method of computing a high-temperature expansion for the anomalous specific heat above the transition temperature.

In this paper we are concerned exclusively with the simple cubic lattice. In this case, on taking the Nth root of each side of (2), we may write

$$\Lambda(1,\eta)=2[(1+\eta)/2]^3L(u), \quad (3)$$

where

$$L(u)=1+\sum \alpha_n u^{2n}, \quad (4)$$

the restriction to even powers of u in (4) arising from the open character of the lattice.

In 1951, Wakefield[2] published the expansion of (4) up to $n=6$ as

$$L(u)=1+3u^4+22u^6+192u^8+2070u^{10}+24\,943u^{12}+\cdots, \quad (5)$$

the coefficients up to and including that of u^8 being derived almost simultaneously by Oguchi.[3] In 1957, Domb and Sykes[4] repeated this calculation and obtained 2046 and 24 861 as the coefficients of u^{10} and u^{12}, respectively. A little later the present authors[5] also reexamined the problem and obtained as these coefficients 2046 and 24 853, the latter value being accepted as correct by Domb and Sykes.[6] We have now derived the coefficient of u^{14} in (4), and believe that the series reads

$$L(u)=1+3u^4+22u^6+192u^8+2046u^{10}+24\,853u^{12}+329\,406u^{14}+\cdots. \quad (6)$$

Much of the difficulty in finding these coefficients comes from the necessity of counting noncrossing, i.e., non-self-intersecting, closed polygons on the lattice. We do not believe that this is the only source of possible error, for what we call below composite (or ambiguous) figures also call for care in their enumeration, but it would indeed be a very formidable task to enumerate graphically all the 14-sided noncrossing polygons on a simple cubic lattice and one of us (J.E.) has devised a program for counting such polygons using high-speed digital computers. Our interest in this work in fact stemmed, in part, from this possibility.

In Sec. 2 we give some details of the calculations leading to the coefficients of Eq. (6), but owing to their lengthiness it is impracticable to report them fully. We then, in Sec. 3, draw attention to an unusual feature of some of the above coefficients (although this may, indeed, be without significance), and present the corresponding expressions for the anomalous specific heat, expanded in powers either of u or of κ $(=w/2kT)$.

In Sec. 4 we discuss the method of machine counting for noncrossing lattice polygons without entering into

[1] B. L. Van der Waerden, Z. Physik 118, 473 (1941).
[2] A. J. Wakefield, Proc. Cambridge Phil. Soc. 47, 419 (1951).
[3] T. Oguchi, J. Phys. Soc. Japan 6, 31 (1951).
[4] C. Domb and M. F. Sykes, Phil. Mag. 2, 733 (1957).
[5] G. S. Rushbrooke and J. Eve, J. Chem. Phys. 31, 1333 (1959).
[6] C. Domb, Phil. Mag. Suppl. 9, 276 (1960).

purely programing considerations. And we report the number of 16-sided noncrossing polygons on the simple cubic lattice.

We defer any numerical discussion of the Ising problem series for a later paper with Dr. H. I. Scoins,[7] in which we shall also deal with the corresponding numerical aspect of low-temperature series valid below the transition point. But in Sec. 5 we give a brief numerical analysis of the polygon counts, also employing the noncrossing chain results of Sykes.[8] This latter discussion closely parallels that recently given by Hiley and Sykes,[9] though our procedure is rather different from theirs. Our conclusions are close to, though not quite identical with, those of Hiley and Sykes.

2. HIGH-TEMPERATURE PARTITION FUNCTION

For the simple cubic lattice, we find that the coefficients a_n in Eq. (2) are given by

$$a_4=3N,\quad a_6=22N,$$
$$a_8=\tfrac{9}{2}N^2+(375/2)N,$$
$$a_{10}=66N^2+1980N,$$
$$a_{12}=\tfrac{9}{2}N^3+(1609/2)N^2+24\,044N,$$
$$a_{14}=99N^3+10\,065N^2+319\,242N,$$

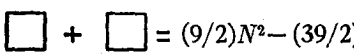
whence (6) follows either by taking the Nth root (by the binomial theorem) or on simply putting $N=1$.

The lattice diagrams contributing to a_4 are, of course, simply squares.

For a_6 we have three diagrams, which may be depicted as

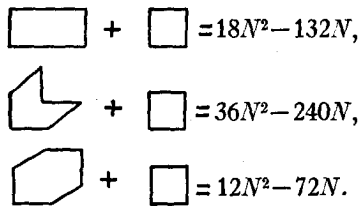
with weights 6, 12, and 4, respectively.

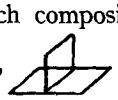
For a_8 we require the 8-sided (noncrossing) polygons, of which there are eleven different types (not distinguishing between mirror images). We shall not illustrate them; their total contribution is $207N$. And we have further to include diagrams which separate into two squares: symbolically, we find



this group including both "separable" as well as "separated" squares, a "separable" diagram decomposing into two allowed graphs on cutting at one vertex only.

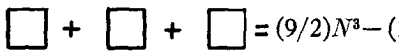
For a_{10} we require the 10-sided (noncrossing) polygons, of which there are seventy three different types (not distinguishing between mirror images). We shall not illustrate them; their total contribution is $2412N$. We have further to include diagrams which separate symbolically, giving contributions as follows:



These groups include "separable" as well as "separated" parts, but exclude "composite" figures which separate ambiguously into two parts on simultaneous cutting at two points. For a_{10} such composite diagrams are of one type only, namely,  , which contributes $12N$.

For a_{12}, the number of 12-sided (noncrossing) polygons has been determined mechanically by computer calculations, as $31\,754N$. The separated, or separable, but noncomposite diagrams give contributions as follows:

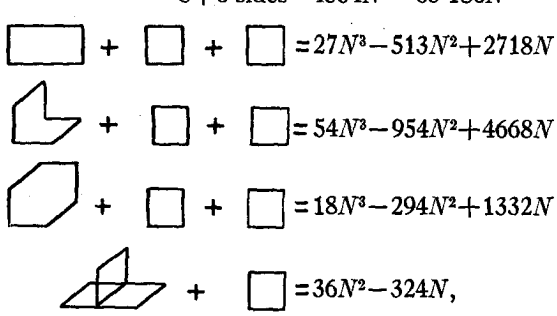
$$8+4\text{ sides}=621N^2-5718N$$
$$6+6\text{ sides}=242N^2-2486N$$



And "composite" figures, of which there are twelve different types, contribute $279N$.

For a_{14}, computer calculations give the number of 14-sided (noncrossing) polygons as $452\,640N$. The other contributions classify symbolically as follows (with the same conventions as employed above):

$$10+4\text{ sides}=7236N^2+84\,144N$$
$$8+6\text{ sides}=4554N^2-63\,180N$$



and one-part "composite" figures, of which there are 142 different types, contribute $5532N$.

We would add only that we both performed the calculations independently, redoing those cases in which there was any initial disagreement.

3. SPECIFIC HEAT AND RELATED SERIES

So far we have found

$$\Lambda(1,\eta)=2[(1+\eta)/2]^3[1+3u^4+22u^6+192u^8+2046u^{10}+24\,853u^{12}+329\,406u^{14}+\cdots]. \quad (7)$$

[7] G. S. Rushbrooke, H. I. Scoins, and J. Eve (to be published).
[8] M. F. Sykes, J. Math. Phys. 2, 52 (1961).
[9] B. J. Hiley and M. F. Sykes, J. Chem. Phys. 34, 1531 (1961).

Of more immediate thermodynamic importance is $\ln\Lambda(1,\eta)$, which can either be obtained algebraically from (7) or by simply retaining the coefficients of N in the above expressions for the quantities a_n. Thus

$$\ln\Lambda(1,\eta)=3\ln(1+\eta)-2\ln 2+3u^4+22u^6+187\tfrac{1}{2}u^8+1980u^{10}+24\,044u^{12}+319\,242u^{14}+\cdots. \quad (8)$$

For calculating the energy and specific heat, it is convenient to avoid odd powers of u and follow Wakefield[10] in writing the partition function as $\eta^{-\frac{1}{2}N}\Lambda(1,\eta)^N$. Then the free energy, A, is given by

$$A=-NkT[\ln 2-\tfrac{3}{2}\ln(1-u^2)+\ln L(u)]$$

$$=-NkT\ln 2-\tfrac{3}{2}NkT\left[u^2+\frac{5}{2}u^4+15u^6+\frac{501}{4}u^8+\frac{6601}{5}u^{10}+\frac{32\,059}{2}u^{12}+\frac{1\,489\,797}{7}u^{14}+\cdots\right], \quad (9)$$

whence the internal energy, U, and specific-heat, C, are given by

$$U=-\tfrac{3}{4}Nw(1-u^2)(d/du)[u^2+\tfrac{5}{2}u^4+\cdots] \quad (10)$$

and

$$C=3Nk(w/2kT)^2[1+11u^2+188u^4+2992u^6+51\,708u^8+930\,436u^{10}+17\,131\,724u^{12}+\cdots] \quad (11)$$

where $w/2kT=\tanh^{-1}u$. Denoting $\tanh^{-1}u$ by κ, in terms of κ

$$C/Nk=3\kappa^2+33\kappa^4+542\kappa^6+\frac{123\,547}{15}\kappa^8+\frac{14\,473\,442}{105}\kappa^{10}+\frac{11\,336\,607\,022}{4725}\kappa^{12}+\frac{605\,523\,385\,244}{14\,175}\kappa^{14}+\cdots. \quad (12)$$

We do not wish to pursue the numerical aspect of these series here, deferring till a later paper (which will discuss also the corresponding "low-temperature" series) any attempt to estimate the critical temperature. But we end this section by drawing attention to an odd feature of the coefficients in (7).

Writing $L(u)=1+\sum\alpha_n u^{2n}$, then for n odd, i.e., when $n=2m+1$, we have

$$\alpha_3 = 22=2\cdot 11,$$
$$\alpha_5 = 2046=2\cdot 3\cdot 11\cdot 31,$$
$$\alpha_7 = 329\,406=2\cdot 3\cdot 7\cdot 11\cdot 23\cdot 31,$$

and, apart from the 2, all the factors thus revealed are integers of the form $4k-1$. Now if the terms with n odd reflect a finite radius of convergence of $L(u)$ this property cannot persist without repetition of factors. It might in fact hardly be worth drawing attention to it were it not that if we assume that ultimately increasing m by unity produces the *four* new factors 3, 3, 7, and 7 (corresponding to $k=1$ and 2 each occurring twice), then the radius of convergence of these odd terms is given by $u^{-4}=(21)^2$, so that the critical temperature corresponds to $\eta=0.6417\cdots$. This is very close both to Wakefield's[10] value ($\eta=0.641$) and to the most recent estimate of Domb and Sykes[11] from susceptibility series[12] ($1/6u=0.7640\pm 0.0010$, i.e., $0.6414\leqslant\eta\leqslant 0.6422$).

4. MACHINE COUNTING OF NONCROSSING POLYGONS

The foregoing calculation of α_7 depends on a knowledge of p_{14}, where Np_n is the number of noncrossing n polygons (i.e., nondirectional and non-self-crossing closed chains of n links joining successive neighboring sites) on a simple cubic lattice of N sites. And any direct counting of p_{14} by graphical classification of polygon types, is virtually impracticable. The direct enumeration of p_{12} is perhaps just feasible, and was attempted by Wakefield, but for both p_{12} and p_{14} we have relied on machine counts (which have also checked the earlier p values). We have previously published[5] these machine values (and also $p_4\cdots p_{18}$ for the plane square lattice, which confirm the values found directly by Domb and coworkers,[9,13]) but have not hitherto described the process by which they were obtained. We have recently used it to find p_{16} for the simple cubic lattice, but although we do not see how to improve it, the procedure is essentially inefficient and further progress is impracticable with this method until very much faster machines are available. We estimate, indeed, that to find p_{18} for the simple cubic lattice would require not less than 120 hours of machine time on an IBM 704 computer.

We observe first that the number actually counted by the machine is t_n, where $t_n=2np_n/z$. This is the number of polygons passing through a given pair of adjacent lattice sites.[14] The essence of the method is now most simply described by considering the plane square lattice. From any point in the lattice we can move to a neighbor by one of four translation vectors, which we may denote by the symbols a, b, c, and d. To find t_n we systematically generate, in dictionary order, n-letter words starting with a, and containing the letters a, b, c, and d. At every stage of adding a letter to an incomplete word we record the coordinates of the

[10] A. J. Wakefield, Proc. Cambridge Phil. Soc. **47**, 799 (1951).

[11] C. Domb and M. F. Sykes, J. Math. Phys. **2**, 63 (1961).

[12] With factors 3, 3, 11, and 17 we should have $\eta=0.7299\cdots$, which is within the range ($0.7296\leqslant\eta\leqslant 0.7301$) given by Domb and Sykes[11] for the body-centered cubic lattice. The numbers are partly suggested by examining the $L(u)$ series for this lattice,[11] but there is rather more guesswork. Indeed, $u^{-2}=41$ would equally well yield $\eta=0.7298\cdots$.

[13] Reference 6, Sec. 5.4–4.

[14] For a given lattice we find that the time required is closely proportional to t_n.

point then reached on the lattice. If this point has already been reached in the present journey (i.e., while forming our present word) then, since self-crossing polygons are not allowed, the letter just recorded is inadmissible and we replace it by the next in alphabetical sequence before proceeding.

As it stands, this process simply generates all non-self-crossing chains (of n links), and we could find t_n by keeping a score of sequences rejected because the last step finished at the origin; but the process is prohibitively lengthy, and has been drastically reduced by two modifications.

The first is the obvious one allowed by the symmetry of the lattice. Calling the direction of the initial link the x axis, then the *first* time we leave the x axis we move off on one side only, and we multiply the final result by 2. This effectively halves the time that would otherwise be required. For the simple cubic lattice, we again allow only one way of initially moving off the x axis (and attach the weight 4 to this) and then allow only one way of initially moving out of the x-y plane thus defined (attaching the weight 2 to such a move). The reduction in time so achieved in the simple cubic case is almost by a factor of an eighth.

The second modification is perhaps less obvious, but equally important. For the plane square lattice, let the coordinates of the point reached after j steps (including the first one) by (x_j,y_j). Then it is impossible to return to the origin and thus have a contribution to t_n in $n-j$ more steps unless

$$|x_j|+|y_j|\leqslant n-j.$$

At each step, therefore, as well as checking that we have not revisited a previous site, we check that this condition is satisfied. For the simple cubic lattice, in an obvious notation, the condition is

$$|x_j|+|y_j|+|z_j|\leqslant n-j.$$

And this second modification effects more than a big reduction in the number of sequences actually generated by the machine. For at the nth (final) step we have only to check that this condition is satisfied to be sure that we have thus completed a noncrossing closed polygon.

With only minor modifications, but some loss of efficiency, the method is applicable to lattices other than the plane square and simple cubic; we have not, however, done machine calculations in other cases.

Some idea of the magnitude (or inefficiency) of the calculations is perhaps best conveyed by observing that for the simple cubic lattice the evaluation of p_{12} and p_{14} took approximately 3 hr and 50 hr, respectively, on a Pegasus computer. On an IBM 704, p_{14} (which we checked) took only 25 min, but p_{16} needed about 7 hr of machine time. The value thus obtained for p_{16} on the simple cubic lattice is 6 840 774.

5. ANALYSIS OF POLYGONS AND CHAINS

We conclude with a brief analysis of the trend of the polygon numbers p_n for the simple cubic lattice and a comparison of this with that of the corresponding chain numbers c_n. Here c_n denotes the number, per lattice site, of directional[15] non-self-crossing n-link chains (whose links join neighboring sites); and for the simple cubic lattice the values of c_n up to $n=11$ have been determined by Sykes,[8] using his elegant chain-counting theorem. An analysis of essentially the same data (lacking only the value of p_{16}, which in fact proves not very helpful) has recently been given by Hiley and Sykes[9]; for which reason we shall keep numerical details to a minimum. Our method of analysis, however, is quite different from that of Hiley and Sykes, and our conclusions differ from theirs slightly, but not significantly.

Of the numbers c_n and p_n (or $u_n\equiv 2np_n$), it is known with certainty from the work of Hammersley that[16]

$$c_n=\exp[kn+o(n)] \quad (n\to\infty)$$

and, for hypercubical lattice,[17]

$$u_n=\exp[kn+o(n)] \quad (n \text{ even}, n\to\infty).$$

We shall follow custom, however,[5,9,18] and assume the stronger, unproved, asymptotic forms

$$c_n\sim An^{\alpha}\mu^n \tag{13}$$

and

$$p_n\sim Bn^{\beta}\mu^n \text{ (for } n \text{ even)}, \tag{14}$$

endeavoring to estimate the parameters μ, α, and β.

For the polygons, the data which we are analyzing are conveniently summarized in the power-series

$$\begin{aligned}P(x)\equiv 1+\sum p_n x^n &= 1+3x^4+22x^6+207x^8+2412x^{10}+31\,754x^{12}\\ &\quad +452\,640x^{14}+6\,840\,774x^{16}+\cdots.\end{aligned}$$

If (14) were true for all n, rather than asymptotically, we should have

$$\tfrac{1}{2}\log(p_{n+2}/p_n)=\log\mu+\tfrac{1}{2}\beta\log[(n+2)/n] \tag{15}$$

and a plot of $\frac{1}{2}\log(p_{n+2}/p_n)$ against $\log[(n+2)/n]$ would be linear, its intersect with the axis $n\to\infty$ determining μ. The plot is indeed quite impressively linear (after the first point), but for an accurate determination of μ we need to extrapolate numerically. Denoting the linear extrapolant of the points corresponding to $n=m$ and $n=m+2$ by μ_{m+2}, we find $\mu_6=4.1588$, $\mu_8=4.9401$, $\mu_{10}=4.7654$, $\mu_{12}=4.6931$ and $\mu_{14}=4.6980$, from which it is clear that these p_n coefficients have not yet settled down to a quite steady

[15] For nondirectional chains we have simply to divide by 2; but we follow the notation of Sykes.
[16] J. M. Hammersley, Proc. Cambridge Phil. Soc. **53**, 642 (1957).
[17] J. M. Hammersley, Proc. Cambridge Phil. Soc. **57**, 516 (1961).
[18] M. E. Fisher and M. F. Sykes, Phys. Rev. **114**, 45 (1959).

behavior, and we may do better to determine μ from the chain data given by Sykes.[8]

By considering simply ratios of successive c's, rather than their logarithms, Hiley and Sykes conclude that certainly $4.676<\mu<4.694$, and adopt the estimate 4.680. We have ourselves used two methods: In the first we suppose that the subsequence c_n for n even should be treated separately from the subsequence c_n for n odd, and for each of these subsequences use the direct analog of the method just described. From this we conclude $\mu<4.690$, but very close thereto. In the second we plot $\log(c_{n+1}/c_n)$ against $\log[(n+1)/n]$ but extrapolate only on alternate points (which is the logarithmic analog of the method used by Hiley and Sykes). From this we conclude $\mu>4.680$, but very close thereto. In these circumstances our final conclusion must be simply

$$4.68<\mu<4.69,$$

with an over-all uncertainty dictated by the essential optimism of extrapolating to asymptotic behavior from limited data.

For estimating α, we have considered the alternative sequences of approximations provided by

$$\alpha_n=[\log(c_{n+2}/c_n)-2\log\mu]/\log[(n+2)/n] \quad (16)$$

and

$$\alpha_n=[\log(c_{n+1}/c_n)-\log\mu]/\log[(n+1)/n], \quad (17)$$

in both cases plotting α_n against $1/n$ to view the trend of successive estimates. We have done this for a range of values of μ, but shall concentrate here on the two values 4.680 and 4.690. If $\mu=4.680$, then we estimate $\alpha=0.179\pm0.002$, while if $\mu=4.690$ we estimate $\alpha=0.138\pm0.001$. Hiley and Sykes suggest $\alpha=0.17\pm0.03$, but we believe it is possible to be more precise (for a given value of μ).

For estimating β we use the analog of (16), p_n replacing c_n. If $\mu=4.680$, then we estimate $\beta=-2.74\pm0.03$, while if $\mu=4.690$ we estimate $\beta=-2.80\pm0.01$. Hiley and Sykes give -2.75 ± 0.05.

We end with two final comments. First, regarding the "index of initial ring closure,"[9] namely $1+\beta-\alpha$. This quantity is relatively insensitive to choice of μ; when $\mu=4.680$ we estimate -1.92 ± 0.03, while if $\mu=4.690$ we would favor a value in the upper third of this range. Hiley and Sykes have -1.92 ± 0.08; subject to the over-all optimism of any extrapolation, we believe it permissible to reduce their limits.

Secondly, regarding Fisher's proposal that α may be the reciprocal of an integer.[18] If this is true, then from the present data it seems impossible to decide between $\alpha=1/7$ ($\mu=4.689$) and $\alpha=1/6$ ($\mu=4.683$). For this, if for no other reason, we believe that even two more terms in the c_n series would be very valuable.[19,20]

ACKNOWLEDGMENTS

We acknowledge with gratitude the facilities provided both by the University of Durham Computing Laboratory (Ferranti Pegasus machine), and by the IBM Research Endowment Scheme, which gave us time on an IBM 704 computer.

[19] *Note added in proof*. Fisher and Hiley[20] have recently re-examined the c values, and conclude $\mu=4.683\pm0.007$, $\alpha=1/6$. From our analysis we believe that the evidence for $\alpha=1/7$ is at least as strong.

[20] M. E. Fisher and B. J. Hiley, J. Chem. Phys. **34**, 1253 (1961)

Surface Vibrational Modes in Crystal Lattices with Complex Interatomic Interactions*

DENOS C. GAZIS
General Motors Research Laboratories, Warren, Michigan

AND

RICHARD F. WALLIS
U. S. Naval Research Laboratory, Washington, D. C.

(Received August 17, 1961)

The investigation of surface vibrational modes in crystal lattices is complicated by the necessity for satisfying the free boundary conditions at the surface if the interatomic interactions include next-nearest neighbor interactions, next-next-nearest neighbor interactions, etc. The number of equations specifying the free boundary conditions may then become rather large. A general method is presented in this paper for the investigation of the normal modes of one-, two-, and three-dimensional lattices which are finite or semi-infinite in one dimension and have free boundary surfaces, assuming interatomic interactions of various ranges. The method is illustrated by calculations of the normal modes for the finite and semi-infinite diatomic linear chains with nearest and next-nearest neighbor interactions. Some remarks are made regarding the applications of the mathematical techniques to the evaluation of certain continuant determinants of large order.

I. INTRODUCTION

An understanding of the effect of free surfaces on the normal vibrational modes of a crystal is useful in the study of surface effects on experimentally measurable quantities such as specific heat and electrical conductivity. Investigations of surface modes of vibration using the continuum point of view have recently been reported for cubic crystals by Stoneley[1] and by Gazis, Herman, and Wallis.[2] A general treatment of surface waves for discrete lattices has been given by Lifshitz and Pekar.[3] Calculations based on specific lattice models have been given by Wallis[4,5] for diatomic one-, two-, and three-dimensional NaCl-type lattices with nearest neighbor interactions only, and by Kaplan[6] for the monatomic one-dimensional lattice with nearest and next-nearest neighbor interactions. Gazis, Herman, and Wallis[2] treated the semi-infinite three-dimensional monatomic simple cubic lattice with nearest and next-nearest neighbor central forces and with angular stiffness forces involving consecutive nearest neighbors forming a right angle at equilibrium.

The work of Wallis[4,5] and of Kaplan[6] makes use of certain mathematical methods for handling continuant determinants developed by Rutherford.[7,8] These methods, however, are useful only for very simple models and become quite cumbersome if not impossible to use when the model involves many types of atoms with many types of interactions. In the present paper a rather general method is presented for investigating the normal modes of vibration of one-, two-, and three-dimensional lattices which are finite or semi-infinite in one dimension and have free boundary surfaces. The method can handle models having more than one atom per unit cell and having many types of interactions between atoms. The method is illustrated by calculations of the normal modes for finite and semi-infinite diatomic linear chains with free ends and with nearest and next-nearest neighbor interactions.

For clarity in presentation the complete discussion of the one-dimensional lattice is given first and generalizations to the two- and three-dimensional lattices are indicated afterwards. Some remarks are made on the applicability of the present methods to the evaluation of certain continuant determinants of large order.

II. FORMULATION FOR THE ONE-DIMENSIONAL LATTICE

Consider a one-dimensional diatomic lattice formed by equidistant particles of alternating mass m and M. Assuming central force interactions of nearest and next-nearest neighbors, the equations of motion are

$$m\ddot{x}_n=\alpha(x_{n+1}+x_{n-1}-2x_n)+\beta(x_{n+2}+x_{n-2}-2x_n), \quad (1)$$

for particles of mass m, and

$$M\ddot{x}_n=\alpha(x_{n+1}+x_{n-1}-2x_n)+\gamma(x_{n+2}+x_{n-2}-2x_n), \quad (2)$$

for particles of mass M. In Eqs. (1) and (2) α, β, and γ are the force constants associated with central force interactions of nearest neighbors, next-nearest neighbors of mass m, and next-nearest neighbors of mass M, respectively. It might be noted that the distance between particles does not enter explicitly in Eqs. (1)

* A preliminary account of this work was presented at the New York meeting of the Americal Physical Society, February 1–4 (1961) [Bull. Am. Phys. Soc. 6, 31 (1961)].

[1] R. Stoneley, Proc. Roy. Soc. (London) A232, 447 (1955).
[2] D. C. Gazis, R. Herman, and R. F. Wallis, Phys. Rev. 119, 533 (1960).
[3] I. M. Lifshitz and S. I. Pekar, Uspekhi Fiz. Nauk 56, 531 (1955).
[4] R. F. Wallis, Phys. Rev. 105, 540 (1957).
[5] R. F. Wallis, Phys. Rev. 116, 302 (1959).
[6] H. Kaplan, Bull. Am. Phys. Soc. II, 2, 147 (1957).
[7] D. E. Rutherford, Proc. Roy. Soc. Edinburgh A62, 229 (1947).
[8] D. E. Rutherford, Proc. Roy. Soc. Edinburgh A63, 232 (1951).

TABLE I. Nature of roots of secular equation.

Case	y	θ
1	$0\leq y\leq 1$	$\theta=$real
2	$y>1$	$\theta=\pi/2+ib$, $b=$real
3	$y<0$	$\theta-ib$, $b=$real
4	$y=$complex	$\theta=$complex

and (2) but only affects the value of the force constants. As a consequence the present discussion would also apply to non-equidistant particles if one assumed that the interactions of particles of different masses were described by the same force constant on both sides of a given particle in spite of the different equilibrium distances.

Assuming a periodic solution given by

$$x_n=\begin{pmatrix}U_m\\U_M\end{pmatrix}\exp[i(\omega t+n\theta)], \tag{3}$$

Eqs. (1) and (2) yield

$$\begin{aligned}(m\omega^2-2\alpha-2\beta+2\beta\cos 2\theta)U_m+2\alpha\cos\theta U_M=0,\\(M\omega^2-2\alpha-2\gamma+2\gamma\cos 2\theta)U_M+4\alpha\cos\theta U_m=0,\end{aligned} \tag{4}$$

where U_m and U_M are the amplitudes of the particles having masses m and M, respectively.

The system of Eqs. (4) is homogeneous in U_m and U_M, and has a nontrivial solution if

$$(uv-4)-4(Bv+Cu-1)y+16BCy^2=0, \tag{5}$$

where

$$\begin{aligned}u=(m\omega^2/\alpha)-2,\quad v=(M\omega^2/\alpha)-2,\\B=\beta/\alpha,\quad C=\gamma/\alpha,\quad y=\sin^2\theta.\end{aligned} \tag{6}$$

The amplitude ratio of an eigenmode of the system is given by

$$\frac{U_m}{U_M}=-\frac{2\cos\theta}{u-4B\sin^2\theta}=-\frac{v-4C\sin^2\theta}{2\cos\theta}=k. \tag{7}$$

where the values of frequency ω and wave number θ, entering Eqs. (7), satisfy the secular equation given by Eq. (5). The preceding equations describe waves and vibrations in an infinite lattice. It is seen that Eq. (5) is a quadratic in ω^2 or $\sin^2\theta$. Hence, to a given frequency ω correspond two wave numbers θ, which may be real, imaginary, or complex depending on the nature of the roots y of Eq. (5), the dependence being shown in Table I.

In an infinite lattice only real wave numbers are physically acceptable since an imaginary or complex θ results in unbounded amplitudes of displacement at infinity. Thus the frequency ranges for which θ is not real are forbidden bands for the infinite lattice. However, nonreal θ's are acceptable for bounded lattices such as a semi-infinite or a finite lattice. It will be seen that an appropriate linear combination of the two modes associated with the two wave numbers satisfies the boundary conditions of such bounded lattices at some distinct eigenfrequencies.

The boundary conditions are derived by considering the bounded lattice as formed from an infinite one after the removal of an appropriate number of interactions at the boundary regions. Since nearest and next-nearest neighbor interactions have been assumed, a breakage of the lattice affects the interactions of the two last boundary particles. Thus two boundary conditions are associated with each boundary. The formulation of the eigenvalue problem of the bounded lattice can thus proceed as follows:

A. Finite Lattice

The finite lattice may contain an even number $2N$, or an odd number $2N+1$, of particles. For an odd number of particles it is expedient to assume the origin at the geometric center of the lattice, with the boundary particles at the $\pm N$th position. Thus we can take advantage of the fact that the symmetric and antisymmetric modes are uncoupled due to the complete symmetry of the system. For an even number of particles we do not have such symmetry and the origin shall coincide with one of the boundary particles in order to simplify certain expressions of the boundary conditions. However, it might be mentioned here that during the numerical computations for a moderately large number of particles the occurrence of imaginary and complex wave numbers necessitated the shifting of the origin near the center of the lattice in order to avoid certain computational difficulties pertaining to the involvement of very large numbers.

For an odd number of particles the end particles may be assumed of mass m without loss of generality. Let

$$x_n=\sum_{j=1}^{2}\begin{pmatrix}k_j\\1\end{pmatrix}A_j\cos n\theta_j e^{i\omega t}, \tag{8}$$

for the antisymmetric modes, and

$$x_n=\sum_{j=1}^{2}\begin{pmatrix}k_j\\1\end{pmatrix}B_j\sin n\theta_j e^{i\omega t}, \tag{9}$$

for the symmetric modes,[9] where θ_j are the two wave numbers associated with ω and k_j the corresponding amplitude ratios given by Eqs. (7). In Eqs. (8) and (9) the amplitudes A_j, B_j are multiplied by k_j or unity in order to give the displacements of particles having mass m or M, respectively.

[9] The symmetric modes are those in which the symmetry of the lattice about the center is preserved throughout the motion. This definition differs from that of reference 4.

The boundary conditions are obtained by setting the interactions of the Nth and $(N-1)$th particles with some fictitious extensions of the lattice equal to zero, namely,

$$\alpha(x_{N+1}-x_N)+\beta(x_{N+2}-x_N)=0,$$
$$\gamma(x_{N+1}-x_{N-1})=0. \tag{10}$$

Due to the symmetry the boundary conditions are also satisfied at the other end of the lattice.

Substituting Eqs. (8) and (9) in (10) we obtain two homogeneous linear equations in A_1, A_2 and B_1, B_2, respectively. A nontrivial solution is obtained when the determinants of these systems are zero. Thus one obtains the frequency equation

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}=0, \tag{11}$$

where

$$a_{1j}=\cos(N+1)\theta_j-k_j\cos N\theta_j$$
$$-2Bk_j\sin(N+1)\theta_j\sin\theta_j, \tag{12}$$
$$a_{2j}=\sin N\theta_j\sin\theta_j, \quad (j=1,2)$$

for the antisymmetric modes, and

$$a_{1j}=\sin(N+1)\theta_j-k_j\sin N\theta_j$$
$$+2Bk_j\cos(N+1)\theta_j\sin\theta_j, \tag{13}$$
$$a_{2j}=\cos N\theta_j\sin\theta_j, \quad (j=1,2)$$

for the symmetric modes.

The left-hand side of Eq. (11) is a transcendental function of the frequency only, since all the other quantities are either physical constants or functions of the frequency given by Eqs. (5) and (7). The roots of Eq. (11) have been evaluated numerically and the results of this evaluation are discussed in subsequent sections.

For an even number of particles it is necessary to include both symmetric and antisymmetric distributions of displacements. Furthermore, the boundary conditions are neither symmetric nor antisymmetric and, hence, all four of them, two for either end, are taken into account. The displacements are given by

$$x_n=\sum_{j=1}^{2}\binom{k_j}{1}(A_j\cos n\theta_j+B_j\sin n\theta_j)e^{i\omega t}, \tag{14}$$

and the boundary conditions for the 0th, 1st, $(2N-2)$th, and $(2N-1)$th particles are

$$\begin{aligned}\alpha(x_{-1}-x_0)+\beta(x_{-2}-x_0)&=0,\\ \gamma(x_{-1}-x_1)&=0,\\ \beta(x_{2N}-x_{2N-2})&=0,\\ \alpha(x_{2N}-x_{2N-1})+\gamma(x_{2N+1}-x_{2N-1})&=0.\end{aligned} \tag{15}$$

Substituting Eqs. (14) in (15) we obtain, after certain reductions using Eqs. (7), a homogeneous system in A_j, B_j. A nontrivial solution is obtained if the determinant of this system is zero. This condition yields the frequency equation

$$\begin{vmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \\ a_{31} & a_{32} & b_{31} & b_{32} \\ a_{41} & a_{42} & b_{41} & b_{42} \end{vmatrix}=0, \tag{16}$$

where

$$\begin{aligned} a_{1j}&=\cos\theta_j-k_j-2Bk_j\sin^2\theta_j\\ b_{1j}&=-(\sin\theta_j+Bk_j\sin 2\theta_j)\\ a_{2j}&=0\\ b_{2j}&=\sin\theta_j\\ a_{3j}&=k_j\sin(2N-1)\theta_j\sin\theta_j \qquad (j=1,2)\\ b_{3j}&=-k_j\cos(2N-1)\theta_j\sin\theta_j\\ a_{4j}&=k_j\cos 2N\theta_j-\cos(2N-1)\theta_j-2C\sin 2N\theta_j\sin\theta_j\\ b_{4j}&=k_j\sin 2N\theta_j-\sin(2N-1)\theta_j+2C\cos 2N\theta_j\sin\theta_j. \end{aligned} \tag{17}$$

Equation (16) yields the natural frequencies of the system. For every such frequency the three amplitude ratios involving A_j, B_j can be computed from the compatible homogeneous linear system corresponding to Eqs. (15).

B. Semi-Infinite Lattice

The investigation of the semi-infinite lattice is performed for the purpose of investigating surface modes, i.e., modes with displacement amplitudes decaying rapidly away from the free end. If surface modes are physically possible, they may be obtained as limiting cases of modes of a finite lattice when the number of particles is increased indefinitely. However, the study of the semi-infinite lattice yields a more direct investigation of these modes.

The origin is taken at the end particle which can be assumed of mass m without loss of generality. Let the displacements be given by

$$x_n=\binom{U_m}{U_M}\exp(-qn+i\omega t). \tag{18}$$

Equation (18) can be obtained from (3) by means of the substitution

$$\theta=iq. \tag{19}$$

Consequently Eq. (5) yields the attenuation constants q_j, as functions of the frequency, where now

$$y=-\sinh^2 q. \tag{20}$$

For the amplitude ratio U_m/U_M we have, instead of

Eqs. (7), the equations,

$$\frac{U_m}{U_M}=-\frac{2\cosh q}{u+4B\sinh^2 q}=-\frac{v+4C\sinh^2 q}{2\cosh q}=k. \quad (21)$$

We take the displacements to be given by a linear superposition of the two modes associated with the two values of q_j obtained for a given frequency, namely,

$$x_n=\sum_{j=1}^{2}\binom{k_j}{1}A_j\exp(-q_j n+i\omega t). \quad (22)$$

The boundary conditions are

$$\begin{aligned}\alpha(x_{-1}-x_0)+\beta(x_{-2}-x_0)&=0,\\ \gamma(x_{-1}-x_1)&=0,\end{aligned} \quad (23)$$

which yield, in conjunction with Eq. (22), the frequency equation

$$[e^{q_1}-k_1(1+B)+Bk_1e^{2q_1}]\sinh q_2 -[e^{q_2}-k_2(1+B)+Bk_2e^{2q_2}]\sinh q_1=0 \quad (24)$$

in the manner already discussed for the case of the finite lattice.

It may be pointed out that a surface mode may be obtained only if both attenuation constants q_j are real or complex. Inspection of Table I and Eq. (19) indicates that this corresponds to cases 2, 3, or 4, i.e., the frequency lies in a forbidden band of the infinite lattice.

From the preceding discussion it is seen that all the boundary value problems for bounded lattices will be solved by an appropriate linear superposition of the two modes of the infinite lattice. Therefore, before the discussion of any particular boundary value problem, an investigation of the frequency dispersion curves of the infinite lattice is given in the following section.

III. FREQUENCY DISPERSION CURVES OF THE INFINITE DIATOMIC LATTICE

At this point the question of stability of the lattice will be discussed briefly. Stability is assumed to be associated with a positive definite quadratic form for the strain energy. In the case of a one-dimensional continuum, the stability condition degenerates simply to the requirement of a positive Young's modulus. We can express Young's modulus as a function of the force constants entering Eqs. (1). This is accomplished by expanding the displacements in Taylor series and retaining terms up to quadratic in the interatomic distance. The condition of positive Young's modulus yields sufficient conditions involving the force constants, namely,

$$\begin{aligned}&\alpha>0,\\ &B=\beta/\alpha>-0.25,\\ &C=\gamma/\alpha>-0.25.\end{aligned} \quad (25)$$

Now, after introduction of a dimensionless frequency parameter, λ, and the mass ratio, μ, given by

$$\lambda=m\omega^2/\alpha, \quad \mu=M/m, \quad (26)$$

Eq. (5) can be written as a quadratic either in λ or in y, namely,

$$\mu\lambda^2-2[(\mu+1)+2(B\mu+C)y]\lambda +4y[4BCy+2(B+C)+1]=0, \quad (27)$$

or

$$16BCy^2-4[B(\mu\lambda-2)+C(\lambda-2)-1]y +\lambda[\mu\lambda-2(\mu+1)]=0. \quad (28)$$

Equation (26) yields

$$\lambda=(1/\mu)\{[(\mu+1)+2(B\mu+C)y] \pm[(\mu-1+2B\mu y-2Cy)^2+4\mu(1-y)]^{\frac{1}{2}}\}. \quad (29)$$

The discriminant of Eq. (29) is always positive for real y between 0 and 1; hence, one obtains in this range two real branches of the frequency spectrum, the well-known acoustical and optical ones.

For infinitely long wavelengths, i.e., $y=0$, Eq. (27) yields

$$\lambda_1=0, \quad \lambda_2=2(1+\mu^{-1}). \quad (30)$$

Therefore,

$$\omega_1=0, \quad \omega_2=[2\alpha(m^{-1}+M^{-1})]^{\frac{1}{2}}. \quad (31)$$

For $y=1$ the two values of λ are

$$\lambda_1=2(1+2B), \quad \lambda_2=2(1+2C)/\mu, \quad (32)$$

and the corresponding cutoff frequencies at the boundary of the first Brillouin zone are

$$\begin{aligned}\omega_1'&=[2(\alpha+2\beta)/m]^{\frac{1}{2}},\\ \omega_2'&=[2(\alpha+2\gamma)M]^{\frac{1}{2}}.\end{aligned} \quad (33)$$

The following remarks may be made on the basis fo Eqs. (30) to (33):

(a) For $y\to 0$ the frequency of the optical branch, ω_2, henceforth referred to as the dispersion frequency, is independent of the next-nearest neighbor interactions. This is as it should be, since there is no relative motion of next-nearest neighbors in this limit.

(b) For β/α and γ/α going to infinity the cut-off frequencies tend to the values

$$\omega_1^*=2(\beta/m)^{\frac{1}{2}}, \quad \omega_2^*=2(\gamma/M)^{\frac{1}{2}}, \quad (34)$$

i.e., the cutoff frequencies of the two monatomic lattices obtained from Eqs. (1) by setting $\alpha=0$. That is, for α very small in comparison with β and γ the two lattices of particles m and M are essentially uncoupled except for very long wavelengths. There is no forbidden band between the acoustical and optical branches since both cut-off frequencies are higher than the dispersion frequency, ω_2, of Eq. (31).

(c) A forbidden band between the acoustical and

optical branches exists if the frequencies ω_1' and ω_2' given by Eq. (33) are different and at least one of them is smaller than ω_2 of Eq. (31). The forbidden band disappears if

$$\omega_1'=\omega_2', \tag{35}$$

which yields

$$C=\tfrac{1}{2}(\mu-1)+B\mu. \tag{36}$$

For example, there is no forbidden band for $\mu=1$ and $B=C$, i.e., a monatomic lattice.

(d) The forbidden band also disappears if the cutoff frequency of the acoustical branch is greater than or equal to the dispersion frequency. This is so when

$$1+2B>1+1/\mu, \tag{37a}$$

or

$$1+2C>\tfrac{1}{2}(\mu+1), \tag{37b}$$

depending on whether the acoustical cutoff is given by the first or second of Eqs. (33).

Additional information about the character of the frequency dispersion curves is obtained by the solution of Eq. (28), which yields

$$y=(1/8BC)\{[B(\mu\lambda-2)+C(\lambda-2)-1]\pm D_1^{\frac{1}{2}}\}, \tag{38}$$

where the discriminant, D_1, is given by

$$D_1=[B(\mu\lambda-2)+C(\lambda-2)-1]^2 -4BC\lambda(\mu\lambda-2\mu-2). \tag{39}$$

For $\omega\ll[\alpha/m]^{\frac{1}{2}}$, i.e., $\lambda\ll 1$,

$$y_1\approx\frac{(\mu+1)\lambda}{2(2B+2C+1)}, \qquad y_2\approx-\frac{2B+2C+1}{4BC}+\left[\frac{B\mu+C}{4BC}-\frac{(\mu+1)}{2(2B+2C+1)}\right]\lambda, \tag{40}$$

omitting second and higher powers of λ.

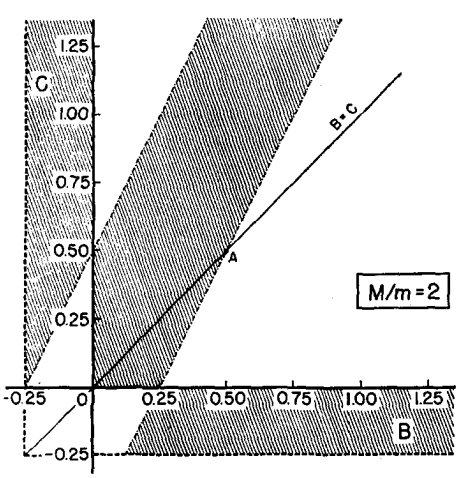


FIG. 1. The region of stability in the plane of the force-constant ratios $B=\beta/\alpha$ and $C=\gamma/\alpha$ lies above and to the right of the heavy dashed line for all mass ratios. The shaded area corresponds to values of B and C for which two stationary frequencies exist in the frequency dispersion curves of the infinite lattice, assuming mass ratio $M/m=2$. The two stationary frequencies coincide for points such as A.

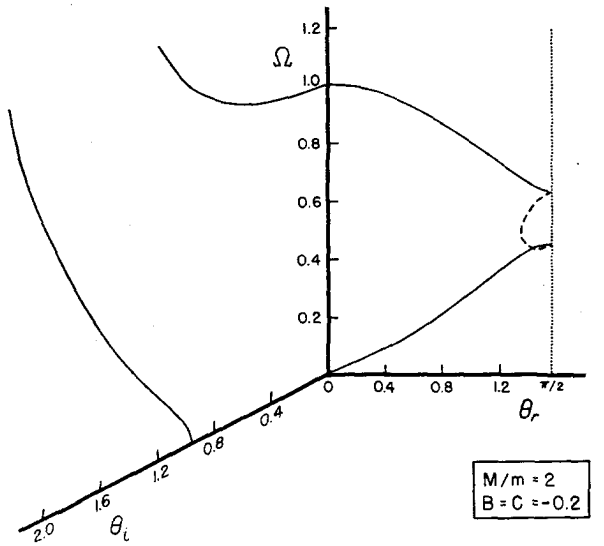


FIG. 2. Frequency dispersion curves for mass ratio $M/m=2$ and force-constant ratios $B=C=-0.2$. There are no stationary frequencies. The two pure imaginary branches above the optical dispersion frequency extend to infinity.

The root y_1 corresponds to the acoustical branch, the dispersion curve of which is approximated at low frequencies by

$$\theta_1\approx\left[\frac{(m+M)}{2(\alpha+2\beta+2\gamma)}\right]^{\frac{1}{2}}\omega. \tag{41}$$

The nature of the second root y_2 depends on the sign of the product BC. For $BC>0$ the root y_2 tends, for $\lambda\rightarrow 0$, to a negative value, since

$$2B+2C+1>0, \tag{42}$$

on the strength of the stability conditions given in Eqs. (25). The corresponding wave number θ_2 is pure imaginary and tends to the imaginary infinity when either one or both of B and C go to zero. For $BC<0$ one obtains

$$-\frac{2B+2C+1}{4BC}-1=-\frac{(2B+1)(2C+1)}{4BC}>0. \tag{43}$$

Hence y_2 tends to a value greater than unity when ω goes to zero. This corresponds to a wave number $\theta_2=\pi/2+ib$ where b goes to infinity when either one or both of B and C go to zero.

From Eq. (28) one obtains, in general, two distinct values of y, or θ, for a given value of the frequency ω. There may be, however, certain stationary values of the frequency (maximum or minimum) for which the two values of y are identical. These stationary values of the frequency are obtained by setting the discriminant D_1, given in Eq. (39), equal to zero. This procedure yields a quadratic in λ, namely,

$$a_1\lambda^2+a_2\lambda+a_3=0, \tag{44}$$

where

$$a_1=(B\mu-C)^2,$$
$$a_2=(2B-2C+1)(B\mu-C)+2C, \quad (45)$$
$$a_3=(2B+2C+1)^2.$$

Equation (44) yields

$$\lambda=(B\mu-C)^{-2}\{[B(2B\mu+\mu-2C)]^{\frac{1}{2}} \pm[C(2C+1-2B\mu)]^{\frac{1}{2}}\}^2, \quad (46)$$

from which one deduces the stationary values of the frequency and then, from Eq. (38), the corresponding wave numbers. A real positive λ, and consequently a real frequency, is obtainable from Eq. (46) if both of the radicals are positive. In the B, C plane each of the radicals is a product of linear functions of B and C, and hence it is positive inside two plane sectors described by

$$B(2\mu B+\mu-2C)>0, \quad (47a)$$

and

$$C(2C+1-2\mu B)>0, \quad (47b)$$

respectively. An example is shown in Fig. 1 for the mass ratio $\mu=2$, where the shaded areas correspond to force constant ratios B and C satisfying both the inequalities given in Eqs. (47).

For B and C going to zero the frequency parameter, λ, tends to

$$\lambda\approx[(B\mu)^{\frac{1}{2}}\pm(C)^{\frac{1}{2}}]^{-2}, \quad 0<(B;C)\ll 1, \quad (48)$$

and the corresponding wave-number parameters are

$$y\approx(\tfrac{1}{8}BC)\{(B\mu+C)[(B\mu)^{\frac{1}{2}}\pm(C)^{\frac{1}{2}}]^{-2}-1\}. \quad (49)$$

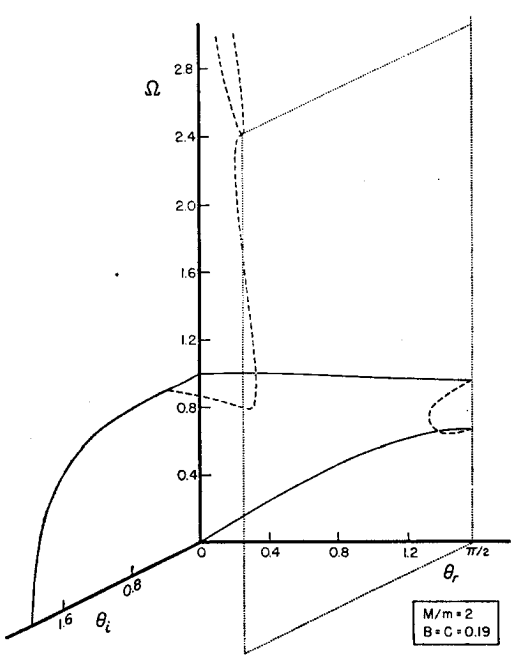


FIG. 3. Frequency dispersion curves for mass ratio $M/m=2$ and force-constant ratios $B=C=0.19$. A complex branch connects the maximum of a pure imaginary branch and the minimum of a branch lying on the plane $\theta_r=\pi/2$.

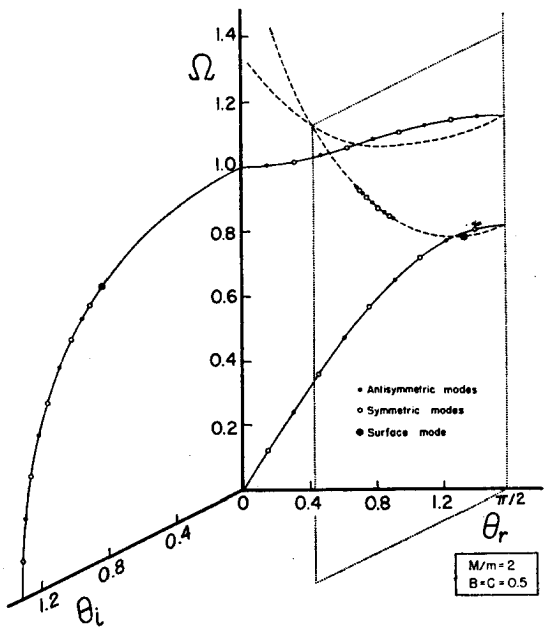


FIG. 4. Frequency dispersion curves for mass ratio $M/m=2$ and force-constant ratios $B=C=0.5$. There are two intersecting branches, emanating from the cutoff frequencies, which lie entirely on the plane $\theta_r=\pi/2$. Also shown in this figure are the symmetric and antisymmetric modes of a chain with 21 particles. In the acoustical frequency range each mode is associated with one real and one imaginary wave number; in the optical frequency range with one real and one complex wave number of the type $\pi/2+ib$. The surface mode is associated with one imaginary and one complex wave number $\pi/2+ib$.

It may be ascertained that one of the roots y given by Eq. (49) is positive and the other one negative, and they are both very large in absolute value when B and C are small. The negative y corresponds to the smaller one of the two frequencies given by Eq. (48) which is a maximum of an imaginary branch, i.e., a branch associated with pure imaginary wave numbers. The positive y corresponds to the higher stationary frequency which is a minimum of a branch associated with wave numbers of the type $(\pi/2+ib)$. This branch extends from the minimum frequency given by Eq. (48) to infinity.

As an illustration we discuss the case $B=C$ on the basis of the preceding remarks. The frequency dispersion curves obtained by numerical computations for mass ratio $\mu=2$ are shown in Figs. 2–4 for B and C equal to -0.2, 0.19, and 0.5, respectively. The ordinate of these figures is the frequency normalized with respect to the optical dispersion frequency, namely,

$$\Omega=\omega[2\alpha(m^{-1}+M^{-1})]^{-\frac{1}{2}}, \quad (50)$$

and θ_r and θ_i are the real and imaginary axes for the wave numbers. All the solid lines lie on the principal planes (Ω,θ_r) and (Ω,θ_i) which are hereafter referred to as the real and imaginary planes, respectively.

When B and C are negative, Fig. 2, the acoustical and optical branches have approximately the shape of

the corresponding branches for $B=C=0$. There is a complex loop connecting the cutoff frequencies. There is also a pure imaginary branch extending through the entire range of frequencies from zero to infinity. This branch moves along the axis of imaginary wave numbers towards infinity as B and C tend to zero. The optical branch continues into the imaginary plane as a second imaginary branch for frequencies higher than the dispersion frequency. As $B=C$ becomes positive the two imaginary branches loop into a maximum. A complex branch originates from this maximum, Fig. 3, and goes into the minimum of another complex branch which lies on the plane $\theta=\pi/2$, i.e., one associated with wave numbers of the type $(\pi/2+ib)$. As B and C increase, the maximum of the imaginary loop is lowered and approaches the ω axis. For $B=C=0.2$ and $\mu=2$ it coincides with the dispersion frequency of the optical branch. For $B=C>0.2$ the maximum moves into the real plane and becomes a maximum of the optical branch. At the same time the cutoff of the optical branch rises until it becomes equal to the dispersion frequency for $B=C=0.25$. For B and C increasing further the maximum of the optical branch moves in the direction of the real wave number axis toward $\theta=\pi/2$ unitil it coincides with the cutoff when $B=C\approx 0.315$. This is the value of the force constant ratios such that the wave number parameter, given by Eq. (38), corresponding to the lower stationary frequency, given by Eq. (46), is equal to unity. For $B=C>0.315$ this lower stationary frequency moves into the complex loop which connects the cutoff frequencies. The two stationary frequencies approach each other, for increasing B and C, until they coincide for $B=C=0.5$, and the two branches in the plane $\theta_r=\pi/2$ intersect. The latter case is shown in Fig. 4

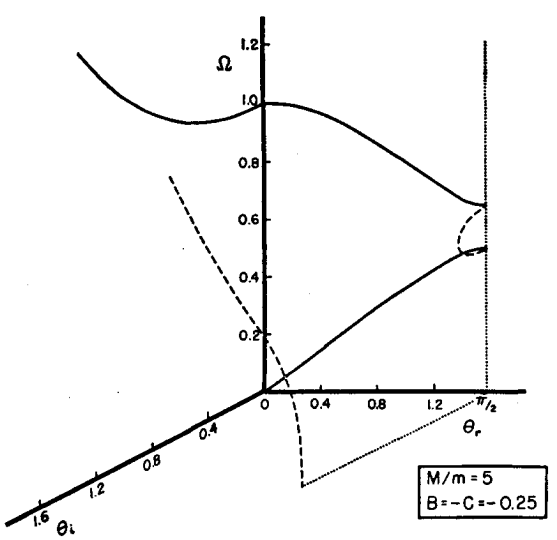


FIG. 5. Frequency dispersion curves for mass ratio $M/m=5$ and force-constant ratios $B=-C=-0.25$. There are no stationary frequencies. The pure imaginary branch above the optical dispersion frequency extends to infinity. Another branch lies entirely on the plane $\theta_r=\pi/2$ for frequencies from zero to infinity.

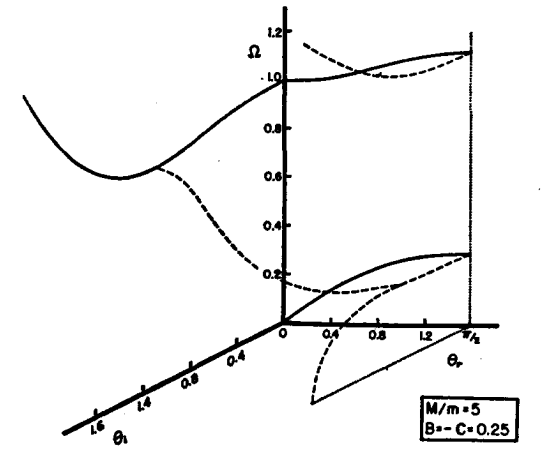


FIG. 6. Frequency dispersion curves for mass ratio $M/m=5$ and force-constant ratios $B=-C=+0.25$. A complex branch connects the maximum of a branch lying on the plane $\theta_r=\pi/2$ with the minimum of a pure imaginary branch which extends to infinity. Another branch originating from the optical cutoff lies entirely on the plane $\theta_r=\pi/2$ and also extends to infinity.

and corresponds to the point A of Fig. 1 on the boundary between sectors for which stationary frequencies are possible and impossible. For $B=C>0.5$ no stationary frequencies exist.

In Figs. 5 and 6 are given the dispersion curves for $B=-C=\pm 0.25$ and mass ratio $\mu=5$ as an illustration of the various types of branches that may be obtained. Of particular importance in the study of surface modes are the forbidden bands of the infinite lattice. As seen from Figs. 2 to 6 the wave numbers within these bands may be pure imaginary, complex, or special complex of the type $(\pi/2+ib)$. All these three types of wave numbers may give rise to surface modes. However, it will be seen that in the case of free boundaries considered in this paper no surface modes are found above the minimum frequency of the optical branch.

IV. SURFACE MODES IN FINITE AND SEMI-INFINITE LATTICES

The evaluation of the roots of Eqs. (11), (16), and (24) was done numerically using an IBM 704 computer. Equations (11) and (16) yield the normal mode frequencies of a finite lattice, whereas Eq. (24) yields the frequency of the surface mode of a semi-infinite lattice when such a mode exists. As is to be expected a surface mode, when possible, may be obtained through a limiting process, namely, by increasing the number of particles of a finite lattice indefinitely. Table II shows the evolution of the surface mode from symmetric or antisymmetric modes of a chain with an odd number of particles, as the number of particles increases from three to infinity. The table gives the normalized frequency, Ω, and the two associated wave numbers, θ_j, of the normal mode obtained within the forbidden band between the frequency cut-offs, assuming a mass ratio $M/m=2$, particles of mass m at the ends, and

TABLE II. Evolution of the surface mode from symmetric or antisymmetric modes of a finite chain with $2N+1$ particles and particles of mass m at the ends. ($M/m=2$, $B=C=0.25$, $\Omega=\omega[2\alpha(m^{-1}+M^{-1})]^{-\frac{1}{2}}$.)

	Symmetric mode			Antisymmetric mode		
N	Ω	$\theta_1-\frac{\pi}{2}$	θ_2	Ω	$\theta_1-\frac{\pi}{2}$	θ_2
1	...	...	...	0.81650	0.46854i	1.34991i
2	0.80769	0.45265i	1.36480i	0.76376	0.35071i	1.42982i
3	0.78100	0.39591i	1.40596i	0.79834	0.43433i	1.37986i
4	0.79426	0.42584i	1.38621i	0.78688	0.40965i	1.39736i
5	0.78919	0.41484i	1.39391i	0.79236	0.42179i	1.38911i
7	0.79054	0.41782i	1.39188i	0.79113	0.41912i	1.39098i
10	0.79086	0.41853i	1.39139i	0.79081	0.41842i	1.39147i
∞	0.79084	0.41847i	1.39143i	0.79084	0.41847i	1.39143i

$B=C=0.25$. Figure 4 shows the complete frequency spectrum of a symmetric lattice with 21 particles, for $B=C=0.5$ and $M/m=2$.

A more complete investigation of the surface modes can be obtained from a consideration of the semi-infinite lattice. Such a lattice has, of course, an infinite number of normal modes. However, the restriction to displacements given by Eq. (18) makes possible the direct determination of the surface mode alone.

For $B=C=0$, Eq. (24) degenerates into the equation

$$uv+u+v=0 \tag{51}$$

also given by Wallis.[4] The frequency squared of the surface mode is given in this case, by

$$\omega_0^2=[\alpha(1/m+1/M)] \tag{52}$$

and is equal exactly to the mean-square frequency of the gap between the cutoff frequencies of the acoustical and optical modes.

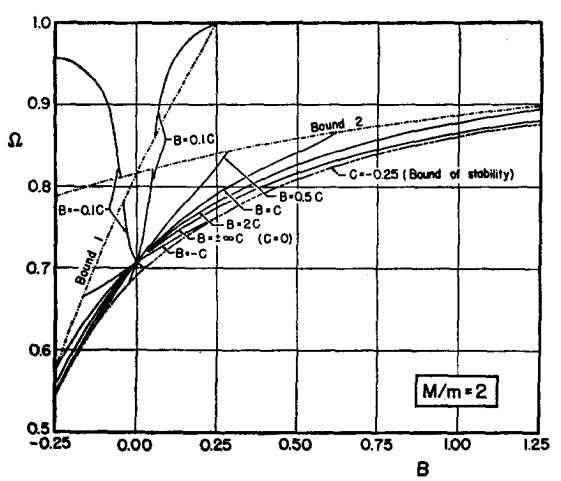


FIG. 7. Normalized frequency of surface modes, Ω, versus force-constant ratio B for mass ratio $M/m=2$. The surface modes have been obtained for B and C along straight lines $B=$constant$\times C$. Surface modes exist below both bounds 1 and 2, and above the bound of stability ($C=-0.25$), and also above both bounds 1 and 2, and below the optical dispersion frequency ($\Omega=1$).

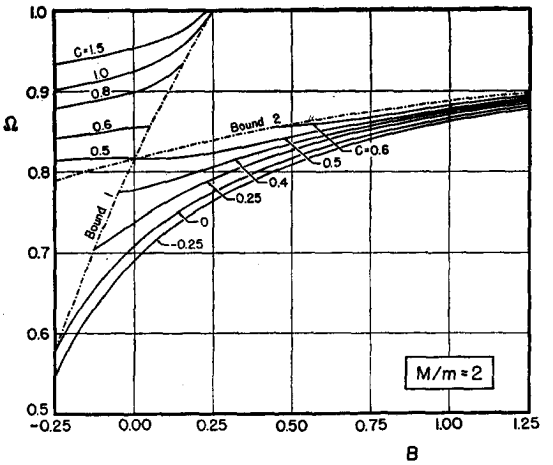


FIG. 8. Normalized frequency of surface modes, Ω, versus force-constant ratio B for mass ratio $M/m=2$. The results are the same as for Fig. 7 but plotted for $C=$constant.

For B and C going to zero, we obtain by a perturbation procedure approximate expressions for the frequency squared of the surface mode, ω_s^2, and the frequency squared of the center of the gap ω_c^2, namely,

$$\omega_s^2\approx\omega_0^2[1+B+C(m/M)^2], \tag{53}$$

and

$$\omega_c^2=\tfrac{1}{2}(\omega_1'^2+\omega_2'^2)$$
$$\approx\omega_0^2[1+2(1+m/M)^{-1}(B+Cm/M)], \tag{54}$$

where ω_0^2 is given by Eq. (52), and ω_1', ω_2' are the cutoff frequencies given by Eqs. (33).

The numerical solution of Eq. (24) was obtained taking B and C to lie along center straight lines in the B, C plane, namely, $B=$constant$\times C$, $B=$constant, and $C=$constant. Examples of the results are shown in Figs. 7–9 where the frequency of the surface mode, when such a mode exists, is plotted against B or C.

As has already been mentioned, a surface mode may exist if there is a forbidden range of frequencies for the infinite lattice. However, the existence of such a range is a necessary condition but not a sufficient one. This is seen, for example, from the computations for $B=C$ and $M/m=2$ (Fig. 7). For very small values of $B=C$ there exists a surface mode which is approximated very well by Eq. (53). A surface mode also exists for the entire range of negative B's satisfying the stability condition, as well as for positive values of B up to about 0.63. As the value of B increases from 0 to $\sim$0.63 the cutoff frequency of the acoustical branch rises more rapidly than that of the surface mode until they coincide when $B\approx0.63$. For $B>0.63$ there is no surface mode even though the forbidden gap between the acoustical and optical branches exists for values of $B=C$ up to unity. Similar results are obtained when B and C are taken along other straight lines on the B, C plane. On the basis of the numerical computations, we have plotted in Fig. 10 the boundaries between zones of the

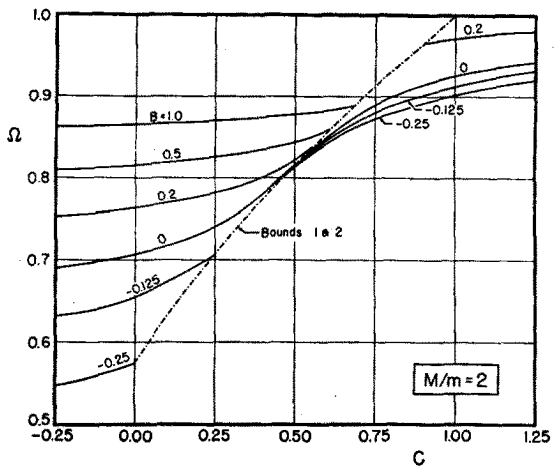


FIG. 9. Normalized frequency of surface modes, Ω, versus force-constant ratio C for mass ratio $M/m=2$, and for various constant values of B. The bounds 1 and 2 coincide in this representation. Surface modes exist above both the line marked bounds 1 and 2 and the bound of stability ($B=-0.25$).

B, C plane in which surface modes are possible and zones where they are impossible, for various values of the mass ratio μ. These boundaries are lines along which the frequency of the surface mode coincides with a cutoff frequency. In particular, the line marked bound 1 corresponds to the case of the disappearance of the forbidden band which takes place when the two cut-off frequencies coincide. Accordingly, bound 1 is described by the equations

$$\lambda=2(1+2B)=2(1+2C)/\mu. \tag{55}$$

It may be ascertained that Eqs. (55) correspond to a degenerate case for which the boundary conditions for a surface mode are satisfied.

The other bound, namely, bound 2, corresponds to

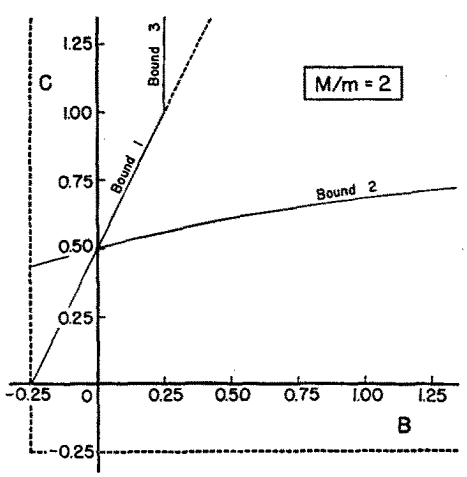


FIG. 10. Region of stability in the plane of the force-constant ratios B and C, showing regions of existence of surface modes, for mass ratio $M/m=2$, and end particle of mass m. Surface modes exist below both bounds 1 and 2, and also above these bounds and to the left of bound 3.

the case when the cut-off frequency given by the second of Eqs. (33) coincides with the frequency of the surface mode. In this case the secular and frequency equations yield the relationships

$$\lambda=2(1+2C)/\mu,$$
$$B=\frac{(2C+1)(4C-\mu)}{4C(\mu-2C)}, \tag{56}$$
$$y_1=1, \quad y_2=\frac{(\mu-2C)^2}{\mu(4C-\mu)}.$$

The two bounds are shown in Figs. 7–10, and they coincide in Fig. 9. Bound 3 shown in Fig. 10 corresponds to the case of coincidence of the acoustical cut-off and the optical dispersion frequency.

The meaning of the bounds can be illustrated by considering the case of $B=-0.125$ and a varying C, for $\mu=2$ (Fig. 8). A surface mode exists when $-0.25 \leqslant C<0.25$, when the acoustical cutoff corresponds to $\lambda=1+2C$. The surface mode vanishes when the acoustical cutoff coincides with the optical cut-off, at $C=0.25$ (bound 1). Above bound 1 the frequency cutoff corresponding to $\lambda=1+2C$ is that of the optical branch. There can be no surface mode between bounds 1 and 2, i.e., $0.25 \leqslant C \leqslant \sim 0.47$. For $C \approx 0.47$, (bound 2), the surface mode frequency coincides with the optical cutoff, and for $C>\sim 0.47$ the surface mode is unveiled by the rising optical cutoff and is found in the frequency range between the acoustical cutoff and the optical dispersion frequency.

In Fig. 11 are plotted the bounds 1, 2, and 3 for $M/m=0.5$, i.e., assuming a mass ratio equal to two but a heavy particle at the free end. It is seen that a surface mode is again possible, in this case, when B and C are below both bounds 1 and 2, or above these bounds and to the left of bound 3. For example, one obtains a surface mode at $\Omega=0.978$ for $B=-0.2$ and $C=0.5$, and at $\Omega=0.878$ for $B=1.2$ and $C=-0.1$. No surface mode is obtainable for $B=C=0$, i.e., nearest neighbor interactions only, in agreement with the results of Wallis.[4]

It may be asked whether a surface mode may appear in the range of frequencies above the optical branch, which is also a forbidden range for the infinite lattice. No surface modes have been found in that range numerically. Furthermore, one does not have to search for such modes at very high frequencies. Since a surface mode may be obtained as a limiting case of a mode of the finite lattice, the only way for a surface mode to appear above the optical branch is by a body mode "leaking" into this range. It has been ascertained that this does not happen as one increases B and C. In the limit $B \to \infty$ and $C \to \infty$ one obtains two essentially uncoupled monatomic lattices which, of course, do not exhibit surface modes.

V. GENERALIZATIONS AND APPLICATIONS TO THE EVALUATION OF ROOTS OF DETERMINANTS

In the preceding sections a method was given for the evaluation of the eigenmodes of a bounded lattice with free boundaries and, in particular, surface modes. The evaluation was accomplished by superposition of modes of the infinite lattice which were computed assuming a real frequency but a complex wave number, in general. It may be observed that a sufficient number of modes was always available so that a linear superposition of them might satisfy the boundary conditions of the lattice. This is always the case in the problems of this nature, regardless of the number of interactions assumed, for the following reason: The boundary conditions involve more and more particles away from the end particle as the number of interactions is increased to include interactions of farther and farther neighbors. This increases the number of boundary conditions which must be satisfied. At the same time these more complex interactions increase the order of the secular equation of the infinite lattice and, consequently, the number of modes of the infinite lattice at a given frequency. Thus it is possible to extend the present discussion and include moderately long-range interactions. The only difficulty lies in the fact that the order of both the secular determinant of the infinite lattice and the frequency determinant of the finite lattice increases as one increases the number of interactions considered.

The discussion can be generalized to two or three dimensions, with one important restriction: Only plane-front waves can be treated and plane boundaries admitted. If the lattice is bounded by two planes intersecting at a finite distance a treatment similar to the present one cannot yield a closed solution. The present method has already been used for the investigation of surface waves along a plane of symmetry of a simple cubic lattice, assuming central-force interactions between nearest and next-nearest neighbors and an angular stiffness of a right angle formed by three consecutive nearest neighbors.[2] In this case the order of the secular equation of the infinite lattice is, in general three, and the boundary conditions are three. If one applies the same treatment to a face-centered or body-centered lattice both the order of the secular equation of the infinite lattice and the number of boundary conditions increase to six, in general. Work is currently underway on such lattices.

Finally, some remarks may be made regarding the possibility of application of the present discussion to the evaluation of roots of certain determinants of large order. This application stems from the fact that the treatment of the finite lattice given in this paper should and does yield the same eigenfrequencies as would be obtained if one considered explicitly the equations of motion of all the particles, including the boundary particles, and found time-periodic solutions for their displacements. The latter treatment has been used, for example, by Wallis[4] in his discussion of the finite diatomic lattice with nearest-neighbor interactions only. The normal mode frequencies of the finite lattice have been obtained by Wallis through the evaluation of a secular determinant which is of order equal to the number of particles and has only three non-zero terms in every row, about the main diagonal. Rutherford[7,8] has given some techniques for evaluating such determinants. The same results may be obtained very directly following the method of this paper by setting $B=C=0$ and eliminating half of the boundary conditions, i.e., those pertaining to the next to the end particles. The advantages of the present method are even greater if one includes next-nearest neighbor or more complex interactions. A Rutherford-type approach for the evaluation of the secular equation of a multi-particle finite lattice would be rather cumbersome. However, since the boundary value formulation is a mathematically

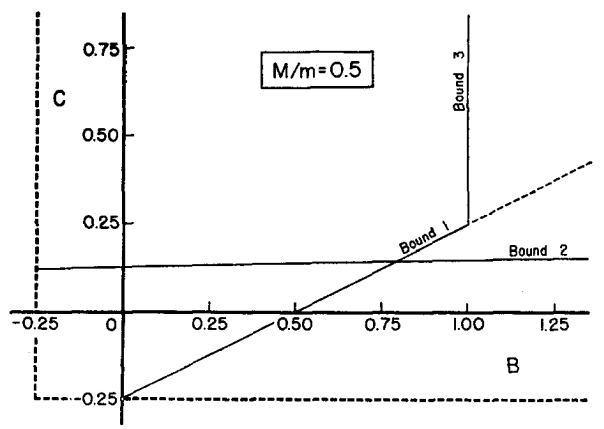


FIG. 11. Region of stability in the plane of the force-constant ratios B and C, showing regions of existence of surface modes, for mass ratio $M/m=0.5$ and end particle of mass m. Surface modes exist below both bounds 1 and 2, and also above these bounds, and to the left of bound 3.

equivalent one, we may use it in order to evaluate indirectly roots of the corresponding secular determinants. For example the boundary value approach given in the early sections for $2N$ particles yields the roots of a determinant of the type,

$$\begin{vmatrix} u+1+B & 1 & B & & & & & \\ 1 & v+C & 1 & C & & & & \\ B & 1 & u & 1 & B & & & \\ & C & 1 & v & 1 & C & & \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & \\ & & & & B & 1 & u+B & 1 \\ & & & & & C & 1 & v+1+C \end{vmatrix}_{2N} = 0.$$

ACKNOWLEDGMENT

We wish to thank Dr. Robert Herman, General Motors Research Laboratories, for a number of valuable discussions and criticisms in the course of this work and for reading the manuscript.

Vibrational Modes of Disordered Linear Chains

HERBERT B. ROSENSTOCK AND RICHARD E. MCGILL
U. S. Naval Research Laboratory, Washington, D. C.
(Received July 24, 1961)

The kth normal mode of vibration (i.e., the atomic displacements) of a disordered one-dimensional lattice (masses and force constants completely arbitrary) with nearest-neighbor interaction has precisely $k-1$ nodes. *A fortiori*, the same is true for ordered one-dimensional lattices with any number of atoms per unit cell. This theorem exhibits a close relationship between eigenfunctions in monatomic ordered lattices (to which its application has been known for many years) and disordered lattices; a relationship which appears surprising in view of recent demonstrations of the gross differences exhibited in the distribution of eigenvalues. It is thus suggested that some basic concepts of ordered lattice dynamics—propagation vector, phonon momentum, etc.—may retain some simple validity for disordered solids as well. Some numerical examples are given.

I. INTRODUCTION AND DISCUSSION

THE vibrations of ordered lattices have been studied for many years, and some qualitative understanding has been attained. By contrast, intensive work on disordered lattices began only fairly recently, and our understanding is confined mostly to restricted one-dimensional models, and is quantitative (rather than analytical) in nature. Also, work on disordered lattices has been directed mostly to the determination of the distribution of frequencies (eigenvalues). The behavior of the atomic displacements (normal modes, or eigenvectors) is, however, also of mathematical and physical interest, and this paper is devoted to them.

We consider a linear chain of N particles. Let them interact with their nearest neighbors only by Hooke's law forces, and let the force constants and the masses be entirely arbitrary.[1] Call the displacements of the ith particle u_i $(i=1, 2, \cdots, N)$. We know that N eigenvalues (squared frequencies) will exist and we call them $\lambda^{(k)}$, $k=1, 2, \cdots, N$, and we order them according to magnitude so that $\lambda^{(l)} \leqslant \lambda^{(m)}$ if $l<m$. We define the normal mode, or eigenfunction, as the continuous function of position that is obtained if one plots the displacement amplitude u_i of the ith atom in a one-dimensional chain versus i, and then connects adjacent points by straight lines.[2] We shall then show that the normal mode corresponding to $\lambda^{(k)}$ has precisely $k-1$ nodes. That is, there is exactly one normal mode with no zeros, one mode with one zero, one mode with two zeros, $\cdots$, and one with $N-1$ zeros, and the larger the number of zeros the higher the frequencies to which the modes correspond. Since an ordered lattice is just a special case of a disordered lattice, the theorem applies, *a fortiori*, also to all ordered lattices, whatever the number of atoms per unit cell.

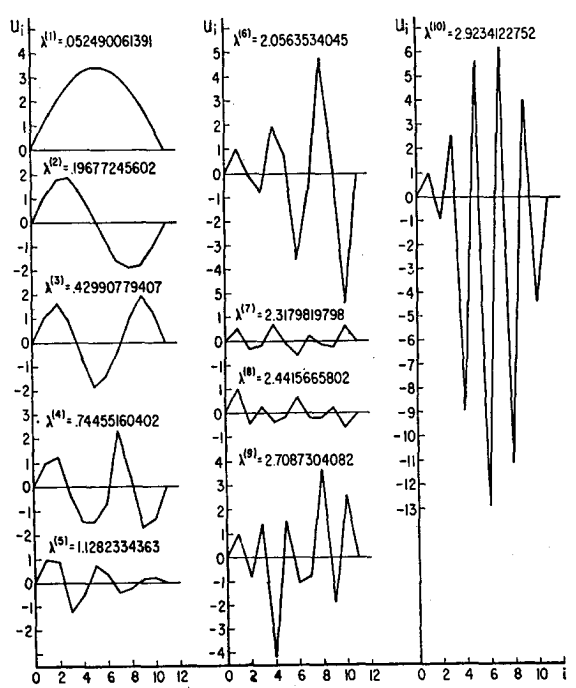


FIG. 1. Normal modes of a disordered lattice [Eq. (4)]. N=10, all $k_i=1$, m_1 to m_{10} equal to 1, 2, 2, 1, 2, 1, 2, 1, 2, 1.

It has been well known for a long time that this result is valid for ordered monatomic lattices,[3] but its application to disordered lattices, or indeed to ordered polyatomic lattices, appears to be new, and indeed surprising in view of recent work[4] showing that the distribution of eigenvalues of disordered lattices behaves very differently from that of an ordered lattice. It is interesting that a frequency distribution as irregular as Dean's[4] should be associated with normal modes as regular as the ones exhibited here.

In Sec. II our theorem is proved in a mathematically very simple way. In Sec. III we present some numerical examples in graphic form to show that our results are not only correct but also immediately apparent to the eye.

The mathematical content of Sec. II can be formu-

[1] Provided only that the masses and force constants are positive.
[2] This convention serves mostly to simplify our language: we can now speak of "modes" and "eigenfunctions" rather than "eigenvectors", and of "zeros" or "nodes" rather than "sign changes."
[3] M. Born and Th. von Karman, Physik Z. 13, 297 (1912).
[4] P. Dean, Proc. Roy. Soc. (London) A254, 507 (1960) has made extensive numerical calculations for various special cases of disordered lattices and has determined that their frequency distributions contain many sharp peaks and other features not found in the frequency distributions of ordered lattices.

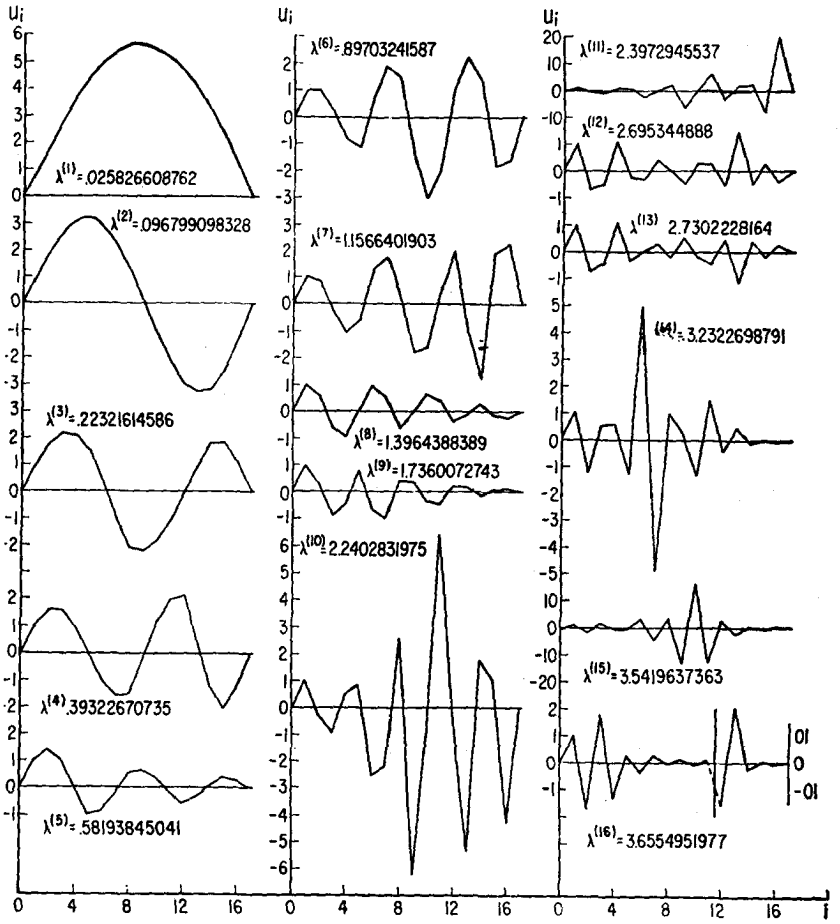


FIG. 2. Normal modes of a disordered lattice [Eq. (4)]. N=16, all $k_i=1$, m_1 to m_{16} equal to 1, 1, 1, 1, 2, 1, 1, 2, 1, 1, 1, 2, 1, 2, 2, 1.

lated as a theorem on "normal Jacobian matrices," and as such is not new.[5] The physical implications of it, particularly in solid-state theory, appear not to be generally realized, however. The physical implications are therefore discussed in Sec. IV.

II. PROOF OF THE THEOREM

Newton's equation of motion for the ith particle in the model described in Sec. I is

$$k_i(u_{i-1}-u_i)+k_{i+1}(u_{i+1}-u_i)=m_i\ddot{u}_i, \tag{1}$$

$i=1, 2, \cdots, N$, with[6]

$$u_0=0, \quad u_{N+1}=0. \tag{2}$$

Here m_i is the mass of the ith atom and k_i the restoring force constant between atoms i and $i-1$. After inserting harmonic time dependence, (1) can be written as

$$-k_{i+1}u_{i+1}+(k_i+k_{i+1}-m_i\lambda)u_i-k_iu_{i-1}=0, \tag{3}$$

$i=1, 2, \cdots, N$, again with (2); here $\lambda^{\frac{1}{2}}$ is the frequency. The usual way of computing the N eigenvalues of the system (3) would be to divide each equation by m_i and equate the determinant of the coefficients of the u's to 0; but an equally correct, and here preferable, way is to write (3) in the form

$$k_2u_2=(k_1+k_2-m_1\lambda)u_1, \tag{4.1}$$

$$k_3u_3=(k_2+k_3-m_2\lambda)u_2-k_2u_1, \tag{4.2}$$

$$k_4u_4=(k_3+k_4-m_3\lambda)u_3-k_3u_2, \tag{4.3}$$

$$k_Nu_N=(k_{N-1}+k_N-m_{N-1}\lambda)u_{N-1}-k_{N-1}u_{N-2}, \tag{4.N-1}$$

$$k_{N+1}u_{N+1}=(k_N+k_{N+1}-m_N\lambda)u_N-k_Nu_{N-1}. \tag{4.N}$$

Here we have used $u_0=0$ [from (2)] in (4.1). We can now set $u_1=1$ (this is permissible since any eigenvalue will fix the eigenfunctions only to within a constant factor), pick a value of λ, use (4.1) to compute u_2, (4.2) to compute $u_3, \cdots$, (4.N−1) to compute u_N, and (4.N) to compute u_{N+1}. The question now is whether

[5] F. R. Gantmacher and M. G. Krein, *Oszillationsmatrizen, Oszillationskerne, und kleine Schwingungen Mechanischer Systeme* (Akademie Verlag, Berlin, Germany, 1960), p. 80.

[6] Equation (2) corresponds to the choice of "fixed" boundary conditions. The boundary conditions enter explicitly into our formulation of the proof, but do not affect the result at all. The proof could be carried through for "free" or "cyclic" boundary conditions with only trivial changes. In fact, fixed boundary conditions may be considered a special case of cyclic ones if only nearest neighbors interact. The present formulation has the advantage that questions of degeneracy, which complicate the application of Sturm's theorem, do not arise.

the u_{N+1} so computed is zero [as it is supposed to be according to (2)], or not. If not, we throw away the u_i's we have computed, and repeat our computation using a different value of λ; if so, then the λ we have chosen is an eigenvalue of the set, and the u_i we computed are eigenfunctions. In this manner all eigenvalues and their corresponding eigenfunctions can be computed, or, more precisely, can be approximated to any required degree of precision.

For the purposes of our proof, we note that the set (4) with $u_1=1$ enables us to consider u_{N+1}, u_N, u_{N-1}, $\cdots$, u_3, u_2, $u_1=1$ as a set of functions of the variable λ. Upon inspection it is found that they do indeed form a "Sturm series"[7]; that is to say, Sturm's theorem applies to them. For our purposes, Sturm's theorem may be stated thus: "The number of variations in sign that occur if the functions u_1, u_2, $\cdots$, u_{N+1} are successively evaluated at one given value of λ is equal to the number of roots of u_{N+1} that are smaller than that λ."[8] Thus if the u_i are evaluated at the first root of u_{N+1} (i.e., at our lowest eigenvalue), there will be zero variations in sign (i.e., our eigenfunction, which is identical with the first N Sturm functions $u_1, u_2, \cdots, u_N$ will have no zeros); if the u_i are evaluated at the second root u_{N+1} (i.e., at our second eigenvalue), there will be one variation in sign in the Sturm's series (i.e., one node in our eigenfunction); and generally if the Sturm's series is evaluated at the kth root of u_{N+1} (i.e., at our kth eigenvalue) the Sturm functions will show $k-1$ variations in sign. Thus the eigenfunction of the kth eigenvalue $\lambda^{(k)}$ has $k-1$ zeros. This is what we set out to prove.

III. NUMERICAL EXAMPLES

For purposes of illustration we exhibit here the results of some numerical calculations of eigenfunctions, performed on the NAREC on the set (4). The method of calculation was precisely the one described in Sec. II.[9,10] We used the "isotope" model of the lattice; that is, all the k_i were taken as equal, and only two masses were allowed; their ratio was 1 to 2. Figure 1 shows a calculation with $N=10$, with five heavy and five light masses, and Fig. 2 shows one for $N=16$, with five heavy and eleven light masses. The order of masses was determined by a random process. The applicability of our theorem is strikingly obvious to the eye.

IV. CONCLUDING REMARKS

The result here obtained is of the nature of an "adiabatic theorem": it asserts that a certain property—viz., the number of nodes of eigenfunctions, and thus, in a general sense, the concept of propagation vector—remains invariant under certain changes of the physical parameters. The theorem is quite general in some respects—there are no restrictions at all on the lattice spacing, or on the masses and the harmonic force constants (other than the physically trivial one that their ratios be positive.) On the other hand, we are restricted to nearest-neighbor interaction, and to one-dimensional crystals. One is tempted to conjecture that the first restriction is a human, rather than a physical one—that our theorem should in fact hold, perhaps in a somewhat modified way, for lattices with longer-range interaction as well, and that it is only the mathematical proof that is lacking at this time. We would not, on the other hand, guess in what sense, if any, the theorem might be correct for two- or three-dimensional lattices. A persistence of the theorem—and thus a persistence of the concepts of wave number, propagation vector, phonon momentum, etc.—for disordered two- and three-dimensional crystals would be of particular interest.

ACKNOWLEDGMENTS

We should like to thank Dr. Richard F. Wallis for a critical discussion, and Dr. Robert J. Rubin for calling reference 5 to our attention.

[7] For a definition of Sturm's series and a particularly clear proof of Sturm's theorem, see J. V. Uspensky, *Theory of Equation*. (McGraw-Hill Book Company, Inc., New York, 1948), Chap. 8*s*

[8] Our statement of Sturm's theorem differs from Uspensky's[7] only in terminology, with 3 slight exceptions. First, Uspensky's property 1 (p. 138) is satisfied in our case if $-\lambda$ is taken as the variable, not λ; as a result, the expression $v(a)-v(b)$ on p. 142 is to be replaced by its negative. Second, we have omitted reference to $v(a)$ which in our case is zero [all the u_i are positive when $\lambda=0$, as is easily verified from (4)]. Third, the trivial restriction that b not be a root of u_{N+1} has been removed in our case by not including the point of evaluation in the root count.

[9] The method appears to be both simple and efficient for practical calculations. For a more elaborate computational scheme, see T. O'Callaghan, J. Ind. and Appl. Math. **9**, 294 (1961).

[10] The algorithm (4) is very similar to the one used by Dean[4] in his calculation of the eigenvalue spectrum, but is slightly simpler from the computational standpoint. We are informed that our algorithm has now in fact been used for such calculations [A. A. Maradudin and P. A. Flinn (to be published)].

Evolution of a Crystal in an Oscillating Field

LESSER BLUM
Catedra de Fisicoquimica, Facultad de Ciencias Exactas, Buenos Aires, Argentina
(Received June 26, 1961)

Using Prigogine and Henin's method for the asymptotic solution of Liouville's equation, a master equation for the anharmonic homogeneous crystal is obtained; this equation is valid up to the order $\gamma^2\lambda^2 t$, where γ, and λ are the coupling parameters.

1. INTRODUCTION

IN this paper we obtain the evolution equation of the homogeneous anharmonic crystals using Prigogine and Henin's technique.[1] The notations and diagrams are similar to those of Prigogine. Furthermore to avoid repetition we will not explain the method. The path followed is the same, once a model Hamiltonian is proposed, in picking out the dominant diagrams in the iteration series of the transformed Liouville equation, and then repass to a differential form.

We suppose that the oscillating field of frequency ω_0 is coupled to the mth normal mode, the coupling parameter is γ.

$$H=H+\gamma H(t), \tag{1.1}$$

$$H=H_0+\lambda V=\sum_{k,s}\omega_{k,s}J_{k,s}+\lambda\sum V_{ks,k's',k''s''}\times\left(\frac{J_{ks}J_{k's'}J_{k''s''}}{\omega_{ks}\omega_{k's'}\omega_{k''s''}}\right)^{\frac{1}{2}}, \tag{1.2}$$

and

$$H(t)=\omega_0\cos(\omega,t)(J_{ms}/\omega_{ms})^{\frac{1}{2}}(e^{i\alpha_m}+e^{-i\alpha_m}). \tag{1.3}$$

In these formulas, J and ω are the angular and action variables corresponding to the normal mode k with polarization s (in what follows, we shall drop the s).

2. LIOUVILLE EQUATION

In a formal way we may write for the Liouville equation

$$\partial\rho/\partial t=(L+\delta L^+)\rho, \tag{2.1}$$

where

$$\delta L^+=\gamma\sum_i^N\left\{\frac{\partial H(t)}{\partial\alpha_i}\frac{\partial}{\partial J_i}-\frac{\partial H(t)}{dJ_i}\frac{\partial}{\partial\alpha_i}\right\} \tag{2.2}$$

$$=\gamma\omega_0\cos\omega_0 t\left(\frac{J_m}{\omega_m}\right)^{\frac{1}{2}}\left\{e^{i\alpha_m}\left(i\frac{\partial}{\partial J_m}-\frac{1}{2J_m}\frac{\partial}{\partial\alpha_m}\right)+e^{-i\alpha_m}\left(-i\frac{\partial}{\partial J_m}-\frac{1}{2J_m}\frac{\partial}{\partial\alpha_m}\right)\right\}. \tag{2.3}$$

If we pass to the "interaction representation,"[1] we obtain for the Fourier components of the distribution function, the following iterated equation:

$$\rho\{n_k\}^{(t)}=\rho\{n_k\}^{(0)}+\lambda\sum_{\{nk'\}}\int_0^t dt_1\exp[i\textstyle\sum_k(n_k-n_k')\omega_k t]\times\langle\{n_k\}|\delta L+\delta L^+|\{n_k'\}\rangle\rho\{n_{k'}\}^{(0)}+\lambda^2\sum_{\{nk'\}\{nk''\}}\int_0^t dt_1\int_0^{t_1}dt_2\times\exp[i\textstyle\sum_k(n_k-n_k')\omega_k t_1]\times\exp[i\textstyle\sum_k(n_k'-n_k'')\omega_k t_2]\times\langle\{n_k\}|\delta L+\delta L^+|\{n_k'\}\rangle\times\langle\{n_k'\}|\delta L+\delta L^+|\{n_k''\}\rangle\rho\{n_{k''}\}^{(0)}+\cdots. \tag{2.4}$$

The operators

$$\tilde{a}\equiv\lambda\langle\{n_k\}|\delta L|\{n_k'\}\rangle,\quad \tilde{b}\equiv\gamma\langle\{n_k\}|\delta L^+|\{n_k'\}\rangle, \tag{2.5}$$

may be written in a more explicit way as

$$\tilde{a}\to -iV_{\pm k\pm k'\pm k''}\left[\frac{1}{2}\left(\frac{n_k}{J_k}+\frac{n_{k'}}{J_{k'}}+\frac{n_{k''}}{J_{k''}}\right)\pm\frac{\partial}{\partial J_k}\pm\frac{\partial}{\partial J_{k'}}\pm\frac{\partial}{\partial J_{k''}}\right]\left(\frac{J_kJ_{k'}J_{k''}}{\omega_k\omega_{k'}\omega_{k''}}\right)^{\frac{1}{2}} \tag{2.6}$$

with $\{n_e\}=\{n_e'\}$, $n_k'=n_k\pm1$, etc. $e\neq k,k',k''$

$$\tilde{b}=i\gamma\omega_0\cos(\omega_0 t)\left(\frac{J_m}{\omega_m}\right)^{\frac{1}{2}}\left[\mp\frac{\partial}{\partial J_m}+\frac{n_m\pm1}{2J_m}\right], \tag{2.7}$$

3. DIAGRAM TECHNIQUE

To investigate the behavior of the terms in the series (2.4) we represent them by Prigogine's diagrams. The diagrams corresponding to the operator $\tilde{b}$ are shown in Fig. 1. The cycle will be given by

$$\tilde{c}=\tfrac{1}{2}\sum_{\text{sign}}\langle 0|\delta L^+|\pm1\rangle\langle\pm1|\delta L^+|0\rangle\times[\delta_+(\pm\omega_m-\omega_0)+\delta_+(\pm\omega_m+\omega_0)]. \tag{3.1}$$

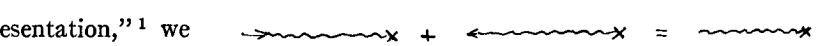


FIG. 1. Prigogine's diagrams corresponding to the operator $\tilde{b}$.

[1] I. Prigogine and F. Henin, J. Math. Phys. 1, 349 (1960).

4. INVESTIGATION OF THE ASYMPTOTIC BEHAVIOR

We impose the following conditions

$$N\to\infty;\quad \gamma\to O(1);\quad \gamma^2 t\to\infty;\quad \gamma/\lambda\to O(N^n);$$
$$t\to\infty;\quad \lambda\to 0;\quad \lambda^2 t\to \text{finite}. \quad (4.1)$$

This choice will produce divergence in the series if $n>2$. If $n<2$ there will be no divergences greater than that of the "cycle." The physical meaning of this is that the parameter γ must be great enough to produce any noticeable effect. Too great a γ, however, will produce a resonance disaster because it cannot be dissipated by the anharmonic forces.

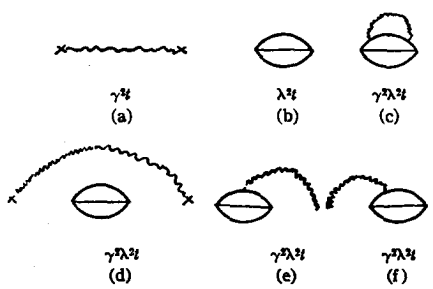


FIG. 2. Time dependence results for the most important diagrams.

For the time dependence the analysis is simple but rather lengthy; the results for the most important diagrams are shown in Fig. 2. The results of the asymptotic integrations are

$$(a)\to \tfrac{1}{2}\sum_{\text{sign}}\langle 0|\delta L^+|\pm 1\rangle\langle\pm 1|\delta L^+|0\rangle \times[\zeta_+(\pm\omega_m+\omega_0)+\zeta_+(\pm\omega_m-\omega_0)],$$

$$(b)\to \sum_{\text{sign}}\sum_{kk'k''} i\zeta(\omega_k+\omega_{k'}+\omega_{k''}) \times\langle 0|\delta L|-1_k-1_{k'}1_{k''}\rangle\langle -1_k-1_{k'}-1_{k''}|\delta L|0\rangle,$$

$$(c)\to \sum\langle 0|\delta L|1_e1_{e'}1_{e''}\rangle\langle 1_e1_{e'}1_{e''}|\delta L^+|1_e1_{e'}1_{e''}1_m\rangle \times\langle 1_m1_e1_{e'}1_{e''}|\delta L^+|1_e1_{e'}1_{e''}\rangle\langle 1_e1_{e'}1_{e''}|\delta L|0\rangle \times\left(\frac{-i}{4}\right)\zeta'(\omega_e+\omega_{e'}+\omega_{e''}) \times[\zeta(\omega_e+\omega_{e'}+\omega_{e''}+\omega_m+\omega_0) +\zeta(\omega_e+\omega_{e'}+\omega_{e''}+\omega_m-\omega_0)],$$

$$(d)\to \sum\langle 0|\delta L^+|1_m\rangle\langle 1_m|\delta L|1_m1_e1_{e'}1_{e''}\rangle \times\langle 1_m1_e1_{e'}1_{e''}|\delta L|1_m\rangle\langle 1_m|\delta L^+|0\rangle \times\left(\frac{-i}{4}\right)[\zeta'(\omega_m+\omega_0)\zeta(\omega_e+\omega_{e'}+\omega_{e''}+\omega_m+\omega_0) +\zeta'(\omega_m-\omega_0)\zeta(\omega_e+\omega_{e'}+\omega_{e''}+\omega_m-\omega_0)], \quad (4.2)$$

$$(e)\to \sum\langle 0|\delta L|1_e1_{e'}1_{e''}\rangle\langle 1_e1_{e'}1_{e''}|\delta L^+|1_e1_{e'}1_{e''}1_m\rangle \times\langle 1_e1_{e'}1_{e''}1_m|\delta L|1_m\rangle\langle 1_m|\delta L^+|0\rangle \times(i/4)\zeta(\omega_e+\omega_{e'}+\omega_{e''})[\zeta(\omega_e+\omega_{e'}+\omega_{e''}+\omega_m+\omega_0) \times\zeta(\omega_m+\omega_0)+\zeta(\omega_e+\omega_{e'}+\omega_{e''}+\omega_m-\omega_0) \times\zeta(\omega_m-\omega_0)],$$

$$(f)\to \sum\langle 0|\delta L^+|1_m\rangle\langle 1_m|\delta L|1_m1_e1_{e'}1_{e''}\rangle \times\langle 1_m1_e1_{e'}1_{e''}|\delta L^+|1_e1_{e'}1_{e''}\rangle\langle 1_e1_{e'}1_{e''}|\delta L|0\rangle \times(i/4)\zeta(\omega_e+\omega_{e'}+\omega_{e''})[\zeta(\omega_e+\omega_{e'}+\omega_{e''}+\omega_m+\omega_0) \times\zeta(\omega_m+\omega_0)+\zeta(\omega_e+\omega_{e'}+\omega_{e''}+\omega_m-\omega_0) \times\zeta(\omega_m-\omega_0)].$$

We may also show that to the order considered, the most important diagrams are built from these contributions. For example the diagram of Fig. 3 is negligible.

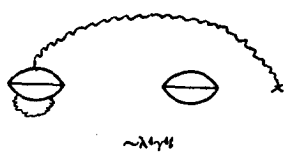


FIG. 3. Example of a negligible diagram.

5. MASTER EQUATION

With the conditions imposed, we have for the homogeneous component of the distribution function the following asymptotic expression:

$$\rho_0(t)=\left\{1+\sum_m^\infty \frac{t^m}{m!}(\gamma^2\bar{c}+\lambda^2(O_0)_2+\gamma^2\lambda^2(\Omega_{22}))^m\right\}\rho_0(0), \quad (5.1)$$

where

$$(O_0)_2=(b);\quad (\Omega_{22})=(c)+(d)+(e)+(f). \quad (5.2)$$

Differentiating, we get

$$\partial\rho/\partial t=[\gamma^2\bar{c}+\lambda^2(O_0)_2+\lambda^2\gamma^2(\Omega_{22})]. \quad (5.3)$$

With the conditions imposed in (4), $n<2$, we have, to the order $(\lambda^2\gamma^2 t)^n$

$$\partial\rho/\partial t=\lambda^2[(O_0)_2+\gamma^2(\Omega_\gamma)]. \quad (5.4)$$

ACKNOWLEDGMENTS

I am indebted to Professor I. Prigogine who proposed this problem to me, and to the Consejo Nacional de Investigaciones cientificas of Argentina for material support during my stay in Brussels.

Erratum: Dynamics of Nonlinear Stochastic Systems

[J. Math. Phys. 2, 124 (1961)]

ROBERT H. KRAICHNAN
Institute of Mathematical Sciences, New York University, New York

THE treatment in Secs. 8 and 10 of the interaction between mean and fluctuating amplitudes contains several errors. (These errors do not affect the treatment of the mean field $\bar{u}$ in Sec. 11.) Equation (10.24) actually holds only for $\mathbf{x}=\mathbf{x}'$. Consequently, if $\bar{\psi}(\mathbf{x},t)\neq 0$, a correct closed set of equations which determine $\Psi(\mathbf{x}, t; \mathbf{x}', t')$ in the random-coupling model cannot be obtained in complete analogy to the treatment in Sec. 8. Instead, one must carry out the diagram expansion of $\langle v_\alpha(\mathbf{x})\psi_\alpha^*(\mathbf{x}',t)\rangle$ and thereby obtain simultaneous equations which determine this quantity and $\Psi(\mathbf{x}, t; \mathbf{x}', t')$. The equations are then completed by (10.18), which is correct as given. For models other than the random-coupling model defined by (5.1) and (5.2), the existence of nonvanishing mean amplitudes implies the appearance, in (8.22) and (10.23), of additional classes of terms which were ignored in the text discussion. The omitted terms are bilinear functionals of $\langle b_\alpha q_\alpha^*(t)\rangle$ and $\langle v_\alpha(\mathbf{x})\psi_\alpha^*(\mathbf{x}',t)\rangle$, respectively. They may be obtained explicitly by straightforward extension of the methods described in the text. The diagrammatic representation of the additional terms is simplified if the full symmetry constraints (11.5) are imposed.

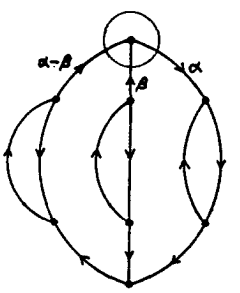


FIG. 1. An example of a diagram in the turbulence problem which does not arise in the random oscillator problem.

In Sec. 11, the paragraph immediately following (11.6) is misleading as worded. It is true, as stated, that the rules for constructing the $C_{2n;\ p}(\alpha, \beta, \alpha-\beta)$ associated with any diagram are the same as in Sec. 4. However, there are additional classes of diagrams in the turbulence problem which do not arise in Sec. 4. These diagrams do not require special treatment. Certain (reducible) classes of such diagrams occur in the random-coupling model, and they are included in the closed equations for this model derived in Sec. 11. An example is shown in Fig. 1.

First Reprinting, November 1962