JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 4, NUMBER 9 SEPTEMBER 1963

Asymptotic Wave Vector and Nonrelativistic Perturbation Theory

S. TEITLER
U. S. Naval Research Laboratory, Washington, D. C.
(Received 21 January 1963)

The dual nature of perturbations in causing both transitions and persistent effects is investigated. A wave vector initially represented in terms of a set of unperturbed eigenvectors is subjected to a nondissipative perturbation. After a long time, it is assumed that this unperturbed wave vector evolves into an asymptotic wave vector in which both persistent and transition effects are present. These two effects are considered separately and the asymptotic wave vector is formally expressed as an expansion in terms of asymptotically stationary states. A time-operator form of nonrelativistic perturbation theory which formally is very similar to the resolvent formalism is presented. In a manner similar to the resolvent formalism, diagonal and nondiagonal contributions to the development operator are considered separately. A further classification of the development operator into asymptotic and nonasymptotic parts is made. This latter form is used to obtain an explicit form for the asymptotic wave vector, and the asymptotically stationary states are identified.

I. INTRODUCTION

SOME time ago Van Hove[1,2] in a series of remarkable papers emphasized the dual role of perturbations in causing transitions on the one hand, and persistent effects on the other. Persistent effects themselves also have a dual character, for under appropriate circumstances, they may be related to cloud effects with accompanying self-energy and renormalization, or to dissipative behavior as in the description of metastable states or both.

In the next section, we express this dual role of perturbations by assuming that the effect of a nondissipative perturbation a very long time after it has become effective may be represented by a so-called asymptotic wave vector. To evaluate the form this physical assumption takes, we develop a perturbation formalism in the third section, which is closely related to the resolvent formalism of Van Hove. The resolvent formalism is based on an integral transform of the development operator. One calculates the resolvent by a perturbative technique and then returns to the time representation by means of an inverse transformation. The resolvent formalism has many advantages in collecting terms to identify persistent and transition effects.[3]

In the present method, we use a time-operator technique which is formally very similar to the resolvent formalism and has many of its advantages. However, one replaces the performance of an integral transform and its inverse by a formulation in terms of time operators and then evaluation of a formula including time operators. This latter procedure presents an advantage for certain problems when the details of the calculation depend in a complicated way on the analytic properties of the resolvent on which the inverse integral transformation must be performed. An example of this will be discussed in Sec. III.

We use operator techniques provided by Mikusin-

[1] L. Van Hove, Physica **21**, 901 (1955), hereafter referred to as I.

[2] L. Van Hove, Physica **22**, 343 (1956), hereafter referred to as II.

[3] See also S. Teitler and R. F. Wallis, J. Math. Phys. **1**, 372 (1960).

ski in his masterful book.[4] To make sense of the procedures to follow, and since Mikusinski's method is apparently not well-known, we recount here some basic concepts of the operational calculus, referring the interested reader to Mikusinski's book for necessary details, further development of his method, and many applications different from our own.

Mikusinski starts from an algebra of convolutions. The convolution $\{c(t)\}$ of the functions $\{a(t)\}$ and $\{b(t)\}$ is defined by the integral

$$\{c(t)\} = \int_0^t a(t-\tau)b(\tau)\,d\tau \equiv \{a(t)\}\{b(t)\}, \qquad (1.1)$$

where $\{a(t)\}$ and $\{b(t)\}$ are functions of class C which are defined and continuous for $0 \leq t < \infty$. Their product is defined to be the convolution. The convolution product is commutative, associative, and distributive.

Note that $\{1\}$ is just the integral operator since, by Eq. (1.1),

$$\{1\}\{f(t)\} = \left\{\int_0^t f(\tau)\,d\tau\right\}. \qquad (1.2)$$

Further, it can be proved that if $\{f(t)\}$ and $\{g(t)\}$ are not identically equal to zero, then neither is their convolution identically equal to zero. This theorem allows the definition of a unique inverse operation to convolution. Thus if

$$\{a\} = \{b\}\{c\},$$

then

$$\{a\}/\{b\} = \{c\}. \qquad (1.3)$$

In general, however, $\{a\}/\{b\}$ will not be a function of class C, so that Mikusinski here introduces the concept of operator to cover this class. To make this class useful he defines the following operations on operators:

$$\{a\}/\{b\} = \{c\}/\{d\}$$

$$\text{if and only if} \quad \{a\}\{d\} = \{b\}\{c\}, \qquad (1.4a)$$

$$(\{a\}/\{b\})(\{c\}/\{d\}) = \{a\}\{c\}/\{b\}\{d\}, \qquad (1.4b)$$

$$\frac{\{a\}}{\{b\}} + \frac{\{c\}}{\{d\}} = \frac{\{a\}\{d\} + \{b\}\{c\}}{\{b\}\{d\}}, \qquad (1.4c)$$

where it is assumed that $\{b\}$ and $\{d\}$ are not identically equal to zero and therefore $\{b\}\ \{d\}$ is not identically equal to zero. The complete analogy between operators and fractions of arithmetic allows operations on operators to be performed in the same way as those on ordinary fractions.

[4] Jan Mikusinski, *Operational Calculus* (Pergamon Press, Inc., New York, 1959).

We consider some useful examples of operators. First consider $\{\alpha\}/\{1\}$ where $\{\alpha\}$ is an arbitrary constant function which has the value α everywhere. Then $\{\alpha\}/\{1\}$ is called a numerical operator and has all the properties of ordinary numbers and may be dealt with as such, e.g.,

$$(\{\alpha\}/\{1\})\{f(t)\} = \alpha\{f(t)\} = \{\alpha f(t)\}. \qquad (1.5)$$

Now consider the inverse of the integral operator $\{1\}$ which is denoted by s:

$$s\{1\} = \{1\}s = 1. \qquad (1.6)$$

It follows that if a function $a = \{a(t)\}$ has a derivative $a' = \{a'(t)\}$, continuous for $0 \leq t < \infty$, then

$$sa = a' + a(0), \qquad (1.7)$$

where $a(0)$ is *not* a function but is the value of the function a at the point $t = 0$. Thus, for example,

$$s\{e^{\alpha t}\} = 1 + \alpha\{e^{\alpha t}\},$$

or

$$\{e^{\alpha t}\} = 1/(s-\alpha). \qquad (1.8)$$

We mean by a rational operator, the fraction

$$\frac{\gamma_m s^m + \cdots + \gamma_1 s + \gamma_0}{\delta_n s^n + \cdots + \delta_1 s + \delta_0}, \qquad (m < n), \qquad (1.9)$$

where $\gamma_m, \cdots, \gamma_0, \delta_n, \cdots, \delta_0$ are complex numbers and $\delta_n \neq 0$. Using the operations (1.4) it can be shown that if two polynomials of the operator s are equal, then their coefficients are respectively equal. From this equality, it may be shown that the equality of two rational operators remains valid if s is replaced by any number (real or complex) for which the denominators do not vanish. Frequently a rational operator may be expanded, but every such expansion must be individually justified.

In the next section, as indicated previously, we discuss asymptotically valid contributions to a perturbed wave vector. The asymptotic wave vector is defined and its relation to asymptotically stationary states and their corresponding self-energies is indicated. In the third section we develop the present method of perturbation theory. Diagonal and nondiagonal contributions to the development operator are obtained in a manner similar to that used by Van Hove, employing the resolvent formalism. A further classification of the development operator into asymptotic and nonasymptotic parts is made and expressed in terms of the diagonal and nondiagonal contributions. In the fourth section, we restrict ourselves to nondissipative

behavior and find the explicit form of the asymptotic wave vector. In particular, we identify there the asymptotic stationary states.

II. UNPERTURBED AND PERTURBED WAVE VECTORS

We consider a large quantum system characterized by a Hamiltonian

$$H = H^0 + \lambda V, \tag{2.1}$$

where H^0 and V are independent of time. The unperturbed Hamiltonian H^0 has orthonormal eigenstates $|\alpha\rangle$ with eigenvalues E^0_α. The Schrödinger equation for a ket state, setting $\hbar = 1$, is

$$(H^0 + \lambda V)\ |\psi(t)\rangle = i(\partial/\partial t)\ |\psi(t)\rangle. \tag{2.2}$$

We may expand $|\psi(t)\rangle$:

$$|\psi\rangle = \int d\alpha_1 c(\alpha_1, t)\ |\alpha\rangle. \tag{2.3}$$

At $t = 0$, we have

$$|\psi(0)\rangle = \int d\alpha_1 c(\alpha_1)\ |\alpha_1\rangle. \tag{2.4}$$

If $\lambda = 0$, then

$$|\psi(t)\rangle_u = \int d\alpha_1 c(\alpha_1) \exp(-iE^0_{\alpha_1} t)\ |\alpha_1\rangle. \tag{2.5}$$

To treat the case when $\lambda \neq 0$, we suppose now that λ jumps from zero to its value λ at $t = 0$, and $|\psi(0)\rangle$ is given by Eq. (2.4). We use the techniques discussed in the introduction to solve the Schrödinger equation for positive times. Although we shall develop the perturbation formalism to include both cloud and dissipative persistent effects in the present paper, we confine our discussion to nondissipative cloud effects.

We assume that after a sufficiently long time the wave vector $|\psi(t)\rangle$ evolves from the original wave vector in the sense that it may be represented as follows

$$|\psi(t)\rangle = \int d\alpha c(\alpha_1) e^{-iE_{\alpha_1} t}\ |\alpha\rangle_{\mathrm{LT}}. \tag{2.6}$$

Here the label LT stands for long time. We designate $|\psi(t)\rangle$ the asymptotic wave vector. While $|\psi(t)\rangle$ has the same form as $|\psi(t)\rangle_u$ given in Eq. (2.5), it should be noted that E_{α_1} is not in general equal to $E^0_{\alpha_1}$ and the $|\alpha\rangle_{\mathrm{LT}}$ are not the same as the $|\alpha\rangle$.

The asymptotic wave vector $|\psi(t)\rangle$ is closely related to the asymptotic wave packet of Van Hove I, which is an expansion in which only the persistent effects are included. The asymptotic wave packet $|\Phi(t)\rangle$ may be written in the form

$$|\Phi(t)\rangle = \int d\alpha c(\alpha) e^{-iE_\alpha t}\ |\alpha\rangle_{\mathrm{AS}}. \tag{2.7}$$

Here the $|\alpha\rangle_{\mathrm{AS}}$ are the asymptotically stationary states representing the effect of the cloud on the respective α states. The $|\alpha\rangle_{\mathrm{AS}}$ are asymptotically orthonormal and have the properties indicated by Van Hove, (I, Sec. 5.) Van Hove writes down the asymptotic wave packets at both $t = -\infty$ and $t = +\infty$. He then relates the transition aspect of the perturbation to changes in the spectral composition of the asymptotic wave packet as t passes from $t = -\infty$ to $t = +\infty$, thereby identifying the S matrix.

On the other hand, we are emphasizing the relation between the asymptotic wave vector which includes effects of both the transition and persistent aspects of the perturbation, and the unperturbed wave vector. We follow directly the effect of the perturbation on an unperturbed wave vector after a very long time as it evolves into what might be termed its "dressed" configuration. In general the $|\alpha\rangle_{\mathrm{LT}}$ contain both transition and cloud effects. We denote the states which are coupled to the α state by transitions as the α star. We may formally separate cloud and star effects by introducing the star coefficient operator $c_{\mathrm{op}}(\alpha)$. We write

$$|\psi(t)\rangle = \int d\alpha c_{\mathrm{op}}(\alpha) e^{-iE_\alpha t}\ |\alpha\rangle_{\mathrm{AS}}, \tag{2.8}$$

where

$$c(\alpha)\ |\alpha\rangle_{\mathrm{LT}} = c_{\mathrm{op}}(\alpha)\ |\alpha\rangle_{\mathrm{AS}}. \tag{2.9}$$

When there are no transition effects, $|\psi(t)\rangle$ must be equal to the asymptotic wave packet, i.e., $c_{\mathrm{op}}(\alpha)$ reduces to $c(\alpha)$.

After carrying through our perturbative procedure we will be able to explicitly identify $c_{\mathrm{op}}(\alpha)$, $|\alpha\rangle_{\mathrm{AS}}$ and E_α in a straightforward manner. The quantities E_α may be recognized as the energy characteristic of the $|\alpha\rangle_{\mathrm{AS}}$ states and are called the complete energies. The difference $E_\alpha - E^0_\alpha$ may be identified as the self-energy.

III. PERTURBATION THEORY

We define the development operator A by the equation

$$|\psi(t)\rangle = A\ |\psi(0)\rangle. \tag{3.1}$$

Using Eq. (2.2), we then obtain

$$(H^0 + \lambda V)A = i\, dA/dt = dA/du, \tag{3.2}$$

where $u = -it$, and we have for notational simplicity written $\{f(u)\} = f(u)$. Employing the methods discussed in the introduction and writing matrix elements explicitly, we obtain

$$(S - E^0_{\alpha_1})A_{\alpha_1\alpha_2}(s) = A_{\alpha_1\alpha_2}(0) + \lambda \int d\alpha_3 V_{\alpha_1\alpha_3} A_{\alpha_3\alpha_2}(s). \quad (3.3)$$

We note that $A_{\alpha_1\alpha_2}(0) = \delta(\alpha_1 - \alpha_2)$ and write

$$A_{\alpha_1\alpha_2}(s) = \frac{1}{s - E^0_{\alpha_2}} \delta(\alpha_1 - \alpha_2) + \frac{\lambda}{s - E^0_{\alpha_1}} \int d\alpha_3 V_{\alpha_1\alpha_3} A_{\alpha_3\alpha_2}(s). \quad (3.4)$$

We solve Eq. (3.4) using an iterative procedure equivalent to the time iteration of the development operator:

$$A^{(0)}_{\alpha_1\alpha_2} = [1/(s - E^0_{\alpha_2})]\delta(\alpha_1 - \alpha_2), \quad (3.5a)$$

$$A^{(1)}_{\alpha_1\alpha_2} = [\lambda/(s - E^0_{\alpha_1})]V_{\alpha_1\alpha_2}[1/(s - E^0_{\alpha_2})], \quad (3.5b)$$

$$A^{(2)}_{\alpha_1\alpha_2} = \frac{\lambda^2}{s - E^0_{\alpha_1}} \times \int d\alpha_3 V_{\alpha_1\alpha_3} \frac{1}{s - E^0_{\alpha_3}} V_{\alpha_3\alpha_2} \frac{1}{s - E^0_{\alpha_2}}, \quad (3.5c)$$

etc. Then,

$$A_{\alpha_1\alpha_2}(s) = \sum_{n=0}^{\infty} A^{(n)}_{\alpha_1\alpha_2}(s). \quad (3.5)$$

As can be immediately seen by substituting for the operators involving s their values as functions of $u = -it$ and carrying out the indicated convolutions, this iteration procedure is equivalent to the time iteration of the development operator. We shall assume convergence for such a procedure.

The s-operator form of the development operator, Eq. (3.5), is very remindful of the resolvent of Van Hove. We may be guided by this similarity in form to consider separately the diagonal and nondiagonal parts of $A_{\alpha_1\alpha_2}$. It is here that we diverge from the standard time-iterative techniques, since their use makes it difficult to identify persistent effects. We seek to alleviate this difficulty by writing

$$A_{\alpha_1\alpha_2} = A^{\mathrm{D}}_{\alpha_1\alpha_2} + A^{\mathrm{ND}}_{\alpha_1\alpha_2}, \quad (3.6)$$

where D denotes diagonal and ND, nondiagonal part.

Using Eq. (3.5), we may write

$$A^{\mathrm{D}}_{\alpha_1\alpha_2} = \left[1 + \lambda \frac{1}{s - E^0_{\alpha_2}} \Sigma^{\mathrm{D}}_{\alpha_2}\right] \frac{1}{s - E^0_{\alpha_2}} \delta(\alpha_1 - \alpha_2), \quad (3.7)$$

where

$$\lambda\Sigma^{\mathrm{D}}_{\alpha_2} = \left[\lambda V_{\alpha_2\alpha_2} + \lambda^2 \times \int d\alpha_3 V_{\alpha_2\alpha_3} \frac{1}{s - E^0_{\alpha_3}} V_{\alpha_3\alpha_2} + \cdots\right]_{\mathrm{D}} \quad (3.8)$$

has contributions from all iterations for which the far right- and the far left-hand states are the same. It should be noted that in defining corresponding quantities, Van Hove does not use the first-order term $\lambda V_{\alpha\alpha}$, so that his quantities are of order λ^2. Also, where we have the time operator $(s - E^0_\alpha)^{-1}$, Van Hove has the quantum-mechanical operator $(l - H^0)^{-1}$ where l is the transform variable.

Define now

$$\lambda G_{\alpha_2}(s) = \left[\lambda V_{\alpha_2\alpha_2} + \lambda^2 \times \int d\alpha_3 V_{\alpha_2\alpha_3} \frac{1}{s - E^0_{\alpha_3}} V_{\alpha_3\alpha_2} + \cdots\right]_{\mathrm{SD}}, \quad (3.9)$$

where SD means simply diagonal in the sense there are no intermediate states equal to the far left- (or right-) hand state. We may write

$$\lambda\Sigma^{\mathrm{D}}_{\alpha_2} = \frac{(s - E^0_{\alpha_2})(\lambda G_{\alpha_2})}{s - E^0_{\alpha_2} - \lambda G_{\alpha_2}(s)}. \quad (3.10)$$

Thus

$$A^{\mathrm{D}}_{\alpha_1\alpha_2} = \frac{1}{s - E^0_{\alpha_2} - \lambda G_{\alpha_2}(s)} \delta(\alpha_1 - \alpha_2). \quad (3.11)$$

In I and II, Van Hove collected terms in a different manner than indicated here. He defined G, not in simply diagonal terms, but in so-called irreducibly diagonal terms in which all diagonal components of the intermediate states are also abstracted out and collected. Then the final expression is one in which none of the intermediate states are equal to each other. The physical problem will determine which expansion should be used, for in given circumstances one or the other leads to more rapid convergence of the perturbation procedure. However it might be noted that, in using irreducibly diagonal definitions, some care must be employed inasmuch as terms which do not actually occur are usually added to compensate for those already abstracted in any collection of diagonal terms prior to a given collection as one iterates to reach the irreducibly diagonal form. For many potentials, however, this latter difficulty causes no real problem and for lower-order expansions, in general, there is no real difficulty. However, for definiteness and simplicity, we choose to carry through our formula-

tion assuming that the physical problem of interest is appropriate to the use of the simple diagonal G_α. In general, then, we will make no collections of diagonal components between intermediate states. Actually, Van Hove in later work[5] has also used the simply diagonal definition for G when the physical situation indicated its use.

The nondiagonal part of $A_{\alpha_1\alpha_2}$ may be written

$$A^{\mathrm{ND}}_{\alpha_1\alpha_2} = \frac{1}{s - E^0_{\alpha_1}} \lambda \Sigma^{\mathrm{ND}}_{\alpha_1\alpha_2} \frac{1}{s - E^0_{\alpha_2}}, \qquad (3.12)$$

where $\lambda \sum^{\mathrm{ND}}_{\alpha_1\alpha_2}$ has the same form as $\lambda \sum^{\mathrm{D}}_{\alpha_2}$ given in Eq. (3.8), except that it has contributions from all iterations for which the far right- and far left-hand states are not the same. Equation (3.12) may be put into a suggestive form if all the terms in $\lambda \sum^{\mathrm{ND}}_{\alpha_1\alpha_2}$, with intermediate states diagonal with the far right-hand state, are abstracted and collected in a manner similar to Eq. (3.10). Then

$$A^{\mathrm{ND}}_{\alpha_1\alpha_2} = \frac{1}{s - E^0_{\alpha_1}} \Sigma^{\mathrm{RND}}_{\alpha_1\alpha_2} \frac{1}{s - E^0_{\alpha_2} - \lambda G_{\alpha_2}(s)}, \qquad (3.13)$$

where RND means right nondiagonal in the sense that the term is nondiagonal, but also there are no intermediate states equal to the far right-hand state, here α_2.

$A_{\alpha_1\alpha_2}$ now takes the form

$$\begin{aligned} A_{\alpha_1\alpha_2} &= A^{\mathrm{D}}_{\alpha_1\alpha_2} + A^{\mathrm{ND}}_{\alpha_1\alpha_2} \\ &= \left[\delta(\alpha_1 - \alpha_2) + \frac{\lambda}{s - E^0_{\alpha_1}} \Sigma^{\mathrm{RND}}_{\alpha_1\alpha_2}\right] \\ &\quad \times \left[\frac{1}{s - E^0_{\alpha_2} - \lambda G_{\alpha_2}(s)}\right]. \end{aligned} \qquad (3.14)$$

We note that as $\lambda \to 0$, the time dependence of $A_{\alpha_1\alpha_2}$ is controlled by the last factor in the brackets and has a value consistent with Eq. (2.5). Comparison of the asymptotic wave vector when $\lambda \neq 0$ with the unperturbed wave vector leads us to expect that this last factor in Eq. (3.14) will play a dominant role in the time evolution of the asymptotic wave vector or any vector similar to it in which dissipative effects are included. We seek a solution for $A_{\alpha_1\alpha_2}$ of the form

$$A_{\alpha_1\alpha_2} = [A_1]_{\alpha_1\alpha_2} \frac{1}{s - E^0_{\alpha_2} - \lambda G_{\alpha_2}(s)} + [A(s)]^{\mathrm{NAS}}_{\alpha_1\alpha_2}, \qquad (3.15)$$

where $[A_1]_{\alpha_1\alpha_2}$ is constant in time and $[A(s)]^{\mathrm{NAS}}_{\alpha_1\alpha_2}$ is a function of time. We call the first term the asymptotic contribution, and the second term the nonasymptotic contribution to the development operator. We distinguish between the two parts because we expect the closure of $|\psi(t)\rangle = A|\psi(0)\rangle$ with an asymptotic wave packet, i.e., $\langle\Phi(t)|\psi(t)\rangle$, to have a nonvanishing contribution in the long-time limit only for that part of $|\psi(t)\rangle$ which arises from the development due to the asymptotic contribution. We shall return to this point later when we discuss the form of $[A(s)]^{\mathrm{NAS}}_{\alpha_1\alpha_2}$.

We may equate Eq. (3.14) and Eq. (3.15) so that

$$\begin{aligned} &[A_1]_{\alpha_1\alpha_2} + [A]^{\mathrm{NAS}}_{\alpha_1\alpha_2}[s - E^0_{\alpha_2} - \lambda G_{\alpha_2}(s)] \\ &\quad = \delta(\alpha_1 - \alpha_2) + \frac{\lambda}{s - E^0_{\alpha_1}} \Sigma^{\mathrm{RND}}_{\alpha_1\alpha_2}(s). \end{aligned} \qquad (3.16)$$

From our discussion in the introduction, we know that Eq. (3.16) is valid when any number (real or complex) for which the individual terms are defined is substituted for s. We may then determine $[A_1]_{\alpha_1\alpha_2}$ by using the number for which the condition

$$s - E^0_{\alpha_2} - \lambda G_{\alpha_2}(s) = 0 \qquad (3.17)$$

is valid. We label this number E_{α_2} and assume this solution is unique. We shall suppose that $[A]^{\mathrm{NAS}}_{\alpha_1\alpha_2}$ has no cancelling zero in the denominator. Also, since it is required that $\alpha_1 \neq \alpha_2$ in the second term on the right-hand side of Eq. (3.16), it may then be shown that E_{α_2} cannot equal $E^0_{\alpha_1}$ for the class of perturbations being discussed, so that this term is defined. This follows since if $\sum^{\mathrm{RND}}_{\alpha_1\alpha_2}$ does not vanish, we expect $|\alpha_1\rangle$ will occur as an intermediate state in the definition of $G_{\alpha_2}(s)$. If such is the case and any approximation to E_{α_2} can equal $E^0_{\alpha_1}$, there would be an imaginary part of G_{α_2} and hence an imaginary part of the final E_{α_2}. E_{α_2} with an imaginary part cannot equal $E^0_{\alpha_1}$ which is real. We may write then

$$[A_1]_{\alpha_1\alpha_2} = \delta(\alpha_1 - \alpha_2) + \frac{\lambda}{E_{\alpha_2} - E^0_{\alpha_1}} \Sigma^{\mathrm{RND}}_{\alpha_1\alpha_2}(E_{\alpha_2}). \qquad (3.18)$$

Let us consider the condition Eq. (3.17) more closely. We know from Van Hove's work that $G_{\alpha_2}(s)$ is holomorphic for s complex, and with Van Hove, we may assume $G_{\alpha_2}(E_{\alpha_2})$ holomorphic when G_{α_2}, and therefore E_{α_2}, are real. In this latter case, Eq. (3.17) is equivalent to Van Hove's condition for singularity of the resolvent at E_{α_2}. The situation is different when $G_{\alpha_2}(s)$ is complex so that E_{α_2} is complex. Then Eq. (3.17) is not equivalent to Van Hove's singularity condition. Instead it corresponds to the case when the diagonal part of the resolvent is bounded and has a finite dis-

[5] L. Van Hove, Lecture Notes, University of Washington, Seattle, Washington (1958).

continuity across the real axis. This is the dissipative case which in practice is usually discussed to restricted order in the perturbation. For example, one usually treats the case of small dissipations in the neighborhood of what would have been a pole on the real axis if the dissipation didn't exist.[6] This leads to the concept of quasipole in the resolvent inasmuch as the resolvent must be holomorphic off the real axis. On the other hand, the occurrence of the zero of Eq. (3.17) for E_{α_2} off the real axis provides no special difficulty in the evaluation of $[s - E^0_{\alpha_2} - \lambda G_{\alpha_2}(s)]^{-1}$, except that this term is defined only for s an operator where $G_{\alpha_2}(s)$ is the time-operator correspondence of its function counterpart.

Thus to sum up, both the present method and the resolvent formalism require an investigation of the analytic properties of G. The resolvent formalism further requires the knowledge of the analytic properties of the resolvent itself in order to carry out the inverse transform and this may present added difficulties. On the other hand, it is more cumbersome to use the present method to treat both positive and negative times together. In general it will be the choice of problem that will determine which method is most advantageous.

A further approximation is usually made in the asymptotic limit. Not only does the time dependence depend on the complete energy in an essential way, but also it has an exponential form. This approximation may be expressed in the present case by requiring that G_{α_2} be linear in s. Then in terms of time operators we may write

$$\begin{aligned}[s - E^0_{\alpha_2} - \lambda G_{\alpha_2}(s)]^{-1} &= (s - E_{\alpha_2})^{-1}[1 - \lambda G'_{\alpha_2}(E_{\alpha_2})]^{-1} \\ &= \frac{1}{1 - \lambda G'_{\alpha_2}(E_{\alpha_2})}\{e^{-iE_{\alpha_2}t}\}. \end{aligned} \tag{3.19}$$

Here the prime signifies the derivative with respect to the complex variable s and we have made a correspondence between G_{α_2} in analytic and operator form. Then the asymptotic contribution to the development operator has an exponential time dependence. If there is a dissipative part, it may be shown that, barring exceptional situations, the definition of $G_{\alpha_2}(s)$ assures that this will indeed be a decaying solution. We define the asymptotic development matrix:

$$A^{\mathrm{As}}_{\alpha_1\alpha_2}(t) = [A_1]_{\alpha_1\alpha_2}\frac{1}{1 - \lambda G'_{\alpha_2}(E_{\alpha_2})}e^{-iE_{\alpha_2}t}. \tag{3.20}$$

[6] See, e.g., N. M. Hugenholtz, *The Many Body Problem* (John Wiley & Sons, Inc., New York, 1959), p. 33.

We investigate now the nature of $[A]^{\mathrm{NAS}}_{\alpha_1\alpha_2}$. To do this we return to Eq. (3.14) and rewrite it in the form

$$\begin{aligned}A_{\alpha_1\alpha_2} = \left[\delta(\alpha_1 - \alpha_2) + \frac{\lambda}{s - E^0_{\alpha_1}}\Sigma^{\mathrm{RND}}_{\alpha_1\alpha_2}(s)\right] \\ \times \frac{1}{1 - \lambda G'_{\alpha_2}(E_{\alpha_2})}\frac{1}{s - E_{\alpha_2}}.\end{aligned} \tag{3.21}$$

We note that $\lambda \sum^{\mathrm{RND}}_{\alpha_1\alpha_2}(s)$ has a form similar to $\lambda G(s)$ and we assume that it too is linear in s in the asymptotic limit of t very large. We further assume it has a derivative at E_{α_2} so that

$$\Sigma^{\mathrm{RND}}_{\alpha_1\alpha_2}(s) = \Sigma^{\mathrm{RND}}_{\alpha_1\alpha_2}(E_{\alpha_2}) + \Sigma^{\mathrm{RND}'}_{\alpha_1\alpha_2}(E_{\alpha_2})[s - E_{\alpha_2}]. \tag{3.22}$$

We also note that

$$\begin{aligned}\frac{1}{s - E^0_{\alpha_1}}\frac{1}{s - E_{\alpha_2}} \\ = \frac{1}{E_{\alpha_2} - E^0_{\alpha_1}}\left[\frac{1}{s - E_{\alpha_2}} - \frac{1}{s - E^0_{\alpha_1}}\right].\end{aligned} \tag{3.23}$$

Then, recalling Eq. (3.18), we may identify $[A]^{\mathrm{NAS}}_{\alpha_1\alpha_2}$ from Eq. (3.21):

$$\begin{aligned}[A]^{\mathrm{NAS}}_{\alpha_1\alpha_2} &= \frac{\lambda}{s - E^0_{\alpha_1}}\left[\frac{-1}{E_{\alpha_2} - E^0_{\alpha_1}}\Sigma^{\mathrm{RND}}_{\alpha_1\alpha_2}(E_{\alpha_2}) \right. \\ &\quad \left. + \Sigma^{\mathrm{RND}}_{\alpha_1\alpha_2}(E_{\alpha_2})\right][1 - \lambda G'_{\alpha_2}(E_{\alpha_2})]^{-1} \\ &\equiv \lambda[A_2]_{\alpha_1\alpha_2}e^{-iE^0_{\alpha_1}t},\end{aligned} \tag{3.24}$$

where $[A_2]_{\alpha_1\alpha_2}$ is a constant. Closure of that part of the wave vector which evolves by means of $[A]^{\mathrm{NAS}}$ with an asymptotic wave packet, provides, in general, an oscillating term within a sum over states when G_{α_1} is nonvanishing and has a real part. We shall assume that the perturbations of interest are such that this is the case except perhaps for a subset of α_1 states of measure zero. Then this closure, $\langle\Phi(t)|\,[A]^{\mathrm{NAS}}\,|\psi(0)\rangle$ is negligible in the limit of long times. Further, the cancellation of such an oscillating term when the linerar approximation is not made is most unlikely, so that we may assume this closure vanishes in all cases.

It is interesting to note that the form of Eq. (3.24) also implies that the supposition following Eq. (3.17), namely that $[A]^{\mathrm{NAS}}_{\alpha_1\alpha_2}$ has no cancelling zero in denominator, is correct.

A similar perturbation scheme may be carried out for the adjoint development operator. Thus, corresponding to Eq. (3.11), we would have

$$A^{\dagger\mathrm{D}}_{\alpha_1\alpha_2} = \frac{1}{s + E^0_{\alpha_1} + \lambda\mathcal{G}_{\alpha_1}(s)}\delta(\alpha_1 - \alpha_2), \tag{3.25}$$

where now α_2 is the dummy index and

$$(-\lambda)\mathcal{G}_{\alpha_1}(s) = \Big[(-\lambda)V_{\alpha_1\alpha_1} + (-\lambda)^2 \times \int d\alpha_3 V_{\alpha_1\alpha_3}\frac{1}{s+E^0_{\alpha_3}}V_{\alpha_3\alpha_1} + \cdots\Big]_{SD}. \quad (3.26)$$

Similarly, corresponding to Eq. (3.13), we would have

$$A^{+ND}_{\alpha_1\alpha_2} = \frac{1}{s+E^0_{\alpha_1}+\lambda\mathcal{G}_{\alpha_1}(s)}(-\lambda) \times \Gamma^{LND}_{\alpha_1\alpha_2}\frac{1}{s+E^0_{\alpha_2}}, \quad (3.27)$$

where

$$(-\lambda)\Gamma^{LND}_{\alpha_1\alpha_2} = \Big[(-\lambda)V_{\alpha_1\alpha_2} + (-\lambda)^2 \times \int d\alpha_3 V_{\alpha_1\alpha_3}\frac{1}{s+E^0_{\alpha_3}}V_{\alpha_3\alpha_2} + \cdots\Big]_{LND}. \quad (3.28)$$

Here, LND means nondiagonal but also no intermediate states equal to the far left-hand state.

We again seek a form of the development operator more compatible with the asymptotic wave vector. We write

$$A^{\dagger}_{\alpha_1\alpha_2} = \frac{1}{s+E^0_{\alpha_1}+\lambda\mathcal{G}_{\alpha_2}(s)}[A^{\dagger}]_{\alpha_1\alpha_2} + [A^{\dagger}(s)]^{NAS}_{\alpha_1\alpha_2}, \quad (3.29)$$

where

$$[A^{\dagger}]_{\alpha_1\alpha_2} = \delta(\alpha_1-\alpha_2) - \lambda\Gamma^{LND}_{\alpha_1\alpha_2}(-\mathcal{E}_{\alpha_1})\frac{1}{-\mathcal{E}_{\alpha_1}+E^0_{\alpha_2}}. \quad (3.30)$$

Here $s = -\mathcal{E}_{\alpha_1}$ satisfies

$$s + E^0_{\alpha_1} + \lambda\mathcal{G}_{\alpha_1}(s) = 0. \quad (3.31)$$

Again we assume the asymptotic limit of an exponential time dependence, so that

$$\frac{1}{s+E^0_{\alpha_1}+\lambda\mathcal{G}_{\alpha_1}(s)} = \frac{1}{1+\lambda\mathcal{G}'_{\alpha_1}(-\mathcal{E}_{\alpha_1})}\{e^{i\mathcal{E}_{\alpha_1}t}\}. \quad (3.32)$$

Here again when dissipative effects are present, the properties of $\mathcal{G}_{\alpha_1}(s)$ generally assure a decaying solution. We are led then to define the adjoint asymptotic development matrix:

$$(A^{\dagger})^{AS}_{\alpha_1\alpha_2} = \{e^{i\mathcal{E}_{\alpha_1}t}\}\frac{1}{1+\lambda\mathcal{G}'_{\alpha_1}(-\mathcal{E}_{\alpha_1})}[A^{\dagger}_1]_{\alpha_1\alpha_2}. \quad (3.33)$$

IV. THE ASYMPTOTIC WAVE VECTOR

We have completed our use of the operational calculus and deal now with ordinary functions of time. We investigate the result of the application of the asymptotic development operator to the initial wave vector given by Eq. (2.4) in the case when $\mathcal{G}_\alpha(-\mathcal{E}_\alpha)$ and $G_\alpha(E_\alpha)$ are pure real, i.e., when there are no dissipative effects present. As Van Hove pointed out (see especially I, Sec. 4), this reality and the assumed holomorphic property of G_{α_2} implies that there are restrictions on the integrals over intermediate state arising in the definition of G_α. These restrictions are in general different for different α. We will proceed with the understanding that such restrictions are included as a matter of course when we specify that G_α and $\mathcal{G}_\alpha$ are real and holomorphic. Then $G_\alpha(E_\alpha)$ and $\mathcal{G}_\alpha(-\mathcal{E}_\alpha)$ are equal and real so that $E_\alpha = \mathcal{E}_\alpha$ is real. Also it follows that $-G'_\alpha(E_\alpha) = \mathcal{G}'(-\mathcal{E}_\alpha)$ so that Eq. (3.19) and Eq. (3.32) have the same real factor multiplying the exponential term.

We may write an (unrenormalized) asymptotic wave vector $|\Xi(t)\rangle$:

$$|\Xi(t)\rangle = A^{As}|\psi(0)\rangle = \int d\alpha_2 \frac{c(\alpha_2)e^{-iE\alpha_2 t}}{1-\lambda G'_{\alpha_2}(E_{\alpha_2})} \times\Big[1+\frac{\lambda}{E_{\alpha_2}-H^0}\Sigma^{RND}(E_{\alpha_2})\Big]|\alpha_2\rangle. \quad (4.1)$$

Similarly we have

$$\langle\Xi(t)| = \int d\alpha_1 \frac{c^{\dagger}(\alpha_1)e^{iE\alpha_1 t}}{1-\lambda G'_{\alpha_1}(E_{\alpha_1})}\langle\alpha_1| \times\Big[1-\lambda\Gamma^{LND}\frac{1}{H^0-E_{\alpha_1}}\Big]. \quad (4.2)$$

We consider first the simple case when the complete energies of different states cannot be equal, i.e., $E_{\alpha_1} \neq E_{\alpha_2}$ unless α_1 and α_2 are the same. Then assuming t is sufficiently large that we may use Riemann–Lebesgue lemma to eliminate oscillating terms, we find

$$\langle\Xi(t)\mid\Xi(t)\rangle = \int d\alpha \frac{c^{\dagger}(\alpha)c(\alpha)}{[1-\lambda G'_\alpha(E_\alpha)]^2} \times\Big[1-\int d\alpha_3\lambda^2\Gamma^{LND}_{\alpha\alpha_3}(-E_\alpha) \times\frac{1}{E^0_{\alpha_3}-E_\alpha}\frac{1}{E_\alpha-E^0_{\alpha_3}}\Sigma^{RND}_{\alpha_3\alpha}(E_\alpha)\Big]. \quad (4.3)$$

However, we note that

$$\lambda G'_\alpha(E_\alpha) = \int d\alpha_3 \lambda^2 \Gamma^{\mathrm{LND}}_{\alpha\alpha_3}(-E_\alpha) \times \frac{1}{E^0_{\alpha_3} - E_\alpha} \frac{1}{E_\alpha - E^0_{\alpha_3}} \Sigma^{\mathrm{RND}}_{\alpha_3\alpha}(E_\alpha). \quad (4.4)$$

Then Eq. (4.3) becomes

$$\langle \Xi(t) \mid \Xi(t) \rangle = \int \frac{c^\dagger(\alpha)c(\alpha)}{1 - \lambda G'(E_\alpha)} d\alpha. \quad (4.5)$$

Except for the occurrence of the factor

$$[1 - \lambda G'(E_\alpha)]^{-1}$$

within the itegral, Eq. (4.5) has the form of the closure of an asymptotic wave packet. This is not surprising since we are dealing with the case in which a given α state is not coupled to any other state. To arrive at a formulation in more complete agreement with the asymptotic wave packet in this case, we consider the renormalized asymptotic development matrix

$$D_{\alpha_1\alpha_2} = A^{\mathrm{AS}}_{\alpha_1\alpha_2}[1 - \lambda G'_{\alpha_2}(E_{\alpha_2})]^{\frac{1}{2}}. \quad (4.6)$$

Then we write the (renormalized) asymptotic wave vector in the form

$$|\psi(t)\rangle = D\,|\psi(0)\rangle = \int d\alpha_2\, |\alpha_2\rangle_{\mathrm{LT}} c(\alpha_2) e^{-iE_{\alpha_2}t}, \quad (4.7)$$

where $|\alpha\rangle_{\mathrm{LT}}$ represents the following states:

$$|\alpha_2\rangle_{\mathrm{LT}} = N^{\frac{1}{2}}(\alpha_2) \times \left[1 + \frac{\lambda}{E_{\alpha_2} - H^0} \Sigma^{\mathrm{RND}}(E_{\alpha_2})\right] |\alpha_2\rangle. \quad (4.8)$$

Here

$$N^{\frac{1}{2}}(\alpha_2) = [1 - \lambda G'_{\alpha_2}(E_{\alpha_2})]^{-\frac{1}{2}} \quad (4.9)$$

represents the so called eigenvector renormalization factor.

An exactly similar procedure leads us to consider the (renormalized) asymptotic adjoint development operator, and allows us to write

$$\langle \psi(t)| = \int d\alpha_1\, {}_{\mathrm{LT}}\langle \alpha_1|\, c^\dagger(\alpha_1) e^{iE_{\alpha_1}t}. \quad (4.10)$$

Here

$${}_{\mathrm{LT}}\langle \alpha_1| = N^{\frac{1}{2}}(\alpha_1)\langle \alpha_1| \times \left[1 - \lambda \Gamma^{\mathrm{LND}}(-E_{\alpha_1}) \frac{1}{H^0 - E_{\alpha_1}}\right]. \quad (4.11)$$

Again we consider the case when the complete energies of different states cannot be equal. Then, for t sufficiently large,

$$\langle \psi(t) \mid \psi(t) \rangle = \int d\alpha c^\dagger(\alpha) c(\alpha). \quad (4.12)$$

Thus in this case if we identify $|\alpha\rangle_{\mathrm{LT}}$ as $|\alpha\rangle_{\mathrm{AS}}$, and $c(\alpha)$ as $c_{\mathrm{op}}(\alpha)$, we see that Eq. (4.7) now has the form of the asymptotic wave packet discussed in Sec. II. Of course we have yet to show that the $|\alpha\rangle_{\mathrm{AS}}$ so identified actually have the properties of asymptotically stationary states. We will return to this point later after discussing the more general case. For the present however, we can note that the $|\alpha\rangle_{\mathrm{AS}}$ defined here are indeed asymptotically orthonormal as indicated by Eq. (4.12).

It should be emphasized that in obtaining the property of asymptotic orthonormality, we have primarily used the Riemann–Lebesgue lemma rather than the vector properties of the unperturbed states. In the more general case when complete energies of different states may be equal, the Riemann–Lebesgue lemma does not insure asymptotically orthonormality of the $|\alpha\rangle_{\mathrm{LT}}$ and they cannot be identified with the $|\alpha_{\mathrm{AS}}$. In using the Riemann–Lebesgue lemma there are significant formal complications related to the fact that the complete energy restrictions arise only in closure. These complete energy restrictions lead to nonvanishing terms only when the complete energies of the far right-hand and far left-hand states are equal. This can occur in two ways, i.e., when the term is diagonal and when the two states belong to the same star. We may write then

$$e^{-iE_\alpha t} \to e^{-iE_\alpha t}[\mathfrak{D}_\alpha + \mathfrak{N}_\alpha]. \quad (4.13)$$

Here the $\mathfrak{D}$ and $\mathfrak{N}$ project out far right-hand and far left-hand states, which are respectively diagonal and nondiagonal, with equal complete energies. They have the following properties:

$$\mathfrak{D}_\alpha = \begin{cases} 1 & \text{if far left-hand and far right-hand states are equal to } \alpha, \\ 0 & \text{otherwise,} \end{cases} \quad (4.14a)$$

$$\mathfrak{N}_\alpha = \begin{cases} 1 & \text{if far left-hand and far right-hand states are different with same complete energy } E_\alpha, \\ 0 & \text{otherwise,} \end{cases} \quad (4.14b)$$

$$\mathfrak{D}_{\alpha_1}\mathfrak{D}_{\alpha_2} = \mathfrak{D}_{\alpha_1}\delta(\alpha_1 - \alpha_2), \quad (4.14c)$$

$$\mathfrak{D}\mathfrak{N} = \mathfrak{N}\mathfrak{D} = 0, \quad (4.14d)$$

$$\mathfrak{N}_{\alpha_1}\mathfrak{N}_{\alpha_2} = \begin{cases} 1 & \text{if the far left-hand state } \alpha_1 \text{ and the far right-hand state } \alpha_2 \text{ are different but have equal complete energies,} \\ 0 & \text{otherwise.} \end{cases} \quad (4.14e)$$

With these operators we may write the asymptotic wave vector in the form

$$|\psi(t)\rangle = \int d\alpha c(\alpha) e^{-iE_\alpha t} N^{\frac{1}{2}}(\alpha) \times \left\{ \left[1 + \frac{\lambda}{E_\alpha - H^0} \Sigma^{\mathrm{RND}}(E_\alpha) \mathfrak{D}_\alpha \right] + \frac{\lambda}{E_\alpha - H^0} \Sigma^{\mathrm{RND}}(E_\alpha) \mathfrak{N}_\alpha \right\} |\alpha\rangle. \quad (4.15)$$

We see that the part of the right-hand side of Eq. (4.15) in the square brackets has the same properties as the complete $|\alpha\rangle_{\mathrm{LT}}$ when the complete energies of different states could not be equal. We identify it as the asymptotic stationary state,

$$|\alpha\rangle_{\mathrm{AS}} = N^{\frac{1}{2}}(\alpha) \times \left[1 + \frac{\lambda}{E_\alpha - H^0} \Sigma^{\mathrm{RND}}(E_\alpha) \mathfrak{D}_\alpha \right] |\alpha\rangle. \quad (4.16)$$

The last term on the right-hand side of Eq. (4.15) serves to couple different states with the same complete energies and clearly should be associated with transitions; i.e., it determines the α star. Since $\mathfrak{N}\mathfrak{D} = 0$, we may write

$$|\psi(t)\rangle = \int d\alpha e^{-iE_\alpha t} \left\{ c(\alpha) \times \left[1 + \frac{\lambda}{E_\alpha - H^0} \Sigma^{\mathrm{RND}}(E_\alpha) \mathfrak{N}_\alpha \right] \right\} |\alpha\rangle_{\mathrm{AS}}. \quad (4.17)$$

Equations (4.14) assure us that $\langle\psi(t)|\psi(t)\rangle$ has the same value using either Eq. (4.17) or Eq. (4.15). If we identify the quantity in parenthesis on the right-hand side of Eq. (4.17) as the star coefficient operator $c_{\mathrm{op}}(\alpha)$, we see that Eq. (4.17) now has the form of the asymptotic wave vector assumed in Eq. (2.8). The properties of c_{op} may be obtained by considering the closure of the asymptotic wave packet and the asymptotic wave vector, i.e., $\langle\Phi(t)|\psi(t)\rangle$. However we have yet to establish that the $|\alpha\rangle_{\mathrm{AS}}$ as defined here have the properties of asymptotically stationary states described by Van Hove (I, Sec. 5).

The form of Eq. (4.16) apparently differs from that of Van Hove [I, Eq. (5.12)]. Firstly, we have implicitly allowed for the restrictions over intermediate states, whereas Van Hove writes these down explicitly, Secondly, we have explicitly included the $\mathfrak{D}$ operation in our definition, whereas Van Hove includes it implicitly. These two differences are nonessential and only a matter of notation. An important difference does arise because we have assumed the physical problem of interest was such that it was not appropriate to collect diagonal contributions between intermediate states. However, it is clear that if we carried through Van Hove's procedures with this assumption, we would establish that $|\alpha\rangle_{\mathrm{AS}}$ as given in Eq. (4.16) have all the properties indicated for asymptotically stationary states by Van Hove.

It seems worthwhile to emphasize the formal relation of the present scheme of perturbation theory with the resolvent formalism. The development operator and the resolvent are integral transforms of one another. The present method deals with the development operator in s-operator form, which corresponds to the resolvent written in terms of the transform variable. One is then able, in the present method, to treat diagonal and nondiagonal contributions separately, and is concerned with the mathematical properties of the self-energy in a way which is equivalent to the resolvent formalism. However, one is no longer concerned with the analytical properties of the resolvent itself since its transform is not needed. For certain problems, this may be a distinct advantage. However, the present formulation is somewhat cumbersome in treating both positive and negative times together. Thus the choice of problem and perhaps, individual preference, will determine which method is advantageous.

ACKNOWLEDGMENT

The writer gratefully acknowledges his many helpful discussions with Dr. R. F. Wallis.

Bases for the Irreducible Representations of the Unitary Groups and Some Applications*

MARCOS MOSHINSKY†
Instituto de Física, Universidad de México, México, D. F.
(Received 22 May 1963)

In this paper we show that sets of polynomials in the components of $(2j + 1)$-dimensional vectors, solutions of certain invariant partial differential equations, form bases for all the irreducible representations of the unitary group U_{2j+1}. These polynomials will play, for the group U_{2j+1}, the same role that the solid spherical harmonics (themselves polynomials in the components of a three-dimensional vector) play for the rotation group R_3. With the help of these polynomials we define and determine the reduced Wigner coefficients for the unitary groups, which we then use to derive the Wigner coefficients of U_{2j+1} by a factorization procedure. An ambiguity remains in the explicit expression for the Wigner coefficients as the Kronecker product of two irreducible representations of U_{2j+1} is not, in general, multiplicity-free. We show how to eliminate this ambiguity with the help of operators that serve to characterize completely the rows of representations of unitary groups for a particular chain of subgroups. The procedure developed to determine the polynomial bases of U_{2j+1} seems, in principle, generalizable to arbitrary semisimple compact Lie groups.

1. INTRODUCTION

THE purpose of this paper is to determine the polynomial bases for the irreducible representations of unitary groups of an arbitrary number of dimensions, and to use the bases for a general discussion of Wigner coefficients of these groups. The reader may immediately ask himself whether the first part of the program is necessary, as many of the great names associated with the development of group theory have contributed to the discussion of the bases and the subject was masterfully completed and summarized by Weyl in his book on Classical Groups.[1]

To indicate why we think a discussion of the bases is still important we could compare, for example, the very clear derivation of the bases for irreducible representations of the general linear groups, and hence also of the unitary unimodular groups, given recently by Hamermesh,[2] with the derivation of the bases of the rotation group as given by Wigner.[3]

Both the unitary unimodular and the three-dimensional rotation groups are semisimple compact Lie Groups, and yet their bases are discussed from very different viewpoints. For the unitary unimodular group, one makes use of its relations with the symmetric group to build up irreducible tensors that constitute a basis. For the rotation group, one looks for polynomials, the solid spherical harmonics, that are solutions of certain partial differential equations invariant under rotation, and then proves that these polynomials are bases for irreducible representations.

At first sight it would seem that the procedure followed for the unitary unimodular groups is by far the most general, as it can also be extended to their subgroups such as the orthogonal or symplectic groups in any number of dimensions.

We shall show in this paper that Wigner's procedure for the rotation group is equally general. Specifically, we obtain expressions that can be interpreted as partial differential equations and are invariant under unitary transformations, and show that the polynomial solutions of these equations are bases for the irreducible representations of the unitary groups, i.e., we get what could be called solid unitary harmonics. Furthermore, we show that these procedures for deriving the bases suggest a direct way of deriving the Wigner coefficients for unitary groups, coefficients that have a very wide range of applications.[4]

We shall start our discussion by reviewing Wigner's work on the rotation group in a language that makes his ideas generalizable to the unitary groups.

2. THE THREE-DIMENSIONAL ROTATION GROUP

The group of rotations in a three-dimensional space of coordinates x, y, z has, as operators asso-

* Work supported by the Comisión Nacional de Energía Nuclear.
† Alfred P. Sloan Research Fellow.
[1] H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, New Jersey, 1946), Chap. IX.
[2] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1962), Chap. X.
[3] E. P. Wigner, *Group Theory and its Applications to Quantum Mechanics* (Academic Press Inc., New York, 1959), pp. 153–156.
[4] M. Moshinsky, Rev. Mod. Phys. 34, 813 (1962).

ciated with infinitesimal rotations, the components of the vector

$$\mathbf{L} = (\mathbf{r}\times\mathbf{p}), \qquad \mathbf{p} = i^{-1}\nabla, \tag{2.1}$$

which satisfy the commutation rules

$$[L_0, L_\pm] = \pm L_\pm, \qquad [L_+, L_-] = 2L_0;$$
$$L_\pm \equiv L_x \pm iL_y, \qquad L_0 \equiv L_z. \tag{2.2}$$

We would like to determine the set of polynomials $P(x, y, z)$ that would be bases for irreducible representations of the rotation group. To characterize these polynomials we shall first look into the invariants with respect to rotations, i.e., operators that commute with $\mathbf{L}$, that we can form from $\mathbf{r}$ and $\mathbf{p}$ and in lowest (second) order we have three:

$$I_+ \equiv \tfrac{1}{2}p^2; \qquad I_0 \equiv (1/4i)(\mathbf{r}\cdot\mathbf{p} + \mathbf{p}\cdot\mathbf{r})$$
$$= -\tfrac{1}{2}(\mathbf{r}\cdot\nabla + \tfrac{3}{2}); \qquad I_- \equiv \tfrac{1}{2}r^2. \tag{2.3}$$

These operators form an algebra as all products and linear combinations of the I's commute with $\mathbf{L}$, and, in fact, form a Lie algebra as from the commutation rules of $\mathbf{r}$ and $\mathbf{p}$ we get

$$[I_\pm, I_0] = \mp I_\pm, \qquad [I_+, I_-] = 2I_0. \tag{2.4}$$

The Casimir[5] operator for this commutator Lie Algebra is

$$I^2 \equiv I_-I_+ + I_0(I_0 + 1) = \tfrac{1}{4}r^2p^2$$
$$+ \tfrac{1}{4}(\mathbf{r}\cdot\nabla + \tfrac{3}{2})(\mathbf{r}\cdot\nabla - \tfrac{1}{2}) = \tfrac{1}{4}L^2 - \tfrac{3}{16}, \tag{2.5}$$

where L^2 is the square of the vector (2.1). The relations between L^2 and I^2 suggests that the polynomials $P(x, y, z)$ forming a basis for an irreducible representation of R_3 should be eigenpolynomials of I^2 which, from the definition (2.5), can be achieved when

$$I_0P = \kappa P, \qquad I_+P = 0. \tag{2.6}$$

Equations (2.6) imply that the polynomials are homogeneous of degree l related to κ by[6]

$$\kappa = -\tfrac{1}{2}(l + \tfrac{3}{2}), \tag{2.7}$$

and that they satisfy the Laplace equation.

The set of linearly independent polynomials satisfying (2.6) form a basis for an irreducible representation, as only multiples of the unit matrix commute with all the matrices of the representation.[3] We would like though to prove this result in a more direct way which could be generalized to unitary groups. This can be achieved by using Cartan's theorem[7] stating that the highest-weight polynomial in a basis for an irreducible representation of a semisimple Lie group is unique.

To apply Cartan's theorem to the rotation group, we consider the polynomial solution of (2.6) characterized further by the equation

$$L_0P = mP, \tag{2.8}$$

which can always be applied as L_0 commutes with $I_\pm$, I_0. We refer to m as the weight of the polynomial.[7] Because of the commutation relation (2.2), L_+P which also satisfies (2.6), has weight $m + 1$, and so the polynomial of highest weight will be characterized by

$$L_0P = MP, \qquad L_+P = 0. \tag{2.9}$$

We shall now prove by *reductio ad absurdum*, that the polynomial solutions of (2.6) are a basis for an irreducible representation of R_3. First, as the operators $I_\pm$, I_0 are invariant under rotations, it is clear that the set of all linearly independent polynomial solutions of (2.6) form a basis for a representation of R_3. Let us assume that this representation is reducible. We can then choose linear combinations of these polynomials so that in the new basis the representation is explicitly reduced. Each subset of polynomials that is a basis for an irreducible representation will have a unique term of highest weight. This would imply existence of several polynomial solutions satisfying simultaneously (2.6) and (2.9) for a fixed κ. But from (2.5) we get

$$\kappa(\kappa + 1) = \tfrac{1}{4}M(M + 1) - \tfrac{3}{16},$$
$$\text{or} \quad l(l + 1) = M(M + 1), \tag{2.10}$$

and as both l, M are nonnegative, the only possible solution corresponds to $l = M$, for which we have the single polynomial

$$P_l^l(x, y, z) = (-1)^l[(2l + 1)!/4\pi]^{\frac{1}{2}}(2^l l!)^{-1}(x + iy)^l$$
$$= r^l Y_{ll}(\theta, \varphi), \tag{2.11}$$

thus contradicting our hypothesis.

To get the other polynomials of the basis for an irreducible representation, we can start from the highest-weight polynomial and decrease its weight with the operator L_-, and so the normalized poly-

[5] G. Racah, "Group Theory and Spectroscopy," Lecture notes, Institute of Advanced Study, Princeton, New Jersey (1951), p. 44.

[6] The reader may be puzzled by the fact that κ is not a nonnegative integer or semi-integer, but from (2.3) one can see that I_0 and $I_2 \equiv (2i)^{-1}(I_+ - I_-)$ are non-Hermitian operators, and so the usual discussion [E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, Cambridge, England, 1935), pp. 46, 47] does not apply to the eigenvalues of I^2, I_0.

[7] Reference 5, p. 37.

nomials of the basis can be written as

$$P_m^l(x, y, z) = [(l + m)!]^{\frac{1}{2}}[(l - m)!\,(2l)!]^{-\frac{1}{2}} \times L_-^{l-m}P_l^l(x, y, z) = r^l Y_{lm}(\theta, \varphi). \quad (2.12)$$

We shall now use the basis (2.12) to derive the Wigner coefficients of the R_3 group.[8] We consider polynomials $P(\mathbf{r}_1, \mathbf{r}_2)$ in two vectors that are basis for irreducible representations of independent rotations in $\mathbf{r}_1$ and $\mathbf{r}_2$, i.e., polynomials that satisfy the equations

$$I_0^{(1)}P = -\tfrac{1}{2}(l_1 + \tfrac{3}{2})P, \qquad I_+^{(1)}P = 0, \quad (2.13a)$$

$$I_0^{(2)}P = -\tfrac{1}{2}(l_2 + \tfrac{3}{2})P, \qquad I_+^{(2)}P = 0, \quad (2.13b)$$

where the upper index in the I's refers to the index of the vector. We would like to choose subsets of P's that would transform irreducibly under simultaneous rotation of $\mathbf{r}_1$ and $\mathbf{r}_2$. Each of these subsets will have a polynomial of highest weight determined by the equation

$$L_0P = lP, \qquad L_+P = 0, \quad (2.14a, b)$$

where

$$L_q = L_q^{(1)} + L_q^{(2)}, \qquad q = \pm, 0. \quad (2.15)$$

Using (2.12) we see that the most general solution of (2.13) and (2.14a) is

$$P(\mathbf{r}_1, \mathbf{r}_2) = \sum_{m_1} A_{m_1} r_1^{l_1} Y_{l_1 m_1}(\theta_1, \varphi_1) \times r_2^{l_2} Y_{l_2 l-m_1}(\theta_2, \varphi_2), \quad (2.16)$$

where A_{m_1} is an arbitrary constant. Applying now (2.14b), we get for A_{m_1} the recurrence relation

$$\frac{A_{m_1+1}}{A_{m_1}} = -\left[\frac{(l_1 - m_1)(l_1 + m_1 + 1)}{(l_2 - l + m_1 + 1)(l_2 + l - m_1)}\right]^{\frac{1}{2}}. \quad (2.17)$$

If we disregard for the moment $r_1^{l_1}$, $r_2^{l_2}$ and normalize the polynomial (2.16) with respect to the angular variables, the A_{m_1} will be completely determined and, in fact, correspond to the Wigner coefficient[8]

$$A_{m_1} = \langle l_1 l_2 m_1 l - m_1 \mid ll\rangle. \quad (2.18)$$

To get now the polynomials of arbitrary weight in the irreducible representation l, we apply to (2.16) the descending operator in (2.12) with L_- defined by (2.15). The coefficients of the spherical harmonics in the polynomial of weight m will then be the general Wigner coefficients and, in fact, appear in one of the explicit forms proposed by Racah.[9]

We have obtained the bases for the irreducible representations of the R_3 group, and from it we determined the Wigner coefficients of R_3. The analysis developed in this section will be generalized step by step to unitary groups of arbitrary dimension.

3. BASES FOR THE IRREDUCIBLE REPRESENTATIONS OF UNITARY GROUPS

We will consider the unitary group in $2j + 1$ dimensions denoted by U_{2j+1} where j is either integer or semi-integer.[10] This group will consist of all unitary matrices in a $(2j + 1)$-dimensional space. Following a procedure originally introduced by Schwinger[11] for the SU_2 group, it is very convenient to denote the vectors in this vector space as the creation operators $a_{\mu s}^+$, where $\mu = 1, 2, \cdots, 2j + 1$ is the index of the components of the vectors and $s = 1, 2, \cdots, n$, is an index that distinguishes between the vectors themselves. Under unitary transformations the vectors $a_{\mu s}^+$ transform into

$$a_{\mu' s}^{\prime +} = \sum_{\mu=1}^{2j+1} U_{\mu'}^{\mu} a_{\mu s}^+. \quad (3.1)$$

The main purpose of this section will be to find the sets of linearly independent polynomials

$$P(a_{\mu s}^+) = P(a_{11}^+, a_{21}^+, \cdots, a_{2j+1n}^+), \quad (3.2)$$

forming bases for the irreducible representations of U_{2j+1}.

The initial step in this program should be to find the operators which play for U_{2j+1}, the role that $L_q (q = \pm, 0)$ of (2.2) play for R_3. For this purpose we first introduce the annihilation operator $a^{\mu}{}_s$ defined by the commutation relations

$$[a^{\mu}{}_s, a_{\mu' s'}^+] = \delta_{\mu'}^{\mu}\delta_{ss'}. \quad (3.3)$$

From (3.3) we see that when applied to the polynomial expression (3.2), the annihilation operators could be interpreted as differential operators, i.e.,

$$a^{\mu}{}_s = \partial/\partial a_{\mu s}^+. \quad (3.4)$$

Furthermore, from (3.3) or (3.4) we have that when the $a_{\mu s}^+$ transform as in (3.1), the $a^{\mu}{}_s$ transform into

[8] Reference 3, p. 192.

[9] G. Racah, Phys. Rev. **62**, 438 (1942).

[10] The notation U_{2j+1} is used instead of U_n, as in applications unitary groups frequently appear in relation to the 2j + 1 functions in a shell of angular momentum j.

[11] J. Schwinger, "On Angular Momentum," U. S. Atomic Energy Commission Rept. NYO-3071, 1952 (Unpublished). V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960).

$$a'^{\mu'}{}_s = \sum_\mu (U^+)^{\mu'}_\mu a^\mu{}_s, \tag{3.5}$$

where U^+ is the transposed conjugate of the matrix U.

We also introduce the concept of vacuum state $|0\rangle$ by the definition

$$a^\mu{}_s\,|0\rangle = 0 \quad \text{for all} \quad \mu, s. \tag{3.6}$$

With the help of $|0\rangle$ we can define the scalar product of two polynomials of the type (3.2) as

$$(P, P') \equiv \langle 0|\, P^+ P'\, |0\rangle, \tag{3.7}$$

where P^+ is obtained by replacing all $a^+_{\mu s}$ in P by $a^\mu{}_s$, and we use the commutation relation (3.3) and the definition (3.6) to evaluate (3.7).

We now define the operators associated with the infinitesimal unitary transformations[12] in a space of $2j+1$ dimensions as

$$\mathfrak{C}_\mu{}^{\mu'} = \sum_{s=1}^{n} a^+_{\mu s} a^{\mu'}{}_s, \tag{3.8}$$

and from (3.3) they satisfy the commutation rules

$$[\mathfrak{C}_\mu{}^{\mu'}, \mathfrak{C}_{\mu''}{}^{\mu'''}] = \mathfrak{C}_\mu{}^{\mu'''}\delta^{\mu'}_{\mu''} - \mathfrak{C}_{\mu''}{}^{\mu'}\delta^{\mu'''}_{\mu}, \tag{3.9}$$

which are associated with the generators[13] of U_{2j+1}.

The operators $\mathfrak{C}_\mu{}^{\mu'}$ play for the U_{2j+1} group the role that $L_q (q = \pm, 0)$ play for R_3, while $a^+_{\mu s}$, $a^\mu{}_s$ have for U_{2j+1} a meaning similar to the one $\mathbf{r}$ and ∇ have for R_3. In analogy with the previous section, we could now ask which are the invariants with respect to U_{2j+1}, i.e., operators that commute with the $\mathfrak{C}_\mu{}^{\mu'}$, we can form from $a^+_{\mu s}$ and $a^\mu{}_s$. From (3.1), (3.5), the lowest- (second-) order ones are

$$C_{ss'} = \sum_{\mu=1}^{2j+1} a^+_{\mu s} a^\mu{}_{s'}, \tag{3.10}$$

which certainly satisfy

$$[\mathfrak{C}_\mu{}^{\mu'}, C_{ss'}] = 0, \tag{3.11}$$

as can also be seen by using (3.3). The invariant operators that satisfy (3.11) form an algebra, and in fact, a Lie algebra, as from (3.3) we obtain

$$[C_{ss'}, C_{s''s'''}] = C_{ss'''}\delta_{s''s'} - C_{s''s'}\delta_{ss'''}. \tag{3.12}$$

When applying functions of the operators (3.8), (3.10) to polynomials of the type (3.2), we shall always think of $\mathfrak{C}_\mu{}^{\mu'}$, $C_{ss'}$ as first-order partial differential operators in the variables $a^+_{\mu s}$ according to the rule (3.4).

We shall now show that the three subsets of the set of operators (3.10),

$$\begin{aligned} &C_{ss'} \quad \text{with} \quad s < s' = 2, \cdots n, \\ &C_{ss} \quad \text{with} \quad s = 1, \cdots n, \\ &C_{ss'} \quad \text{with} \quad s > s' = 1, \cdots n-1, \end{aligned} \tag{3.13}$$

play for U_{2j+1} the same role as I_+, I_0, I_- play for R_3. This implies proving the following:

Theorem: The linearly independent polynomials of the type (3.2) *that satisfy the equations*

$$C_{ss}P = h_s P, \qquad C_{ss'}P = 0, \qquad s < s', \tag{3.14}$$

form a basis for an irreducible representation of U_{2j+1} *characterized by the set of nonnegative intergers* $[h_1 h_2 \cdots]$ *giving the degree of the polynomials in the components of each one of the vectors* $a^+_{\mu s}$.

We shall use Cartan's theorem for the proof, and to apply it we first consider the polynomial solution of (3.14) characterized further by the equations

$$\mathfrak{C}_\mu{}^\mu P = k_\mu P, \qquad \mu = 1, 2, \cdots, 2j+1, \tag{3.15}$$

which can always be applied as from (3.9) and (3.11), the $\mathfrak{C}_\mu{}^\mu$ commute among themselves and with the $C_{ss'}$. We refer to the set of numbers $[k_1 \cdots k_{2j+1}]$ as the weight of the polynomial P. If we have another polynomial P' of weight $[k'_1 \cdots k'_{2j+1}]$ then we say that P is of higher weight[14] than P' if in $[(k_1 - k'_1), \cdots, (k_{2j+1} - k'_{2j+1})]$ the first nonvanishing component is positive.[7]

If P satisfies (3.14), so does $\mathfrak{C}_\mu{}^{\mu'}P$, but from the commutation rules (3.9), $\mathfrak{C}_\mu{}^{\mu'}P$ with $\mu < \mu'$ has weight $[k_1, \cdots, k_\mu + 1, \cdots, k_{\mu'} - 1, \cdots, k_{2j+1}]$ and so it is of higher weight than P. Clearly then, the polynomial of highest weight is characterized by

$$\begin{aligned} &\mathfrak{C}_\mu{}^\mu P = K_\mu P, \qquad \mathfrak{C}_\mu{}^{\mu'}P = 0, \\ &\qquad \mu < \mu';\ \mu, \mu' = 1, \cdots 2j+1. \end{aligned} \tag{3.16a, b}$$

From the discussion of the previous section we conclude that our theorem will be proved if the polynomial satisfying both (3.14) and (3.16) for a given $[h_1 h_2 \cdots]$ is unique.[15]

In a previous paper[16] we discussed the most

[12] M. Moshinsky, "Group Theory and Collective Motions" Lecture notes, Latin American School of Physics, México, 1962, pp. 19, 20.

[13] Reference 5, p. 29.

[14] Reference 5, p. 33.

[15] This implies that Cartan's theorem is valid for U_{2j+1} which is not a semisimple group as $\mathrm{tr}\,\|\mathfrak{C}_\mu{}^{\mu'}\| \equiv H$ commutes with all the operators $\mathfrak{C}_\mu{}^{\mu'}$. The generators of the corresponding semisimple group SU_{2j+1} are constructed by subtracting $(2j+1)^{-1} H\delta^{\mu\mu'}$ from $\mathfrak{C}_\mu{}^{\mu'}$, and so give the same commutation relations (3.9). The polynomial of maximum weight for SU_{2j+1} is then also given by (3.16). Therefore, Racah's proof of Cartan's theorem (reference 5, p. 37) will apply to U_{2j+1} and so we shall deal systematically with this group rather than with SU_{2j+1}.

[16] M. Moshinsky, Nucl. Phys. **31**, 384 (1962).

general polynomial solutions of (3.14) and we obtained the following results:

(1) The polynomials satisfying (3.14) can, at most, be functions of the first $2j+1$ vectors $a^+_{\mu s}$, i.e.,

$$h_s = 0 \quad \text{if} \quad s > 2j+1. \tag{3.17}$$

(2) The degrees h_s of the polynomials in each of the $2j+1$ vectors $a^+_{\mu s}$, satisfy the inequality

$$h_1 \geq h_2 \geq h_3 \geq \cdots \geq h_{2j+1} \geq 0. \tag{3.18}$$

(3) The solution of (3.14) can be given in terms of the determinants

$$\Delta^{s_1 s_2 \cdots s_r}_{\mu_1 \mu_2 \cdots \mu_r} = \sum_{\mathfrak{p}} [(-1)^{\mathfrak{p}} \mathfrak{p} a^+_{\mu_1 s_1} a^+_{\mu_2 s_2} \cdots a^+_{\mu_r s_r}], \tag{3.19}$$

(where $\mathfrak{p}$ stands for a permutation of $s_1, \cdots s_r$) as

$$P = (\Delta^1_1)^{h_1-h_2}(\Delta^{12}_{12})^{h_2-h_3}\cdots(\Delta^{1\cdots 2j+1}_{1\cdots 2j+1})^{h_{2j+1}} \times Z\left(\frac{\Delta^1_\mu}{\Delta^1_1}, \frac{\Delta^{12}_{1\mu}}{\Delta^{12}_{12}}, \cdots, \frac{\Delta^{12\cdots 2j}_{12\cdots \mu}}{\Delta^{12\cdots 2j}_{12\cdots 2j}}\right), \tag{3.20}$$

where Z is an arbitrary polynomial in the ratios indicated, subject only to the condition that P should be a polynomial in $a^+_{\mu s}$.

We now require that the polynomials (3.20) should also satisfy (3.16). The effect of $\mathcal{C}_\mu{}^{\mu'}$ with $\mu < \mu'$ on a typical determinant in (3.20) is

$$\mathcal{C}_\mu{}^{\mu'} \Delta^{12\cdots r-1 r}_{12\cdots r-1 \mu'} = \delta^{\mu'}_{\mu'} \Delta^{12\cdots r-1 r}_{12\cdots r-1 \mu};$$

$$\mu < \mu', \qquad r = 1, \cdots 2j+1. \tag{3.21}$$

From (3.21) we see that

$$\mathcal{C}_1{}^{\mu} P = \frac{P}{Z}\frac{\partial Z}{\partial(\Delta^1_\mu/\Delta^1_1)} = 0, \quad \mu = 2, \cdots 2j+1, \tag{3.22a}$$

and so Z is independent of the ratio $(\Delta^1_\mu/\Delta^1_1)$. Applying now $\mathcal{C}^{\mu}{}_2$ to this restricted polynomial, we get

$$\mathcal{C}_2{}^{\mu} P = \frac{P}{Z}\frac{\partial Z}{\partial(\Delta^{12}_{1\mu}/\Delta^{12}_{12})} = 0, \quad \mu = 3, \cdots 2j+1, \tag{3.22b}$$

so that Z is also independent of $(\Delta^{12}_{1\mu}/\Delta^{12}_{12})$. Continuing with $\mathcal{C}_3{}^{\mu}$ etc., we obtain finally that Z is independent of all the ratios, i.e., Z is a constant. The polynomial satisfying (3.14) and (3.16b) is then unique up to a constant, and denoting it by $\mathcal{P}$, we have

$$\mathcal{P} = (\Delta^1_1)^{h_1-h_2}(\Delta^{12}_{12})^{h_2-h_3}\cdots(\Delta^{12\cdots 2j+1}_{12\cdots 2j+1})^{h_{2j+1}}. \tag{3.23}$$

Furthermore, by applying (3.16a) to (3.23) we get

$$K_1 = h_1, \quad K_2 = h_2, \cdots, K_{2j+1} = h_{2j+1}. \tag{3.24}$$

We have thus proved that the set of linearly independent polynomials (3.20) is a basis for an irreducible representation of U_{2j+1}, as the highest-weight polynomial in this set is unique, and, in fact, we get the explicit form (3.23) for it.

In the next section we shall indicate how to generate all the polynomials for an irreducible basis from the highest-weight polynomial, using the lowering operators $\mathcal{C}_\mu{}^{\mu'}$ with $\mu > \mu'$.

4. CONSTRUCTION OF THE BASIS FOR AN IRREDUCIBLE REPRESENTATION OF U_{2j+1} FROM THE HIGHEST WEIGHT POLYNOMIAL

In Sec. 2 we obtained the highest-weight polynomial (2.11) for R_3 and then derived the full basis (2.12) for an irreducible representation of R_3 by applying the lowering operator L_- to the highest-weight polynomial.

We would like to get a similar result for U_{2j+1}, i.e., to construct the basis for an irreducible representation of U_{2j+1} in terms of polynomials

$$P = \mathcal{R}(\mathcal{C}_\mu{}^{\mu'})\mathcal{P}, \tag{4.1}$$

where the $\mathcal{R}$'s are polynomial functions of the $\mathcal{C}_\mu{}^{\mu'}$ and $\mathcal{P}$ is the highest-weight polynomial (3.23). From Eqs. (3.16) and the commutation rules (3.9) we clearly see that we can restrict $\mathcal{R}$ to be a function only of the $\mathcal{C}_\mu{}^{\mu'}$ with $\mu > \mu'$. Furthermore, from (3.11), all polynomials (4.1) satisfy Eqs. (3.14) that define a basis for an irreducible representation.

We shall give the explicit derivation of $\mathcal{R}(\mathcal{C}_\mu{}^{\mu'})$ for the U_3 group, and then sketch the generalization of the analysis to any U_{2j+1}. The rows of a basis for an irreducible representation of a given group are usually characterized by a chain of subgroups, e.g., the rows of the basis for an irreducible representation l of R_3 are characterized by the index $m = l, \cdots, -l$ which corresponds to irreducible representations of the subgroup R_2 of rotations around the z axis. For the unitary groups, the natural chain of subgroups[4] is formed when we diminish the dimension of the group by one in each step, i.e., for U_3 the chain of subgroups will be

$$\begin{bmatrix} U^1_1 & U^2_1 & U^3_1 \\ U^1_2 & U^2_2 & U^3_2 \\ U^1_3 & U^2_3 & U^3_3 \end{bmatrix} \supset \begin{bmatrix} U^1_1 & U^2_1 & 0 \\ U^1_2 & U^2_2 & 0 \\ 0 & 0 & U^3_3 \end{bmatrix} \supset \begin{bmatrix} U^1_1 & 0 & 0 \\ 0 & U^2_2 & 0 \\ 0 & 0 & U^3_3 \end{bmatrix}. \tag{4.2}$$

If for physical reasons we are interested in any other chain of subgroups, for example $U_3 \supset R_3$, we could pass from the basis whose rows are characterized by (4.2), to the basis whose rows correspond to the $U_3 \supset R_3$ chain with the help of appropriate coefficients.[4]

While the generators of the U_3 group are $\mathcal{C}_\mu{}^{\mu'}$ with $\mu, \mu' = 1, 2, 3$, those of the subgroup U_2 in the chain (4.2) will be the subset $\mathcal{C}^1_1$, $\mathcal{C}^2_1$, $\mathcal{C}^1_2$, $\mathcal{C}^2_2$. We could now characterize the rows of the polynomials (4.1) by requiring that they be of definite

weight in the subgroup U_2, which for U_3 is a condition equivalent to (2.8) for R_3. Because of the homomorphism between SU_2 and R_3, we know that by applying powers[4] of $\mathcal{C}_2^1$ to the highest-weight polynomial in U_2 we can get polynomials of any weight in this subgroup. We therefore will restrict ourselves to characterizing the rows by the highest weight in U_2, i.e., we require that the polynomials of (4.1) satisfy

$$\mathcal{C}_1^1 P = [\mathcal{C}_1^1, \mathfrak{R}]\mathcal{P} + h_1\mathfrak{R}\mathcal{P} = q_1 P, \tag{4.3a}$$

$$\mathcal{C}_2^2 P = [\mathcal{C}_2^2, \mathfrak{R}]\mathcal{P} + h_2\mathfrak{R}\mathcal{P} = q_2 P, \tag{4.3b}$$

$$\mathcal{C}_1^2 P = [\mathcal{C}_1^2, \mathfrak{R}]\mathcal{P} = 0, \tag{4.3c}$$

where q_1, q_2, being degrees of the polynomial in the components 1, 2 of the vectors, are nonnegative integers.

For U_3 the most general polynomial function $\mathfrak{R}(\mathcal{C}_\mu^{\mu'})$ is

$$\mathfrak{R}(\mathcal{C}_\mu^{\mu'}) = \sum_{\alpha\beta\gamma} A_{\alpha\beta\gamma}(\mathcal{C}_2^1)^\alpha(\mathcal{C}_3^2)^\beta(\mathcal{C}_3^1)^\gamma, \tag{4.4}$$

and from Eqs. (4.3a,b) and the commutation relations (3.9) we immediately get

$$\gamma = h_1 - q_1 - \alpha, \qquad \beta = h_2 - q_2 + \alpha. \tag{4.5}$$

The coefficient of the $\mathcal{C}_\mu^{\mu'}$ in (4.4) can thus be denoted as A_α, and applying (4.3c), we get the recurrence relation

$$\frac{A_{\alpha+1}}{A_\alpha} = \frac{(h_1 - q_1 - \alpha)}{(\alpha + 1)(q_1 - q_2 + \alpha + 2)}. \tag{4.6}$$

The lowering function $\mathfrak{R}(\mathcal{C}_\mu^{\mu'})$ has then the form

$$\mathfrak{R}(\mathcal{C}_\mu^{\mu'}) = \sum_\alpha \left[\frac{(\mathcal{C}_2^1)^\alpha(\mathcal{C}_3^2)^{h_2-q_2+\alpha}(\mathcal{C}_3^1)^{h_1-q_1-\alpha}}{\alpha!\,(q_1-q_2+\alpha+1)!\,(h_1-q_1-\alpha)!}\right], \tag{4.7}$$

and applying it to the polynomial of maximum weight we get

$$\begin{aligned} P = \mathfrak{R}(\mathcal{C}_\mu^{\mu'})\mathcal{P} = {} & \{(h_2 - h_3)!\,(h_1 - h_2)! \\ & \times (h_1 - h_3 + 1)!\,[(-h_3 + q_2)!\,(h_1 - q_2 + 1)! \\ & \times (-h_2 + q_1)!\,(h_1 - q_1)!\,(-h_3 + q_1 + 1)!]^{-1} \\ & \times (\Delta_1^1)^{-h_2+q_1}(\Delta_3^1)^{h_1-q_1}(\Delta_{12}^{12})^{-h_2+q_2} \\ & \times (\Delta_{13}^{12})^{h_2-q_2}(\Delta_{123}^{123})^{h_3}\}, \end{aligned} \tag{4.8}$$

which is seen immediately to satisfy both (3.14) and (4.3). Furthermore, as P is a polynomial, all exponents in (4.8) must be nonnegative and so we get the inequalities

$$h_1 \geq q_1 \geq h_2 \geq q_2 \geq h_3 \geq 0, \tag{4.9}$$

that give well-known restrictions[4] on the highest weights of U_2 contained in U_3.

The polynomial (4.8) is then part of a basis of an irreducible representation $[h_1h_2h_3]$ of U_3 and at the same time, the highest-weight polynomial in a basis for the irreducible representation of U_2 characterized by $[q_1q_2]$. To get all the weights in U_2 we notice that from (3.9),

$$\mathcal{C}_1^2, \quad \tfrac{1}{2}(\mathcal{C}_1^1 - \mathcal{C}_2^2), \quad \mathcal{C}_2^1 \tag{4.10}$$

satisfy the same commutation relations (2.2) as do L_+, L_0, L_-. From (2.12) we see then that we would get all weights in U_2 if to the polynomials (4.8) we apply the operator

$$[(t + \tau)!]^{\frac{1}{2}}[(t - \tau)!\,(2t)!]^{-\frac{1}{2}}(\mathcal{C}_2^1)^{t-\tau}, \tag{4.11a}$$

where

$$t = \tfrac{1}{2}(q_1 - q_2), \quad \tau = t, t-1, \cdots, -t. \tag{4.11b}$$

From (4.11) we notice that there are $2t + 1 = q_1 - q_2 + 1$ polynomials in the basis for the irreducible representation $[h_1h_2h_3]$ of U_3 whose row indices are $[q_1q_2]$. As the values of $[q_1q_2]$ are limited by the inequality (4.9), we conclude that the dimension of the representation is

$$\begin{aligned} \sum_{q_1=h_2}^{h_1}\sum_{q_2=h_3}^{h_2}(q_1 - q_2 + 1) = {} & \tfrac{1}{2}(h_1 - h_2 + 1) \\ & \times (h_1 - h_3 + 2)(h_2 - h_3 + 1), \end{aligned} \tag{4.11c}$$

which is the well-known Weyl formula for[17] U_3.

Combining all of the previous results and normalizing the polynomials according to the definition (3.7) of the scalar product, we obtain for the bases of the irreducible representations of U_3, the expressions

$$\begin{aligned} P_{q_1q_2\tau}^{h_1h_2h_3} = {} & \mathfrak{R}_{q_1q_2\tau}^{h_1h_2h_3}(\mathcal{C}_\mu^{\mu'})\{[(h_1 - h_2 + 1) \\ & \times (h_1 - h_3 + 2)(h_2 - h_3 + 1)]^{\frac{1}{2}} \\ & \times [(h_1 + 2)!\,(h_2 + 1)!\,h_3!]^{-\frac{1}{2}} \\ & \times (\Delta_1^1)^{h_1-h_2}(\Delta_{12}^{12})^{h_2-h_3}(\Delta_{123}^{123})^{h_3}\}, \end{aligned} \tag{4.12a}$$

where $\mathfrak{R}_{q_1q_2\tau}^{h_1h_2h_3}$ is the operator

$$\begin{aligned} \mathfrak{R}_{q_1q_2\tau}^{h_1h_2h_3}(\mathcal{C}_\mu^{\mu'}) = {} & \{[(q_1 - q_2 + 1)(-h_3 + q_2)! \\ & \times (h_1 - q_2 + 1)!\,(-h_2 + q_1)!\,(h_1 - q_1)! \\ & \times (-h_3 + q_1 + 1)!\,(\tfrac{1}{2}q_1 - \tfrac{1}{2}q_2 + \tau)!]^{\frac{1}{2}} \\ & \times [(h_1 - h_2)!\,(h_1 - h_3 + 1)!\,(h_2 - h_3)! \\ & \times (h_2 - q_2)!\,(\tfrac{1}{2}q_1 - \tfrac{1}{2}q_2 - \tau)!]^{-\frac{1}{2}} \\ & \times \sum_\alpha \left[\frac{(\mathcal{C}_2^1)^{\frac{1}{2}(q_1-q_2)-\tau+\alpha}(\mathcal{C}_3^2)^{h_2-q_2+\alpha}(\mathcal{C}_3^1)^{h_1-q_1-\alpha}}{\alpha!\,(q_1 - q_2+\alpha+1)!\,(h_1 - q_1 - \alpha)!}\right]\Bigg\}. \end{aligned} \tag{4.12b}$$

[17] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover Publications, Inc., New York, 1955), p. 383.

The polynomials (4.12) are for U_3, in the $U_3 \supset U_2$ chain, equivalent to the polynomials (2.12) for R_3. As the indices q_1, q_2, τ are associated with Hermitian operators,[4] polynomials (4.12) that differ in any of them will be orthogonal, so that they form a basis for a *unitary* irreducible representation characterized by the partition $[h_1h_2h_3]$.

The construction of the basis for an irreducible representation of U_3 from its highest-weight polynomial which was given above, can be immediately generalized to U_{2j+1}. For example, for U_4 we would first get a function $\mathfrak{R}(\mathcal{C}_\mu^{\mu'})$ that gives us from the polynomial of highest weight $\mathcal{P}$, the polynomials of highest weight in the subgroup U_3 of U_4. This implies that $\mathfrak{R}\mathcal{P}$ satisfies equations similar to (4.3), but where the indices take now the values 1, 2, 3 instead of only 1, 2. Once we have these polynomials, we can lower the weight in U_3, but still have them of highest weight in the subgroup U_2 of U_3, if we apply (4.7). Finally we would get the full set of polynomials for the irreducible basis of U_4 in the $U_4 \supset U_3 \supset U_2$ chain if we apply (4.11).

Another procedure of getting the basis of U_{2j+1} would be to apply the equations

$$\mathcal{C}_\mu^{\mu}P = q_\mu P, \qquad \mathcal{C}_\mu^{\mu'}P = 0, \qquad \mu < \mu'; \qquad \mu, \mu' = 1, 2, \cdots 2j, \tag{4.13}$$

giving the highest weight in the subgroup U_{2j}, directly to the solution (3.20), from which we obtain immediately

$$P = (\Delta_1^1)^{-h_2+q_1}(\Delta_{2j+1}^1)^{h_1-q_1}(\Delta_{12}^{12})^{-h_3+q_2} (\Delta_{12j+1}^{12})^{h_2-q_2} \cdots (\Delta_{12\cdots 2j+1}^{12\cdots 2j+1})^{h_{2j+1}}. \tag{4.14}$$

Equation (4.14) is the generalization to U_{2j+1} of (4.8), and from it we obtain the inequalities

$$h_1 \geq q_1 \geq h_2 \geq q_2 \cdots \geq q_{2j} \geq h_{2j+1} \geq 0, \tag{4.15}$$

which, by an analysis similar to (4.11c), lead to the general Weyl formula for the dimensionality of irreducible representations of unitary groups.[17]

5. WIGNER COEFFICIENTS FOR THE UNITARY GROUPS

We have obtained the highest-weight polynomial $\mathcal{P}$ of a basis for an irreducible representation characterized by $[h_1h_2 \cdots h_{2j+1}]$ of U_{2j+1}, and we showed in the previous section how to obtain the lowering operators with whose help we can derive from $\mathcal{P}$ the full basis. Following the program outlined for R_3 in Sec. 2, we shall use these results to derive the Wigner coefficients of the unitary groups.

Before proceeding on this program, we would like to connect the basis for the irreducible representations of U_{2j+1} discussed in the previous sections with the basis for the unitary unimodular group SU_{2j+1}. An arbitrary unitary matrix can always be written as the product of a unitary matrix whose determinant is 1 (i.e., unimodular) by a matrix

$$\exp(i\delta)I, \tag{5.1}$$

where I is the $(2j + 1)$-dimensional unit matrix, and δ is an arbitrary real constant. If we apply the transformation (5.1) to the basis (3.20), we see that all terms of the basis are then multiplied by

$$\exp(ih\delta), \qquad h = h_1 + h_2 + \cdots + h_{2j+1}. \tag{5.2}$$

Clearly then, a basis for an irreducible representation of U_{2j+1} will also be a basis for an irreducible representation of SU_{2j+1}, but in view of the fact that the determinant

$$\Delta_{12\cdots 2j+1}^{12\cdots 2j+1} \tag{5.3}$$

in (3.20) is an invariant under SU_{2j+1}, we have that all bases of U_{2j+1}, with the same values for

$$[h_1 - h_{2j+1}, h_2 - h_{2j+1}, \cdots, h_{2j} - h_{2j+1}], \tag{5.4}$$

are equivalent under SU_{2j+1}. As the full structure of the representation is already in SU_{2j+1}, we shall restrict ourselves in the following analysis to representations of U_{2j+1} in which $h_{2j+1} = 0$.

We shall now explicitly discuss the procedure for deriving the Wigner coefficients for the U_3 group and later sketch the generalization of the analysis to any U_{2j+1}. According to the remarks of the previous paragraph, for U_3 we can restrict ourselves to partitions in which the third integer is zero and so the basis for two irreducible representations could be characterized by $[h_1'h_2'0]$ and $[h_3'h_4'0]$, respectively. For short, we shall denote the representations by $[h_1'h_2']$ and $[h_3'h_4']$, and the polynomials for their bases are given by (4.12) in which we denote the corresponding vectors by $a_{\mu 1}^+$, $a_{\mu 2}^+$ and $a_{\mu 3}^+$, $a_{\mu 4}^+$, respectively.

From the theorem of Sec. 3, the most general polynomials in the four vectors $a_{\mu s}^+$, $s = 1, 2, 3, 4$ that transform irreducibly, according to the partitions $[h_1'h_2']$ and $[h_3'h_4']$, in the first two and the last two vectors, satisfy the equations

$$C_{11}P = h_1'P, \quad C_{22}P = h_2'P, \quad C_{12}P = 0, \text{ (5.5a, b, c)}$$

$$C_{33}P = h_3'P, \quad C_{44}P = h_4'P, \quad C_{34}P = 0. \text{ (5.5d, e, f)}$$

Equations (5.5) are, for U_3, equivalent to Eqs. (2.13) for R_3.

We should like now to choose subsets of the P's that transform irreducibly under simultaneous unitary transformations on all four vectors $a^+_{\mu s}$. Each of these subsets will have polynomials of highest weight determined by the equations

$$\mathcal{C}_\mu{}^\mu P = h_\mu P, \quad \mathcal{C}_\mu{}^{\mu'} P = 0, \qquad \mu < \mu'; \quad \mu, \mu' = 1, 2, 3, \qquad (5.6a, b)$$

where

$$\mathcal{C}_\mu{}^{\mu'} = \sum_{s=1}^{4} a^+_{\mu s} a^{\mu'}{}_s. \qquad (5.7)$$

Equations (5.6) are then, for U_3, equivalent to Eqs. (2.14) for R_3. The operators $\mathcal{C}_\mu{}^{\mu'}$ associated with the infinitesimal transformations of U_3 involve now all four vectors, i.e., the index s in (5.7) goes from 1 to 4.

Following the outline of Sec. 2, our next objective would be to find the polynomial solutions of Eqs. (5.5), (5.6). We shall do this in full detail in the next section where we shall also discuss the problem of the multiplicity of solutions and the way of distinguishing between them. For the moment we assume that we have the polynomials $P(a^+_{\mu s})$ satisfying Eqs. (5.5) and (5.6), which we furthermore normalize according to the definition (3.7) of the scalar product. Applying to these highest-weight polynomials, the lowering operators $\mathfrak{R}$ of (4.12), we can construct the polynomial of arbitrary weight:

$$P^{h_1 h_2 h_3}_{q_1 q_2 \tau}(a^+_{\mu_1}, a^+_{\mu_2}, a^+_{\mu_3}, a^+_{\mu_4}) \equiv \mathfrak{R}^{h_1 h_2 h_3}_{q_1 q_2 \tau}(\mathcal{C}_\mu{}^{\mu'}) P(a^+_{\mu s}). \qquad (5.8)$$

The Wigner coefficients of U_3 are then given by the scalar product of (5.8) into the polynomial

$$P^{h_1' h_2' 0}_{q_1' q_2' \tau'}(a^+_{\mu 1}, a^+_{\mu 2}) P^{h_3' h_4' 0}_{q_1'' q_2'' \tau''}(a^+_{\mu 3}, a^+_{\mu 4}), \qquad (5.9)$$

where each of the factors in (5.9) is given by (4.12). These coefficients will also be Wigner coefficients for SU_3 and, in fact, the general Wigner coefficients, as the partitions in SU_3 have at most two rows.

It is convenient to define the concept of reduced Wigner coefficients which will provide us with a factorization procedure for the Wigner coefficients of unitary groups. We first introduce the polynomial

$$\Pi^{h_1' h_2', h_3' h_4'}_{q_1' q_2', q_1'' q_2'', q_1 q_2 \tau}(a^+_{\mu 1}, a^+_{\mu 2}, a^+_{\mu 3}, a^+_{\mu 4}) \equiv \sum_{\tau' \tau''} \{\langle \tfrac{1}{2}(q_1' - q_2') \tfrac{1}{2}(q_1'' - q_2'') \tau' \tau'' \mid \tfrac{1}{2}(q_1 - q_2) \tau \rangle \times P^{h_1' h_2' 0}_{q_1' q_2' \tau'}(a^+_{\mu 1}, a^+_{\mu 2}) P^{h_3' h_4' 0}_{q_1'' q_2'' \tau''}(a^+_{\mu 3}, a^+_{\mu 4})\}, \qquad (5.10)$$

where $\langle | \rangle$ stands for the usual Wigner coefficient in SU_2. The reduced Wigner coefficient is then defined by the scalar product

$$\langle h_1' h_2' q_1' q_2', h_3' h_4' q_1'' q_2'') h_1 h_2 h_3, q_1 q_2 \rangle \equiv (P^{h_1 h_2 h_3}_{q_1 q_2 \frac{1}{2}(q_1 - q_2)}, \Pi^{h_1' h_2', h_3' h_4'}_{q_1' q_2', q_1'' q_2'', q_1 q_2 \frac{1}{2}(q_1 - q_2)}), \qquad (5.11)$$

where we have taken $\tau = \frac{1}{2}(q_1 - q_2)$ in both polynomials, since from (5.8) and (5.10) they are bases for irreducible representations of SU_2, and so, their scalar product[18] is independent of τ.

Because of the orthonormality property of the Wigner coefficients of SU_2, we could express the product (5.9) in terms of (5.10) and therefore, the Wigner coefficients of U_3 could be expressed as a product of the reduced Wigner coefficients (5.11) and the Wigner coefficients of SU_2.

The generalization of the present ideas to U_{2j+1} is straightforward. Again, as the structure of the group is given by SU_{2j+1}, we restrict ourselves to two representations characterized by the partitions $[h_1' h_2' \cdots h_{2j}']$, $[h_{2j+1}' h_{2j+2}' \cdots h_{4j}']$. The polynomials in $4j$ vectors must now satisfy equations similar to (5.5), where in the first set s, $s' = 1, 2, \cdots 2j$ and in the second, s, $s' = 2j + 1, \cdots, 4j$. We get the polynomials of highest weight by requiring that they also satisfy (5.6) with $\mu, \mu' = 1, 2, \cdots 2j+1$ and where, in the definition of the $\mathcal{C}_\mu{}^{\mu'}$ in (5.7), we have $s = 1, 2, \cdots 4j$.

At this step, instead of applying the full lowering operator for U_{2j+1}, equivalent to $\mathfrak{R}^{h_1 h_2 h_3}_{q_1 q_2 \tau}(\mathcal{C}_\mu{}^{\mu'})$ of (4.12) for U_3, we apply the operator in U_{2j+1} that lowers to maximum weight in U_{2j}, equivalent to $\mathfrak{R}(\mathcal{C}_\mu{}^{\mu'})$ of (4.7) for U_3. The polynomials thus obtained would be part of a basis for an irreducible representation of both U_{2j+1} and its subgroup U_{2j}, but of highest weight in the latter, i.e., they represent for U_{2j+1}, what (5.8) with $\tau = \frac{1}{2}(q_1 - q_2)$ represents for U_3.

As the polynomials associated with the partitions $[h_1' \cdots h_{2j}']$, $[h_{2j+1}' \cdots h_{4j}']$ have their rows labeled by the chain of subgroups $U_{2j} \supset U_{2j-1} \cdots$ of U_{2j+1}, we could use the Wigner coefficients of U_{2j} to construct, from the products of the two polynomials, one of highest weight in U_{2j}. For U_{2j+1} this would be the analogous polynomial to what (5.10), with $\tau = \frac{1}{2}(q_1 - q_2)$ is for U_3.

The reduced Wigner coefficient for U_{2j+1} will then be defined by the scalar product of the polynomials mentioned in the two previous paragraphs. The full Wigner coefficients of U_{2j+1}, which, as in the case of U_3 and SU_3, will be the most general Wigner coefficient of SU_{2j+1}, can then be factorized in terms of products of reduced Wigner coefficients of U_{2j+1}, U_{2j}, etc.

---

[18] Reference 3, p. 115.

6. DETERMINATION OF THE POLYNOMIAL SOLUTIONS

The remaining problem we had for the Wigner coefficients of U_3, was to determine the polynomials satisfying (5.5) and (5.6). We shall first see that this problem can be reinterpreted in a way that shows the existence of, in general, more than one linearly independent solution, which implies that the Kronecker product of two irreducible representations of U_3 is not multiplicity-free, i.e., each irreducible representation contained in the product may appear more than once.

We notice from the commutation relations (3.12), that $C_{ss'}$ with $s, s' = 1, 2, 3, 4$ could be considered as operators associated with the infinitesimal unitary transformations in a four-dimensional space of the vector indices. From (3.11) we can then interpret $\mathcal{C}_\mu{}^{\mu'}$ as the invariants associated with the operators $C_{ss'}$. We can now interchange the interpretation given to the indices μ and s in $a^+_{\mu s}$ and Eqs. (5.6) and (5.5) tell us that we are looking for the polynomials in a basis for an irreducible representation of U_4, characterized by the partition $[h_1h_2h_3]$, in which the subgroups U_2, whose generators are C_{11}, C_{12}, C_{21}, C_{22}, and C_{33}, C_{34}, C_{43}, C_{44}, respectively, are explicitly reduced so that the polynomials correspond to highest weight in these subgroups.

We could characterize the rows for the irreducible representations of U_4 by different chains of subgroups; for example,

$$U_4 \supset \begin{bmatrix} U_3 & 0 \\ 0 & U_1 \end{bmatrix}, \quad U_4 \supset \begin{bmatrix} U_2 & 0 \\ 0 & U_2 \end{bmatrix}, \quad U_4 \supset R_4, \text{ etc.}, \tag{6.1a, b, c}$$

but, as was discussed in Sec. 4, only in the case (6.1a) will the subgroups completely determine the rows.[4,19] As in our present problem the subgroup is the one in (6.1b), we will need, besides Eqs. (5.5), some other operator equations to define completely the polynomials.

The restatement of the problem that we achieved for U_3 can clearly be extended to U_{2j+1}. The equations corresponding to (5.6) and (5.5) for this case, tell us that we are looking for the polynomials in a basis for an irreducible representation of U_{4j}, characterized by the partition $[h_1h_2 \cdots h_{2j+1}]$, in which the subgroups U_{2j}, whose generators are $C_{ss'}$ with $s, s' = 1, \cdots 2j$, and $C_{ss'}$ with $s, s' = 2j + 1, \cdots, 4j$, are explicitly reduced, the polynomials being of highest weight in these subgroups.

[19] V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961).

By the same reasoning as in the previous paragraph, we conclude that these polynomials are, in general, nonunique, i.e., the Kronecker product of two irreducible representations of U_{2j+1} is not multiplicity-free.

We shall now proceed to obtain explicitly the solution of (5.6) and (5.5) making use of this reinterpretation. In Sec. 3 we indicated that the polynomial (3.20) was the most general solution of (3.14). If we now want the most general polynomial solution P of (5.6), we only need to interchange the roles of the upper and lower indices in the determinants in (3.20) to obtain

$$P = (\Delta_1^1)^{h_1-h_2}(\Delta_{12}^{12})^{h_2-h_3}(\Delta_{123}^{123})^{h_3} \times Z\left(\frac{\Delta_1^2}{\Delta_1^1}, \frac{\Delta_1^3}{\Delta_1^1}, \frac{\Delta_1^4}{\Delta_1^1}, \frac{\Delta_{12}^{13}}{\Delta_{12}^{12}}, \frac{\Delta_{12}^{14}}{\Delta_{12}^{12}}, \frac{\Delta_{123}^{124}}{\Delta_{123}^{123}}\right), \tag{6.2}$$

where Z is an arbitrary polynomial in the ratios indicated, subject only to the condition that P be a polynomial in the $a^+_{\mu s}$.

We now apply Eqs. (5.5) to (6.2). We start with (5.5c) where, since C_{12} changes upper index 2 into 1, we get

$$C_{12}P = \frac{P}{Z}\frac{\partial Z}{\partial(\Delta_1^2/\Delta_1^1)} = 0,$$

$$\text{or} \quad P \quad \text{independent of} \quad (\Delta_1^2/\Delta_1^1). \tag{6.3}$$

We then make use of the following identities between determinants:

$$\Delta_1^3\Delta_{12}^{14} - \Delta_1^4\Delta_{12}^{13} = \Delta_1^1\Delta_{12}^{34}, \tag{6.4a}$$

$$\Delta_{12}^{14}\Delta_{123}^{123} - \Delta_{12}^{13}\Delta_{123}^{124} = -\Delta_{12}^{12}\Delta_{123}^{134}, \tag{6.4b}$$

to express the ratios $(\Delta_{12}^{14}/\Delta_{12}^{12})$, $(\Delta_{123}^{124}/\Delta_{123}^{123})$ in terms of the ratios already present in (6.2) and of $(\Delta_{12}^{34}/\Delta_{12}^{12})$ and $(\Delta_{123}^{134}/\Delta_{123}^{123})$, so that the P that satisfies (6.3) can also be written as

$$P = (\Delta_1^1)^{h_1-h_2}(\Delta_{12}^{12})^{h_2-h_3}(\Delta_{123}^{123})^{h_3} \times Z\left(\frac{\Delta_1^3}{\Delta_1^1}, \frac{\Delta_1^4}{\Delta_1^1}, \frac{\Delta_{12}^{13}}{\Delta_{12}^{12}}, \frac{\Delta_{12}^{34}}{\Delta_{12}^{12}}, \frac{\Delta_{123}^{134}}{\Delta_{123}^{123}}\right). \tag{6.5}$$

Applying now (5.5f), we see that as C_{34} changes upper index 4 into 3, it implies that Z is independent of (Δ_1^4/Δ_1^1). We now expand the polynomial Z of (6.5) in powers of the four remaining ratios and apply Eqs. (5.5a, b, d, e), which just indicate that the degrees of the polynomial in the components of the vectors $a^+_{\mu s}$ are h'_s, $s = 1, 2, 3, 4$. Only three of these four equations are independent as

$$\sum_{s=1}^{4} C_{ss} = \sum_{\mu=1}^{3} \mathcal{C}_\mu{}^\mu, \tag{6.6}$$

and so the integer exponent in one of the ratios, say $(\Delta_{12}^{34}/\Delta_{12}^{12})$, is still undetermined and will be denoted by q. The most general solution of (5.6) and (5.5) becomes then

$$P \equiv \sum_q B_q P_q = (\Delta_{12}^{13})^{h_2-h_2'-h_4'} \times \sum_q \{B_q(\Delta_1^1)^{h_1'-h_2+q}(\Delta_1^3)^{h_1-h_1'-q}(\Delta_{12}^{12})^{h_2'+h_4'-h_3-q} \times (\Delta_{12}^{34})^q(\Delta_{123}^{123})^{h_3-h_4'+q}(\Delta_{123}^{134})^{h_4'-q}\}, \tag{6.7}$$

where B_q is an arbitrary constant.

We now have two cases to discuss, the first one when

$$h_2 - h_2' - h_4' \geq 0, \tag{6.8a}$$

where we shall get polynomial solutions P_q for all q's for which the exponents in (6.7) are nonnegative, i.e., for q's satisfying the inequalities

$$\begin{aligned} h_2 - h_1' &\leq q \leq h_1 - h_1', \\ h_4' - h_3 &\leq q \leq h_2' + h_4' - h_3, \\ 0 &\leq q \leq h_4'. \end{aligned} \tag{6.8b}$$

The second case corresponds, of course, to

$$h_2 - h_2' - h_4' < 0, \tag{6.9a}$$

and then the B_q's must be chosen in such a way that the polynomial expression in (6.7) is divisible by the appropiate power of Δ_{12}^{13}. We shall show how to find these B_q's by first remarking that the P of (6.7) can be written as a product of given powers of the determinants multiplied by the polynomial

$$\sum_q B_q y^q, \quad \text{where} \quad y \equiv \frac{\Delta_1^1\Delta_{12}^{34}\Delta_{123}^{123}}{\Delta_1^3\Delta_{12}^{12}\Delta_{123}^{134}}. \tag{6.10}$$

By using the identity

$$\Delta_1^1\Delta_{123}^{123} = \Delta_{12}^{12}\Delta_{13}^{13} - \Delta_{12}^{13}\Delta_{13}^{12}, \tag{6.11}$$

and an equivalent one for $\Delta_1^3\Delta_{123}^{134}$, we can write y as

$$y = -1 + (\Delta_{12}^{13}s)(\Delta_1^3\Delta_{12}^{12}\Delta_{123}^{134})^{-1}, \tag{6.12}$$

where

$$s \equiv \Delta_{12}^{12}\Delta_{13}^{34} - \Delta_{12}^{34}\Delta_{13}^{12}. \tag{6.13}$$

As the B_q's are so far arbitrary, we can rewrite the expression (6.10) as

$$(1 + y)^{h_2'+h_4'-h_2} \sum_q B_q' y^q, \tag{6.14}$$

where, from (6.9a), the exponent of $(1 + y)$ is positive. Replacing (6.10) by (6.14) in the polynomial (6.7), we obtain an alternative expression for this polynomial:

$$P \equiv \sum_q B_q' P_q' = s^{h_2'+h_4'-h_2} \times \sum_q \{B_q'(\Delta_1^1)^{h_1'-h_2+q}(\Delta_1^3)^{h_3'-h_3-q}(\Delta_{12}^{12})^{h_2-h_3-q} \times (\Delta_{12}^{34})^q(\Delta_{123}^{123})^{h_3-h_4'+q}(\Delta_{123}^{134})^{h_3-h_2'-q}\}. \tag{6.15}$$

From (6.9a), the exponent of s is positive, and we shall therefore get polynomial solutions P_q' for all q's for which the exponents in (6.15) are nonnegative, i.e., for q's that satisfy the inequalities

$$\begin{aligned} h_2 - h_1' &\leq q \leq h_2 - h_2', \\ h_4' - h_3 &\leq q \leq h_3' - h_3, \\ 0 &\leq q \leq h_2 - h_3. \end{aligned} \tag{6.9b}$$

The inequalities (6.8) and (6.9) can be interpreted in two ways. First, if $h_1'h_2'$, $h_3'h_4'$ are given, then we have polynomial solutions for all $h_1h_2h_3$ for which there is, at least, one q that satisfies the inequalities (6.8b) or (6.9b), depending on whether $h_2 - h_2' - h_4'$ is nonnegative or negative. These inequalities are then equivalent to Littlewood's[20] rules that give us the irreducible representations $[h_1h_2h_3]$ of U_3 contained in a Kronecker product of the irreducible representation $[h_1'h_2']$, $[h_3'h_4']$. If there is more than one q satisfying the inequalities, it follows that there is more than one linearly independent highest-weight polynomial corresponding to the irreducible representation $[h_1h_2h_3]$, i.e., the Kronecker product is not multiplicity-free. As an example of the application of the inequalities, we obtain the following expansion for the Kronecker product:

$$[21] \times [21] = [42] + [33] + [411] + 2[321] + [222], \tag{6.16a}$$

or

$$\{21\} \times \{21\} = \{42\} + \{33\} + \{3\} + 2\{21\} + \{0\}, \tag{6.16b}$$

where the square brackets correspond to representations of U_3 and the curly brackets to SU_3 in which all columns of three rows are supressed.

Another interpretation of the inequalities (6.8) and (6.9) is obtained when we consider $h_1h_2h_3$ as given. In this case there will be polynomial solutions for all $h_1'h_2'$, $h_3'h_4'$ for which there is, at least, one q that satisfies the inequalities (6.8b) or (6.9b), depending on whether $h_2 - h_2' - h_4'$ is nonnegative or negative. This then gives us the irreducible representations of the U_2 subgroups in the chain

[20] D. E. Littlewood, *The Theory of Group Characters* (Oxford University Press, New York, 1940), p. 94.

(6.1b) contained in an irreducible representation $[h_1h_2h_3]$ of U_4. If there is more than one q satisfying the inequalities, it implies that the rows of the representation $[h_1h_2h_3]$ of U_4 are not completely determined by this chain of subgroups.

What happens to the reduced Wigner coefficients in the case when we have more than one q satisfying either (6.8) or (6.9)? The answer is that we will have as many linearly independent polynomials P in the scalar product (5.11) as we have values of q, and so in the Wigner coefficient we must add a symbol q which distinguishes between the linearly independent bases for equivalent irreducible representations. This would be a natural way to characterize the Wigner coefficients, but it has the disadvantage that the coefficients will not satisfy orthonormality properties in the index q, as the polynomials corresponding to different q's are not, in general, orthogonal.

To have polynomials of highest weight that will be orthogonal, we could again turn to our reinterpretation of Eqs. (5.6) and (5.5) and ask how could we completely characterize the rows of an irreducible representation $[h_1h_2h_3]$ of U_4 in which we take the chain of subgroups (6.1b). A problem of this type was already considered by Bargmann and Moshinsky[19] when discussing the bases for the irreducible representations of U_3 in the chain $U_3 \supset R_3$. Following the reasoning developed in that paper, we need Casimir[5] operators of the *subgroup* that are formed from the generators of the *full group*. In this case we must look for Hermitian polynomial functions of $C_{ss'}$, $s, s' = 1, 2, 3, 4$ which will commute with the generators C_{11}, C_{12}, C_{21}, C_{22}, and C_{33}, C_{34}, C_{43}, C_{44} of the two U_2 groups in the chain (6.1b). The operators must be independent of the Casimir operators either of the full group U_4 or of the subgroups U_2, since for these, from (5.6) and (5.5), the P_q, P'_q would already be eigenpolynomials with eigenvalues independent of q.

A systematic procedure for finding these operators will be to consider polynomial functions of increasing degree in the $C_{ss'}$. It is easily seen that there are no polynomials of first and second degree that satisfy the requirements of the previous paragraph. For the third degree, a polynomial of the $C_{ss'}$ satisfying the requirement is

$$\mathrm{X} = \sum_{s,s',s''=1}^{2} C_{s+2s'}C_{s's''}C_{s''s+2}, \tag{6.17}$$

and it is possible to show that any other polynomial of third degree in the $C_{ss'}$ that satisfies the same restrictions can be expressed in terms of X and the Casimir operators of U_4 and its subgroups U_2.

Will the operator X completely define the rows of the representation $[h_1h_2h_3]$ of U_4 in the chain (6.1b)? To answer this question we again follow the procedure of reference 19 for the $U_3 \supset R_3$ chain. We first apply the operator X to P_q or P'_q, depending on whether $h_2 - h'_2 - h'_4$ is nonnegative or negative, and we obtain straightforwardly

$$\mathrm{X}P_q = \sum_r \beta_{rq}P_r, \tag{6.18a}$$

$$\mathrm{X}P'_q = \sum_r \beta'_{rq}P'_r, \tag{6.19a}$$

where the β's are only different from zero for $r = q \pm 1, q$ and for those cases

$$\begin{aligned}
\beta_{q+1q} &= -(h_1 - h'_1 - q)(h'_2 + h'_4 - h_3 - q)(h'_4 - q),\\
\beta_{qq} &= (h_1 - h'_1 - q)(h'_2 + h'_4 - h_3 - q)\\
&\times (h'_1 - h_2 + 1 + q) + q(h_3 - h'_4 + q)\\
&\times (h'_1 - h'_2 - h'_4 + 1 + q) + (h'_1 + 3)\\
&\times [(h_1 - h'_1)(h'_1 - h_2 + 1 + q)\\
&+ q(h_2 - h_3 + 1 - q)] + (h'_2 + 1)\\
&\times [(h'_2 + h'_4 - h_3 + 1 - q)(h_1 + h_2 - h'_1 - h'_2 - q)\\
&+ h'_4(h_3 - h'_4 + 1 + q)],\\
\beta_{q-1q} &= q(h'_1 - h_2 + q)(h_3 - h'_4 + q),
\end{aligned} \tag{6.18b}$$

and

$$\begin{aligned}
\beta'_{q+1q} &= -(h'_3 - h_3 - q)(h_2 - h_3 - q)(h_2 - h'_2 - q),\\
\beta'_{qq} &= (h'_1 - h_2 + q + 1)[(2h_3 - h_2 - h'_4 + 2q)q\\
&+ (h'_3 - h_3)(h'_1 - h'_2 - h'_4 + 2h_2 - h_3 - q + 3)]\\
&+ (h_2 - h_3 - q + 1)[(h'_1 - h'_2 + 2)q\\
&+ (h_2 - h'_4 + 1)(h'_3 - h_3 + h_2 - h'_2)]\\
&+ (h_2 - h'_4 + 1)(h_2 - h'_2)(h_3 - h'_4 + q + 1)\\
&+ (h'_2 + h'_4 - h_2)[(h'_1 + h'_2 - h_2 + q + 2)\\
&\times (h'_1 + h'_3 - h_3 + 3) + (h'_3 - h_3)(h'_1 - h_2)\\
&+ (h_2 - h'_2)(h_3 - h'_4) + (h'_3 - h_3 + h_2 - h'_2)\\
&\times (h_2 - h_3 + 2)],\\
\beta'_{q-1q} &= q(h_3 - h'_4 + q)(h'_1 - h_2 + q).
\end{aligned} \tag{6.19b}$$

From (6.18b) we see that the β_{rq} have the following properties: First, if we denote by q', q'' the minimum and the maximum value of q, according to the

inequalities (6.8b), then

$$\beta_{q-1q} = 0 \quad \text{for} \quad q = q', \qquad \beta_{q+1q} = 0 \quad \text{for} \quad q = q'', \tag{6.20a}$$

so that the sum in (6.18a) ranges over the r in the interval $q' \leq r \leq q''$ as it should. Second, in the allowed range,

$$\beta_{qq+1} \neq 0. \tag{6.20b}$$

From (6.19b), similar properties hold also for β'_{rq}.

The properties of the β_{rq} of the previous paragraph allow us to show, by exactly the same reasoning as in reference 19, the uniqueness of the polynomial solutions of

$$XP_\chi \equiv X \sum_q (B_q P_q) = \chi \sum_q (B_q P_q), \tag{6.21}$$

where the B_q, χ satisfy the linear equations

$$\sum_{r=q'}^{q''} \beta_{qr} B_r = \chi B_q. \tag{6.22}$$

A similar result holds of course for P'_q, β'_{rq}.

The operator X completely defines the rows of the irreducible representation $[h_1 h_2 h_3]$ of U_4 in the chain of subgroups (6.1b). We could then, instead of the polynomials P_q, P'_q, use the polynomials P_χ, P'_χ defined by (6.21) and (6.22) and the corresponding expressions for P'_q, β'_{rq}.

The reduced Wigner coefficient (5.11) could then be written

$$\langle h'_1 h'_2 q'_1 q'_2,\ h'_3 h'_4 q''_1 q''_2) h_1 h_2 h_3,\ q_1 q_2 \chi \rangle, \tag{6.23}$$

and as the operator (6.17) is Hermitian, we have the orthonormality properties with respect to index χ guaranteed.

The problem of the reduced Wigner coefficients of U_3 is then completely solved. The scalar product (5.11) may present some difficulties as regards its general evaluation, but at least for the case when $h'_4 = 0$, [which is multiplicity-free, as from (6.8b), $q = 0$] it has been explicitly carried out giving the algebraic expression (3.25) of reference 4.

For U_{2j+1}, the analysis is, in principle, the same. The polynomials in $4j$ vectors satisfying (3.16a, b) would have the form (3.20) in which the role of upper and lower indices in the determinants is interchanged. The polynomial Z depends then on $(3j - 1)(2j + 1)$ ratios. The equations equivalent to (5.5) for the U_{2j+1} group will be $2j(2j + 1)$ in number, and as

$$\sum_{\mu=1}^{2j+1} \mathcal{C}_\mu{}^\mu = \sum_{s=1}^{4j} C_{ss}, \tag{6.24}$$

one of them is redundant. We expect therefore to need

$$(j - 1)(2j + 1) + 1 \tag{6.25}$$

commuting operators of the form X, to completely define the bases. While a procedure similar to the one followed for U_3 could determine these operators, no general technique for deriving them *a priori* is, as yet, available.

7. CONCLUSION

In the present paper we obtained the bases for all the irreducible representations of unitary groups, and from them developed a procedure for evaluating their Wigner coefficients. Can the present analysis be extended from the unitary group to its subgroups? The generators of the subgroups will be linear combinations[16] of the $\mathcal{C}_\mu{}^{\mu'}$ of (3.8) and a suggestive procedure would be to look for the invariants with respect to these generators. If the invariants, which necessarily include the $C_{ss'}$ of (3.10), form a Lie Algebra, we could classify them in three sets similar to (3.13) and characterize the bases as in (3.14). In fact, we could say that this procedure was actually followed in Sec. 2 for the subgroup R_3 of U_3. This extension is being investigated so as to determine the bases for all irreducible representations of the semisimple compact Lie groups. From the bases we could then obtain the Wigner coefficients, and, by a recoupling procedure, the Racah coefficients[21] for these groups—a general program which, from another viewpoint, has also been initiated by L. C. Biedenharn.[22]

ACKNOWLEDGMENTS

The author is grateful to his colleagues in México for many discussions, suggestions, and for their encouragement. He is also indebted to Professor V. Bargmann and Professor B. Bayman for helpful discussions.

[21] G. Racah, Phys. Rev. **63**, 367 (1943).
[22] L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963).

Spin-Component Analysis of Single-Determinant Wavefunctions*

FUKASHI SASAKI† AND KIMIO OHNO‡
Quantum Chemistry Group, University of Uppsala, Uppsala, Sweden
(Received 16 April 1963)

The weights of spin components involved in a single-determinant wavefunction are obtained. The behavior of the weights for a system of a large number of electrons is discussed.

1. INTRODUCTION

A SINGLE-determinant form for wavefunctions has been used extensively in the quantum mechanics of many-electron systems.

In the usual Hartree–Fock method, single-determinant wavefunctions are used with a restriction that two electrons with spins α and β are put into the same space orbital. A single-determinant wavefunction with this restriction is an eigenfunction of the total spin operator $\mathbf{S}^2$. The Hartree–Fock method has proved to be very useful in the theory of atoms, molecules and solids.

In the unrestricted Hartree–Fock method, single-determinant wavefunctions without the above-mentioned restriction are used. This allows us to treat, in a compact form, the exchange polarization and the correlation of electrons with antiparallel spins. However, the unrestricted Hartree–Fock method has a disadvantage. Wavefunctions used in the method are in general not eigenfunctions of the total spin but are linear combinations of eigenfunctions which have different eigenvalues.

It is of some interest to see how much of each spin eigenfunction is contained in the single-determinant wavefunction.

2. FORMULATION OF THE PROBLEM

Any antisymmetric wavefunction of an N-electron system can be written in the following form:

$$\Phi(\xi_1, \xi_2, \cdots, \xi_N) = \mathcal{A}[\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)\Theta(\sigma_1, \sigma_2, \cdots, \sigma_N)], \quad (2.1)$$

where ξ_i stands for the space and spin coordinates $\mathbf{r}_i$ and σ_i, respectively, of the ith electron. $\mathcal{A}$ is the idempotent antisymmetrizing operator. By using permutation operators P and their parities ϵ_P, the antisymmetrizing operator $\mathcal{A}$ is expressed as

$$\mathcal{A} = \frac{1}{N!}\sum \epsilon_P P. \quad (2.2)$$

The wavefunction Φ is in general a linear combination of pure spin states,

$$\Phi = \sum_{S,M} \Phi_{S,M}, \quad (2.3)$$

where $\Phi_{S,M}$ is an eigenfunction of $\mathbf{S}^2$ and $\mathbf{S}_z$ with the eigenvalues $S(S+1)$ and M respectively ($\hbar = 1$).

This decomposition is of physical importance if the Hamiltonian does not involve spin operators. An expectation value of a spin-free operator f can be expressed as

$$\langle f\rangle = \frac{\langle\Phi, f\Phi\rangle}{\langle\Phi, \Phi\rangle} = \sum_{S,M} \omega_{S,M}\langle f\rangle_{S,M}, \quad (2.4)$$

where

$$\omega_{S,M} = \frac{\langle\Phi_{S,M}, \Phi_{S,M}\rangle}{\langle\Phi, \Phi\rangle} \geqq 0, \quad \left(\sum_{S,M} \omega_{S,M} = 1\right), \quad (2.5)$$

and

$$\langle f\rangle_{S,M} = \langle\Phi_{S,M}, f\Phi_{S,M}\rangle/\langle\Phi_{S,M}, \Phi_{S,M}\rangle, \quad (2.6)$$

when $\langle\Phi_{S,M},\Phi_{S,M}\rangle$ is not zero. When the Hamiltonian of the system is spin-free, we have

$$E = \sum_{S,M} \omega_{S,M}E_{S,M}. \quad (2.7)$$

This equation shows that at least one of the energy expectation values $E_{S,M}$ is lower than E unless all $E_{S,M}$ are equal to E. By selecting from Φ a suitable spin component $\Phi_{S,M}$, we have a wavefunction which is not only a spin eigenfunction but which has a lower energy expectation value.

The analysis is also useful in interpreting the function Φ, (2.1). This is in some cases (e.g., a single Slater determinant) much easier to handle than its components $\Phi_{S,M}$. If we know the values

* This work was sponsored in part by the King Gustaf VI Adolf's 70-Years Fund for Swedish Culture, Knut and Alice Wallenberg's Foundation, the Swedish Natural Science Research Council, and in part by the Aeronautical Research Laboratory, OAR, through the European Office, Aerospace Research, United States Air Force.

†‡ The authors are on leave of absence (1961) from Department of Physics, University of Tokyo, Tokyo, Japan.

of $\omega_{S,M}$ and $\langle f\rangle_{S,M}$, we can better interpret the nature of the simple form (2.1). One of the basic problems here is to determine the weight $\omega_{S,M}$.

In the following we shall investigate a special case in which the spin part Θ of the wavefunction (2.1) is a simple product of spin functions $\gamma_i(\sigma_i)$ which are either $\alpha(\sigma_i)$ or $\beta(\sigma_i)$:

$$\Theta(\sigma_1, \sigma_2, \cdots, \sigma_N) = \gamma_1(\sigma_1)\gamma_2(\sigma_2)\cdots\gamma_N(\sigma_N). \quad (2.8)$$

The spin function Θ is an eigenfunction of S_z. If we denote the number of α and β functions in (2.8) by N_α and N_β respectively, the eigenvalue of S_z (M) is expressed as

$$M = \tfrac{1}{2}(N_\alpha - N_\beta). \quad (2.9)$$

In order to evaluate $\omega_{S,M}$, it is convenient to introduce the spin operator $\mathcal{O}_{S,M}$ which projects out the component of the pure spin state:

$$\begin{aligned}[\mathsf{S}^2 - S(S+1)]\mathcal{O}_{S,M} &= 0,\\ [\mathsf{S}_z - M]\mathcal{O}_{S,M} &= 0,\\ \sum_{S,M}\mathcal{O}_{S,M} &= 1.\end{aligned} \quad (2.10)$$

We note that this operator works only on the spin part of a wavefunction. For any wavefunction

$$\Phi = \sum C_i\Psi_i(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)\Theta_i(\sigma_1, \sigma_2, \cdots, \sigma_N),$$

$\mathcal{O}_{S,M}\Phi$ is expressed as

$$\mathcal{O}_{S,M}\Phi = \sum_i C_i\Psi_i(\mathcal{O}_{S,M}\Theta_i).$$

Since the wavefunction under consideration is an eigenfunction of S_z with the eigenvalue M, we shall drop the subscript M in the following unless it causes some ambiguity.

The weight ω_S for the wavefunction Φ can be written as

$$\begin{aligned}\omega_S &= \frac{\langle\mathcal{O}_S\Phi, \mathcal{O}_S\Phi\rangle}{\langle\Phi, \Phi\rangle} = \frac{\langle\Psi\Theta, \mathcal{O}_S\mathcal{A}\Psi\Theta\rangle}{\langle\Psi\Theta, \mathcal{A}\Psi\Theta\rangle}\\ &= \frac{\sum_P \epsilon_P\langle\Psi, P^x\Psi\rangle\langle\Theta, \mathcal{O}_SP^\sigma\Theta\rangle}{\sum_P \epsilon_P\langle\Psi, P^x\Psi\rangle\langle\Theta, P^\sigma\Theta\rangle},\end{aligned} \quad (2.11)$$

where P^x and P^σ denote the corresponding permutations of the space and the spin coordinates respectively. Similarly we obtain the expectation value of a spin-free operator f,

$$\begin{aligned}\langle f\rangle_S &= \frac{\langle\mathcal{O}_S\Phi, f\mathcal{O}_S\Phi\rangle}{\langle\mathcal{O}_S\Phi, \mathcal{O}_S\Phi\rangle} = \frac{\langle\Psi\Theta, f\mathcal{O}_S\mathcal{A}\Psi\Theta\rangle}{\langle\Psi\Theta, \mathcal{O}_S\mathcal{A}\Psi\Theta\rangle}\\ &= \frac{\sum_P \epsilon_P\langle\Psi, fP^x\Psi\rangle\langle\Theta, \mathcal{O}_sP^\sigma\Theta\rangle}{\sum_P \epsilon_P\langle\Psi, P^x\Psi\rangle\langle\Theta, \mathcal{O}_SP^\sigma\Theta\rangle}.\end{aligned} \quad (2.12)$$

We shall derive an explicit formula for $\langle\Theta, \mathcal{O}_SP^\sigma\Theta\rangle$ in the next section. In Secs. 4 and 5, the weights ω_S derived from a single Slater determinant are given, and the behavior of the ω_S values for large N is discussed.

3. CALCULATIONS OF $\langle\Theta, \mathcal{O}_SP^\sigma\Theta\rangle$

Because we assumed the form (2.8) for the spin part Θ, for any permutation P we can find a permutation $Q(P)$ which brings Θ and $P^\sigma\Theta$ into the following forms:

$$Q^\sigma\Theta = \alpha(1)\alpha(2)\cdots\alpha(n+M)\beta(n+M+1)\cdots \times\beta(2n) \equiv f_0, \quad (3.1)$$

$$\begin{aligned}Q^\sigma P^\sigma\Theta = {}&\alpha(1)\cdots\alpha(n+M-i)\\ &\times\beta(n+M-i+1)\cdots\beta(n+M)\\ &\times\alpha(n+M+1)\cdots\alpha(n+M+i)\\ &\times\beta(n+M+i+1)\cdots\beta(2n) \equiv f_i\\ &(2n = N).\end{aligned} \quad (3.2)$$

Here the integer $i(P)$ is the number of α functions in Θ which are changed to β functions in $P^\sigma\Theta$. The number $i(P)$ is uniquely determined by the given permutation P.

Using the commutability of $\mathcal{O}_S$ and Q^σ, we obtain

$$\begin{aligned}\langle\Theta, \mathcal{O}_SP^\sigma\Theta\rangle &= \langle Q^\sigma\Theta, Q^\sigma\mathcal{O}_SP^\sigma\Theta\rangle\\ &= \langle Q^\sigma\Theta, \mathcal{O}_SQ^\sigma P^\sigma\Theta\rangle\\ &= \langle f_0, \mathcal{O}_Sf_i\rangle \equiv c_{S,i}.\end{aligned} \quad (3.3)$$

In order to calculate the value $c_{S,i}$, it is convenient to divide the total number of electrons into four groups A, B, C, and D. A, B, C, and D stand for the first $n+M-i$ electrons, the second i electrons, the third i electrons and the last $n+M-i$ electrons, respectively:

$$\begin{aligned}A&\quad 1, 2, \cdots, n+M-i,\\ B&\quad n+M-i+1, n+M-i+2, \cdots, n+M,\\ C&\quad n+M+1, n+M+2, \cdots, n+M+i,\\ D&\quad n+M+i+1, n+M+i+2, \cdots, 2n.\end{aligned}$$

For example when we write CD, this means the combined groups of C and D, i.e., the last $n-M$ electrons.

We introduce the symbol $Y_K(s, m; \mu)$ for one element of an orthonormal complete set of simultaneous spin eigenfunctions of S^2 and S_z with the eigenvalues $s(s+1)$ and m, respectively. K denotes

an electron system which can be any of the groups defined above. μ specifies one of the spin functions with common s and m, to differentiate degenerate functions. We adopt the usual convention for the relative phase of these functions[1]:

$$(\mathsf{S}_x \pm i\mathsf{S}_y)\,Y_K(s, m; \mu) = [(s \mp m)(s \pm m + 1)]^{\frac{1}{2}} Y_K(s, m \pm 1; \mu). \quad (3.4)$$

As the element of the set which has the highest eigenvalue of S^2 ($s = \frac{1}{2}N_K$, N_K is the number of electrons in the group K), we choose the following function:

$$Y_K(\tfrac{1}{2}N_K, \tfrac{1}{2}N_K) = \alpha\alpha \cdots \alpha. \quad (3.5)$$

We may drop μ in this case since (3.5) is the only function in the set with the eigenvalues $s = m = \frac{1}{2}N_K$. From (3.4) and (3.5), we obtain

$$Y_K(\tfrac{1}{2}N_K, -\tfrac{1}{2}N_K) = \beta\beta \cdots \beta. \quad (3.6)$$

An orthonormal set of spin functions of the system $ABCD$ can be obtained by coupling Y_{AB} and Y_{CD} in the following manner:

$$\begin{aligned} Y_{ABCD}(s, m; \mu) &= Y_{ABCD}(s, m; s's''\mu'\mu'') \\ &= \sum_{m'm''} Y_{AB}(s', m'; \mu')\, Y_{CD}(s'', m''; \mu'') \\ &\quad \times (s', m', s'', m'' \mid s', s'', s, m). \end{aligned} \quad (3.7)$$

Here $(s', m', s'', m'' \mid s', s'', s, m)$ is the vector coupling coefficient[2]:

$$\begin{aligned} (s', m', s'', m'' \mid s', s'', s, m) &= \delta(m' + m'' - m) \\ &\times \left[\frac{(2s+1)(s'+s''-s)!\,(s'-m')!\,(s''-m'')!\,(s+m)!\,(s-m)!}{(s'+s''+s+1)!\,(s'-s''+s)!\,(-s'+s''+s)!\,(s'+m')!\,(s''+m'')!}\right]^{\frac{1}{2}} \\ &\times \sum_{\nu} (-1)^{\nu+s'-m'} \frac{(s'+m'+\nu)!\,(s''+s-m'-\nu)!}{\nu!\,(s'-m'-\nu)!\,(s-m-\nu)!\,(s''-s+m'+\nu)!}. \end{aligned} \quad (3.8)$$

Since

$$\mathcal{O}_{S,M} = \sum_{\mu} Y_{ABCD}(S, M; \mu)\rangle\langle Y_{ABCD}(S, M; \mu), \quad (3.9)$$

we obtain

$$c_{S,i} = \sum_{\mu} \langle f_0, Y_{ABCD}(S, M; \mu)\rangle \times \langle Y_{ABCD}(S, M; \mu), f_i\rangle. \quad (3.10)$$

The functions f_0 and f_i are expressed as

$$f_0 = Y_{AB}\left(\frac{n+M}{2}, \frac{n+M}{2}\right) \times Y_{CD}\left(\frac{n-M}{2}, -\frac{n-M}{2}\right), \quad (3.11)$$

and

$$f_i = Y_A\left(\frac{n+M-i}{2}, \frac{n+M-i}{2}\right) Y_B\left(\frac{i}{2}, -\frac{i}{2}\right) \times Y_C\left(\frac{i}{2}, \frac{i}{2}\right) Y_D\left(\frac{n-M-i}{2}, -\frac{n-M-i}{2}\right). \quad (3.12)$$

Therefore, the terms of the right side of (3.10) vanish except for the following Y_{ABCD}:

$$\begin{aligned} &Y_{ABCD}\left(S, M; \frac{n+M}{2}, \frac{n-M}{2}\right) \\ &= \sum_{m'm''} Y_{AB}\left(\frac{n+M}{2}, m'\right) Y_{CD}\left(\frac{n-M}{2}, m''\right) \\ &\quad \times \left(\frac{n+M}{2}, m', \frac{n-M}{2}, m'' \,\middle|\, \frac{n+M}{2}, \frac{n-M}{2}, S, M\right). \end{aligned} \quad (3.13)$$

Here,

$$\begin{aligned} &Y_{AB}\left(\frac{n+M}{2}, m\right) \\ &= \sum_{m'm''} Y_A\left(\frac{n+M-i}{2}, m'\right) Y_B\left(\frac{i}{2}, m''\right) \\ &\quad \times \left(\frac{n+M-i}{2}, m', \frac{i}{2}, m'' \,\middle|\, \frac{n+M-i}{2}, \frac{i}{2}, \frac{n+M}{2}, m\right), \end{aligned} \quad (3.14)$$

and

[1] See, for example, E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, Cambridge, England, 1957).

[2] See, for example, A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1957), p. 44.

$$Y_{CD}\left(\frac{n-M}{2}, m\right)$$
$$= \sum_{m'm''} Y_C\left(\frac{i}{2}, m'\right) Y_D\left(\frac{n-M-i}{2}, m''\right)$$
$$\times \left(\frac{i}{2}, m', \frac{n-M-i}{2}, m'' \middle| \frac{i}{2}, \frac{n-M-i}{2}, \frac{n-M}{2}, m\right). \quad (3.15)$$

Using (3.8), (3.11), (3.12), (3.13), (3.14), and (3.15), we find

$$c_{S,i} = \left\langle f_0, Y_{ABCD}\left(S, M; \frac{n+M}{2}, \frac{n-M}{2}\right)\right\rangle$$
$$\times \left\langle Y_{ABCD}\left(S, M; \frac{n+M}{2}, \frac{n-M}{2}\right), f_i\right\rangle$$
$$= \left(\frac{n+M}{2}, \frac{n+M}{2}, \frac{n-M}{2}, -\frac{n-M}{2} \middle| \frac{n+M}{2}, \frac{n-M}{2}, S, M\right)$$
$$\times \left(\frac{n+M-i}{2}, \frac{n+M-i}{2}, \frac{i}{2}, -\frac{i}{2} \middle| \frac{n+M-i}{2}, \frac{i}{2}, \frac{n+M}{2}, \frac{n+M-2i}{2}\right)$$
$$\times \left(\frac{i}{2}, \frac{i}{2}, \frac{n-M-i}{2}, -\frac{n-M-i}{2} \middle| \frac{i}{2}, \frac{n-M-i}{2}, \frac{n-M}{2}, -\frac{n-M-2i}{2}\right)$$
$$\times \left(\frac{n+M}{2}, \frac{n+M-2i}{2}, \frac{n-M}{2}, -\frac{n-M-2i}{2} \middle| \frac{n+M}{2}, \frac{n-M}{2}, S, M\right)$$
$$= (2S+1)\frac{(n-M-i)!\,i!}{(n+1\ S+1!}\sum_{\nu}(-1)^{i-\nu}$$
$$\times \frac{(n+M-i+\nu)!\,(S-M+i-\nu)!}{\nu!\,(i-\nu)!\,(S-M-\nu)!\,(n-S-i+\nu)!} \quad (3.16)$$

$$= (2S+1)\frac{(n-M-i)!\,(S+M)!}{(S-M)!}\sum_{\nu}(-1)^{\nu}$$
$$\times \frac{\{(S-M+\nu)!\}^2}{\nu!(S-M+\nu-i)!(n-S-\nu)!(2S+1+\nu)!}. \quad (3.17)$$

It is convenient to use the expression (3.16) in calculating the values $c_{S,i}$ for some special cases. For example,

$$c_{S,0} = (2S+1)\frac{(n-M)!\,(n+M)!}{(n+S+1)!\,(n-S)!},$$
$$c_{M,i} = (-1)^i(2M+1)\frac{i!\,(n+M-i)!}{(n+M+1)!},$$
$$c_{n,i} = \frac{(n+M)!\,(n-M)!}{(2n)!}.$$

4. APPLICATION TO A SINGLE-DETERMINANT WAVEFUNCTION

When a wavefunction Φ is a single-determinant wavefunction, the space part $\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$ is a product of one-electron functions:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = \psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)\cdots\psi_N(\mathbf{r}_N). \quad (4.1)$$

We denote orbitals associated with α spins by $\phi_1, \phi_2, \cdots, \phi_{n+M}$ and those with β spins by $\varphi_1, \varphi_2, \cdots, \varphi_{n-M}$. Without changing the total wavefunction Φ, we can transform the orbitals $\{\phi\}$ and $\{\varphi\}$ to $\{\phi'\}$ and $\{\varphi'\}$ so that the only overlap remaining is between the pairs ϕ_i' and φ_i'[3]:

$$\phi_i' = \sum_{j=1}^{n+M} a_{ij}\phi_j, \qquad i = 1, 2, \cdots, n+M,$$
$$\varphi_i' = \sum_{j=1}^{n-M} b_{ij}\varphi_j, \qquad i = 1, 2, \cdots, n-M,$$
$$\langle\phi_i', \phi_j'\rangle = \langle\varphi_i', \varphi_j'\rangle = \delta_{ij}, \quad (4.2)$$
$$\langle\phi_i', \varphi_j'\rangle = \delta_{ij}\lambda_i^{\frac{1}{2}}, \qquad \lambda_i \geq 0.$$

Using these transformed orbitals for the space part Ψ, we find that the inner product $\langle\Psi, P^x\Psi\rangle$ vanishes except when P^x is a product of some interchanges of the pairs $\{\phi_i', \varphi_i'\}$. When P interchanges the electrons of t pairs $\{\phi_{k1}', \varphi_{k1}'\}, \{\phi_{k2}', \varphi_{k2}'\}, \cdots, \{\phi_{kt}', \varphi_{kt}'\}$, we see that

$$\langle\Psi, P^x\Psi\rangle = \lambda_{k1}\lambda_{k2}\cdots\lambda_{kt},$$
$$\epsilon_P = (-1)^t, \qquad i(P) = t. \quad (4.3)$$

It follows from (2.11), (3.3), and (4.3) that

$$\omega_S = \sum_{k=0}^{n-M}(-1)^k A_k c_{S,k}, \quad (4.4)$$

where A_k is defined by the coefficients of the polynomial

$$\prod_{k=1}^{n-M}(1-\lambda_k x) = \sum_{k=0}^{n-M}(-1)^k A_k x^k. \quad (4.5)$$

[3] A. T. Amos and G. G. Hall, Proc. Roy. Soc. (London) **A263**, 483 (1961).

For example,

$$A_0 = 1,$$

$$A_1 = \sum_k \lambda_k,$$

$$A_2 = \tfrac{1}{2}(\sum_k \lambda_k)^2 - \sum_k \lambda_k^2 = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \cdots + \lambda_1\lambda_{n-M} + \lambda_2\lambda_3 + \cdots + \lambda_{n-M-1}\lambda_{n-M},$$

$$\vdots$$

$$A_{n-M} = \prod_k \lambda_k.$$

In order to calculate (4.4), it is convenient to introduce the following polynomial $W(x)$:

$$W(x) = \prod_{k=1}^{n-M} [1 - (1 - \lambda_k)x] = \sum_{k=0}^{n-M} (-1)^k B_k x^k. \quad (4.6)$$

By comparing (4.5) and (4.6), it is found that

$$A_k = \sum_{j=0}^{k} (-1)^j \frac{(n - M - j)!}{(n - M - k)!\,(k - j)!} B_j. \quad (4.7)$$

The weights ω_S of a single-determinant wavefunction are, therefore, expressed as

$$\omega_S = \sum_{k=0}^{n-M} \sum_{j=0}^{k} (-1)^{j+k} \frac{(n - M - j)!}{(n - M - k)!\,(k - j)!} B_j c_{S,k} = (2S + 1) \frac{(S + M)!}{(S - M)!} \sum_{\nu=0}^{n-S} \sum_{j=0}^{n-M} \sum_{k=j}^{S-M+\nu} (-1)^{\nu+j+k} \times \frac{[(S - M + \nu)!]^2 (n - M - j)!}{(n - S - \nu)!\,(2S + 1 + \nu)!\,\nu!\,(k - j)!\,(S - M + \nu - k)!} B_j$$

$$= (2S + 1) \frac{(S + M)!}{(S - M)!} \sum_{\nu=0}^{n-S} (-1)^\nu \frac{[(S - M + \nu)!]^2}{(2S + 1 + \nu)!\,\nu!} B_{S-M+\nu}. \quad (4.8)$$

This can be written as an integral,

$$\omega_S = (-1)^{S-M} \frac{2S + 1}{(S - M)!} \times \int_0^1 x^{S-M}(1 - x)^{S+M} \frac{d^{S-M}W(x)}{dx^{S-M}}\, dx. \quad (4.9)$$

When $W(x)$ is given, we can derive ω_S using the above Eq. (4.9).

Since the first-order density matrix completely determines the original single-determinant wavefunction, it should also determine the weight ω_S. We derive an explicit expression for $W(x)$ in terms of the first-order density matrix. ω_S can be derived from $W(x)$ by using Eq. (4.9). The first-order density matrix of the wavefunction under consideration has the following form:

$$\rho(\xi, \xi') = \rho_+(\mathbf{r}, \mathbf{r}')\alpha(\sigma)\alpha(\sigma') + \rho_-(\mathbf{r}, \mathbf{r}')\beta(\sigma)\beta(\sigma'),$$

where

$$\rho_+ = \sum_{i=1}^{n+M} \phi_i'\rangle\langle\phi_i', \qquad \rho_- = \sum_{i=1}^{n-M} \varphi_i'\rangle\langle\phi_i'.$$

Since

$$\langle\varphi_i', [(1 - x)\rho_- + x\rho_-\rho_+]\varphi_i'\rangle = \delta_{ii}[1 - (1 - \lambda_i)x],$$

$W(x)$ can be expressed as an expectation value of an $n - M$ particle operator $K(x)$:

$$W(x) = \langle\psi^-, K(x)\psi^-\rangle,$$

where

$$K(x: \mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_{n-M}: \mathbf{r}_1', \mathbf{r}_2', \cdots, \mathbf{r}_{n-M}') = \prod_{i=1}^{n-M} [(1 - x)\rho_-(\mathbf{r}_i, \mathbf{r}_i') + x\rho_-\rho_+(\mathbf{r}_i, \mathbf{r}_i')]$$

and

$$\psi^-(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_{n-M}) = [(n - M)!]^{\frac{1}{2}} \mathcal{A}\varphi_1'(\mathbf{r}_1)\varphi_2'(\mathbf{r}_2) \cdots \varphi_{n-M}'(\mathbf{r}_{n-M}).$$

It is seen that $\mathcal{A}K(0)$ is a projection operator,

$$\mathcal{A}K(0) = \psi^-\rangle\langle\psi^-,$$

and $K(0)K(x) = K(x)$. Therefore, we obtain an expression of $W(x)$ in terms of ρ_+ and ρ_- only:

$$W(x) = \langle\psi^-, K(x)\psi^-\rangle = \mathrm{tr}\, \mathcal{A}K(0)K(x) = \mathrm{tr}\, \mathcal{A}K(x).$$

5. BEHAVIOR OF ω_S FOR LARGE N

We discuss the case when all λ_k's are equal to λ. Then

$$W(x) = [1 - (1 - \lambda)x]^{n-M}. \quad (5.1)$$

Putting (5.1) into (4.9), we obtain

$$\omega_S(\lambda) = (2S + 1) \frac{(n - M)!\,(1 - \lambda)^{S-M}}{(S - M)!\,(n - S)!} \times \int_0^1 x^{S-M}(1 - x)^{S+M}\{1 - (1 - \lambda)x\}^{n-S}\, dx. \quad (5.2)$$

The asymptotic form of (5.2) is derived in the appendix when $S - M$ is small compared with $N^{\frac{1}{2}}$,

$$\omega_S(\lambda) \sim \frac{(2S+1)(n-M)^{S-M}}{(1-\lambda)\left[n-M+\frac{2M}{(1-\lambda)}\right]^{S-M+1}}$$
$$\times \exp\left[-\frac{1}{2(n-M)}(S-M-1)(S-M)\right.$$
$$-\frac{2M+(1-\lambda)^2(n-M)}{2\{2M+(1-\lambda)(n-M)\}^2}$$
$$\times (S-M+1)(S-M+2)$$
$$\left.-\frac{\lambda}{2M+(1-\lambda)(n-M)}\right.$$
$$\left.\times (S-M)(S-M+1)\right]. \quad (5.3)$$

Equation (5.3) may be regarded valid over all possible values of S, since both left- and right-hand sides of (5.3) decrease rapidly as $S-M$ becomes large compared with $N^{\frac{1}{2}}$.

A. Orthogonal Case

Putting $\lambda = 0$ in the expression (5.2), we obtain

$$\omega_S(0) = (2S+1)\frac{(n+M)!\,(n-M)!}{(n-S)!\,(n+S+1)!}. \quad (5.4)$$

It is interesting to note that $\omega_S(0)$ is proportional to the number, $f_{n,S}$, of linearly independent spin functions for given $n(=\frac{1}{2}N)$ and S,

$$\omega_S(0) = f_{n,S} \Big/ \sum_{S'=M}^{n} f_{n,S'}. \quad (5.5)$$

The asymptotic form of $\omega_S(0)$ is, from (5.3),

$$\omega_S(0) \sim (2S+1)\frac{(n-M)^{S-M}}{(n+M)^{S-M+1}}$$
$$\times \exp\left[-\frac{1}{2(n-M)}(S-M-1)(S-M)\right.$$
$$\left.-\frac{1}{2(n+M)}(S-M+1)(S-M+2)\right]. \quad (5.6)$$

For $M = kn(0 < k < 1)$, ω_S can be approximated by a geometrical sequence,

$$\omega_S \sim (2S+1)\frac{(n-M)^{S-M}}{(n+M)^{S-M+1}}$$
$$\sim \frac{2k}{1+k}\left(\frac{1-k}{1+k}\right)^{S-M}. \quad (5.7)$$

It should be noted that (5.7) does not contain the number of electrons explicitly. Therefore, for a fixed k, ω_S does not change much when N increases.

For $M = 0$,

$$\omega_S \sim [(2S+1)/n]\exp[-(S^2+S+1)/n]. \quad (5.8)$$

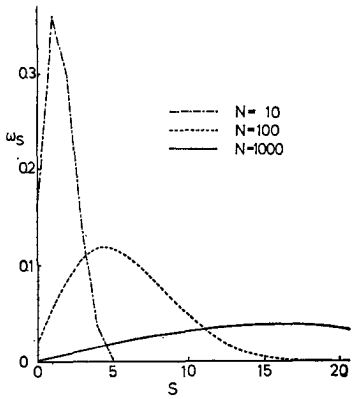

FIG. 1. The weights $\omega_S(0)$ as a function of S.

From (5.8), the expectation values of S, S^2, and S^3 can be calculated as

$$\langle S\rangle = \sum S\omega_S \sim \int \frac{2S^2}{n}\exp\left(-\frac{S^2}{n}\right)dS = \tfrac{1}{2}(n\pi)^{\frac{1}{2}},$$

$$\langle S^2\rangle = \sum S^2\omega_S \sim \int \frac{2S^3}{n}\exp\left(-\frac{S^2}{n}\right)dS = n,$$

$$\langle S^3\rangle = \sum S^3\omega_S \sim \int \frac{2S^4}{n}\exp\left(-\frac{S^2}{n}\right)dS = \tfrac{3}{4}(n^3\pi)^{\frac{1}{2}}.$$

The exact values are as follows:

$$\langle S\rangle = \tfrac{1}{2} + 2^{2n-1}(n!)^2/(2n)!,$$
$$\langle S^2\rangle = n + 1 - \langle S\rangle,$$
$$\langle S^3\rangle = -\tfrac{1}{4} + 2^{2n-2}(6n-5)n(n!)^2/(2n+1)!.$$

In Figs. 1 and 2, ω_S is plotted as a function of S and S/n, respectively, for $N = 10$, 100, and 1000, by using (5.8). We see from these figures that when N increases, ω_S spreads towards bigger N. However, for large N, appreciable weight ω_S is localized around the value $S_{\max} \sim 0.5N^{\frac{1}{2}}$. The second moment of the distribution around the average $\langle S\rangle$ is

$$\langle S^2\rangle - \langle S\rangle^2 \sim (1-\tfrac{1}{4}\pi)n \sim 0.107N.$$

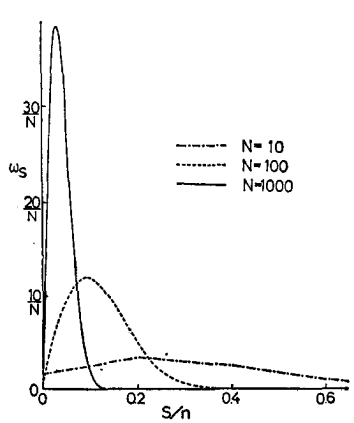

FIG. 2. The weights $\omega_S(0)$ as a function of S/n.

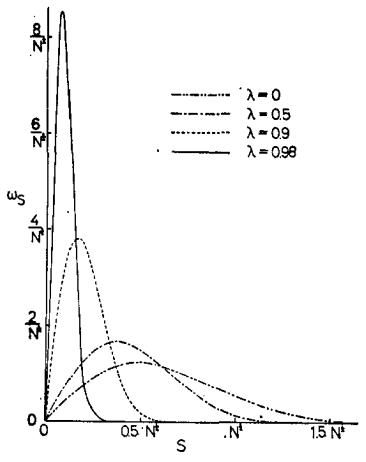

FIG. 3. The weights ω_s as a function of S when $M = 0$ for several values of the overlap integral.

B. Nonorthogonal Case

For $M \neq 0$,

$$\omega_S(n, M, \lambda) \sim \omega_S[(1-\lambda)(n-M)+M, M, 0], \quad (5.9)$$

since the factor

$$(2S+1)(n-M)^{S-M}/(1-\lambda)(n-M+2M/1-\lambda)^{S-M+1}$$

decreases rapidly compared with the exponential part in (5.3), which may therefore be regarded as 1.

For $M = 0$,

$$\omega_S(n, 0, \lambda) \sim \frac{2S+1}{(1-\lambda)n} \exp\left[-\frac{S^2+S+1-\lambda}{(1-\lambda)n}\right] \sim \omega_S\{(1-\lambda)n, 0, 0\}. \quad (5.10)$$

It follows from (5.9) and (5.10) that the weight in the nonorthogonal case for N electrons can be approximated by the weight for $2[(1-\lambda)(n-M)+M]$ electrons in the orthogonal case:

$$\omega_S(n, M, \lambda) \sim \omega_S\{(1-\lambda)(n-M)+M, M, 0\}. \quad (5.11)$$

When the overlap λ increases, the distribution shrinks. This can be seen from Figs. 3 and 4, where ω_S for some λ values are plotted as a function of S for $M = 0$ and $M = 0.2n$, respectively.

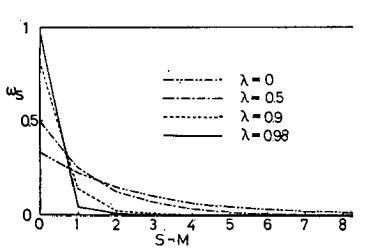

FIG. 4. The weights ω_s as a function of $S - M$ when $M = 0.2n$ for several values of the overlap integral.

ACKNOWLEDGMENT

The problem was suggested to us by Professor Per-Olov Löwdin. We would like to thank him for valuable discussions and for his generous hospitality extended to us in Uppsala.

We are indebted to Dr. J. W. Eastman for reading and criticizing the manuscript.

APPENDIX

We define a function I by the integral

$$I_{\alpha,\beta,\gamma}(z) \equiv \frac{1}{\alpha!}\int_0^1 x^\alpha(1-x)^\beta(1-zx)^\gamma\,dx. \quad (A1)$$

Then the weight (5.2) can be expressed as

$$\omega_S(\lambda) = (2S+1)\frac{(n-M)!\,(1-\lambda)^{S-M}}{(n-S)!} \times I_{S-M,S+M,n-S}(1-\lambda). \quad (A2)$$

The maximum of the integrand in (A1) is given by one of the roots of the following equation:

$$\frac{\alpha}{x_0} - \frac{\beta}{1-x_0} - \frac{z\gamma}{1-zx_0} = 0. \quad (A3)$$

Therefore,

$$x_0 = [(1+z)\alpha+\beta+z\gamma - \{[(1+z)\alpha+\beta+z\gamma]^2 - 4\alpha z(\alpha+\beta+\gamma)\}^{\frac{1}{2}}][2z(\alpha+\beta+\gamma)]^{-1}. \quad (A4)$$

The order of x_0 is the same as $\alpha/(\beta+z\gamma)$.

Expanding $(1-x)^\beta(1-zx)^\gamma \exp(\alpha x/x_0)$ in terms of $x - x_0$, we obtain

$$\frac{1}{\alpha!}x^\alpha(1-x)^\beta(1-zx)^\gamma = \{a_0 + a_2(x-x_0)^2 + a_3(x-x_0)^3 + \cdots\}(x^\alpha/\alpha!)\exp(-\alpha x/x_0). \quad (A5)$$

Here

$$a_0 = (1-x_0)^\beta(1-zx_0)^\gamma \exp\alpha, \quad (A6)$$
$$a_2 = -\frac{a_0}{2}\left(\frac{\beta}{(1-x_0)^2} + \frac{z^2\gamma}{(1-zx_0)^2}\right),$$

and

$$a_k = a_0 O(\alpha^{[\frac{1}{2}k]} x_0^{-[\frac{1}{2}k]}). \quad (A7)$$

Integrating (A5) over the range $(x = 0, x = \infty)$, we arrive at an asymptotic expansion

$$I \sim a_0 b_0 + a_2 b_2 + \cdots, \quad (A8)$$

where

$$b_k = \frac{1}{\alpha!}\int_0^\infty x^\alpha(x-x_0)^k \exp(-\alpha x/x_0)\,dx, \quad (A9)$$

$$b_0 = (x_0/\alpha)^{\alpha+1}, \tag{A10}$$

$$b_2 = b_0 x_0^2(\alpha + 2)/\alpha^2,$$

$$b_k = b_0 O(x_0^k \alpha^{-[\frac{1}{2}(k+1)]}). \tag{A11}$$

From (A5), (A7), and (A11), we find the order of magnitude of $a_k b_k$:

$$\frac{a_k b_k}{a_0 b_0} = \begin{cases} O(x_0^{\frac{1}{2}k}) & k: \text{ even.} \\ O(x_0^{\frac{1}{2}(k-1)}/\beta + z\gamma) & k: \text{ odd.} \end{cases} \tag{A12}$$

Remembering that the order of x_0 is the same as that of $\alpha/(\beta + z\gamma)$ and using (A12), we can take the first two terms in (A8) in order to calculate $\log \omega_S$ with the accuracy of order $(\beta + Z\gamma)^{-1}$. Then we obtain

$$I \sim \left\{1 - \frac{x_0^2(\alpha+2)}{2\alpha^2}\left[\frac{\beta}{(1-x_0)^2} + \frac{z^2\gamma}{(1-zx_0)^2}\right]\right\} \times \frac{x_0^{\alpha+1}}{\alpha^{\alpha+1}}(1-x_0)^\beta(1-zx_0)^\gamma \exp\alpha. \tag{A13}$$

The expression (A13) is substituted into (A2), and after some manipulation, we obtain (5.3).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 9 SEPTEMBER 1963

Lower Bounds for Energy Levels of Molecular Systems*

NORMAN W. BAZLEY
Battelle Memorial Institute, Geneva, Switzerland
AND
DAVID W. FOX
Battelle Memorial Institute, Geneva, Switzerland, and *Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland*
(Received 26 October 1962)

It is shown that the Hamiltonian H for a molecule in which the nuclei are regarded as fixed has the form $H = \sum H_\alpha + H'$, where H' is positive and each H_α has known eigenvalues and eigenvectors. A procedure is developed for the calculation of lower bounds to the eigenvalues of H. The optimization of the procedure and the existence of the eigenvalues are discussed.

I. INTRODUCTION

THIS paper is devoted to the development of procedures for the calculation of lower bounds for energies of electronic states in molecules and molecular ions. We take for our model of these systems the nonrelativistic quantum-mechanical Hamiltonian stated by Eq. (2.9), in which the nuclei are regarded as fixed and magnetic interactions are neglected. Our development consists of two main parts. The first shows how the Hamiltonian H for any such molecular system can be decomposed into the sum of Hamiltonians with known spectral families and a positive Hamiltonian, i.e.,

$$H = \sum H_\alpha + H', \tag{1.1}$$

in which each Hamiltonian H_α has known eigenvalues and eigenvectors and H' is positive. The second main part discusses the extension of lower-bound procedures to include these cases.

Section II gives the details of the decompositions, starting first from the simplest case, the H_2^+ molecular ion, and then treating the general molecular system. The nature of the spectra of the operators H_α that arise in the decompositions is also sketched.

Section III describes the extension of the lower-bound procedures. Truncations of the operators H_α and techniques of approximation of H' by the methods of intermediate problems are employed to construct new operators that give lower bounds from finite matrix problems. Section IV discusses how the results of Secs. II and III are to be used in the calculations of lower bounds for molecular systems. Optimization of the procedure and questions related to the existence of point eigenvalues in the first part of the spectra of these systems are also considered.

* This work was supported in part by the Department of the Navy under Contract NOrd 7386 with the Bureau of Naval Weapons.

$$b_0 = (x_0/\alpha)^{\alpha+1}, \tag{A10}$$

$$b_2 = b_0 x_0^2(\alpha + 2)/\alpha^2,$$

$$b_k = b_0 O(x_0^k \alpha^{-[\frac{1}{2}(k+1)]}). \tag{A11}$$

From (A5), (A7), and (A11), we find the order of magnitude of $a_k b_k$:

$$\frac{a_k b_k}{a_0 b_0} = \begin{cases} O(x_0^{\frac{1}{2}k}) & k: \text{ even.} \\ O(x_0^{\frac{1}{2}(k-1)}/\beta + z\gamma) & k: \text{ odd.} \end{cases} \tag{A12}$$

Remembering that the order of x_0 is the same as that of $\alpha/(\beta + z\gamma)$ and using (A12), we can take the first two terms in (A8) in order to calculate $\log \omega_S$ with the accuracy of order $(\beta + Z\gamma)^{-1}$. Then we obtain

$$I \sim \left\{1 - \frac{x_0^2(\alpha+2)}{2\alpha^2}\left[\frac{\beta}{(1-x_0)^2} + \frac{z^2\gamma}{(1-zx_0)^2}\right]\right\} \times \frac{x_0^{\alpha+1}}{\alpha^{\alpha+1}}(1-x_0)^\beta(1-zx_0)^\gamma \exp\alpha. \tag{A13}$$

The expression (A13) is substituted into (A2), and after some manipulation, we obtain (5.3).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 9 SEPTEMBER 1963

Lower Bounds for Energy Levels of Molecular Systems*

NORMAN W. BAZLEY
Battelle Memorial Institute, Geneva, Switzerland
AND
DAVID W. FOX
Battelle Memorial Institute, Geneva, Switzerland, and *Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland*
(Received 26 October 1962)

It is shown that the Hamiltonian H for a molecule in which the nuclei are regarded as fixed has the form $H = \sum H_\alpha + H'$, where H' is positive and each H_α has known eigenvalues and eigenvectors. A procedure is developed for the calculation of lower bounds to the eigenvalues of H. The optimization of the procedure and the existence of the eigenvalues are discussed.

I. INTRODUCTION

THIS paper is devoted to the development of procedures for the calculation of lower bounds for energies of electronic states in molecules and molecular ions. We take for our model of these systems the nonrelativistic quantum-mechanical Hamiltonian stated by Eq. (2.9), in which the nuclei are regarded as fixed and magnetic interactions are neglected. Our development consists of two main parts. The first shows how the Hamiltonian H for any such molecular system can be decomposed into the sum of Hamiltonians with known spectral families and a positive Hamiltonian, i.e.,

$$H = \sum H_\alpha + H', \tag{1.1}$$

in which each Hamiltonian H_α has known eigenvalues and eigenvectors and H' is positive. The second main part discusses the extension of lower-bound procedures to include these cases.

Section II gives the details of the decompositions, starting first from the simplest case, the H_2^+ molecular ion, and then treating the general molecular system. The nature of the spectra of the operators H_α that arise in the decompositions is also sketched.

Section III describes the extension of the lower-bound procedures. Truncations of the operators H_α and techniques of approximation of H' by the methods of intermediate problems are employed to construct new operators that give lower bounds from finite matrix problems. Section IV discusses how the results of Secs. II and III are to be used in the calculations of lower bounds for molecular systems. Optimization of the procedure and questions related to the existence of point eigenvalues in the first part of the spectra of these systems are also considered.

* This work was supported in part by the Department of the Navy under Contract NOrd 7386 with the Bureau of Naval Weapons.

II. DECOMPOSITIONS OF HAMILTONIANS OF MOLECULAR SYSTEMS

In this section we show how the Hamiltonians for molecular systems can be decomposed into the sum of operators H_α, each of which has a known spectral family, and a positive operator H'.

We consider a molecular system that consists of m electrons about n fixed nuclei of charges Z_α, $\alpha = 1, 2, \cdots, n$, in which we denote the position vectors of the electrons by r_i, $i = 1, 2, \cdots, m$, and those of the nuclei by R_α, $\alpha = 1, 2, \cdots, n$. The auxiliary vectors that are needed are defined by

$$R_{\alpha\beta} = R_\alpha - R_\beta, \qquad \alpha, \beta = 1, 2, \cdots, n; \tag{2.1}$$

$$r_{ij} = r_i - r_j, \qquad i, j = 1, 2, \cdots, m; \tag{2.2}$$

and

$$\mathfrak{r}_{i\alpha} = r_i - R_\alpha, \qquad i = 1, 2, \cdots, m; \quad \alpha = 1, 2, \cdots, n. \tag{2.3}$$

The vectors (2.1) define the nuclear arrangement, those of (2.2) give the interelectronic positions, and those of (2.3) are the position vectors of the electrons referred to the various nuclei. The magnitudes of vectors are indicated by the sign of absolute value.

For simplicity we begin by considering the simplest molecular system, the H_2^+ ion. Its Hamiltonian has the form[1] (in atomic units)

$$H = -\tfrac{1}{2}\Delta_1 - 1/|\mathfrak{r}_{11}| - 1/|\mathfrak{r}_{12}|, \tag{2.4}$$

in which Δ_1 denotes the Laplacian in the coordinates r_1 of the electron. Since the Laplacian is invariant with respect to translations of the coordinate origin, we may write

$$\Delta_1 = a_{11}\Delta_{11} + a_{12}\Delta_{12}, \tag{2.5}$$

in which a_{11} and a_{12} are any positive real numbers that satisfy

$$a_{11} + a_{12} = 1, \tag{2.6}$$

and Δ_{11} and Δ_{12} are the Laplacian expressed in the coordinates $\mathfrak{r}_{11}$ and $\mathfrak{r}_{12}$, respectively. From (2.4) and (2.5), the operator H takes the form

$$H = H_1 + H_2, \tag{2.7}$$

where

$$H_\alpha = -\tfrac{1}{2}a_{1\alpha}\Delta_{1\alpha} - 1/|\mathfrak{r}_{1\alpha}|, \qquad \alpha = 1, 2. \tag{2.8}$$

The operators H_1 and H_2 are Hamiltonians for hydrogenlike atoms and consequently their spectral families are completely known. Thus the Hamiltonian of the H_2^+ ion can be decomposed as we have asserted. In this case the term H' is missing.

For more general molecular systems, the decomposition is made in a similar way. H has the form

$$H = -\sum_{i=1}^{m} \tfrac{1}{2}\Delta_i - \sum_{\alpha=1}^{n}\sum_{i=1}^{m} \frac{Z_\alpha}{|\mathfrak{r}_{i\alpha}|} + \sum_{i>j=1}^{m} \frac{1}{|r_{ij}|}, \tag{2.9}$$

in which Δ_i denotes the Laplacian in the coordinates r_i of the ith electron. We introduce the positive constants $a_{i\alpha}$ that satisfy the equations

$$\sum_{\alpha=1}^{n} a_{i\alpha} = 1, \qquad i = 1, 2, \cdots, m, \tag{2.10}$$

and we write the Laplacians Δ_i as

$$\Delta_i = \sum_{\alpha=1}^{n} a_{i\alpha}\Delta_{i\alpha}, \qquad i = 1, 2, \cdots, m, \tag{2.11}$$

in which $\Delta_{i\alpha}$ means the Laplacian of the ith electron expressed in terms of the coordinates $\mathfrak{r}_{i\alpha}$, which refer to the αth nucleus as origin. When the expressions (2.11) for Δ_i are inserted in (2.9), H takes the form

$$H = \sum_{\alpha=1}^{n}\sum_{i=1}^{m} \left(-\frac{a_{i\alpha}\Delta_{i\alpha}}{2} - \frac{Z_\alpha}{|\mathfrak{r}_{i\alpha}|}\right) + \sum_{i>j=1}^{m} \frac{1}{|r_{ij}|}. \tag{2.12}$$

We define the operators H_α by

$$H_\alpha = \sum_{i=1}^{m} \left(-\frac{a_{i\alpha}\Delta_{i\alpha}}{2} - \frac{Z_\alpha}{|\mathfrak{r}_{i\alpha}|}\right), \qquad \alpha = 1, 2, \cdots, n, \tag{2.13}$$

and H' by

$$H' = \sum_{i>j=1}^{m} \frac{1}{|r_{ij}|}. \tag{2.14}$$

In terms of these new operators, H has the desired form

$$H = \sum_{\alpha=1}^{n} H_\alpha + H'. \tag{2.15}$$

For convenience we will designate by H^0 the sum

$$H^0 = \sum_{\alpha=1}^{n} H_\alpha. \tag{2.16}$$

Since the operator H' consists of a multiplication by a positive real function, it is clear that H' is a positive operator. Further, as we require, each operator H_α has completely known spectral families, for each is just the Hamiltonian for m uncorrelated electronic particles of masses $m_{i\alpha}$,

$$m_{i\alpha} = (a_{i\alpha})^{-1}; \qquad i = 1, 2, \cdots, m; \quad \alpha = 1, 2, \cdots, n, \tag{2.17}$$

[1] Here and in the following discussions we will omit the constant terms arising from nuclear repulsions, except when their inclusion is specifically stated.

about a single nucleus of charge Z_α. In fact, the spectrum of each H_α is expressible in terms of the spectrum of the hydrogen atom. Each eigenfunction Ψ^α of H_α has the form

$$\Psi^\alpha = \prod_{i=1}^{m} \Psi_{n_i,l_i,m_i}(m_{i\alpha} Z_\alpha \mathbf{r}_{i\alpha}), \tag{2.18}$$

in which Ψ_{n_i,l_i,m_i} is a hydrogen wavefunction. The corresponding eigenvalue E^α is given by

$$E^\alpha = -\frac{Z_\alpha^2}{2} \sum_{i=1}^{m} \frac{m_{i\alpha}}{n_i^2}. \tag{2.19}$$

The initial part of the spectrum of each H_α is discrete and consists of infinitely many eigenvalues converging to the first limit point E_*^α, which is given by

$$E_*^\alpha = -\frac{Z_\alpha^2}{2} \sum_{i=1}^{m}{}' m_{i\alpha}, \tag{2.20}$$

where $\sum'$ indicates that the smallest $m_{i\alpha}$ has been omitted from the summation. Further, a continuous spectrum extends from E_*^α to plus infinity, and between E_*^α and zero there are infinitely many eigenvalues which have as limit points the numbers

$$-\frac{Z_\alpha^2}{2} \sum_{i=1}^{m}{}'' \frac{m_{i\alpha}}{n_i^2}, \tag{2.21}$$

where $\sum''$ indicates that at least one term has been omitted in the sum.

The decomposition of H into the form (2.15) groups all of the electron repulsion terms together to form H'. The decomposition of the remaining part H^0 into the form (2.16) may be interpreted as a distribution of the kinetic energies of the electrons to the various nuclei. In fact, when each H_α is considered separately, one observes that the fraction $a_{i\alpha}$ of the kinetic energy of the ith electron has been associated with the αth nucleus.

Whenever the interelectronic forces do not appear or are neglected, the decomposition gives

$$H = \sum_{\alpha=1}^{n} H_\alpha. \tag{2.22}$$

This happens naturally for one-electron systems, e.g., the H_2^+ ion, and also when for comparison with other calculations, the model considered regards the electrons as uncorrelated.

III. LOWER-BOUND PROCEDURES

In this Section we extend previous work of the authors[2,3] in such a way that lower bounds to the eigenvalues can be calculated for the molecular systems under consideration. In those papers, we introduced modifications of the Weinstein–Aronszajn method of intermediate problems[4,5] such that the calculations for the lower bounds involve only finite matrices. The operators involved were of the form

$$H = H^0 + H' \tag{3.1}$$

in which H^0 had known spectral families and H' was positive, or

$$H = H_1 + H_2 \tag{3.2}$$

in which H_1 and H_2 each had known spectral families. One of the principal techniques was the use of *truncations* of one or both of the operators involved in order to construct new operators smaller than H for which the spectral problem could be solved easily. Here we use again truncations of operators to obtain our extensions to the cases in which the operators have the form (2.15) or (2.22). For the sake of efficiency and generality, our developments will be given in terms of operators in Hilbert space; nevertheless, the resulting procedures and formulas will be directly applicable to molecular systems. We begin by considering operators of the form (2.22), and later consider these of the more general form (2.15).

We suppose that H is a self-adjoint operator[6] with domain $\mathfrak{D}_H$ in a separable Hilbert space $\mathfrak{H}$ in which the inner product is (u, v) and that the lowest part of the spectrum of H consists of eigenvalues of finite multiplicity given in nondecreasing order by

$$E_1 \leq E_2 \leq \cdots \leq E_*, \tag{3.3}$$

where E_* is the first limit point of the spectrum of H. The corresponding orthonormal eigenvectors are denoted by $\Psi_1, \Psi_2, \cdots$. We assume that H has the form (2.22), that is,

$$H = H^0 = \sum_{\alpha=1}^{n} H_\alpha,$$

in which the operators H_α are of the same general type as H^0 and have known spectral families. The domain of H^0 is given in terms of the domains $\mathfrak{D}_{H_\alpha}$ of H_α by

$$\mathfrak{D}_{H^0} = \bigcap_{\alpha=1}^{n} \mathfrak{D}_{H_\alpha}. \tag{3.4}$$

[2] N. Bazley and D. W. Fox, J. Res. Natl. Bur. Std. **B65**, 105 (1961).

[3] N. Bazley and D. W. Fox, Phys. Rev. **124**, 483 (1961).

[4] A. Weinstein, Mém. Sci. Math. No. **88** (1937).

[5] N. Aronszajn, Proceedings of the Oklahoma Symposium on Spectral Theory and Differential Problems, Stillwater, Oklahoma, 1950, (Department of Mathematics, Oklahoma Agricultural and Mechanical College, Stillwater, Oklahoma, 1955) pp. 179–202.

[6] T. Kato [Trans. Am. Math. Soc. **70**, 195 (1951)] has shown that the operators of the form we consider in Sec. II are essentially self-adjoint.

As for H^0, we denote the lowest-ordered eigenvalues and first limit point of each operator H_α by

$$E_1^\alpha \leq E_2^\alpha \leq \cdots \leq E_*^\alpha, \tag{3.5}$$

and the corresponding orthonormal eigenvectors by $\Psi_1^\alpha, \Psi_2^\alpha, \cdots$.

We introduce the *truncated operators* $H_\alpha^{l_\alpha,0}$ defined by

$$H_\alpha^{l_\alpha,0}\Psi = \sum_{\nu=1}^{l_\alpha} (\Psi, \Psi_\nu^\alpha)E_\nu^\alpha\Psi_\nu^\alpha + E_{l_\alpha+1}^\alpha\left[\Psi - \sum_{\nu=1}^{l_\alpha} (\Psi, \Psi_\nu^\alpha)\Psi_\nu^\alpha\right], \tag{3.6}$$

in which l_α is a positive integer for each α. Each operator $H_\alpha^{l_\alpha,0}$ has the same first l_α eigenvalues as H_α; the rest of its spectrum consists of the point $E_{l_\alpha+1}^\alpha$, an eigenvalue of infinite multiplicity. The operators $H_\alpha^{l_\alpha,0}$ are bounded and satisfy the inequalities[7]

$$H_\alpha^{l_\alpha,0} \leq H_\alpha^{l_\alpha+1,0} \leq H_\alpha, \quad \alpha = 1, 2, \cdots, n. \tag{3.7}$$

In terms of these truncations, we define new operators $H^{l,0}$ by

$$H^{l,0} = \sum_{\alpha=1}^{n} H_\alpha^{l_\alpha,0}, \tag{3.8}$$

where the symbol l means the index vector $(l_1, l_2, \cdots, l_n)$. We will say $l^1 \leq l^2$ if and only if $l_\alpha^1 \leq l_\alpha^2$, $\alpha = 1, 2, \cdots, n$, and we will designate by $|l|$ the sum $\sum_{\alpha=1}^n l_\alpha$. The collection[8] of all l's will be denoted by L. From (3.7) it follows that the operators $H^{l,0}$ satisfy the inequalities

$$H^{l^1,0} \leq H^{l^2,0} \leq H^0, \tag{3.9}$$

whenever $l^1 \leq l^2$. Consequently, the ordered eigenvalues of $H^{l,0}$ give lower bounds to those of H^0 and obey the parallel inequalities

$$E_\nu^{l^1,0} \leq E_\nu^{l^2,0} \leq E_\nu^0, \qquad \nu = 1, 2, \cdots, \tag{3.10}$$

and

$$E_*^{l^1,0} \leq E_*^{l^2,0} \leq E_*^0. \tag{3.11}$$

The operators $H^{l,0}$ have the explicit expression

$$H^{l,0}\Psi = \sum_{\alpha=1}^{n}\left[\sum_{\nu=1}^{l_\alpha} (\Psi, \Psi_\nu^\alpha) \times (E_\nu^\alpha - E_{l_\alpha+1}^\alpha)\Psi_\nu^\alpha + E_{l_\alpha+1}^\alpha\Psi\right]. \tag{3.12}$$

From (3.12) it is clear that each operator $H^{l,0}$ is reduced by a finite-dimensional subspace $\mathfrak{M}^l$ of $\mathfrak{H}$ spanned by the vectors Ψ_ν^α, $\nu = 1, 2, \cdots, l_\alpha$, $\alpha = 1, 2, \cdots, n$. In fact, if Ψ is orthogonal to $\mathfrak{M}^l$ then $H^{l,0}\Psi = \sum_{\alpha=1}^n E_{l_\alpha+1}\Psi$, and if Ψ is in $\mathfrak{M}^l$ then $H^{l,0}\Psi$ is also in $\mathfrak{M}^l$. On $\mathfrak{M}^l$ the spectral problem for $H^{l,0}$ is equivalent to a matrix eigenvalue problem. Let us introduce the variables γ_ν^α defined by

$$\gamma_\nu^\alpha = (\Psi, \Psi_\nu^\alpha)(E_\nu^\alpha - E_{l_\alpha+1}^\alpha); \quad \nu = 1, 2, \cdots, l_\alpha; \qquad \alpha = 1, 2, \cdots, n. \tag{3.13}$$

In terms of these, the eigenvalue problem for $H^{l,0}$ has the form

$$H^{l,0}\Psi - E\Psi = \sum_{\alpha=1}^{n}\sum_{\nu=1}^{l_\alpha} \gamma_\nu^\alpha\Psi_\nu^\alpha - \left(E - \sum_{\alpha=1}^{n} E_{l_\alpha+1}^\alpha\right)\Psi = 0. \tag{3.14}$$

On taking inner products with the vectors Ψ_μ^β we come to the equivalent matrix eigenvalue problem,

$$\sum_{\alpha=1}^{n}\sum_{\nu=1}^{l_\alpha}\gamma_\nu^\alpha\left[(\Psi_\nu^\alpha, \Psi_\mu^\beta) - \left(E - \sum_{\alpha=1}^{n} E_{l_\alpha+1}^\alpha\right) \times \frac{\delta_{\nu\mu}\,\delta_{\alpha\beta}}{E_\nu^\alpha - E_{l_\alpha+1}^\alpha}\right] = 0, \quad \mu = 1, 2, \cdots, l_\beta; \quad \beta = 1, 2, \cdots, n. \tag{3.15}$$

This matrix problem of order $|l|$ gives eigenvalues which are smaller than or equal to $\sum_{\alpha=1}^n E_{l_\alpha+1}^\alpha$. Those eigenvalues that are strictly less than $\sum_{\alpha=1}^n E_{l_\alpha+1}^\alpha$ correspond to eigenvectors of $H^{l,0}$ which in turn are given (not normalized) by

$$\Psi = \sum_{\alpha=1}^{n}\sum_{\nu=1}^{l_\alpha}\gamma_\nu^\alpha\Psi_\nu^\alpha, \tag{3.16}$$

where the constants γ_ν^α are determined by (3.15). The remaining point in the spectrum of $H^{l,0}$ is $\sum_{\alpha=1}^n E_{l_\alpha+1}$, an eigenvalue of infinite multiplicity, for which the characteristic subspace consists of all vectors orthogonal to those given by (3.16). Consequently, the limit point $E_*^{l,0}$ equals $\sum_{\alpha=1}^n E_{l_\alpha+1}^\alpha$.

Thus, according to (3.10), the operators $H^{l,0}$ give improvable lower bounds to the eigenvalues of an operator H of the form (2.22), and the calculation of these lower bounds may be made from the matrix problem (3.15).

In the more general cases, the operators for which we wish lower bounds have the form (2.15), that is,

$$H = \sum_{\alpha=1}^{n} H_\alpha + H',$$

with domain $\mathfrak{D}_H$ given by

[7] For symmetric operators A and B, the inequality $A \leq B$ means that $\mathfrak{D}_B \subset \mathfrak{D}_A$ and $(A\Psi, \Psi) \leq (B\Psi, \Psi)$ for each Ψ in $\mathfrak{D}_B$.

[8] The set L, under the given ordering, forms a directed set.

$$\mathfrak{D}_H = \bigcap_{\alpha=1}^{n} \mathfrak{D}_{H\alpha} \bigcap \mathfrak{D}_{H'}. \tag{3.17}$$

As has just been shown, the operators $H^{l,0}$ are smaller than $\sum_{\alpha=1}^{n} H_\alpha$ and can be regarded as having known spectral families. As before,[2,3] we follow the procedure of Aronszajn[5] in introducing operators smaller than H'. In fact, let $\{p_1, p_2, \cdots\}$ be a sequence of vectors in $\mathfrak{D}_{H'}$ for which the vectors $\{H'p_1, H'p_2, \cdots\}$ form a linearly independent set, and introduce on $\mathfrak{D}_{H'}$ the auxiliary inner product $[\varphi, \Psi]$ defined by

$$[\varphi, \Psi] = (H'\varphi, \Psi). \tag{3.18}$$

Let P^k be the orthogonal projection with respect to this inner product on the subspace spanned by the first k vectors, $p_1, p_2, \cdots, p_k$. As has been shown,[2,3,5] the bounded symmetric operators $H'P^k$ satisfy the inequalities

$$0 \leq H'P^k \leq H'P^{k+1} \leq H', \tag{3.19}$$

and have the form

$$H'P^k\Psi = \sum_{i=1}^{k} \alpha_i H'p_i, \tag{3.20}$$

where

$$\sum_{i=1}^{k} \alpha_i (H'p_i, p_j) = (\Psi, H'p_j), \qquad j = 1, 2, \cdots, k. \tag{3.21}$$

Parallel to our earlier procedures,[2,3] we introduce the operators $H^{l,k}$ defined by

$$H^{l,k} = H^{l,0} + H'P^k. \tag{3.22}$$

These are bounded symmetric operators, and according to (3.9) and (3.19), they satisfy the inequalities

$$H^{l,k_1} \leq H^{l,k_2} \leq H \tag{3.23}$$

for $k_1 \leq k_2$, and

$$H^{l^1,k} \leq H^{l^2,k} \leq H \tag{3.24}$$

for $l^1 \leq l^2$. Consequently, the eigenvalues $E^{l,k}_\nu$ of $H^{l,k}$ satisfy the parallel inequalities

$$E^{l,k_1}_\nu \leq E^{l,k_2}_\nu \leq E_\nu, \qquad \nu = 1, 2, \cdots, \tag{3.25}$$

and

$$E^{l^1,k}_\nu \leq E^{l^2,k}_\nu \leq E_\nu, \qquad \nu = 1, 2, \cdots, \tag{3.26}$$

so that these operators give improvable lower bounds to the eigenvalues of H. The first limit points of the operators also satisfy the inequalities

$$E^{l,k_1}_* \leq E^{l,k_2}_* \leq E_*, \tag{3.27}$$

and

$$E^{l^1,k}_* \leq E^{l^2,k}_* \leq E_*. \tag{3.28}$$

The spectral families of the operators $H^{l,k}$ can be determined without difficulty since these operators are constructed so that they are reduced by finite-dimensional subspaces of $\mathfrak{H}$. In fact, for an operator $H^{l,k}$ the subspace $\mathfrak{M}^{l,k}$ spanned by the vectors Ψ^α_ν, $\nu = 1, 2, \cdots, l_\alpha$, $\alpha = 1, 2, \cdots, n$, and $H'p_i$, $i = 1, 2, \cdots, k$, is such a reducing subspace. This can be seen directly by writing

$$H^{l,k}\Psi = \sum_{\alpha=1}^{n} \sum_{\nu=1}^{l_\alpha} (\Psi, \Psi^\alpha_\nu)(E^\alpha_\nu - E^\alpha_{l_\alpha+1})\Psi^\alpha_\nu + \sum_{i=1}^{k} \alpha_i H'p_i + E^{l,0}_*\Psi, \tag{3.29}$$

in which the values α_i are determined from (3.21). If Ψ is in $\mathfrak{M}^{l,k}$, then $H^{l,k}\Psi$ is just a linear combination of vectors in $\mathfrak{M}^{l,k}$, and thus is also in $\mathfrak{M}^{l,k}$; on the other hand, if Ψ is orthogonal $\mathfrak{M}^{l,k}$, then $H^{l,k}\Psi = E^{l,0}_*\Psi$. On $\mathfrak{M}^{l,k}$ the spectral problem for $H^{l,k}$ is equivalent to a matrix eigenvalue problem. One form of this matrix problem is easily obtained from the eigenvalue equation

$$H^{l,k}\Psi - E\Psi = 0 \tag{3.30}$$

by taking inner products with the vectors spanning $\mathfrak{M}^{l,k}$ and using the quantities γ^α_ν defined by (3.13). The resulting matrix problem of order $|l| + k$ is stated by

$$\sum_{\alpha=1}^{n} \sum_{\nu=1}^{l_\alpha} \gamma^\alpha_\nu (\Psi^\alpha_\nu, \Psi^\beta_\mu) + \sum_{i=1}^{k} \alpha_i (H'p_i, \Psi^\beta_\mu) - (E - E^{l,0}_*) \sum_{\alpha=1}^{n} \sum_{\nu=1}^{l_\alpha} \gamma^\alpha_\nu \frac{\delta_{\nu\mu}\,\delta_{\alpha\beta}}{E^\alpha_\nu - E^\alpha_{l_\alpha+1}} = 0, \qquad \mu = 1, 2, \cdots, l_\beta; \quad \beta = 1, 2, \cdots, n, \tag{3.31}$$

and

$$\sum_{\alpha=1}^{n} \sum_{\nu=1}^{l_\alpha} \gamma^\alpha_\nu (\Psi^\alpha_\nu, H'p_j) + \sum_{i=1}^{k} \alpha_i (H'p_i, H'p_j) - (E - E^{l,0}_*) \sum_{i=1}^{k} \alpha_i (H'p_i, p_j) = 0, \qquad j = 1, 2, \cdots, k. \tag{3.32}$$

The eigenvalues of $H^{l,k}$ that differ from $E^{l,0}_*$ arise from the matrix problem (3.31), (3.32); the corresponding eigenvectors (not normalized) have the form

$$\Psi = \sum_{\alpha=1}^{n} \sum_{\nu=1}^{l_\alpha} \gamma^\alpha_\nu \Psi^\alpha_\nu + \sum_{i=1}^{k} \alpha_i H'p_i, \tag{3.33}$$

where the constants γ_ν^α and α_i are determined by (3.31) and (3.32). The remaining point in the spectrum is $E_*^{l,0}$, an eigenvalue of infinite multiplicity, for which the corresponding characteristic subspace consists of the orthogonal complement to those eigenvectors given by (3.33). Hence $E_*^{l,k}$ is equal to $E_*^{l,0}$.

Thus the operators $H^{l,k}$ give improvable lower bounds to those in the initial part of the spectrum of an operator of the form (2.15), and the calculation of the lower bounds may be made from the matrix problem (3.31), (3.32).

Parallel to our earlier results,[3] it is also possible to obtain lower bounds for operators of the form (2.15) by introducing the operators $\tilde{H}^{l,k}$ defined by

$$\tilde{H}^{l,k} = H^{l,0} + (H - H^{l,0})P_l^k, \qquad (3.34)$$

in which P_l^k denotes the projection on the span of the vectors p_i, which must now belong to $\mathfrak{D}_H$, with respect to the inner product $[\varphi, \Psi]_l$ given by

$$[\varphi, \Psi]_l = ([H - H^{l,0}]\varphi, \Psi). \qquad (3.35)$$

The operators $\tilde{H}^{l,k}$ have the property that they agree with H on the span of the p's. The procedures for determining the eigenvalues and eigenvectors of these operators are parallel to those for $H^{l,k}$. In fact, it is necessary only to replace H' by $H - H^{l,0}$ in the formulas (3.31), (3.32), and (3.33).

The procedures of this section, as well as those of H. F. Weinberger,[9] can be deduced from an operator inequality of quite general form which will be reported by the authors in a forthcoming publication.[10]

IV. LOWER BOUNDS FOR MOLECULAR ELECTRONIC ENERGY LEVELS

The decompositions of molecular Hamiltonians given in Sec. II, and the extensions of the lower-bound techniques sketched in Sec. III, make it possible to calculate lower bounds to the eigenvalues of molecular systems. In the cases when the decomposed Hamiltonian has the form (2.22), the only inner products that need to be computed are those between pairs of hydrogenic wavefunctions with different centers. When the Hamiltonian has the form (2.15), addditional inner products involving the essentially arbitrary vectors p_i must be carried out as well.

In both cases, the inner products and the resulting lower bounds determined from the matrix calculations depend parametrically on the values of the constants $a_{i\alpha}$ as well as on the indices k and l, and on the choice of the vectors p_i. It is natural to ask how the numbers $a_{i\alpha}$ should be chosen to give optimum results in our lower-bound procedures. A partial answer is given by demonstrating that one choice makes the lowest eigenvalues and eigenvectors of $H^{l,0}$ agree exactly with those of H^0 when the nuclei are brought together, and that the same choice maximizes the value of $\lim_L E_*^{l,0}$.

When the nuclei are brought together we have the corresponding *united atom* of nuclear charge $\sum_{\alpha=1}^n Z_\alpha$. The Hamiltonian H^0 then takes the form

$$H^0 = \sum_{\alpha=1}^n H_\alpha = -\frac{1}{2}\sum_{i=1}^m \Delta_i - \sum_{\beta=1}^n Z_\beta \cdot \sum_{i=1}^m \frac{1}{|\mathfrak{r}_i|}, \qquad (4.1)$$

in which $|\mathfrak{r}_i|$ is the distance from the ith electron to the united nucleus. The spectral problem for H^0 is explicitly solvable and gives the eigenvalues

$$E = -\frac{1}{2}\left(\sum_{\beta=1}^n Z_\beta\right)^2 \cdot \sum_{i=1}^m \frac{1}{n_i^2}, \qquad (4.2)$$

with the corresponding eigenvectors

$$\Psi = \prod_{i=1}^m \Psi_{n_i, l_i, m_i}\left(\sum_{\beta=1}^n Z_\beta \cdot \mathfrak{r}_i\right), \qquad (4.3)$$

and a continuous spectrum extending from $E_*^0(0)$, which is given by

$$E_*^0(0) = -\frac{m-1}{2}\left(\sum_{\beta=1}^n Z_\beta\right)^2 \qquad (4.4)$$

to plus infinity. A comparison of the eigenvectors (4.3) with those of (2.18) suggests the choice of constants $a_{i\alpha}$ given by

$$a_{i\alpha} = Z_\alpha\left(\sum_{\beta=1}^n Z_\beta\right)^{-1}, \qquad \alpha = 1, 2, \cdots, n;$$
$$i = 1, 2, \cdots, m, \qquad (4.5)$$

which is independent of i. In fact, with the choice (4.5) for $a_{i\alpha}$, each H_α given by (2.13) becomes

$$H_\alpha = Z_\alpha\left(\sum_{\beta=1}^n Z_\beta\right)^{-1}\left[-\frac{\Delta_{i\alpha}}{2} - \sum_{\beta=1}^n Z_\beta \cdot \sum_{i=1}^m \frac{1}{|\mathfrak{r}_{i\alpha}|}\right],$$
$$\alpha = 1, 2, \cdots, n, \qquad (4.6)$$

so that for the united atom we have

$$H_\alpha = Z_\alpha\left(\sum_{\beta=1}^n Z_\beta\right)^{-1} \cdot \sum_{\beta=1}^n H_\beta$$
$$= Z_\alpha\left(\sum_{\beta=1}^n Z_\beta\right)^{-1} H^0. \qquad (4.7)$$

When each l_α is chosen equal to a common value l_0, and we designate $(l_0, l_0, \cdots, l_0)$ by l^0, then (4.7)

[9] H. F. Weinberger, Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland, Technical Note BN-183 (1959).

[10] N. Bagley and D. W. Fox, Battelle Memorial Institute, Geneva, Switzerland, Technical Note, July, 1963.

shows that for the united atom

$$H^{l_0,0} = \sum_{\alpha=1}^{n} H_\alpha^{l_0} = (H^0)^{l_0}, \tag{4.8}$$

where $(H^0)^{l_0}$ means the truncation of H^0 of order l_0. Consequently, the choice (4.5) makes the first l_0 eigenvalues of $H^{l_0,0}$ agree with those of H^0. When l is arbitrary, we let $l_0 = \min l_\alpha (\alpha = 1, 2, \cdots, n)$ and observe that, according to (3.9),

$$H^{l_0,0} \leq H^{l,0} \leq H^0. \tag{4.9}$$

Since the operators on the right and left in (4.9) have the same first l_0 eigenvalues and eigenvectors, the characterization of the eigenvalues of $H^{l,0}$ as recursive minima[11] shows that $H^{l,0}$ has the same l_0 first eigenvalues and eigenvectors as well.

The methods of Sec. III show how the operators $H^{l,k}$ lead to the computations of lower bounds for the eigenvalues of H that lie below $E_*^{l,0}$. Thus, no matter how l is chosen, it is impossible by these methods to find lower bounds to eigenvalues of H that lie above $\lim_L E_*^{l,0}$.[12] To show that the choice (4.5) maximizes this limit, which for convenience we designate by C, we recall that according to (3.11),

$$E_*^{l,0}(a_{i\alpha}) \leq E_*^0(R_{\alpha\beta}), \tag{4.10}$$

where we note that $E_*^{l,0}$ depends on $a_{i\alpha}$ but not on $R_{\alpha\beta}$, while E_*^0 depends on $R_{\alpha\beta}$ and not on $a_{i\alpha}$. Hence we have for the limit,

$$C(a_{i\alpha}) \leq E_*^0(R_{\alpha\beta}). \tag{4.11}$$

In particular, (4.11) holds for the united atom, that is,

$$C(a_{i\alpha}) \leq E_*^0(0) = -\frac{m-1}{2}\left(\sum_{\beta=1}^{n} Z_\beta\right)^2. \tag{4.12}$$

However, for the united atom and the choice (4.5), we have the equality

$$C\left[Z_\alpha\left(\sum_{\beta=1}^{n} Z_\beta\right)^{-1}\right] = E_*^0(0), \tag{4.13}$$

and so the choice (4.5) maximizes C. We remark that, from (4.11) and (4.13), it follows that

$$-\frac{m-1}{2}\left(\sum_{\beta=1}^{n} Z_\beta\right)^2 = E_*^0(0) \leq E_*^0(R_{\alpha\beta}), \tag{4.14}$$

and consequently the minimum value of $E_*^0(R_{\alpha\beta})$ occurs for the united atom, that is, when $R_{\alpha\beta} = 0$.

According to (4.14), the number $E_*^0(0)$ is a lower bound for the first limit point of H^0; further, the inequality (3.28) implies that[13]

$$-\frac{m-1}{2}\left(\sum_{\beta=1}^{n} Z_\beta\right)^2 = E_*^0(0) \leq E_*. \tag{4.15}$$

Thus H can have only point spectrum to the left of $E_*^0(0)$. With this fact, the Rayleigh–Ritz procedure can establish the existence of the eigenvalues of H that lie below $E_*^0(0)$. In fact, if $\mathfrak{M}$ is a Rayleigh–Ritz manifold such that the quadratic form of H is less than $E_*^0(0)(\Psi, \Psi)$ for each Ψ in $\mathfrak{M}$, then H has at least as many eigenvalues below $E_*^0(0)$ as the dimension of $\mathfrak{M}$. If $\mathfrak{M}$ is infinite-dimensional,[14] the initial part of the spectrum of H consists of a point spectrum converging to $E_*^0(0)$.

[11] See, for example, S. H. Gould, *Variational Methods for Eigenvalue Problems* (The University of Toronto Press, Toronto; 1957).

[12] $\lim_L$ is the Moore–Smith limit and is the same as $\lim_{l_\alpha \to \infty} (\alpha = 1, 2, \cdots, n)$.

[13] When the nuclear-repulsion terms are included in H, we have

$$\sum_{\alpha>\beta=1}^{n} \frac{Z_\alpha Z_\beta}{|R_{\alpha\beta}|} - \frac{m-1}{2}\left(\sum_{\beta=1}^{n} Z_\beta\right)^2 \leq E^*.$$

[14] T. Kato [Trans. Am. Math. Soc. **70**, 212 (1951)] has given such manifolds for the isoelectronic series of helium for nuclear charges greater than one; others can be easily constructed for one-electron molecular ions.

An Exactly Soluble Model of a Many-Fermion System*

J. M. LUTTINGER
Department of Physics, Columbia University, New York, New York
(Received 2 April 1963)

An exactly soluble model of a one-dimensional many-fermion system is discussed. The model has a fairly realistic interaction between pairs of fermions. An exact calculation of the momentum distribution in the ground state is given. It is shown that there is no discontinuity in the momentum distribution in this model at the Fermi surface, but that the momentum distribution has infinite slope there. Comparison with the results of perturbation theory for the same model is also presented, and it is shown that, for this case at least, the perturbation and exact answers behave qualitatively alike. Finally, the response of the system to external fields is also discussed.

I. INTRODUCTION

WE shall be concerned in this paper with a model of a many-fermion system which is exactly soluble. The model is quite unrealistic for two reasons: it is one-dimensional and the fermions are massless. On the other hand, it has the realistic feature that there is a true pair interaction between the particles. It is very closely related to the well-known Thirring Model[1] in field theory, though slightly more general. Our main interest in the model is in connection with the question of whether or not a sharp Fermi Surface (F.S.) exists in the exact ground state.

This question has only been investigated previously[2] by a special sort of many-body perturbation theory, when it has been shown for the usual realistic three-dimensional many-fermion system that each term of the series does give rise to a sharp F.S. This, of course, proves nothing about the entire series unless one can also prove something about its convergence, which has not been possible so far. The main point of this investigation therefore is to see if in this soluble model the exact solution and the perturbation solution (via propagators) behave in an essentially different fashion.

We now consider the exact formulation of the model. Consider first the case of no interaction between the particles. These are taken to be spinless, massless, fermions moving in a one-dimensional space. The analogue of the relativistic Dirac Hamiltonian is $v_0\sigma_3 p$ (σ_3 is the usual Pauli spin matrix; units such that $\hbar = 1$ are chosen). v_0 is the velocity of the particles, which would be c in the relativistic case. Then the Hamiltonian is

$$H_0 = v_0 \int_0^L \psi^+(x)\sigma_3 p\psi(x)\, dx. \tag{1}$$

Here ψ is the two component spinor

$$\psi = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}, \tag{2}$$

and we are assuming that the particles are confined to a length L along the x axis. The quantity p is of course the ordinary momentum operator $1/i\,\partial/\partial x$.

Written out, (1) becomes

$$H_0 = v_0 \int_0^L [\psi_1^+ p\psi_1 - \psi_2^+ p\psi_2]\, dx. \tag{3}$$

If we go into momentum space via

$$\psi_i(x) = \sum_k a_{ik}e^{ikx}/L^{\frac{1}{2}} \tag{4}$$

(where the allowed values of k are

$$k = (2\pi/L)n, \qquad n = 0, \pm1, \pm2, \cdots, \pm\infty \tag{5}$$

since we shall impose periodic boundary conditions on our sample), we obtain

$$H_0 = v_0 \sum_k (a_{1k}^+ a_{1k} - a_{2k}^+ a_{2k})k. \tag{6}$$

The creation and destruction operators a, a^+ satisfy the commutation relationship

$$a_{jk}^+ a_{j'k'} + a_{jk'}a_{jk}^+ = \delta_{jj}\delta_{kk'}. \tag{7}$$

Since the allowed values of $a_{jk}^+ a_{jk}$ are zero and unity, the lowest state of H_0 is $-\infty$ since we can choose all the $j = 1$, $k < 0$ and the $j = 2$, $k > 0$ states occupied. This is the usual problem occurring in Dirac theory and requires a redefinition of the creation and destruction operators so that we deal only with "particles" and "holes". Define

* Work supported in part by the Office of Naval Research.
[1] W. Thirring, Ann. Phys. **3**, 91 (1958). See also V. Glaser, Nuovo Cimento **9**, 990 (1958); T. Pradhan, Nucl. Phys. **9**, 124 (1961); K. Johnson, Nuovo Cimento **21**, 773 (1961).
[2] J. M. Luttinger and J. C. Ward, Phys. Rev. **118**, 1417 (1960); J. M. Luttinger, *ibid.* **119**, 1153 (1960); **121**, 942 (1961).

$$\begin{aligned} a_{1k} &= b_k \quad k > 0 \\ &= c_k^+ \quad k < 0, \\ a_{2k} &= b_k \quad k < 0 \\ &= c_k^+ \quad k > 0. \end{aligned} \tag{8}$$

We may also write this as

$$b_k = \theta_k^+ a_{1k} + \theta_k^- a_{2k}, \qquad c_k = \theta^- a_{1k}^+ + \theta_k^+ a_{2k}^+, \tag{9}$$

where

$$\begin{aligned} \theta_k^+ &= 1 \quad k > 0 \\ &= 0 \quad k < 0, \\ \theta_k^- &= 1 \quad k < 0 \\ &= 0 \quad k > 0. \end{aligned} \tag{10}$$

From (9) we see at once that b_k, c_k also have the commutation rules of fermions, i.e.,

$$b_k^+ b_{k'} + b_{k'} b_k^+ = \delta_{kk'}, \quad c_k^+ c_{k'} + c_{k'} c_k^+ = \delta_{kk'}, \tag{11}$$

and all the rest anticommute.

Inserting (8) in (6) we obtain

$$H_0 = v_0 \sum_k (b_k^+ b_k + c_k^+ c_k)\ |k| + v_0 (\sum_{k<0} k - \sum_{k>0} k). \tag{12}$$

The last term is infinite, but a constant, and as usual we simply redefine H_0 without it, i.e., we take

$$H_0 = v_0 \sum_k (b_k^+ b_k + c_k^+ c_k)\ |k|. \tag{13}$$

We shall call the operators b_k and c_k the destruction operators for particles and holes respectively. The vacuum state ϕ_0 is clearly defined by

$$b_k \phi_0 = 0, \qquad c_k \phi_0 = 0. \tag{14}$$

The interaction Hamiltonian H' is taken to be (this special choice is what makes the model soluble)

$$H' = 2\lambda v_0 \iint_0^L \psi_1^+(x)\psi_1(x) V(x-y) \times \psi_2^+(y)\psi_2(y)\, dx\, dy. \tag{15}$$

$V(x - y)$ is an arbitrary two-body potential at this point. If we write this in momentum space [assuming also that $V(x - y)$ satisfies periodic boundary conditions], we obtain

$$H' = \frac{2\lambda v_0}{L} \sum_{k_1 \cdots k_4} \delta_{k_1 - k_2 + k_3 - k_4, 0} \times v(k_3 - k_4) a_{1k_1}^+ a_{1k_2} a_{2k_3}^+ a_{2k_4}, \tag{16}$$

where

$$\begin{aligned} V(x) &= \frac{1}{L} \sum_k v(k) e^{-ikx}, \\ v(k) &= \int_0^L dx e^{ikx} V(x). \end{aligned} \tag{17}$$

The term in (16) corresponding to $k_3 = k_4$, say H'', is given by

$$H'' = \frac{2\lambda}{L} v_0 v(0) (\sum_k a_{1k}^+ a_{1k})(\sum_k a_{2k}^+ a_{2k}). \tag{18}$$

This term clearly gives rise to divergent effects, since for the unperturbed vacuum the number of "1" and "2" particles are infinite. To avoid this difficulty, we shoose $v(0)$ to be zero, which is the same as taking the average value of the potential ($\bar{V}$) equal to zero. We also express this by saying that in (15) we replace $V(x - y)$ by $V(x - y) - \bar{V}$.

The total Hamiltonian of the problem is now given by

$$H = H_0 + H'. \tag{19}$$

II. EXACT SOLUTION OF THE MODEL

We shall show that (19) can be diagonalized by a very simple canonical transformation. Consider

$$\tilde{H} = e^{i\lambda S} H e^{-i\lambda S}, \tag{20}$$

where

$$S \equiv \iint_0^L dx\, dy \psi_1^+(x)\psi_1(x) E(x-y) \psi_2^+(y)\psi_2(y). \tag{21}$$

Here $E(x)$ is defined by

$$dE(x)/dx = V(x) - \bar{V}. \tag{22}$$

Writing

$$V(x) - \bar{V} = \frac{1}{L} \sum_k{}' v(k) e^{-ikx}, \tag{23}$$

we obtain

$$E(x) = \frac{i}{L} \sum_k{}' \frac{v(k)}{k} e^{-ikx}. \tag{24}$$

Let us define

$$N_i(x) = \psi_i^+(x)\psi_i(x). \tag{25}$$

Then from the commutation rules it follows at once that

$$(N_i(x), N_{i'}(x')) = 0, \tag{26}$$

so that

$$e^{i\lambda S} H' e^{-i\lambda S} = H'. \tag{27}$$

(In the non-second quantized version of the theory H' and S are just functions of position.)

Therefore

$$\tilde{H} = H + i\lambda(S, H_0) + [(i\lambda)^2/2'](S, (S, H_0)) + \cdots . \quad (28)$$

Using the commutation rules for the ψ_j, we obtain at once

$$(S, H_0) = v_0 \int_0^L\!\!\int dx\, dy \left\{ \frac{\partial N_1(x)}{\partial x} E(x-y) N_2(y) - N_1(x) E(x-y) \frac{\partial N_2(y)}{\partial y} \right\}. \quad (29)$$

Integrating by parts and using the periodic boundary conditions to drop the surface terms, we obtain

$$\begin{aligned}(S, H_0) &= -\frac{2v_0}{i} \int_0^L\!\!\int dx\, dy N_1(x) E'(x-y) N_2(y) \\ &= -\frac{2v_0}{i} \int_0^L\!\!\int dx\, dy N_1(x)(V(x-y) - \bar{V}) N_2(y) \\ &= -\frac{1}{i\lambda} H'. \end{aligned} \quad (30)$$

Since this commutes with S, there are no higher terms in the series (28), and we obtain

$$\tilde{H} = \tilde{H} - H' = H_0. \quad (31)$$

(Again these results are seen very easily by going over to the non-second quantized representation.)

Now $\tilde{H}$ is trivial to diagonalize, just being the noninteracting Hamiltonian. Therefore, all the energy levels of H are the same as those of H_0. (This is very unrealistic indeed.) On the other hand, the wavefunctions of H are very different from the free-particle ones. If ψ_n^0 is a wavefunction of H_0 corresponding to energy E_n^0, then the corresponding wavefunction for H (say, ψ_n) is

$$\psi_n = e^{-i\lambda S} \psi_n^0. \quad (32)$$

Therefore, although the energy levels do not change as a result of the interaction, other properties depending on more details of the wavefunction may be profoundly altered.

We next want to formulate the many-body problem for our system. We at once have the following problem: since particle-hole pairs can be produced by the interaction, the number of particles in an eigenstate of H is not fixed. However, we clearly must have that the number of particles minus the number of holes (call this n) is fixed in an eigenstate. Writing

$$n = \sum_k (b_k^+ b_k - c_d^+ c_k), \quad (33)$$

we can easily verify by direct calculation that n is a constant of the motion.

The noninteracting case for the N-particle problem is clearly the case of n having the eigenvalue N. Similarly, we define the N-particle problem for the interacting case as the system for which n has the value N. There will always be a certain number of holse present, but the smaller the interaction, the smaller this number will be.

The exact ground state of the N-particle system may be obtained as follows. Certainly the lowest state (ψ_N^0) of $\tilde{H}$ for which $n = N$ is obtained by having no holes present. Then the first N particle states will be occupied. That is

$$\begin{aligned} b_k^+ \psi_N^0 &= 0, \quad |k| < k_F, \\ b_k \psi_N^0 &= 0, \quad |k| > k_F, \\ c_k \psi_N^0 &= 0, \end{aligned} \quad (34)$$

where the Fermi momentum k_F is determined by

$$N = \sum_{|k|<k_F} 1 = \frac{L}{2\pi} \int_{-k_F}^{k_F} dk = \frac{L}{\pi} k_F. \quad (35)$$

We may also write

$$\psi_N^0 = b_{k_N}^+ \cdots b_{k_1}^+ \phi_0, \quad (36)$$

where ϕ_0 is the unperturbed vacuum and $k_1 \cdots k_N$ are the N allowed momenta between $-k_F$ and k_F.

Therefore the exact ground-state wavefunction (ψ_N) is given by

$$\psi_N = e^{-i\lambda S} \psi_N^0. \quad (37)$$

In order to study the sharpness of the F.S., we must investigate[2] the mean number of particles with momentum k, say $\bar{n}_k$. We have, of course,

$$n_k = b_k^+ b_k,$$

so

$$\bar{n}_k = (\psi_N, b_k^+ b_k \psi_N) = (\psi_N^0, e^{i\lambda S} b_k^+ b_k e^{-i\lambda S} \psi_N^0). \quad (38)$$

If we wanted to know the average number of holes $\bar{N}_h$ present we may use

$$\bar{N}_h = \sum_k \bar{n}_k - N. \quad (39)$$

Clearly $\bar{n}_k$ is an even function of k, so we shall restrict ourselves to $k > 0$. Then, by (9),

$$b_k^+ b_k = a_{1k}^+ a_{1k} = \frac{1}{L}\int_0^L\int_0^L d\xi\, d\eta e^{ik(\xi-\eta)}\psi_1^+(\xi)\psi_1(\eta), \quad (40)$$

as one sees, by direct integration,

$$\bar{n}_k = \frac{1}{L}\int_0^L\int_0^L d\xi\, d\eta e^{ik(\xi-\eta)} \times (\psi_N^0 | e^{i\lambda S}\psi_1^+(\xi)\psi_1(\eta)e^{-i\lambda S} | \psi_N^0). \quad (41)$$

Now we have the following operator identity

$$\exp\left[i\lambda\int_0^L g(x)N_1(x)\,dx\right]\psi_1(\eta) \times \exp\left[-i\lambda\int_0^L g(x)N_1(x)\,dx\right] = e^{-i\lambda g(\eta)}\psi_1(\eta), \quad (42)$$

if $g(x)$ commutes with $\psi_1(\eta)$. This is most easily proved by differentiating with respect to λ and making use of the fact that

$$(\psi_1(\eta), N_1(x)) = \delta(x-\eta)\psi_1(\eta). \quad (43)$$

Using (42), (41) becomes

$$\bar{n}_k = \frac{1}{L}\int_0^L\int_0^L d\xi\, d\eta e^{ik(\xi-\eta)} \times \left(\psi_N^0 \left| \psi_1^+(\xi)\psi_1(\eta)\exp\left\{i\lambda\int_0^L dyN_2(y) \times [E(\xi-y)-E(\eta-y)]\right\}\right| \psi_N^0\right). \quad (44)$$

Expressed in terms of a_{jk}, (34) becomes

$$\begin{aligned} a_{1k}^+\psi_N^0 &= 0, \quad k < k_F, \\ a_{1k}\psi_N^0 &= 0, \quad k > k_F, \end{aligned} \quad (45)$$

and

$$\begin{aligned} a_{1k}^+\psi_N^0 &= 0, \quad k > -k_F, \\ a_{2k}\psi_N^0 &= 0, \quad k < -k_F, \end{aligned} \quad (46)$$

Writing $\psi_N^0 = \Psi_1\Psi_2$ where Ψ_1, depends on the variables of the field "1" and is given by (45), and Ψ_2 depends on the variables of the field "2" and is given by (46), we have

$$\bar{n}_k = \frac{1}{L}\int_0^L\int_0^L d\xi\, d\eta e^{ik(\xi-\eta)}(\Psi_1 | \psi_1^+(\xi)\psi_1(\eta) | \Psi_1) \times \left(\Psi_2 \left| \exp\left\{i\lambda\int_0^L dyN_2(y) \times [E(\xi-y)-E(\eta-y)]\right\}\right| \Psi_2\right). \quad (47)$$

From (45) we have at once

$$(\Psi_1 | \psi_1^+(\xi)\psi_1(\eta) | \Psi_1) = \frac{1}{L}\sum_{k'<k_F} e^{-ik'(\xi-\eta)}. \quad (48)$$

The second factor in (47) is also not difficult to reduce to simpler form. We have, in fact,

$$\left(\Psi_2 \left| \exp\left\{i\lambda\int_0^L dyN_2(y) \times [E(\xi-y)-E(\eta-y)]\right\}\right| \Psi_2\right) = \mathrm{Det}\,(g). \quad (49)$$

Det (g) is the determinant of the matrix $g_{\alpha\alpha'}$, where

$$g_{\alpha\alpha'} = \frac{1}{L}\int_0^L dy e^{-i(k_\alpha - k_{\alpha'})y} \times \exp\{i\lambda[E(\xi-y)-E(\eta-y)]\}, \quad (50)$$

the k_α being the occupied states of the "2" particles in (46), i.e., the k_α are the set of discrete allowed k values greater than $-k_F$. The proof of (50) is given in the Appendix. The remarkable thing is that this (infinite) determinant can in fact be evaluated and the answer reduced to quadratures.

Writing Det $(g) = G(\xi, \eta)$, (47) becomes

$$\begin{aligned} \bar{n}_k &= \frac{1}{L^2}\int_0^L\int_0^L d\xi\, d\eta \sum_{k'<k_F} e^{+i(k-k')(\xi-\eta)}G(\xi,\eta) \\ &\equiv \frac{2\pi}{L}\sum_{k'<k_F} F(k-k'), \end{aligned} \quad (51)$$

where

$$F(\kappa) = \frac{1}{2\pi L}\int_0^L\int_0^L d\xi\, d\eta e^{i\kappa(\xi-\eta)}G(\xi,\eta). \quad (52)$$

III. EXPLICIT EVALUATION OF MOMENTUM DISTRIBUTION

We now must consider the determinant $G(\xi, \eta)$ in more detail. Since $k_\alpha = (2\pi/L)n$,

$$k_\alpha - k_{\alpha'} = (2\pi/L)(n-n');$$
$$n, n' = -n_F, -n_F+1, \cdots, \infty, \quad (53)$$

we may write

$$G = \begin{vmatrix} g_0 & g_{-1} & g_{-2} & \cdots \\ g_1 & g_0 & g_{-1} & \cdots \\ g_2 & g_1 & g_0 & \cdots \\ g_3 & g_2 & g_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{vmatrix}, \quad (54)$$

where

$$g_m = \frac{1}{L}\int_0^L dy e^{-2\pi i m y/L} \times \exp\{i\lambda[E(\xi - y) - E(\eta - y)]\}. \quad (55)$$

[(54) incidentally, is independent of k_F.]

This type of determinant has been studied extensively, and is known as a *Toeplitz* determinant.[3] For very large order, an asymptotic formula can be given for them, which in our case (infinite-determinant) becomes exact. The result is the following: for a finite Toeplitz determinant

$$D_M = \begin{vmatrix} g_0 & g_{-1} & \cdots & g_{-M} \\ g_1 & g_0 & \cdots & g_{-M+1} \\ \vdots & \vdots & & \vdots \\ g_M & g_{M-1} & \cdots & g_0 \end{vmatrix},$$

we have[4]

$$\lim_{M\to\infty} \frac{D_M}{D^{M+1}} = \exp\left(\sum_{l=1}^{\infty} K_l K_{-l} l\right), \quad (56)$$

where

$$D = \exp\left[\frac{1}{2\pi}\int_0^{2\pi} d\theta \log f(\theta)\right], \quad (57)$$

$$K_l = \frac{1}{2\pi}\int_0^{2\pi} d\theta e^{-il\theta} \log f(\theta), \quad (58)$$

$$f(\theta) = \sum_{m=-\infty}^{\infty} g_m e^{im\theta}. \quad (59)$$

In the proof, $\log f(\theta)$ is defined by

$$\log f(\theta) = \log\{1 - [1 - f(\theta)]\} = -\sum_{n=1}^{\infty} \frac{[1 - f(\theta)]^n}{n}, \quad (60)$$

and it is assumed that this series converges.

In our case, this leads to particularly simple results. Changing variables in (55) from y to θ where

$$\theta = 2\pi y/L, \qquad 0 \le \theta \le 2\pi,$$

we obtain

$$g_m = \frac{1}{2\pi}\int_0^{2\pi} d\theta e^{-im\theta} \times \exp\left\{i\lambda\left[E\left(\xi - \frac{L\theta}{2\pi}\right) - E\left(\eta - \frac{L\theta}{2\pi}\right)\right]\right\}. \quad (61)$$

Therefore from (59) we see at once that

$$f(\theta) = \exp\left\{i\lambda\left[E\left(\xi - \frac{L\theta}{2\pi}\right) - E\left(\eta - \frac{L\theta}{2\pi}\right)\right]\right\}. \quad (62)$$

Thus for sufficiently small λ, (60) is clearly satisfied since as one easily sees from (23) or (24), $E(x)$ is a bounded function of x. We shall for simplicity assume that λ is sufficiently small, and therefore we may write

$$\log f(\theta) = i\lambda\left[E\left(\xi - \frac{L\theta}{2\pi}\right) - E\left(\eta - \frac{L\theta}{2\pi}\right)\right]. \quad (63)$$

Now

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta \log f(\theta) = \frac{i\lambda}{L}\int_0^L dy[E(\xi - y) - E(\eta - y)] = 0, \quad (64)$$

since, by (24), the average of $E(x)$ is zero. Therefore

$$D = 1. \quad (65)$$

Further,

$$K_l = \frac{1}{2\pi}\int_0^{2\pi} d\theta e^{-il\theta} \log f(\theta) = \frac{i\lambda}{L}\int_0^L dy[E(\xi - y) - E(\eta - y)]e^{-2\pi i l y/L} = \frac{\lambda}{L}\left[(e^{-ik\eta} - e^{-ik\xi})\frac{v(k)}{k}\right]_{k=2\pi l/L}. \quad (66)$$

Then

$$\sum_{l=1}^{\infty} K_l K_{-l} l = -\frac{\lambda^2}{2\pi}\frac{1}{L}\sum_{k>0}^{\infty} \frac{|v(k)|^2\,|e^{-i\eta k} - e^{-i\xi k}|^2}{k} = -\frac{\lambda^2}{\pi}\frac{1}{L}\sum_{k>0}^{\infty} |v(k)|^2 \frac{1 - \cos k(\xi - \eta)}{k} \equiv -Q(\xi - \eta). \quad (67)$$

So finally we have

$$G(\xi, \eta) = e^{-Q(\xi - \eta)}. \quad (68)$$

Using the periodicity of Q in ξ and η, we see that (52) may be written

$$F(\kappa) = \frac{1}{2\pi}\int_{-\frac{1}{2}L}^{\frac{1}{2}L} d\xi e^{+L\kappa\xi} e^{-Q(\xi)} = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\xi e^{i\kappa\xi} e^{-Q(\xi)}. \quad (69)$$

Finally, replacing the sum by an integral in (67) we obtain

$$Q(\xi) = \frac{\lambda^2}{2\pi^2}\int_0^{\infty} dk \frac{1 - \cos k\xi}{k} |v(k)|^2. \quad (70)$$

[3] See, for example, V. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, (University of California Press, Berkeley and Los Angeles, 1958), especially P. 176 ff. See also M. Kac, *Probability and Related Topics in Physical Sciences*, (Interscience Publishers, London and New York, 1959), p. 60 ff.

[4] The formula given in Grenander and Szegö, (reference 3) contains K_L^* instead of K_{-L} as given in (56). I am indebted to Professor M. Kac for pointing out to me that if $f(\theta)$ is complex, rather than real as Grenander and Szegö assume, this simple change is all that is necessary.

We cannot go further in the evaluation of $Q(\xi)$ without some further information on the potential. However, the nature of the discontinuity at the F.S. can be investigated.

We may write

$$\bar{n}_k = \frac{2\pi}{L}\sum_{k'<k_F} F(k-k') = \int_{k-k_F}^{\infty} d\kappa F(\kappa) \qquad (71)$$
$$= \int_0^{\infty} d\kappa F(\kappa) + \int_{k-k_F}^{0} d\kappa F(\kappa).$$

The first term of (71) is a constant. To study the behavior of $\bar{n}_k$ near the F.S. ($k \cong k_F$) we therefore need $F(\kappa)$ only for very small κ. This in turn, from (69), requires the behavior of $Q(\xi)$ for large ξ. Since $Q(\xi)$ is an even function of ξ, we consider it for large positive ξ. We have

$$\frac{\partial Q(\xi)}{\partial \xi} = \frac{\lambda^2}{2\pi^2}\int_0^{\infty} dk \sin k\xi \, |v(k)|^2$$
$$= \frac{\lambda^2}{2\pi^2}\left[|v(0)|^2 \frac{1}{\xi} + O\left(\frac{1}{\xi^2}\right)\right], \qquad (72)$$

by successive integrations by parts. Integrating, we get

$$Q(\xi) = (\lambda^2/2\pi^2)[|v(0)|^2 \log \xi + C + O(1/\xi^2)], \qquad (73)$$

where C is a constant which is in principle calculable from the potential. This may be written in the following way:

$$Q(\xi) = \alpha \log (\xi/a)^2 + O(1/\xi^2), \qquad (74)$$

where

$$\alpha \equiv (\lambda^2/4\pi^2)\,|v(0)|^2, \qquad (75)$$

and a is a constant with the dimensions of a length, which depends only on the shape of the potential (it is a measure of its range). Inserting (74) into (69) we obtain, for $|\kappa a| \ll 1$,

$$F(\kappa) = \frac{1}{\pi}\int_0^{\infty} d\xi \frac{\cos \kappa\xi}{(\xi/a)^{2\alpha}}$$
$$= \frac{\Gamma(1-2\alpha)\sin \pi\alpha}{\pi}\frac{a^{2\alpha}}{|\kappa|^{1-2\alpha}}. \qquad (76)$$

Thus we obtain, for $|k - k_F|\, a \ll 1$,

$$\int_{k-k_F}^{0} d\kappa F(\kappa) = -\frac{\Gamma(1-2\alpha)\sin \pi\alpha}{2\pi\alpha}$$
$$\times |(k-k_F)a|^{2\alpha}\, \sigma(k-k_F), \qquad (77)$$

where

$$\sigma(x) = 1, \quad x > 0$$
$$= -1, \quad x < 0.$$

Therefore we see that there is, for $\alpha \neq 0$, no discontinuity at the F.S. (because the factor $|(k-k_F)a|^{2\alpha}$ vanishes there) though the slope is infinite at this point. On the other hand, if $\alpha = 0$, (77) behaves like $-\frac{1}{2}\sigma(k-k_F)$, which just gives the usual discontinuity at the F.S. Thus, in this model, the smallest amount of interaction *always* destroys the discontinuity of $\bar{n}_k$ at the F.S.

The behavior of $\bar{n}_k$ for large k[i.e., $(k-k_F)a \gg 1$] is also not difficult to obtain. From (71) we need $F(\kappa)$ for large κ, which is the same as knowing $Q(\xi)$ for small ξ. From (70) this may be obtained by expanding

$$Q(\xi) = \frac{\lambda^2}{2\pi^2}\frac{1}{2}\int_0^{\infty} dk k |v(k)|^2 \xi^2 + \cdots, \qquad (78)$$

as long as the integral converges, which we shall assume. Writing this as

$$Q(\xi) \cong \tfrac{1}{2}\xi^2/b^2,$$
$$b^2 \equiv \frac{\lambda^2}{2\pi^2}\int_0^{\infty} dk \cdot k \cdot |v(k)|^2, \qquad (79)$$

we obtain

$$F(\kappa) = b/(2\pi)^{\frac{1}{2}} e^{-b^2\kappa^2/2}. \qquad (80)$$

Therefore, for large k, we have

$$\bar{n}_k \cong \frac{b}{(2\pi)^{\frac{1}{2}}}\int_k^{\infty} d\kappa e^{-\kappa^2 b^2/2} \cong \frac{1}{(2\pi)^{\frac{1}{2}}}\frac{1}{kb} e^{-k^2b^2/2}. \qquad (81)$$

Therefore the momementum distribution decreases exponentially for large k.

For k close to the origin we may write

$$\bar{n}_k = \int_{k-k_F}^{\infty} F(\kappa)\, d\kappa$$
$$= \int_{-\infty}^{\infty} F(\kappa)\, d\kappa + \int_{-(k\ -k)}^{-\infty} F(\kappa)\, d\kappa. \qquad (82)$$

From (69),

$$\int_{-\infty}^{\infty} d\kappa F(\kappa) = \int_{-\infty}^{\infty} d\xi\, \delta(\xi) e^{-Q(\xi)} = e^{-Q(0)} = 1. \qquad (83)$$

Further, $F(\kappa)$ is an even function of κ. Thus

$$\bar{n}_k = 1 - \int_{k_F-k}^{\infty} F(\kappa)\, d\kappa = 1 - \bar{n}_{2k_F-k}, \qquad (84)$$

$$\bar{n}_0 = 1 - \bar{n}_{2k_F}. \qquad (85)$$

Therefore $\bar{n}_0 < 1$. If the interaction is such that $k = 2k_F$ is already in the asymptotic region for large k, then $\bar{n}_{2k_F}$ is exponentially small, and $\bar{n}_0$ is very close to unity.

Finally, we should like to conclude this section

with a remark about the case where $V(x) = \delta(x)$, the Dirac δ function. In this case $v(k)$ is a constant, so that (70) diverges logarithmically.

If one regards the δ function as the limit of a smooth function [a very convenient choice, with which one can calculate explicitly, is $v(k) = e^{-|k|a/2}$, letting a approach zero in the final answer], it is easy to see that the result is simply $\bar{n}_k = \frac{1}{2}$. The anomalous behavior of the δ-function case is not surprising as it looks at first. Since the particle mass is zero and λ (as may easily be verified) is dimensionless, the only length which can come into the problem is the mean distance between particles or, equivalently k_F^{-1}. However, from (54), k_F does not enter into $F(\kappa)$, so that $\bar{n}_k$ is a function of $k - k_F$ alone, which must be dimensionless. One such example is the unperturbed distribution, which depends only on whether $|k| > k_F$ or not. Another is a constant, which is what we actually obtain for the δ-function potential. The physical origin of this distribution which extends to infinite k, is that the high fourier components of the δ function produce infinitely many pairs, so that infinitely many particles are present.

IV. COMPARISON WITH PERTURBATION THEORY

According to the general formulas[2] the momentum distribution in the ground state is given by

$$\bar{n}_k = \frac{1}{2\pi i}\int_{\mu-i\infty}^{\mu+i\infty} d\zeta \, \frac{e^{\zeta 0^+}}{\zeta - \epsilon_k - G_k(\zeta)}, \tag{86}$$

where $G_k(\zeta)$ is the proper self-energy part of the particle propagator. In this formalism one should calculate the correct propagator at finite temperature (including "anomalous" diagrams) and also use the correct chemical potential μ. It was found there that if the F.S. does not distort (spherical case) this is the same as using ordinary Goldstone perturbation theory (no anamolous diagrams) and the the unperturbed chemical potential. We shall assume that this is also the case here, there being nothing comparable to F.S. distortion in one dimension. Then we replace μ by $v_0 k_F$ and take for $G_k(\zeta)$ the lowest nonvanishing contribution. This is second order. A straightforward calculation yields, for $k > 0$,

$$G(\zeta) = -\frac{1}{2}\left(\frac{\lambda}{\pi}\right)^2 v_0\left(\int_{|k'|}^{\infty} + \int_{-\infty}^{-|k'|}\right) d\kappa \times \frac{|v[\frac{1}{2}(\kappa + |k'|)]|^2 \, |\frac{1}{2}(\kappa + |k'|)|}{z + \kappa}, \tag{87}$$

where

$$k' = k_\perp - k_F, \qquad z_\perp = (\zeta - k_F v_0) \mid v_0.$$

This function is analytic in the cut z plane, the cuts extending from $-\infty$ to $-|k'|$ and from $|k'|$ to ∞.

If this is inserted in (86) (with μ replaced by $v_0 k_F$), the resulting integral is quite complicated to discuss, even in the neighborhood of $k = k_F$, for an arbitrary potential, and we shall limit ourselves to a special case.

Writing $z = x - i0^+$, we have

$$G_k(\zeta) = v_0[K_{k'}(x) + iJ_{k'}(x)]. \tag{88}$$

It is easy to see that by suitably deforming the contour in (86) we may write

$$\bar{n}_k = \frac{1}{2\pi i}\int_{-\infty}^{0} dx \times \left[\frac{1}{x - k' - K_{k'}(x) - iJ_{k'}(x)} - \text{c.c.}\right] \quad k' > 0 \tag{89}$$

$$= 1 - \frac{1}{2\pi i}\int_{0}^{\infty} dx \times \left\{\frac{1}{x - k' - K_{k'}(x) - iJ_{k'}(x)} - \text{c.c.}\right\} \quad k' < 0. \tag{90}$$

Now choosing

$$|v(\kappa)|^2 = 1 \quad |\kappa| < \tfrac{1}{2}q$$
$$= 0 \quad |\kappa| > \tfrac{1}{2}q, \tag{91}$$

one easily sees

$$K_{k'}(x) = \alpha\,|k'|\left[-2 + \left(1 - \frac{x}{|k'|}\right) \times \log\left|\frac{q^2 - (x - |k'|)^2}{x^2 - k'^2}\right|\right],$$
$$J_{k'}(x) = 0 \quad \text{unless} \quad -q + |k'| < x < -|k'|, \quad \text{or} \quad |k'| < x < q + |k'|$$
$$= \pi\alpha\,|x - |k'|| \quad \text{otherwise.} \tag{92}$$

We want to investigate $\bar{n}_k$ for small k'. It is not difficult, using (92), to show that, for small α and $|k'|$, $\bar{n}_k$ takes the form

$$\bar{n}_k = \tfrac{1}{2}\{1 - \sigma(k')/(1 - 2\alpha \log|k'a|)\}, \tag{93}$$

where $a = 1/q$.

This expression is, just as the exact expression, continuous at $k = k$, and has infinite slope there. In fact if we write

$$|k'a|^{2\alpha} = (e^{-2\alpha \log|k'a|})^{-1} = (1 - 2\alpha \log|k'a| + \cdots)^{-1},$$

forcing an expansion of the exact result (77) for

small α, we see that in this sense (93) agrees exactly with the exact answer to the order involved.

Thus, unlike the realistic three-dimensional case, perturbation theory predicts no discontinuity at the F.S. Since the exact answer behaves in the same way, perturbation theory (for the proper self-energy part) in this problem at least is a reliable guide to the behavior of $\bar{n}_k$.

V. RESPONSE TO EXTERNAL FIELDS

If one considers particles to have a charge e, we can induce currents to flow by applying an external field. It follows at once from the commutation relationships that

$$\dot{\rho} + \partial j/\partial x = 0, \tag{94}$$

where

$$\rho(x) = e\psi^+(x)\psi(x) = e[N_1(x) + N_2(x)], \tag{95}$$

$$j(x) = ev_0\psi^+(x)\sigma_3 \quad \psi(x) = ev_0[N_1(x) - N_2(x)], \tag{96}$$

$$\dot{\rho} \equiv i[H, \rho]. \tag{97}$$

This is clearly the equation of continuity of charge, and we can identify ρ and j with the charge and current densities, respectively.[5]

Suppose we couple to our system an external electric field described by a potential $\varphi(x, t)$. The interaction is described by a Hamiltonian H_{ext} given by

$$H_{\text{ext}} = \int_0^L \rho(x)\varphi(x, t)\, dx, \tag{98}$$

$$H_T = H + H_{\text{ext}}. \tag{99}$$

If we again make the canonical transformation (20),

$$\tilde{H}_T = e^{i\lambda S}H_T e^{-i\lambda S}, \tag{100}$$

we find, since S commutes with H_{ext},

$$\tilde{H}_T = H_0 + H_{\text{ext}}. \tag{101}$$

Therefore, for a static field, all the energy levels are identical with the noninteracting case. In particular, this means that the Kohn effect[6] (which predicts a logarithmic singularity in $|q - 2k_F|$ for the change in energy of the system in the presence of an external field of wavenumber q) is completely unaltered by the interaction, this, in spite of the fact that the behavior of $\bar{n}_k$ in the neighborhood of $k = k_F$ is profoundly altered.

If we calculate the linear response, (i.e., the current that flows to terms linear in the external field) by means of (say) the Kubo formula,[7] then one sees immediately that the result is the same as in the unperturbed case. Again this result is due to the fact that both the charge and current densities depend only on N_1 and N_2, which commute with S. This is also true if the external field couples to the current or when there are impurities present which act on the individual particles.

Finally, we may consider "positron annihilation" in this model.[8] Usually this is thought of as an effect which gives a direct experimental measurement of $\bar{n}_k$. In the one-dimensional case one cannot measure an angular correlation between the photons which come out. However, one can ask questions about the probability of one of them having a momentum between q and $q + dq$. We do not want to enter into a long discussion of the various possibilities here. We mention, however, that if one couples massless "photons" described by a scalar field ϕ having velocities $u_0(<v_0)$, via an effective interaction for pair annihilation,

$$H''' = g\int_0^L dx\rho(x)\phi^2(x); \tag{102}$$

then again only the unperturbed momentum distribution plays a role. However, if one takes more complicated couplings (depending for example on other bilinear expressions than ρ or j) one can get a large effect from the interaction.

Thus we see that although the momentum distribution is very much altered by the interaction in this model, it is by no means true that effects due to "particles at the Fermi Surface" are correspondingly altered. In other words, the naive association of the existence of a discontinuity in the momentum distribution, and the quasiparticle-like behavior of a weakly excited system of interacting fermions is shown to be unjustified for this model.

APPENDIX

We want to evaluate expressions of the following type:

$$I = (\Psi, \Lambda\Psi),$$
$$\Lambda \equiv \exp\left[i\int_0^L Q(y)\psi^+(y)\psi(y)\, dy\right], \tag{A1}$$

where $Q(y)$ is an ordinary function, and where Ψ

[5] In reality these definitions should be modified by the subtraction of infinite constants corresponding to the redefinition of the vacuum state as that with no holes and no elections. We imagine this done in what follows.

[6] W. Kohn, Phys. Rev. Letters, **2**, 393 (1959).

[7] R. Kubo, Can. J. Phys. **34**, 1274 (1956).

[8] See, for example, R. Ferrell, Rev. Mod. Phys. **28**, 308 (1956).

represents a wavefunction in which the single-particle states $n = 1, 2, \cdots, M$ are occupied. If we write

$$\psi(y) = \sum_n a_n \varphi_n(y), \tag{A2}$$

where the $\varphi_n(y)$ are a complete orthonormal set of single-particle states, then clearly

$$\Psi = a_1^+ \cdots a_M^+ \Psi_0. \tag{A3}$$

Ψ_0 is the unperturbed vacuum.

We may write (A1) as

$$I = (\Psi_0, a_M \cdots a_1 \Lambda a_1^+ \cdots a_M^+ \Psi_0). \tag{A4}$$

Writing

$$a_1 = \int_0^\infty dz_1 \varphi_1^*(z_1)\psi(z_1), \tag{A5}$$

we get, making use of (42),

$$\begin{aligned} a_1\Lambda &= \int dz_1 \varphi_1^*(z_1)\psi(z_1)\Lambda \\ &= \Lambda \int dz_1 \varphi_1^*(z_1) e^{iQ(z_1)}\psi(z_1). \end{aligned} \tag{A6}$$

Therefore, (A4) becomes

$$\begin{aligned} I = \Big(\Psi_0, \Lambda \int d^M z \Big(\prod_{n=1}^M \varphi_n^*(z_n) e^{iQ(z_n)}\Big) \\ \times \psi(z_M) \cdots \psi(z_1) a_1^+ \cdots a_M^+ \mid \Psi_0). \end{aligned} \tag{A7}$$

Since Ψ_0 is the unperturbed vacuum,

$$\Lambda \Psi_0 = \Psi_0, \tag{A8}$$

so that (A7) becomes

$$\begin{aligned} I &= \int d^M z \Big(\prod_{n=1}^M \varphi_n^*(z_n) e^{iQ(z_n)}\Big) \\ &\times (\Psi_0, \psi(z_M) \cdots \psi(z_1) a_1^+ \cdots a_M^+ \Psi_0) \\ &= \int d^M z\, d^M z' \Big(\prod_{n=1}^M \varphi_n^*(z_n)\varphi_n(z_n') e^{iQ(z_n')}\Big) \\ &\times (\Psi_0 \,|\psi(z_M) \cdots \psi(z_1)\psi^+(z_1') \cdots \psi^+(z_M')|\, \Psi_0). \end{aligned} \tag{A9, A10}$$

The expectation value in (A10) is a familiar one in the many-body problem. It can be obtained by taking the sum of the products of the corresponding expectation value for all possible ψ, ψ^+ pairs. The sign of each term is given by a plus if the permutation necessary to bring them to the required position is even, a minus if it is odd. Clearly then

$$\begin{aligned} I = \int d^M z\, d^M z' \sum_P (-)^P P\Big(\prod_1^M \varphi_n^*(z_n)\varphi_n(z_n') e^{iQ(z_n)}\Big) \\ \times (\Psi_0 \,|\psi(z_1)\psi^+(z_1')|\, \Psi_0) \cdots (\Psi_0 \,|\psi(z_M)\psi^+(z_M')|\, \Psi_0). \end{aligned} \tag{A11}$$

The sum on P is over all possible permutations of the variables. Now

$$\begin{aligned} &(\Psi_0 \,|\psi(z_1)\psi^+(z_1')|\, \Psi_0) \\ &\quad = (\Psi_0 \,|\psi(z_1)\psi^+(z_1') + \psi^+(z_1')\psi(z_1)|\, \Psi_0) \\ &\quad = \delta(z_1 - z_1')(\Psi_0, \Psi_0) = \delta(z_1 - z_1'), \end{aligned} \tag{A12}$$

$$\begin{aligned} I = \int d^M z \sum_P (-)^P \varphi_1^*(z_1) e^{iQ(z_1)} \cdots \varphi_M^*(z_M) e^{iQ(z_m)} \\ \times \varphi_{i_1}(z_1) \cdots \varphi_i\ (z_M), \end{aligned} \tag{A13}$$

where

$$P(1, 2, \cdots, M) = (i_1, i_2, \cdots, i_M),$$

or

$$I = \sum_P (-)^P (\varphi_1, e^{iQ(z)}\varphi_{i_1}) \cdots (\varphi_M, e^{iQ(z)}\varphi_i\ ). \tag{A14}$$

This, however, is just the definition of the determinant of the matrix g_{nn}, where

$$g_{nn'} = (\varphi_n, e^{iQ(z)}\varphi_{n'}). \tag{A15}$$

Therefore,

$$I = \mathrm{Det}\,(g).$$

If we take for the φ_n plane wave states, we get just the result used in the text.

Incidently, if one does this in configuration space and uses determinental wavefunctions, this becomes a well-known theorem about the integral over products of determinants.

ACKNOWLEDGMENTS

I should like to express my appreciation to the Brookhaven National Laboratory for its hospitality during the summer of 1962, when this work was initiated. I should also like to express my thanks to Professor W. Kohn for many valuable conversations relating to the general question of the Fermi surface, out of which this work originated.

Relativistic Kinetic Theory of a Simple Gas

WERNER ISRAEL
Mathematics Department, University of Alberta, Edmonton, Alberta, Canada
(Received 25 February 1963)

A consistent relativistic theory of transport processes in a simple gas is developed. The approach is the four-dimensional geometric one due to Synge. Scalar and vector eigenfunctions of the linearized collision operator are derived when the scattering cross section is a simple separable function of scattering angle and relative velocity (Maxwellian particles). The bulk viscosity and thermal conductivity are computed explicitly for this case.

1. INTRODUCTION

THE study of relativistic kinetic theory has recently received a new impetus, partly on account of developments in plasma physics.[1] During the last few years derivations of the relativistic form of Boltzmann's transport equation, with[2,3] or without[4,5] collision term, have been presented independently by several authors. However, the only applications so far considered have been concerned with the equilibrium state. Apart from Lee's[6] estimate of the thermal conductivity of a relativistic Fermi gas, the relativistic theory of transport processes seems to be an unexplored field. It is the aim of this paper to develop the elements of such a theory for a relativistic Boltzmann gas.

A "conventional" approach to relativistic kinetic theory (e.g., Clemmow and Willson[5]) is hampered by the burdensome calculations needed to check the Lorentz invariance of the formalism. To avoid these difficulties, we shall adopt the elegant four-dimensional geometrical point of view which has been consistently and fruitfully exploited by Synge.[4,7,8,9] This makes Lorentz invariance manifest at every stage by relating all quantities entering the formulas to geometrical objects in space–time. Tauber and Weinberg achieve the same effect by the use of a powerful but rather elaborate eight-dimensional formalism. The present approach is simpler and equally effective, at least for the simple applications we have in view.

Because of our different point of view, it has seemed worthwhile to develop the theory *ab initio*, even though in the early sections this entails some overlap with previous work.

The first part of the paper (Secs. 2–9) is concerned with the general theory of a simple gas, neglecting quantum effects and regarding gravitation and binary collisions as the only interactions. After some preparatory work, the Boltzmann collision equation is obtained (Sec. 4) and used to derive the conservation equations and the Boltzmann H theorem (Sec. 5). The equilibrium state is briefly considered in Sec. 6. Since this has been extensively studied elsewhere,[2,3,4] we present only enough of the theory to display a number of formulas needed later.

Sections 7, 8, and 9 are devoted to the study of a gas close to equilibrium. Expressions are derived for the heat-flow vector and the viscosity tensor. It is confirmed that the heat flow is determined, not by the temperature gradient, but by the gradient of *thermal potential*, defined as

$$[1 + c^{-2}\ (\text{chemical potential})]/(\text{temperature}).$$

(This result first emerged clearly in the phenomenological theory of Landau and Lifshitz,[10] although it is implicit in earlier work by Eckart.[11]) It is also shown that the relativistic gas possesses a bulk viscosity which vanishes only in the classical limit.

Finally, in the second part of the paper (Secs. 10–15), explicit expressions for thermal conductivity and bulk viscosity are obtained when the scattering cross section for collisions has a simple separable form ("relativistic Maxwellian particles").

[1] See the articles by O. Bunemann in *Plasma Physics*, edited by J. E. Drummond (McGraw-Hill Book Company, Inc., New York, 1961), Chaps. 7 and 10.
[2] G. E. Tauber and J. W. Weinberg, Phys. Rev. **86**, 621 (1952); **122**, 1342 (1961).
[3] N. A. Chernikov, Soviet Physics—Doklady **2**, 248 (1957); **5**, 764, 786 (1960); **7**, 397, 414, 428 (1962).
[4] J. L. Synge, *The Relativistic Gas* (North-Holland Publishing Company, Amsterdam, 1957).
[5] P. C. Clemmow and A. J. Willson, Proc. Cambridge Phil. Soc. **53**, 222 (1957).
[6] T. D. Lee, Astrophys. J. **111**, 625 (1950).
[7] J. L. Synge, Trans. Roy. Soc. Canada III **28**, 127 (1934).
[8] J. L. Synge, *Relativity: The Special Theory* (North-Holland Publishing Company, Amsterdam, 1956).
[9] J. L. Synge, *Relativity: The General Theory* (North-Holland Publishing Company, Amsterdam, 1960).
[10] L. Landau and E. Lifshitz, *Fluid Mechanics* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1959), p. 505.
[11] C. Eckart, Phys. Rev. **58**, 919 (1940).

PART I: GENERAL THEORY

2. 4-Momentum Space

Our attention will be confined throughout to a *simple gas*, i.e., an assemblage of material particles with a continuous distribution of velocities, all having the same proper mass m. The only interactions we shall deal with are gravitation, treated as a self-consistent background field, and elastic binary collisions at close encounters. The history of the assemblage is thus to be conceived as a network of timelike world lines in a Riemannian space–time. Each world line has a denumerable set of kinks, corresponding to collisions; between kinks the world line is a geodesic.

The unit tangent p^μ to a world line, satisfying

$$g_{\mu\nu}p^\mu p^\nu = -1, \tag{2.1}$$

may be regarded as a normalized 4-momentum or 4-velocity (the true 4-momentum is mcp^μ). [The metric is assumed to have signature $+2$, so that in a local Minkowskian (inertial) frame (x, y, z, ct), the metric tensor reduces to diag $(1, 1, 1, -1)$.] We define the *4-momentum space* at a world point x^μ to be a Galilean 4-space tangent to space–time at x^μ. Each curvilinear coordinate net in space–time induces in the tangent 4-momentum space a natural set of (in general, oblique) pseudocartesian axes, whose origin we take to be the point of tangency. We may now visualize the 4-momenta p^μ of particles at a given world point as line segments in the local 4-momentum space, emanating from the origin and terminating on the unit pseudosphere (2.1). (Cf. reference 4, Chap. I).

Employing a local Minkowskian frame, we can define polar coordinates (χ, θ, φ) on the unit pseudosphere by

$$p^\mu = (\sinh\chi \sin\theta\cos\varphi, \sinh\chi\sin\theta\sin\varphi, \sinh\chi\cos\theta, \cosh\chi). \tag{2.2}$$

The intrinsic metric of the pseudosphere is given by

$$dp_1^2 + dp_2^2 + dp_3^2 - dp_4^2 = d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta\, d\varphi^2),$$

so that the 3-dimensional element of area or solid angle is

$$d\omega = d\chi(\sinh\chi\, d\theta)(\sinh\chi\sin\theta\, d\varphi). \tag{2.3}$$

The projection of $d\omega$ onto the 3-flat of the Cartesian axes p_1, p_2, p_3 is

$$dp_1\, dp_2\, dp_3 = d\omega\cosh\chi. \tag{2.4}$$

The number χ, which measures the inclination of p^μ to the time axis, exemplifies what we shall call a pseudoangle. More generally, the pseudoangle ψ between p^μ and a unit timelike vector n^μ is defined by $\cosh\psi = |n_\mu p^\mu|$. In the local Minkowskian frame whose time axis is parallel to n^μ, $\cosh\psi$ reduces to p^4 or $(1 - v^2/c^2)^{-\frac{1}{2}}$, where v is the 3-velocity. We record a useful result called the *projection theorem* (reference 8, p. 275): *the projection of a flat 3-area S, with normal p^μ, onto the 3-flat normal to n^μ, is $S\cosh\psi$*. Equation (2.4) represents an obvious special case of this, since $d\omega$ is normal to the radial vector p^μ.

3. Dynamics of Collisions

We consider an elastic collision between two particles having the same rest-mass m. The initial momenta p_μ, $'p_\mu$ and the final momenta p^*_μ, $'p^*_\mu$ all satisfy normalization conditions of the form (2.1). Conservation of 4-momentum is expressed by

$$p_\mu + {}'p_\mu = p^*_\mu + {}'p^*_\mu. \tag{3.1}$$

Define the *relative 4-momenta* of the particles before and after collision to be the spacelike vectors

$$g_\mu = {}'p_\mu - p_\mu, \qquad g^*_\mu = {}'p^*_\mu - p^*_\mu. \tag{3.2}$$

It follows from (2.1) and (3.1) that these vectors have the same magnitude:

$$g \equiv (g_\mu g^\mu)^{\frac{1}{2}} = (g^*_\mu g^{\mu *})^{\frac{1}{2}} \equiv g^*. \tag{3.3}$$

The vector

$$\bar{p}_\mu \equiv (4 + g^2)^{-\frac{1}{2}}(p_\mu + {}'p_\mu) = \bar{p}^*_\mu \tag{3.4}$$

is tangent to the world line of the centroid. We have

$$\bar{p}_\alpha\bar{p}^\alpha = -1, \qquad g_\alpha\bar{p}^\alpha = g^*_\alpha\bar{p}^\alpha = 0, \tag{3.5}$$

so that g_α and g^*_α lie in the 3-flat in 4-momentum space normal to $\bar{p}_\mu$—the *centroidal 3-flat*. In the center-of-mass frame, g_α and g^*_α are thus purely spatial vectors, equal to the relative 3-momenta of the particles.

Let ϑ, ϵ and ϑ^*, ϵ^* be polar angles in the centroidal 3-flat which measure the orientation of g_α and g^*_α relative to any orthonormal triad fixed in the 3-flat. Such a triad might, for instance, be constructed by projecting the cartesian axes of p_1, p_2, p_3 onto the centroidal 3-flat and performing a Schmidt orthonormalization. Physically, the parameters mcg, ϑ, ϵ fix the magnitude and direction of relative 3-momentum in the center-of-mass frame. The momenta of the particles before collision are completely described *either* (i) by the six polar angles χ, θ, φ, $'\chi$, $'\theta$, $'\varphi$ of the unit vectors p_μ, $'p_\mu$ [see (2.2)];

or (ii) by the three polar angles $\tilde{\chi}$, $\tilde{\theta}$, $\tilde{\varphi}$ of $\tilde{p}_\mu$, together with g, ϑ, ϵ. Computing the Jacobian of the transformation between these two sets of parameters, we find (see Appendix)

$$d\omega\, d'\omega = (1 + \tfrac{1}{4}g^2)^{\frac{1}{2}}\, d\tilde{\omega} g^2\, dg \sin\vartheta\, d\vartheta\, d\epsilon. \tag{3.6}$$

Similarly, for the momenta after collision,

$$d\omega^*\, d'\omega^* = (1 + \tfrac{1}{4}g^{*2})^{\frac{1}{2}}\, d\tilde{\omega}^* g^{*2}\, dg^* \times \sin\vartheta^*\, d\vartheta^*\, d\epsilon^*. \tag{3.7}$$

Noting (3.3) and (3.4), we obtain *Liouville's theorem* in the form

$$d\omega\, d'\omega \sin\vartheta^*\, d\vartheta^*\, d\epsilon^* = d\omega^*\, d'\omega^* \sin\vartheta\, d\vartheta\, d\epsilon. \tag{3.8}$$

4. Distribution Function. Boltzmann's Collision Equation

Let dS be a flat 3-dimensional target at the world point x^μ with unit normal n^μ. The *indicator* of n^μ is

$$\epsilon(n) \equiv n_\mu n^\mu = \begin{cases} 1 & \text{for spacelike } n^\mu \\ -1 & \text{for timelike } n^\mu \end{cases}.$$

Consider the tube of parallel world lines with tangent p^μ which intersect dS. The normal section of this tube has 3-area $dS\,|n_\mu p^\mu|$. It is easily verified that the world lines cross in the positive or negative sense of the normal accordingly as $\epsilon(n) n_\mu p^\mu$ is positive or negative. Hence, for the *nett* number of world lines making a positive transit across dS, whose tangents lie within a solid angle $d\omega$ surrounding the vector p^μ, we may write

$$N(x, p)\epsilon(n) n_\mu p^\mu\, dS\, d\omega. \tag{4.1}$$

The *distribution function* $N(x^\mu, p_\nu)$ is a relativistic invariant, as is evident from the geometrical character of its definition (cf. reference 4, Chap. II).

To interpret (4.1) in familiar three-dimensional terms, we choose n^μ as time axis of a local Minkowskian frame. We then have, recalling (2.4),

$$dS = dx\, dy\, dz, \qquad -(n_\mu p^\mu)\, d\omega = dp_x\, dp_y\, dp_z.$$

Hence, (4.1) reduces to

$$N(x, p)\, dx\, dy\, dz\, dp_x\, dp_y\, dp_z,$$

giving the number of particles in the volume $dx\, dy\, dz$ having the indicated range of 3-momenta.

The nett number of world lines escaping from the infinitesimal 4-volume $d\tau$ in space–time bounded by a closed 3-space S, and associated with the range of 4-momenta $(p^\mu, d\omega)$, is

$$p^\mu\, d\omega \int_S N(x, p)\epsilon(n) n_\mu\, dS = p^\mu (\partial N(x, p)/\partial x^\mu)\, d\omega\, d\tau, \tag{4.2}$$

by Gauss's divergence theorem (reference 8, p. 276). In the absence of collisions and external forces (other than gravitation), the number of particles having a given 4-momentum stays constant as we follow the particles. We then obtain

$$p^\mu \partial_\mu N(x, p) = 0 \qquad (\partial_\mu \equiv \partial/\partial x^\mu).$$

To allow for collisions, let

$$W(p, {}'p; p^*, {}'p^*) N(x, p) \times N(x, {}'p)\, d\omega\, d'\omega\, d\omega^*\, d'\omega^*\, d\tau \tag{4.3}$$

be the number of binary collisions in the 4-volume $d\tau$ at x^μ between particles having initial momenta in the ranges $(p^\alpha, d\omega)$, $({}'p^\alpha, d'\omega)$, and final momenta in the ranges $(p^*_\alpha, d\omega^*)$, $({}'p^*_\alpha, d'\omega^*)$. The number of particles having a given range of 4-momenta $(p^\alpha, d\omega)$ is then depleted through collisions in $d\tau$ by

$$N(x, p)\, d\omega\, d\tau \iiint W(p, {}'p; p^*, {}'p^*) \times N(x, {}'p)\, d'\omega\, d\omega^*\, d'\omega^*, \tag{4.4}$$

and augmented by

$$d\omega\, d\tau \iiint W(p^*, {}'p^*; p, {}'p) N(x, p^*) \times N(x, {}'p^*)\, d\omega^*\, d'\omega^*\, d'\omega. \tag{4.5}$$

Equating (4.2) to the difference of (4.5) and (4.4), we obtain *Boltzmann's collision equation in the absence of (nongravitational) external forces*:

$$p^\mu \partial_\mu N(x, p) = \iiint W(p, {}'p; p^*, {}'p^*)[N(x, p^*) \times N(x, {}'p^*) - N(x, p)N(x, {}'p)]\, d'\omega\, d\omega^*\, d'\omega^*. \tag{4.6}$$

In writing down (4.6), we have used the symmetry property

$$W(p, {}'p; p^*, {}'p^*) = W(p^*, {}'p^*; p, {}'p), \tag{4.7}$$

which expresses the assumption of microscopic reversibility of the collisions. We note also the trivial symmetry

$$W(p, {}'p; p^*, {}'p^*) = W({}'p, p; {}'p^*, p^*), \tag{4.8}$$

which follows by inspection of (4.3). The properties (4.7), (4.8) of the probability amplitude [for the conservation laws, (4.8) alone suffices], together with its positive-definite character, are sufficient to establish all the general theorems (conservation laws, H theorem, etc.) of Secs. 5–9.

For the explicit calculations of Secs. 10–15 we shall postulate in addition that the interaction between two particles has axial symmetry about their

line of centers. We now express this as a condition on W. Observe first that, since only collisions satisfying (3.1)—or, equivalently $\bar{p}_\alpha = \bar{p}^*_\alpha$, $g = g^*$—can occur, W must have the form

$$W(p, 'p; p^*, 'p^*) = W_1 \delta_\omega(\bar{p}^*_\alpha - \bar{p}_\alpha)\delta(g^* - g).$$

Here, δ is the Dirac delta function, and δ_ω the 3-dimensional delta function on the pseudosphere $p_\alpha p^\alpha = -1$:

$$\int f(p^*_\mu)\ \delta_\omega\ (p^*_\alpha - p_\alpha)\ d\omega^* = f(p_\mu).$$

In a local Minkowskian frame, the function δ_ω may be expressed in terms of the Dirac delta function by

$$\delta_\omega(p^*_\alpha - p_\alpha) = -p_4\delta(p^*_1 - p_1)\delta(p^*_2 - p_2)\delta(p^*_3 - p_3).$$

Our assumption of axial symmetry is embodied in the statement that the scalar W_1 can depend only on the two scalar collision invariants g and Θ, where

$$\cos\Theta \equiv g^*_\alpha g^\alpha / g^2. \tag{4.9}$$

We can now write

$$W(p, 'p; p^*, 'p^*) = (1 + \tfrac{1}{4}g^2)^{-\frac{1}{2}} g^{-1}\sigma(g, \Theta) \times \delta_\omega(\bar{p}^*_\alpha - \bar{p}_\alpha)\delta(g^* - g). \tag{4.10}$$

Substituting (4.10) in (4.3), and noting (3.7), we find that [abbreviating $N(x, p) = N$, $N(x, 'p) = 'N$, etc.]

$$(N\ d\omega)('N\ d'\omega)g\sigma(g, \Theta)(\sin\vartheta^*\ d\vartheta^*\ d\epsilon^*)\ d\tau$$

represents the number of particles with relative 3-momentum mcg which are deflected through the angle Θ and scattered into the solid angle $\sin\vartheta^*\ d\vartheta^*\ d\epsilon^*$ in the center-of-mass frame. The scalar $\sigma(g, \Theta)$ is thus to be interpreted as the scattering cross section observed in the center-of-mass frame. Boltzmann's collision equation, expressed in terms of σ, reads

$$p^\mu \partial_\mu N(x, p) = \int_0^\pi \int_0^{2\pi} \int g\sigma(g, \Theta) \times ('N^*N^* - 'NN) \sin\vartheta^*\ d\vartheta^*\ d\epsilon^*\ d'\omega. \tag{4.11}$$

5. Conservation Laws. H Theorem

Let $\Psi(x, p)$ be an arbitrary tensor function of x^μ, p, and abbreviate $\Psi(x, p) = \Psi$, $\Psi(x, p^*) = \Psi^*$, etc. We define

$$J(\Psi) \equiv \iiiint W(p, 'p; p^*, 'p^*)\Psi(x, p) \times ('N^*N^* - 'NN)\ d\omega\ d'\omega\ d\omega^*\ d'\omega^* \tag{5.1}$$

$$= \frac{1}{4}\iiiint W(p, 'p; p^*, 'p^*)[\Psi + '\Psi - \Psi^* - '\Psi^*]('N^*N^* - 'NN)\ d\omega\ d'\omega\ d\omega^*\ d'\omega^*, \tag{5.2}$$

where we have used the symmetry properties (4.7) and (4.8). From (5.2) and (3.1),

$$J(\Psi) = 0 \quad \text{if} \quad \Psi = 1 \quad \text{or} \quad p^\nu. \tag{5.3}$$

Multiplication of the collision equation (4.6) by $\Psi(x, p)$ and integration yields

$$\int \Psi p^\mu\ \partial_\mu N(x, p)\ d\omega = J(\Psi). \tag{5.4}$$

The choices $\Psi = 1$, $\Psi = p^\nu$ now lead to the five *conservation laws*

$$M^\mu{}_{|\mu} = 0, \qquad T^{\mu\nu}{}_{|\nu} = 0. \tag{5.5}$$

Here, the vertical bar denotes covariant differentiation, and

$$M^\mu(x) \equiv mc \int N(x, p)p^\mu\ d\omega, \tag{5.6}$$

$$T^{\mu\nu}(x) \equiv mc^2 \int N(x, p)p^\mu p^\nu\ d\omega \tag{5.7}$$

define the *numerical flux vector* and the *stress energy–momentum tensor.*

The *entropy-flux vector* is defined by

$$S^\mu(x) \equiv -k \int N(x, p)\ \log\ [AN(x, p)]p^\mu\ d\omega, \tag{5.8}$$

where k is Boltzmann's constant, and A a constant having the dimensions of N^{-1}. Using (5.4) and (5.2) with $\Psi = \log N(x, p)$, we find

$$S^\mu{}_{|\mu} = \tfrac{1}{4}k \iiiint W(p, 'p; p^*, 'p^*) \times [\log ('N^*N^*) - \log ('NN)] \times ('N^*N^* - 'NN)\ d\omega\ d'\omega\ d\omega^*\ d'\omega^* \tag{5.9}$$

$$\geq 0, \tag{5.10}$$

which expresses *Boltzmann's H theorem.*

6. Equilibrium

The equilibrium state is characterized by $S^\mu{}_{|\mu} = 0$, which, by (5.9), holds if and only if

$$\log N + \log 'N = \log N^* + \log 'N^*.$$

Since 1 and p^μ are essentially the only additive collision invariants, $\log N$ must be a linear combination

of them. Writing $N_0(x, p)$ for the equilibrium distribution, we therefore have

$$N_0(x, p) = \alpha(x) \exp [\beta_\mu(x)p^\mu]. \quad (6.1)$$

In order that $\int N_0\, d\omega$ be finite, it is necessary that $\beta_\mu(x)$ be timelike. We set

$$\beta^2 = -\beta_\mu(x)\beta^\mu(x), \qquad \beta_\mu(x) = c^{-1}\beta u_\mu(x),$$
$$u_\mu u^\mu = -c^2. \quad (6.2)$$

The collision equation (4.6) reduces to

$$p^\mu \partial_\mu N_0(x, p) = 0. \quad (6.3)$$

Substituting (6.1) into (6.3) yields

$$p^\mu \partial_\mu \log \alpha(x) + p^\mu p^\nu \beta_{\mu|\nu} = 0,$$

an identity for arbitrary unit, timelike vectors p^μ. We deduce

$$\alpha(x) = \alpha = \text{const}, \qquad \beta_{\mu|\nu} + \beta_{\nu|\mu} = 0. \quad (6.4)$$

Anticipating the fact that u^μ is the macroscopic 4-velocity of the gas, we see that *the most general motion of a gas in thermal equilibrium is a Killing rigid motion or "group motion."*[12] The relations

$$\dot{u}_\mu \equiv u_{\mu|\nu}u^\nu = \partial_\mu \log \beta(x), \quad u^\mu \partial_\mu \log \beta(x) = 0, \quad (6.5)$$

follow immediately from (6.4) and (6.2).

For a more detailed discussion of thermal equilibrium in a gravitational field, we refer to reference 2. From this point onwards we neglect gravitational effects and consider space–time to be flat. The general solution of (6.4) in Minkowskian coordinates is then

$$\beta_\mu(x) = \omega_{\mu\nu}x^\nu + \beta_\mu \qquad (\omega_{\mu\nu} = -\omega_{\nu\mu}), \quad (6.6)$$

where $\omega_{\mu\nu}$ and β_μ are constants. We shall assume further that $\omega_{\mu\nu} = 0$, thus excluding rigid rotations. The equilibrium distribution function now takes the simple form

$$N_0(p) = \alpha \exp [\beta_\mu p^\mu]. \quad (6.7)$$

The numerical flux M_0^μ and the energy tensor $T_0^{\mu\nu}$ for equilibrium are most easily obtained from *the generating function*

$$Z_0(\alpha, \beta_\mu) \equiv \int N_0(p)\, d\omega = \alpha \int \exp [\beta_\mu p^\mu]\, d\omega. \quad (6.8)$$

We have, in fact,

$$M_0^\mu = mc\partial Z_0/\partial\beta_\mu, \qquad T_0^{\mu\nu} = mc^2\partial^2 Z_0/\partial\beta_\mu\partial\beta_\nu. \quad (6.9)$$

[12] Cf. G. Salzman and A. H. Taub, Phys. Rev. **95**, 1659 (1954). The above theorem appears to have been first enunciated by Tauber and Weinberg, reference 2, p. 1358. Its classical counterpart is given by R. H. Fowler, *Statistical Mechanics* (Cambridge University Press, New York, 1929), 1st ed., p. 430.

The integral in (6.8) is readily evaluated by choosing polar coordinates (2.2) in 4-momentum space with the axis $\chi = 0$ parallel to u^μ:

$$Z_0(\alpha, \beta_\mu) = \alpha \int_0^\infty \int_0^\pi \int_0^{2\pi} e^{-\beta \cosh \chi} \times \sinh^2 \chi\, d\chi \sin\theta\, d\theta\, d\varphi$$
$$= 4\pi\alpha K_1(\beta)/\beta. \quad (6.10)$$

We recall the definition of the modified Bessel function of the second kind:

$$K_n(\beta) = \frac{\beta^n}{1\cdot 3 \cdots (2n-1)} \times \int_0^\infty e^{-\beta \cosh \chi} \sinh^{2n} \chi\, d\chi. \quad (6.11)$$

Noting

$$(d/d\beta)[\beta^{-n}K_n(\beta)] = -\beta^{-n}K_{n+1}(\beta), \quad (6.12)$$
$$\partial\beta/\partial\beta_\mu = -c^{-1}u^\mu, \qquad \partial\beta^\nu/\partial\beta_\nu = g_{\mu\nu},$$

we easily obtain from (6.9),

$$M_0^\mu = \rho u^\mu, \quad (6.13)$$
$$T_0^{\mu\nu} = (\mu + c^{-2}P)u^\mu u^\nu + Pg^{\mu\nu} = \mu u^\mu u^\nu + P\Delta^{\mu\nu}, \quad (6.14)$$

where

$$\Delta^{\mu\nu} \equiv g^{\mu\nu} + c^{-2}u^\mu u^\nu$$
$$= \text{diag } (1, 1, 1, 0) \text{ in proper coordinates;}$$
$$\rho = 4\pi m\alpha K_2(\beta)/\beta, \quad P/c^2 = 4\pi m\alpha K_2(\beta)/\beta^2, \quad (6.15)$$
$$\mu + c^{-2}P = 4\pi m\alpha K_3(\beta)/\beta. \quad (6.16)$$

Formulas (6.15) and (6.16) for a relativistic Boltzmann gas were first given by Jüttner.[13] The interpretation of u^μ as the hydrodynamical 4-velocity, and of P, μc^2, and ρ as the pressure, energy density, and density of proper mass measured in the rest frame, are obvious from (6.13) and (6.14). Also, comparing the relation

$$P/c^2 = \rho/\beta \quad (6.17)$$

with the classical gas law $P = (k/m)\rho T$, we infer

$$\beta = mc^2/(kT). \quad (6.18)$$

This equation will be adopted as the definition of the relativistic temperature T.

[13] F. Jüttner, Ann. Phys. (Leipzig) **34**, 856 (1911); **35**, 145 (1911). Relativistic Bose and Fermi gases are treated in a later paper: F. Jüttner, Z. Phys. **47**, 542 (1928). See also S. Chandrasekhar, *Introduction to the Study of Stellar Structure* (1939) (The University of Chicago Press, Chicago, 1939), Chap. X; J. L. Synge, *The Relativistic Gas* (North-Holland Publishing Company, Amsterdam, 1957), Chap. IV.

Higher moments of the distribution function can be computed in similar fashion. If we define generally

$$U^{\lambda\mu\nu} \equiv mc^3 \int N(x, p)p^\lambda p^\mu p^\nu \, d\omega, \tag{6.19}$$

then we have in equilibrium,

$$\begin{aligned} U_0^{\lambda\mu\nu} &= mc^3\partial^3 Z_0/\partial\beta_\lambda\partial\beta_\mu\partial\beta_\nu \\ &= (\xi + 3c^{-2}\zeta)u^\lambda u^\mu u^\nu \\ &\qquad + \zeta(g^{\lambda\mu}u^\nu + g^{\nu\lambda}u^\mu + g^{\mu\nu}u^\lambda) \\ &= \xi u^\lambda u^\mu u^\nu + 3\zeta\Delta^{(\lambda\mu}u^{\nu)}, \end{aligned} \tag{6.20}$$

with

$$\begin{aligned} \xi + 3c^{-2}\zeta &= 4\pi m\alpha K_4(\beta)/\beta, \\ c^{-2}\zeta &= 4\pi m\alpha K_3(\beta)/\beta^2. \end{aligned} \tag{6.21}$$

Parentheses enclosing a group of indices indicate "mixing", e.g.,

$$\begin{aligned} A^{(\alpha\beta\gamma)} \equiv (1/3!)(&A^{\alpha\beta\gamma} + A^{\alpha\gamma\beta} \\ &+ A^{\beta\gamma\alpha} + A^{\beta\alpha\gamma} + A^{\gamma\alpha\beta} + A^{\gamma\beta\alpha}). \end{aligned}$$

The new functions ξ and ζ can be expressed in terms of parameters already introduced:

$$\xi = \rho + 3c^{-2}\eta P, \qquad \zeta = \eta P. \tag{6.22}$$

The relativistic *enthalpy* η is defined by

$$\eta = (\mu + c^{-2}P)/\rho = K_3(\beta)/K_2(\beta); \tag{6.23}$$

it corresponds to [1 + c^{-2} (classical enthalpy)].

For the entropy flow in equilibrium we have, from (5.6), (5.7), and (5.8),

$$\begin{aligned} S_0^\mu &= -k \int (\beta_\lambda p^\lambda + \log A\alpha)N_0 p^\mu \, d\omega \\ &= -(k/mc^2)\beta_\lambda T_0^{\lambda\mu} - (k/mc)\log(A\alpha)M_0^\mu \\ &= (k/mc)(\beta\mu - \rho\log A\alpha)u^\mu, \end{aligned} \tag{6.24}$$

where we have used (6.2), (6.13), and (6.14). The *entropy per unit mass* S, defined by

$$cS_0^\mu = \rho S u^\mu, \tag{6.25}$$

has the value

$$S = (k/m)(\eta\beta - \log\alpha) + \text{const.} \tag{6.26}$$

This equation enables us to express α in terms of familiar macroscopic parameters. Defining the *thermal potential* θ by

$$\theta = \frac{1}{c^2 T}\left(\frac{\mu c^2}{\rho} + \frac{P}{\rho} - TS\right) = \frac{\eta - c^{-2}TS}{T} \tag{6.27}$$

= [1 + c^{-2} (chemical potential)]/T,

we have

$$\log\alpha = (mc^2/k)\theta + \text{const.} \tag{6.28}$$

From (6.15) and (6.12),

$$\begin{aligned} \frac{1}{\rho}dP &= \frac{c^2}{\beta}\left[\frac{d\alpha}{\alpha} - \frac{K_3(\beta)}{K_2(\beta)}\,d\beta\right] \\ &= c^2 T\,d\theta - \eta c^2\,d\beta/\beta. \end{aligned} \tag{6.29}$$

Hence (6.27) yields the basic phenomenological relation

$$\begin{aligned} T\,dS &= c^2\,d\eta - (dP)/\rho \\ &= d(\mu c^2/\rho) + P\,d(1/\rho). \end{aligned} \tag{6.30}$$

For future reference, we record some formulas involving the specific heats. Defining

$$c_P = T(\partial S/\partial T)_P, \qquad c_v = T(\partial S/\partial T)_\rho, \tag{6.31}$$

we have from (6.30), (6.23), and (6.18),

$$c_P = c^2\,d\eta/dT = -(k/m)\beta^2\,d\eta/d\beta, \tag{6.32}$$

and similarly,

$$c_v = c^2\frac{d\eta}{dT} - \frac{d}{dT}\left(\frac{P}{\rho}\right) = -\frac{k}{m}\left(1 + \beta^2\frac{d\eta}{d\beta}\right). \tag{6.33}$$

We thus recover Robert Mayer's relation

$$c_P - c_v = k/m. \tag{6.34}$$

The ratio of specific heats $\gamma = c_P/c_v$ is given by

$$\gamma/(\gamma - 1) = -\beta^2\,d\eta/d\beta; \tag{6.35}$$

γ decreases monotonically from $\frac{5}{3}$ in the classical limit ($\beta \to \infty$) to $\frac{4}{3}$ in the ultrarelativistic limit ($\beta \to 0$). The ratio $\gamma/(\gamma - 1)$ has been tabulated by Chandrasekhar (reference 13, p. 397).

From (6.23) and the well-known formulas for the derivatives of the Bessel functions, one readily obtains

$$d\eta/d\beta = \eta^2 - (5/\beta)\eta - 1. \tag{6.36}$$

The following identities now follow easily from (6.17), (6.22), (6.23), (6.35), and (6.36):

$$\begin{aligned} &\rho\zeta - \mu P = P^2/c^2, \\ &\rho\xi - \mu^2 = c^{-4}P^2/(\gamma - 1), \\ &\mu\zeta - \xi P = (P^2/\beta c^2)[\eta\beta - \gamma/(\gamma - 1)]. \end{aligned} \tag{6.37}$$

7. Off-Equilibrium

The equilibrium state was distinguished by the existence of a unique, timelike vector u^μ which could be identified with the hydrodynamical 4-velocity. The various tensors characterizing the

macroscopic properties of the gas—the entropy flow S^μ, the numerical flux M^μ, the energy tensor $T^{\mu\nu}$, and all higher moments of the distribution function—were completely determined by u^μ and a pair of scalars. Away from equilibrium, this simple situation no longer prevails. While M^μ, $T^{\mu\nu}$, and S^μ are still unambiguously defined by Eqs. (5.6) to (5.8), they are no longer expressible in terms of a *single* 4-vector. The definition of a macroscopic 4-velocity now becomes largely a matter of arbitrary choice. Thus, we might define the mean motion of the gas in terms of the stream of particles (choose u^μ parallel to M^μ), or in terms of the flow of energy (choose u^μ as the timelike eigenvector of $T^{\mu\nu}$), or in any other way consistent with the equilibrium definition.

This indeterminacy of the macroscopic rest frame rests basically on the relativistic equivalence of mass and energy, and has no exact analogue in the classical theory of a simple gas. It means that *local* thermodynamic parameters such as the energy density μc^2 or the specific entropy S cannot be defined in an absolute manner. The conventional hydrodynamical description can be salvaged only if one assumes—as we do here—that the deviation from local equilibrium is small.

The theory here developed is primarily intended to apply to a gas close to equilibrium, and without circulatory motion, in flat space–time. It may be remarked, however, that the theory is also applicable, with no formal changes, to a gas in a weak, slowly varying, *external* gravitational field. The self-field of the gas is neglected; its consideration would require the introduction of Einstein's field equations.

We assume that the actual distribution function is approximated at each world point by a local relativistic Maxwellian distribution

$$N_0(x, p) = \alpha(x) \exp [\beta_\mu(x) p^\mu], \tag{7.1}$$

in which the five functions α, β_μ are disposable parameters. We set

$$N(x, p) = N_0(x, p)[1 + f(x, p)], \tag{7.2}$$

where f, which measures deviation from the local Maxwellian distribution, is assumed small compared with unity; we neglect quantities of order f^2. It will also be assumed that the spatio-temporal variation $\partial_\mu f$ over macroscopic distances is small compared with f, and can be neglected.

Substituting (7.2) into (5.6), (5.7), and (5.8), we obtain

$$\begin{aligned} M^\mu(x) &= M_0^\mu(x) + M_1^\mu(x), \\ T^{\mu\nu}(x) &= T_0^{\mu\nu}(x) + T_1^{\mu\nu}(x), \\ S^\mu(x) &= S_0^\mu(x) + S_1^\mu(x), \end{aligned} \tag{7.3}$$

where M_0^μ, $T_0^{\mu\nu}$, and S_0^μ are associated with the Maxwellian distribution function $N_0(x, p)$. They are given by the formulas of Sec. 6 [Eqs. (6.13) *et seq.*], but now the 4-velocity and all the thermodynamic parameters have to be regarded as functions of x^μ. The first-order correction terms are

$$M_1^\mu(x) = mc \int N_0(x, p) f p^\mu \, d\omega, \tag{7.4}$$

$$T_1^{\mu\nu}(x) = mc^2 \int N_0(x, p) f p^\mu p^\nu \, d\omega, \tag{7.5}$$

$$\begin{aligned} S_1^\mu(x) &= -k \int N_0(\log N_0 + \text{const}) f p^\mu \, d\omega \\ &= -(c\theta + \text{const}) M_1^\mu - (cT)^{-1} u_\lambda T_1^{\lambda\mu}; \end{aligned} \tag{7.6}$$

they embrace the various transport effects.

So far, we have not specified precisely how the local Maxwellian distribution N_0 is to be fitted to the actual distribution function N. It will be assumed that the five parameters α, β_μ have been chosen to satisfy the two conditions

$$M_1^\mu u_\mu = 0, \qquad T_1^{\mu\nu} u_\mu u_\nu = 0, \tag{7.7}$$

of which

$$S_1^\mu u_\mu = 0 \tag{7.8}$$

is a consequence. This leaves three degrees of freedom (essentially the three independent components of u_μ), which we shall not need to pin down further. [The conditions (7.7) could, for instance, be supplemented by postulating

either $\quad M_1^\mu \Delta_\mu^\lambda = 0$, or $\quad T_1^{\mu\nu} \Delta_\mu^\lambda u_\nu = 0$.

In combination with (7.7) this implies

either $\quad M^\mu = \rho u^\mu$, or $\quad T^{\mu\nu} u_\nu = -(\mu c^2) u^\mu$,

and one other condition. If we impose the first of these restrictions, we are driven on the path corresponding to Eckart's phenomenological theory; the second choice corresponds to the Landau–Lifshitz formalism. There are, of course, many other possibilities].

8. The Linearized Boltzmann Equation

Inserting (7.2) into (4.6), and retaining only terms of first order in f, leads to the *linearized Boltzmann equation*

$$p^\mu \partial_\mu N_0(x, p) = -N_0(x, p) \mathfrak{L}[f], \tag{8.1}$$

where the linear integral operator $\mathfrak{L}$ has the form

$$\mathfrak{L}[f] \equiv \iiint W(p, {}'p; p^*, {}'p^*) \times {}'N_0 \, \delta(f) \, d'\omega \, d\omega^* \, d'\omega^*, \tag{8.2}$$

$$\delta(f) \equiv f(x,p) + f(x,'p) - f(x,p^*) - f(x,'p^*). \quad (8.3)$$

For explicit calculations, the form [cf. (4.10), (4.11)]

$$\mathfrak{L}[f] = \int_0^{\pi} \int_0^{2\pi} \int g\sigma(g,\Theta)\, 'N_0\, \delta(f) \times \sin\vartheta^*\, d\vartheta^*\, d\epsilon^*\, d'\omega, \quad (8.2')$$

which expresses the linearized collision operator in terms of the scattering cross section σ, is more convenient.

It will be convenient to define the *inner product* of any two functions $F(x,p)$, $G(x,p)$ by

$$(F, G) \equiv \int N_0(x,p)G(x,p)\mathfrak{L}[F]\, d\omega \quad (8.4)$$

$$= \frac{1}{4}\iiiint WN_0\, 'N_0 \times \delta(F)\, \delta(G)\, d\omega\, d'\omega\, d\omega^*\, d'\omega^* \quad (8.5)$$

$$= (G, F). \quad (8.6)$$

From (8.5) we have

$$(f, f) \geq 0, \quad (8.7)$$

with equality if and only if $\delta(f) = 0$. Thus the only independent solutions of the homogeneous integral equation $\mathfrak{L}[f] = 0$ are the collision invariants 1, p^μ. The positive-definite property (8.7) obviously extends to the "square" of any *spacelike tensor* (i.e., a tensor all of whose components involving an index 4 vanish in a suitable frame of reference):

$$(\Phi^{\lambda_1\cdots\lambda_n}, \Phi_{\lambda_1\cdots\lambda_n}) \geq 0 \quad \text{for spacelike} \quad \Phi. \quad (8.8)$$

The linearized Boltzmann equation has to be solved for $f(x,p)$, given $\partial_\mu N_0$, i.e., the space–time gradients of α, β_μ. Not all these gradients can be freely assigned; they are subject to the five conditions

$$M_0{}^{\mu}{}_{|\mu} = 0, \qquad T_0{}^{\mu\nu}{}_{|\nu} = 0, \quad (8.9)$$

which have

$$S_0{}^{\mu}{}_{|\mu} = 0 \quad (8.10)$$

as a consequence.

Equations (8.9) are simply another form of the othogonality relations

$$(f, 1) = \int N_0\mathfrak{L}[f]\, d\omega = 0, \qquad (f, p^\mu) = \int N_0\mathfrak{L}[f]p^\mu\, d\omega = 0, \quad (8.11)$$

which follow from (8.5), and express the condition that there is a five-parameter family of solutions of the inhomogeneous integral equation (8.1). The conditions (8.9) can also be deduced directly from (5.5), recalling that $\partial_\mu f$ is assumed negligible. Similarly, (8.10) follows directly if we observe that (5.9) can be written correctly to the *second* order in f:

$$S^{\mu}{}_{|\mu} = k(f, f). \quad (8.11')$$

Following the classical approach of Chapman and Enskog, we proceed to eliminate the five time derivatives

$$d\alpha/d\tau \equiv \dot\alpha \equiv u^\mu\partial_\mu\alpha, \qquad d\beta_\mu/d\tau \equiv \dot\beta_\mu \equiv u^\lambda\beta_{\mu|\lambda}$$

from the left-hand side of (8.1) by use of (8.9). The equations

$$M_0{}^{\mu}{}_{|\mu} = \Delta_{\lambda\mu}T_0{}^{\lambda\nu}{}_{|\nu} = 0 \quad \text{yield}$$

$$\dot\rho = -\rho u^{\alpha}{}_{|\alpha}, \quad (8.12)$$

$$\dot u_\mu = -(\mu + c^{-2}P)^{-1}\Delta_\mu^\lambda\partial_\lambda P. \quad (8.13)$$

From (6.25), (8.10), and (8.12), it follows that entropy is conserved along the streamlines:

$$\dot S = 0. \quad (8.14)$$

Hence, noting (6.30) and (6.17),

$$\dot\eta = \left(\frac{\partial\eta}{\partial P}\right)_S \dot P = \frac{1}{\rho}\frac{d}{d\tau}\left(\frac{\rho}{\beta}\right) = -(\dot\beta/\beta^2) - \beta^{-1}u^{\alpha}{}_{|\alpha}.$$

On the other hand, from (6.35),

$$\dot\eta = -\frac{\gamma\dot\beta}{(\gamma-1)\beta^2}.$$

Solving for $\dot\beta$ yields

$$\dot\beta = \beta(\gamma-1)u^{\alpha}{}_{|\alpha}, \quad (8.15)$$

and hence, by (6.26),

$$\frac{d}{d\tau}(\log\alpha) = \frac{mc^2}{k}\,\dot\theta = \frac{d}{d\tau}\left(\eta\beta - \frac{mS}{k}\right) = [\eta\beta(\gamma-1) - \gamma]u^{\alpha}{}_{|\alpha}. \quad (8.16)$$

Equations (8.13), (8.15), and (8.16) furnish the time derivatives of the five basic parameters in terms of the purely spatial derivatives $\Delta_\mu^\lambda\,\partial_\lambda P$ and $u^{\alpha}{}_{|\alpha} = \Delta^{\lambda\mu}u_{\lambda|\mu}$. Substituting into the left-hand side of (8.1), we obtain

$$\begin{aligned}-\mathfrak{L}[f] &= p^\mu[\Delta_\mu^\lambda\partial_\lambda\log N_0 - c^{-2}u_\mu(d/d\tau)\log N_0] \\ &= (mc^2/k)(\partial_\lambda\theta)\Delta_\mu^\lambda p^\mu + c^{-1}\beta u_{\mu|\lambda}\Delta_\nu^\lambda p^\mu p^\nu \\ &\quad + c^{-3}\beta[c^2\beta^{-1}\partial_\lambda\beta + (\mu + c^{-2}P)^{-1}\partial_\lambda P]\Delta_\mu^\lambda u_\nu p^\mu p^\nu \\ &\quad - c^{-2}[\eta\beta(\gamma-1) - \gamma]u^{\alpha}{}_{|\alpha}u_\mu p^\mu \\ &\quad - c^{-3}\beta(\gamma-1)u^{\alpha}{}_{|\alpha}u_\mu u_\nu p^\mu p^\nu.\end{aligned}$$

We infer from (6.29) that

$$c^2\beta^{-1}\partial_\lambda\beta + (\mu + c^{-2}P)^{-1}\partial_\lambda P = (mc^4/k\eta\beta)\partial_\lambda\theta,$$

and we adopt the abbreviations

$$\epsilon_{\mu\nu} \equiv \Delta^\beta_{(\nu}u_{\mu)\,|\beta} - \tfrac{1}{3}\Delta_{\mu\nu}u^\alpha{}_{|\alpha}$$
$$= \tfrac{1}{2}\Delta^\alpha_\mu\Delta^\beta_\nu(u_{\alpha|\beta} + u_{\beta|\alpha} - \tfrac{2}{3}\Delta_{\alpha\beta}u^\lambda{}_{|\lambda}), \quad (8.17)$$

$$\mathbf{p}^2 \equiv \Delta_{\mu\nu}p^\mu p^\nu, \text{ so that } c^{-1}u_\mu p^\mu = -(1+\mathbf{p}^2)^{\frac{1}{2}}, \quad (8.18)$$

$$\langle p^\mu p^\nu\rangle \equiv \Delta^\mu_\alpha\Delta^\nu_\beta p^\alpha p^\beta - \tfrac{1}{3}\Delta^{\mu\nu}\mathbf{p}^2. \quad (8.19)$$

Then the above equation reduces to

$$\mathfrak{L}[f] = c^{-1}B(\mathbf{p}^2,\beta)u^\alpha{}_{|\alpha} + (mc^2/k)[\eta^{-1}(1+\mathbf{p}^2)^{\frac{1}{2}} - 1]$$
$$\times(\partial_\lambda\theta)\Delta^\lambda_\mu p^\mu - c^{-1}\beta\epsilon_{\mu\nu}\langle p^\mu p^\nu\rangle, \quad (8.20)$$

where

$$B(\mathbf{p}^2,\beta) \equiv -\beta(\tfrac{4}{3}-\gamma)\mathbf{p}^2$$
$$-[\eta\beta(\gamma-1)-\gamma](1+\mathbf{p}^2)^{\frac{1}{2}} + \beta(\gamma-1). \quad (8.21)$$

To discern the general form of the solution of (8.20), we note that the four components $\beta_\mu = \beta u_\mu/c$ are the only parameters nontrivially involved in the integral operator $\mathfrak{L}$. Hence the effect of $\mathfrak{L}$ on various functions of p^μ must be as follows:

$$\mathfrak{L}[F(\mathbf{p}^2)] = F_1,$$
$$\mathfrak{L}[F(\mathbf{p}^2)p^\mu] = F_2p^\mu + F_3u^\mu,$$
$$\mathfrak{L}[F(\mathbf{p}^2)p^\mu p^\nu] = F_4p^\mu p^\nu + 2F_5u^{(\mu}p^{\nu)}$$
$$+ F_6u^\mu u^\nu + F_7\Delta^{\mu\nu},$$

where $F_i = F_i(\mathbf{p}^2,\beta)$ $(i = 1, \cdots, 7)$ are undetermined functions. From the linearity of $\mathfrak{L}$, it then follows that

$$\mathfrak{L}[F(\mathbf{p}^2)\Delta^\lambda_\mu p^\mu] = F_2(\mathbf{p}^2,\beta)\Delta^\lambda_\mu p^\mu,$$
$$\mathfrak{L}[F(\mathbf{p}^2)\langle p^\mu p^\nu\rangle] = F_4(\mathbf{p}^2,\beta)\langle p^\mu p^\nu\rangle.$$

The solution of (8.1) *accordingly has the form*

$$f(x,p) = c^{-1}A_1(\mathbf{p}^2,\beta)u^\alpha{}_{|\alpha}$$
$$+ (mc^2/k)A_2(\mathbf{p}^2,\beta)(\partial_\lambda\theta)\Delta^\lambda_\mu p^\mu$$
$$- c^{-1}\beta A_3(\mathbf{p}^2,\beta)\epsilon_{\mu\nu}\langle p^\mu p^\nu\rangle + f_{\text{hom}}(x,p), \quad (8.22)$$

where the functions A_i *are solutions of the integral equations*

$$\mathfrak{L}[A_1(\mathbf{p}^2,\beta)] = B(\mathbf{p}^2,\beta), \quad (8.23)$$

$$\mathfrak{L}[A_2(\mathbf{p}^2,\beta)\Delta^\lambda_\mu p^\mu] = [\eta^{-1}(1+\mathbf{p}^2)^{\frac{1}{2}} - 1]\Delta^\lambda_\mu p^\mu, \quad (8.24)$$

$$\mathfrak{L}[A_3(\mathbf{p}^2,\beta)\langle p^\mu p^\nu\rangle] = \langle p^\mu p^\nu\rangle, \quad (8.25)$$

and f_{hom} *is the general solution of the homogeneous equation* $\mathfrak{L}[f] = 0$:

$$f_{\text{hom}}(x,p) = C_\mu(x)p^\mu + C_5(x). \quad (8.26)$$

In conformity with our original assumptions concerning f, the five functions C_i are to be considered small and slowly varying.

9. Heat Flow and Viscosity

Insertion of (8.22) into the first-order perturbations (7.4) and (7.5) for the numerical flux and energy tensor yields

$$M_1^\nu = c^{-4}u^\alpha{}_{|\alpha}\chi^\nu_{(1)}$$
$$+ (mc^{-2}/k)(\partial_\lambda\theta)\Delta^\lambda_\mu\kappa_{(2)}{}^{\mu\nu} + M_{\text{hom}}{}^\nu, \quad (9.1)$$

$$T_1^{\nu\rho} = c^{-4}u^\alpha{}_{|\alpha}\kappa_{(1)}{}^{\nu\rho} + (mc^{-2}/k)(\partial_\lambda\theta)\Delta^\lambda_\mu\lambda_{(2)}{}^{\mu\nu\rho}$$
$$- c^{-2}\beta\epsilon_{\lambda\mu}\nu_{(3)}{}^{\lambda\mu\nu\rho} + T_{\text{hom}}{}^{\nu\rho}. \quad (9.2)$$

We have abbreviated, for $i = 1, 2, 3$,

$$\chi_{(i)}{}^\nu = mc^4\int N_0A_ip^\nu\,d\omega,$$
$$\kappa_{(i)}{}^{\mu\nu} = mc^5\int N_0A_ip^\mu p^\nu\,d\omega,$$
$$\lambda_{(i)}{}^{\mu\nu\rho} = mc^6\int N_0A_ip^\mu p^\nu p^\rho\,d\omega, \quad (9.3)$$
$$\nu_{(i)}{}^{\lambda\mu\nu\rho} = mc^3\int N_0A_ip^\lambda p^\mu p^\nu p^\rho\,d\omega,$$

and

$$M_{\text{hom}}{}^\nu \equiv mc\int N_0f_{\text{hom}}p^\nu\,d\omega = c^{-1}C_\mu T_0{}^{\mu\nu} + C_5M_0{}^\nu$$
$$= (\mu c^{-1}C_\mu u^\mu + \rho C_5)u^\nu + c^{-1}P\Delta^{\mu\nu}C_\mu, \quad (9.4a)$$

$$T_{\text{hom}}{}^{\nu\rho} \equiv mc^2\int N_0f_{\text{hom}}p^\nu p^\rho\,d\omega = c^{-1}C_\mu U_0{}^{\mu\nu\rho} + C_5T_0{}^{\nu\rho}$$
$$= (\xi c^{-1}C_\mu u^\mu + \mu C_5)u^\nu u^\rho$$
$$+ 3\zeta c^{-1}C_\mu u^{(\mu}\Delta^{\nu\rho)} + PC_5\Delta^{\nu\rho}, \quad (9.4b)$$

by (7.25), (5.6), (5.7), (6.13), (6.14), and (6.20). The integrals in (9.3) have the following general form:

$$\chi_{(i)}{}^\nu = \chi_iu^\nu,$$
$$\kappa_{(i)}{}^{\mu\nu} = \kappa_{i1}u^\mu u^\nu + \kappa_{i2}\Delta^{\mu\nu},$$
$$\lambda_{(i)}{}^{\mu\nu\rho} = \lambda_{i1}u^\mu u^\nu u^\rho + 3\lambda_{i2}u^{(\mu}\Delta^{\nu\rho)}, \quad (9.5)$$
$$\nu_{(i)}{}^{\lambda\mu\nu\rho} = \nu_{i1}u^\lambda u^\mu u^\nu u^\rho$$
$$+ 6\nu_{i2}\Delta^{(\lambda\mu}u^\nu u^{\rho)} + \tfrac{3}{2}\nu_{i3}\Delta^{(\lambda\mu}\Delta^{\nu\rho)}.$$

The coefficients χ_i, κ_{ij}, etc. are functions of β.

Since $\epsilon_{\lambda\mu}\Delta^{\lambda\mu} = \epsilon_{\lambda\mu}u^\mu = 0$, (9.1) and (9.2) simplify to

$$M_1^\nu = c^{-4}\chi_1u^\nu u^\alpha{}_{|\alpha}$$
$$+ (mc^{-2}/k)\kappa_{22}\Delta^{\lambda\nu}(\partial_\lambda\theta) + M_{\text{hom}}{}^\nu, \quad (9.6)$$

$$T_1^{\nu\rho} = c^{-4}\kappa_{11}u^\nu u^\rho u^\alpha{}_{|\alpha} + 2(mc^{-2}/k)\lambda_{22}u^{(\rho}\Delta^{\nu)\lambda}(\partial_\lambda\theta) - c^{-2}\beta\nu_{33}\epsilon^{\nu\rho} + c^{-4}\kappa_{12}\Delta^{\nu\rho}u^\alpha{}_{|\alpha} + T_{\text{hom}}{}^{\nu\rho}. \quad (9.7)$$

The conditions of fit, (7.7), fix two of the five functions C_i. We find

$$-u_\nu M_1^\nu = c^{-2}\chi_1 u^\alpha{}_{|\alpha} - u_\nu M_{\text{hom}}{}^\nu = 0,$$

$$u_\nu u_\rho T_1^{\nu\rho} = \kappa_{11}u^\alpha{}_{|\alpha} + u_\nu u_\rho T_{\text{hom}}{}^{\nu\rho} = 0,$$

with the solution

$$c^{-1}C_\mu u^\mu = -c^{-4}(\rho\xi - \mu^2)^{-1}(\rho\kappa_{11} - \mu\chi_1)u^\alpha{}_{|\alpha}, \quad (9.8)$$

$$C_5 = c^{-4}(\rho\xi - \mu^2)^{-1}(\mu\kappa_{11} - \xi\chi_1)u^\alpha{}_{|\alpha}.$$

The three spatial components $\Delta_\mu^\lambda C^\mu$ remain arbitrary, but have no effect on the expressions for viscosity and heat flow [see (9.13) and (9.17) below].

Substituting (9.4) and (9.8) into (9.6) and (7.7), we have finally

$$M_1^\nu = (mc^{-2}/k)\kappa_{22}\Delta^{\lambda\nu}(\partial_\lambda\theta) + c^{-1}P\Delta^{\lambda\nu}C_\lambda, \quad (9.9)$$

$$T_1^{\nu\rho} = 2(mc^{-2}/k)\lambda_{22}u^{(\rho}\Delta^{\nu)\lambda}(\partial_\lambda\theta) - c^{-2}\beta\nu_{33}\epsilon^{\nu\rho} - \kappa\Delta^{\nu\rho}u^\alpha{}_{|\alpha} + 2c^{-1}\eta Pu^{(\rho}\Delta^{\nu)\lambda}C_\lambda, \quad (9.10)$$

where

$$\kappa = -c^{-4}\kappa_{12} - c^{-4}(\rho\xi - \mu^2)^{-1} \times [(\mu\zeta - P\xi)\chi_1 - (\rho\zeta - \mu P)\kappa_{11}]$$

$$= -c^{-4}\kappa_{12} - c^{-2}[\eta(\gamma - 1) - \beta^{-1}\gamma]\chi_1 + c^{-2}(\gamma - 1)\kappa_{11}, \quad (9.11)$$

by use of (6.37).

The *heat flux* q^μ, defined as the flow of energy relative to the particle stream, is given by

$$q^\mu = -c(-M_\alpha M^\alpha)^{-\frac{1}{2}}T^{\lambda\nu}M_\nu \times [\delta_\lambda^\mu - (M_\beta M^\beta)^{-1}M_\lambda M^\mu]. \quad (9.12)$$

In a local Minkowskian frame with time axis parallel to M^μ, this reduces to

$$q^\mu = c[T^{14}, T^{24}, T^{34}, 0].$$

To the first order in f, (9.12) can be written

$$q^\mu = -\eta c^2 M_1{}^\mu - \Delta_\rho^\mu T_1{}^{\rho\nu}u_\nu = (m/k)(\lambda_{22} - \eta\kappa_{22})\Delta^{\mu\lambda}(\partial_\lambda\theta). \quad (9.13)$$

From (9.5) and (9.3), we have

$$\kappa_{22} = \tfrac{1}{3}\Delta_{\mu\nu}\kappa_{(2)}{}^{\mu\nu} = \tfrac{1}{3}mc^5\int N_0A_2\mathbf{p}^2\,d\omega,$$

$$\lambda_{22} = -\tfrac{1}{3}c^{-2}u_\lambda\Delta_{\mu\nu}\lambda_{(2)}{}^{\lambda\mu\nu} = \tfrac{1}{3}mc^5\int N_0A_2\mathbf{p}^2(1+\mathbf{p}^2)^{\frac{1}{2}}\,d\omega.$$

Hence,

$$(k/m)\lambda \equiv \lambda_{22} - \eta\kappa_{22} = \tfrac{1}{3}mc^5\int N_0A_2\mathbf{p}^2[(1+\mathbf{p}^2)^{\frac{1}{2}} - \eta]\,d\omega \quad (9.14)$$

$$= \tfrac{1}{3}mc^5\eta\int N_0(\Delta_\mu^\lambda p^\mu A_2)[\eta^{-1}(1+\mathbf{p}^2)^{\frac{1}{2}} - 1]\Delta_\lambda^\nu p_\nu\,d\omega$$

$$= \tfrac{1}{3}mc^5\eta(\Delta_\mu^\lambda p^\mu A_2,\ \Delta_\lambda^\nu p_\nu A_2) > 0. \quad (9.15)$$

where we have used (8.24), (8.4), and (8.8).

Summarizing, *we have obtained Fourier's law of heat conduction in the form*

$$q^\mu = \lambda\Delta^{\mu\nu}(\partial_\nu\theta), \quad (9.16)$$

in which the thermal conductivity λ *is given by* (9.14). In accordance with the second law of thermodynamics as expressed by (5.10) or (8.11), λ is positive-definite. It is noteworthy that the flow of heat is determined by the gradient of thermal potential [see (6.27)] rather than temperature.

We turn to the *viscous stress tensor* $\tau_{\mu\nu}$. To the first order in f, this may be defined unambiguously by

$$\tau_{\alpha\beta} \equiv \Delta_{\alpha\rho}\Delta_{\beta\nu}T_1{}^{\nu\rho} = -c^{-2}\beta\nu_{33}\epsilon_{\alpha\beta} - \kappa\Delta_{\alpha\beta}u^\mu{}_{|\mu}.$$

Hence the viscous stresses are given in terms of the spatial velocity gradient by

$$\tau_{\mu\nu} = -\nu\Delta_\mu^\alpha\Delta_\nu^\beta(u_{\alpha|\beta} + u_{\beta|\alpha} - \tfrac{2}{3}\Delta_{\alpha\beta}u^\lambda{}_{|\lambda}) - \kappa\Delta_{\mu\nu}u^\lambda{}_{|\lambda}, \quad (9.17)$$

where ν, *the first coefficient of viscosity, is*

$$\nu = \tfrac{1}{2}c^{-2}\beta\nu_{33} = \tfrac{1}{15}c^{-2}\beta\Delta_{\lambda\mu}\Delta_{\nu\rho}\nu_{(3)}{}^{\lambda\mu\nu\rho},$$

i.e.,

$$kT\nu = \tfrac{2}{15}m^2c^3\int N_0A_3\mathbf{p}^4\,d\omega. \quad (9.18)$$

Noting that

$$\langle p^\mu p^\nu\rangle\langle p_\mu p_\nu\rangle = \tfrac{2}{3}\mathbf{p}^4,$$

and using (8.25), (8.4), and (8.8), we may throw this expression into a form which brings out the positive character of ν:

$$kT\nu = \tfrac{1}{5}m^2c^3\int N_0A_3\langle p^\mu p^\nu\rangle\mathfrak{L}[A_3\langle p_\mu p_\nu\rangle]\,d\omega$$

$$= \tfrac{1}{5}m^2c^3(A_3\langle p^\mu p^\nu\rangle,\ A_3\langle p_\mu p_\nu\rangle) > 0. \quad (9.19)$$

The relativistic gas departs from its classical limit in having a nonvanishing volume viscosity. The second coefficient of viscosity has the value [cf. (9.11)]

$$m^{-1}c^{-1}\kappa = -\frac{1}{3}\int N_0A_1\mathbf{p}^2\,d\omega - [\eta(\gamma-1)-\beta^{-1}\gamma]\int N_0A_1(1+\mathbf{p}^2)^{\frac{1}{2}}\,d\omega + (\gamma-1)\int N_0A_1(1+\mathbf{p}^2)\,d\omega,$$

or, by (8.21),

$$\kappa = (mc/\beta)\int N_0A_1B\,d\omega. \tag{9.20}$$

By virtue of (8.23), this can also be written

$$\kappa = (kT/c)\int N_0A_1\mathfrak{L}[A_1]\,d\omega = (kT/c)(A_1, A_1) > 0, \tag{9.21}$$

in agreement with the second law of thermodynamics.

PART II: RELATIVISTIC MAXWELLIAN PARTICLES

10. A Special Form of Cross Section

Our formal expressions for viscosity and thermal conductivity can be explicitly evaluated only when the collision mechanism, as specified by the scattering cross section $\sigma(g, \Theta)$, is known. In the remainder of this paper we shall attempt to push through the calculations for a case where $\sigma(g, \Theta)$ has a special, mathematically tractable form. The results may be useful as a qualitative guide to what can be expected in the general case, and could also serve as the ground approximation for various perturbational procedures.

In the classical theory, the linearized collision operator is known[14] to possess simple eigenfunctions (the Laguerre polynomials) when $\sigma(g, \Theta)$ is a separable function of the form $g^{-1}\Gamma(\Theta)$. This form arises, for instance, in the case of Maxwellian particles, repelling according to the inverse fifth power of the distance.

By analogy, we shall refer to relativistic Maxwellian particles when the scattering cross section is given by

$$g\sigma(g, \Theta) = (1+\tfrac{1}{4}g^2)^{-\frac{1}{2}}\Gamma(\Theta). \tag{10.1}$$

The factor $(1+\frac{1}{4}g^2)^{-\frac{1}{2}}$ has been included for mathematical convenience.

Inserting (10.1) into the expression (8.2′) for the linearized collision operator, we obtain

$$\mathfrak{L}[f] = \alpha\int\int_0^{\pi}\int_0^{2\pi}\exp(\beta_\mu{}'p_\mu)\,\delta(f) \times (1+\tfrac{1}{4}g^2)^{-\frac{1}{2}}\Gamma(\Theta)\sin\vartheta^*\,d\vartheta^*\,d\epsilon^*\,d'\omega. \tag{10.2}$$

Our problem is to solve the integral equations (8.23), (8.24), and (8.25) when the integral operator has the special form (10.2). The thermal conductivity and coefficients of viscosity then follow immediately from (9.13), (9.18), and (9.20). Only the solution of (8.23) and (8.24) will be considered in this paper. The third equation can be handled by similar techniques, but the computations are somewhat lengthier.

11. The Centroidal Triad

The integral operator $\mathfrak{L}$ involves two integrations: first, with respect to the unit vector g^*_α/g giving the direction of recoil in the center-of-mass frame, and secondly, with respect to one of the initial momenta ${}'p_\alpha$. The other, p_α, is kept fixed throughout. In the *first* integration, we fix ${}'p_\alpha$ and hence also $\bar{p}_\alpha = (p_\alpha + {}'p_\alpha)/|p_\mu + {}'p_\mu|$. Recalling that $g^*_\alpha\bar{p}^\alpha = 0$, we see that we are integrating over the unit sphere in a *fixed* 3-space, the centroidal 3-flat. We shall now prepare the way for this integration by constructing, in the centroidal 3-flat, a convenient orthonormal triad to which the angles ϑ^*, ϵ^* can be referred.

Accordingly, let n_α be an arbitrary fixed unit timelike vector, which will later be chosen parallel to the macroscopic 4-velocity u_α. Introduce the spacelike vectors

$$X_\alpha = n_\alpha - g^{-2}(n_\mu g^\mu)g_\alpha + (n_\mu\bar{p}^\mu)\bar{p}_\alpha, \qquad Y_\alpha = (-\det g_{\mu\nu})^{\frac{1}{2}}\epsilon_{\alpha\beta\gamma\delta}n^\beta g^\gamma\bar{p}^\delta \tag{11.1}$$

($\epsilon_{\alpha\beta\gamma\delta}$ is the Levi–Civita permutation symbol), and the corresponding unit vectors

$$I_\alpha = X_\alpha/X, \quad J_\alpha = Y_\alpha/Y, \quad K_\alpha = g_\alpha/g, \tag{11.2}$$

where

$$X \equiv (X_\alpha X^\alpha)^{\frac{1}{2}}, \qquad Y \equiv (Y_\alpha Y^\alpha)^{\frac{1}{2}}.$$

From (11.1) and (3.5),

$$\bar{p}_\alpha g^\alpha = \bar{p}_\alpha X^\alpha = \bar{p}_\alpha Y^\alpha = 0, \qquad g_\alpha X^\alpha = g_\alpha Y^\alpha = X_\alpha Y^\alpha = 0, \tag{11.3}$$

$$n_\alpha Y^\alpha = 0, \tag{11.4}$$

so that I_α, J_α, K_α form an orthonormal triad in the 3-flat normal to $\bar{p}_\alpha$.

[14] C. S. Wang and G. E. Uhlenbeck: Engineering Research Institute, University of Michigan, Project M199(1952). An elegant exposition is given by L. Waldmann, *Encyclopedia of Physics*, edited by S. Flügge (Julius Springer-Verlag, Berlin, 1958), Vol. XII, p. 370.

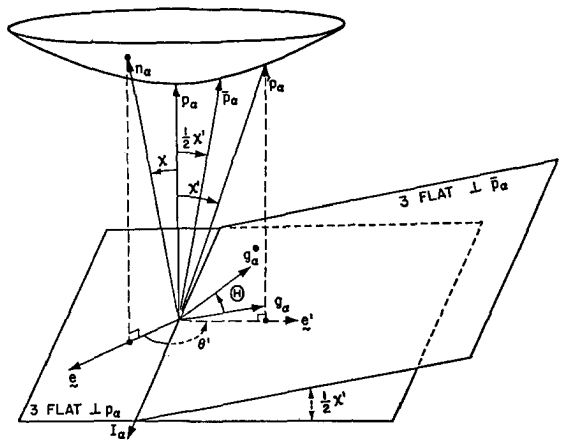

FIG. 1. The centroidal 3-flat, showing two vectors of the centroidal triad. One space dimension (the direction of vector J_α) has been suppressed.

The calculations are somewhat simplified in a Minkowskian frame whose 4-axis is directed along the fixed vector p_α and such that n^α lies in the 2-flat of the axes 3, 4. In this frame, we have

$$p^\alpha = [0, 0, 0, 1], \tag{11.5}$$

and, using pseudopolar coordinates [cf (2.2)],

$$n^\alpha = [\mathbf{e} \sinh \chi, \cosh \chi], \qquad \mathbf{e} = (0, 0, 1), \tag{11.6}$$

$$'p^\alpha = [\mathbf{e}' \sinh \chi', \cosh \chi'],$$

$$\mathbf{e}' = (\sin \theta' \cos \varphi', \sin \theta' \sin \varphi', \cos \theta']. \tag{11.7}$$

From (11.5), (11.7),

$$g^\alpha = {}'p^\alpha - p^\alpha$$

$$= 2 \sinh \tfrac{1}{2}\chi' [\mathbf{e}' \cosh \tfrac{1}{2}\chi', \sinh \tfrac{1}{2}\chi'], \tag{11.8}$$

$$g = 2 \sinh \tfrac{1}{2}\chi', \qquad (1 + \tfrac{1}{4}g^2)^{-\frac{1}{2}} = \operatorname{sech} \tfrac{1}{2}\chi', \tag{11.9}$$

$$\tilde{p}^\alpha = [\mathbf{e}' \sinh \tfrac{1}{2}\chi', \cosh \tfrac{1}{2}\chi']. \tag{11.10}$$

Equations (11.1) yield, after some routine work,

$$X^\alpha = \sinh \chi [\mathbf{e} - \mathbf{e}' \cos \theta', 0],$$

$$Y^\alpha = 2 \sinh \chi \sinh \tfrac{1}{2}\chi' [\mathbf{e} \times \mathbf{e}', 0],$$

whence it follows that

$$I^\alpha = [\mathbf{e} \operatorname{cosec} \theta' - \mathbf{e}' \cot \theta', 0],$$

$$J^\alpha = [\mathbf{e} \times \mathbf{e}' \operatorname{cosec} \theta', 0], \tag{11.11}$$

$$K^\alpha = [\mathbf{e}' \cosh \tfrac{1}{2}\chi', \sinh \tfrac{1}{2}\chi'].$$

We refer the unit vector g^*_α/g to our orthonormal triad by setting

$$g^*_\alpha/g = I_\alpha \sin \Theta \cos \Phi + J_\alpha \sin \Theta \sin \Phi + K_\alpha \cos \Theta. \tag{11.12}$$

This definition of Θ accords with our previous one, (4.9). In the integration over g^*_α/g in (10.2) we can now replace ϑ^*, ϵ^* by Θ, Φ. The situation is shown schematically in Fig. 1, with the angle Φ suppressed.

Let r_α be an arbitrary fixed unit vector in the 3-flat of the axes 1, 3, 4:

$$r^\alpha = [\mathbf{e}_r r_I, r_{II}], \qquad \mathbf{e}_r = (\sin \theta, 0, \cos \theta), \tag{11.13}$$

where

$$r_I = \sinh \chi_r, \quad r_{II} = \cosh \chi_r \quad \text{if} \quad r_\alpha r^\alpha = -1,$$

$$r_I = \cosh \chi_r, \quad r_{II} = \sinh \chi_r \quad \text{if} \quad r_\alpha r^\alpha = 1.$$

We then have

$$r^\alpha p_\alpha = -r_{II}, \tag{11.14a}$$

$$r^\alpha {}'p_\alpha = -r_{II} \cosh \chi' + r_I \sinh \chi'$$

$$\times (\cos \theta \cos \theta' + \sin \theta \sin \theta' \cos \varphi'). \tag{11.14b}$$

Noting from (3.1) and (3.2) that

$$^{(\prime)}p^*_\alpha = \tfrac{1}{2}(p_\alpha + {}'p_\alpha)(\overline{+})\tfrac{1}{2}g^*_\alpha,$$

and computing $r^\alpha g^*_\alpha$ from (11.9), (11.11), and (11.12), we obtain

$$r^\alpha p^*_\alpha = -r_{II}(1 + 2 \sinh^2 \tfrac{1}{2}\chi' \sin^2 \tfrac{1}{2}\Theta)$$

$$+ r_I \sinh \chi' \sin^2 \tfrac{1}{2}\Theta (\cos \theta \cos \theta' + \sin \theta \sin \theta'$$

$$\times \cos \varphi') - r_I \sinh \tfrac{1}{2}\chi' \sin \Theta (\cos \theta \sin \theta' \cos \Phi$$

$$- \sin \theta \cos \theta' \cos \varphi' \cos \Phi - \sin \theta \sin \varphi' \sin \Phi). \tag{11.14c}$$

The corresponding expression for $r^\alpha {}'p^*_\alpha$ is just (11.14c) with Θ, Φ replaced by $\pi - \Theta$, $\pi + \Phi$ respectively.

12. Scalar Eigenfunctions

As a possible generating function for the scalar eigenfunctions of the collision operator $\mathfrak{L}$, let us consider $\exp(\sigma c^{-1} u_\alpha p^\alpha)$, where σ is a free parameter.

In the formulas of Sec. 11, we choose the arbitrary vectors n_α, r_α parallel to the macroscopic 4-velocity u_α, so that

$$r_\alpha = n_\alpha = c^{-1} u_\alpha, \qquad \theta = 0, \qquad \chi_r = \chi. \tag{12.1}$$

From (10.2), we then have

$$\mathfrak{L}[\exp(\sigma n_\alpha p^\alpha)] = \alpha \int_0^\pi (I + {}'I - I^* - {}'I^*) \Gamma(\Theta) \sin \Theta \, d\Theta, \tag{12.2}$$

where

$$I = e^{-\sigma\cosh\chi}\iint_0^{2\pi}\exp(\beta n_\mu' p^\mu) \times \operatorname{sech}\tfrac{1}{2}\chi'\, d'\omega\, d\Phi, \quad (12.3)$$

$$'I = \iint_0^{2\pi}\exp[(\beta+\sigma)n_\mu' p^\mu] \times \operatorname{sech}\tfrac{1}{2}\chi'\, d'\omega\, d\Phi, \quad (12.4)$$

$${}^{(\prime)}I^* = \iint_0^{2\pi}\exp[\beta n_\mu' p^\mu + \sigma n^{\mu(\prime)} p_\mu^*] \times \operatorname{sech}\tfrac{1}{2}\chi'\, d'\omega\, d\Phi. \quad (12.5)$$

We recall that

$$\int\cdots d'\omega \equiv \int_0^\infty\int_0^\pi\int_0^{2\pi}\cdots \sinh^2\chi' \times \sin\theta'\, d\chi'\, d\theta'\, d\varphi'$$

[cf (2.3)], and that $n_\mu' p^\mu$, $n^{\mu(\prime)} p_\mu^*$ are given by (11.14b) and (11.14c).

To evaluate (12.3), introduce the new variables of integration

$$x = \sinh\tfrac{1}{2}\chi'\sin\theta'\cos\Phi,$$
$$y = \sinh\tfrac{1}{2}\chi'\sin\theta'\sin\Phi,$$
$$z = \sinh\tfrac{1}{2}\chi'\cos\theta', \quad (12.6)$$

so that

$$dx\,dy\,dz = \sinh^2\tfrac{1}{2}\chi'\sin\theta'\, d(\sinh\tfrac{1}{2}\chi')\, d\theta'\, d\Phi,$$
$$dx\,dy\,dz\,d\phi' = \tfrac{1}{8}\operatorname{sech}\tfrac{1}{2}\chi'\, d'\omega\, d\Phi.$$

The integration over φ' is trivial, and we find

$$I = e^{-\sigma\cosh\chi}\,16\pi\iiint_{-\infty}^{\infty}\exp[-\beta\cosh\chi(1+2x^2 + 2y^2+2z^2) + 2\beta\sinh\chi(1+x^2+y^2+z^2)^{\frac{1}{2}}z]\,dx\,dy\,dz.$$

Replacing z by a new variable t according to

$$z = (1+x^2+y^2)^{\frac{1}{2}}\sinh\tfrac{1}{2}t, \quad (12.7)$$

reduces this to

$$I = e^{-\sigma\cosh\chi}\,8\pi\int\!\!\iint_{-\infty}^{\infty}\exp[-\beta\cosh\chi(x^2+y^2) - \beta(1+x^2+y^2)\cosh(t-\chi)] \times (1+x^2+y^2)^{\frac{1}{2}}\cosh\tfrac{1}{2}t\,dt\,dx\,dy.$$

Now,

$$\int_{-\infty}^{\infty} e^{-a\cosh(t-\chi)}\cosh\tfrac{1}{2}t\,dt = 2\cosh\tfrac{1}{2}\chi\int_0^\infty e^{-a\cosh t}\cosh\tfrac{1}{2}t\,dt = 2\cosh\tfrac{1}{2}\chi\cdot(2a/\pi)^{-\frac{1}{2}}e^{-a}. \quad (12.8)$$

Hence we have finally

$$I = e^{-\sigma\cosh\chi}\cdot 16\pi\cosh\tfrac{1}{2}\chi\cdot(2\beta/\pi)^{-\frac{1}{2}}e^{-\beta} \times\int_{-\infty}^{\infty}\!\!\int e^{-\beta(1+\cosh\chi)(x^2+y^2)}\,dx\,dy = e^{-\sigma\cosh\chi}\cdot 16\pi\operatorname{sech}\tfrac{1}{2}\chi\cdot(2\beta/\pi)^{-\frac{3}{2}}e^{-\beta}. \quad (12.9)$$

$'I$ is just the integral of (12.3) with β replaced by $(\beta+\sigma)$:

$$'I = 16\pi\operatorname{sech}\tfrac{1}{2}\chi\cdot[2(\beta+\sigma)/\pi]^{-\frac{1}{2}}e^{-(\beta+\sigma)}. \quad (12.10)$$

The evaluation of ${}^{(\prime)}I^*$ is slightly more elaborate. In accordance with (11.14) and (12.1), the integrand involves

$$\beta n_\mu' p^\mu + \sigma n^{\mu(\prime)} p_\mu^* = -(\beta+\sigma)\cosh\chi - 2\,{}^{(\prime)}\beta^*\sinh^2\tfrac{1}{2}\chi' + {}^{(\prime)}\beta^*\sinh\chi\sinh\chi'\cos\theta'(\overline{+}) (\overline{+})\,\sigma\sinh\chi\sin\Theta\sinh\tfrac{1}{2}\chi'\sin\theta'\cos\Phi,$$

with

$$\beta^* \equiv \beta + \sigma\sin^2\tfrac{1}{2}\Theta, \quad {}'\beta^* \equiv \beta + \sigma\cos^2\tfrac{1}{2}\Theta. \quad (12.11)$$

The substitutions (12.6) and (12.7) convert this into

$$\beta n_\mu' p^\mu + \sigma n^\mu\,{}^{(\prime)}p_\mu^* = -\,{}^{(\prime)}\sigma^*\cosh\chi - {}^{(\prime)}\beta^*\cosh\chi\cdot(x^2+y^2) - {}^{(\prime)}\beta^*(1+x^2+y^2) \times\cosh(t-\chi)(\overline{+})\sigma\sinh\chi\sin\Theta\cdot x, \quad (12.12)$$

where

$$\sigma^* = \sigma\cos^2\tfrac{1}{2}\Theta, \qquad {}'\sigma^* = \sigma\sin^2\tfrac{1}{2}\Theta, \quad (12.13)$$

and also

$$\operatorname{sech}\tfrac{1}{2}\chi'\, d'\omega\, d\Phi = 8(1+x^2+y^2)^{\frac{1}{2}} \times\cosh\tfrac{1}{2}t\,dx\,dy\,dt\,d\varphi'. \quad (12.14)$$

Using (12.8), we obtain without difficulty

$${}^{(\prime)}I^* = e^{-({}^{(\prime)}\sigma^*\cosh\chi}16\pi\cosh\tfrac{1}{2}\chi\cdot[2\,{}^{(\prime)}\beta^*/\pi]^{-\frac{1}{2}}e^{-({}^{(\prime)}\beta^*} \times\iint_{-\infty}^{\infty}\exp[-{}^{(\prime)}\beta^*(1+\cosh\chi) \times(x^2+y^2)(\overline{+})\sigma\sinh\chi\sin\Theta\cdot x]\,dx\,dy$$

$$= 16\pi \operatorname{sech} \tfrac{1}{2}\chi \cdot [2^{(\prime)}\beta^*/\pi]^{-\frac{1}{2}} e^{-(\beta+\sigma)}$$

$$\times \exp\left[-(2^{(\prime)}\sigma^*\beta/{}^{(\prime)}\beta^*)\sinh^2 \tfrac{1}{2}\chi\right]. \tag{12.15}$$

This completes the evaluation of the integrals involved in (12.2).

It is convenient to define the parameters s, ${}^{(\prime)}s^*$ by

$$\sigma = \beta s/(1-s), \qquad s^* = s\cos^2 \tfrac{1}{2}\Theta, \qquad {}'s^* = s\sin^2 \tfrac{1}{2}\Theta, \tag{12.16}$$

so that

$${}^{(\prime)}\beta^*/\beta = (1 - {}^{(\prime)}s^*)/(1-s), \qquad {}^{(\prime)}\sigma^*/{}^{(\prime)}\beta^* = {}^{(\prime)}s^*/(1 - {}^{(\prime)}s^*), \tag{12.17}$$

and to define

$$\Psi_p(\tau, s) \equiv (1-s)^{-(p+1)} e^{-\tau s/(1-s)}. \tag{12.18}$$

Multiplying both sides of (12.2) by $(1-s)^{-\frac{1}{2}}e^{\sigma}$, we obtain

$$\mathfrak{L}[\Psi_{\frac{1}{2}}(\tau, s)] = (2\pi)^{5/2}\beta^{-3/2}\alpha e^{-\beta} \operatorname{sech} \tfrac{1}{2}\chi \times \int_0^{\pi} [\Psi_{\frac{1}{2}}(\tau, s) + 1 - \Psi_{\frac{1}{2}}(\tau, s^*) - \Psi_{\frac{1}{2}}(\tau, {}'s^*)]\Gamma(\Theta)\sin\Theta\, d\Theta, \tag{12.19}$$

where we have defined

$$\tau \equiv -\beta(1 + n_\alpha p^\alpha) = \beta[(1+\mathbf{p}^2)^{\frac{1}{2}} - 1] = 2\beta\sinh^2 \tfrac{1}{2}\chi, \tag{12.20}$$

recalling (8.18).

If (12.19) is differentiated n times with respect to the parameter s, and if we note the definition of the Laguerre polynomials

$$n!\, L_n^p(\tau) = [(\partial/\partial s)^n \Psi_p(\tau, s)]_{s=0}, \tag{12.21}$$

it follows that

$$\mathfrak{L}[L_n^{\frac{1}{2}}(\tau)] = E_n(1 + \tfrac{1}{2}\tau/\beta)^{-\frac{1}{2}} L_n^{\frac{1}{2}}(\tau). \tag{12.22}$$

As in the classical theory, *the Laguerre polynomials are scalar eigenfunctions of the linearized collision operator for Maxwellian particles.* The eigenvalues E_n are given by

$$E_0 = 0,$$

$$E_n = (2\pi)^{5/2}\beta^{-3/2}\alpha e^{-\beta} \int_0^{\pi} (1 - \cos^{2n}\tfrac{1}{2}\Theta - \sin^{2n}\tfrac{1}{2}\Theta) \times \Gamma(\Theta)\sin\Theta\, d\Theta \qquad (n = 1, 2, \cdots). \tag{12.23}$$

13. Vector Eigenfunctions

The correspondence between the classical and relativistic eigenfunctions, very simple for scalars, is unfortunately much less simple for tensorial eigenfunctions of higher order. In this section we shall attempt to obtain a set of vector eigenfunctions.

We begin by considering $\mathfrak{L}[r_\mu p^\mu \exp(\sigma n_\alpha p^\alpha)]$, where the vectors n_α, r_α are chosen respectively parallel and perpendicular to the macroscopic 4-velocity u_α. Thus,

$$n_\alpha = c^{-1}u_\alpha, \qquad r_\alpha n^\alpha = 0, \tag{13.1}$$

implying, by virtue of (11.6) and (11.13),

$$\sinh\chi\cosh\chi_r\cos\theta = \cosh\chi\sinh\chi_r. \tag{13.2}$$

From (10.2),

$$[r_\mu p^\mu \exp(\sigma n_\alpha p^\alpha)] = \alpha\int_0^{\pi} (J + {}'J - J^* - {}'J^*)\Gamma(\Theta)\sin\Theta\, d\Theta, \tag{13.3}$$

where

$$J = [r_\mu p^\mu \exp(\sigma n_\alpha p^\alpha)] \times \iint_0^{2\pi} \exp(\beta n_\alpha {}'p^\alpha)\operatorname{sech}\tfrac{1}{2}\chi'\, d'\omega\, d\Phi, \tag{13.4}$$

$${}'J = \iint_0^{2\pi} r_\mu {}'p^\mu \exp[(\beta+\sigma)n_\alpha {}'p^\alpha] \times \operatorname{sech}\tfrac{1}{2}\chi'\, d'\omega\, d\Phi, \tag{13.5}$$

$${}^{(\prime)}J^* = \iint_0^{2\pi} r^{\mu\,(\prime)}p^*_\mu \exp[\beta n_\alpha {}'p^\alpha + \sigma n^{\alpha\,(\prime)}p^*_\alpha] \times \operatorname{sech}\tfrac{1}{2}\chi'\, d'\omega\, d\Phi. \tag{13.6}$$

The various scalar products involved in these integrals are given by the general formulas (11.14).

The integral in (13.4) has already been evaluated [cf (12.3)]:

$$J = e^{-\sigma\cosh\chi} r_\mu p^\mu 16\pi \operatorname{sech}\tfrac{1}{2}\chi(2\beta/\pi)^{-\frac{1}{2}}e^{-\beta}. \tag{13.7}$$

For ${}'J$, we have

$${}'J = \iint_0^{2\pi} [\cosh\chi_r\sinh\chi'(\cos\theta\cos\theta' + \sin\theta \times \sin\theta'\cos\varphi') - \sinh\chi_r\cosh\chi'] \times \exp[-(\beta+\sigma)(\cosh\chi\cosh\chi' - \sinh\chi\sinh\chi'\cos\theta')]\operatorname{sech}\tfrac{1}{2}\chi'\cdot d'\omega\, d\Phi.$$

The term involving $\cos\varphi'$ has a vanishing integral with respect to φ'. Making the substitutions (12.6), (12.7), and using (13.2), we eventually find

$${}'J = 8\pi\sinh\chi_r \times \int_{-\infty}^{\infty}\iiint \left[\frac{\sinh(t-\chi)}{\sinh\chi}(1 + x^2 + y^2) - (x^2+y^2)\right]$$

$$\times \exp\{-(\beta+\sigma)[\cosh\chi\cdot(x^2+y^2) + (1+x^2+y^2)\cosh(t-\chi)]\} \times (1+x^2+y^2)^{\frac{1}{2}}\cosh\tfrac{1}{2}t\,dt\,dx\,dy.$$

Integration with respect to t is simple if we use (12.8) and the formula

$$\int_{-\infty}^{\infty} e^{-a\cosh(t-\chi)} a\sinh(t-\chi)\cosh\tfrac{1}{2}t\,dt = \sinh\tfrac{1}{2}\chi\cdot(2a/\pi)^{-\frac{1}{2}}e^{-a}, \qquad (13.8)$$

which follows from (12.8) by differentiation with respect to χ. We find

$$'J = 8\pi\sinh\chi_r[2(\beta+\sigma)/\pi]^{-\frac{1}{2}}e^{-(\beta+\sigma)} \times \iint_{-\infty}^{\infty}[\tfrac{1}{2}(\beta+\sigma)^{-1}\operatorname{sech}\tfrac{1}{2}\chi - 2\cosh\tfrac{1}{2}\chi\cdot(x^2+y^2)] \times \exp[-(\beta+\sigma)(1+\cosh\chi)(x^2+y^2)]\,dx\,dy.$$

Noting (13.11) and (13.12) below, with $b = 0$, we finally arrive at

$$'J = 8\pi r_\mu p^\mu[2(\beta+\sigma)/\pi]^{-3} \times \operatorname{sech}^3\tfrac{1}{2}\chi(\beta+\sigma)^{-1}e^{-(\beta+\sigma)}. \qquad (13.9)$$

Turning to (13.6), we find, after making the substitutions (12.6) and (12.7), that the various factors in the integrand are given by (12.12), (12.14), and [cf (11.14c)]

$$r^\alpha p^*_\alpha = -\sinh\chi_r\Big[\cos^2\tfrac{1}{2}\Theta + \sin^2\tfrac{1}{2}\Theta\cdot(x^2+y^2) + \coth\chi\cdot\sin\Theta\cdot x - \sin^2\tfrac{1}{2}\Theta \times \frac{\sinh(t-\chi)}{\sinh\chi}(1+x^2+y^2)\Big] + \cdots, \qquad (13.10)$$

where we have made use of (13.2). The dots in (13.10) indicate terms whose integrals with respect to φ' vanish. The corresponding expression for $r^{\alpha\prime}p^*_\alpha$ is obtainable from (13.10) by the replacements

$$\Theta \to \pi - \Theta, \qquad x \to -x, \qquad y \to -y.$$

Integration with respect to t proceeds uneventfully with the aid of (12.8) and (13.8):

$$J^* = -8\pi e^{-(\beta^*+\sigma^*\cosh\chi)}(2\beta^*/\pi)^{-\frac{1}{2}}\sinh\chi_r \times \iint_{-\infty}^{\infty} 2\cosh\tfrac{1}{2}\chi[\cos^2\tfrac{1}{2}\Theta + \sin^2\tfrac{1}{2}\Theta\cdot(x^2+y^2) + \coth\chi\cdot\sin\Theta\cdot x - \tfrac{1}{4}\beta^{*-1}\operatorname{sech}^2\tfrac{1}{2}\chi\cdot\sin^2\tfrac{1}{2}\Theta] \times \exp[-\beta^*(1+\cosh\chi)(x^2+y^2) - \sigma\sinh\chi\cdot\sin\Theta\cdot x]\,dx\,dy.$$

Now make use of the formulas

$$I(a,b) \equiv \iint_{-\infty}^{\infty} e^{-a(x^2+y^2)-bx}\,dx\,dy = \pi a^{-1}e^{b^2/4a}, \qquad (13.11)$$

$$\iint_{-\infty}^{\infty} e^{-a(x^2+y^2)-bx}(x^2+y^2)\,dx\,dy = -\frac{\partial I}{\partial a} = \frac{\pi}{a}\left(\frac{1}{a}+\frac{b^2}{4a^2}\right)e^{b^2/4a}, \qquad (13.12)$$

$$\iint_{-\infty}^{\infty} e^{-a(x^2+y^2)-bx}x\,dx\,dy = -\partial I/\partial b = -(\tfrac{1}{2}\pi b/a^2)e^{b^2/4a}.$$

After some tedious but routine simplification, one gets

$$J^* = -8\sinh\chi_r\operatorname{sech}^3\tfrac{1}{2}\chi(2\beta^*/\pi)^{-5/2}e^{-(\beta+\sigma)} \times [4\beta\cos^2\tfrac{1}{2}\Theta\cosh^2\tfrac{1}{2}\chi + \sin^2\tfrac{1}{2}\Theta - (\beta\sigma\sin^2\Theta/\beta^*)\sinh^2\tfrac{1}{2}\chi] \times \exp[-2(\sigma^*\beta/\beta^*)\sinh^2\tfrac{1}{2}\chi]. \qquad (13.13)$$

Making the substitutions (12.16), we have finally

$$J^* = 16\pi r_\mu p^\mu(2\beta/\pi)^{-\frac{3}{2}}\operatorname{sech}^3\tfrac{1}{2}\chi \times \exp\{-(\beta+\sigma) - 2\beta[s^*/(1-s^*)]\sinh^2\tfrac{1}{2}\chi\} \times \left(\frac{1-s}{1-s^*}\right)^{5/2}\Big[\cos^2\tfrac{1}{2}\Theta\Big(1+\frac{1-s}{1-s^*}\sinh^2\tfrac{1}{2}\chi\Big) + \tfrac{1}{4}\beta^{-1}\sin^2\tfrac{1}{2}\Theta\Big]. \qquad (13.14)$$

The expression for $'J^*$ follows from this by replacing s^* by $'s^*$, and Θ by $\pi - \Theta$.

This completes the evaluation of the integrals involved in (13.3). Multiplying this equation through by $(1-s)^{-5/2}e^{\sigma}$, we find, with the definitions (12.18) and (12.20),

$$\mathfrak{L}[\Psi_{\frac{3}{2}}(\tau,s)r_\mu p^\mu] = 16\pi(2\beta/\pi)^{-\frac{3}{2}}\alpha e^{-\beta}\operatorname{sech}^3\tfrac{1}{2}\chi\cdot r_\mu p^\mu \times \int_0^\pi [\cdots]\Gamma(\Theta)\sin\Theta\,d\Theta, \qquad (13.15)$$

where

$$[\cdots] = \Psi_{\frac{3}{2}}(\tau,s) - \Psi_{\frac{3}{2}}(\tau,s^*)\cos^2\tfrac{1}{2}\Theta - \Psi_{\frac{3}{2}}(\tau,'s^*)\sin^2\tfrac{1}{2}\Theta + \tfrac{1}{2}(\tau/\beta)[\Psi_{\frac{3}{2}}(\tau,s) - \Psi_{\frac{3}{2}}(\tau,s^*) - \Psi_{\frac{3}{2}}(\tau,'s^*) + \Psi_{5/2}(\tau,s^*)\sin^2\tfrac{1}{2}\Theta + \Psi_{5/2}(\tau,'s^*) \times \cos^2\tfrac{1}{2}\Theta] + \tfrac{1}{4}\beta^{-1}[1 - \Psi_{\frac{3}{2}}(\tau,s^*)\sin^2\tfrac{1}{2}\Theta - \Psi_{\frac{3}{2}}(\tau,'s^*)\cos^2\tfrac{1}{2}\Theta]. \qquad (13.16)$$

If we differentiate n times with respect to s, and

note (12.21), this becomes

$$\mathfrak{L}[L_n^{\frac{3}{2}}(\tau)r_\mu p^\mu] = (2\pi)^{5/2}\beta^{-3/2}\alpha e^{-\beta}(1 + \tfrac{1}{2}\tau/\beta)^{-3/2}r_\mu p^\mu \times \int_0^\pi Q_n(\tau, \Theta)\Gamma(\Theta)\sin\Theta\, d\Theta, \qquad (13.17)$$

with

$$\begin{aligned} Q_0(\tau, \Theta) &= 0, \\ Q_n(\tau, \Theta) &= [1 - C^{2n+2} - S^{2n+2} \\ &\quad + (\tau/\beta)(1 - C^{2n} - S^{2n}) \\ &\quad - \tfrac{1}{4}\beta^{-1}(S^2C^{2n} + C^2S^{2n})]L_n^{\frac{3}{2}}(\tau) \\ &\quad + \tfrac{1}{2}(\tau/\beta)(S^2C^{2n} + C^2S^{2n})L_n^{5/2}(\tau) \\ &\qquad (n = 1, 2, \cdots). \end{aligned} \qquad (13.18)$$

We have abbreviated

$$C = \cos\tfrac{1}{2}\Theta, \qquad S = \sin\tfrac{1}{2}\Theta.$$

From the recursion formulas for the Laguerre polynomials, we have

$$\begin{aligned} \tau L_n^{5/2}(\tau) &= (1 - \Lambda)[(n + \tfrac{5}{2})L_n^{\frac{3}{2}}(\tau) - (n + 1)L_n^{\frac{3}{2}}(\tau)] \\ &\quad + \Lambda[(n + \tfrac{3}{2})L_{n-1}^{\frac{3}{2}}(\tau) - (n - \tau)L_n^{\frac{3}{2}}(\tau)] \\ &\qquad (n = 1, 2, \cdots), \end{aligned}$$

for arbitrary Λ. Substitute this in (13.18), choosing Λ so as to remove the terms involving τ as a factor:

$$\Lambda = -(1 - C^{2n} - S^{2n})/(S^2C^{2n} + C^2S^{2n}).$$

We thus reach the "pseudoeigenvalue equation"

$$\begin{aligned} \mathfrak{L}[L_n^{\frac{3}{2}}(\tau)r_\mu p^\mu] &= (1 + \tfrac{1}{2}\tau/\beta)^{-\frac{3}{2}}r_\mu p^\mu\{E_{n+1}L_n^{\frac{3}{2}}(\tau) \\ &\quad + \beta^{-1}[(\tfrac{1}{2}n + 1)E_{n+1} + \tfrac{1}{4}(2n + 1)E_n]L_n^{\frac{3}{2}}(\tau) \\ &\quad - \beta^{-1}\tfrac{1}{2}(n + 1)E_{n+1}L_{n+1}^{\frac{3}{2}}(\tau) - \beta^{-1}\tfrac{1}{2}(n + \tfrac{3}{2})E_nL_{n-1}^{\frac{3}{2}}(\tau)\} \\ &\qquad (n = 1, 2, \cdots), \end{aligned} \qquad (13.19)$$

where E_n is defined by (12.23). In the classical limit, this reduces to the orthodox eigenvalue equation

$$\mathfrak{L}[L_n^{\frac{3}{2}}(\tau)r_\mu p^\mu] = E_{n+1}L_n^{\frac{3}{2}}(\tau)r_\mu p^\mu \qquad (\beta = \infty),$$

in agreement with known results.[14]

For general β, (13.19) may be used to solve the integral equation

$$\mathfrak{L}[G(\tau)r_\mu p^\mu] = F(\tau)r_\mu p^\mu, \qquad (13.20)$$

by a method to be explained presently.

Observe first that the five equations (cf. Sec. 8)

$$\int N_0F(\tau)r_\mu p^\mu\, d\omega = 0, \qquad \int N_0F(\tau)r_\mu p^\mu p^\nu\, d\omega = 0$$

express the necessary and sufficient conditions for the existence of a solution to (13.20); they are equivalent to

$$(1, G(\tau)r_\mu p^\mu) = 0, \qquad (p^\nu, G(\tau)r_\mu p^\mu) = 0.$$

Only one of them is nontrivial:

$$\int N_0F(\tau)(r_\mu p^\mu)^2\, d\omega = 0. \qquad (13.21)$$

Choose the third and fourth axes of a Minkowskian frame parallel to r^μ and u^μ, respectively, and set

$$p^\mu = [\sinh\chi\sin\theta\cos\varphi, \sinh\chi\sin\theta\sin\varphi, \sinh\chi\cos\theta, \cosh\chi].$$

Then, by virtue of (2.3) and (12.20),

$$\begin{aligned} \beta\, d\omega &= (2\tau/\beta)^{\frac{1}{2}}(1 + \tfrac{1}{2}\tau/\beta)^{\frac{1}{2}}\, d\tau\sin\theta\, d\theta\, d\varphi, \qquad (13.22) \\ (r_\mu p^\mu)^2 &= (2\tau/\beta)(1 + \tfrac{1}{2}\tau/\beta)\cos^2\theta. \end{aligned}$$

Note also

$$N_0 = \alpha e^{-(\beta+\tau)}.$$

The solubility condition (13.21) can therefore be written

$$\int_0^\infty (1 + \tfrac{1}{2}\tau/\beta)^{\frac{3}{2}}\tau^{\frac{3}{2}}F(\tau)e^{-\tau}\, d\tau = 0. \qquad (13.21')$$

It will be assumed that the given function $F(\tau)$ satisfies this condition.

Expand $F(\tau)$ according to

$$(1 + \tfrac{1}{2}\tau/\beta)^{\frac{3}{2}}F(\tau) = \sum_{n=0}^\infty F_nL_n^{\frac{3}{2}}(\tau), \qquad (13.23)$$

so that

$$\Gamma(n + \tfrac{5}{2})F_n = n!\int_0^\infty (1 + \tfrac{1}{2}\tau/\beta)^{\frac{3}{2}}\tau^{\frac{3}{2}} \times F(\tau)L_n^{\frac{3}{2}}(\tau)e^{-\tau}\, d\tau. \qquad (13.24)$$

From (13.21'), it follows that

$$F_0 = 0. \qquad (13.25)$$

We try the following expansion for the unknown function $G(\tau)$:

$$G(\tau) = \sum_{n=1}^\infty G_nL_n^{\frac{3}{2}}(\tau) + (\text{arb const}). \qquad (13.26)$$

Substituting (13.26) into (13.20), applying (13.19), and equating coefficients of $L_n^{\frac{3}{2}}(\tau)$, yields the following three-term recursion formula for the coefficients G_n:

$$\begin{aligned} \beta(F_n - E_{n+1}G_n) &= [(\tfrac{1}{2}n + 1)E_{n+1} \\ &\quad + \tfrac{1}{4}(2n + 1)E_n]G_n - \tfrac{1}{2}nE_nG_{n-1} \\ &\quad - \tfrac{1}{2}(n + \tfrac{5}{2})E_{n+1}G_{n+1} \quad (n = 1, 2, \cdots). \end{aligned} \qquad (13.27)$$

The first of these relations (note $E_1 = 0$),

$$\beta(F_1 - E_2 G_1) = E_2(\tfrac{3}{2}G_1 - \tfrac{7}{4}G_2),$$

involves *two* unknown coefficients, so it is not clear how to proceed in general. However, if β is moderately large, we can take the classical solution

$$G_n = F_n/E_{n+1} \qquad (\beta = \infty)$$

as a first approximation. Substituting this into the right-hand side of (13.2), we obtain the next approximation

$$G_n \approx \frac{F_n}{E_{n+1}} + \frac{1}{\beta}\left\{\left[(\tfrac{1}{2}n + 1) + \tfrac{1}{4}(2n+1)\frac{E_n}{E_{n+1}}\right]\frac{F_n}{E_{n+1}} - \tfrac{1}{2}n\frac{F_{n-1}}{E_{n+1}} - \tfrac{1}{2}(n + \tfrac{5}{2})\frac{F_{n+1}}{E_{n+1}}\right\}. \tag{13.28}$$

This iterative procedure can obviously be continued, and leads to an asymptotic expansion for G_n in inverse powers of β. We can expect this asymptotic approximation to be satisfactory if (i) to a given order of approximation, $O(\beta^{-N})$, the series (13.23) is truncated, and (ii) $N/\beta \ll 1$. The second condition would require an upper bound of about $10^{10\,\circ}$ for the temperature of an electron gas.

14. Bulk Viscosity

The scalar eigenfunctions obtained in Sec. 12 will now be employed to solve the integral equation (8.23) for Maxwellian particles, and thus derive an explicit expression for the bulk viscosity.

We have to solve

$$\mathfrak{L}[A_1(\tau, \beta)] = B(\tau, \beta), \tag{14.1}$$

where, because of (8.21) and (12.20), the function B can be expressed as

$$B(\tau, \beta) = \beta^{-1}(\gamma - \tfrac{4}{3})\tau^2 + [2(\gamma - \tfrac{4}{3}) - \eta(\gamma - 1) + \beta^{-1}\gamma]\tau + \gamma - \beta(\eta - 1)(\gamma - 1). \tag{14.2}$$

Let us expand

$$(1 + \tfrac{1}{2}\tau/\beta)^{\frac{1}{2}}B(\tau, \beta) = \sum_{n=0}^{\infty} B_n L_n^{\frac{1}{2}}(\tau), \tag{14.3}$$

so that

$$\Gamma(n + \tfrac{3}{2})B_n = n!\int_0^\infty (1 + \tfrac{1}{2}\tau/\beta)^{\frac{1}{2}}\tau^{\frac{1}{2}} \times B(\tau)L_n^{\frac{1}{2}}(\tau)e^{-\tau}\,d\tau. \tag{14.4}$$

The solubility conditions of (14.1),

$$\int N_0 B\,d\omega = 0, \qquad \int N_0 B u_\mu p^\mu\,d\omega = 0$$

[cf (8.9) and (8.9′)], yield

$$B_0 = 0, \qquad B_1 = 0, \tag{14.5}$$

as can also be verified directly from (14.4) and (14.2).

Our eigenvalue equation (12.22) now shows that the solution of (14.1) is

$$A_1(\tau, \beta) = \sum_{n=2}^{\infty} E_n^{-1}B_n L_n^{\frac{1}{2}}(\tau) + (\text{arb const}). \tag{14.6}$$

We substitute the two series (14.3), (14.6) into the expression (9.20) for the bulk viscosity κ, and make use of (13.22) and the orthogonality relations

$$n!\int_0^\infty L_m^\alpha(\tau)L_n^\alpha(\tau)\tau^\alpha e^{-\tau}\,d\tau = \Gamma(n + \alpha + 1)\,\delta_{mn}. \tag{14.7}$$

The result is

$$\kappa = \frac{mc}{2\pi\beta}\sum_{n=2}^{\infty}\frac{3\cdot 5\cdots(2n+1)}{2\cdot 4\cdots 2n}\frac{B_n^2}{e_n}, \tag{14.8}$$

where [cf. (12.23)]

$$e_n \equiv \int_0^\pi (1 - \cos^{2n}\tfrac{1}{2}\Theta - \sin^{2n}\tfrac{1}{2}\Theta) \times \Gamma(\Theta)\sin\Theta\,d\Theta. \tag{14.9}$$

Equation (14.8) reduces to a particularly simple form under "moderate" relativistic conditions (large β). From (6.36) and (6.35) we easily obtain the expansions

$$\eta = 1 + 5/2\beta + 15/8\beta^2 - 15/8\beta^3 + \cdots, \qquad \gamma = \tfrac{5}{3}(1 - \beta^{-1} + 4\beta^{-2} + \cdots). \tag{14.10}$$

Hence (14.2) yields

$$B(\tau) = \beta^{-1}(\tfrac{1}{3}\tau^2 - \tfrac{5}{3}\tau + \tfrac{5}{4}) + O(\beta^{-2}) = \tfrac{2}{3}\beta^{-1}L_2^{\frac{1}{2}}(\tau) + O(\beta^{-2}). \tag{14.11}$$

Correct to the lowest nonvanishing order, we therefore have

$$B_2 = \tfrac{2}{3}\beta^{-1}, \qquad B_n = 0 \qquad (n \neq 2),$$

and our expression for the bulk viscosity reduces to

$$\kappa = 5mc/12\pi\beta^3 e_2 \tag{14.12}$$

This is of the order of $(1/\beta^2)$ times the first coefficient of viscosity ν. Under moderate relativistic conditions—in the case of an electron gas, up to temperatures of about $10^{9\,\circ}$—the bulk viscosity can be neglected entirely.

15. Thermal Conductivity

We proceed to the solution of the integral equation (8.24), using the method outlined in Sec. 13.

Our integral equation may be written

$$\mathfrak{L}[A_2(\tau, \beta)r_\mu p^\mu] = F(\tau, \beta)r_\mu p^\mu, \qquad (15.1)$$

where r_μ is a fixed vector normal to the 4-velocity u_μ, and

$$F(\tau, \beta) \equiv \eta^{-1}(1 + \tau/\beta) - 1. \qquad (15.2)$$

One easily verifies, either from (8.9′) or by direct substitution, that $F(\tau)$ satisfies the solubility condition (13.21′). This implies the feasability of the expansion (13.23) with $F_0 = 0$.

In what follows, we retain only the first-order relativistic correction, $O(\beta^{-1})$. (Higher-order terms may be generated with no difficulty). Using (14.10), we find

$$\begin{aligned}(1 + \tfrac{1}{2}\tau/\beta)^{\frac{3}{2}}F(\tau) &= -\beta^{-1}(\tfrac{5}{2} - \tau) + \tfrac{3}{2}\beta^{-2}(\tfrac{1}{2}\tau^2 - \tfrac{35}{12}\tau + \tfrac{35}{12}) + O(\beta^{-3}) \\ &= -\beta^{-1}L_1^{\frac{3}{2}}(\tau) + \tfrac{3}{2}\beta^{-2}[L_2^{\frac{3}{2}}(\tau) - \tfrac{7}{12}L_1^{\frac{3}{2}}(\tau)] + O(\beta^{-3}).\end{aligned} \qquad (15.3)$$

Then (13.28) yields, for the coefficients G_n in the Laguerre expansion of $G(\tau) \equiv A_2(\tau)$,

$$E_2G_1 = -\beta^{-1} - \tfrac{19}{8}\beta^{-2} + O(\beta^{-3}),$$

$$E_3G_2 = \tfrac{5}{2}\beta^{-2} + O(\beta^{-3}), \qquad E_{n-1}G_n = O(\beta^{-3}) \quad (n = 3, 4, \cdots).$$

Substituting into (9.14), and using (13.22) and (14.7), we finally obtain for the thermal conductivity,

$$k\lambda = \tfrac{4}{3}\pi m^2c^5\beta^{-1}\alpha e^{-\beta}\eta \int_0^\infty (1 + \tfrac{1}{2}\tau/\beta)^{\frac{3}{2}} \times (2\tau/\beta)^{\frac{3}{2}}F(\tau)G(\tau)e^{-\tau}\, d\tau,$$

i.e.,

$$\begin{aligned}\lambda &= \frac{8\sqrt{2}\,\pi}{3}\frac{m^2c^5}{k}\,\eta\beta^{-5/2}\sum_{n=1}^{\infty}\frac{\Gamma(n + \frac{5}{2})}{n!}F_nG_n \\ &= \frac{5}{4\pi}\frac{k^2}{mc}\,\eta T^3\,\frac{1}{e_2}\left[1 + \frac{13}{4\beta} + O\left(\frac{1}{\beta^2}\right)\right].\end{aligned} \qquad (15.4)$$

In lowest approximation, this agrees with the classical result.[15] It should perhaps be noted that our definition of thermal conductivity [see (9.16)] differs from the conventional definition by what reduces to a factor of T^2 in the classical limit.

APPENDIX: PROOF OF (3.6)

Since a direct calculation of the Jacobian is very laborious, we proceed geometrically as follows.

On the unit pseudosphere in 4-momentum space, consider two points with Minkowskian coordinates p^μ, $'p^\mu$, and pseudopolar coordinates [see (2.2)] (χ, θ, φ), $(\chi', \theta', \varphi')$. The radius midway between them is $\bar{p}^\mu$, and the vector $g^\mu = {}'p^\mu - p^\mu$, orthogonal to $\bar{p}^\mu$, represents the joining chord. Let ϑ, ϵ be polar angles giving the direction of g^μ in the spacelike 3-flat orthogonal to $\bar{p}^\mu$.

We permit p^μ, $'p^\mu$ to vary independently, tracing out 3-areas $d\omega$, $d'\omega$ [see (2.3)] on the unit pseudosphere. It is geometrically evident that these variations are related to the corresponding variations in $\bar{p}^\mu$, g^μ by

$$d\omega\, d'\omega = A\, d\bar{\omega}g^2\, dg \sin\vartheta\, d\vartheta\, d\epsilon, \qquad (A1)$$

in which the undetermined factor A can depend on g only; $A = A(g)$.

Let χ, χ' assume pure imaginary values, setting $\chi = i\chi_1$, $\chi' = i\chi_1'$. Then

$$p^k = i\,p_1^k, \quad p^4 = p_1^4, \qquad g^k = i\,g_1^k, \quad g^4 = g_1^4, \qquad (k = 1, 2, 3),$$

where

$$p_1^\mu = (\sin\chi_1 \sin\theta\cos\varphi, \sin\chi_1 \sin\theta \sin\varphi, \sin\chi_1\cos\theta, \cos\chi_1),$$

and all quantities with suffix 1 are real. The pseudosphere

$$-1 = p^\mu p_\mu = -\sum_{k=1}^{3}(p_1^k)^2 - (p_1^4)^2$$

can now be regarded as a closed unit sphere S: $0 \le \chi_1 \le \pi$, $0 \le \theta \le \pi$, $0 \le \varphi < 2\pi$, in the *Euclidean* 4-space of the cartesian vectors p_1^μ. We have

$$g_1^2 \equiv \sum_{k=1}^{3}(g_1^k)^2 + (g^4)^2 = -g_\mu g^\mu = -g^2,$$

and

$$i\, d\omega = \sin^2\chi_1 \sin\theta\, d\chi_1\, d\theta\, d\varphi \equiv d\omega_1.$$

(A1) becomes

$$d\omega_1\, d'\omega_1 = A(i\,g_1)\, d\bar{\omega}_1 g_1^2\, dg_1 \sin\vartheta\, d\vartheta\, d\epsilon. \qquad (A2)$$

To find A, we evaluate (in two different ways) the integral

$$I \equiv \iint_S g_1^{-2}F(g_1)\, d\omega_1\, d'\omega_1,$$

where $F(g)$ is an arbitrary function, and both integrations are performed over the unit sphere S.

In the integration over $'p_1^\mu$, choose coordinates in

[15] Reference 14, p. 383.

which $\chi_1 = 0$, so that $p^\alpha = (0, 0, 0, 1)$, $g_1 = 2 \sin \frac{1}{2}\chi_1'$. We find

$$I = \int_S \int_0^\pi \int_0^{2\pi} d\omega_1 \sin\theta' \, d\theta' \, d\varphi' \times \int_0^\pi \tfrac{1}{4} \operatorname{cosec}^2 \tfrac{1}{2}\chi_1' F(g_1) \sin^2 \chi_1' \, d\chi_1'$$

$$= 8\pi^3 \int_0^2 F(g_1)(1 - \tfrac{1}{4}g_1^2)^{\frac{1}{2}} \, dg_1. \tag{A3}$$

On the other hand, by virtue of (A2),

$$I = \int_S \int_0^\pi \int_0^{2\pi} d\bar{\omega}_1 \sin\vartheta \, d\vartheta \, d\epsilon \times \int_0^2 F(g_1) A(i\, g_1) \, dg_1$$

$$= 8\pi^3 \int_0^2 F(g_1) A(i\, g_1) \, dg_1. \tag{A4}$$

Comparing (A3) and (A4), we see that

$$A(g) = (1 + \tfrac{1}{4}g^2)^{\frac{1}{2}}.$$

Inserting this into (A1), we arrive at (3.6).

Kinetic Equations for Electrons and Phonons*

AMIRAM RON†
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey
(Received 11 April 1963)

A method is presented for deriving a set of kinetic equations for a system of electrons and phonons in a simplified model of a metal. By employing the second quantization representation for the creation and annihilation operators of the electrons and the phonons, an hierarchy of equations for the distribution functions and correlation functions is introduced. This hierarchy is studied, using an approach originally developed by Bogoliubov, where both truncation of the hierarchy and irreversibility are achieved under general assumptions. A set of kinetic equations is obtained for an homogeneous system, where the electrons dynamically shield both each other and the phonons.

I. INTRODUCTION

ATTEMPTS have been made in recent years to study the approach to equilibrium and transport properties of electron–phonon system by means of kinetic equations for *both* the electrons and the phonons. Klimontrovich and Ternko[1] derived classical and semiclassical kinetic equations starting from the Fröhlich Hamiltonian.[2] They have written the Liouville equation for the system and have derived from it an hierarchy of equations for the distribution functions. By employing Bogoliubov's[3] method for the truncation of the hierarchy and for the introduction of irreversibility they have obtained Folker–Planck-type kinetic equations. An approach very close to reference 1, but purely quantum mechanical, was reported by Gurzhi.[4] Both references 1 and 4 have *excluded* electron–electron interactions. Another treatment of the electron–phonon kinetic equations was given by Pines and Schrieffer.[5] They have started from the Bohm–Pines[6] Hamiltonian and have treated the system under consideration as composed of electrons, phonons and plasmons. They have *postulated* kinetic equations by adopting the method of *transition probabilities*.[7,8] An interesting paper devoted to the derivation of kinetic equation for the electrons *only* in an electron–phonon system was reported by Mori.[9] Here the irreversibility and the truncation have been introduced by Kirkwood's method.[10]

The purpose of the present paper is to derive a *set* of kinetic equations for *both* the electrons and the phonons in a simplified model of a metal, described by the Hamiltonian of Bardeen and Pines.[11] We attempt to construct a quantum mechanical kinetic theory on the basis of assumptions more general than those employed in previous works.

We start with the Bardeen–Pines[11] Hamiltonian written in the second quantization representation of both the electrons and the phonons. From here we derive an *hirerarchy* of equations for the quantized density operators and correlation operators without resorting to the Liouville equation, but instead employ the Heisenberg equations of motion for the operators. Distribution functions are introduced taking into account the statistics of the particles. This procedure departs from the derivation of the hierarchy as was given by references 1 and 4; rather, it follows the spirit of Mori's paper,[9] but generalizes it to obtain an hierarchy of equations.

The hierarchy is truncated in a *systematic* manner by employing Bogoliubov method to obtain a *closed* set of equations for the distribution functions and the correlation functions. The truncation scheme rests on the assumption that the more particles involved in the correlation, the weaker the correlation function. In this formulation one does not have to introduce a separate description for the

* This work was performed under the auspices of the United States Atomic Energy Commission, Contract AT(30-1)-1238.
† On leave of absence from Technion–Israel Institute of Technology, Haifa, Israel. Present address: Department of Physics, University of California, San Diego, La Jolla, California.
[1] Y. Klimontovich and S. V. Temko, Zh. Eksperim. i Teor. Fiz. **35**, 1141 (1958) [English transl.; Soviet Phys.—JETP **8**, 799 (1959)].
[2] H. Frohlich, Proc. Roy. Soc. (London) **A215**, 291 (1952).
[3] N. N. Bogoliubov, *Studies of Statistical Mechanics*, edited by J. deBoer and G. E. Uhlenbeck (North-Holland Publishing Company, Amsterdam, 1962).
[4] R. N. Gurzhi, Zh. Eksperim. i Teor. Fiz. **33**, 451 (1957) [English transl.: Soviet Phys.—JETP **6**, 352 (1958)].
[5] D. Pines and J. R. Schriffer, Phys. Rev. **125**, 804 (1962).
[6] D. Bohm and D. Pines, Phys. Rev. **92**, 609 (1953).
[7] R. E. Peierls, *Quantum Theory of Solids* (Oxford University Press, New York, 1955).
[8] J. M. Ziman, *Electrons and Phonons*, (Oxford University Press, New York, 1960).
[9] H. Mori, Progr. Theoret. Phys. (Kyoto) **9**, 473 (1953).
[10] J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946); **15**, 72 (1947).
[11] J. Bardeen and D. Pines, Phys. Rev. **99**, 1140 (1955); J. Bardeen, in *Encyclopedia of Physics*, edited by S. Flügge (Julius Springer-Verlag, Berlin, 1956), Vol. XV, p. 274.

plasmons,[8] as the collective features of the system are embedded in the theory.

The irreversibility is introduced in the spirit of Bogoliubov[3]; that is, we assume that correlation functions relax to their *asymptotic* long-time value in a time short compared to the time scale of the distribution functions.[12] This assumption is shared with all other theories of electron–phonon interaction, e.g., the transition probability approach, Kirkwood[10] time average, Van Hove[13] transition singularities, etc. One of the advantages of the present scheme is that this multiple time-scale assumption can be abandoned.[14]

The kinetic equations *derived* from the hierarchy in this paper differ from those obtained by other authors in that they contain *dynamic* shielding of the interactions.

II. FORMULATION OF THE PROBLEM

In this section we first review briefly the Bardeen–Pines Hamiltonian[11] to make clear what assumptions are involved and then introduce the hierarchy and show how the truncation of it comes about.

A. Electron–Phonon Hamiltonian

Following Bardeen and Pines,[11] we confine ourselves to a simplified model of a metal. We assume a monoatomic crystal of N ions of valence Z and mass M, and ZN valence electrons with mass m and charge $-e$. We now introduce phonon coordinates to represent ion motion, assuming that the lattice waves are either longitudinal or transverse, and that the electrons interact only with the longitudinal phonons. This approximation is valid only for long wavelength, but is incorrect for short wavelength. Furthermore, we make another assumption which is a result of the Bardeen–Pines treatment, namely, that the phonon frequencies are renormalized. This amounts to the assumption that the renormalization processes are completed before any other processes are considered, and it is consistent with our treatment of the system to be homogenous in space. The Hamiltonian

$$H = H_e + H_{ph} + H_{ep} \tag{1}$$

is composed of three parts, the electron H_e, the phonon H_p, and the electron–phonon interaction H_{ep}. In the second quantization representation we have for the electrons

$$H_e = \sum_s \int d\mathbf{r}\psi_s^+(\mathbf{r})\left[-\frac{1}{2m}\frac{\partial^2}{\partial \mathbf{r}^2} + V(\mathbf{r})\right]\psi_s(\mathbf{r}) + \frac{1}{2}\sum_{s,s'}\int d\mathbf{r}\,d\mathbf{r}'\psi_s^+(\mathbf{r})\psi_{s'}^+(\mathbf{r}') \times \phi(\mathbf{r}-\mathbf{r}')\psi_{s'}(\mathbf{r}')\psi_s(\mathbf{r}). \tag{2}$$

$\psi_s^+(\mathbf{r})$ and $\psi_s(\mathbf{r})$ are, respectively, the creation and annihilation operators of the electrons with spin s and position $\mathbf{r}$, satisfying the anticommutation relations

$$[\psi_s(\mathbf{r}), \psi_{s'}^+(\mathbf{r}')]_+ = \psi_s(\mathbf{r})\psi_{s'}^+(\mathbf{r}') + \psi_{s'}^+(\mathbf{r}')\psi_s(\mathbf{r}) = \delta_{ss'}\delta(\mathbf{r}-\mathbf{r}')$$
$$[\psi_s(\mathbf{r}), \psi_{s'}(\mathbf{r}')]_+ = [\psi_s^+(\mathbf{r}), \psi_{s'}^+(\mathbf{r}')]_+ = 0, \tag{3}$$

with the usual notations for δ Kronecker and δ Dirac. $V(\mathbf{r})$ stands for the effective potential in which the electrons move, taking into account the equilibrium positions of the ions compensated by a uniform negative charge. The Coulomb potential between the electrons is denoted by

$$\phi(r) = e^2/r. \tag{4}$$

We also choose $\hbar = 1'$.

The phonons Hamiltonian is

$$H_p = \sum_{\mathbf{q}} \omega_q(b_{\mathbf{q}}^+ b_{\mathbf{q}} + \tfrac{1}{2}), \tag{5}$$

where ω_q is the renormalized frequency of the longitudinal phonon with wave vector $\mathbf{q}$, and it depends only on the absolute value of $\mathbf{q}$. $b_{\mathbf{q}}^+$ and $b_{\mathbf{q}}$ are respectively the creation and annihilation operators of a $\mathbf{q}$ phonon, and they satisfy the commutation relations

$$[b_{\mathbf{q}}, b_{\mathbf{q}'}^+] = b_{\mathbf{q}}b_{\mathbf{q}'}^+ - b_{\mathbf{q}'}^+ b_{\mathbf{q}} = \delta_{\mathbf{q}\mathbf{q}'},$$
$$[b_{\mathbf{q}}, b_{\mathbf{q}'}] = [b_{\mathbf{q}'}^+, b_{\mathbf{q}}^+] = 0. \tag{6}$$

The $\mathbf{q}$ summation in Eq. (5) is carried out only over the first Brillouin zone and the transverse phonons are not considered here.

The interaction Hamiltonian is given by

$$H_{ep} = \sum_{\mathbf{q}} b_{\mathbf{q}}\alpha_{\mathbf{q}} \cdot \sum_s \int d\mathbf{r}\psi_s^+(\mathbf{r})\mathbf{E}(r)\psi_s(\mathbf{r}) + \sum_{\mathbf{q}} b_{\mathbf{q}}^+\alpha_{\mathbf{q}} \cdot \sum_s \int d\mathbf{r}\psi_s^+(\mathbf{r})\mathbf{E}^*(\mathbf{r})\psi_s(\mathbf{r}), \tag{7}$$

where

$$\alpha_{\mathbf{q}} = \alpha_{-\mathbf{q}} = -(2MN\omega_q)^{-\frac{1}{2}}\hat{q} \tag{8}$$

stands for the polarization, $\hat{q}$ is a unit vector in the direction of the wave, and

$$\mathbf{E}_{\mathbf{q}}(r) = \frac{\partial}{\partial \mathbf{r}}\sum_{\mathbf{R}_i} v(\mathbf{r}-\mathbf{R}_i)e^{-i\mathbf{q}\cdot\mathbf{R}_i} \tag{9}$$

[12] R. L. Guernsey, Phys. Rev. **127**, 1446 (1962). E. A. Frieman, J. Math. Phys. **4**, 410 (1963).
[13] L. Van Hove, Physica **21**, 517 (1955); **23**, 441 (1957); **25**, 268 (1959).
[14] R. L. Guernsey, Phys. Fluids **5**, 322 (1962).

$v(\mathbf{r} - \mathbf{R}_i)$ being the effective interaction between an electron in $\mathbf{r}$ and an ion in equilibrium position $\mathbf{R}_i$. In Eq. (7) the $\mathbf{q}$ summation is over all $\mathbf{q}$'s, but with the usual interpretation for $b_\mathbf{q}$ and $b_\mathbf{q}^+$ to stay in the first Brillouin zone.

We shall later use a "momentum"-type representation for the electrons, and it is convenient to introduce it here. The Bloch equation for the one-electron functions is

$$[-(1/2m)(\partial^2/\partial\mathbf{r}^2) + V(\mathbf{r})]\phi_\mathbf{k}(\mathbf{r}) = E(\mathbf{k})\phi_\mathbf{k}(\mathbf{r}), \tag{10}$$

where $\phi_\mathbf{k}(\mathbf{r})$ enjoy the following properties:

$$\int d\mathbf{r}\phi^*_{\mathbf{k}'}(\mathbf{r})\phi_\mathbf{k}(\mathbf{r}) = \delta_{\mathbf{k}\mathbf{k}'},$$
$$\sum_\mathbf{k} \phi^*_\mathbf{k}(\mathbf{r}')\phi_\mathbf{k}(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}'), \tag{11}$$

and

$$\phi^*_\mathbf{k}(\mathbf{r}) = \phi_{-\mathbf{k}}(\mathbf{r}). \tag{12}$$

We describe the electrons in the extended zone scheme with $\mathbf{k}$ running also out of the first Brillouin zone. The creation and annihilation operators for an electron in the state $\mathbf{k}$, $a^+_{\mathbf{k}s}$, and $a_{\mathbf{k}s}$, are defined by

$$\psi_s(\mathbf{r}) = \sum_\mathbf{k} \phi_\mathbf{k}(\mathbf{r})a_{\mathbf{k}s}, \tag{13}$$

and

$$\psi^+_s(\mathbf{r}) = \sum_\mathbf{k} \phi^*_\mathbf{k}(\mathbf{r})a^+_{\mathbf{k}s}, \tag{14}$$

and thus satisfy

$$[a_{\mathbf{k}s}, a^+_{\mathbf{k}'s'}]_+ = \delta_{ss'}\delta_{\mathbf{k}\mathbf{k}'},$$
$$[a_{\mathbf{k}s}, a_{\mathbf{k}'s'}]_+ = [a^+_{\mathbf{k}s}, a_{\mathbf{k}'s'}]_+ = 0. \tag{15}$$

The Coulomb interaction between the electrons [the second term in Eq. (2)] may be expressed in the form

$$H_C = \tfrac{1}{2}\sum_\mathbf{k}{}' \phi(k) \sum_{\substack{\mathbf{p},\mathbf{p}'\\ s,s'}} a^+_{\mathbf{p}+\frac{1}{2}\mathbf{k},s} \times a^+_{\mathbf{p}'-\frac{1}{2}\mathbf{k},s'}a_{\mathbf{p}'+\frac{1}{2}\mathbf{k},s'}a_{\mathbf{p}-\frac{1}{2}\mathbf{k},s}, \tag{16}$$

where

$$\phi(k) = \int d\mathbf{r}\, d\mathbf{r}'\phi^*_{\mathbf{p}+\frac{1}{2}\mathbf{k}}(\mathbf{r})\phi^*_{\mathbf{p}'-\frac{1}{2}\mathbf{k}}(\mathbf{r}') \times \phi(\mathbf{r} - \mathbf{r}')\phi_{\mathbf{p}'+\frac{1}{2}\mathbf{k}}(\mathbf{r}')\phi_{\mathbf{p}-\frac{1}{2}\mathbf{k}}(\mathbf{r}), \tag{17}$$

and we have assumed that the matrix elements depend only on the wave-vector difference between initial and final states, and similar to the free-electron case ($\phi(k) = 4\pi e^2/k^2$), $\phi(k)$ depends only on the absolute value of $\mathbf{k}$. These assumptions are in the spirit of the Bloch theory and amount essentially to neglect of small inhomogeneities. This implies that an electron in the $\mathbf{k}$ state is treated much like a free electron with the same wave vector. We apply the same arguments to obtain

$$H_{ep} = \sum_\mathbf{q} v_\mathbf{q}[b_\mathbf{q} + b^+_{-\mathbf{q}}] \sum_{s,\mathbf{p}} a^+_{\mathbf{p}+\frac{1}{2}\mathbf{q},s}a_{\mathbf{p}-\frac{1}{2}\mathbf{q},s}, \tag{18}$$

where

$$v_\mathbf{q} = \int d\mathbf{r}\phi^*_{\mathbf{p}+\frac{1}{2}\mathbf{q}}(\mathbf{r})\boldsymbol{\alpha}_\mathbf{q}\cdot\mathbf{E}_\mathbf{q}(\mathbf{r})\phi_{\mathbf{p}-\frac{1}{2}\mathbf{q}}(\mathbf{r}) \tag{19}$$

depends only on $\mathbf{q}$, and

$$v^*_{-\mathbf{q}} = v_\mathbf{q}. \tag{20}$$

In Eq. (18) $\mathbf{q}$ runs over all values and is not restricted to the first Brillouin zone, while $b_\mathbf{q}$ and $b^+_\mathbf{q}$ *are* restricted to the reduced $\mathbf{q}$ vectors in the first zone. By this scheme we guarantee the inclusion of Peierls–Umklapp processes.

B. Electron–Phonon Hierarchy

For shorthand notation we introduce

$$u_\mathbf{q}(\mathbf{r}) = \boldsymbol{\alpha}_\mathbf{q}\cdot\mathbf{E}_\mathbf{q}(\mathbf{r}). \tag{21}$$

The equations of motion for the creation and annihilation operators in the Heisenberg representation read

$$\frac{\partial\psi_s(\mathbf{r})}{\partial t} = -\frac{1}{i}\left[\frac{1}{2m}\frac{\partial^2}{\partial\mathbf{r}^2} - V(\mathbf{r})\right]\psi_s(\mathbf{r}) + \frac{1}{i}\sum_{s'}\int d\mathbf{r}'\phi(\mathbf{r} - \mathbf{r}')\psi^+_{s'}(\mathbf{r}')\psi_{s'}(\mathbf{r}')\psi_s(\mathbf{r}) + \frac{1}{i}\sum_\mathbf{q}[u_\mathbf{q}(\mathbf{r})b_\mathbf{q} + u^*_\mathbf{q}(\mathbf{r})b^+_\mathbf{q}]\psi_s(\mathbf{r}), \tag{22a}$$

$$\frac{\partial\psi^+_s(\mathbf{r})}{\partial t} = \frac{1}{i}\left[\frac{1}{2m}\frac{\partial^2}{\partial\mathbf{r}^2} - V(\mathbf{r})\right]\psi^+_s(\mathbf{r}) - \frac{1}{i}\psi^+_s(\mathbf{r})\sum_{s'}\int d\mathbf{r}'\phi(\mathbf{r} - \mathbf{r}')\psi^+_{s'}(\mathbf{r}')\psi_{s'}(\mathbf{r}') - \frac{1}{i}\sum_\mathbf{q}(u_\mathbf{q}(\mathbf{r})b_\mathbf{q} + u^*_\mathbf{q}(\mathbf{r})b^+_\mathbf{q})\psi^+_s(\mathbf{r}), \tag{22b}$$

$$\frac{\partial b_\mathbf{q}}{\partial t} = -i\omega_\mathbf{q}b_\mathbf{q} + \frac{1}{i}\sum_s\int d\mathbf{r}\, u^*_\mathbf{q}(\mathbf{r})\psi^+_s(\mathbf{r})\psi_s(\mathbf{r}), \tag{22c}$$

and

$$\frac{\partial b^+_\mathbf{q}}{\partial t} = i\omega_\mathbf{q}b^+_\mathbf{q} - \frac{1}{i}\sum_s\int d\mathbf{r}\, u_\mathbf{q}(\mathbf{r})\psi^+_s(\mathbf{r})\psi_s(\mathbf{r}). \tag{22d}$$

In Eqs. (22), the time t is suppressed for simplicity, as we deal with all the operators at the *same* time.

The one-electron density matrix is defined by

$$F_s(1, 1') = F_s(\mathbf{r}, \mathbf{r}'; t) = \mathrm{tr}\{\psi^+_s(\mathbf{r}', t)\psi_s(\mathbf{r}, t)D\} = \langle\psi^+_s(\mathbf{r}', t)\psi_s(\mathbf{r}, t)\rangle, \tag{23}$$

where D is the density matrix of the whole system. In the Heisenberg representation, D is independent of time, and the time dependence of $\psi_s^+(\mathbf{r})$ and $\psi_s(\mathbf{r})$ is determined by Eqs. (22a) and (22b). In the same way, one defines the two-electron density matrix by

$$F_{ss'}(1, 2; 1', 2') = \langle \psi_s^+(\mathbf{r}_1', t)\psi_{s'}^+(\mathbf{r}_2', t)\psi_{s'}(\mathbf{r}_2, t)\psi_s(\mathbf{r}_1, t)\rangle, \quad (24)$$

and so forth, and thus constructs an hierarchy of density matrices for the electrons. The same procedure is applied to the phonons. But as we are interested here only in homogeneous (but not necessarily isotropic) systems, we define the phonon distribution function, i.e., the number of phonons of wave vector $\mathbf{q}$, rather than the density matrix:

$$n(\mathbf{q}, t) = \langle b_\mathbf{q}^+(t)b_\mathbf{q}(t)\rangle. \quad (25)$$

One can proceed to define higher-order phonon distribution functions in the same manner, but our theory is not concerned with these functions. We now define mixed density matrices for both electrons and phonons. The most important ones are given by

$$G_s^{(1)}(1, 1'; \mathbf{q}) = \langle b_\mathbf{q}(t)\psi_s^+(\mathbf{r}_1', t)\psi_s(\mathbf{r}_1, t)\rangle, \quad (26a)$$

and

$$G_s^{(2)}(1, 1'; \mathbf{q}) = \langle b_\mathbf{q}^+(t)\psi_s^+(\mathbf{r}_1', t)\psi_s(\mathbf{r}_1, t)\rangle. \quad (26b)$$

These functions represent correlations between electrons and phonons and play an important role in the following.

The equations of motion for the functions defined above by Eqs. (23)–(26), are now obtained by employing Eqs. (22). We first introduce the shorthand formulas

$$T_i = \frac{1}{i}\left\{\frac{1}{2m}\left(\frac{\partial^2}{\partial \mathbf{r}_i^2} - \frac{\partial^2}{\partial \mathbf{r}_i'^2}\right) - [V(\mathbf{r}_i) - V(\mathbf{r}_i')]\right\}, \quad (27)$$

$$w_{ij} = \frac{1}{i}[\phi(\mathbf{r}_i - \mathbf{r}_j) - \phi(\mathbf{r}_i' - \mathbf{r}_j')],$$
$$\bar{w}_{ij} = \frac{1}{i}[\phi(\mathbf{r}_i - \mathbf{r}_j) - \phi(\mathbf{r}_i' - \mathbf{r}_j)], \quad (28)$$

and

$$U_i(\mathbf{q}) = \frac{1}{i}[u_\mathbf{q}(\mathbf{r}_i) - u_\mathbf{q}(\mathbf{r}_i')]. \quad (29)$$

We obtain the following equations:

(1) For the one-electron density matrix:

$$\left(\frac{\partial}{\partial t} + T_1\right)F_s(1, 1') = \sum_{s'}\int d\mathbf{r}_2\bar{w}_{12}F_{ss'}(1, 2; 1', 2) + \sum_\mathbf{q}\{U_i(\mathbf{q})G_s^{(1)}(1, 1'; \mathbf{q}) - U_i^*(\mathbf{q})G_s^{(2)}(1, 1'; \mathbf{q})\}. \quad (30)$$

(2) For the phonon distribution function:

$$\frac{\partial n(\mathbf{q})}{\partial t} = -\frac{1}{i}\sum_s\int d\mathbf{r}_1[u_\mathbf{q}(\mathbf{r}_1)G_s^{(1)}(1, 1'; \mathbf{q}) - u_\mathbf{q}^*(\mathbf{r}_1)G_s^{(2)}(1, 1'; \mathbf{q})], \quad (31)$$

(3) For the two-electron density matrix:

$$\left(\frac{\partial}{\partial t} + T_1 + T_2 - w_{12}\right)F_{ss'}(1, 2; 1', 2') = \sum_{s''}\int d\mathbf{r}_3(\bar{w}_{13} + \bar{w}_{23})F_{ss's''}(1, 2, 3; 1', 2', 3) + \sum_\mathbf{q}[U_1(\mathbf{q}) + U_2(\mathbf{q})]\langle b_\mathbf{q}\psi_s^+(\mathbf{r}_1')\psi_{s'}^+(\mathbf{r}_2')\psi_{s'}(\mathbf{r}_2)\psi_s(\mathbf{r}_1)\rangle - \sum_\mathbf{q}[U_1^*(\mathbf{q}) + U_2^*(\mathbf{q})] \times \langle b_\mathbf{q}^+\psi_s^+(\mathbf{r}_1')\psi_{s'}^+(\mathbf{r}_2')\psi_{s'}(\mathbf{r}_2)\psi_s(\mathbf{r}_1)\rangle, \quad (32)$$

(4) For the mixed electron–phonon density matrix:

$$\left(\frac{\partial}{\partial t} + i\omega_q + T_1\right)G_s^{(1)}(1, 1'; \mathbf{q}) = \sum_{s'}\int d\mathbf{r}_2u_\mathbf{q}^*(\mathbf{r}_2)F_{ss'}(1, 2; 1', 2) + \frac{1}{i}u_\mathbf{q}^*(\mathbf{r}_1)F_s(1, 1') + \sum_{s'}\int d\mathbf{r}_2\bar{w}_{12} \times \langle b_\mathbf{q}\psi_s^+(\mathbf{r}_1')\psi_{s'}^+(\mathbf{r}_2')\psi_{s'}(\mathbf{r}_2)\psi_s(\mathbf{r}_1)\rangle + \sum_{\mathbf{q}'}U_1(\mathbf{q})\langle b_\mathbf{q}b_{\mathbf{q}'}\psi_s^+(\mathbf{r}_1')\psi_s(\mathbf{r}_1)\rangle - \sum_{\mathbf{q}'}U_1^*(\mathbf{q})\langle b_\mathbf{q}b_{\mathbf{q}'}^+\psi_s^+(\mathbf{r}_1')\psi_s(\mathbf{r}_1)\rangle, \quad (33a)$$

and

$$\left(\frac{\partial}{\partial t} - i\omega_q + T_1\right)G_s^{(2)}(1, 1'; q) = -\sum_{s'}\int d\mathbf{r}_2u_\mathbf{q}(\mathbf{r}_2)F_{ss'}(1, 2; 1', 2) - \frac{1}{i}u_\mathbf{q}^*(\mathbf{r}_1)F_s(1, 1') + \sum_{s'}\int d\mathbf{r}_2\bar{w}_{12} \times \langle b_\mathbf{q}^+\psi_s^+(\mathbf{r}_1')\psi_{s'}^+(\mathbf{r}_2')\psi_{s'}(\mathbf{r}_2)\psi_s(\mathbf{r}_1)\rangle + \sum_{\mathbf{q}'}U_1(\mathbf{q})\langle b_\mathbf{q}^+b_{\mathbf{q}'}\psi_s^+(\mathbf{r}_1')\psi_s(\mathbf{r}_1)\rangle - \sum_{\mathbf{q}'}U_1^*(\mathbf{q})\langle b_\mathbf{q}b_{\mathbf{q}'}\psi_s^+(\mathbf{r}_1')\psi_s(\mathbf{r}_1)\rangle. \quad (33b)$$

Equations (30)–(33) represent the lowest members of an hierarchy of equations. Each member is given in terms of higher members, i.e., in terms of functions of more particles correlated in a more

complicated way. It is a hopeless task to solve this hierarchy exactly, and thus one must resort to approximate methods in order to truncate this hierarchy and obtain a closed set of equations from which kinetic equations can be obtained.

Before we turn to the discussion of our scheme of approximations, we introduce the symmetry requirements on the functions studied in Eqs. (30)–(33). It is convenient to introduce an antisymmetrization operator for the electrons (we do not discuss the phonon symmetry, because in the present theory, we restrict ourselves to one-phonon functions):

$$\gamma_n = \prod_{j=2}^{n}\left[1 - \sum_{k=1}^{j-1}\delta_{s_j s_k}P_{j,k}\right], \tag{34}$$

where $P_{j,k}$ permutes the variables $\mathbf{r}_j$ and $\mathbf{r}_k$. The γ_n satisfies the relations

$$\gamma_{n+1} = \gamma_n\left[1 - \sum_{j=1}^{s}\delta_{s_j,s_{n+1}}P_{j,n+1}\right], \tag{35}$$

and commutes with the operators T, w, U of Eqs. (27)–(29). Now we define new functions by means of γ_n:

$$\begin{aligned} F_s(1,1') &= \gamma_1 f_s(1,1') = f_s(1,1'),\\ G_s^{(i)}(\mathbf{1},1',\mathbf{q}) &= \gamma_1 g_s^{(i)}(1,1',\mathbf{q})\\ &= g_s^{(i)}(1,1',\mathbf{q}), \qquad i = 1, 2,\\ F_{ss'}(1,2;1',2') &= \gamma_2 f_{ss'}(1,2;1',2'), \end{aligned} \tag{36}$$

and so on. If we now substitute these relations into Eqs. (30)–(33) we obtain a set of equations for the new functions f_s, $f_{ss'}$, g_s, etc. In the following we are concerned with the properly symmetrized functions thus obtained.

C. Truncation of the Hierarchy

We employ here a generalized version of a method originally developed by Bogoliubov,[3] which has proved successful in the classical theory of interacting particles[14,15] and also for degenerate Coulomb gas.[12] The method rests on the observation that, for large class of initial conditions, the intrinsic correlation functions between particles can be treated as small, compared to the product of one-particle distribution functions. For example, if one defines the two-body correlation function $g(1, 2)$ by the equation

$$f_2(1,2) = f_1(1)f_1(2) + g(1,2),$$

where $f_2(1, 2)$ is the joint distribution function for two particles and $f_1(1)$ is the one-particle distribution function, then $g(1, 2)$ can be treated as small compared to $f_1(1)f_1(2)$. The argument is that $g(1, 2)$ is roughly proportional to the ratio between the average potential energy and the average kinetic energy per particle, and dies out when the two particles are far apart (roughly when the separation is of the order of the range of the interparticle potential or larger). From the previous discussion one concludes that the more particles involved in the correlation, the smaller the magnitude of this correlation function.

To carry out this program, we start with the definition of the two-electron correlation function $g_{ss'}(1, 2; 1', 2')$ by

$$\begin{aligned} f_{ss'}(1,2;1',2') &= f_s(1,1')f_{s'}(2,2')\\ &\quad + g_{ss'}(1,2;1',2'), \end{aligned} \tag{37}$$

and we thus make the assumption that g is a first-order quantity while f_s is a zero-order quantity. We are concerned here only with quantities of the zero and first order, therefore, we write

$$\begin{aligned} f_{ss's''}(1,2;3;1',2',3') &\approx f_s(1,1')f_{s'}(2,2')f_{s''}(3,3')\\ &\quad + f_s(1,1')g_{s's''}(2,3;2',3')\\ &\quad + f_{s'}(2,2')g_{ss''}(1,3;1',3')\\ &\quad + f_{s''}(3,3')g_{ss'}(1,2;1',2'). \end{aligned} \tag{38}$$

We also consider the phonon distribution function, $n(\mathbf{q})$, as zero-order quantity, and the phonon–electron functions $g^{(1)}$ and $g^{(2)}$ as first-order quantities (the latter functions cannot be factorized and, therefore, are correlation-type functions). If we make use of the fact that the phonons are not coupled in the Hamiltonian, and that the only way they can be correlated is via the electrons, we can write, to first order,

$$\begin{aligned} &\langle b_{\mathbf{q}}\psi_s^+(\mathbf{r}_1')\psi_{s'}^+(\mathbf{r}_2')\psi_{s'}(\mathbf{r}_2)\psi_s(\mathbf{r}_1)\rangle\\ &\quad\approx \gamma_2\{g_s^{(1)}(1,1',\mathbf{q})f_{s'}(2,2') + g_{s'}^{(1)}(2,2';\mathbf{q})f_s(1,1')\},\\ &\langle b_{\mathbf{q}}^+\psi_s^+(\mathbf{r}_1')\psi_{s'}^+(\mathbf{r}_2')\psi_{s'}(\mathbf{r}_2)\psi_s(\mathbf{r}_1)\rangle\\ &\quad\approx \gamma_2\{g_s^{(2)}(1,1',\mathbf{q})f_{s'}(2,2') + g_{s'}^{(2)}(2,2';\mathbf{q})f_s(1,1')\}, \end{aligned} \tag{39a}$$

and

$$\langle b_{\mathbf{q}'}^+ b_{\mathbf{q}}\psi_s^+(\mathbf{r}_1')\psi_s(\mathbf{r}_1)\rangle \approx \delta_{\mathbf{q}\mathbf{q}'}n(\mathbf{q})f_s(1,1'), \tag{39b}$$

and all the other functions involve two-phonon and two-electron operators and are of higher order.

It should be pointed out that although we assume the electron–electron interaction to be small, the *long range* of the Coulomb potential produces

[15] T. H. Dupree, Phys. Fluids 4, 696 (1961).

"shielding" effects and thus "self-consistent" terms (where there is an integral over the Coulomb potential) are *not* small. This does not apply, of course, to the "exchange" terms where the range of the force is reduced (to the order of de Broglie wavelength). The electron–phonon interaction is also considered to be small, but terms where integrals over $u_q(\mathbf{r})$ are involved are not small, due to the fact that there are many electrons in one wavelength of the phonons.

Thus, for homogeneous systems, we obtain the following equations:

(1) The distribution functions satisfy

$$\frac{\partial}{\partial t} f_s(1, 1') = \sum_{s'} \int d\mathbf{r}_2 \bar{w}_{12} g_{ss'}(1, 2; 1', 2) + \sum_{\mathbf{q}} \{U_i(\mathbf{q}) g_s^{(1)}(1, 1'; \mathbf{q}) - U_i^*(\mathbf{q}) g_s^{(2)}(1, 1'; \mathbf{q})\}, \tag{40a}$$

and

$$\frac{\partial n(\mathbf{q})}{\partial t} = -\frac{1}{i} \sum_{s} \int d\mathbf{r}_1 \{u_{\mathbf{q}}(\mathbf{r}_1) g_s^{(1)}(1, 1'; \mathbf{q}) - u_{\mathbf{q}}^*(\mathbf{r}_1) g_s^{(2)}(1, 1'; \mathbf{q})\}. \tag{40b}$$

(2) The correlation functions satisfy

$$\left(\frac{\partial}{\partial t} + T_1 + T_2\right) g_{ss'}(1, 2; 1', 2') = w_{12} f_s(1, 1') f_{s'}(2, 2') - \sum_{s''} \int d\mathbf{r}_3 [\bar{w}_{13} P_{23} \delta_{s's''} + \bar{w}_{23} \delta_{ss''} P_{13}] f_s(1, 1') f_{s'}(2, 2') f_{s''}(3, 3) + \sum_{s''} \int d\mathbf{r}_3 \{\bar{w}_{13} g_{s's''}(2, 3; 2', 3) f_s(1, 1') + \bar{w}_{23} g_{ss''}(1, 3; 1', 2) f_{s'}(2, 2')\}, \tag{41}$$

$$\left(\frac{\partial}{\partial t} + i\omega_q + T_1\right) g_s^{(1)}(1, 1', \mathbf{q}) = \frac{1}{i} \{u_{\mathbf{q}}^*(\mathbf{r}_1)[n(\mathbf{q}) + 1] - u_{\mathbf{q}}^*(\mathbf{r}_1') n(\mathbf{q})\} f_s(1, 1') - \frac{1}{i} \int d\mathbf{r}_2 u_{\mathbf{q}}^*(\mathbf{r}_2) f_s(1, 2) f_s(2, 1') + \sum_{s'} \int d\mathbf{r}_2 \bar{w}_{12} g_{s'}^{(1)}(2, 2; \mathbf{q}) f_s(1, 1'), \tag{42a}$$

and

$$\left(\frac{\partial}{\partial t} - i\omega_q + T_1\right) g_s^{(2)}(1, 1'; \mathbf{q}) = \frac{1}{i} \{u_{\mathbf{q}}(\mathbf{r}_1) n(\mathbf{q}) - u_{\mathbf{q}}(\mathbf{r}_1')[1 + n(\mathbf{q})]\} f_s(1, 1') - \frac{1}{i} \int d\mathbf{r}_2 u_{\mathbf{q}}(\mathbf{r}_2) f(1, 2) f(2, 1') + \sum_{s'} \int d\mathbf{r}_2 \bar{w}_{12} g_{s'}^{(2)}(2, 2; \mathbf{q}) f_s(1, 1'). \tag{42b}$$

The remainder of the report is devoted to the solution of these equations, leading to kinetic equations for the electrons and the phonons. We wish also to point out that Eqs. (41) and (42) are not coupled, because we have assumed ω_q to represent the renormalized frequencies of the phonons, and then restricted ourselves to homogeneous systems.

III. KINETIC EQUATIONS

In the present section we solve Eqs. (41) and (42) in terms of the electron and phonon distribution functions $f(\mathbf{p})$ and $n(\mathbf{k})$ and substitute the time-asymptotic solutions of the correlation functions into Eqs. (40a) and (40b) to obtain the kinetic equations. The justification of this substitution is based on the *physical* assumption, which is originally due to Bogoliubov,[3] that the correlation functions resulting from "collision" reach asymptotic values in time which is very short when compared with the time over which the distribution functions vary. This procedure makes the final form of Eqs. (40a) and (40b) irreversible.

It is now convenient to employ a "momentum" representation for the electrons. We introduce the following functions:

(1) The electron distribution function

$$f_s(\mathbf{p}) = \int d\mathbf{r}_1 \, d\mathbf{r}_1' \phi_{\mathbf{p}}(\mathbf{r}_1') \phi_{\mathbf{p}}^*(\mathbf{r}_1) f_s(\mathbf{r}_1, \mathbf{r}_1'), \tag{43}$$

where, in our special homogeneous case, $f_s(\mathbf{r}_1, \mathbf{r}_1') = f_s(\mathbf{r}_1 - \mathbf{r}_1')$, and

$$f_s(1, 1') = \sum_{\mathbf{p}} \phi_{\mathbf{p}}^*(\mathbf{r}_1') \phi_{\mathbf{p}}(\mathbf{r}_1) f_s(\mathbf{p}); \tag{44}$$

we can also write

$$f_s(\mathbf{p}) = \langle a_{\mathbf{p}s}^+ a_{\mathbf{p}s} \rangle, \tag{45}$$

with the normalization (assuming that the ions are singly ionized)

$$N = \sum_{s} \int d\mathbf{p} f_s(\mathbf{p});$$

(2) The electron–electron correlation function

$$g_{ss'}(\mathbf{p}, \mathbf{p}'; \mathbf{k}) = \int d\mathbf{r}_1 \, d\mathbf{r}_2 \, d\mathbf{r}_1' \, d\mathbf{r}_2' \phi_{\mathbf{p}+\frac{1}{2}\mathbf{k}}(\mathbf{r}_1') \phi_{\mathbf{p}'-\frac{1}{2}\mathbf{k}}(\mathbf{r}_2') \times \phi_{\mathbf{p}'+\frac{1}{2}\mathbf{k}}^*(\mathbf{r}_2) \phi_{\mathbf{p}-\frac{1}{2}\mathbf{k}}^*(\mathbf{r}_1) g_{ss'}(1, 2; 1', 2'); \tag{46}$$

(3) The electron–phonon correlation functions

$$g_s^{(1)}(\mathbf{p},\mathbf{q}) = \int d\mathbf{r}_1\, d\mathbf{r}_1' \phi_{\mathbf{p}+\frac{1}{2}\mathbf{q}}(\mathbf{r}_1') \times \phi^*_{\mathbf{p}-\frac{1}{2}\mathbf{q}}(\mathbf{r}_1) g_s^{(1)}(1,1';\mathbf{q}), \quad (47a)$$

$$g_s^{(2)}(\mathbf{p},\mathbf{q}) = \int d\mathbf{r}_1'\, d\mathbf{r}_1 \phi_{\mathbf{p}-\frac{1}{2}\mathbf{q}}(\mathbf{r}_1') \times \phi^*_{\mathbf{p}+\frac{1}{2}\mathbf{q}}(\mathbf{r}_1) g_s^{(2)}(1,1';\mathbf{q}). \quad (47b)$$

If we now use Eqs. (17), (19), and (20), we can transform Eqs. (40)–(42) to the form:

(1) The electron distribution function

$$\frac{\partial f_s(\mathbf{p})}{\partial t} = \left[\frac{\partial f_s(\mathbf{p})}{\partial t}\right]_{\text{electron}} + \left[\frac{\partial f_s(\mathbf{p})}{\partial t}\right]_{\text{phonon}}, \quad (48a)$$

$$\left[\frac{\partial f_s(\mathbf{p})}{\partial t}\right]_{\text{electron}} = -i \sum_{\mathbf{k}}{}' \phi(k) \times \sum_{\mathbf{p}',s'} \{g_{ss'}(\mathbf{p}-\tfrac{1}{2}\mathbf{k},\mathbf{p}';\mathbf{k}) - g_{ss'}(\mathbf{p}+\tfrac{1}{2}\mathbf{k},\mathbf{p}';\mathbf{k})\}, \quad (48b)$$

and

$$\left[\frac{\partial f_s(\mathbf{p})}{\partial t}\right]_{\text{phonon}} = -i \sum_{\mathbf{q}} u_{\mathbf{q}}[g_s^{(1)}(\mathbf{p}-\tfrac{1}{2}\mathbf{q},\mathbf{q}) - g_s^{(1)}(\mathbf{p}+\tfrac{1}{2}\mathbf{q},\mathbf{q})] + i \sum_{\mathbf{q}} u^*_{\mathbf{q}}[g_s^{(2)}(\mathbf{p}-\tfrac{1}{2}\mathbf{q},\mathbf{q}) - g_s^{(2)}(\mathbf{p}+\tfrac{1}{2}\mathbf{q},\mathbf{q})]; \quad (48c)$$

(2) The phonon distribution function

$$\frac{\partial n(\mathbf{q})}{\partial t} = i \sum_{\mathbf{p},s} [u_{\mathbf{q}} g_s^{(1)}(\mathbf{p},\mathbf{q}) - u^*_{\mathbf{q}} g_s^{(2)}(\mathbf{p},\mathbf{q})]; \quad (49)$$

(3) The electron–electron correlation function

$$\left[\frac{\partial}{\partial t} - i\Delta(\mathbf{p},\mathbf{k}) - i\Delta(\mathbf{p}',-\mathbf{k})\right] g_{ss'}(\mathbf{p},\mathbf{p}',\mathbf{k}) + i\phi(k) H_{s'}(\mathbf{p}_1' - \mathbf{k}) \sum_{\mathbf{p}'',s''} g_{ss''}(\mathbf{p},\mathbf{p}'';\mathbf{k}) + i\phi(k) H_s(\mathbf{p},\mathbf{k}) \sum_{\mathbf{p}'',s''} g_{s's''}(\mathbf{p}',\mathbf{p}'',-\mathbf{k}) = -iI_{ss'}(\mathbf{p},\mathbf{p}';\mathbf{k}), \quad (50)$$

where we have introduced the notations

$$\Delta(\mathbf{p},\mathbf{k}) = E(\mathbf{p}+\tfrac{1}{2}\mathbf{k}) - E(\mathbf{p}-\tfrac{1}{2}\mathbf{k}), \quad (51)$$

$$H_s(\mathbf{p},\mathbf{k}) = f_s(\mathbf{p}+\tfrac{1}{2}\mathbf{k}) - f_s(\mathbf{p}-\tfrac{1}{2}\mathbf{k}), \quad (52)$$

and

$$I_{ss'}(\mathbf{p},\mathbf{p}';\mathbf{k}) = \{F_s(\mathbf{p},\mathbf{k})F_{s'}(\mathbf{p}',-\mathbf{k}) - F_s(\mathbf{p},-\mathbf{k})F_{s'}(\mathbf{p}',\mathbf{k})\}\phi(\mathbf{k}), \quad (53)$$

with

$$F_a(\mathbf{p},\mathbf{k}) = f_s(\mathbf{p}\pm\tfrac{1}{2}\mathbf{k})[1 - f_s(\mathbf{p}\mp\tfrac{1}{2}\mathbf{k})]. \quad (54)$$

Finally, for the electron–phonon correlations functions, we obtain

$$\left[\frac{\partial}{\partial t} + i\omega_q - i\Delta(\mathbf{p},\mathbf{q})\right] g_s^{(1)}(\mathbf{p},\mathbf{q}) + i\phi(q) H_s(\mathbf{p},\mathbf{q}) \sum_{\mathbf{p}',s'} g_{s'}^{(1)}(\mathbf{p}',\mathbf{q}) = -iu^*_{\mathbf{q}} n(\mathbf{q}) H_s(\mathbf{p},\mathbf{q}) - iu^*_{\mathbf{q}} f_s(\mathbf{p}+\tfrac{1}{2}\mathbf{q})[1 - f_s(\mathbf{p}-\tfrac{1}{2}\mathbf{q})], \quad (55a)$$

and

$$\left[\frac{\partial}{\partial t} - i\omega_q + i\Delta(\mathbf{p},\mathbf{q})\right] g_s^{(2)}(\mathbf{p},\mathbf{q}) - i\phi(q) H_s(\mathbf{p},\mathbf{q}) \times \sum_{\mathbf{p}',s'} g_s^{(2)}(\mathbf{p}',\mathbf{q}) = +iu_{\mathbf{q}} n(\mathbf{q}) H_s(\mathbf{p},\mathbf{q}) + iu_{\mathbf{q}} f_s(\mathbf{p}+\tfrac{1}{2}\mathbf{q})[1 - f_s(\mathbf{p}-\tfrac{1}{2}\mathbf{q})]. \quad (55b)$$

Our program is now to solve Eqs. (50) and (55) in terms of $f_s(\mathbf{p})$ and $n(\mathbf{q})$, assuming that the latter functions are not changing in time, and then substitute the asymptotic time solutions into (48) and (49) to obtain the kinetic equations.

A. Electron–Electron Correlation Function

Equation (50) is very similar to the equation one meets in the theory of electron gas[14,16] and has been solved by various authors using different methods; therefore, we only sketch briefly one method of solution which is due to Dupree[15] (also see Wolff[17]). We define a time-independent operator

$$H_s(\mathbf{p},\mathbf{k}) = -i\Delta(\mathbf{p},\mathbf{k}) + i\phi(k) H_s(\mathbf{p},\mathbf{k}) \sum_{\mathbf{p},s}. \quad (56)$$

Equation (50) is now given by

$$[\partial/\partial t + H_s(\mathbf{p},\mathbf{k}) + H_{s'}(\mathbf{p}',-\mathbf{k})] g_{ss'}(\mathbf{p},\mathbf{p}';\mathbf{k}\cdot t) = -iI_{ss'}(\mathbf{p},\mathbf{p}';\mathbf{k}). \quad (57)$$

From Eq. (57) one sees that $H_s(\mathbf{p},\mathbf{k})$ operates on $\mathbf{p}$ only, and $H_{s'}(\mathbf{p}',-\mathbf{k})$ operates on $\mathbf{p}'$ only, and thus they commute with each other. If we now introduce an operator $R_s(\mathbf{p},\mathbf{k},t)$ by the equation

$$(\partial/\partial t)R_s(\mathbf{p},\mathbf{k},t) + H_s(\mathbf{p},\mathbf{k})R_s(\mathbf{p},\mathbf{k},t) = 0, \quad (58)$$

with $R_s(\mathbf{p},\mathbf{k};0) = 1$, then

$$g_{ss'}(\mathbf{p},\mathbf{p}';\mathbf{k}) = -i\int_0^\infty d\tau R_s(\mathbf{p},\mathbf{k};\tau) \times R_{s'}(\mathbf{p}',-\mathbf{k};\tau) I_{ss'}(\mathbf{p},\mathbf{p}';\mathbf{k}), \quad (59)$$

and the problem is reduced to that of solving Eq. (58) for $R_s(\mathbf{p},\mathbf{k})$. The solution of Eqs. (58)

[16] H. W. Wyld, Jr., and B. D. Fried (to be published).
[17] P. A. Wolff, Phys. Fluids 5, 316 (1962).

is readily

$$R_s(\mathbf{p},\mathbf{k},\tau)h_s(\mathbf{p},\mathbf{k}) = \int_{-\infty}^{\infty} d\omega e^{-i\omega\tau}\frac{1}{\Delta(\mathbf{p},\mathbf{k})+\omega+i\epsilon}\left\{h_s(\mathbf{p},\mathbf{k}) + \frac{\phi(k)}{\epsilon(\mathbf{k},\omega)}H_s(\mathbf{p},\mathbf{k})\sum_{\mathbf{p},s}\frac{h_s(\mathbf{p},\mathbf{k})}{\Delta(\mathbf{p},\mathbf{k})+\omega+i\epsilon}\right\}, \quad (60)$$

where $\epsilon(\mathbf{k},\omega)$, the "dielectric function", is defined by

$$\epsilon(\mathbf{k},\omega) = 1 - \phi(k)\sum_{\mathbf{p},s}\frac{H_s(\mathbf{p},\mathbf{k})}{\Delta(\mathbf{p},\mathbf{k})+\omega+i\epsilon}, \quad (61)$$

and $h_s(\mathbf{p},\mathbf{k})$ is any function of $\mathbf{p}$, $\mathbf{k}$, and s. Finally we obtain

$$\left[\frac{\partial f_s(\mathbf{p})}{\partial t}\right]_{\text{electron}} = 2\pi\sum_{\mathbf{k},\mathbf{p}',s}\left|\frac{\phi(k)}{\epsilon[\mathbf{k},E(\mathbf{p})-E(\mathbf{p}+\mathbf{k})]}\right|^2 \times\{f(\mathbf{p})f(\mathbf{p}')[1-f(\mathbf{p}+\mathbf{k})][1-f(\mathbf{p}'-\mathbf{k})] - f(\mathbf{p}+\mathbf{k})f(\mathbf{p}'-\mathbf{k})[1-f(\mathbf{p})][1-f(\mathbf{p}')]\} \times\delta\{E(\mathbf{p}+\mathbf{k})-E(\mathbf{p})+E(\mathbf{p}'-\mathbf{k})-E(\mathbf{p}')\}. \quad (62)$$

In order to obtain this result we have made the assumption that the electron distribution function, $f(\mathbf{p})$, is such that the zeros of $\epsilon(\mathbf{k},\omega)$ all lie in the upper half-plane, i.e., that there are no unstable oscillations.

Equation (62) represents the contribution of the direct electron–electron interaction to the collision integral of the kinetic equation for the electrons. It amounts both to the individual and the collective aspects of the electron interactions, so that the electron–electron potential is dynamically shielded by the simultaneous motion of all the electrons of the system. The "collision integral" of Eq. (62) is the same as that of a quantum electron plasma[12,16] as was expected on the basis of the model employed for the metal. Thus, as far as the electrons are concerned, the higher their density, the better the approximation we have made.

B. Electron–Phonon Correlation Functions

The solution of Eqs. (55) is somewhat simpler than that of Eq. (50). We first take a one-sided Fourier transform in time of Eq. (55a):

$$[-i\omega+\epsilon+i\omega_q-i\Delta(\mathbf{p},\mathbf{q})]g_s^{(1)}(\mathbf{p},\mathbf{q};\omega) + i\phi(q)H_s(\mathbf{p},\mathbf{q})\sum_{\mathbf{p}',s'}g_{s'}^{(1)}(\mathbf{p}',\mathbf{q};\omega) = -\frac{iu_q^*}{-i\omega+\epsilon}\{n(\mathbf{q})H_s(\mathbf{p},\mathbf{q}) + f_s(\mathbf{p}+\tfrac{1}{2}\mathbf{q})[1-f_s(\mathbf{p}-\tfrac{1}{2}\mathbf{q})]\}, \quad (63)$$

where the initial values of $g^{(1)}$ are ignored, due to the fact that they do not contribute to the asymptotic value of $g^{(1)}$ to be used in the kinetic equations. We have also introduced a positive small factor ϵ to choose the path of integration later. To obtain the time assymptotic value of $g_s^{(1)}(\mathbf{p},\mathbf{q})$ we solve Eq. (63) and apply the theorem

$$g_s^{(1)}(\mathbf{p},\mathbf{q}) = \lim_{\omega\to 0}(-i\omega)g_s^{(1)}(\mathbf{p},\mathbf{q},\omega). \quad (64)$$

The treatment of Eq. (55b) in the limit of Eq. (64) is similar to that of Eq. (55a). We find that $g^{(2)}$ is the complex conjugate of $g^{(1)}$:

$$[g_s^{(1)}(\mathbf{p},\mathbf{q})]^* = g_s^{(2)}(\mathbf{p},\mathbf{q}). \quad (65)$$

If we use this relation in Eqs. (48) and (49), we obtain

$$\left[\frac{\partial f_s(\mathbf{p})}{\partial t}\right]_{\text{ph}} = -2\,\text{Im}\{\sum_{\mathbf{q}}u_{\mathbf{q}}[g_s^{(1)}(\mathbf{p}+\tfrac{1}{2}\mathbf{q};\mathbf{q}) - g_s^{(1)}(\mathbf{p}-\tfrac{1}{2}\mathbf{q},\mathbf{q})]\}, \quad (66)$$

and

$$\frac{\partial n(\mathbf{q})}{\partial t} = -\text{Im}\{\sum_{\mathbf{p},s}u_{\mathbf{q}}g_s^{(1)}(\mathbf{p},\mathbf{q})\}, \quad (67)$$

where Im (f) denotes the imaginary part of f.

It is instructive to solve Eq. (63) in the limit of Eq. (64) in two steps. First we assume that the self-consistent term (screening term) $i\phi(q)H_s(\mathbf{p},\mathbf{q})$, $\sum_{\mathbf{p}',s'}g_s^{(1)}(\mathbf{p}',\mathbf{q},\omega)$ is negligible. In that case we obtain

$$g_s^{(1)}(p,q) = u_{\mathbf{q}}^*n(\mathbf{q})\frac{H_s(\mathbf{p},\mathbf{q})}{\Delta(\mathbf{p},\mathbf{q})-\omega_q+i\epsilon} + u_{\mathbf{q}}^*\frac{f_s(\mathbf{p}+\tfrac{1}{2}\mathbf{q})[1-f_s(\mathbf{p}-\tfrac{1}{2}\mathbf{q})]}{\Delta(\mathbf{p},\mathbf{q})-\omega_q+i\epsilon}, \quad (68)$$

and upon substitution in Eqs. (66) and (67), we find

$$\left[\frac{\partial f_s(\mathbf{p})}{\partial t}\right]_{\text{ph}} = (2\pi)\sum_{\mathbf{q}}|u_{\mathbf{q}}|^2 n(\mathbf{q}) \times\{[f_s(\mathbf{p}+\mathbf{q})-f_s(\mathbf{p})]\delta[E(\mathbf{p}+\mathbf{q})-E(\mathbf{p})-\omega_q] - [f_s(\mathbf{p})-f_s(\mathbf{p}-\mathbf{q})]\delta[E(\mathbf{p})-E(\mathbf{p}-\mathbf{q})-\omega_q] + (2\pi)\sum_{\mathbf{q}}|u_{\mathbf{q}}|^2\{f_s(\mathbf{p}+\mathbf{q}) \times[1-f_s(\mathbf{p})]\delta[E(\mathbf{p}+\mathbf{q})-E(\mathbf{p})-\omega_q] - f_s(\mathbf{p})[1-f_s(\mathbf{p}-\mathbf{q})]\delta[E(\mathbf{p})-E(\mathbf{p}-\mathbf{q})-\omega_q]\}, \quad (69)$$

and

$$\frac{\partial n(\mathbf{q})}{\partial t} = |u_{\mathbf{q}}|^2 n(\mathbf{q})\sum_{\mathbf{p},s}[f_s(\mathbf{p}+\mathbf{q})-f_s(\mathbf{p})] \times\delta[E(\mathbf{p}+\mathbf{q})-E(\mathbf{p})-\omega_q] + |u_{\mathbf{q}}|^2\sum_{\mathbf{p},s}f_s(\mathbf{p}+\mathbf{q})$$

$$\times [1 - f_s(\mathbf{p})]\delta[E(\mathbf{p}+\mathbf{q}) - E(\mathbf{p}) - \omega_q]. \quad (70)$$

Equations (69) and (70) are the well-known electron–phonon kinetic equations (see e.g., Peierls[7]), which are obtained under the assumption that the electron–electron interactions can be neglected.

We now turn to the solution of Eq. (63) without the restriction we have imposed before. In the limit of Eq. (64) we obtain

$$g_s^{(1)}(\mathbf{p},\mathbf{q}) = \frac{u_q^*}{\epsilon(\mathbf{q},-\omega_q)}\frac{1}{\Delta(\mathbf{p},\mathbf{q}) - \omega_q + i\epsilon}$$
$$\times \{n(\mathbf{q})H_s(\mathbf{p},\mathbf{q}) + F_s(\mathbf{p},\mathbf{q}) + H_s(\mathbf{p},\mathbf{q})I(\mathbf{q})$$
$$+ F_s(\mathbf{p},\mathbf{q})[\epsilon(\mathbf{q},-\omega_q) - 1]\}, \quad (71)$$

where $\epsilon(\mathbf{k},\omega)$ is given by Eq. (61), $F_s(\mathbf{p},\mathbf{q})$ is defined by Eq. (54), and

$$I(\mathbf{q},\omega_q) = \phi(q)\sum_{\mathbf{p}',s'}\frac{F_s(\mathbf{p},\mathbf{q})}{\Delta(\mathbf{p},\mathbf{q}) - \omega_q + i\epsilon}. \quad (72)$$

If we now denote by

$$I_1(\mathbf{q},\omega_q) = \phi(q)\sum_{\mathbf{p},s}{}'\frac{F_s(\mathbf{p},\mathbf{q})}{\Delta(\mathbf{p},\mathbf{q}) - \omega_q},$$
$$I_2(\mathbf{q},\omega_q) = -\phi(q)\sum_{\mathbf{p},s}F_s(\mathbf{p},\mathbf{q})$$
$$\times \delta[\Delta(\mathbf{p},\mathbf{q}) - \omega_q + i\epsilon], \quad (73)$$

and

$$\epsilon_1(\mathbf{q},\omega) = -\phi(q)\sum_{\mathbf{p},s}{}'\frac{f_s(\mathbf{p}+\frac{1}{2}\mathbf{q}) - f_s(\mathbf{p}-\frac{1}{2}\mathbf{q})}{E(\mathbf{p}+\frac{1}{2}\mathbf{q}) - E(\mathbf{p}-\frac{1}{2}\mathbf{q}) + \omega},$$
$$\epsilon_2(\mathbf{q},\omega) = \phi(q)\sum_{\mathbf{p},s}[f_s(\mathbf{p}+\tfrac{1}{2}\mathbf{q}) - f_s(\mathbf{p}-\tfrac{1}{2}\mathbf{q})]$$
$$\times \delta[E(\mathbf{p}+\tfrac{1}{2}\mathbf{q}) - E(\mathbf{p}-\tfrac{1}{2}\mathbf{q}) + \omega], \quad (74)$$

where $\sum'$ stands for the principal value summation' we obtain

$$\left.\frac{\partial f(\mathbf{p})}{\partial t}\right|_{\mathrm{ph}} = 2\pi\sum_{\mathbf{q}}\frac{|u_q|^2}{|\epsilon(\mathbf{q},\omega_q)|^2}(\{[n(\mathbf{q}) + I_1(\mathbf{q},\omega_q)]$$
$$\times \epsilon_1(\mathbf{q},-\omega_q) + I_2(\mathbf{q},\omega_q)\epsilon_2(\mathbf{q},-\omega_q)\}$$
$$\times \{[f_s(\mathbf{p}+\mathbf{q}) - f_s(\mathbf{p})]\delta[E(\mathbf{p}+\mathbf{q}) - E(\mathbf{p}) - \omega_q]$$
$$- [f_s(\mathbf{p}) - f_s(\mathbf{p}-\mathbf{q})]\delta[E(\mathbf{p}) - E(\mathbf{p}-\mathbf{q}) - \omega_q]\}$$
$$+ \{[n(\mathbf{q}) + I_1(\mathbf{q},\omega_q)]\epsilon_2(\mathbf{q},-\omega_q)$$
$$- I_2(\mathbf{q},\omega_q)\epsilon_1(\mathbf{q},-\omega_q)\}$$
$$\times \{[f_s(\mathbf{p}+\mathbf{q}) - f_s(\mathbf{p})][E(\mathbf{p}+\mathbf{q}) - E(\mathbf{p}) - \omega_q]^{-1}$$
$$- [f_s(\mathbf{p}) - f_s(\mathbf{p}-\mathbf{q})][E(\mathbf{p}) - E(\mathbf{p}-\mathbf{q}) - \omega_q]\})$$
$$+ 2\pi\sum_{\mathbf{q}}u_q^2\{f_s(\mathbf{p}+\mathbf{q})[1 - f_s(\mathbf{p})]$$
$$\times \delta[E(\mathbf{p}+\mathbf{q}) - E(\mathbf{p}) - \omega_q] - f_s(\mathbf{p})$$
$$\times [1 - f_s(\mathbf{p}-\mathbf{q})]\delta[E(\mathbf{p}) - E(\mathbf{p}-\mathbf{q}) - \omega_q]\}, \quad (75)$$

and

$$\frac{\partial n(\mathbf{q})}{\partial t} = 2\pi\frac{|u_q|^2}{|\epsilon(\mathbf{q},\omega_q)|^2}\{[n(\mathbf{q}) + I_1(\mathbf{q},\omega_q)]$$
$$\times \sum_{\mathbf{p},s}[f_s(\mathbf{p}+\mathbf{q}) - f_s(\mathbf{p})]\delta[E(\mathbf{p}+\mathbf{q}) - E(\mathbf{p}) - \omega_q]$$
$$+ \epsilon_1(\mathbf{q},\omega_q)\sum_{\mathbf{p}',s'}f_s(\mathbf{p},\mathbf{q})$$
$$\times [1 - f_s(\mathbf{p})]\delta[E(\mathbf{p}+\mathbf{q}) - E(\mathbf{p}) - \omega_q]\}. \quad (76)$$

Equations (62), (75), and (76) are our final kinetic equations for the electron–phonon system. They all include both the *individual* and *collective* aspects of the electron interactions. The simultaneous motion of the electrons causes a dynamic shielding of each other and of the phonons.

IV. SUMMARY

The present paper represents a method of derivation of kinetic equations for electron–phonon system. We have assumed a simplified model for the system, and started from the Hamiltonian of Bardeen and Pines, modified in such a way to include the "real" frequencies of the phonons. Using the second qualization representation, we have introduced an hierarchy of equations for the electron and phonon distribution functions and correlation functions. The lower members of this hierarchy have been studied employing a generalized version of Bogoliubov procedure of truncating the hierarchy and introducing irreversible kinetic equations. A set of kinetic equations was obtained for homogeneous system taking into account the individual and collective aspects of the electron interactions.

The coupled set of equations, Eqs. (62), (75), and (76), for the electron–phonon system, provide a discription both for the approach to equilibrium of the distribution functions and for the study of transport phenomena of metals. Although a complete solution of these nonlinear equations appears to be extremely difficult because of the formidable nature of the expression, it is hoped that they may prove useful in the approximate determination of transport properties like relaxation times and conductivity. A more detailed study of these problems is reserved for future work.

ACKNOWLEDGMENTS

The author wishes to acknowledge helpful discussions with Dr. C. Oberman. Thanks are also due to Mrs. Kuckes for helping with the English formulation.

The Number of Distinct Sites Visited in a Random Walk on a Lattice*

GEORGE H. VINEYARD
Brookhaven National Laboratory, Upton, Long Island, New York
(Received 20 May 1963)

A general formalism is developed from which the average number of distinct sites visited in n steps by a random walker on a lattice can be calculated. The asymptotic value of this number for large n is shown to be $(8n/\pi)^{\frac{1}{2}}$ for a one-dimensional lattice and cn for lattices of three or more dimensions. The constant c is evaluated exactly, with the help of Watson's integrals, for the simple cubic, body-centered cubic, and face-centered cubic lattices. An analogy is drawn with an electrical network in which unit resistors replace all near-neighbor bonds in a lattice, and the resistance of such a network on each of the three cubic lattices is evaluated.

1. INTRODUCTION

THE random walk on a lattice generates statistical problems which have had recurrent attention for a number of years.[1] The simple question of the average number of distinct sites visited by a walker in n steps, S_n, however, has not been fully treated. Dvoretzky and Erdös[2] have found asymptotic forms of S_n, in the limit of large n, for walks on the simple square lattice, the simple cubic lattice, and the simple hypercubic lattices. For lattices of 3 or more dimensions, they find that S_n approaches the form cn for large n, but do not evaluate the constants c. In the simple square lattice they find the asymptotic form $\pi n/\log n$ for S_n. Beeler and Delaney[3] have studied random walks by Monte Carlo methods on a computing machine, and have deduced approximate asymptotic values of S_n for certain two- and three-dimensional lattices.

The problem of the number of distinct sites visited has practical importance in the theory of annealing of point defects in crystals. A defect such as an interstitial or a vacancy diffuses by a random walk on a lattice, and the rate at which defects are annihilated at point sinks is proportional to the average rate at which defects are arriving at fresh sites on the lattices, that is, at sites which have not been visited previously. The physical side of this problem will be treated at some length in a forthcoming book by Damask and Dienes.[4]

In this note we give a new and simple formulation of the problem of determining S_n, examine its limiting behavior for large n in one, three, and more than three dimensions, and present exact numerical results for the three cubic lattices.

2. GENERAL FORMALISM

Consider a random walk on a Bravais lattice of any number of dimensions. Let the coordination number of the lattice be z. The walker is allowed to step only to nearest-neighbor sites, and to step to each of these with probability $1/z$. If a site is considered to be marked with a "footprint" as soon as the walker visits it, a cloud of footprints develops in the lattice as the walk progresses. On the average this cloud will have the symmetry of the lattice, and, if viewed from the current position of the walker at any stage, will also, on the average, possess the lattice symmetry. Our strategy is to calculate the average density of this cloud of footprints, for the rate at which fresh sites are being visited is just the probability that a site adjoining the walker does not bear a footprint.

Thus, define the probability $p_n(\mathbf{r})$ that, after n steps, the site at $\mathbf{r}$ from the *present position* of the walker has been visited at least once. These probabilities obey the following relations:

$$p_{n+1}(\mathbf{r}) = \frac{1}{z}\sum_{\mathbf{b}} p_n(\mathbf{r}+\mathbf{b}), \qquad \mathbf{r} \neq 0,$$

$$n = 0, 1, 2, \cdots, \tag{1}$$

$$p_n(0) = 1, \qquad n = 0, 1, 2, \cdots, \tag{2}$$

$$p_0(\mathbf{r}) = 0, \qquad \mathbf{r} \neq 0, \tag{3}$$

where $\mathbf{b}$ denotes a nearest-neighbor displacement and $\sum_{\mathbf{b}}$ means summation over the set of z nearest-neighbor displacements.

* Work performed under the auspices of the U. S. Atomic Energy Commission.

[1] For reviews see S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943); E. Montroll, J. Soc. Ind. Appl. Math. **4**, 241 (1956).

[2] A. Dvoretzky and P. Erdös, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman (University of California Press, Berkeley and Los Angeles, 1951), p. 353.

[3] J. R. Beeler, Jr., and J. A. Delaney (unpublished).

[4] A. C. Damask and G. J. Dienes, *Point Defects in Metals* (to be published).

Equations (2) and (3) are obvious. To prove Eq. (1), observe that in the step $n + 1$ the walker might be displaced by a vector $\mathbf{b}$, in which case the site previously at $\mathbf{r} + \mathbf{b}$ relative to the walker becomes the site at $\mathbf{r}$. Since $\mathbf{r} \neq 0$ the walker has not moved on to this site, and so the probability that the site has been visited has not changed. Allowing the probability $1/z$ for the particular displacement $\mathbf{b}$, and summing over the possible displacements, one arrives at Eq. 1.

Equations (1) and (2) and the initial conditions (3) determine the entire set $p_n(\mathbf{r})$.

Let S_n be the average number of distinct sites visited in n steps. The increase, in step $n + 1$, in average number of distinct sites visited is just the probability that any nearest-neighbor site $\mathbf{b}$ has not yet been visited by step n, namely $1 - p_n(\mathbf{b})$ [note that symmetry makes $p_n(\mathbf{b})$ the same for all nearest-neighbor sites]. Thus,

$$S_{n+1} - S_n = 1 - p_n(\mathbf{b}), \qquad n = 0, 1, 2, \cdots ; \qquad S_0 = 1. \tag{4}$$

Equations (1) and (4) determine the set S_n.

Directly from the definitions of S_n and $p_n(\mathbf{r})$ one can also write the relation

$$S_n = \sum_{\mathbf{r}} p_n(\mathbf{r}), \tag{5}$$

where the summation is over all lattice vectors. The consistency of Eqs. (4) and (5) is easy to prove with the aid of Eqs. (1) and (2).

Without seeking explicit solutions of Eqs. (1)–(4), certain general conclusions can be drawn. For n large, $p_n(\mathbf{r})$ becomes indepent of n. Let this limiting value be called $p_\infty(\mathbf{r})$. From Eq. (1), $p_\infty(\mathbf{r})$ is determined by

$$p_\infty(\mathbf{r}) = \frac{1}{z} \sum_{\mathbf{b}} p_\infty(\mathbf{r} + \mathbf{b}), \qquad \mathbf{r} \neq 0, \tag{6}$$

$$p_\infty(0) = 1.$$

These equations can be understood more easily by the following analogy: If an electrical network is constructed[5] with nodes at the lattice points and unit resistors replacing all near-neighbor bonds, and if the nodes at infinity are grounded and the node at the origin is held at unit potential, $p_\infty(\mathbf{r})$ will be the potential of the node at $\mathbf{r}$. From this consideration one can demonstrate that, for networks of 3 or more dimensions, $0 < p_\infty(\mathbf{b}) < 1$.

Equations (4) lead, in three or more dimensions, to a limiting form, for large n,

$$S_n = a + [1 - p_\infty(\mathbf{b})]n, \tag{7}$$

where a is a constant.

For one- or two-dimensional lattices, the electrical network analogy shows that $p_\infty(\mathbf{b}) = 1$, and here S_n must increase less rapidly than linearly with n.

3. LIMITING VALUES IN ONE DIMENSION

For one dimension, the limiting growth is found by passing from Eq. (1) to a differential equation for $p_n(\mathbf{r})$, valid in the limit of large n:

$$2[\partial p_n(x)/\partial n] = \partial^2 p_n(x)/\partial x^2, \qquad p_n(0) = 1. \tag{8}$$

The solution of (8) is

$$p_n(x) = 1 - \operatorname{erf}[x/(2n)^{\frac{1}{2}}].$$

In the same limit,

$$\frac{dS_n}{dn} = 1 - p_n(1) = \operatorname{erf}\left[\frac{1}{(2n)^{\frac{1}{2}}}\right] \underset{n\to\infty}{\to} \left(\frac{2}{\pi n}\right)^{\frac{1}{2}}. \tag{9}$$

Equation (9) has the solution

$$S_n = a + 2(2n/\pi)^{\frac{1}{2}} \qquad (n \to \infty), \tag{10}$$

showing a square-root growth of S_n with n. The meaning of this is evident from the consideration that the rms excursion of the walker is proportional to $n^{\frac{1}{2}}$, and, in one dimension, while sites inside this distance will almost always have been visited, sites outside it will not.

4. LIMITING VALUES IN THREE DIMENSIONS

Consider a three-dimensional cubic Bravais lattice. In the simple cubic lattice let the cell edge be the unit of length; in the body-centered and face-centered cubic lattices, let half the cubic cell edge be the unit of length. Let the Cartesian components of a near-neighbor vector $\mathbf{b}$ be denoted b_1, b_2, b_3, and of a lattice vector $\mathbf{r}$ be denoted r_1, r_2, r_3; all of these components will be integers. The general solution of the Eqs. (6) for $p_\infty(\mathbf{r})$ can now be written down:

$$p_\infty(\mathbf{r}) = \frac{1}{F} \int_0^\pi \int_0^\pi \int_0^\pi \frac{\cos(ur_1)\cos(vr_2)\cos(wr_3)}{1 - z^{-1}\sum_{\mathbf{b}} \cos(ub_1)\cos(vb_2)\cos(wb_3)}\, du\, dv\, dw, \tag{11}$$

where

$$F = \int_0^\pi \int_0^\pi \int_0^\pi \frac{du\, dv\, dw}{1 - z^{-1}\sum_{\mathbf{b}} \cos(ub_1)\cos(vb_2)\cos(wb_3)}. \tag{12}$$

[5] Electrical networks have also been employed in random-walk problems by K. Compaan and Y. Havens, Trans. Faraday Soc. 52, 786 (1956).

To demonstrate that (11) is a solution of (6), write the latter in the form

$$\Omega p_\infty(\mathbf{r}) = 0, \qquad \mathbf{r} \neq 0,$$

where Ω is an operator defined by

$$\Omega f(\mathbf{r}) = f(\mathbf{r}) - z^{-1} \sum_{\mathbf{b}} f(\mathbf{r} + \mathbf{b}).$$

Observe that, by virtue of the cubic symmetry, $\cos(ur_1)\cos(vr_2)\cos(wr_3)$ is an eigenfunction of Ω with eigenvalue

$$1 - z^{-1} \sum_{\mathbf{b}} \cos(ub_1)\cos(vb_2)\cos(wb_3).$$

Then, applying Ω to the expression (11), for $p_\infty(\mathbf{r})$ one finds

$$\Omega p_\infty(\mathbf{r}) = \frac{1}{F}\int_0^\pi\int_0^\pi\int_0^\pi \cos(ur_1)\cos(vr_2) \times \cos(wr_3)\,du\,dv\,dw$$

$$= \frac{1}{F}\prod_{i=1}^{3}\frac{\sin \pi r_i}{r_i} = 0, \qquad (\mathbf{r} \neq 0).$$

Finally, from the definition of F, it is evident that $p_\infty(0) = 1$, which completes the demonstration.

For the three lattices, simple cubic (sc), body-centered cubic (bcc), and face-centered cubic (fcc), $p_\infty(\mathbf{b})$ may readily be evaluated. One uses the relation $z^{-1}\sum_{\mathbf{b}'} p_\infty(\mathbf{b}') = p_\infty(\mathbf{b})$ to rewrite (11) as

$$p_\infty(\mathbf{b}) = \frac{1}{F}\int_0^\pi\int_0^\pi\int_0^\pi \frac{z^{-1}\sum_{\mathbf{b}'}\cos(ub_1')\cos(vb_2')\cos(wb_3')}{1 - z^{-1}\sum_{\mathbf{b}'}\cos(ub_1')\cos(vb_2')\cos(wb_3')}\,du\,dv\,dw$$

$$= \frac{1}{F}\int_0^\pi\int_0^\pi\int_0^\pi \left[\frac{1}{1 - z^{-1}\sum_{\mathbf{b}'}\cos(ub_1')\cos(vb_2')\cos(wb_3')} - 1\right] du\,dv\,dw$$

$$= \frac{1}{F}[F - \pi^3].$$

Finally, for sc lattices, $F = 3\pi^3 I_3$; for bcc, $F = \pi^3 I_1$; and for fcc, $F = 3\pi^3 I_2$, where I_1, I_2, and I_3 are integrals which have been evaluated by Watson.[6]

The asymptotic rate of change of S_n with n, as seen from Eq. (7), is $1 - p_\infty(\mathbf{b})$. Values of this quantity, as determined here, and also as found by Beeler and Delaney in their Monte Carlo treatments of diffusion, are given in Table I.

TABLE I. $\lim_{n\to\infty}(dS_n/dn)$, where S_n is average number of distinct sites visited in n steps.

Lattice	Present calculations	Beeler and Delaney
sc	0.659 462 670	0.667
bcc	0.717 770 010	0.725
fcc	0.743 681 763	0.756

An independent way of deriving the asymptotic rate of change of S_n with n is the following: In the paper by Dvoretzky and Erdös,[2] it is demonstrated that $1 - p_n(\mathbf{b})$ equals the probability that the walker does not return to the origin at any time during the first n steps. Then $1 - p_\infty(\mathbf{b})$ is the so-called escape probability, the probability that the walker never returns to the origin. Montroll[1] has evaluated the escape probabilities for sc, bcc, and fcc lattices; his results agree with ours in Table I. Also, the demonstration given above in Sec. 2 that, for one- or two-dimensional lattices,

$$\lim_{n\to\infty}\frac{dS_n}{dn} = 0$$

accords with the previously known fact[7] that in these lattices the escape probability is zero.

Finally, one notes that, from the electrical network analogue of Eq. (6) cited earlier, the resistance from a node to infinity in a lattice in which unit resistors have replaced all near-neighbor bonds can be written

$$\frac{1}{z[1 - p_\infty(\mathbf{b})]}.$$

This resistance is 0.25273, 0.17415, and 0.11206 Ω for the sc, bcc, and fcc lattices, respectively.

5. ACKNOWLEDGMENTS

Dr. A. H. Schoen and Dr. Elliott Montroll have contributed helpful discussions. The writer is also indebted to Dr. Beeler and Dr. Delaney for sending him their Monte Carlo results in advance of publication.

[6] G. N. Watson, Quart. J. Math. **10**, 266 (1939).

[7] G. Polya, Math. Ann. **84**, 149 (1921).

Observable Properties of Large Relativistic Masses*

A. W. K. METZNER†
Department of Physics, Cornell University, Ithaca, New York
(Received 26 December 1962)

Interstellar space may be full of very dense and very faint stars (supernova remnants, very old white dwarfs, etc.), on whose surfaces the gravitational field intensity is very high. What, according to general relativity, are their observable properties, and what are the maximum gravitational effects as a matter of principle? To answer this, the exact test particle orbits (geodesics) of the Schwarzschild metric are derived, classified and described, on the basis of an exact classical model. The latter is of considerable help in making the properties of the orbits immediately evident, and represents the principal advantage of the present derivation over previous ones. Also, as radial coordinate, instead of the usual largely arbitrary r, the metric coefficient $g_{00} = A^2(r)$ is used. A^2 has immediate physical significance as the red shift ratio $\nu(r)/\nu(\infty)$, and its use also simplifies the formulas. The test-particle scattering angle (generalizing the results of Darwin), and the differential scattering cross section, as well as the capture cross section of a sphere of radius $A^2(r) = A^2(R) = A_R^2$, are calculated as a function of test-particle energy, and presented graphically. The observed, augmented, angular diameter of a sphere is calculated in terms of A_R^2, and some peculiar lenslike effects of masses, first discussed by Einstein, are reviewed. It is then pointed out that an important result of Curtis for the interior field, implies $A_R^2 > 0.514$, so that the exterior field extends at most down to $A^2 = 0.514$, and the region $A^2 < 0.514$—which would be the most pathological region of the exterior—cannot in fact exist, if Curtis' arguments are assumed valid. As a result, quasihyperbolic test-particle orbits exist with pericentrum equal to the radius of any conceivable spherical mass, and the radius can therefore in principle always be determined by an asymptotic scattering experiment. As a further result, the maximum photon-scattering angle can be no more than about 110° (and larger for slower particles), providing a cutoff at this point in the photon-scattering cross section, and allowing a massive star to produce at most one secondary image of another star. As a practical matter, none of the effects discussed here seems large enough to be measurable today, with the sole well-known exception of the red shift, which for the Curtis limiting sphere would be large enough to shift a visible spectrum into the near infrared.

INTRODUCTION

THE subject of this paper is the exact general-relativistic (GR) behavior of test particles (TP's) in the Schwarzschild exterior field. The purpose is to determine the observable properties, both in principle and in an astronomical sense, of large, dense masses—in particular, of the largest masses apparently possible in a consistent GR description.

The motivation for this investigation comes from the old,[1] and recently revived,[2] interest in the ultimate fate of large, gravitating masses, i.e., the final evolutionary stages of white dwarfs. We have here the following dilemma: On the one hand, according to GR, the well-known Schwarzschild singularity sets a lower limit to the size of a given mass, as Wheeler has shown.[2] On the other hand, there appears to be nothing in present theories of the structure of matter, which could in fact prevent an indefinite gravitational collapse.[3] We shall not here be concerned with finding a way out of this dilemma. We merely note that, whatever its solution, there is in any case the strong suggestion that interstellar space is abundantly populated with very old, very dense, and very faint white dwarfs. If this is so, what, according to GR, are their effects on a distant observer? It is to the answer to that question that we wish to contribute here.

A TP in this context is a particle whose own contribution to the field is negligible, and which therefore moves along the geodesics (null-geodesics for TP's with zero rest mass) of the given metric.

The discussion will be based almost entirely on the exact geodesics, since we are interested in the

* Some of this material was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in physics at Cornell University in 1963.

† Present address: 10 Saint Paul Street, Cambridge, Massachusetts.

[1] J. R. Oppenheimer and R. Serber, Phys. Rev. **54**, 540 (1938); **55**, 374 (1939); **56**, 455 (1939); G. M. Volkoff, Phys. Rev. **55**, 413 (1939); R. C. Tolman, Phys. Rev. **55**, 364 (1939); F. Zwicky, Astrophys. J. **88**, 522 (1938); Phys. Rev. **55**, 726 (1939); *Morphological Astronomy* (Julius Springer-Verlag, Heidelberg, 1957), pp. 255–6.

[2] J. A. Wheeler, "Some Implications of General Relativity for the Structure and Evolution of the Universe," *Institut International de Physique Solvay, Onzième Conseil de Physique, June 1958* (Editions Stoops, Bruxelles, Belgium); A. G. W. Cameron, Astrophys. J. **130**, 884 (1959).

[3] See for example, S. A. Colgate, W. H. Grasberger, and R. H. White, "The Dynamics of a Supernova Explosion", 1961 (unpublished).

strong field limit and in questions of principle. The exact TP orbits in the Schwarzschild exterior metric were first derived by Hagihara.[4] More recently they were treated by Darwin,[5,6] and parts of the present work may be viewed as extensions of parts of Darwin's work. The principal novel feature here is that the orbits are derived and classified with the help of an easily visualizable classical model (Sec. IC). In addition, we use the metric coefficient g_{00} as the radial coordinate instead of the usual distance coordinate r, for reasons that will be explained below (Sec. IA).

Since the Schwarzschild field is static, we are in effect restricted to those effects that can be observed during a time interval short compared to the evolutionary time scale of the source matter. To take account of the evolutionary nature of real masses, suitable nonstatic solutions must first be discovered and investigated. The fact that the Schwarzschild field leads straight to the above-mentioned dilemma, makes this all the more urgent.

Also left out of account here are the details of the structure of the source matter. Thus, quantum mechanics, elementary particle theory, etc., and their possible reconciliation with GR, will not be discussed, except to point out possible overall limits placed on the size of the source by GR itself (see below).

The paper falls into two main parts. In Sec. I the exact behavior of TP's in the empty Schwarzschild exterior field is worked out. Among other things we shall here find, as Darwin[5,6] did, that there are TP orbits that can attain the center of force. In part II, however, we shall introduce the fact that the source of the field itself must occupy some of the space. Hence the exterior orbits cannot, in fact, approach the center of force arbitrarily closely, because the source body is in the way. Part II deals with the effect this has on the observable properties of the field.

In the concluding part of Sec. II we discuss the result of Curtis[7] which showed that GR itself puts a lower limit on the possible radius of a given mass, which is much more restrictive than the Schwarzschild limit. The existence of the Curtis limit would mean that the exterior field cannot extend down to where the Schwarzschild singularity would be, and has an important bearing on the question of what can be measured in principle: We shall see that, with the Curtis limit, the radius of any source body can in principle always be determined by an asymptotic scattering experiment, which would not necessarily be the case if masses could go beyond the Curtis limit.

I. BEHAVIOR OF TEST PARTICLES IN AN EMPTY, STATIC, SPHERICALLY SYMMETRIC GRAVITATIONAL FIELD

A. Metric of the Exterior Field

The most general form of the Schwarzschild metric is[8]

$$ds^2 = A^2(r)\,dt^2 - d\ell^2, \tag{1}$$

where

$$d\ell^2 = B^2(r)\,dr^2 + C^2(r)r^2\,d\omega^2, \tag{1a}$$

and

$$d\omega^2 = d\theta^2 + \sin^2\theta\,d\phi^2.$$

Here θ and ϕ are the usual uniquely defined spherical angular coordinates. The radial distance coordinate r, however, is arbitrary, in the sense that any substitution of the form $r = \bar{r}f(\bar{r})$, with arbitrary differentiable $f(\bar{r})$, leaves the form of the metric invariant.

Now it is well known[9] that, for the exterior solution, the weak field limit, the Newtonian limit, and the limit $r \to \infty$ are identical, namely,

$$A^2(r) \to (1 - 2m/r), \quad B^2(r) \to 1, \quad C^2(r) \to 1, \tag{2}$$

where $m = GM/c^2$, $G = 6.67 \times 10^{-8}\ \mathrm{c^3\ g^{-1}\ sec^{-2}}$, provided M is the mass as measured by the sphere's gravitational attraction at infinitely large distances, and provided r is taken as the familiar Newtonian distance coordinate in the same limit. Under these conditions the usual exterior solution is

$$C(r) = (2m/r)(1 - A^2)^{-1}, \quad B(r) = A^{-1}(drC/dr), \tag{3}$$

and the spatial part (1a) becomes

$$(d\ell)^2/4m^2 = A^{-2}(1 - A^2)^{-4}(dA^2)^2 + (1 - A^2)^{-2}\,d\omega^2, \tag{1b}$$

where r has been eliminated as radial coordinate in favor of $g_{00} \equiv A^2$. The usual procedure is to make some more or less arbitrary choice of the function $A^2(r)$, consistent with the limit [Eq. (2)]. For example, with $A^2 = 1 - 2m/r$, we get the

[4] Y. Hagihara, Japan J. Astron. Geophys. **8**, 67 (1931). This old article was discovered after the present work was substantially completed; the only reference to it is apparently in the book of Synge (reference 9).

[5] C. Darwin, Proc. Roy. Soc. (London) **A249**, 180 (1959).

[6] C. Darwin, Proc. Roy. Soc. (London) **A263**, 39 (1962).

[7] A. R. Curtis, Proc. Roy. Soc. (London) **A200**, 248 (1949).

[8] V. Fock, *The Theory of Space Time and Gravitation*, (Pergamon Press, Inc., New York, 1959).

[9] See for example, G. C. McVittie, *General Relativity and Cosmology* (John Wiley & Sons, Inc., New York, 1956); C. Møller, *The Theory of Relativity* (Oxford University Press, London, 1952); J. L. Synge, *Relativity: The General Theory* (North-Holland Publishing Company, Amsterdam, 1960).

familiar "Schwarzschild" form of the metric, which we shall here call the S metric. Other forms that have found some use are the isotropic[10] and the harmonic.[8] We perfer to use A^2 itself, because, in contrast with any of these r's, it has a direct physical meaning, and its value at the surface of the source body (say, at $r = R$ or $A^2 = A_R^2$) is perhaps the most readily measurable property of the source. As is well known, $A^2(r)$ is simply the ratio, $\nu(r)/\nu(\infty)$, of the frequency of a photon (say, H_α) emitted at r and received at a large distance to the frequency of a similar photon (i.e. also H_α) emitted close to the receiver. Thus A_R^2 gives the red shift of the source body; and in conjunction with a mass measurement (by gravitational attraction at large distances) it gives the radius R for any given radial coordinate choice. As $A_R^2 \to 0$, the observed frequency of emitted photons goes to zero, and the rate of reception of energy, and of information, goes to zero. One could, therefore, never "see" a sphere shrink to this size.

B. Orbits of Test Particles

The TP orbits are given by the geodesics of the metric (1), which in the plane $\theta = \frac{1}{2}\pi$ are well-known[9] to yield,

$$(rC)^2(d\phi/ds) = m\alpha, \qquad A^2(dt/ds) = \gamma, \tag{4}$$

where $m\alpha$ and γ are two constants of the motion, which in the limit $r \to \infty$ are unambiguously the special-relativistic angular momentum per unit TP rest mass, and the TP Lorentz factor. Thus (4) corresponds to the classical equations of conservation of angular momentum and total energy. As alternatives to α and γ, it will be convenient also to introduce β, the TP velocity at infinity; and δ, the TP impact parameter (in units of m):

$$\alpha^2 = \delta^2(\gamma^2 - 1) = \delta^2\beta^2(1 - \beta^2)^{-1}. \tag{5}$$

The differential equation for the orbits is now obtained by eliminating t, B, C from (3), (1a), and (4), which gives

$$[d(A^2)/d\phi]^2 = (4/\alpha^2)(\gamma^2 - A^2) - A^2(1 - A^2)^2. \tag{6}$$

This in turn may be put in a standard mathematical form[11] by introducing yet another alternative radial coordinate z,

$$z = \tfrac{1}{12}(2 - 3A^2), \qquad A^2 = \tfrac{2}{3} - 4z, \tag{7}$$

so that (6) becomes

$$\begin{aligned}(dz/d\phi) &= 4z^3 - g_2 z - g_3 \\ &= 4(z - z_1)(z - z_2)(z - z_3),\end{aligned} \tag{6a}$$

where

$$\begin{aligned} g_2 &= -4(z_2z_3 + z_3z_1 + z_1z_2) = \tfrac{1}{12}(1 - 12/\alpha^2), \\ g_3 &= 4z_1z_2z_3 = (\tfrac{1}{216})[1 + 18(2 - 3\gamma^2)/\alpha^2], \\ 0 &= z_1 + z_2 + z_3, \end{aligned} \tag{8}$$

and where we assume $z_1 \geq z_2 \geq z_3$ when all three are real. The standard solution of (6a) is the Weierstrassian elliptic function $p(\phi - \phi_0; g_2, g_3)$; in general, a complex, doubly periodic function of a complex argument with two real parameters. Naturally only real values of the angle ϕ are involved here, but the constant of integration may be complex. If $2\omega_i$ ($i = 1, 2, 3$) are the periods of $p(\phi)$, then

$$\begin{aligned} p(\phi) &= p(-\phi) = p(\phi + 2\omega_i), \\ p(\omega_i) &= z_i, \qquad \omega_1 + \omega_2 + \omega_3 = 0. \end{aligned} \tag{9}$$

The character of a particular orbit is largely determined by the sign of the discriminant Δ of the cubic on the right-hand side of (6a), which is defined by

$$\Delta = g_2^3 - 27g_3^2,$$

so that from (8), after some manipulation,

$$\Delta(\alpha, \gamma) = (27/16\alpha^4)(\gamma_u^2 - \gamma^2)(\gamma^2 - \gamma_s^2), \tag{10}$$

$$\Delta(\beta, \delta) = (1/16\delta^6)(\delta^2 - \delta_u^2)(\delta^2 - \delta_s^2), \tag{11}$$

where

$$\gamma_{u,s}^2 = (\alpha^2/54)[(1 + 36/\alpha^2) \pm (1 - 12/\alpha^2)^{\frac{3}{2}}], \tag{10a}$$

$$\delta_{u,s}^2 = (1/2\beta^4)[(8\beta^4 + 20\beta^2 - 1) \pm (8\beta^2 + 1)^{\frac{3}{2}}]. \tag{11a}$$

For each particular ($\pm$) sign of Δ, the solution $p(\phi)$ of (6a) can be written as an expression involving only Jacobian elliptic functions, which, roughly speaking, are generalizations of trigonometric and hyperbolic functions (to which they reduce for $\Delta = 0$), with real periods ranging from the trigonometric limit 2π to the hyperbolic limit (∞). Functions like this had to appear, of course, to give the perihelion advance.

The roots of the cubic on the right-hand side of (6a) are evidently related to the possible orbital turning points (where $dr/d\phi = dA/d\phi = dz/d\phi = 0$). These points will be denoted by r_i, A_i, or z_i ($i =$

[10] A. S. Eddington, *The Mathematical Theory of Relativity*, (Cambridge University Press, Cambridge, England, 1957).

[11] See for example, L. M. Milne-Thomson, *Jacobian Elliptic Function Tables* (Dover Publications, Inc., New York, 1950), p. 23; or, E. Jahnke and F. Emde, *Tables and Functions* (Dover Publications, Inc., New York, 1945), p. 98. The results of Sec. IC, D, E were obtained in 1959 before the author became aware of the related work of Darwin (reference 5).

1, 2, 3). Only real positive values of r_i can, of course, qualify as physical turning points. By (8), the z_i (only two of which are independent) are uniquely determined for any TP with given constants of the motion. In fact, each TP can now be associated with a variety of alternative pairs of orbital invariants: either (α, γ), or (β, δ), or (g_2, g_3), or two of the three r_i, A_i, z_i, or ω_i. Yet another pair of invariants is (e, ℓ), the eccentricity and the latus rectum, which were used by Darwin,[6] working in the S metric, to write (6a) in the following form (in our notation):

$$[d(A^2)/d\phi]^2 = (A_1^2 - A^2)(A_2^2 - A^2)(A_3^2 - A^2),$$

with

$$A_1^2 = 4m/\ell, \qquad A_2^2 = 1 - 2m(1 + e)/\ell,$$
$$A_3^2 = 1 - 2m(1 - e)/\ell.$$

The degenerate orbits with $\Delta = 0$ have only one degree of freedom: Any one orbital invariant is in this case uniquely determined in terms of any other. Thus, since two of the z_i are equal in this case, one has from (7), (8), (10a), and (11a),

$$\begin{aligned} A_{u,s}^2 &= \tfrac{1}{3}[2 \mp (1 - 12/\alpha^2)^{\frac{1}{2}}], \\ &= \tfrac{1}{4}(1 - \beta^2)^{-1}[3 \mp (8\beta^2 + 1)^{\frac{1}{2}}], \end{aligned} \tag{12}$$

where $A_u = A_1 = A_2$ and $A_s = A_2 = A_3$.

C. Exact Classical Model

Consider a classical particle with angular momentum $m\alpha$ per unit mass and with total energy (excluding rest energy) $\frac{1}{2}(\gamma^2 - 1)$ per unit mass, moving in the central potential:

$$V(r) = -(m/r)(1 + m^2\alpha^2/r^2). \tag{13}$$

The classical conservation laws give

$$m\alpha = r^2(d\phi/dt), \tag{14}$$

$$\gamma^2 - 1 = (dr/dt)^2 + r^2(d\phi/dt)^2 + 2V, \tag{15}$$

and therefore,

$$\gamma^2 - 1 = (dr/dt)^2 + m^2\alpha^2/r^2 + 2V. \tag{15a}$$

After eliminating t and putting $u = m/r$, the classical equation of motion becomes

$$(du/d\phi)^2 = (\gamma^2 - 1)/\alpha^2 + 2u/\alpha^2 - u^2 + 2u^3. \tag{6b}$$

But this is precisely the form taken by the orbit equation (6) in its S-metric form, i.e., with $A^2 = 1 - 2m/r$. Hence such a classical particle has precisely the Schwarzschild orbits.

To make this model more instructive, we note

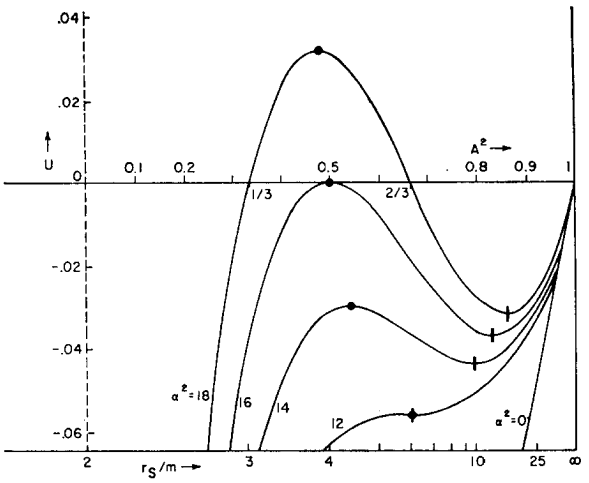

FIG. 1. The equivalent classical one-dimensional potential for the motion of test particles in the Schwarzschild metric vs the radial coordinate $A^2(=g_{00})$, both dimensionless [Eq. (16a)]. The parameter α corresponds to the test-particle angular momentum per unit rest mass. The maxima (dots) and the minima (bars) are at the positions defined by (12).

that the central potential $V(r)$ is equivalent to the one-dimensional potential[12]

$$U(r) = V(r) + m^2\alpha^2/2r^2, \tag{16}$$

in the sense that, if the particle were constrained to move in one dimension in the potential U, then its energy equation would be just (15a) again.

The relations (15a) and (16) show that only those ranges of r which satisfy $\frac{1}{2}(\gamma^2 - 1) \geq U(r)$, are accessible to a particle with given (α, γ); and $U(r)$ depends only on α^2. Hence on a plot of $U(r)$ vs r, the intersections of the straight line $U = \frac{1}{2}(\gamma^2 - 1)$ with the curve $U = U(r; \alpha)$ give the orbital turning points r_i, and the parts of the straight line above the curve give the accessible ranges of r. Instead of $U(r)$ one can consider $U(A^2)$, which then applies directly to the orbit equation (6). Thus from (13) and (16), with $A^2 = 1 - 2m/r$, one obtains

$$U = -\tfrac{1}{2}(1 - A^2) + (\tfrac{1}{8}\alpha^2)A^2(1 - A^2)^2, \tag{16a}$$

which is shown in Fig. 1 for various values of α^2.

It is easy to verify that $U(A^2)$ has extrema at $A^2 = A_{u,s}^2$, where $U = \frac{1}{2}(\gamma_{u,s}^2 - 1) \equiv U_{u,s}$, as expected from (10a) and (12), the subscripts u and s corresponding, respectively, to a maximum and a minimum of U. This also explains the choice of these subscripts, for it is evident from the figure that unstable (stable) circular orbits are possible for a TP with $\gamma^2 = \gamma_u^2(\gamma^2 = \gamma_s^2)$, i.e., when the

[12] See for example, H. Goldstein, *Classical Mechanics* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1953), p. 63.

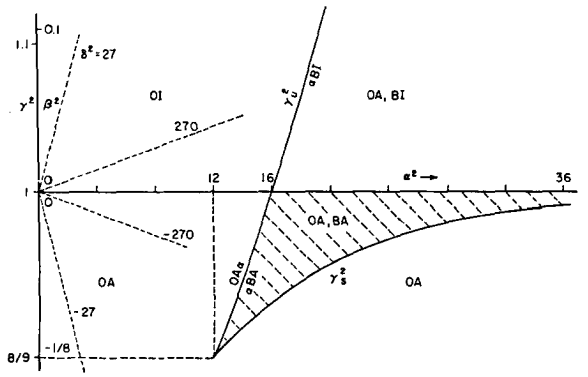

FIG. 2. The (α^2, γ^2) plane with the curves $\gamma_{u,s}^2$ of (10a) (the β^2 ordinate is not on a linear scale), showing also several of the curves $\delta^2 = \text{const}$ given by (5). For a test particle at infinity, γ, β, α, δ are, respectively, the Lorentz factor, the velocity in units of c, the angular momentum per unit rest mass, and the impact parameter in units of the graviational radius. The capital letters identify the domains of the various types of orbits defined in Table I.

line $U = \frac{1}{2}(\gamma^2 - 1)$ is tangent to the maximum (minimum) of the curve $U = U(A^2; \alpha^2)$. With the condition $\gamma^2 = \gamma_u^2$, it is possible in addition to have orbits that spiral up to or down to the point $A^2 = A_u^2$ (Types aBI, aBA, OAa; see below).

D. Classification of Orbits

Figure 2 gives a plot of $\gamma_{u,s}^2$ vs α^2 as in (10a). The TP orbits in the metric (1a) form a two-parameter family of curves, every orbit corresponding to a unique point in the (α^2, γ^2) plane. However, more than one physically distinct orbit may correspond to a given point in this plane, since in Fig. 1 the straight line $U = \frac{1}{2}(\gamma^2 - 1)$ may have two distinct segments above the curve $U = U(A^2)$. In such a case, the choice of the constant of integration ϕ_0 in the solution $p(\phi - \phi_0)$ of Eq. (6a) determines one or the other of the two possibilities.

Figure 2 also shows the line $\gamma^2 = 1$. Orbits below this line are bound orbits, of course. In Fig. 1 this corresponds to the fact that an orbit with $\gamma^2 < 1$ has a straight line $U = \frac{1}{2}(\gamma^2 - 1)$ that lies below the abscissa, and which therefore must intersect the curve $U = U(A^2)$ at a positive value of A_3^2 less than one (since these curves all approach $U = 0$ from below as $A^2 \rightarrow 1$).

Using Fig. 1 we therefore arrive at the classification of orbits summarized in Table I and on Fig. 2. Our nomenclature has the following significance: "O" means that the orbit spirals into the origin, $r = 0$ (which may be anywhere between $A^2 = 0$ and $A^2 = -\infty$); "I" means that the orbit attains infinity ($r = \infty$ or $A^2 = 1$); "A" ("B") means that the orbit is bounded above (below) by a circle to which it is tangent; "a" means that it approaches this bounding circle in an asymptotic spiral. The general appearance of the orbits is therefore self-explanatory.

Orbits of types OI, OA, aBI, aBA, and OAa have no analogues in Newtonian gravitation. BI orbits correspond to the Newtonian hyperbolas (parabolas when $\gamma^2 = 1$), and BA orbits are the precessing ellipses that give rise to the advance of the perihelion. Finally, of course, there is the possibility of purely radial motion, on OI or OA orbits in the limit $\alpha^2 = 0$.

E. Precessing Ellipses: Perihelion Advance

The BA orbits are best written in the form

$$A^2(\phi) = A_3^2 - (A_3^2 - A_2^2)\,\mathrm{sn}^2(b\phi, k), \qquad (17)$$

where

$$2b^2 = A_3^2 + \tfrac{1}{2}A_2^2 - 1,$$
$$k^2 = (A_3^2 - A_2^2)(2A_3^2 + A_2^2 - 2)^{-1}.$$

Since sn ( ) lies between ± 1, this clearly shows that A^2 lies between A_2^2 and A_3^2. Fig. 1 shows that the minimum possible value of the perihelion A_2^2 for a BA orbit is $\frac{1}{2}$ (or $r_2 = 4m$, when $A^2 = 1 - 2m/r$, as Darwin[5] showed). The perihelion advance from (9) is $2(\omega_2 - \pi)$, where in this case,[11] $\omega_2 = 2(A_3^2 - A_1^2)^{-\frac{1}{2}}K(k)$ and $K(k)$ is the complete elliptic

TABLE I. Classification of orbits.

Type	Δ	Range of A^2	Solution of orbit Eq. (6) or (6a)
BI	+	$A_2^2 < A^2 < 1$	$z(\phi) = z_3 + (z_2 - z_3)\,\mathrm{sn}^2(b\phi, k)$ $b^2 = z_1 - z_3, \quad k^2 = (z_2 - z_3)/b^2$
OI	−	$-\infty < A^2 < 1$	$z(\phi) = z_1 + H^2(1 + \mathrm{cn}\,2H\phi)(1 - \mathrm{cn}\,2H\phi)^{-1}$ $H^4 = 2z_1^2 + g_3/4z_1, \quad k^2 = \frac{1}{2} - \frac{3}{4}z_1/H^2$
BA	+	$A_2^2 < A^2 < A_3^2$	same form as BI
OA	$\left\{\begin{matrix}-\\0\\+\end{matrix}\right\}$	$-\infty < A^2 < A_1^2$	same form as OI $A^2(\phi) = 1 - \frac{1}{2}A_1^2 - 4b^2\csc^2 b\phi, \quad 8b^2 = 2 - 3A_1^2$ same form as BI or BA
aBI	0	$A_u^2 < A^2 < 1$	$A^2(\phi) = 2(1 - A_u^2) - 4b^2\tanh^2 b\phi, \quad 4b^2 = 2 - 3A_u^2$
aBA	0	$A_u^2 < A^2 < A_3^2$	same form as aBI
OAa	0	$-\infty < A^2 < A_u^2$	replace $b\phi$ by $b\phi + \frac{1}{2}i\pi$ in aBI

integral of the first kind with modulus k. For the planets, $\gamma^2 \lesssim 1$ and $\alpha^2 \gg 1$, so that one can readily show (e.g., Synge[9] and Darwin[5]) that, to first order in α^{-2}, the advance per revolution comes out to $6\pi/\alpha^2$. This is the formula that has been applied to the orbit of Mercury.[9]

F. Photon Orbits

Photons have zero rest mass, and therefore are simply TP's with $\beta^2 = 1$. For instance, by taking the ratio of the two expressions in (4), rewriting (α^2, γ^2) in terms of (β^2, δ^2) by means of (5), and then setting $\beta^2 = 1$, one gets

$$A^{-2}(rC)^2(d\phi/dt) = m\delta, \qquad (4')$$

which is just the well-known result[9] for the null-geodesics of the metric (1a). In the same way, the photon orbits are given by (6a) and (7), but with

$$g_2 = 1/12, \qquad g_3 = (1/216)(1 - 54/\delta^2), \qquad (8')$$

$$\Delta = (1/16\delta^4)(\delta^2 - 27), \quad \delta_u^2 = 27, \quad \delta_s^2 = 0. \qquad (11')$$

Degenerate photon orbits ($\Delta = 0$) have $A_u^2 = \frac{1}{3}$ and A_s^2 does not exist, as follows from (12). Therefore photons can move in unstable circular orbits with radius $A^2 = \frac{1}{3}$, but not in stable circular orbits.

As for a classical model, it is easy to verify that a classical particle with total energy $\frac{1}{2}$ per unit mass and total angular momentum $m\delta$ moving in the central potential $V = -m^3\alpha^2/r^3$ has the photon equation of motion (6a) with (8'), if, as before, r is defined by $A^2 = 1 - 2m/r$. The equivalent one-dimensional potential is then $U = (\delta^2/8)A^2(1 - A^2)^2$. This is negative for $A^2 < 0$, zero at $A^2 = 0$, has a maximum $U = U_u = \frac{1}{2} = (\delta^2/54) = (\delta_u^2/54)$ at $A^2 = A_u^2 = \frac{1}{3}$, and approaches zero again at $A^2 = 1$. [Note, however, that this model potential is not a limiting form of the model (16a) for the general case.] Clearly, there can be no photon orbits of types BA or aBA. But photons can have all the other types of orbits listed in Table I.

G. Circular Orbits

These have been frequently discussed before,[5,6,13,14] so we shall merely briefly show here how these previous results emerge in the present treatment.

The existence of stable and unstable circular orbits has already been pointed out, and their radii are given by (12). Now, by putting $dr = 0$ in (1a), and eliminating α^2 and $\gamma^2 = \gamma_{u,s}^2$ from (3), (4), (5), and (10a), one gets

$$(dt/d\phi)^2 = 8m^2(1 - A_{u,s}^2)^{-3},$$

$$(ds/d\phi)^2 = \tfrac{1}{2}(3A_{u,s}^2 - 1)(dt/d\phi)^2.$$

If, following McCrea[13] and Darwin,[6] the coordinate period T and the proper period S are now defined by the changes in t and s, respectively, for a change of 2π in ϕ, then it follows from the above and from (12) that

$$\begin{aligned} T &= 4\sqrt{2}\,\pi m(1 - A_{u,s}^2)^{-\frac{3}{2}}, \\ &= \sqrt{2}\,\pi m\alpha^3[1 \mp (1 - 12/\alpha^2)^{\frac{1}{2}}]^{\frac{3}{2}} \\ &= \sqrt{2}\,\pi m(-\beta)^{-3}[1 - 4\beta^2 \mp (8\beta^2 + 1)^{\frac{1}{2}}]^{+\frac{3}{2}}, \\ S &= \sqrt{2}\,(3A_{u,s}^2 - 1)^{\frac{1}{2}}T, \\ &= \sqrt{2}\,[1 \mp (1 - 12/\alpha^2)^{\frac{1}{2}}]^{\frac{1}{2}}T \\ &= \tfrac{1}{4}(1 - \beta^2)^{-\frac{1}{2}}[3 \mp (8\beta^2 + 1)^{\frac{1}{2}}]T. \end{aligned}$$

These formulas show once again, what was already evident from Figs. 1 and 2—that both stable and unstable circular orbits have a minimum possible "angular momentum" given by $\alpha^2 = 12$, where $\beta^2 = -\frac{1}{8}$ and $A^2 = \frac{1}{2}$. For the stable ones, we get the correct classical limits $S \to T$ and $T \to 2\pi m^{\frac{1}{2}}r^{\frac{3}{2}} \cong 2\pi m(-\beta)^3 \cong 2\pi m\alpha^3$, as $\alpha^2 \to \infty$, or $\beta^2 \to -0$, or $A_s^2 \to 1 - 2m/r$ as in (2). And for the unstable ones, we get a minimum possible radius of $A_u^2 = \frac{1}{3}$ as $\beta^2 \to 1$, or $\alpha^2 \to \infty$, or $S \to 0$.

H. Scattering Angle

Consider a TP incident from infinity with velocity β and impact parameter δ. What is the scattering angle? For photons the question was answered by Darwin,[5] working in the S metric. In the general case we start with the BI-orbit equation in Table I. The azimuth of the pericentrum is ω_2, and if ϕ_0 denotes the azimuth of the incident direction $(r = \infty, A^2 = 1, z = -1/12)$, then

$$b\phi_0 = \mathrm{sn}^{-1}(n, k), \qquad b\omega_2 = K(k),$$

where

$$n^2 = -(z_3 + 1/12)(z_2 - z_3)^{-1}. \qquad (18)$$

Hence the scattering angle θ is given by

$$\theta + \pi = 2(\omega_2 - \phi_0) = 2F/b, \qquad (19)$$

where

$$F = K(k) - \mathrm{sn}^{-1}(n, k) = \mathrm{sn}^{-1}(h, k),$$

$$k^2 = (1 - n^2)(1 - k^2n^2)^{-1}.$$

Thus, for a given (δ^2, β^2) one must first solve

[13] W. H. McCrea, Helv. Phys. Acta **29**, Suppl. 4, 121 (1956).
[14] P. Goldhammer, Nuovo Cimento **20**, 1205 (1961).

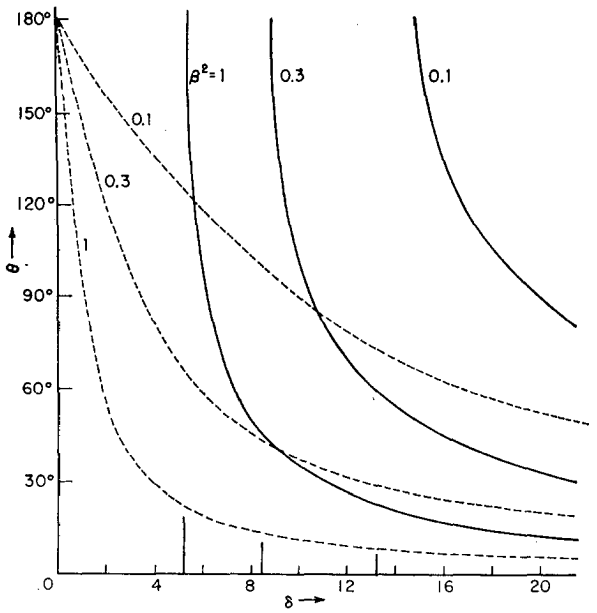

FIG. 3. The test-particle scattering angle as a function of the dimensionless impact parameter for three values of the total energy parameter β^2 (19). The dotted curves give the corresponding classical scattering angles (20). The vertical lines along the abscissa give the asymptotes δ_u^2 (11a) for the solid curves. The curve $\beta^2 = 1$ is for photons.

the cubic in (16a), then calculate k, b, n, or h, and finally substitute in (19). This, and the computation of the elliptic integrals, is readily done by machine,[15] and some typical results are shown in Fig. 3, which includes the photon case $\beta^2 = 1$, and for comparison, also the classical scattering angle

$$\theta_{\text{class}} = 2 \cot^{-1} (\delta\beta^2). \tag{20}$$

If we want to use the pericentrum A_2^2 instead of δ as orbital invariant, then from (5) to (8),

$$z_2 = (1/12)(2 - 3A_2^2), \quad z_{1,3} = -\tfrac{1}{2}(z_2 \mp b^2),$$
$$16b^4 = [1 - (1 - \beta^2)A_2^2]^{-1}[A_2^2(4 - 3A_2^2) - (1 - \beta^2)A_2^2(2 - A_2^2)^2], \tag{18a}$$
$$2k^2 = 1 + (2 - 3A_2^2)(4b^2)^{-1},$$
$$n^2 = (4b^2 - A_2^2)(4b^2 + 2 - 3A_2^2)^{-1},$$

which substituted in (19) gives the scattering angle as a function of β^2 and A_2^2, where the corresponding classical expression is now [with $A_2^2 = 1 - 2m/r_2$ because of (2)]

$$\theta_{\text{class}} = 2 \cot^{-1} [\beta^2(r_2/m)(1 + 2m/\beta^2 r_2)^{\frac{1}{2}}]. \tag{20a}$$

Curves of $\theta(\beta^2, A_2^2)$ look very much like curves of $\theta(\beta^2, \delta^2)$.

[15] The programming was done for an IBM 7090 by Ellen J. Metzner, to whom I would like to express my sincere thanks.

The most notable feature of the results of Fig. 3 is that θ becomes infinite as $\delta \to \delta_u$ (11a), or as $A_2^2 \to A_u^2$ (12), whereas θ_{class} approaches 180° as $\delta \to 0$ for all β. This is due to the existence of aBI and OI orbits: a TP with impact parameter $\delta \leq \delta_u$, or with pericentrum $A_2^2 \leq A_u^2$, will always spiral into the center of force. As Darwin already showed, these minimum-approach distances are smallest for photons, for which $\delta_u^2 = 27$ and $A_u^2 = \frac{1}{3}$.

Darwin[5] also gave approximate expressions for the photon ($\beta^2 = 1$) scattering angle in the two limits $\delta \to \infty$ and $\delta \to \delta_u$. The former is, of course, the well-known formula that gives the measurable bending of light in the sun's field,

$$\theta \cong 4/\delta \cong 4m/r_2 \ll 1, \tag{19a}$$

obtained from (19) by using (2) and expanding in inverse powers of δ. In the other limit we put $\delta = 27^{\frac{1}{2}}(1 + \epsilon)$ and expand in powers of ϵ, which gives Darwin's result,

$$\delta(27)^{\frac{1}{2}} = 1 + 216(7 + 4\sqrt{3})^{-1} \times \exp[-(\pi + \theta)]. \tag{19b}$$

I. Differential Scattering Cross Section

Classically, and in units of m^2, this cross section is[12]

$$\sigma(\theta) = \csc\theta\, f(\theta), \qquad f(\theta) = |\delta(d\delta/d\theta)|. \tag{21}$$

Now in the last part of this section we saw that θ becomes infinite at a finite value of δ. This means first of all, that in addition to the usual infinity in the forward direction, there is also one in the backward direction. This is entirely due to the csc θ factor: The element of solid angle into which TP's are scattered becomes zero as θ approaches 180°, while the area of the annulus about the incident direction, from which they came, remains finite; hence the number per unit solid angle goes to infinity. Secondly, we have the fact that the TP's scattered into a particular element of solid angle, $2\pi \sin\theta\, d\theta$, do not all come from a unique annulus, $2\pi\delta\, d\delta$, but from an infinite sequence of ever smaller annuli with radii tending to δ_u: TP's with δ close to δ_u may make several complete revolutions about the scattering center before emerging at some net angle between 0 and π. Thus, with $0 < \theta < \pi$,

$$\sin\theta\, \sigma(\theta) = f(\theta) + \sum_{n=1}^{\infty} [f(2n\pi - \theta) + f(2n\pi + \theta)]. \tag{21a}$$

Figure 4 shows a plot of (21a) for photons, with (upper curve) and without (lower curve) the sum

terms. The principal term $f(\theta)$ was machine calculated from the data of the last part of this section[15]; and the sum terms, only the first of which is numerically significant, were calculated with the use of Darwin's approximation (19b). The dotted curve in Fig. 4 gives the corresponding classical Rutherford cross section

$$\sin\theta\,\sigma_{\text{class}} = \sin\theta\,(4\beta^2\sin^4\tfrac{1}{2}\theta)^{-1}, \qquad (22)$$

evaluated for a particle moving with the velocity of light. Both σ and σ_{class} have a θ^{-4} singularity in the forward direction, where $\sigma \longrightarrow \sigma_{\text{class}}$.

II. SOURCE OF THE GRAVITATIONAL FIELD: EFFECT OF ITS SPATIAL DIMENSIONS ON TEST-PARTICLE BEHAVIOR

A. Capture Cross Section of Source

Consider a parallel monchromatic beam of TP's incident from infinity with velocity β. We now introduce for the first time the fact that the source of the field must itself occupy some of the space. Thus TP orbits cannot come arbitrarily close to the center of force without intersecting the spherical surface (say at $r = R$ or $A^2 = A_R^2$) of the source body. In fact, a TP will be captured if its pericentrum is less than the radius of the source body, i.e., if $A_2^2 < A_R^2$. The maximum impact parameter δ_c of those TP's that are captured is therefore given by substituting (β^2, δ^2) for (α^2, γ^2) in (6) from (5), putting the left-hand side of (6) equal to zero, and replacing A^2 by A_R^2, which results in

$$\delta_c^2 = 4[\beta A_R(1 - A_R^2)]^{-2}[1 - (1 - \beta^2)A_R^2]. \qquad (23)$$

As expected, this has a minimum δ_u^2(11a) at $A_R^2 = A_u^2$ (12), so that TP's with $\delta < \delta_u$ will be captured no matter what the radius. This means that the radius of a body with $A_R^2 \leq A_u^2(\beta)$ cannot be measured by an asymptotic-scattering experiment using TP's with total energy parameter β; and if $A_R^2 \leq \frac{1}{3}$, the radius cannot be measured by this means at all, not even with photons. The capture cross section is therefore

$$\Sigma(R, \beta^2) = \begin{cases} 2\pi\delta_u^2(\beta^2), & A_R^2 \leq A_u^2, \quad (24) \\ 2\pi\delta_c^2(R, \beta^2), & A_R^2 \geq A_u^2. \quad (24a) \end{cases}$$

This is shown in Fig. 5 for several values of β. For comparison, the classical cross section is shown

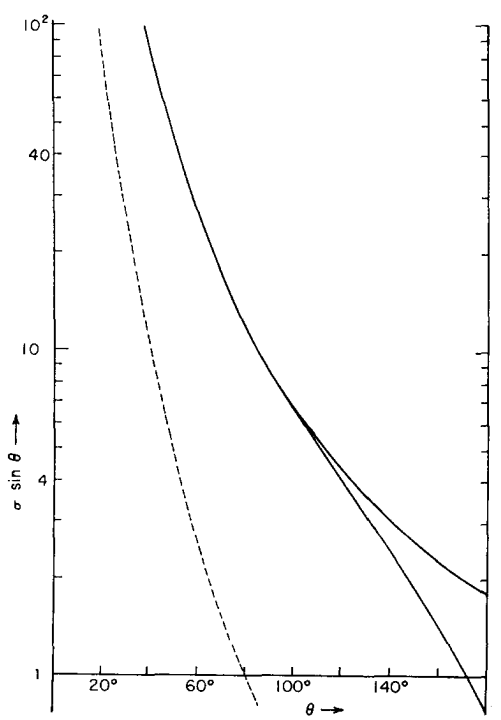

FIG. 4. The photon differential scattering cross section ($\beta^2 = 1$) from (19), (21), and (21a), both with (upper curve) and without (lower curve) the flux of those photons that are scattered through more than 180°. The dotted curve gives the Rutherford cross section (22) for a particle moving with the velocity of light. The units are the square of the gravitational radius.

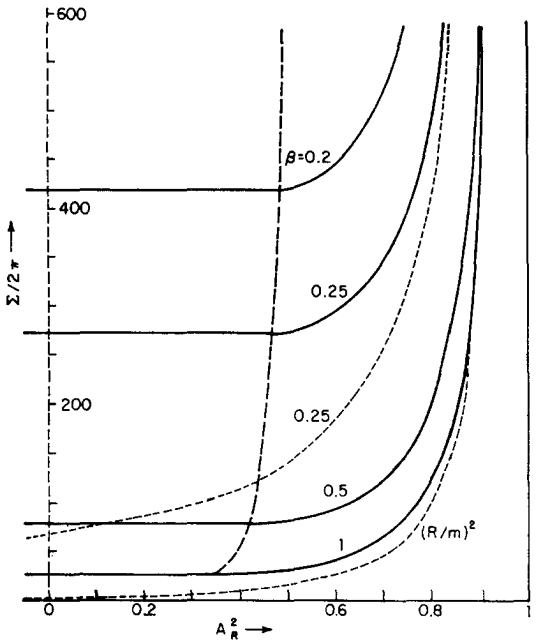

FIG. 5. The capture cross section (24) and (24a), in units of the square of the gravitational radius, for a body of radius R, as a function of $A^2(R)$, for various values of the incident velocity β of the particles being captured. The dotted curve gives the corresponding classical capture cross section (25) for $\beta = 0.25$. The curve $\beta = 1$ is for photons. The curve $(R/m)^2$ gives the geometrical cross section of the sphere. The broken curve starting at $A_R^2 = 1/3$ and asymptotic to $A_R^2 = \frac{1}{2}$ gives A_u^2 as in (12).

for the case $\beta = 0.25$:

$$\Sigma_{\text{class}} = 2\pi(R/m)^2(1 + 2m/R\beta^2), \qquad (25)$$

and also the geometrical cross section $2\pi(R/m)^2$.

B. Surface Emission: Escape Velocity

Following McVittie,[9] define the coordinate velocity of a TP in its orbit by $q = d\ell/dt$, so that from (1), (4), (5),

$$q^2 = A^2 - (ds/dt)^2 = A^2[1 - (1 - \beta^2)A^2]. \qquad (26)$$

A TP is able to escape to infinity if $\beta^2 \geq 0$. Hence the escape velocity from the surface of the source body is given by putting $\beta^2 = 0$ and $A^2 = A_R^2$ in (26),

$$q^2_{\text{escape}} = A_R^2(1 - A_R^2), \qquad (27)$$

which has a maximum of $\frac{1}{4}$ at $A_R^2 = \frac{1}{2}$. Thus it might be tempting to say that, for a sphere shrinking in size in the range $A_R^2 < \frac{1}{2}$, it becomes easier and easier for particles to evaporate off its surface as the radius gets smaller and smaller. This is not really the case, however. The velocities defined in (26) and (27) are coordinate velocities only. A more physical measure of TP velocity is the ratio of the TP coordinate velocity to the photon coordinate velocity. Since the latter, by (26), is simply A^2, this ratio is $1 - (1 - \beta^2)A^2$, and the corresponding escape velocity is $(1 - A_R^2)$, both of which are properly monotonically decreasing.

C. Surface Emission: Cone Effect

TP's leaving the surface of the source body at a small angle with the vertical are on OI orbits, and have less angular momentum than TP's leaving almost tangentially, which must be on BI orbits. However, if $A_R^2 < A_u^2$, then there is no BI orbit tangent to the surface, and, *a fortiori*, no BI orbit that intersects the surface at all, so that in this case the TP must have less than a certain maximum angular momentum in order to be able to escape on an OI orbit. Hence the TP must leave the surface within a certain cone about the vertical, otherwise it will fall back again on an OA orbit. From the point of view of Fig. 1, this simply means that the TP's potential α^2 curve must be such that its maximum lies below the TP's horizontal energy γ^2 line. The semiangle of the cone, $\chi(R)$, is defined by the limiting OAa orbit, an orbit which neither escapes nor falls back, but which asymptotically approaches $A^2 = A_u^2$ from below.

Now, if the orbit equation is $r = r(\phi)$, or $A^2 = A^2(\phi)$, then simple geometry gives

$$\tan\chi = r(dr/d\phi)^{-1} = r(dA^2/dr)(dA^2/d\phi)^{-1}. \qquad (28)$$

With the equation of the OAa orbit from Table I, this gives, at $r = R$, the required semiangle as a function of R and β^2,

$$\tan\chi(R, \beta^2) = R(dA_R^2/dR)(A_u^2 - A_R^2)^{-1}(2 - 2A_u^2 - A_R^2)^{-\frac{1}{2}}.$$

As $A_R^2 \to A_u^2$ from below, the cone angle approaches 90°, i.e., the effect vanishes. Since the largest A_u^2 is $\frac{1}{2}$ (namely for $\beta^2 = 0$), this "cone effect" plays no role for spheres with radii such that $A_R^2 > \frac{1}{2}$.

The following short table gives two numerical examples (for $\beta^2 = 0$ and $\beta^2 = 1$) of the variation with radius of the cone angle in the S metric (i.e., with $A^2 = 1 - 2m/r$):

R/m	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
$\chi(R, 0)$	0°	30°	43°	54°	63°	71°	78°	86°	90°
$\chi(R, 1)$	0°	30°	44°	57°	75°	80°	90°	···	···

This is not an effect that can be "observed" in any sense by a distant observer. In fact, the formula shows that the magnitude of χ depends on the choice of the radial coordinate function $A^2(r)$. By doctoring up $A^2(r)$ in the neighborhood of any given point, we can make χ anything we please at that radius. We have discussed the effect here for its intrinsic oddity; it does not seem to have been treated quantitatively before.

There is another, related, even odder effect. Consider the photon orbits, and imagine an observer on the surface of a sphere of radius $A_R^2 < \frac{1}{3} = A_u^2$. As he lifts his gaze above the geometrical horizon, he will see regions of the surface beyond the horizon by light travelling on OA orbits. At a certain elevation he will see the antipodal point (assuming enough light and sufficient resolving power) as a ring around the horizon. At a still higher elevation he will see a ring image of the back of his head, then a second fainter image of the antipodal point, then another of his head, and so on, an infinite sequence converging in elevation to the angle defined by the cone effect. Since the "sky" will be visible only within the overhead cone, the observer will no doubt ask himself whether he is really on the "surface" and not deep in some circular well. This is perhaps a typical example of the ambiguity that many familiar concepts acquire in a realm far from the classical.

D. Surface Emission: Enhanced Angular Radius

Classically, photons travel in straight lines, and the angular radius of a star at a large distance r would be

$$\chi_{\text{class}} \cong R/r. \tag{29}$$

Because of the bending of light, however, the photons that form the outer edge of the observer's disc image of the sphere actually left the surface of the sphere slightly beyond the geometrical limb, and arrive on a curved trajectory making a greater angle with the sphere–observer axis than the straight line from the limb. These photons are on those BI orbits that have a pericentrum distance equal to the radius of the sphere, as long as that radius has $A_R^2 > \frac{1}{3}$. For $A_R^2 < \frac{1}{3}$, and the photons that form the edge of the observer's image are those OI photons that are emitted just inside the cone of the previous section; i.e., those that just manage to clear the potential hump in Fig. 1. Hence the observed angular radius, when the real radius is in this latter range, is independent of the real radius, and is given by the angle that the OAa photons make at the observer (which is the limit of the angles made by BI photons as $A_2^2 \to \frac{1}{3}$).

The observed angular radius then is given by (28), where r is now the observer distance. Since r is large, (2) applies, and (28) becomes

$$\chi_{\text{obs}} \cong (2m/r) \lim_{A^2 \to 1} |dA^2/d\phi|^{-1}. \tag{30}$$

The BI orbit with pericentrum at A_R^2 is given in Table I and by (18a) with $A_2^2 = A_R^2$. By differentiating it with respect to ϕ, and reexpressing $dA^2/d\phi$ as a function of A^2, and then taking the limit $A^2 \to 1$, one readily obtains

$$\chi_{\text{obs}} \cong (2m/r) A_R^{-1} (1 - A_R^2)^{-1}, \tag{31}$$

and

$$\chi_{\text{obs}}/\chi_{\text{class}} \cong A_R^{-1}. \tag{32}$$

In the limit of large R the ratio (32) tends to unity by (2), as it should. In the limit $A_R^2 \to \frac{1}{3}$, the angular radius in (31) tends to $27^{\frac{1}{2}}/r$, and the ratio in (32) tends to $\sqrt{3}$; these are, respectively, the smallest possible observable angular radius, and the largest possible magnification over the classical angular radius.

This effect does not seem to have been discussed before. It is extremely small in all practial cases. For example, for the sun it comes to about 10^{-2} seconds of arc in about 10^3 sec total, and for a typical white dwarf at a distance of one light year it is about 10^{-6} sec out of 10^{-3} sec total; for the former, the effect is too small, and for the latter the angular radius is too small in the first place. In addition, there is no way of comparing the observed angular radius with what it would have been in the absence of the bending of light. It might, however, be possible to get around[16] this last difficulty by looking at the intensity distribution across the disc image, especially toward the limb; or by measuring the radar-signal travel time. In the latter case, as the radar is beamed at points closer and closer to the limb, the travel time would become larger than one would expect classically, and indeed infinite if $A_R^2 < \frac{1}{3}$.

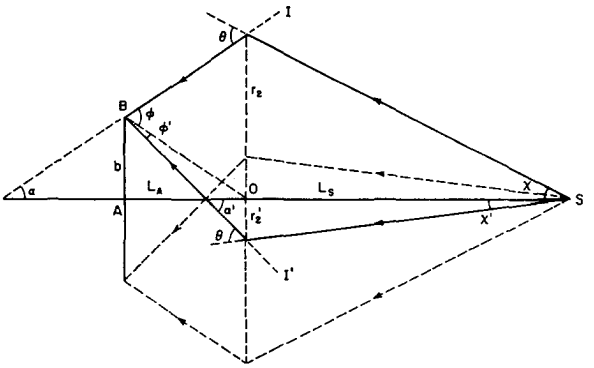

FIG. 6. Model for the formation of two images, I and I′, of the star S at B in the field of the gravitating body O. AB = b, OS = L_S, OA = L_A. All angles are vastly exaggerated.

E. Gravitational Lens Action

Lenslike effects of masses were already discussed in 1936 by Einstein[17] and Zwicky,[18] and again more recently by Darwin.[5] We discuss them here merely to show the relation between the results of Einstein and Darwin, and to point out an error in the work of the latter.

Consider the situation depicted in Fig. 6: A star S emits photons which are "bent" around the field source O and reach an observer B. Assuming $b \ll L_A \ll L_S$, all angles are small. We may therefore dispense with the exact photon equations, and consider a model in which each light ray is bent through the angle (19a) at a point on the perpendicular through O, as indicated in the figure. Thus, from the figure,

$$\theta = \alpha + \chi \cong 4m/r_2, \tag{33}$$

$$\chi \cong r_2/L_S, \qquad \alpha \cong (r_2 - b)/L_A, \tag{33'}$$

[16] I owe these suggestions to Professor P. Morrison and Professor P. L. Hartman.

[17] A. Einstein, Science **84**, 506 (1936).

[18] F. Zwicky, Phys. Rev. **51**, 290 (1937); and pp. 216–219 of his book cited in reference 1.

for the primary image, and similar relations with primed quantities for the secondary image. Defining $\alpha_0^2 \equiv 4m/L_A$ and $a^2 \equiv 4mL_A$, one finds from (33) and (33′),

$$\alpha, \alpha' \cong [(1 + b^2/4a^2)^{\frac{1}{2}} \mp b/2a], \tag{34}$$

and

$$\phi \cong \alpha' \cong \theta' \cong \chi(L_S/L_A), \tag{35}$$

$$\phi' \cong \alpha \cong \theta \cong \chi'(L_S/L_A). \tag{35'}$$

Thus when the observer is in the symmetric position ($B \to A$), then $b = 0$, $\alpha = \alpha' = \alpha_0$, and the angular radius of A's "ring image" of S is α_0, as Einstein, Zwicky, and Darwin found.

Let J be the absolute intensity of S. Then without O, its apparent brightness would be $I_0 = J/4\pi L_S^2$. But with O, it is

$$I = (J/4\pi)2\pi \sin\chi \, d\chi/2\pi \cos\alpha \, b \, db$$
$$\cong (J/4\pi) \, |\chi \, d\chi/b \, db|, \tag{36}$$

and similarly for the secondary image (I'). Therefore,

$$e \equiv (I + I')/I_0 \cong (a/b)(1 + b^2/2a^2) \times (1 + b^2/4a^2)^{-\frac{1}{2}}, \tag{37}$$

$$I'/I \cong |\chi' \, d\chi'/\chi \, d\chi|$$
$$\cong |\phi' \, d\phi'/\phi \, d\phi| \cong (\alpha_0/\phi)^4 \cong (\phi'/\phi)^2$$
$$\cong [(1 + b^2/4a^2)^{\frac{1}{2}} - b/2a]^4. \tag{38}$$

The expression (37) for the enhancement factor e of the total apparent brightness, is Einstein's result. Darwin calculated (I'/I), but because he worked with angles rather than solid angles in (36), he obtained only the square root of (38). Both Einstein and Darwin have pointed out that the magnitudes involved make these results more or less irrelevant from the point of view of observational astronomy.

F. Conclusion: The Ultimate in Gravitational Field Strength

We have seen that in the region $A^2 < \frac{2}{3}$ the field rapidly becomes pathological, while for $A^2 > \frac{2}{3}$ it differs only very little from the Newtonian gravitational field. The question now is, how small can A^2 really be on the surface of the source body without some contradiction within the framework of general relativity? A precise answer to this was given by Curtis,[7] who found that, if the trace of the energy–momentum tensor is regarded as a measure of the baryon number density in the interior of the source body (and this is the natural interpretation), and if this sphere is so dense that the velocity of sound has its maximum GR value $c/\sqrt{3}$ throughout the interior, then A_R^2 must be greater than about 0.514, or otherwise the pressure would become infinite at the center. According to this argument, it is therefore a necessary conclusion within the framework of GR that the exterior field can extend only down to $A^2 = 0.514$ at most, and the ultimate in gravitational field strength is to be found in regions of space immediately outside this.

While we believe Curtis' arguments to be correct, it is perhaps well to remember what was already pointed out in the introduction, namely that present theories of the structure of matter do not seem to provide a mechanism that could prevent a gravitational collapse. Once the properties of real matter and radiation at these high densities are understood, either such a mechanism will be provided, or it may well be that either Curtis' result on the velocity of sound or his interpretation of the energy–momentum tensor may have to be modified.

By way of conclusion we shall now enumerate the principal properties of Curtis' limiting sphere.

(i) *Red shift.* According to the Curtis limit, the greatest red shift that could ever be observed is $\Delta\nu/\nu_\infty = 1 - A^2 = 0.486$. Thus the visible spectrum of the source could at most be shifted into the near infrared, and a possible blackbody temperature could at most be halved. Nevertheless, the red shift is still the most important and the most readily accessible measurement that a distant observer can make on a large gravitating body.

(ii) *Orbits.* For the limiting Curtis sphere there can be no circular TP orbits unstable with respect to escape orbits (see Fig. 1), but in the range $0.514 < A^2 < \frac{2}{3}$ there could be circular orbits unstable with respect to precessing ellipses. TP's would have to have "angular momentum" $\alpha < (15.2)^{\frac{1}{2}}$ in order to move on OA orbits (from Fig. 1); in particular, there would be no photon OA and OAa orbits.

(iii) *Scattering angles.* The greatest possible scattering angle for BI orbits with pericentrum $A_2^2 = 0.514$ comes to about 740°, 285°, 110° for $\beta^2 = 0$, 0.2, 1, respectively, from (18a) and (19). Darwin's approximation (19b) for photons has no practical significance for the Curtis limiting sphere. Also, the extra terms in (21a) are zero, and the photon-scattering cross section is given by the lower curve in Fig. 4, except that it drops to zero

at $\theta = 110°$. For slower TP's, however, the extra terms might still have to be taken into account.

(iv) *Capture cross section.* Since by (12), $\frac{1}{3} < A_u^2 < \frac{1}{2}$, the range $A^2 < A_u^2$ is not part of the exterior. Hence the capture cross section is given by (24) without (24a), and does not become independent of A_R^2. If the Curtis limit is valid, then the radius of any conceivable body can always be determined by an asymptotic scattering experiment.

(v) *Surface emission.* The "cone effect" plays no role. The observed angular radius cannot become independent of the actual radius, which can therefore be determined in principle by this method too. From (32) the Curtis sphere has $\chi_{obs}/\chi_{class} \cong 1.4$.

(vi) *Gravitational lens action.* An observer looking at Curtis' limiting sphere (O) could in principle see a secondary image of a star (S) 20° behind him. All other stars lying between S and O would also have one secondary image each within the angle α_0 (34) of O. (If the Curtis limit is ignored, and A^2 is assumed less than $\frac{1}{3}$, then the whole sky will have an infinite sequence of images around O. This is the situation discussed by Darwin.[5]) A search for such a weighted, composite, spectral image was suggested already long ago by Zwicky.[19]

ACKNOWLEDGMENTS

I should like to record my indebtedness to Professor Philip Morrison for guiding and encouraging this work, and to Professor Wolfgang Rindler for his critical interest.

[19] F. Zwicky, Phys. Rev. 59, 221 (1941).

Mathematical Analysis on the Effect of a Prolate Spheroidal Core in a Magnetic Dipole Field*

MOHAMMED AZIZUL ISLAM
Advanced Communications Section, Avco Corporation, Wilmington, Massachusetts
(Received 26 April 1963)

The magnetic vector potential has been obtained for the case of a circular loop of current surrounding a material core of a prolate spheroidal shape, by solving Maxwell's equations and suitable boundary conditions. It is shown that this vector potential consists of two parts; the first part is that due to the loop alone, the second part being due to the presence of the core.

I. INTRODUCTION

IN many problems relating to the scattering and diffraction of electromagnetic waves by obstacles, and also in the field of antennas, one attempts to find the effect of a circular material cylinder of finite length on an existing electromagnetic field. The cylinder will, in general, have permeability and permittivity different from those in free space and may also be lossy. For example, one may wish to know the effect of dielectric or magnetic loading with a finite cylinder on a loop antenna. The author discussed the case of loading a loop antenna with a cylinder of infinite length in earlier papers[1,2] where it has been shown that there is a decided gain obtainable from such loading. Since an infinite cylinder is never realizable in practice, an attempt was made to analyze the case of a finite cylinder. Because of complexities involved at the end of a cylinder, the problem becomes very complicated. The nearest approximation to a finite cylinder is a prolate spheroid; and by adjusting the different parameters involved in a prolate spheroid, it is possible to obtain a simple analytical expression for the surface of a finite cylinder to a reasonably good approximation. The present paper will discuss methods of obtaining the electromagnetic field quantities due to a circular loop of current surrounding a prolate spheroidal core. The dimension of the loop has been assumed small compared to the free-space wavelength of the field quantities involved, so the current through the loop may be considered uniform.

For simplicity, the static case will be taken up first and then the time-varying case will be discussed. The static formulation is useful in cases where one is interested in finding the low-frequency impedance of the system, the time-varying case is, however, applicable to radiation problems.

The electromagnetic fields (both electric and magnetic) may be derived using equations involving the desired fields directly, or one may use suitably defined vector and scalar potentials, and then derive the fields from the potentials thus obtained. This latter method was used in the present case.

II. REVIEW OF ELECTROMAGNETIC FUNDAMENTALS

Definition of Terms

- $\mathbf{H}$ magnetic field intensity, A/m
- $\mathbf{B}$ magnetic flux density, V sec/m^2 $= \mu\mathbf{H}$
- $\mathbf{E}$ electric field intensity, V/m
- $\mathbf{D}$ electric flux density $= \epsilon\mathbf{E}$
- ϵ dielectric constant, F/m
- μ permeability, H/m
- $\mathbf{A}$ vector magnetic potential
- ϕ scalar electric potential
- v $1/(\mu\epsilon)^{\frac{1}{2}}$ = phase velocity of propagating wave in a medium of permeability μ, and dielectric constant ϵ.
- k ω/v = wavenumber
- ω radian frequency
- $\hat{\mathbf{n}}$ unit vector normal to the boundary surface

Other terms will be defined as and when required.

Description of Method for Solution

The method involves the solution of the homogeneous wave equation (time variation of the field quantities according to the factor $e^{-i\omega t}$ understood),

$$\nabla^2\mathbf{A} + k^2\mathbf{A} = 0, \tag{1}$$

in terms of a complete set of harmonic functions with unknown coefficients. Then proper boundary

* A major portion of this work was supported by Laboratory For Electronics, Inc., Boston, Massachusetts, with whom the author was formerly associated.

[1] M. A. Islam, IEEE Trans. Antennas Propagation, 11 162 (1963).

[2] M. A. Islam, "Analysis of an Electrically and Magnetically Loaded Loop Antenna," presented at the twelfth Annual Symposium, USAF Antennas Research and Development Program, University of Illinois, October 16–19, 1962.

conditions are imposed on this set of solutions to determine the value of the unknown coefficients. The boundary conditions are, as usual,

$$\begin{aligned} &\hat{n} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0, \\ &\hat{n} \times (\mathbf{H}_2 - \mathbf{H}_1) = \text{the true surface current.} \end{aligned} \tag{2}$$

where subscripts 2 and 1 represent the values at regions 2 and 1, respectively.

III. STATEMENT AND FORMULATION OF THE PROBLEM

The problem is to find an expression for $\mathbf{A}$ at a point in space having coordinates (η, ξ, ϕ) in the prolate spheroidal coordinate system due to a circular loop located at (η_0, ξ_0) carrying current I (assumed uniform throughout the loop) and perturbed by the presence of a material core (assumed homogeneous and isotropic with permeability μ and permittivity ϵ) located at the origin as shown in Fig. 1. The shape of the core has been assumed to be a prolate spheroid and so the prolate spheroidal coordinate system will be used for convenience. Thus, the problem is to solve (1) subject to the conditions in Eq. (2), and that the solution must be finite, single valued, and an outgoing wave at infinity. The natural division of boundary is also shown in Fig. 1.

IV. DESCRIPTION OF THE PROLATE SPHEROIDAL COORDINATE SYSTEM

A set of prolate spheroids is obtained by revolution about the major axis of a family of confocal ellipses. These coordinates may be obtained by using the conformal transformation in the w plane

$$\begin{aligned} W = z + ir &= f \cosh (u + iv) \\ &= f \cosh u \cos v + if \sinh u \sin v. \end{aligned}$$

Therefore,

$$z = f \cosh u \cos v; \qquad r = f \sinh u \sin v,$$

and

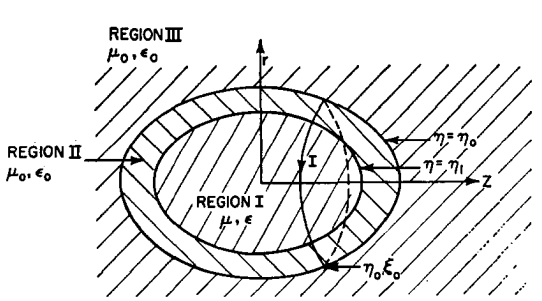

FIG. 1. Permeable prolate spheroid and a current loop. A typical division of regions for applying boundary conditions.

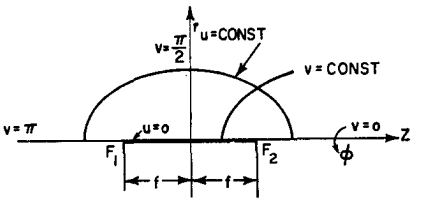

FIG. 2. The prolate spheroidal coordinate system.

$$x = r \cos \phi; \qquad y = r \sin \phi.$$

Since

$$\cosh^2 \theta - \sinh^2 \theta = 1,$$

and

$$\cos^2 \theta + \sin^2 \theta = 1,$$

therefore,

$$\left(\frac{z}{f \cos v}\right)^2 - \left(\frac{r}{f \sin v}\right)^2 = 1,$$

$$\left(\frac{z}{f \cosh u}\right)^2 + \left(\frac{r}{f \sinh u}\right)^2 = 1.$$

These are sets of confocal hyperbolas and ellipses which are shown in Fig. 2.

It is evident that the range of values that u and v may have are $0 \leq v \leq \pi$ and $0 \leq u \leq \infty$. A set of coordinates may now be defined, in terms of u, v, x, and y, as $\eta = \cosh u$; $\xi = \cos v$; $\phi = \tan^{-1} (r/x)$. Therefore,

$$W = z + ir = f\eta\xi + if[(\eta^2 - 1)(1 - \xi^2)]^{\frac{1}{2}},$$

$$dW = r \, d\phi.$$

If the scale factors are now defined as

$$dW^2 = h_\xi^2 \, d\xi^2 + h_\eta^2 \, d\eta^2 + h_\phi^2 \, d^2\phi,$$

then

$$\begin{aligned} h_\xi &= |dW/d\xi| = f[(\eta^2 - \xi^2)/(1 - \xi^2)]^{\frac{1}{2}}, \\ h_\eta &= |dW/d\eta| = f[(\eta^2 - \xi^2)/(\eta^2 - 1)]^{\frac{1}{2}}, \\ h_\phi &= |dW/d\phi| = f[(\eta^2 - 1)(1 - \xi^2)]^{\frac{1}{2}}. \end{aligned} \tag{3}$$

In the simplest case of circular symmetry where a loop of uniform line current is located at (η_0, ξ_0) (the current loop assumed as having linear dimension only, and dimensions of the loop are small compared to wavelength of all field quantities involved, so an assumed uniform distribution of current throughout the loop is reasonable), only the ϕ component of the vector potential will exist, and there will be no variation of any of the quantities with respect to the ϕ coordinate. Under these conditions, the vector Helmholtz equation $(\nabla^2 + k^2)\mathbf{A} = 0$ reduces to

$$\frac{\partial}{\partial\eta}\left[(\eta^2-1)\frac{\partial A_\phi}{\partial\eta}\right]-\frac{1}{\eta^2-1}A_\phi + \frac{\partial}{\partial\xi}\left[(1-\xi^2)\frac{\partial A_\phi}{\partial\xi}\right]-\frac{1}{1-\xi^2}A_\phi + f^2(\eta^2-\xi^2)k^2A_\phi = 0. \quad (4)$$

Static Case

The time-independent equation for the vector potential is obtained from Eq. (4) by letting "k" be equal to zero. One may now use the separation-of-variable method of solution.

Let $A_\phi = U(\eta)V(\xi)$, then

$$\frac{\eta^2-1}{U}\frac{d^2U}{d\eta^2}+\frac{2\eta}{U}\frac{dU}{d\eta}-\frac{1}{\eta^2-1} = -\frac{1-\xi^2}{V}\frac{d^2V}{d\xi^2} + \frac{2\xi}{V}\frac{dV}{d\xi}+\frac{1}{1-\xi^2} = \text{a constant},$$

say N.

Using the standard argument that, since the left side of the above equation is a function of η only, and the right side is a function of ξ only, and that the equation must be valid for all values of η and ξ which can vary independently of each other, the expressions must be equal to a single separation constant N, which is independent of both η and ξ. An examination of the above expression with V and ξ reveals that the solutions for V must be regular at $\xi = \pm 1$. Under this condition, it is convenient to choose the form of the separation constant N as $n(n+1)$, where n is a real positive integer or zero,[3] so the solutions may be written as a combination of the associated Legendre functions.

Therefore, the separated equations are

$$\frac{d}{d\xi}\left[(1-\xi^2)\frac{dV}{d\xi}\right]+\left[n(n+1)-\frac{1}{1-\xi^2}\right]V = 0,$$

$$\frac{d}{d\eta}\left[(1-\eta^2)\frac{dU}{d\eta}\right]+\left[n(n+1)-\frac{1}{1-\eta^2}\right]U = 0.$$

The solutions are, in terms of well-known Legendre associated functions,

$$V = C_nP_n^1(\xi)+D_nQ_n^1(\xi),$$

$$U = E_nP_n^1(\eta)+F_nQ_n^1(\eta).$$

The condition that V must be regular at ± 1, dictates that only $P_n^1(\xi)$ is a permissible solution for V. As for the solutions for U, either $P_n^1(\eta)$ or $Q_n^1(\eta)$ or a combination of both, are permissible solutions, depending on the region of validity for these solutions. Also, since n's have discrete values only (positive integers or zero), the complete solution for U's and V's will have to be summations over all n's.

For example, if it is desired to find the static vector potential due to a small loop at η_0, ξ_0, carrying uniform current I, which is perturbed by the presence of a prolate spheroidal magnetic core of surface η_1 = constant ($\eta_1 \leq \eta_0$), and located symmetrically in the chosen coordinate system, then it is possible to select three regions bounded by the surfaces $\eta = \eta_1$ (Region I); $\eta = \eta_1$, $\eta = \eta_0$ (Region II); $\eta = \eta_0$, $\eta = \infty$ (Region III). These are shown in Fig. 1.

Therefore, it is possible to write the allowable solutions for the static vector potentials for the three regions as

$$A_\phi^{(1)} = \sum_{n=0}^{\infty} B_nP_n^1(\xi)P_n^1(\eta) \quad \text{for} \quad \eta \leq \eta_1,$$

$$A_\phi^{(2)} = \sum_{n=0}^{\infty}\left[C_nP_n^1(\xi)P_n^1(\eta)+D_nP_n^1(\xi)Q_n^1(\eta)\right] \quad \text{for} \quad \eta_1 \leq \eta \leq \eta_0, \quad (5)$$

$$A_\phi^{(3)} = \sum_{n=0}^{\infty} E_nP_n^1(\xi)Q_n^1(\eta) \quad \text{for} \quad \eta \geq \eta_0.$$

The superscripts (1), (2), and (3) with A_ϕ represent the region of validity of the solutions. B_n, C_n, etc., are constants to be determined from the proper boundary conditions, in Eq. (2).

At the boundary surface $\eta = \eta_1$, conditions of Eq. (2) reduce to

$$A_\phi^{(1)} = A_\phi^{(2)},$$

$$\frac{1}{\mu}\frac{\partial}{\partial\eta}(h_\phi A_\phi^{(1)})-\frac{1}{\mu_0}\frac{\partial}{\partial\eta}(h_\phi A_\phi^{(2)}) = 0; \quad (6)$$

using Eqs. (3) and (5), the first of these becomes

$$\sum_{n=0}^{\infty}\left[C_nP_n^1(\xi)P_n^1(\eta)+D_nP_n^1(\xi)Q_n^1(\eta)\right]\Big\}_{\eta=\eta_1} = \sum_{n=0}^{\infty} B_nP_n^1(\xi)P_n^1(\eta)\Big\}_{\eta=\eta_1}.$$

Observing now that, since the $P_n(\xi)$'s are orthogonal to each other, and that the above equation must be satisfied for arbitrary ξ's, it is possible to equate this relationship, term-by-term, yielding

$$C_nP_n^1(\eta_1)+D_nQ_n^1(\eta_1) = B_nP_n^1(\eta_1). \quad (7)$$

Proceeding in an exactly analogous manner, the second part of Eq. (6) yields

[3] See, for example, P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Part II* (McGraw-Hill Book Company, Inc., New York, 1953), p. 1285.

$$B_nP_n(\eta_1) - K_mC_nP_n(\eta_1) - K_mD_nQ_n(\eta_1) = 0, \quad (8)$$

where

$K_m = \mu/\mu_0$ is the relative permeability of the core.

Eliminating B_n from (7) and (8),

$$D_n = C_n \frac{(K_m - 1)P_n^1(\eta_1)P_n(\eta_1)}{P_n(\eta_1)Q_n^1(\eta_1) - K_mP_n^1(\eta_1)Q_n(\eta_1)}. \quad (9)$$

Similarly, at $\eta = \eta_0$, Eq. (2) may be written as

$$A_\phi^{(2)} = A_\phi^{(3)}$$

$$\frac{1}{h_\phi h_\eta}\left[\frac{1}{\mu_0}\frac{\partial}{\partial\eta}(h_\phi A_\phi^{(2)}) - \frac{1}{\mu_0}\frac{\partial}{\partial\eta}(h_\phi A_\phi^{(3)})\right] = \text{the true surface current.} \quad (10)$$

Arguments similar to those for obtaining Eq. (7) yield, from the first relationships of Eq. (10),

$$C_nP_n^1(\eta_0) + D_nQ_n^1(\eta_0) = E_nQ_n^1(\eta_0). \quad (11)$$

Before attempting to simplify the second relationship of Eq. (10), it is necessary to express the loop current in a convenient form. Here it is evident that the loop is located on the surface $\eta = \eta_0$, the other coordinate being $\xi = \xi_0$. Other than this loop current, there is no true surface current on the boundary between Regions II and III. It is possible, therefore, to express the true surface current, using the Dirac delta function, as follows:

$$\text{true surface current} = (I/h_\xi)\,\delta(\xi - \xi_0).$$

The scale factor h_ξ, is required in the above expression to obtain the value of the current in the proper form. The second part of Eq. (10) then becomes

$$\frac{\partial}{\partial\eta}(h_\phi A_\phi^{(2)}) - \frac{\partial}{\partial\eta}(h_\phi A_\phi^{(3)}) = \frac{h_\phi h_\eta}{h_\xi}\mu_0 I\,\delta(\xi - \xi_0).$$

Carrying on the differentiation and utilizing Eqs. (3) and (11), this will reduce to

$$\sum_{n=0}^{\infty}\left[C_nP_n^1(\xi)\frac{\partial}{\partial\eta}P_n^1(\eta) + D_nP_n^1(\xi)\frac{\partial}{\partial\eta}Q_n^1(\eta) - E_nP_n^1(\xi)\frac{\partial}{\partial\eta}Q_n^1(\eta)\right] = [(1 - \xi_0^2)/(\eta_0^2 - 1)]^{\frac{1}{2}}\mu I\,\delta(\xi - \xi_0). \quad (12)$$

It is evident here that, among the functions $P_n^1(\eta)$, $Q_n^1(\eta)$, and $P_n^1(\xi)$, only the angular functions $P_n^1(\xi)$ may be made into eigenfunctions, and the radial functions, $P_n^1(\eta)$, $Q_n^1(\eta)$, must, alone, satisfy the boundary conditions, which is also apparent from relationships Eqs. (7), (8), and (11). Assuming the existence of a complete set of eigenfunctions $P_n^1(\xi)$, the Dirac delta functions $\delta(\xi - \xi_0)$ may be expressed as[4]

$$\delta(\xi - \xi_0) = \sum_{n=0}^{\infty}\frac{r}{N_n^2}P_n^1(\xi)P_n^1(\xi_0),$$

where r is a weight factor which is equal to unity in this case, and N_n is a normalizing factor given by[5]

$$\frac{1}{N_n} = \left[\left(n + \frac{1}{2}\right)\frac{(n-1)!}{(n+1)!}\right]^{\frac{1}{2}} = \left[\frac{1}{2}\frac{2n+1}{n(n+1)}\right]^{\frac{1}{2}}.$$

Therefore,

$$\delta(\xi - \xi_0) = \sum_{n=0}^{\infty}\frac{1}{2}\frac{2n+1}{n(n+1)}P_n^1(\xi)P_n^1(\xi_0).$$

Substituting this value of $\delta(\xi - \xi_0)$ in Eq. (12), and noting that, since the equation must be satisfied for arbitrary ξ, and that the $P_n^1(\xi)$'s are a set of orthogonal functions, the equation must also hold for any particular n. Thus, (at $\eta = \eta_0$),

$$C_n\frac{\partial}{\partial\eta}P_n^1(\eta) + D_n\frac{\partial}{\partial\eta}Q_n^1(\eta) - E_n\frac{\partial}{\partial\eta}Q_n^1(\eta) = \left(\frac{1 - \xi_0^2}{\eta_0^2 - 1}\right)^{\frac{1}{2}}\frac{\mu_0 I}{2}\frac{2n+1}{n(n+1)}P_n^1(\xi_0). \quad (13)$$

Using Eqs. (11) and (13) and the following property of the Legendre associated functions,

$$Q_n^1(\eta)\frac{\partial}{\partial\eta}P_n^1(\eta) - P_n^1(\eta)\frac{\partial}{\partial\eta}Q_n^1(\eta) = \frac{in(n+1)}{\eta^2 - 1},$$

finally, one obtains

$$C_n = -i[(1 - \xi_0^2)(\eta_0^2 - 1)]^{\frac{1}{2}} \times \frac{\mu_0 I}{2}\frac{2n+1}{n^2(n+1)^2}P_n^1(\xi_0)Q_n^1(\eta_0). \quad (14)$$

From Eq. (11),

$$E_n = C_n[P_n^1(\eta_0)/Q_n^1(\eta_0)] + D_n.$$

Using this, plus Eqs. (5), (9), and (14), the value of $A_\phi^{(3)}$ becomes

$$A_\phi^{(3)} = [-i[(1 - \xi_0^2)(\eta_0^2 - 1)]^{\frac{1}{2}} \times \frac{\mu_0 I}{2}\sum_{n=0}^{\infty}\frac{2n+1}{n^2(n+1)^2}P_n^1(\xi_0)P_n^1(\eta_0)P_n^1(\xi)Q_n^1(\eta) - i[(1 - \xi_0^2)(\eta_0^2 - 1)]^{\frac{1}{2}}\frac{\mu_0 I}{2}\sum_{n=0}^{\infty}\frac{2n+1}{n^2(n+1)^2}P_n^1(\xi_0) \times Q_n^1(\eta_0)\cdot\frac{(K_m - 1)P_n^1(\eta_1)P_n(\eta_1)P_n^1(\xi)Q_n^1(\eta)}{P_n(\eta_1)Q_n^1(\eta_1) - K_mP_n^1(\eta_1)Q_n(\eta_1)}\Bigg]. \quad (15)$$

Thus, $A_\phi^{(3)}$ consists of two parts, the first part due

[4] See, for example, P. M. Morse and H. Feshbach, *Methods of Theoretical Physics Part I* (McGraw-Hill Book Company, Inc., New York, 1953), p. 828–832; also p. 729.

[5] See, for example, reference 3, p. 1274.

to the loop only, and the second part due to the presence of the permeable core. This result may be compared with a previous result.[6]

Time-Varying Case

In Eq. (4), by letting $A_\phi = U(\eta)V(\xi)$, the separated equations are obtained. These are

$$\frac{d}{d\eta}\left[(1-\eta^2)\frac{dU}{d\eta}\right] + \left(B - h^2\eta^2 - \frac{1}{1-\eta^2}\right)U = 0,$$

$$\frac{d}{d\xi}\left[(1-\xi^2)\frac{dV}{d\xi}\right] + \left(B - h^2\xi^2 - \frac{1}{1-\xi^2}\right)V = 0,$$

where B is the separation constant and $h = fk$. It may be noted that, even though U and V satisfy similar equations, as in the static case, the solution for U involves its behavior in the range $+1$ to ∞, whereas the solution for V involves its behavior between the singular points -1 and $+1$.

Referring to Eq. (5) and Fig. 1, and following similar arguments as in the static case, solutions can be written for the retarded vector potentials for the three regions, as follows:

$$A_\phi^{(1)} = \sum_l C_l S_{1l}(h, \xi) je_{1l}(h, \eta) \quad \text{for} \quad \eta \leq \eta_1,$$

$$A_\phi^{(2)} = \sum_l [D_l S_{1l}(h_0, \xi) je_{1l}(h_0, \eta) + E_l S_{1l}(h_0, \xi) he_{1l}(h_0, \eta)] \quad \text{for} \quad \eta_0 \geq \eta \geq \eta_1, \tag{16}$$

$$A_\phi^{(3)} = \sum_l F_l S_{1l}(h_0, \xi) he_{1l}(h_0, \eta) \quad \text{for} \quad \eta \geq \eta_0,$$

where $h_0 = fk_0 = f\omega(\mu_0\epsilon_0)^{\frac{1}{2}}$. C_l, D_l, etc., are constants to be determined from the proper boundary conditions.

Functions, $S_{1l}(h, \xi)$ and $S_{1l}(h_0, \xi)$ are known as the spheroidal angle functions, and are related to associated Legendre functions. Functions $je_{1l}(h, \eta)$, $he_{1l}(h, \eta)$, etc., are known as the spheroidal radial functions, and are related to the spherical Bessel and Hankel functions.[7] For convenience, the same notations as used in reference 7 have been used here also; additional information may be obtained in references 8 and 9.

As in the static cases, here again the boundary conditions in Eq. (2) at $\eta = \eta_1$ are applied, and, following similar arguments, the condition

$$A_\phi^{(1)} = A_\phi^{(2)}$$

applied to Eq. (16) reduces to

$$[D_l je_{1l}(h_0, \eta_1) + E_l he_{1l}(h_0, \eta_1)]S_{1l}(h_0, \xi) = C_l je_{1l}(h, \eta_1)S_{1l}(h, \xi). \tag{17}$$

Similarly, the second part of Eq. (2) applied to Eq. (16), yields

$$\begin{aligned} C_l S_{1l}(h, \xi)&\left[\frac{d}{d\eta} je_{1l}(h, \eta) + je_{1l}(h, \eta)\frac{\eta}{\eta^2 - 1}\right]_{\eta=\eta_1} \\ &= K_m S_{1l}(h_0, \xi)\Big\{D_l\Big[\frac{d}{d\eta} je_{1l}(h_0, \eta) + \frac{\eta}{\eta^2 - 1} je_{1l}(h_0, \eta)\Big] \\ &\quad + E_l\Big[\frac{d}{d\eta} he_{1l}(h_0, \eta) + \frac{\eta}{\eta^2 - 1} he_{1l}(h_0, \eta)\Big]\Big\}_{\eta=\eta_1}. \end{aligned} \tag{18}$$

Eliminating C_l from Eq. (17) and Eq. (18), finally,

$$\begin{aligned} E_l = D_l\Big\{&je_{1l}(h_0, \eta)S_{1l}(h_0, \xi)\Big[\frac{d}{d\eta} je_{1l}(h, \eta) + \frac{n}{\eta^2 - 1} je_{1l}(h, \eta)\Big]S_{1l}(h, \xi) \\ &- K_m S_{1l}(h_0, \xi)\Big[\frac{d}{d\eta} je_{1l}(h_0, \eta) + \frac{\eta}{\eta^2 - 1} je_{1l}(h_0, \eta)\Big]S_{1l}(h, \xi) je_{1l}(h, \eta)\Big\} \\ &\times \Big\{K_m S_{1l}(h_0, \xi)\Big[\frac{d}{d\eta} he_{1l}(h_0, \eta) + \frac{\eta}{\eta^2 - 1} he_{1l}(h_0, \eta)\Big]S_{1l}(h, \xi) je_{1l}(h, \eta) \\ &- S_{1l}(h, \xi) he_{1l}(h_0, \eta)S_{1l}(h_0, \xi) \\ &\times \Big[\frac{d}{d\eta} je_{1l}(h, \eta) + \frac{\eta}{\eta^2 - 1} je_{1l}(h, \eta)\Big]\Big\}^{-1}_{\eta=\eta_1}. \end{aligned} \tag{19}$$

Some comments about the range of validity of (17), (18), and (19) are in order. Since the factor h is unequal to the factor h_0, $S_{1l}(h, \xi)$ is different from $S_{1l}(h_0, \xi)$ in general. They are not orthogonal to each other in any way. Therefore, if from (16) one equates $A_\phi^{(1)} = A_\phi^{(2)}$, the following is obtained for the boundary at $\eta = \eta_1$:

$$\sum_l C_l S_{1l}(h, \xi) je_{1l}(h, \eta_1) = \sum_l [D_l je_{1l}(h_0, \eta_1) + E_l he_{1l}(h_0, \eta_1)]S_{1l}(h_0, \xi).$$

[6] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1950), problem 25, p. 302.

[7] Reference 3, pp. 1502–1511.

[8] J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and F. J. Corbató, *Spheroidal Wave Functions* (The Technology Press, MIT, Cambridge, Massachusetts, and John Wiley & Sons, Inc., New York, 1956).

[9] C. Flammer, *Spheroidal Wave Functions* (Stanford University Press, Stanford, California, 1957).

If both sides of above relationship are multiplied by $S_{1m}(h_0, \xi)$ and both sides are integrated with respect to ξ between the limits -1 to $+1$, then one obtains

$$D_l je_{1l}(h_0, \eta_1) + E_l he_{1l}(h_0, \eta_1) = \frac{1}{\Lambda_{1l}(h_0)} \times \int_{-1}^{1} \sum_l C_l je_{1l}(h, \eta_1) S_{1l}(h, \xi) S_{1m}(h_0, \xi)\, d\xi.$$

This is the result of the orthogonality conditions

$$\int_{-1}^{1} [S_{1l}(h_0, \xi)]^2\, d\xi = \Lambda_{1l}(h_0),$$

$$\int_{-1}^{1} S_{1l}(h_0, \xi) S_{1m}(h_0, \xi)\, d\xi = 0, \quad m \neq l,$$

where $\Lambda_{1l}(h_0)$ is a factor as defined in reference 7.

Recalling that in the present case, only the low-frequency solution is of interest, (in fact, only under this assumption, the vector wave equation was separable in the prolate sphroidal coordinate system), it may be assumed that h^2 is very very much smaller than unity. A study of the behavior of the angular functions $S_{1l}(h, \xi)$ reveal that for very very small h^2, $S_{1l}(h, \xi) \sim S_{1l}(h_0, \xi) \sim P_l^1(\xi)$.

Under this restriction, the coupling between the different modes of the angular functions may be neglected without involving any appreciable error and we obtain the relationships (17), (18), and (19). No such coupling, however, appears at the boundary between Regions II and III. In a similar manner, at $\eta = \eta_0$, the first part of the boundary condition of Eq. (2), i.e.,

$$A_\phi^{(3)} = A_\phi^{(2)},$$

yields

$$D_l je_{1l}(h_0, \eta_0) + E_l he_{1l}(h_0, \eta_0) = F_l he_{1l}(h_0, \eta_0), \quad (20)$$

and the second part of Eq. (2) becomes, as in the static case,

$$\frac{\partial}{\partial \eta} [h_\phi A_\phi^{(2)}] - \frac{\partial}{\partial \eta} [h_\phi A_\phi^{(3)}] = \frac{h_\phi h_\eta}{h_\xi} \mu_0 I \delta(\xi - \xi_0) \quad \text{at} \quad \eta = \eta_0. \quad (21)$$

Here, again, it is evident that only the angular functions $S_{1l}(h_0, \xi)$ could be made into eigenfunctions, and the radial functions $je_{1l}(h_0, \eta)$, $he_{1l}(h_0, \eta)$ must, alone, satisfy the boundary conditions. Hence, exactly as in the static case, it is possible to expand the Dirac delta functions as a complete set of the eigenfunctions[7]:

$$\delta(\xi - \xi_0) = \sum_l \frac{1}{\Lambda_{1l}(h_0)} S_{1l}(h_0, \xi) S_{1l}(h_0, \xi_0),$$

where $\Lambda_{1l}(h_0)$ is the square of the normalizing factor defined by the equation $\Lambda_{1l}(h_0) = \int_{-1}^{1} |S_{1l}|^2\, d\xi$. The notation used is the same as in reference 7, which may be consulted for further details about this function.

Substituting this expression for the delta functions in Eq. (21), and equating term by term, as in the static case,

$$\left\{ -F_l \frac{d}{d\eta} he_{1l}(h_0, \eta) + D_l \frac{d}{d\eta} je_{1l}(h_0, \eta) + E_l \frac{d}{d\eta} he_{1l}(h_0, \eta) \right\}_{\eta=\eta_0} = \left(\frac{1 - \xi_0^2}{\eta_0^2 - 1} \right)^{\frac{1}{2}} \mu_0 I \frac{S_{1l}(h_0, \xi_0)}{\Lambda_{1l}(h_0)}. \quad (22)$$

Solving for D_l from Eqs. (20) and (22) leads to

$$\frac{D_l}{he_{1l}(h_0, \eta_0)} \left[je_{1l}(h_0, \eta) \frac{d}{d\eta} he_{1l}(h_0, \eta) - he_{1l}(h_0, \eta) \frac{d}{d\eta} je_{1l}(h_0, \eta) \right]_{\eta=\eta_0} = -\left(\frac{1 - \xi_0^2}{\eta_0^2 - 1} \right)^{\frac{1}{2}} \mu_0 I \frac{S_{1l}(h_0, \xi_0)}{\Lambda_{1l}(h_0)}. \quad (23)$$

It can easily be shown that the Wronskian

$$je_{1l} \frac{d}{d\eta} he_{1l} - he_{1l} \frac{d}{d\eta} je_{1l} = \frac{i}{h_0(\eta^2 - 1)}.$$

This applied to Eq. (23) immediately yields

$$D_l = i\mu_0 I h_0 [(1 - \xi_0^2)(\eta_0^2 - 1)]^{\frac{1}{2}} \frac{1}{\Lambda_{1l}(h_0)} \times S_{1l}(h_0, \xi_0) he_{1l}(h_0, \eta_0). \quad (24)$$

From Eq. (20),

$$F_l = D_l \frac{je_{1l}(h_0, \eta_0)}{he_{1l}(h_0, \eta_0)} + E_l.$$

Therefore,

$$A_\phi^{(3)} = i\mu_0 I h_0 [(1 - \xi_0^2)(\eta_0^2 - 1)]^{\frac{1}{2}} \times \sum_l \frac{1}{\Lambda_{1l}(h_0)} je_{1l}(h_0, \eta_0) S_{1l}(h_0, \xi_0) \times S_{1l}(h_0, \xi) he_{1l}(h_0, \eta) + \sum_l E_l S_{1l}(h_0, \xi) he_{1l}(h_0, \eta), \quad (25)$$

where E_l is obtained by substituting Eq. (24) into Eq. (19).

Discussion

As in the previous case, it is evident that the first part of the expression for $A_\phi^{(3)}$ in Eq. (25) is the vector potential due to the loop only, and the second part is that due to the presence of the permeable core. If, for example, in the expression for E_l in Eq. (19), h is allowed to equal h_0, then $K_m = 1$, and E_l reduces to zero, and Eq. (25) is left with the expression for the loop only. This is, of course, what would be expected. If it is desirable to find the retarded vector potential due to a loop only, of radius a, carrying current I, and located at $v = \frac{1}{2}\pi$, i.e., $\xi_0 = 0$, then, recalling the equation for the ellipse,

$$[z/(f\cosh u)]^2 + [r/(f\sinh u)]^2 = 1,$$

it is recognizable that

$f\cosh u = \frac{1}{2}$ major axis,

$f\sinh u = \frac{1}{2}$ minor axis.

Since at $\xi_0 = 0$, the loop radius is a, therefore,

$$f\sinh u = a,$$

$$\eta_0 = \cosh u_0 = (\sinh^2 u_0 + 1)^{\frac{1}{2}} = (a^2 + f^2)^{\frac{1}{2}}/f,$$

$$(\eta_0^2 - 1)^{\frac{1}{2}} = a/f, \qquad (1 - \xi_0^2)^{\frac{1}{2}} = 1, \qquad h_0 = fk_0.$$

Substituting these in the first part of Eq. (25), one obtains the expression for $A_\phi^{(3)}$ for the loop only: thus,

$$A_\phi^{(3)} = i\mu_0 I a k_0 \sum_l \frac{1}{\Lambda_{1l}(h_0)} S_{1l}(h_0, \xi_0) \times S_{1l}(h_0, \xi) je_{1l}(h_0, \eta_0) he_{1l}(h_0, \eta). \quad (26)$$

Using the integral representation for the vector potential, and the expansion of the three-dimensional Green's function in a prolate spheroidal coordinate system,[7] it is easily possible to arrive at the same expression as in Eq. (26), as follows.

From reference (7), the expansion of the Green's function is (with $m = 1$, $\eta > \eta_0$)

$$\frac{e^{ik_0R}}{R} = 2ik_0 \sum_l \frac{2}{\Lambda_{1l}(h_0)} S_{1l}(h_0, \xi_0) S_{1l}(h_0, \xi) \times \cos(\phi - \phi_0) je_{1l}(h_0, \eta_0) he_{1l}(h_0, \eta);$$

also, from the integral representation of retarded vector potential ($e^{-i\omega t}$ understood),

$$A_\phi = \frac{\mu_0}{4\pi} \oint_V \frac{[J]_{t-r/v}\, d\tau}{R} = \frac{\mu_0 I}{4\pi} \oint \frac{e^{ik_0R}}{R} dl$$

$$= \frac{\mu_0 I a}{4\pi} \int_0^{2\pi} \frac{e^{ik_0R}}{R} \cos(\phi - \phi_0)\, d(\phi - \phi_0)$$

$$= \frac{\mu_0 I a}{4\pi} 2ik_0 \sum_l \frac{2}{\Lambda_{1l}(h_0)} S_{1l}(h_0, \xi_0) \times S_{1l}(h_0, \xi) je_{1l}(h_0, \eta_0) he_{1l}(h_0, \eta). \times \int_0^{2\pi} \cos^2(\phi - \phi_0)\, d(\phi - \phi_0)$$

$$= i\mu_0 I a k_0 \sum_l \frac{1}{\Lambda_{1l}(h_0)} S_{1l}(h_0, \xi_0) \times S_{1l}(h_0, \xi) je_{1l}(h_0, \eta_0) he_{1l}(h_0, \eta).$$

This is, of course, the same expression as in Eq. (26).

ACKNOWLEDGMENT

The author wishes to acknowledge his indebtedness to Professor H. Feshbach of Massachusetts Institute of Technology, and to many of the members of the technical staff of LFE Electronics, a division of Laboratory For Electronics, Inc., Boston, Massachusetts, for their helpful comments and criticisms on various items in this paper.

Addendum: Combinatorial Aspects of the Ising Model for Ferromagnetism. I. A Conjecture of Feynman on Paths and Graphs

[J. Math. Phys. 1, 202 (1960)]

S. SHERMAN
Department of Mathematics, Wayne State University, Detroit, Michigan

(Received 6 June 1963)

A gap in the proof of Theorem 1 of the cited paper is filled.

PROFESSOR M. P. Schutzenberger has indicated to the author that the argument in the cited article on p. 207, 2nd column is incomplete. The following substitute fills the gap:

On p. 207, 2nd column, in line 13 appears a sentence ending "is greater than one". Continue with "To that end, temporarily consider $\prod^* [1 + W(p)]$, where now the product is over all aperiodic *cdls* and so is the square of the left-hand side of Eq. (1). Introduce the order $D_1 < D_1^{-1} < D_2 < \cdots$ among the directed line segments. With each aperiodic *cdls* is associated a word $w(p)$ in the D's, which is lexicographically least among its cyclic permutations. Let $W(w) = {}_{df}W(p)$. Let A_1 be the set of aperiodic *cdls* p such that D_1 occurs in p. For each $p \in A_1$, the corresponding word admits a unique factorization into words, each of which begins with D_1, and has no other appearances of D_1. Some of these words may be of the form $D_1\alpha$, where neither D_1 nor D_1^{-1} occurs in α and others may be of the form

$$D_1\alpha D_1^{-1}\gamma_1 D_1^{-1}\gamma_2 \cdots D_1^{-1}\gamma_k,$$

where neither D_1 nor D_1^{-1} occurs in $\alpha, \gamma_1, \cdots, \gamma_k$. Since

$$\begin{aligned} W(D_1\alpha D_1^{-1}\gamma_1 \cdots D_1^{-1}\gamma_k) &\\ + W(D_1\alpha^{-1}D_1^{-1}\gamma_1 \cdots D_1^{-1}\gamma_k) &\\ + W(D_1\alpha D_1^{-1}\gamma_1 \cdots D_1^{-1}\gamma_k^{-1}) &\\ + W(D_1\alpha^{-1}D_1^{-1}\gamma_1 \cdots D_1^{-1}\gamma_k^{-1}) &= 0, \end{aligned}$$

the application of the Witt identity[1] yields

$$\prod_{p\in A_1} [1 + W(p)] = 1 + d_1\epsilon_{11},$$

where ϵ_{11} is a formal (possibly infinite) sum of monomials none of which has d_1 as a factor. Let A_2 be the set of aperiodic *cdls* p such that D_1^{-1} occurs in p. Note that in general $A_1 \cap A_2 \neq \phi$. For each $p \in A_2$, the corresponding word admits a unique factorization into words each of which begins with D_1^{-1}, and has no other appearances of D_1^{-1}. Some of these words may be of the form $D_1^{-1}\alpha$, where neither D_1^{-1} nor D_1 occurs in α and others may be of the form $D_1^{-1}\alpha D_1\gamma_1 \cdots D_1\gamma_k$, where neither D_1^{-1} nor D_1 occurs in $\alpha, \gamma_1, \cdots, \gamma_k$. The analogue of the earlier argument yields

$$\prod_{p\in A_2} [1 + W(p)] = 1 + d_1\epsilon_{11},$$

where ϵ_{11} is the same sum as before.

Let B be the set of aperiodic *cdls* in which neither D_1 nor D_1^{-1} appears. Then

$$\prod_{\in Bp} [1 + W(p)] = (1 + \epsilon_{12})^2,$$

where ϵ_{12} is a formal sum of monomials, none of which has d_1 as a factor.

In

$$\{\prod_{p\in A_1} [1 + W(p)]\}\{\prod_{p\in A_2} [1 + W(p)]\} = (1 + d_1\epsilon_{11})^2,$$

cdls $p \in A_1 \cap A_2$ have been counted doubly while *cdls* $pA_1\backslash A_2$ (relative complement) and *cdls* $p \in A_2\backslash A_1$ have been counted singly. By an application of the Witt identity[1]

$$1 + d_1\epsilon_{11} = \prod_{p\in A_1\backslash A_2} [1 + W(p)] = \prod_{p\in A_2\backslash A_1} [1 + W(p)],$$

$$\{\prod_{p\in A_1\cup A_2} [1 + W(p)]\}^2 = \{\prod_{p\in A_1} [1 + W(p)]\} \times \{\prod_{p\in A_2} [1 + W(p)]\}\{\prod_{p\in A_1\backslash A_2} [1 + W(p)]\}$$

$$\{\prod_{p\in A_2\backslash A_1} [1 + W(p)]\} = (1 + d_1\epsilon_{11})^4,$$

$$\prod^* [1 + W(p)] = \{\prod_{p\in A_1\cup A_2} [1 + W(p)]\} \times \{\prod_{p\in B} [1 + W(p)]\} = (1 + d_1\epsilon_{11})^2(1 + \epsilon_{12})^2,$$

[1] S. Sherman, Bull. Am. Math. Soc. 68, 225 (1962), Eq. 1.

and so

$$\prod{}^{*}(1 + W[p]) = (1 + d_1\epsilon_{11})(1 + \epsilon_{12})$$
$$= 1 + d_1(\epsilon_{11} + \epsilon_{11}\epsilon_{12}) + \epsilon_{12}.$$

Thus, there are no monomial summands of the right-hand side having factors of the form d_1^n with $n \geq 2$. Analogous arguments dispose of monomial summands with factors of the form d_j^n, $n \geq 2$ for each j, $1 \leq j \leq a$."

This completes the proof and reader can go to the section "Corollary on Coin Arrangements" on p. 208.