JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 2, NUMBER 4 JULY–AUGUST, 1961

Strict Localization in Quantum Field Theory*

JAMES M. KNIGHT
University of Maryland, College Park, Maryland
(Received February 2, 1961)

A definition of strict localization of states in quantum field theory is presented. This definition is based on considering products of field operators as the primary measurable quantities of the theory. An example of a localized state is given, showing that such a state arises when a free field interacts with an external current that is limited to a bounded region of space-time. It is shown by means of a graphical technique that a state having a finite number of particles cannot satisfy the definition of localization. A simple representation of localized states is investigated, and arguments are given to support its generality and uniqueness.

1. INTRODUCTION

THE research in quantum field theory of the past ten years has centered chiefly about the analytic properties of various quantities appearing in the theory.[1] These properties are derived from very general characteristics of the fields stated in the form of postulates which it is believed that any complete theory must satisfy. It is not altogether clear, however, that these postulates, as formulated for example by Wightman,[2] form a consistent system, or that they form a minimal basis from which the analytic properties follow. It is, therefore, of great interest at the present time to carry out investigations having for their ultimate aim the clarification of these postulates.

The earliest derivations of analytic properties for relativistic field theories were based on the finite propagation velocity of wave disturbances, and took into consideration the scattering of initially separated localized wave packets.[3] These packets propagate towards one another with finite velocity and interact, giving rise to scattered waves. The condition that no scattered waves appear before the initial wave packets have had time to collide is then sufficient to give analytic properties for the scattering amplitude as a function of the wave number k.

In later derivations of analytic properties, carried out in the formalism of quantum field theory, the principle of finite propagation velocity was replaced by the condition of local commutativity of the field operators at space-like separations,[4] together with certain other postulates such as the asymptotic condition. This latter condition is an expression of the circumstance that particles involved in a scattering process behave as separated and noninteracting at times in the distant past and future before and after the scattering has taken place. It thus replaces the wave-packet description of the scattering process used in earlier derivations. The recent work of Haag[5] has gone far in clarifying the status of the asymptotic condition, although it is still not clear in precisely what form this condition is satisfied.

It is clear that the question of localization of states is fundamental to the above considerations. To gain an

* Work supported by the National Science Foundation and the U. S. Air Force Office of Scientific Research. This paper is based on a dissertation submitted to the faculty of the University of Maryland in partial fulfillment of the requirements for a Ph.D. degree.

[1] For a review of this work, see for example, the account of the Verenna course on mathematical problems of the quantum theory of fields, published in Nuovo cimento **14**, Suppl. (1959).

[2] A. S. Wightman, Phys. Rev. **101**, 860 (1956), and Ecole Normale Lecture Notes (1957).

[3] J. S. Toll, thesis, Princeton University (1952), and Phys. Rev. **104**, 1760 (1957); N. G. van Kampen, *ibid.* **89**, 1073 (1953); **91**, 1267 (1953).

[4] M. Gell-Mann, M. L. Goldberger, and W. E. Thirring, Phys. Rev. **91**, 1612 (1954).

[5] R. Haag, *Les Problèmes Mathématique de la Théorie Quantique des Champs* (Lille, 1957); Phys. Rev. **112**, 669 (1958), and the article cited in reference 1. See also D. Kastler, Compt. rend. acad. sci. Paris **245**, 2021 (1957).

idea of what localization means, one must have an idea of the significance of the measuring process. Newton and Wigner took up the question of localization from the point of view of position measurements of a particle.[6] They set up postulates from which the position eigenstates of a particle could be determined. Their postulates were based on the quantum theoretical description of a measuring process and upon relativistic invariance of the wave function describing the particle. In order that the position measurement have meaning, eigenstates corresponding to different spatial positions at a fixed time must be orthogonal. This condition is sufficient to determine the position eigenstates completely. However, the definition of localization implied by these position eigenstates is found not to be preserved in time, i.e., a particle localized at a point at one time will be spread over all space, even outside the light cone of the initial point, at later times. Furthermore, a particle localized in one Lorentz frame is not necessarily localized in another. These rather unsatisfactory features of the results of Newton and Wigner make it difficult to see how a suitable concept of localization of particles may be defined in relativistic field theory.

Haag[5] has discussed the asymptotic condition in field theory by means of a definition of localization in which two states are localized if they become orthogonal as the space-like separation of their respective regions of localization becomes infinite. This definition is less restrictive than the one adopted in this paper, which is based on strict localization of a state determined by measurement.

A basic treatment of measurements in field theory was given by Bohr and Rosenfeld,[7] who showed that averages of the field variables over space-time regions may be taken as the basic measurable quantities of the theory. This view seems appropriate also to the recent postulational developments of field theory, where vacuum expectation values of certain combinations of field variables are treated as fundamental.

The definition of localization given in Sec. 2 is based upon taking products of field operators, instead of particle observables, as the basic measurable quantities. The remainder of the paper is devoted to an analysis of the definition and a discussion of some of the properties of the states satisfying the definition.

2. DEFINITION OF LOCALIZATION

In this section, we present and discuss the definition of localization that forms the basis of this paper. We take the point of view that the basic measurable quantities of the theory are products of field variables $A(x)$. By this we mean that any observable quantity Q shall be expressible in the following form as sums of integrals of such products:

$$Q=q_0+\int dx_1 q_1(x_1)A(x_1)+\int\int dx_1 dx_2 q_2(x_1,x_2) \times A(x_1)A(x_2)+\cdots, \quad (1)$$

q_i being c-number functions of their space-time arguments.[8] The expectation value of Q in any state $|\psi\rangle$ can then be written in terms of the quantities

$$\langle\Psi|\prod_{i=1}^{n}A(x_i)|\Psi\rangle$$

in a form similar to (1). We wish to formulate the condition that $|\Psi\rangle$ represent a state of the field which is strictly localized in a region $\mathfrak{U}$ of space-time. If the phenomena described by and the field quantity $A(x)$ are confined to $\mathfrak{U}$, it should not be possible to detect the presence of any field disturbance by making measurements at points outside of $\mathfrak{U}$. In other words, such measurements should lead to the same results whether the state of the system is $|\Psi\rangle$ or the vacuum state $|0\rangle$. Taking into consideration the nature of the measurable quantities (1), we arrive at the following definition:

Definition of localization. A state $|\Psi\rangle$ of a field $A(x)$ is localized in the region $\mathfrak{U}$ if

$$\langle\Psi|\prod_{i=1}^{n}A(x_i)|\Psi\rangle=\langle 0|\prod_{i=1}^{n}A(x_1)|0\rangle, \quad n=1,2,\cdots, \quad (2)$$

for any product of field operators taken at points x_i all of which lie outside of $\mathfrak{U}$.

If there is more than one independent field in the theory, the definition may be extended to include all possible products of any combination of the fields. The definition is meant to apply to fields satisfying the Bose-Einstein statistics. For Fermi-Dirac fields, the measurable quantities are bilinear expressions[9] in the field operators, such as charge and current densities, and the definition must be modified accordingly. We have limited our investigation to the case of a single field satisfying Bose-Einstein statistics. This case contains the essential elements of the problem, and an extension to other cases should not present any difficulties.

For the most part, the analysis that follows is carried out for the case of free fields. The definition, however, is not so restricted, and it is pointed out explicitly wherever our arguments clearly have more general validity. We do not consider the limitations to free fields to be a very serious one, since we are interested in localized states primarily as initial and final states in scattering processes. We expect that in such cases the particles involved in the process are far separated and are therefore describable by free fields.

[6] R. G. Newton and E. P. Wigner, Revs. Modern Phys. **21**, 400 (1949).

[7] N. Bohr and L. Rosenfeld, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **12**, No. 8 (1933), and Phys. Rev. **78**, 794 (1950).

[8] Four-dimensional integrals will be denoted by the symbol dx, three-dimensional integrals by d^3x.

[9] N. Bohr and L. Rosenfeld, Phys. Rev. **78**, 794 (1950).

It is possible to make a statement about the nature of the region $\mathfrak{U}$ in which a state may be localized without entering into the details of the structure of the state itself. We note that the quantity

$$\Psi(x_1,\cdots,x_n)=\langle\Psi|\prod_{i=1}^{n}A(x_i)|\Psi\rangle-\langle 0|\prod_{i=1}^{n}A(x_i)|0\rangle$$

is a solution of the free field equation in each of its arguments independently. It follows from this that if $f(x)$ is an arbitrary solution of the Klein-Gordon equation vanishing sufficiently rapidly for large spatial separations at a given time, then the expression

$$\int_{x_0=t}d^3x_1\left(f(x_1)\frac{\partial}{\partial x_{10}}\Psi(x_1,\cdots,x_n)-\frac{\partial f(x_1)}{\partial x_{10}}\Psi(x_1,\cdots,x_n)\right)$$

is independent of t for fixed x_2, $\cdots$, x_n. This makes it clear that if $\Psi(x_1,\cdots,x_n)$ is not identically zero, it must be nonvanishing in some region of space at any given time. Since (2) requires $\Psi(x_1,\cdots,x_n)$ to vanish outside $\mathfrak{U}$, this means that $\mathfrak{U}$ cannot be bounded in the time-like direction.

The actual form of the region of localization may be also inferred from similar considerations. The solution $\Psi(x)=\langle\Psi|A(x)|\Psi\rangle-\langle 0|A(x)|0\rangle$ of the Klein-Gordon equation may be expressed in terms of its initial values on a surface $x_0=t$ as follows:

$$\Psi(x)=-\int_{x_0=t}d^3x\left(\Delta(x-x')\frac{\partial\Psi(x')}{\partial x_0'}-\frac{\partial\Delta(x-x')}{\partial x_0'}\Psi(x')\right)$$

by means of the singular function $\Delta(x-x')$.[10] If at a time t, $\Psi(x)$ and $\partial\Psi(x)/\partial x_0$ are confined to a region $\mathfrak{R}$ of space, then $\Psi(x)$ for arbitrary times will be confined to the region $V_+(\mathfrak{R})$ consisting of $\mathfrak{R}$ together with the interiors of all the forward and backward light cones with vertices in $\mathfrak{R}$. This is clear from the property that the Δ function vanishes for space-like argument. Note that $\mathfrak{R}$ may be taken as any bounded space-time region instead of a space-like surface without changing the manifold of possible regions of localization.

3. FIELD INTERACTING WITH CLASSICAL CURRENT DISTRIBUTION

To provide an example of a localized state, we consider the interaction of a field $\phi(x)$ with a classical current distribution $j(x)$. The equation of motion for $\phi(x)$ is[11]

$$(\Box-m^2)\phi(x)=-j(x).$$

We may express $\phi(x)$ as follows in terms of free fields $\phi_{\text{in}}(x)$ and $\phi_{\text{out}}(x)$

$$\phi(x)=\phi_{\text{in}}(x)+\int\Delta_R(x-x')j(x')dx'$$

$$=\phi_{\text{out}}(x)+\int\Delta_A(x-x')j(x')dx',$$

$$\phi_{\text{out}}=\phi_{\text{in}}(x)-\int\Delta(x-x')j(x')dx'.$$

We may now introduce two complete sets of states, obtained from the in and out fields, respectively. Each of these sets may be labeled by the eigenvalues of the number operators and the momentum operators. Because of the action of the current $j(x)$, the particle configuration will change with time so that at large times there will be a different configuration, represented as some linear combination of the outgoing states. For example,[11] if the state of the system is $|0_{\text{in}}\rangle$, i.e., no particles initially, then an emission of quanta will take place such that the probability of finding n outgoing quanta is the state of momentum $\mathbf{k}$ is given by the Poisson law

$$w_n(\mathbf{k})=(n!)^{-1}(\langle n(\mathbf{k})\rangle)^n\exp(-\langle n(\mathbf{k})\rangle),$$

where $\langle n(\mathbf{k})\rangle$, the average number of particles of momentum $\mathbf{k}$, is proportional to the quantity $|j(\mathbf{k},\omega)|^2$, where $\omega=+(\mathbf{k}^2+m^2)^{\frac{1}{2}}$, and $j(\mathbf{k},\omega)$ is the Fourier transform of $j(x)$. Note that the outgoing configuration depends only on those components of the Fourier transform which satisfy $\mathbf{k}^2-\omega^2=-m^2$, and not on components off the mass shell. We shall refer to this fact in Sec. 5.

We now make the following identification:

$$A(x)=\phi_{\text{out}}(x),$$
$$|0\rangle=|0_{\text{out}}\rangle,$$
$$|\Psi\rangle=|0_{\text{in}}\rangle.$$

We will now show that, if $j(x)=0$ outside of the space-time region $\mathfrak{R}$, then $|\Psi\rangle$ is a localized state of the field $A(x)$ in $V_+(\mathfrak{R})$ by definition (2). We evaluate the quantity

$$\langle\Psi|\prod_{i=1}^{n}A(x_i)|\Psi\rangle$$

$$=\langle 0_{\text{in}}|\prod_{i=1}^{n}\phi_{\text{out}}(x_i)|0_{\text{in}}\rangle$$

$$=\langle 0_{\text{in}}|\prod_{i=1}^{n}\left(\phi_{\text{in}}(x_i)-\int dx_i'\Delta(x_i-x_i')j(x_i')\right)|0_{\text{in}}\rangle$$

$$=\langle 0_{\text{in}}|\prod_{i=1}^{n}\phi_{\text{in}}(x_i)|0_{\text{in}}\rangle$$

$$-\sum_{j=1}^{n}\int dx_j'\Delta(x_j-x_j')j(x_j')$$

$$\times\langle 0_{\text{in}}|\prod_{i\neq j=1}^{n}\phi_{\text{in}}(x_i)|0_{\text{in}}\rangle+\cdots. \quad (3)$$

[10] See, for example, G. Källén's article in *Handbuch der Physik* (Springer-Verlag, Berlin, Germany, 1958), Vol. V, Part 1.

[11] See, for example, Källén's article on quantum electrodynamics (reference 10).

The first term is the vacuum expectation value of a product of free fields, and is thus equal to

$$\langle 0_{\text{out}}|\prod_{i=1}^{n}\phi_{\text{out}}(x_i)|0_{\text{out}}\rangle=\langle 0|\prod_{i=1}^{n}A(x_i)|0\rangle.$$

The remaining terms on the right each contain as a factor a quantity $\int dx_j'\Delta(x_j-x_j')j(x_j')$ for some x_j. This quantity vanishes unless $x_j\in V_{\pm}(\mathcal{R})$. Therefore, all of the terms on the right vanish except the first term when all x_j lie outside $V_{\pm}(\mathcal{R})$. For this case, we obtain

$$\langle\Psi|\prod_{i=1}^{n}A(x_i)|\Psi\rangle=\langle 0|\prod_{i=1}^{n}A(x_i)|0\rangle,$$

which is just (2).

The states (3) are thus localized states if the generating current operates in a confined region of space-time. They may be thought of as representatives of the type of state which occurs in a physical scattering process. The field of the apparatus which produces and accelerates the particles to be scattered is the external field $j(x)$. The particles then propagate from the region of production according to the free equations of motion. The apparatus for detecting scattered particles may also be thought of as an arrangement of fields and currents located in a bounded region of space-time. In the theoretical treatment of the scattering of fields, the initial and final states are usually idealized to one-particle states of definite momentum. Such states are not localized, and cannot be produced by apparatus confined to a bounded region of space-time. The localization associated with production and detection is accounted for in single-particle scattering theory by using a wave-packet description of initial and final states. It is in this sense that the localized states defined here provide a certain field theoretic analog of the wave-packet description.

4. LOCALIZED STATES IN TERMS OF NUMBER AND MOMENTUM EIGENFUNCTIONS

Having found that localized states may be generated by an external source interacting with the field within a bounded domain of space-time, we pass on to a further analysis of the structure of states satisfying the criterion of localization (2). For this purpose, we consider an expansion of the states in terms of the complete set associated with the momentum and number operators. Such an expansion has the form

$$|\Psi\rangle=\sum_{n=0}^{\infty}\sum_{\mathbf{k}_1\cdots\mathbf{k}_n}V^{-n/2}\psi_n(k_1,\cdots,k_n)|k_1,\cdots,k_n\rangle, \quad (4)$$

in which $|k_1,\cdots,k_n\rangle$ denotes a state containing n particles of momenta $k_1,\cdots,k_n$, and V denotes a finite volume of enclosure. This expansion is always possible for states of a free field. The four-vectors k_i must satisfy the conditions

$$k_i^2=\mathbf{k}_i^2-k_{i0}^2=-m^2, \quad k_{i0}>0, \quad (5)$$

the first in virtue of the field equations, and the second because of the physical requirement that the particles have positive energy. In addition, the coefficient $\psi_n(k_1,\cdots,k_n)$ must be symmetric in all its arguments in order to satisfy the Pauli principle. The states $|k_1,\cdots,k_n\rangle$ appearing in the expansion (4) may be obtained by applying the creation operators $a^{\dagger}(k)$ to the vacuum. They satisfy the normalization condition

$$\langle k_1,\cdots,k_n|k_1',\cdots,k_m'\rangle=\delta_{nm}\sum_{P(i)}\delta(\mathbf{k}_1-\mathbf{k}_{i_1}')\cdots\delta(\mathbf{k}_n-\mathbf{k}_{i_n}), \quad (6)$$

the sum being taken over all permutations of the indices labelling the k'.

We retain the normalization (6) even in the case where two or more of the k are equal in order not to be forced to take these states into account separately in the calculations. This differs from the usual normalization by a factor $[\Pi(n_i!)]^{-1}$ if n_1, n_2, etc., k's are equal. In order to avoid singularities from these states because of products of δ functions with equal arguments appearing in (6), we have enclosed the system in a finite spatial volume V. Then the allowable values of k form a discrete manifold over which the summation in (4) is taken. The δ functions on the right of (6) are then to be interpreted as Kronecker symbols instead of Dirac δ functions. It may be shown that the quantities ψ_n are independent of V when we are dealing with localized states. In the final expressions for expectation values of products of fields, the volume may be taken infinite without these singularities reappearing, and we may then replace the summations by integrations:

$$\sum_{\mathbf{k}}=\int(d^3n/d^3k)d^3k=[V/(2\pi)^3]\int d^3k.$$

We have already introduced the creation operators $a^{\dagger}(k)$, in terms of which the field operator may be given as follows:

$$A(x)=\sum_{\mathbf{k}}\frac{1}{(2\omega V)^{\frac{1}{2}}}[a(k)e^{ikx}+a^{\dagger}(k)e^{-ikx}]$$

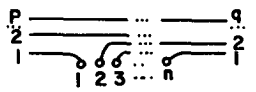
$$kx=\mathbf{k}\cdot\mathbf{x}-\omega x_0, \quad \omega=(\mathbf{k}^2+m^2)^{\frac{1}{2}}.$$

The normalization is chosen so that

$$[a(k),a^{\dagger}(k')]=\delta(\mathbf{k}-\mathbf{k}').$$

The quantities $\Psi(x_1,\cdots,x_n)=\langle\Psi|A(x_1)\cdots A(x_n)|\Psi\rangle-\langle 0|A(x_1)\cdots A(x_n)|0\rangle$ must vanish outside of a region $V_{\pm}(\mathcal{R})$ if (2) is to be satisfied. We wish to express this condition in terms of the $\psi_n(k_1,\cdots,k_n)$. This is greatly facilitated by the introduction of a graphical method of representating the matrix elements involved. From (4)

FIG. 1.



we obtain

$$|\Psi\rangle=\sum_{n=0}^{\infty}\sum_{k_1,\cdots,k_n}V^{-n/2}\psi_n(k_1,\cdots,k_n)a^\dagger(k_1)\cdots a^\dagger(k_n)|0\rangle$$

$$=\sum_{n=0}^{\infty}\sum_{k_1,\cdots,k_n}V^{-n/2}\psi_n(k_1,\cdots,k_n)(2^n\omega_1\cdots\omega_nV^{-n})^{\frac{1}{2}}$$

$$\times\int\cdots\int d^3x_1\cdots d^3x_n e^{ik_1x_1+\cdots+ik_nx_n}$$

$$\times A^{(-)}(x_1)\cdots A^{(-)}(x_n)|0\rangle.$$

$A^{(-)}(x)$ is the creation part of the field operator $A(x)$:

$$A^{(-)}(x)=\sum_{\mathbf{k}}\frac{1}{(2\omega V)^{\frac{1}{2}}}a^\dagger(k)e^{-ikx}.$$

We also list the following for future reference:

$$A^{(+)}(x)=A^{(-)\dagger}(x),\quad A(x)=A^{(+)}(x)+A^{(-)}(x),$$
$$[A^{(-)}(x),A^{(+)}(x')]=i\Delta^{(+)}(x-x'),$$
$$[A^{(+)}(x),A^{(+)}(x')]=[A^{(-)}(x),A^{(-)}(x')]=0.$$

On introducing the Fourier transform of $\psi_n(k_1,\cdots,k_n)$,

$$\psi_n(x_1,\cdots,x_n)=V^{-n}\sum_{k_1,\cdots,k_n}(2^n\omega_1\cdots\omega_n)^{\frac{1}{2}}$$

$$\times\psi_n(k_1,\cdots,k_n)e^{ik_1x_1+\cdots+ik_nx_n},\quad(7)$$

we have

$$|\Psi\rangle=\sum_{n=0}^{\infty}\int\cdots\int d^3x_1\cdots d^3x_n\psi_n(x_1,\cdots,x_n)$$

$$\times A^{(-)}(x_1)\cdots A^{(-)}(x_n)|0\rangle.\quad(8)$$

The expectation value of $A(x_1)\cdots A(x_n)$ becomes

$$\langle\Psi|A(x_1)\cdots A(x_n)|\Psi\rangle$$

$$=\sum_{p,q=0}^{\infty}\int\cdots\int d^3z_1\cdots d^3z_p d^3y_1\cdots d^3y_q$$

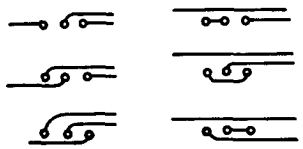
$$\times\psi_p^*(z_1,\cdots,z_p)\psi_q(y_1,\cdots,y_q)\langle0|A^{(+)}(z_p)\cdots$$

$$\times A^{(+)}(z_1)A(x_1)\cdots A(x_n)A^{(-)}(y_1)\cdots$$

$$\times A^{(-)}(y_q)|0\rangle.\quad(9)$$

The matrix element $\langle0|A^{(+)}(z_p)\cdots A^{(+)}(z_1)A(x_1)\cdots A(x_n)A^{(-)}(y_1)\cdots A^{(-)}(y_n)|0\rangle$ may be represented graphically as follows (Fig. 1). First, n points are drawn in a horizontal line, representing the n field operators $A(x_i)$. Then, lines representing particles are drawn, either connecting these points together in pairs, or extending to the left or right. Lines may also extend from left to right without touching one of the n points.

FIG. 2.



Exactly q lines must extend to the right and p lines to the left. The resulting figure describes the sequences of operations involved in computing the matrix element. The graphs, as well as the matrix elements are read from right to left. First the q creation operators $A^{(-)}(y_i)$ act on the vacuum to give a q-particle state. Then the field operators act in turn, either destroying a particle already present, or creating another particle. This results in a state of p particles, which are annihilated by the $A^{(+)}(z_i)$, giving the vacuum. Note that exactly one line emerges from each point of the graph. It may extend to the right or left representing destruction or creation, respectively, of the particle. If the figure cannot be drawn for the given values of n, p, and q, then the matrix element vanishes. It is usually possible to draw more than one graph for the given values of p, q, and n. For example, in the case $n-3$, $p-1$, $q-2$, 12 different diagrams are possible. Six of these are shown in Fig. 2. The other six are obtained by crossing the two lines extending to the right.[12]

The quantity

$$\langle0|A^{(+)}(z_p)\cdots A^{(+)}(z_1)A(x_1)\cdots$$
$$\times A(x_n)A^{(-)}(y_1)\cdots A^{(-)}(y_q)|0\rangle\quad(10)$$

is represented by the sum of all these possible graphs. For a given graph, the line connecting a point P_1 to a point P_2 to the right of P_1 contributes a factor $i\Delta^{(+)}(P_1-P_2)$. We will not bother here to give a proof of all these statements, but only exhibit a set of formulas upon which an inductive proof may be based. These are simple consequences of (7).

$$\langle0|A^{(+)}(z_p)\cdots A^{(+)}(z_1)A(x_1)\cdots A(x_n)A^{(-)}(y_1)\cdots$$
$$\times A^{(-)}(y_q)|0\rangle$$
$$=\sum_{r=1}^{n}i\Delta^{(+)}(z_1-x_r)\langle0|A^{(+)}(z_p)\cdots A^{(+)}(z_2)A(x_1)\cdots$$
$$\times A(x_{r-1})A(x_{r+1})\cdots A(x_r)A^{(-)}(y_1)\cdots A^{(-)}(y_q)|0\rangle$$
$$+\sum_{r=1}^{q}i\Delta^{(+)}(z_1-y_r)\langle0|A^{(+)}(z_p)\cdots A^{(+)}(z_2)A(x_1)\cdots$$
$$\times A(x_n)A^{(-)}(y_1)\cdots A^{(-)}(y_{r-1})A^{(-)}(y_{r+1})\cdots$$
$$\times A^{(-)}(y_q)|0\rangle.$$

$$\langle0|A(x_1)\cdots A(x_n)A(y_1)\cdots A(y_q)|0\rangle$$
$$=\sum_{r=1}^{n}i\Delta^{(+)}(x_r-y_1)\langle0|A(x_1)\cdots A(x_{r-1})A(x_{r+1})\cdots$$
$$\times A(x_n)A^{(-)}(y_2)\cdots A^{(-)}(y_q)|0\rangle.$$

[12] These graphs are similar to Feynman graphs with only one line terminating at each vertex.

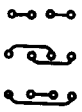


FIG. 3.

As an example of the graphical technique, we may consider the vacuum expectation value of a product of fields. This is the case $p=q=0$, so that no $A^{(+)}$ or $A^{(-)}$ operators appear in (10), and no lines extend to the right or left in the graphs. It is clear that the vacuum expectation value of an odd number of operators vanishes. We obtain

$$\langle 0|A(x_1)\cdots A(x_n)|0\rangle = \begin{cases} 0, & n \text{ odd} \\ \Sigma i^{n/2}\Delta^{(+)}(x_{i1}-x_{i2})\cdots\Delta^{(+)}(x_{in-1}-x_{in}), & n \text{ even,} \end{cases}$$

where the sum is taken over all possible ways to form $n/2$ pairs from the n points, preserving the original order of the x_i within each pair. In the case of 4-fields, the graphs are shown in Fig. 3.

When we are evaluating the quantity (7), many of the graphs will give the same contribution because of the symmetry of $\psi_n(x_1,\cdots,x_n)$. For example, in Fig. 2 only the graphs shown give an independent contribution. The six graphs obtained from them by permuting the terminal points of the lines at the right each give the same contribution to (9) as the corresponding unpermuted graphs. Another example is the normalization integral of the state (8): $\langle\Psi|\Psi\rangle=1$. The independent graphs are shown in Fig. 4. The contribution of each graph must be multiplied by $p!$, the number of possible graphs with crossed lines corresponding to the given graph. Figure 4 (a) is associated with the vacuum component of $|\Psi\rangle$. The resulting condition is

$$\sum_{p=0}^{\infty}\int\cdots\int dz_1\cdots dz_p dy_1\cdots dy_p p!\psi_p^*(z_1,\cdots,z_p) \times\psi_p(y_1,\cdots,y_p)i^p\Delta^{(+)}(z_1-y_1)\cdots\Delta^{(+)}(z_p-y_p)=1,$$

a further condition on the ψ_p.

We may now prove the following theorem:

No state of the form (4) or (8) can be localized if $\psi_n=0$ for all $n>N$, an arbitrary integer.

This means that a localized state must have an "infinite number of particles," i.e., there must be a nonzero probability of finding more than N particles in the state, however large N may be. The proof of this theorem depends upon the structure of expectation values of products of field operators, and on certain analyticity properties of these quantities considered as functions of the variables x_i. These analyticity properties follow from the spectral conditions (5).

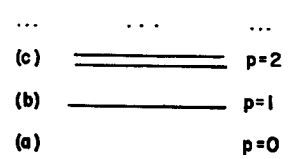


FIG. 4.

Before giving the general proof, we prove the simpler theorem that a state of the form

$$|\Psi\rangle=\psi_0|0\rangle+\textstyle\sum_k V^{-\frac{1}{2}}\psi_1(k)|k\rangle,$$

with ψ_1 not identically zero, cannot be localized. This is a special case of the general theorem with $N=1$. Consider the quantity $\langle\Psi|A(x_1)A(x_2)|\Psi\rangle$. There are four diagrams contributing to this quantity, as shown in Fig. 5. The sum of (a) and (b) is $i\Delta^{(+)}(x_1-x_2)$ times the normalization integral of the state $|\Psi\rangle$, which is equal to unity. Thus the quantity $\Psi(x_1,x_2)$ which must be localized is the sum of (c) and (d):

$$\begin{aligned}\Psi(x_1,x_2)&=i\int d^3z\psi_1^*(z)\Delta^{(+)}(z-x_1)\\ &\quad\cdot i\int d^3y\Delta^{(+)}(x_2-y)\psi_1(y)\\ &\quad+i\int d^3z\psi_1^*(z)\Delta^{(+)}(z-x_2)\\ &\quad\cdot i\int d^3y\Delta^{(+)}(x_1-y)\psi_1(y)\\ &=\Phi(x_1)\Phi^*(x_2)+\Phi(x_2)\Phi^*(x_1),\end{aligned}\tag{11}$$

with

$$\begin{aligned}\Phi(x)&=i\int d^3z\psi_1^*(z)\Delta^{(+)}(z-x)\\ &=-V^{-1}\sum_{\mathbf{k}}\psi_1(\mathbf{k})(2\omega)^{-\frac{1}{2}}e^{-ikx}\\ &=\int dk\Phi(k)e^{-ikx},\end{aligned}$$

$$\Phi(k)=-(2\omega)^{\frac{1}{2}}(2\pi)^{-3}\theta(k)\delta(k^2+m^2)\psi_1^*(k).\tag{12}$$

We have expressed $\Phi(x)$, a solution of the Klein-Gordon equation, as a four-dimensional Fourier transform. The factors $\delta(k^2+m^2)$ and $\theta(k)$ are explicit expressions of the spectral conditions (5). If we replace x by the complex four-vector $z=x+iy$, expression (12) defines a function of the four complex components of z which is analytic for those values of z for which the integral converges. It is easily seen that this is the case in the region $y\in V_+$ and that $\Phi(z)$ is analytic in that region.

As we allow y to approach zero from the forward light cone, we obtain as a boundary value the function (12). If we now hold the space part of z fixed and real, we obtain a function $\Phi(\mathbf{x},z_0)$ which is analytic in the upper half of the z_0 plane, and takes on the value $\Phi(x)$ on the

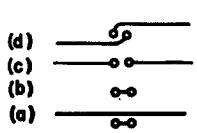


FIG. 5.

real axis. Then, by the uniqueness theorem for analytic functions, we conclude that if Φ vanishes in any time interval at the point x, then it vanishes for all time at that point.

By repeated application of this argument, we may prove that $\Phi(x)$ cannot vanish for all x outside $V_{\pm}(\mathfrak{R})$ unless it is identically zero, that is, unless $\psi_i(\mathbf{k})$ is identically zero. Let x_0 be a point outside the light cone for which $\Phi(x_0)\neq 0$. Then by (11),

$$\langle\Psi|A(x_0)A(x_0)|\Psi\rangle-\langle 0|A(x_0)A(x_0)|0\rangle=2|\Phi(x_0)|^2\neq 0,$$

violating the condition of localization (2), and completing the proof that no state of the form

$$|\Psi\rangle=\psi_0|0\rangle+V^{-\frac{1}{2}}\sum_{\mathbf{k}}\psi_1(\mathbf{k})|\mathbf{k}\rangle$$

can be localized.

We now go on to the proof that no state of the form

$$|\Psi\rangle=\sum_{n=0}^{N}V^{-n/2}\psi_n(k_1,\cdots,k_n)|k_1,\cdots,k_n\rangle$$

can be localized. Consider the expectation value of the product of $2N$ field operators:

$$\langle 0|A(x_1)\cdots A(x_{2N})|0\rangle. \tag{13}$$

The diagrams contributing to this quantity may be classified as follows:

(A) diagrams in which no two points x_i are joined by a line,

(B_p) diagrams in which exactly $2p$ points x_i are joined by a line, $p<N$, and

(C) diagrams in which each x_i is joined to another x_i by a line.[13]

For example, some of the graphs for the case $N-2$ are shown in Fig. 6. Figures 6 (a), (b), and (c), are of type (C); (d) (e), (f), and (g) of type (B_1); and (h) and (j) of type (A). The sum of the contributions of all diagrams of type (C) to the quantity (13) is $\langle 0|A(x_1)\cdots A(x_n)|0\rangle$ time $\langle\Psi|\Psi\rangle$, or simply $\langle 0|A(x_1)\cdots A(x_n)|0\rangle$ since $|\Psi\rangle$ is a normalized state. Therefore, the sum of all contributions from diagrams of types (A) and (B_p) must vanish when all the x_i lie outside $V_{\pm}(\mathfrak{R})$. It is clear that a diagram of type (B_1) gives $i\Delta^{(+)}(x_i-x_j)$ times the sum of all contributions of type (A) to the quantity

$$\langle\Psi|A(x_1)\cdots A(x_{i-1})A(x_{i+1})\cdots \times A(x_{j-1})A(x_{j+1})\cdots A(x_{2N})|\Psi\rangle.$$

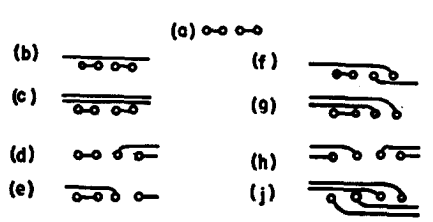
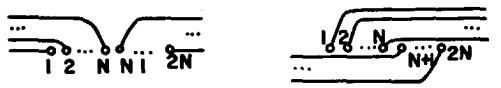


FIG. 6.

[13] We might also call this class B_N.

FIG. 7.



These contributions may be written

$$\langle\Psi|A(x_1)\cdots A(x_{i-1})A(x_{i+1})\cdots \times A(x_{j-1})A(x_{j+1})\cdots A(x_{2N})|\Psi\rangle - \langle 0|A(x_1)\cdots A(x_{i-1})A(x_{i+1})\cdots \times A(x_{j-1})A(x_{j+1})\cdots A(x_{2N})|0\rangle - \Psi^B(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_{j-1},x_{j+1},\cdots,x_{2N}),$$

i.e., the quantity itself minus contributions of types (C) and (B). The term

$$\Psi^B(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_{j-1},x_{j+1},\cdots,x_{2N})$$

may be in turn expressed as a sum of products of $\Delta^{(+)}$ functions with vacuum expectation values of products of a smaller number of fields. It is therefore clear that the (B)-type diagrams will contribute a quantity of the form

$$\sum_{i<j}i\Delta^{(+)}(x_i-x_j)[\langle\Psi|A(x_1)\cdots A(x_{i-1})A(x_{i+1})\cdots \times A(x_{j-1})A(x_{j+1})\cdots A(x_{2N})|\Psi\rangle - \langle 0|A(x_1)\cdots A(x_{i-1})A(x_{i+1})\cdots \times A(x_{j-1})A(x_{j+1})\cdots A(x_{2n})|0\rangle] + \sum_{i<j,k<l}i\Delta^{(+)}(x_i-x_j)\cdot i\Delta^{(+)}(x_k-x_l) \times[\langle\Psi|\cdots|\Psi\rangle-\langle 0|\cdots|0\rangle]+\cdots,$$

involving expectation values of products of less than $2N$ field operators. Now these quantities must all vanish when the x_i lie outside $V_{\pm}(\mathfrak{R})$ if $|\Psi\rangle$ is to be localized. For such x_i then, the only diagrams which need to be considered are those of type (A). Two diagrams of this type are shown in Fig. 7. The contribution of these diagrams is

$$\Phi(x_1,\cdots,x_N)\Phi^*(x_{N+1},\cdots,x_{2N}) + \Phi^*(x_1,\cdots,x_N)\Phi(x_{N+1},\cdots,x_{2N}).$$

$$\Phi(x_1,\cdots,x_N)=i^N\int\cdots\int d^3z_1\cdots d^3z_N\Delta^{(+)}(z_1-x_1)\cdots \times\Delta^{(+)}(z_N-x_N)\psi_N^*(z_1,\cdots,z_N) = (-V)^{-N}\sum_{k_1,\cdots,k_N}(2^N\omega_1\cdots\omega_N)^{-\frac{1}{2}} \times\psi_N^*(k_1,\cdots,k_N)e^{-ik_1x_1-\cdots-ik_Nx_N}. \tag{14}$$

The other type (A) diagrams are obtained by rearranging the points x_i in (14). Note that for each diagram, there occurs another diagram which gives the complex conjugate of the original diagram, as in Fig. 6. If $|\Psi\rangle$ is to be localized, then it is necessary that the

quantity

$$\sum \Phi(x_{i_1},\cdots,x_{i_N})\Phi^*(x_{i_{N+1}},\cdots,x_{i_{2N}}), \qquad (15)$$

be localized in the sense of (2). The summation is taken over all arrangements of the x_i. Now, let us set $x_1=x_{N+1}$ and $x_2=x_3=\cdots=x_N=x_{N+2}=\cdots=x_{2N}$, and let x_2 be a fixed point outside $V_\pm(\mathfrak{R})$. Then (15) becomes

$$4N^2[(2N-2)!](N!)^{-2}|\Phi(x_1,x_2,\cdots,x_2)|^2 + 4N(N-1)[(2N-2)!](N!)^{-2} \times \mathrm{Re}\Phi^*(x_1,x_1,x_2,\cdots,x_2)\Phi(x_2,\cdots,x_2).$$

The symmetry of Φ in its N arguments has been used to obtain this expression. The first term arises from those diagrams in which the points x_1 and x_{N+1} are connected to opposite sides of the diagram, and the second term from those in which they are connected to the same side. The factors $4N^2[(2N-2)!](N!)^{-2}$ and $4N(N-1)[(2N-2)!](N!)^{-2}$ represent the total number of diagrams of each kind. In the second term, complex conjugate diagrams have been combined to give $\mathrm{Re}\Phi^*(x_1,x_1,x_2,\cdots,x_2)\Phi(x_2,\cdots,x_2)$.

The function $\Phi(x_1,x_2,\cdots,x_2)$ may be regarded as a function of x_1 if we hold x_2 fixed. Its Fourier transform with respect to x_1 vanishes except on the forward mass hyperboloid, $k_1^2=-m^2$, and $k_{10}>0$, by virtue of (14) and (5). It is therefore an analytic function in the components of $x_1\to z_1=x_1+iy_1$ in the region $y_1\in V_+$, the forward light cone. Thus, as before, there must be a point $x_1^{(0)}$ outside $V_\pm(\mathfrak{R})$ such that

$$\Phi(x_1^{(0)},x_2,x_2,\cdots,x_2)\neq 0.$$

There are two cases to consider:

(1) $\Phi(x_2,\cdots,x_2)=0$. Then (14) becomes

$$4N^2[(2N-2)!](N!)^{-2}|\Phi(x_1^{(0)},x_2,\cdots,x_2)|^2\neq 0,$$

and the state cannot satisfy the localization criterion; and

(2) $\Phi(x_2,\cdots,x_2)\neq 0$, where we have

$$\langle\Psi|A(x_2)\cdots A(x_2)|\Psi\rangle-\langle 0|A(x_2)\cdots A(x_2)|0\rangle = 2(2N)!(N!)^{-2}|\Phi(x_2,\cdots,x_2)|^2\neq 0.$$

In each case, we have been able to construct points x_i all outside of $V_\pm(\mathfrak{R})$ where $\langle\Psi|A(x_1)\cdots A(x_{2N})|\Psi\rangle \neq\langle 0|A(x_1)\cdots A(x_{2N})|0\rangle$. We have therefore proved that a state of the form

$$|\Psi\rangle=\sum_{n=0}^{N} V^{-n/2}\psi_n(k_1,\cdots,k_n)|k_1,\cdots,k_n\rangle$$

cannot satisfy criterion (2).

The theorem just proved allows us to confine our attention to states having an infinite number of particles in the sense explained above. If we attempt to discuss these states by the diagramatic method of this chapter, we find that an infinite number of graphs contribute to the expectation value of any given number of fields. For example, the sequence of graphs in Fig. 8 contributes to

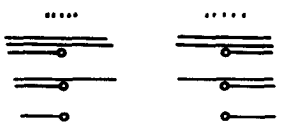


FIG. 8.

$\langle\Psi|A(x)|\Psi\rangle$. This quantity is, therefore, given by an infinite series, of which a typical term is

$$i\int\cdots\int d^3z_1\cdots d^3z_p d^3y_1\cdots d^3y_{p-1}\psi_p^*(z_1,\cdots,z_p) \times\psi_{p-1}(y_1,\cdots,y_{p-1})\Delta^{(+)}(z_p-x).$$

This series involves all the coefficients $\psi_p(z_1,\cdots,z_p)$ for all values of p. The condition that $\langle\Psi|A(x)|\Psi\rangle$ be localized is thus a very complicated condition on the ψ_p, collectively. There are similar conditions for $\langle\Psi|A(x_1)A(x_2)|\Psi\rangle$, etc. The only simplification that appears is that diagrams with two or more points joined together may be omitted. This is similar to the situation above with (B)-type diagrams. Since further analysis based on this method becomes very complicated, we pass to the approach described in the next two sections.

5. STATES OF THE FORM $\exp(iR)|0\rangle$, R LINEAR

Consider the state vector

$$\exp(iR)|0\rangle, \qquad (16)$$

obtained by applying an exponential operator to the vacuum.[14] The fact that such a state always satisfies our condition of localization when R is a Hermitian operator depending only on the field operator $A(x)$ with $x\in\mathfrak{R}$ follows from the property of local commutativity of the fields, and is therefore true in the case of interacting fields having this property as well as for free fields. Our further remarks in this section, however, apply to the free-field case. The precise form of R will be discussed later in this section, but for the present we may imagine R to be of the same form as (1), but with the integrations extending only over $\mathfrak{R}$. To insure the Hermiticity of R, the functions q_i of (1) must be real. Then since any point x_j outside $V_\pm(\mathfrak{R})$ is space-like with respect to the region R over which the integrals defining R are taken, local commutativity insures that $[R,A(x_j)]=0$. It is then easily shown that the state $\exp(iR)|0\rangle$ is localized, for

$$\langle 0|\exp(-iR)\prod_{j=1}^{n}A(x_j)\exp(iR)|0\rangle = \langle 0|\prod_{j=1}^{n}A(x_j)\exp(-iR)\exp(iR)|0\rangle = \langle 0|\prod_{j=1}^{n}A(x_j)|0\rangle$$

[14] The author is greatly indebted to R. Glaser (private communication to J. S. Toll) for calling these states to his attention.

because $\exp(iR)$ is a unitary operator. The fact that $\exp(iR)|0\rangle$ is a localized state considerably enlarges the manifold of such states at our disposal. One is led to conjecture that all localized states may be expressed in this form. Unfortunately, we have been able to neither prove nor to disprove this assertion. We will, therefore, confine ourselves here to as thorough an investigation as practicable of these states, indicating at the end some arguments supporting the conjecture.

First, it can be shown that the Poisson distribution state studied in Sec. 3 can be put into the form $\exp(iR)|0\rangle$, with

$$R=\int j(x)A(x)dx$$
$$=\sum_k(2\omega)^{-\frac{1}{2}}(j^*(\mathbf{k},\omega)a(k)+j(\mathbf{k},\omega)a^\dagger(k)). \quad (17)$$

If we expand $|\Psi\rangle=\exp(iR)|0\rangle$ in momentum eigenstates, we find the expansion coefficient

$$(1/n!)\langle k_1,\cdots,k_n|\Psi\rangle$$
$$=(i^n/n!)(2^n\omega_1\cdots\omega_n)^{-\frac{1}{2}}j(\mathbf{k}_1,\omega_1)\cdots j(\mathbf{k}_n,\omega_n)$$
$$\times\exp(-\tfrac{1}{4}\textstyle\sum_k\omega^{-1}|j(\mathbf{k},\omega)|^2), \quad (18)$$

which is the same as that for the Poisson distribution state.

The property of R essential to give localization is that it involves only $A(x)$ with x restricted to $\mathfrak{R}$. Equation (17) is therefore not the only form possible for R. Another very natural form we might choose is the following:

$$R=\int d^3x g(x)\overset{\bar\partial}{\underset{\partial x_0}{\longleftrightarrow}}A(x)$$
$$\equiv\int d^3x\left(g(x)\overset{\partial}{\underset{\partial x_0}{\longleftrightarrow}}A(x)-A(x)\overset{\partial}{\underset{\partial x_0}{\longleftrightarrow}}g(x)\right), \quad (19)$$

where $g(x)$ is a localized solution of the Klein-Gordon equation. The integral is taken over a space-like surface intersecting the region $\mathfrak{R}$. This surface usually will be chosen as a time plane $x_0=$constant, although this is, of course, not necessary. If we define the Fourier transform of $g(x)$ by

$$g(x)=\textstyle\sum_{\mathbf{k}}(2\omega V)^{-\frac{1}{2}}[g(\mathbf{k})e^{ikx}+g^*(\mathbf{k})e^{-ikx}],$$

we obtain

$$R=-i\textstyle\sum_{\mathbf{k}}[a(k)g^*(\mathbf{k})-a^\dagger(k)g(\mathbf{k})].$$

This is similar to the expression (17), and the expansion of $\exp(iR)|0\rangle$ into number and momentum eigenstates may be performed in the same manner. We find that the expansion coefficients agree with those of the Poisson distribution state (18), if we set

$$g(x)=\int dx'\Delta(x-x')j(x'),$$

which makes explicitly clear the fact that $g(x)$ is a localized solution of the Klein-Gordon equation, since $j(x)$ vanishes outside $\mathfrak{R}$.

We have investigated two different forms for the operator R appearing in (16) in the case where R is linear in $A(x)$. Both of these forms give rise to states of the Poisson distribution type discussed in Sec. 3. Each form may be generalized to involve products of more than one $A(x)$. The generalization of (17) is given by (1). The general form of (19) is

$$R=r_0+\int_{x_0=t}d^3x r_1(x)\overset{\bar\partial}{\underset{\partial x_0}{\longleftrightarrow}}A(x)$$
$$+\int\int_{x_{10}=x_{20}=t}d^3x_1d^3x_2r_2(x_1,x_2)\overset{\bar\partial}{\underset{\partial x_{10}}{\longleftrightarrow}}\overset{\bar\partial}{\underset{\partial x_{20}}{\longleftrightarrow}}$$
$$\times A(x_1)A(x_2)+\cdots. \quad (20)$$

The $r_i(x_1,\cdots,x_n)$ as well as the $q_i(x_1,\cdots,x_n)$ of (1) can always be chosen as symmetric functions of their arguments.

Since we are looking for a general representation of localized states, we wish to decide which of these forms is the more suitable. This is accomplished by requiring that the representation be unique. In the case of the Poisson distribution state, we note that the expansion coefficients ψ_n determine the quantity $j(\mathbf{k},\omega)$ uniquely. This corresponds to the quantity $j(\mathbf{k},\omega)$ if we employ the four-dimensional integration (17), and to $g(\mathbf{k})$ if we use the three-dimensional integral (19) for R. Thus, $j(\mathbf{k},\omega)$ and $g(\mathbf{k})$ are fixed by the state. However, $j(\mathbf{k},\omega)$ is not sufficient to determine the function $j(x)$ appearing in (17), since it fixes only these Fourier components of $j(x)$ which satisfy the restriction $\mathbf{k}^2-k_0^2=-m^2$. The other Fourier components, off the mass shell, are entirely arbitrary inasmuch as they may be varied in any way without affecting the state (16). In fact $j(x)$ need not even vanish outside $\mathfrak{R}$ in order that (16) be localized. It is only necessary that a localized function $j(x)$ exists having the Fourier components $j(\mathbf{k},\omega)$ on the mass shell. On the other hand, the function $g(x)$ appearing in the expression (19) for R is a solution of the Klein-Gordon equation, and is therefore completely determined by its Fourier components on the mass shell. It is thus fixed uniquely once the expansion coefficients (18) of the state are given. For this reason, we will use the form (19) and its generalization (20) for the operator R appearing in (16).

The nonuniqueness of the four-dimensional form can be understood from a different point of view. Equation (17) may be regarded as a linear combination of the field operators at points in the region $\mathfrak{R}$. A linear combination of quantities is unique only if the quantities are linearly independent. The field operators in our extended region of space-time, however, are not linearly independent, since an operator at a given time x_0 can

be expressed through the equation of motion in terms of operators at another time x_0'. Since only points within the light cone contribute, we have a linear relation involving only field operators taken at points of $\mathfrak{R}$. In the other case, the integral (19) involves only field operators and their time derivatives at a given time. These are dynamically independent quantities, and therefore no linear relation between them exists. This explains why we obtain a unique representation with (19) but not with (18).

6. STATES OF THE FORM $\exp(iR)|0\rangle$, GENERAL

Having discussed states of the form (16) with R linear in $A(x)$, we now pass to a consideration of more complicated forms of R. The general form of R is taken to be (20), with $r_n(x_1,\cdots,x_n)$ a localized solution of the Klein-Gordon equation symmetric in all of its arguments. The expectation value

$$\langle\Psi|A(x_1)\cdots A(x_n)|\Psi\rangle = \langle 0|\exp(-iR)A(x_1)\cdots A(x_n)\exp(iR)|0\rangle$$

may be expressed in the form

$$\langle 0|\Gamma(x_1)\cdots\Gamma(x_n)|0\rangle, \tag{21}$$

with

$$\Gamma(x)=\exp(-iR)A(x)\exp(iR)$$
$$=A(x)+\exp(-iR)[A(x),\exp(iR)]$$
$$=A(x)+\sum_{n=1}^{\infty}\frac{i^n}{n!}[[\cdots[A(x),R],R]\cdots], \tag{22}$$

where the nth term contains the n-fold commutator of $A(x)$ with R. Thus, the expectation value of a product of field operators $A(x)$ in a state $|\Psi\rangle$ reduces to the vacuum expectation value of the product of Γ operators (21). $\Gamma(x)$ is given in terms of $A(x)$ by the infinite series (22). The exact form of $\Gamma(x)$ will depend upon the functions $r_n(x_1,\cdots,x_n)$ which define R. Suppose that $r_n=0$ for $n>N$, i.e., R is a sum of terms in each of which at most N operators $A(x)$ appear. We will express this by writing $R=0(A^N)$. It is easily seen that the commutator of an operator of order p with an operator of order q is an operator of order $p+q-2$. It is clear then that the nth term of the series (22) will be of order

$$(1+N-2)+(N-2)+\cdots+(N-2)=1+n(N-2).$$

Thus, if $N>2$, the operator $\Gamma(x)$ is of infinite order in $A(x)$. For $N=1$, the $n=0$ term, which is just $A(x)$, is of order 1, and the term $n=1$ is of order zero. The remaining terms vanish. For $N=2$, each term of the series is of order 1. These two cases are therefore particularly simple. The first, $N=1$, is the Poisson distribution state which we have already discussed at length. For this state, $\Gamma(x)$ is given by

$$\Gamma(x)=A(x)+r_1(x), \tag{23}$$

where the c number $r_1(x)$ is to be identified with the function $g(x)$ of Eq. (19). For the case $N=2$, we take R to be

$$R=\int\int d^3x_1 d^3x_2 r(x_1,x_2)\frac{\bar\partial}{\partial x_{10}}\frac{\bar\partial}{\partial x_{20}}A(x_1)A(x_2). \tag{24}$$

By making use of the fact that $r(x_1,x_2)$ is a symmetrical solution of the Klein-Gordon equation, and therefore satisfies the following equation:

$$r(x_1,x_2)=-\int d^3x_1'\Delta(x_1-x_1')\frac{\bar\partial}{\partial x_{10}'}r(x_1',x_2),$$

we obtain

$$[A(x),R]=2i\int\int d^3x_1 d^3x_2 r(x_1,x_2)\frac{\bar\partial}{\partial x_{10}}\frac{\bar\partial}{\partial x_{20}}$$
$$\times A(x_1)\Delta(x-x_2)$$
$$=2i\int d^3x_1 r(x,x_1)\frac{\bar\partial}{\partial x_{10}}A(x_1),$$

where we have also used the commutation relation $[A(x),A(x')]=i\Delta(x-x')$. The n-fold commutator $[[\cdots[A(x),R],R]\cdots]$ may be computed similarly:

$$[[\cdots[A(x),R],R],\cdots]$$
$$=(2i)^n\int\cdots\int d^3x_1\cdots d^3x_n r(x,x_1)\frac{\bar\partial}{\partial x_{10}}$$
$$\times r(x_1,x_2)\frac{\bar\partial}{\partial x_{20}}\cdots r(x_{n-1},x_n)\frac{\bar\partial}{\partial x_{n0}}A(x_n).$$

We therefore find that

$$\Gamma(x)=A(x)+\int d^3x'\gamma(x,x')\frac{\bar\partial}{\partial x_0'}A(x'), \tag{25}$$

$$\gamma(x,x')=\sum_{n=1}^{\infty}\frac{(-2)^n}{n!}\int\cdots\int d^3x_1\cdots d^3x_n r(x,x_1)\frac{\bar\partial}{\partial x_{10}}$$
$$\times r(x_1,x_2)\cdots\frac{\bar\partial}{\partial x_{n-1,0}}r(x_{n-1},x'), \tag{26}$$

by substituting these expressions into (22). For $N>2$, it is also possible to give the form of $\Gamma(x)$, but we will not do this here.

There is a strong similarity of the relation between $A(x)$ and $\Gamma(x)$ to that between $A_{\text{in}}(x)$ and $A_{\text{out}}(x)$ in a relativistic field theory with interaction. The correspondence is exact if we allow interaction with an external source. This is easily seen from the example given in Sec. 3. The relation

$$\Gamma(x)=\exp(-iR)A(x)\exp(iR)$$

is analogous to the relation

$$A_{\text{out}}(x)=S^{\dagger}A_{\text{in}}(x)S.$$

A relation of the form (22) also exists between in- and out-going fields if we introduce the phase matrix η by $S=\exp(-i\eta)$. There is one point of dissimilarity between $\Gamma(x)$ and the outgoing field of a system such as that described by the quantum electrodynamics where no external current acts. There, the outgoing field is invariant under displacements, whereas in our case the localized nature of the operator R destroys the displacement invariance of $\Gamma(x)$ considered as a free field. Thus the vacuum expectation value of a product of Γ operator does not have the simple properties of the Wightman functions.[2] They are not, for example, functions of the coordinate differences only, and do not have the same simple analyticity properties. It is still possible, however, to draw certain conclusions by means of analyticity properties, as we shall do below in the case where Γ has the form (24).

Before proceeding further, we will note that, since $\Gamma(x)$ is obtained from $A(x)$ by the unitary transformation $\exp(-iR)A(x)\exp(iR)$, it must satisfy the same commutation relation: $[\Gamma(x),\Gamma(y)]=i\Delta(x-y)$. In the case where R is linear in $A(x)$, $N=1$ above, this can also be seen directly from the expression (23) for $\Gamma(x)$, since $r_1(x)$ is a c-number function and thus commutes with $A(x)$. For the case $N=2$, where R has the form (24) and $\Gamma(x)$ the form (25), we find

$$[\Gamma(x),\Gamma(y)]=i\Delta(x-y)-i\gamma(x,y)+i\gamma(y,x)$$
$$+i\int d^3z\gamma(x,z)(\bar{\partial}/\partial z_0)\gamma(z,y).$$

It is therefore necessary that $\gamma(x,y)$ satisfy the relation

$$\gamma(x,y)-\gamma(y,x)-\int d^3z\gamma(x,z)(\bar{\partial}/\partial z_0)\gamma(z,y)=0.$$

This is not an independent relation, but follows identically from (26).

It is clear that the condition of localization places certain restrictions upon the region of space where $\gamma(x,y)$ may differ from zero. It is our purpose to find out what these restrictions are and to find the corresponding conditions for the function $r(x,y)$ in terms of which $\gamma(x,y)$ is defined. Let us assume that on the reference time plane over which the integral (25) is taken, $\gamma(x,y)$ and its derivatives with respect to x_0 and y_0 and its mixed second derivative with respect to x_0 and y_0 all vanish when either x or y lies outside a large but bounded region $\mathcal{S}$ of 3-space. These four quantities, $\gamma(x,y)$ and its 3-time derivatives, when given for all space at a given time, completely determine $\gamma(x,y)$ for all space-time, since they provide the initial values from which $\gamma(x,y)$ may be computed from the Klein-Gordon equation. The region $\mathcal{S}$ is arbitrary, and may be taken much larger than $\mathcal{R}$, the region of localization.

From (25) we obtain

$$\langle 0|\Gamma(x)\Gamma(y)|0\rangle$$
$$=\langle 0|A(x)A(y)|0\rangle$$
$$+i\int d^3y'\gamma(y,y')\frac{\bar{\partial}}{\partial y_0'}\Delta^{(+)}(x-y')$$
$$+i\int d^3x'\gamma(x,x')\frac{\bar{\partial}}{\partial x_0'}\Delta^{(+)}(x'-y)$$
$$+i\int\int d^3x'd^3y'\gamma(x,x')\gamma(x,y')\frac{\bar{\partial}}{\partial x_0'}\frac{\bar{\partial}}{\partial y_0'}$$
$$\times\Delta^{(+)}(x'-y'), \quad (27)$$

where we have used the relation

$$\langle 0|A(x)A(y)|0\rangle=i\Delta^{(+)}(x-y).$$

The sum of the last three terms on the right side of (27) must vanish when both x and y are outside $V_{\pm}(\mathcal{R})$. We denote these terms by a, b, and c, respectively. The Fourier transform of the singular function $\Delta^{(+)}$ appearing in (27) vanishes except on the forward mass hyperboloid $k^2=-m^2$, and $k_0>0$. Therefore, as in Sec. 4, it is an analytic function of the components of the vector x whenever the imaginary part of x lies in the forward light cone. We conclude that the quantity a is analytic in x, and the quantity b is analytic in y. The argument of Sec. 4 can therefore be applied to show that if these functions vanish over any extended region of space-time in their analytic arguments, they must vanish everywhere.

The localization condition requires that the quantity $a+b+c$ together with its three time derivatives, vanish outside the union of the regions $\mathcal{R}\times\mathcal{U}$ and $\mathcal{U}\times\mathcal{R}$. $\mathcal{R}\times\mathcal{U}$ is the cartesian product of the region $\mathcal{R}$ and the whole space,[15] denoted here by $\mathcal{U}$. In consequence of our assumption that $\gamma(x,y)$ vanishes outside of the region $\mathcal{S}\times\mathcal{S}$, it is evident from (27) that $a=0$ outside of $\mathcal{U}\times\mathcal{S}$, that $b=0$ outside of $\mathcal{S}\times\mathcal{U}$, and that $c=0$ outside of $\mathcal{S}\times\mathcal{S}$. Now, consider the region $(\mathcal{U}-\mathcal{S})\times(\mathcal{S}-\mathcal{R})$, shown schematically as the shaded region in Fig. 9.

In this region $a+b+c$ must equal zero. But we have shown that b and c vanish there. Thus, a must also vanish. We can conclude from this by means of the analyticity of a in the variable x that a vanishes for all x when y is in $(\mathcal{S}-\mathcal{R})$. We may, therefore, subtract the region $\mathcal{U}\times(\mathcal{S}-\mathcal{R})$ from the region we have already found, and obtain the result that $a=0$ outside of $\mathcal{U}\times\mathcal{R}$. By an entirely similar argument, $b=0$ outside of $\mathcal{R}\times\mathcal{U}$. The localization condition may now be applied to show that $c=0$ outside the union of $\mathcal{S}\times\mathcal{R}$ and $\mathcal{R}\times\mathcal{S}$. The

[15] In this section, we use $\mathcal{R}$ to denote the three-dimensional region over which the integrals in (5.7) extend, and not the four-dimensional space-time region used previously.

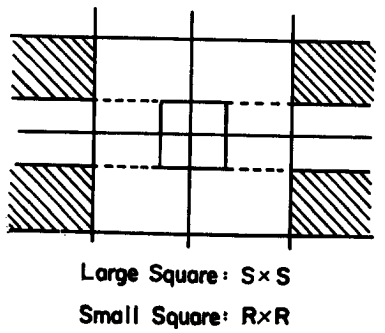


FIG. 9.

same results apply to the three time derivatives of each of these quantities.

It is possible to find a restriction on the region where $\gamma(x,y)$ may differ from zero from the above. From (27) we obtain

$$\mathrm{Im}a=\tfrac{1}{2}\int d^3y'\gamma(y,y')\frac{\bar{\partial}}{\partial y_0'}\Delta(x-y'). \tag{28}$$

Here, we have used $\Delta^{(+)}(x)=\frac{1}{2}(\Delta(x)-i\Delta^{(1)}(x))$. It follows immediately from this and the properties of the Δ function that $\mathrm{Im}a=-\frac{1}{2}\gamma(y,x)$. The corresponding relation between the three time derivatives may be derived in the same way after differentiating both sides of (28). Taking into account that $\mathrm{Im}a$ and its derivatives vanish outside $\mathcal{U}\times\mathcal{R}$, we find

$$\gamma(x,y)=\frac{\partial\gamma(x,y)}{\partial x_0}=\frac{\partial\gamma(x,y)}{\partial y_0}=\frac{\partial^2\gamma(x,y)}{\partial x_0\partial y_0}=0,$$
$$\text{if } (x,y)\notin\mathcal{R}\times\mathcal{U}. \tag{29}$$

Equation (29) implies that $\gamma(x,y)$ for arbitrary y is localized in x in the region $V_\pm(\mathcal{R})$. Our argument shows that (29) is a necessary condition that the expectation value of the product of two fields satisfy the localization criterion. That it is also sufficient is obvious from Eq. (27). Note that we may add a linear term of the form (19) to the expression (24) for R without interfering with the above argument. In fact, the localization condition for $\langle 0|\Gamma(x)|0\rangle$ requires that $g(x)$ and $\partial g(x)/\partial x_0$ be confined to the region $\mathcal{R}$, so that it does not enter into the above considerations.

It has now been shown that condition (29) is equivalent to localization of the state $\exp(iR)|0\rangle$, where R is given by (24). Is it possible to conclude from this that $r(x,y)$ and its time derivatives are zero outside of $\mathcal{R}\times\mathcal{R}$? It is obvious from (26) that this property of $r(x,y)$ is sufficient to guarantee that $\gamma(x,y)$ satisfies (29). Furthermore, if we assume only that $r(x,y)$ is confined, for example, to the cross-shaped region $\mathcal{U}\times\mathcal{R}+\mathcal{R}\times\mathcal{U}$, (29) can only be satisfied if a large-scale cancellation occurs among the terms of the series (26) for $\gamma(x,y)$ at points x lying outside of $\mathcal{R}$. It seems unlikely that this is possible, although we have not been able to prove its impossibility. It is shown in the Appendix that Eq. (26) may be inverted and $r(x,y)$ expressed in terms of $\gamma(x,y)$:

$$r(x,y)=\sum_{n=1}^{\infty}\frac{(-1)^n}{2n}\int\cdots\int d^3x_1\cdots d^3x_{n-1}$$
$$\times\gamma(x,x_1)\frac{\bar{\partial}}{\partial x_{10}}\cdots\frac{\bar{\partial}}{\partial x_{n-1,0}}\gamma(x_{n-1},y). \tag{30}$$

In order to derive (30), certain very restrictive and ad hoc assumptions must be made about the function $r(x,y)$. We cannot, therefore, claim general validity for this relation. However, if these assumptions are satisfied, it is clear from (30) that $r(x,y)$ with its time derivatives must be confined to $\mathcal{R}\times\mathcal{R}$. For, in consequence of conditions (29), each term of the series (30) is equal to zero when x lies outside $\mathcal{R}$, and so also their sum. Then, from the symmetry of $r(x,y)$ in x and y, we conclude that it also vanishes when y is not in $\mathcal{R}$.

We will now give a summary of the results of the last two sections. Our aim in considering states of the form (16) was to provide a general representation of states satisfying the definition of localization (2). Two different forms of the Hermitian operator R were considered, and the first rejected because the corresponding representation proved to be nonunique. The second form of R is determined by a sequence of functions $r_n(x_1,\cdots,x_n)$, each of which is a solution of the Klein-Gordon equation independently in each argument. We have shown that when R is linear in the field, the function $r_1(x)$ must be localized in the region $V_\pm(\mathcal{R})$. When R is quadratic in the fields, we proved that $r_2(x,y)$ must be zero outside of $V_\pm(\mathcal{R})\times V_\pm(\mathcal{R})$, subject to certain assumptions on the nature of $r_2(x,y)$.

In order to give a complete proof of the generality of the representation (16), we would have to show (a) that any state can be written in the form $\exp(iR)|0\rangle$ if $r_n(x_1,\cdots,x_n)$ is not restricted to be confined to any particular region of space, (b) that the localization condition then requires that $r_n(x_1,\cdots,x_n)$ vanish outside $V_\pm(\mathcal{R})\times\cdots\times V_\pm(\mathcal{R})$, and (c) that the representation is unique. The results of this section and the last show that (b) and (c) are true in some simple cases. Although it has not been possible here to give a complete general proof, it is encouraging to find that the conjecture that (16) is a general representation of localized states is substantiated in those cases in which it could be verified.

7. DISCUSSION AND CONCLUSIONS

In this paper, we have introduced a definition of localization based on taking the field variables as the primary measurable quantities. This enables us to avoid some of the difficulties which occur when one formulates such a definition in terms of particle observables. Our definition is completely Lorentz invariant, since it is formulated in terms of Lorentz-invariant quantities and refers to the situation of the field over all space-time. Two states localized in different

regions of space are not expected to be orthogonal because the two regions of localization $V_{\pm}(\mathcal{R}_1)$ and $V_{\pm}(\mathcal{R}_2)$ always overlap, however far $\mathcal{R}_1$ may be removed from $\mathcal{R}_2$. It may be verified that two states of the Poisson distribution type discussed in Chap. III are not orthogonal. Our definition thus seems satisfactory and appropriate to the current formulations of quantum field theory.

We have given results concerning the nature of the states satisfying the localization condition. It was seen in Sec. 5 that there is a large class of such states expressible in the simple form $\exp(iR)|0\rangle$. We have given arguments indicating that *any* localized state may be represented in this form, and proposed a general method of proof. While we have concentrated our attention in a large part of this thesis to the free-field case, this representation offers promise of being valid and general for the case of interacting fields as well.

Our motivation in taking up this problem was the hope that the results might be useful as the basis for a general approach to scattering problems in field theory. It does indeed appear that the states satisfying our definition of localization do provide an idealized description of the initial and final states of scattering processes which is nearer to the physical reality of production and detection than the customary idealization of single particle states. It is possible that by applying the methods of this paper, some of the difficulties associated with the asymptotic condition might be clarified. It is even possible that a description of scattering experiments might be given without appealing to the asymptotic condition. Such a description would necessarily be more complicated in some respects. For example, the theory of Sec. 4 shows that localized states may not contain a finite number of particles. This requires that any description of scattering in terms of localized states involve an infinite number of S-matrix elements between states of definite particle number.

However, it is easily seen from our example in Sec. 3 that if the current $j(x)$ is weak, we obtain a localized state in which the vacuum amplitude dominates, and the one-particle amplitude is much larger than the remaining amplitudes. It may, therefore, be possible to isolate the one-particle scattering terms from the others by a limiting process in which the strength of the current approaches zero.

It is not possible at present to evaluate completely the merits and defects of such a procedure, but it is hoped that the results of this paper will provide the foundation of a useful alternative to the conventional description of the scattering process in quantum field theory.

ACKNOWLEDGMENTS

The author is thankful to Professor Gunnar Källén for constant help and encouragement during the first part of the work of this paper. He acknowledges gratefully the continued help and interest of his advisor Professor John S. Toll, who suggested the problem investigated.

APPENDIX: INVERSION OF EQUATION (26)

In order to invert Eq. (26), we introduce a complete orthonormal set $\{\phi_i(x)\}$ of solutions of the Klein-Gordon equation:

$$\int d^3x\phi_i^*(x)\overleftrightarrow{\frac{\partial}{\partial x_0}}\phi_j(x)=\sigma_i\delta_{ij},$$
$$\sum_i \sigma_i\phi_i(x)\phi_i^*(y)=i\Delta(x-y). \tag{A1}$$

$\sigma_i=\pm 1$ according to whether $\phi_i(x)$ belongs to positive or negative energy. We define the quantity r_{ij} by

$$r_{ij}=\int\int d^3x d^3y\phi_i^*(x)\overleftrightarrow{\frac{\partial}{\partial x_0}}r(x,y)\overleftrightarrow{\frac{\partial}{\partial y_0}}\phi_j(y),$$
$$r(x,y)=-\sum_{i,j}\sigma_i\sigma_j r_{ij}\phi_i(x)\phi_j^*(y). \tag{A2}$$

The second equation (A2) follows from the first upon making use of (A1). The reality and symmetry of $r(x,y)$ require that $r_{ij}=r_{ji}^*$. γ_{ij} is defined in a similar way in terms of $\gamma(x,y)$, but $\gamma_{ij}\neq\gamma_{ji}^*$ because $\gamma(x,y)$ is not symmetric.

We now choose the set $\{\phi_i(x)\}$ in such a way that r_{ij} becomes a diagonal matrix, and set $r_{ij}=r_i\delta_{ij}$. After introduction of the newly defined quantities, (26) is transformed into

$$i\sigma_i\gamma_{ij}=(e^{-2i\sigma_i r_i}-1)\delta_{ij}.$$

Thus, γ_{ij} is also diagonal, with eigenvalues

$$\gamma_i=(1/i\sigma_i)(e^{-2i\sigma_i r_i}-1).$$

This equation may be solved for r_i, yielding

$$-2i\sigma_i r_i=\log(1+i\sigma_i\gamma_i). \tag{A3}$$

The quantities γ_i are complex numbers lying on a circle of unit radius and center at the point $+1$. If r_i is such that γ_i lies within the unit circle centered at the origin, then (A3) may be expanded into a convergent power series in γ_i:

$$2i\sigma_i r_i=\sum_{n=1}^{\infty}\frac{(-1)^n}{n}(i\sigma_i\gamma_i)^n.$$

If this expansion is possible for all r_i, then we may reintroduce $r(x,y)$ and $\gamma(x,y)$ by means of (A2) and obtain Eq. (30):

$$r(x,y)=\sum_{n=1}^{\infty}\frac{(-1)^n}{2n}\int\cdots\int d^3x_1\cdots d^3x_{n-1}$$
$$\times\gamma(x,x_1)\overleftrightarrow{\frac{\partial}{\partial x_{10}}}\cdots\overleftrightarrow{\frac{\partial}{\partial x_{n-1,0}}}\gamma(x_{n-1},y).$$

On the Interpretation of the Einstein-Schrödinger Unified Field Theory

D. W. SCIAMA
Department of Astronomy, Cornell University, Ithaca, New York
(Received December 22, 1960)

Vierbeins are used to analyze the Einstein-Schrödinger unified field theory, in both its real nonsymmetric and complex Hermitian versions. It appears that the skew-symmetric part of the real affine connection is related to spin rather than to electromagnetism. In its complex form the theory does appear to describe the electromagnetic field, since the vierbein analysis shows that gauge transformations arise in a natural way (the spin is then related to the skew-Hermitian part of the affine connection). The resulting identification of the electromagnetic field tensor implies that Einstein's choice of Lagrangian is physically unsatisfactory and an alternative Lagrangian is proposed. Although the electromagnetic field is thus geometrized, neither Lagrangian represents a unification of electromagnetism and gravitation.

1. INTRODUCTION

CURRENT interest in unified fields is centered on the Einstein-Schrödinger theory.[1] This theory exists in two main versions. The first, which is due to Einstein,[2] is based on a metric tensor and an affine connection which are complex and Hermitian. Later, Einstein,[3] following Schrödinger,[4] adopted the second version, in which these quantities are assumed to be real and nonsymmetric. The formalisms of the two versions do not differ greatly, but we shall see that they lend themselves naturally to different interpretations.

A considerable amount of effort has gone into studying the physical significance of these theories. Despite this work, their interpretation remains obscure. The main aim of this paper is to point out that if the theories are rewritten in terms of vierbeins,[5] a rather natural interpretation suggests itself. The main reason for this is that the vierbeins are closely related to a group (the infinitesimal holonomy group[6]) whose physical significance can be recognized from our experience of atomic physics.

In Sec. 2 we study the real nonsymmetric form of the theory. It appears that the skew-symmetric part of the affine connection ($\Gamma^i_{\underset{\sim}{jk}}$) is related to the spin angular momentum of matter rather than to its electromagnetic properties.[7] Since, in addition, atomic physics suggests that gauge transformations are imaginary, it is natural to adopt the complex form of the theory for electromagnetism (the spin then being described by the skew-Hermitian part of the affine connection).

This complex form is examined in Sec. 3. It turns out that the group associated with the vierbeins contains the gauge group as a normal subgroup. This enables one to identify the vector potential and the electromagnetic field tensor. In consequence, Einstein's Lagrangian is seen to be physically unsatisfactory, and an alternative Lagrangian is proposed. Both Lagrangians consist of the sum of two essentially independent terms, so that although the electromagnetic field is geometrized, it is not unified with the gravitational field. If such unification is desired, a new idea will be needed.

2. REAL NONSYMMETRIC THEORY

We first show by means of the vierbein analysis that in a space with a *symmetric* metric tensor g_{ij}, the most natural interpretation of $\Gamma^i_{\underset{\sim}{jk}}$ is that it is essentially the flux of the material spin angular momentum. We then consider the implications of this result for the Einstein-Schrödinger theory.

A detailed discussion of the relation between spin and affine connection has been given elsewhere[7]; here we shall describe only the main points. The essential step is to derive field equations in vierbein form by the Palatini method; that is, by varying a Lagrangian with respect to the vierbein and the affine connection independently. When the Lagrangian contains a term describing a material field with spin, it turns out that Γ^i_{jk} has a skew part $\Gamma^i_{\underset{\sim}{jk}}$ which is simply related to the spin flux S^i_{jk}.

We begin by introducing the vierbeins. These are four linearly independent real vector fields $\mathbf{e}(\alpha)$, which in a given coordinate system have components $e_i(\alpha)$. It is convenient to choose orthonormal vierbeins, so that

$$g_{ij}=\eta(\alpha\beta)e_i(\alpha)e_j(\beta), \tag{2.1}$$

where

$$\eta(\alpha\beta)=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&-1\end{bmatrix},$$

[1] V. Bargmann, Revs. Modern Phys. **29**, 161 (1957); B. Kaufman, Helv. Phys. Acta. Suppl. **4**, 227 (1956); A. Lichnerowicz, *Theories Relativistes de la Gravitation et de L'Electromagnetisme* (Masson, Paris, 1955); M. A. Tonnelat, *La Theorie du Champ Unifie d'Einstein* (Gauthiers-Villars, Paris, 1955).

[2] A. Einstein, Ann. Math. **46**, 578 (1945); A. Einstein and E. G. Straus, Ann. Math. **47**, 731 (1946); A. Einstein, Revs. Modern Phys. **20**, 35 (1948).

[3] A. Einstein, Can. J. Math. **2**, 120 (1950); *The Meaning of Relativity* (Princeton University Press, Princeton, New Jersey), 3rd ed. 1950, 4th ed. 1953, 5th ed. 1955; A. Einstein and B. Kaufman, *Louis de Broglie, Physicien et Penseur* (Albin Michel, Paris, 1952); Ann. Math. **59**, 230 (1954); **62**, 128 (1955).

[4] E. Schrödinger, Proc. Roy. Irish Acad. **51**, 163 (1947); **51**, 205 (1948); **52**, 1 (1948); **56**, 13 (1954); *Space-Time Structure* (Cambridge University Press, New York, 1950).

[5] H Weyl, Z. Physik **15**, 323 (1929).

[6] J. A. Schouten, *Ricci Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed.; A. Lichnerowicz, *Theorie globale des connexions et des groupes d'holonomi* (Cremonese, Rome, 1955).

[7] D. W. Sciama, *Festschrift for Infeld* (Pergamon Press, New York, to be published); T. W. B. Kibble (to be published).

and the summation convention is used for the Greek indexes which number the vierbeins. For a given metric tensor g_{ij}, the vierbeins are defined up to a Lorentz transformation which can vary arbitrarily with position in space time.

We now introduce an affine connection for these vierbeins. To do this we parallelly transfer the vierbeins at one point P to a neighboring point P' by means of the affine connection $\Gamma^i{}_{jk}$. The transferred vierbeins will differ from the local vierbeins at P' by an infinitesimal transformation $O(\alpha\beta)$, say. In an affine space, this transformation will depend linearly on the displacement $PP'(=dx^p)$; that is,

$$O(\alpha\beta)=O_p(\alpha\beta)dx^p.$$

We thus have[8]

$$[\partial e^p(\alpha)/\partial x^q]+\Gamma^p{}_{rq}e^r(\alpha)+O_q(\alpha\beta)e^p(\beta)=0, \quad (2.2)$$

and $O_q(\alpha\beta)$ may be called the vierbein affine connection.

So far we have been dealing with a general affine differential geometry. The basic relation (2.2) can now be specialized in various ways. Of these, the two most important are:

(a) To assume that $\Gamma^p{}_{rq}$ is symmetric. It can then be eliminated from (2.2), to give

$$e^q(\alpha)O_q(\beta\gamma)-e^q(\beta)O_q(\alpha\gamma) = e_p(\gamma)\left\{e^q(\beta)\frac{\partial e^p(\alpha)}{\partial x^q}-e^q(\alpha)\frac{\partial e^p(\beta)}{\partial x^q}\right\}.$$

We shall call this the integrability condition for $O_q(\alpha\beta)$.[7]

(b) To assume that $O_q(\alpha\beta)$ is skew in α and β. This would mean that the transformation between the transferred and the local vierbeins is generated by a skew matrix $O(\alpha\beta)$, so that it would be a Lorentz transformation. This in turn means that the length of a vector is preserved by parallel transfer, which also follows from the fact that when $O_q(\alpha\beta)$ is skew it can be eliminated from (2.2) to give [using (2.1)]

$$g^{ij}{}_{;k}\equiv\frac{\partial g^{ij}}{\partial x^k}+\Gamma^i{}_{lk}g^{lj}+\Gamma^j{}_{lk}g^{il}=0. \quad (2.3)$$

It is important for what follows to realize that conditions (a) and (b) are completely independent of one another. Only when both hold is the geometry Riemannian.

We now rewrite general relativity in the vierbein formalism. This means that our basic variables are $e_i(\alpha)$ and $O_q(\alpha\beta)$. The more familiar variables g_{ij} and $\Gamma^i{}_{jk}$ can then be computed from (2.1) and (2.2). Since the preservation of length by parallel transfer is characteristic of general relativity, we shall assume that $O_q(\alpha\beta)$ is skew. We can define a curvature tensor from the change in an arbitrary vector $A(\alpha)[=A_ie^i(\alpha)]$ when it is parallelly transferred around an infinitesimal closed circuit. The result is

$$R_{pq}(\alpha\beta)=\frac{\partial O_q(\alpha\beta)}{\partial x^p}-\frac{\partial O_p(\alpha\beta)}{\partial x^q} +O_p(\alpha\gamma)O_q(\gamma\beta)-O_q(\alpha\gamma)O_p(\gamma\beta),$$

which is skew in p, q and α, β. This curvature tensor (or rather its projections on surface elements) generates the infinitesimal holonomy group. The curvature scalar R is given by

$$R=e^p(\alpha)e^q(\beta)R_{pq}(\alpha\beta).$$

In the absence of matter, the Lagrangian density is

$$\epsilon R,$$

where

$$\epsilon=\det[e_i(\alpha)].$$

Variation with respect to $e^p(\alpha)$ leads to the gravitational field equations

$$R_p(\alpha)=0,$$

where

$$R_p(\alpha)=e^q(\beta)R_{pq}(\alpha\beta).$$

Variation with respect to $O_p(\alpha\beta)$ leads to the integrability condition for $O_p(\alpha\beta)$.[9] It follows that $\Gamma^p{}_{qr}$ is symmetric and that the geometry is Riemannian.

Now let us suppose that there is matter present, described by an additional Lagrangian which depends on a function ψ and its first derivative, in the manner of unquantized relativistic field theory. Variation with respect to $e^p(\alpha)$ now leads to

$$R_p(\alpha)-\tfrac{1}{2}Re_p(\alpha)=T_p(\alpha)$$

or

$$R(\alpha\beta)-\tfrac{1}{2}R\delta(\alpha\beta)=T(\alpha\beta),$$

where $R(\alpha\beta)=e^p(\beta)R_p(\alpha)$. Here $T(\alpha\beta)$ is not, in general, symmetric unless the material Lagrangian depends on $e^p(\alpha)$ only through the symmetric combination g^{pq}.[10]

The variation with respect to $O_p(\alpha\beta)$ is not quite so straightforward, since it involves the spin flux of matter. This flux is given by

$$S_i(\alpha\beta)=[\partial\mathfrak{L}/\partial(\partial\psi/\partial x^i)]S(\alpha\beta)\psi,$$

where the skew matrix $S(\alpha\beta)$ defines the way ψ transforms under a Lorentz transformation.[11] Now for the matter Lagrangian to be invariant under position-dependent Lorentz transformations of the vierbeins, derivatives of ψ which appear in it must be rewritten as covariant derivatives with respect to the vierbein affine $O_q(\alpha\beta)$.[7] This means that the matter Lagrangian will contain the term

$$S^q(\alpha\beta)O_q(\alpha\beta).$$

[8] It may not be possible to construct a vierbein field *globally* in a nonsingular manner. However, in this paper we are interested only in relations which hold in a neighborhood of any given point.

[9] H. Weyl, Phys. Rev. **77**, 699 (1950).

[10] L. Rosenfeld, Acad. roy. Belg. **18**, No. 6 (1940).

[11] E. M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave-equations* (Blackie and Son Limited, London, 1953).

In consequence, when the variation with respect to $O_q(\alpha\beta)$ is carried out, there will be a contribution from the matter Lagrangian, and the integrability condition for $O_q(\alpha\beta)$ will not be satisfied (as first pointed out by Weyl[9] for the Dirac electron). In this case, $\Gamma^i{}_{jk}$ will no longer be symmetrical, and in fact[7]

$$\Gamma^i{}_{\underset{\sim}{jk}}=S^i{}_{jk}-\tfrac{1}{2}\delta_j{}^i S^l{}_{lk}-\tfrac{1}{2}\delta_k{}^i S^l{}_{jl},$$

where

$$S^i{}_{jk}=e_j(\alpha)e_k(\beta)S^i(\alpha\beta).$$

We may note that this conclusion depends essentially on the fact that we have used the Palatini method of variation. However, we can look at the situation in another way. It will be recalled that in one interpretation of general relativity the field equations are not regarded as a restriction on the geometry of space time. Instead they represent an *identification* of the Einstein tensor as the energy-momentum tensor of matter. The justification for this identification is that the Einstein tensor is the simplest one (apart from g_{ij}) that identically satisfies the required conservation equations. On this view, the introduction of a material Lagrangian merely provides a phenomenological means of describing the Einstein tensor.[12]

We can adopt the same point of view here. If we write $X^p(\alpha\beta)$ for $\delta R/\delta O_p(\alpha\beta)$ [$X^p(\alpha\beta)=0$ is then the integrability condition for $O_p(\alpha\beta)$], we can simply identify $X^p(\alpha\beta)$ as the spin flux of matter, since it satisfies the identity[7]

$$(\partial/\partial x^p)[\epsilon X^p(\alpha\beta)]+2\epsilon X^p(\alpha\rho)O_p(\beta\rho)=\epsilon R(\underset{\sim}{\alpha\beta})=\epsilon T(\underset{\sim}{\alpha\beta}),$$

which will then express the conservation of total angular momentum (spin plus orbital). From this point of view, the introduction of a function ψ with prescribed transformation properties under a Lorentz transformation is just a phenomenological way of describing $X^p(\alpha\beta)$ or $\Gamma^i{}_{\underset{\sim}{jk}}$.

So long as the vierbein affine connection $O_p(\alpha\beta)$ and the curvature tensor $R_{pq}(\alpha\beta)$ are assumed to be skew in α and β and, therefore, the associated infinitesimal holonomy group to be the Lorentz group (or a subgroup), we expect the theory to describe only the spin (and, of course, energetic) properties of matter. To bring in the electromagnetic field, the theory must presumably also incorporate the gauge group. This it would do if we supposed that $O_p(\alpha\beta)$ has a symmetric part, for then $O_p(\alpha\alpha)$ would serve as a vector potential. (Its transformation law is

$$O_p'(\alpha\alpha)=O_p(\alpha\alpha)+(\partial\ \ln\theta/\partial x^p)$$

under a vierbein transformation of determinant θ.) Of course, lengths would no longer be preserved under parallel transfer. The simplest assumption for the symmetric part $O_p(\underline{\alpha\beta})$ of $O_p(\alpha\beta)$ is

$$O_p(\underline{\alpha\beta})=K_p\delta(\alpha\beta).$$

This just corresponds to Weyl's[13] original gauge theory of 1918 (except that Weyl then worked with a symmetrical $\Gamma^i{}_{jk}$). For a general $O_p(\underline{\alpha\beta})$, we retrieve Eddington's[14] theory (again with symmetrical $\Gamma^i{}_{jk}$).

This change in the infinitesimal holonomy group of the space is, in fact, achieved in the Einstein-Schrödinger theory at the cost of assuming that the "metric" tensor g_{ij} is nonsymmetric and that it satisfies the equations

$$g_{+-;k}^{ij}\equiv\frac{\partial g^{ij}}{\partial x^k}+\Gamma^i{}_{lk}g^{lj}+\Gamma^j{}_{kl}g^{il}=0$$

either *a priori*, or from the Palatini variation of a suitably chosen Lagrangian. This equation implies that lengths are not preserved under parallel transfer [since in the last term, the covariant indexes of $\Gamma^i{}_{jk}$ are in reverse order, cf. (2.3)]. It also differs from (2.3) in its implication[1,4] that $\Gamma^i{}_{jk}$ is completely determined by g_{ij}, since (2.3) imposes no restriction on $\Gamma^i{}_{\underset{\sim}{jk}}$.[4] From a geometrical point of view, however, this introduction of $g_{\underset{\sim}{ij}}$ seems quite unnatural. A better formalism can be devised on the basis of our experience with the Weyl-Eddington theory. In that theory, a decisive step forward was made (under the influence of quantum mechanics) when the length change under parallel transfer was converted into a phase change.[15] Correspondingly, the vector potential $O_p(\alpha\alpha)$ became imaginary. This development strongly suggests that if the Einstein-Schrödinger theory contains the electromagnetic field, this fact will appear most naturally in the complex form of the theory. This suggestion is studied in the next section.

3. COMPLEX THEORY

In this theory, the metric tensor g_{ij} is assumed to be complex and Hermitian. The vierbeins are now also complex,[16] and can be chosen to be orthonormal in the sense that

$$g_{ij}=\eta(\alpha\beta)e_i(\alpha)e_j{}^*(\beta), \qquad (3.1)$$

where * means complex conjugate. For a given g_{ij}, the vierbeins are defined up to a (quasi) unitary transformation ["quasi" because the signature of the Hermitian form (3.1) is 2], which can vary arbitrarily with position in space time. The relation between a real nonsymmetric tensor g_{ij} and vierbeins would be more complicated, and at one time Einstein[17] regarded this as a

[12] Of course, the material Lagrangian also describes the nongravitational behavior of matter. One of the goals of unified field theory is to construct a geometrical scheme so complete that no material Lagrangian need be introduced at all.

[13] H. Weyl, *Space-Time-Matter* (Dover Publications, New York, 1951).

[14] A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, New York, 1923).

[15] F. London, Z. Physik **42**, 375 (1927); H. Weyl, *ibid.* **56**, 330 (1929); V. Fock, *ibid.* **57**, 261 (1929).

[16] D. W. Sciama, Nuovo cimento **8**, 417 (1958).

[17] A. Einstein, Revs. Modern Phys. **20**, 35 (1948).

decisive reason for using the complex form of the theory. However, he did not develop this point.

Under parallel transfer to a neighboring point, the vierbeins will differ from the local vierbeins by a complex transformation, leading to a complex vierbein affine connection which we shall call $u_q(\alpha\beta)$. We then have the equation

$$[\partial e^p(\alpha)/\partial x^q]+\Gamma^p{}_{rq}e^r(\alpha)+u_q(\alpha\beta)e^p(\beta)=0, \quad (3.2)$$

where $\Gamma^p{}_{rq}$ is now a complex connection.

We can specialize (3.2) in two main ways:

(a) By assuming that $\Gamma^p{}_{rq}$ is Hermitian. It can then be eliminated from (3.2) to give

$$e^p(\gamma)e^{*q}(\alpha)u_q(\beta\gamma)-e^{*p}(\gamma)e^q(\beta)u_q{}^*(\alpha\gamma)$$
$$=e^q(\beta)[\partial e^{*p}(\alpha)/\partial x^q]-e^{*q}(\alpha)[\partial e^p(\beta)/\partial x^q]\cdots. \quad (3.3)$$

We shall call this the integrability condition for $u_q(\alpha\beta)$.

(b) By assuming that $u_q(\alpha\beta)$ is skew Hermitian in α and β. This would mean that the transformation between the transferred and the local vierbeins is generated by a skew Hermitian matrix $u(\alpha\beta)$, that is, it is a (quasi) unitary transformation. This in turn means that the Hermitian length $(g^{ij}A_iA_j{}^*)$ of a vector A_i is preserved by parallel transfer, which also follows from the fact that when $u_q(\alpha\beta)$ is skew Hermitian it can be eliminated from (3.2) to give [using (3.1)]

$$(\partial g^{ij}/\partial x^k)+\Gamma^i{}_{lk}g^{lj}+\Gamma^*{}_{lk}{}^jg^{il}=0. \quad (3.4)$$

If, in addition, we assume (a), that is, that $\Gamma^p{}_{rq}$ is Hermitian, (3.4) becomes

$$g^{ij}_{+-;p}\equiv(\partial g^{ij}/\partial x^k)+\Gamma^i{}_{lk}g^{lj}+\Gamma^j{}_{kl}g^{il}=0. \quad (3.5)$$

We note for future use that in case (b)

$$u_q(\alpha\alpha)=i\,\mathrm{Im}(\partial\ln\epsilon/\partial x^q)-i\,\mathrm{Im}\,\Gamma^p{}_{pq}. \quad (3.6)$$

We now rewrite the Einstein Lagrangian in vierbein form. The Riemann tensor is given by

$$R_{pq}(\alpha\beta)=[\partial u_q(\alpha\beta)/\partial x^p]-[\partial u_p(\alpha\beta)/\partial x^q]$$
$$+u_p(\alpha\gamma)u_q(\gamma\beta)-u_q(\alpha\gamma)u_p(\gamma\beta),$$

and is skew in p and q and skew Hermitian in α and β. The Lagrangian is then

$$\epsilon_0R,$$

where

$$\epsilon_0=(\det g_{ij})^{\frac{1}{2}}=|\epsilon|$$

and

$$R=e^p(\alpha)e^{*q}(\beta)R_{pq}(\alpha\beta).$$

Note that both ϵ_0 and R are real.

The variation with respect to $e^p(\alpha)$ leads to

$$R_p(\alpha)=0,$$

where

$$R_p(\alpha)=e^{*q}(\beta)R_{pq}(\alpha\beta).$$

The variation with respect to $u_p(\alpha\alpha)$ leads to

$$g^{pq}_{\smile,q}=0$$

(see Appendix). The variation with respect to $u_p(\alpha\beta)$ does not immediately lead to the integrability condition for $u_p(\alpha\beta)$. We must first "extract its trace" by making the transformation

$$u_p'(\alpha\beta)=u_p(\alpha\beta)-\tfrac{1}{4}u_t(\gamma\gamma)\delta(\alpha\beta)+\tfrac{1}{4}i\,\mathrm{Im}(\partial\ln\epsilon/\partial x^p)\delta(\alpha\beta)$$

for which

$$u_p'(\alpha\alpha)=i\,\mathrm{Im}(\partial\ln\epsilon/\partial x^p) \quad (3.7)$$

(cf. Schrödinger's[4] transformation for $\Gamma^p{}_{qr}$). Then $u_p'(\alpha\beta)$ satisfies the integrability conditions (see Appendix), and the associated affine connection $\Gamma'^p{}_{qr}$ is Hermitian and has

$$\mathrm{Im}\,\Gamma'^p{}_{pq}=\Gamma'^p{}_{\underset{\smile}{pq}}=0$$

[from (3.6) and (3.7)]. In terms of the primed variables, we have

$$R_p'(\alpha)+\tfrac{1}{4}e^{*q}(\alpha)R_{pq}(\beta\beta)=0. \quad (3.8)$$

Equations (3.5) [with $\Gamma'^p{}_{qr}$ replacing $\Gamma^p{}_{qr}$], (3.7), and (3.8) are then equivalent to the equations of the Einstein theory:

$$g^{ij}_{+-;k}=0, \quad \Gamma_i=0, \quad R_{\underline{ik}}=0,$$
$$R_{\underset{\smile}{ik}}+\tfrac{2}{3}(\lambda_{i,k}-\lambda_{k,i})=0.$$

We now consider the physical interpretation of this theory. The vector potential is clearly $u_p(\alpha\alpha)$, since (quasi) unitary transformations of the vierbeins of determinant $e^{i\theta}$ induce the transformation[16]

$$u_p'(\alpha\alpha)=u_p(\alpha\alpha)+i(\partial\theta/\partial x^p).$$

Accordingly, the Maxwell field is given by

$$F_{pq}=R_{pq}(\alpha\alpha)$$
$$=R^r{}_{rpq}$$

[from (3.6)].

With this identication in mind, we re-examine the Einstein Lagrangian. It can be written in the following form

$$\epsilon_0R=\epsilon_0\{[e^p(\alpha)e^{*q}(\beta)-e^q(\alpha)e^{*p}(\beta)]$$
$$\times[R_{pq}(\underset{\smile}{\alpha\beta})+R_{pq}(\underline{\alpha\beta})]\}.$$

We now write

$$R_{pq}(\underline{\alpha\beta})=\tfrac{1}{4}F_{pq}\delta(\alpha\beta)+S_{pq}(\alpha\beta)$$

and put

$$R_{pq}(\underset{\smile}{\alpha\beta})+S_{pq}(\alpha\beta)=Q_{pq}(\alpha\beta).$$

Then $Q_{pq}(\alpha\beta)$ and $F_{pq}\delta(\alpha\beta)$ are irreducible under quasi-unitary transformations.[18] In terms of these irreducible

[18] The theory could be somewhat simplified by assuming the Weyl-type relation

$$u_p(\alpha\beta)=\tfrac{1}{4}u_p(\gamma\gamma)\delta(\alpha\beta)$$

to hold. In this case $S_{pq}(\alpha\beta)=0$, and the part of the theory associated with unimodular quasi-unitary transformations would be more closely related to the ordinary gravitational theory.

quantities, the Einstein Lagrangian becomes

$$\epsilon_0 R = \epsilon_0 e^p(\alpha) e^{*q}(\beta) Q_{pq}(\alpha\beta) + \tfrac{1}{2}\mathfrak{g}^{\underset{\sim}{pq}} F_{pq}.$$

The variation of the first term in this Lagrangian with respect to $u_p'(\alpha\beta)$ implies that the Γ' connection associated with $u_p'(\alpha\beta)$ is Hermitian. The second term, which is added to the first in a nonunitary fashion, yields $\mathfrak{g}^{\underset{\sim}{pq}}{}_{,q}=0$ on variation with respect to $u_p(\alpha\alpha)$.[19] It would clearly be a Maxwell-type Lagrangian if $\mathfrak{g}^{\underset{\sim}{pq}}$ could be interpreted as the conjugate of F_{pq} in the sense of the Born-Infeld theory.[20] However, this would require that $g_{\underset{\sim}{pq}}$ should reduce to λF_{pq} for weak fields, which is not, in general, true with our present field equations. It is, therefore, not surprising that the equations of motion of singularities do not contain the Lorentz force.[21]

In order to obtain a correct description of electromagnetism, we must replace the term $\mathfrak{g}^{\underset{\sim}{pq}} F_{pq}$ by one more closely related to the familiar Maxwellian Lagrangian, e.g., $\mathfrak{g}^{pr}g^{qs}F_{rs}F_{pq}$ (cf. reference 22 for another choice). One will then obtain the source-free Maxwell equations and the contribution of the Maxwell stress tensor to the gravitational field equations, which ensures that the Lorentz force is acting on singularities of the electromagnetic field.[23] If a nonsingular description of charges is desired, there are three possible approaches. First, one can introduce Wheeler-Misner[24] wormholes into the topology of space time. Second, one can do without field equations and simply identity $(\mathfrak{g}^{pr}g^{qs}F_{rs})_{,q}$ as the current density j^p. Finally, one can introduce a material wave function ψ and a material Lagrangian L_m. If ψ transforms like a density of weight λ under (quasi) unitary transformations of the vierbeins, then I_m will contain the term[16]

$$\lambda j^p u_p(\alpha\alpha),$$

where

$$j^p = \frac{\partial L_m}{\partial(\partial\psi/\partial x^p)}\psi - \frac{\partial L_m}{\partial(\partial\psi^*/\partial x^p)}\psi^*.$$

The variation with respect to $u_p(\alpha\alpha)$ will then lead to Maxwell's equations with source term. Unless the matter is spinless, that is, unless ψ is a scalar density, L_m will also make a contribution to the variation with respect to $u_p'(\alpha\beta)$. By considerations similar to those of Sec. 2, it can be seen that the spin of matter will give rise to a skew Hermitian part of $\Gamma^i{}_{jk}$. Of course, from the geometrical point of view, this introduction of ψ is purely phenomenological, and has no fundamental significance.

It is to be emphasized that although the electromagnetic field is geometrized in this theory, our Lagrangian, like that of Einstein, is nonunified in the sense that it consists of a sum of unrelated terms. The reason for this is that the infinitesimal holonomy group is nonsimple; it has a normal subgroup consisting of transformations of determinant $e^{i\theta}(\theta\neq 0)$. Correspondingly its generator, the curvature tensor, is not irreducible (as it is in Riemannian geometry), nor is the scalar curvature which acts as the Lagrangian in Einstein's theory. The alternative Lagrangian proposed here does not even try to be irreducible.

Einstein has attempted to overcome this difficulty by introducing new invariance groups which mix up the terms in his Lagrangian, but these new groups do not appear to have physical significance, and the device seems to be a purely formal one. In any case, it is ineffective for the physically reasonable Lagrangians. We conclude that if unification is desired as well as geometrization, a new group will be required. There are some grounds for thinking that this new group may be the symplectic group. This question will be discussed elsewhere.

ACKNOWLEDGMENTS

This work was begun in the Department of Mathematics at King's College, London, with the support of the Aeronautical Research Laboratory, and completed in the Department of Astronomy at Cornell University. I am grateful to Professor H. Bondi and Professor T. Gold for their respective hospitality, and to Dr. R. Penrose for helpful discussions. The work was partially supported by a research contract with the Air Force Office of Scientific Research.

APPENDIX

We here derive the integrability conditions for $u_p'(\alpha\beta)$ from the vanishing of the variation of $\int \epsilon_0 R d\tau$ with respect to $u_p(\alpha\beta)$. For a variation $\delta u_s(\lambda\mu)$

$$\begin{aligned}\delta R = {} & e^p(\lambda)e^{*s}(\mu)\delta u_s(\lambda\mu)_{,p} - e^s(\lambda)e^{*q}(\mu)\delta u_s(\lambda\mu)_{,q} \\ & + e^s(\lambda)e^{*q}(\beta)u_q(\mu\beta)\delta u_s(\lambda\mu) \\ & + e^p(\alpha)e^{*s}(\mu)u_p(\alpha\lambda)\delta u_s(\lambda\mu) \\ & - e^p(\lambda)e^{*s}(\beta)u_p(\mu\beta)\delta u_s(\lambda\mu) \\ & - e^s(\alpha)e^{*q}(\mu)u_q(\alpha\lambda)\delta u_s(\lambda\mu).\end{aligned}$$

We therefore require

$$\int \epsilon_0 \delta R d\tau = 0.$$

This can be simplified by partial integration, which leads to

$$\begin{aligned}&\{\epsilon_0 e^s(\lambda)e^{*p}(\mu) - \epsilon_0 e^p(\lambda)e^{*s}(\mu)\}_{,p} \\ &\quad + u_p(\mu\alpha)\{\epsilon_0 e^s(\lambda)e^{*p}(\alpha) - \epsilon_0 e^p(\lambda)e^{*s}(\alpha)\} \\ &\quad + u_p(\alpha\lambda)\{\epsilon_0 e^p(\alpha)e^{*s}(\mu) - \epsilon_0 e^s(\alpha)e^{*p}(\mu)\} = 0. \quad \text{(A1)}\end{aligned}$$

[19] This shows the group theoretical significance of the u' transformation and the corresponding Schrödinger[4] transformation to his "starred affinity."

[20] M. Born and L. Infeld, Proc. Roy. Soc. (London) **A144**, 425 (1934).

[21] J. Callaway, Phys. Rev. **92**, 1567 (1953); W. B. Bonnor, Ann. inst. Henri Poincaré **15**, 133 (1957).

[22] B. Kursunoglu, Phys. Rev. **88**, 1369 (1952); W. B. Bonnor, Proc. Roy. Soc. (London) **A226**, 366 (1954).

[23] L. Infeld and P. R. Wallace, Phys. Rev. **57**, 797 (1940).

[24] J. A. Wheeler and C. W. Misner, Ann. Phys. **2**, 525 (1957).

If this equation is contracted over μ and λ, we obtain

$$\mathfrak{g}^{sp}{}_{,p}=0.$$

The full equation can be simplified by writing

$$\xi^s(\lambda\mu)=-\epsilon_0 e^p(\lambda)e^{*s}(\alpha)u_p(\mu\alpha)-\epsilon_0 e^p(\lambda)[\partial e^{*s}(\mu)/\partial x^p],$$
$$\chi^s(\lambda\mu)=-\epsilon_0 e^s(\alpha)e^{*p}(\mu)u_p(\alpha\lambda)+\epsilon_0 e^{*p}(\mu)[\partial e^s(\lambda)/\partial x^p].$$

Then (A1) is equivalent to

$$\xi^s{}_{tr}+\chi^s{}_{tr}-\xi^p{}_{pr}\delta_t{}^s-\chi^p{}_{tp}\delta_r{}^s + (\partial\epsilon_0/\partial x^r)\delta_t{}^s-(\partial\epsilon_0/\partial x^t)\delta_r{}^s=0, \quad \text{(A2)}$$

where

$$\xi^s{}_{tr}=e_t(\lambda)e_r{}^*(\mu)\xi^s(\lambda\mu),$$
$$\chi^s{}_{tr}=e_t(\lambda)e_r{}^*(\mu)\chi^s(\lambda\mu).$$

We note that the integrability conditions (3.3) for $u_p(\alpha\beta)$ are equivalent to[25]

$$\xi^s{}_{tr}+\chi^s{}_{tr}=0.$$

We therefore want to avoid the extra terms in (A2). To this end we make the transformation[26]

$$u_t{}'(\mu\alpha)=u_t(\mu\alpha)-\tfrac{1}{4}u_t(\gamma\gamma)\delta(\mu\alpha) -\tfrac{1}{4}i\,\mathrm{Im}(\partial\ln\epsilon/\partial x^t)\delta(\mu\alpha). \quad \text{(A3)}$$

Now

$$\xi^p{}_{rp}=-\epsilon_0 u_r(\alpha\alpha)+\epsilon_0(\partial\ln\epsilon^*/\partial x^r)$$

and

$$\chi^p{}_{pr}=-\epsilon_0 u_r(\alpha\alpha)-\epsilon_0(\partial\ln\epsilon/\partial x^r).$$

Thus

$$\xi^p{}_{rp}-\chi^p{}_{pr}=2(\partial\epsilon_0/\partial x^r)=\xi^p{}_{pr}-\chi^p{}_{rp}$$

from the contraction of (A2). Also, contracting (A2) gives

$$4(\xi^p{}_{pr}+\chi^p{}_{rp})=\xi^p{}_{rp}+\chi^p{}_{pr}.$$

Hence, (A3) can be written

$$\mu_r{}'(\mu\alpha)=u_r(\mu\alpha)+(1/8\epsilon_0)(\xi^p{}_{rp}+\chi^p{}_{pr})\delta(\mu\alpha) =u_r(\mu\alpha)+(1/2\epsilon_0)(\xi^p{}_{pr}+\chi^p{}_{rp})\delta(\mu\alpha).$$

We now introduce $\xi'^s{}_{tr}$ and $\chi'^s{}_{tr}$ defined in terms of $u_r{}'(\mu\alpha)$. They are given by

$$\xi'^s{}_{tr}=\xi^s{}_{tr}-\tfrac{1}{2}(\xi^p{}_{pt}+\chi^p{}_{tp})\delta_r{}^s,$$
$$\chi'^s{}_{tr}=\chi^s{}_{tr}-\tfrac{1}{2}(\xi^p{}_{pr}+\chi^p{}_{rp})\delta_t{}^s.$$

Hence,

$$\begin{aligned}\xi'^s{}_{tr}+\chi'^s{}_{tr}&=\xi^s{}_{tr}+\chi^s{}_{tr}-\delta_r{}^s(\tfrac{1}{2}\xi^p{}_{pt}+\tfrac{1}{2}\chi^p{}_{tp}) -\delta_t{}^s(\tfrac{1}{2}\xi^p{}_{pr}+\tfrac{1}{2}\chi^p{}_{rp})\\ &=\xi^s{}_{tr}+\chi^s{}_{tr}-\delta_r{}^s(\chi^p{}_{tp}-\tfrac{1}{2}\chi^p{}_{tp}+\tfrac{1}{2}\xi^p{}_{pt}) -\delta_t{}^s(\xi^p{}_{pr}-\tfrac{1}{2}\xi^p{}_{pr}+\tfrac{1}{2}\chi^p{}_{rp})\\ &=\xi^s{}_{tr}+\chi^s{}_{tr}-\chi^p{}_{tp}\delta_r{}^s-\xi^p{}_{pr}\delta_t{}^s -(\partial\epsilon_0/\partial x^t)\delta_r{}^s+(\partial\epsilon_0/\partial x^t)\delta_t{}^s\\ &=0\end{aligned}$$

from (A2). This implies that $u_p{}'(\alpha\beta)$ satisfies the integrability conditions, which was to be proved.

[25] Strictly speaking, this gives the integrability conditions for $u_p{}^*(\alpha\beta)$, but this is unimportant since the theory is invariant under complex conjugation.

[26] The last term has the opposite sign from that used in the text since we are here really finding the integrability conditions for $u_p{}'^*(\alpha\beta)$.

Characteristic Coulomb Green's Function and Its Eigenfunction Expansion

ROBERT A. MAPLETON
Geophysics Research Directorate, Air Force Cambridge Research Laboratories, Bedford, Massachusetts
(Received January 3, 1961)

The eigenfunction form of the Green's function is derived from the characteristic form of the Green's function for the nonrelativistic Coulomb operator. This derivation shows how the wave boundary condition of the characteristic form of the Green's function is related to the branch cut in the continuous part of the spectrum. An application of the eigenfunction expansion form of this Green's function is discussed.

IN the theory of scattering, the Green's function is commonly displayed as an eigenfunction expansion. The diverging (or converging) wave character of the Green's function is dictated by the sign of the imaginary component of the energy parameter. In the continuous part of the spectrum, the eigenfunctions may display both the diverging and converging wave behavior for large distances from the scattering center; for example, they may behave as $[\sin(kr-\alpha)]/kr$, with r the distance and k the wave number of the continuum. Consequently, it is not obvious that the expansion form of the Green's function does behave asymptotically as asserted. However, the characteristic form of the Green's function displays the wave boundary conditions explicitly. This is illustrated for the one-dimensional case of the Sturm-Lionville operator on the interval $(0,\infty)$.

$$\left(\frac{d}{dx}p(x)\frac{d}{dx}+q(x)+\lambda\right)G(x,x')=\delta(x-x'),$$

$$G(x,x')=\frac{u(x>;\lambda)v(x<;\lambda)}{p(x)W(u,v)}.$$

λ is a parameter, which is not an eigenvalue, and W is the Wronskian of the two linearly independent solutions u and v appropriate to the boundary conditions of the problem.[1] The eigenfunction expansion form of G is

$$G(x,x')=\sum_n \frac{u_n(x)u_n^*(x')}{\lambda-\lambda_n},$$

with λ_n the eigenvalues such that $\lambda_n<0$ for the discrete part of the spectrum and $\lambda_n>0$ for the continuous part of the spectrum. It is thus interesting to inquire how one can proceed from one form of the Green's function to the other form, and to learn just how the imposed boundary condition at infinity forces a particular sign on the imaginary component of the energy.

The relation between the characteristic form of the Green's function and its spectrum is well known, and this relation leads to the expansion theorem.[1-3] In this paper, the eigenfunction form of the Green's function is derived from the characteristic form of the Green's function for the special case of the Coulomb operator. (The expansion theorem for this operator was derived by Titchmarsh.[4]) The Green's function that satisfies Eq. (1a) is given by Eq. (1b).[5]

$$[\nabla^2+(2\mu e^2/\hbar^2R)+K^2]G^{(\pm)}(\mathbf{R},\mathbf{R}';K)=\delta(\mathbf{R}-\mathbf{R}'), \quad (1a)$$

$$G^{(\pm)}=-(K/4\pi)\sum_{l=0}^{\infty}(2l+1)P_l(\cos\Theta)A_l^{(\pm)},$$

$$P_l(\cos\Theta)=\sum_{m=-l}^{l}\frac{(l-|m|)!}{(l+|m|)!}\times P_l^{|m|}(\cos\theta)P_l^{|m|}(\cos\theta')e^{im(\phi-\phi')},$$

$$A_l^{(\pm)}=H_l^{(\pm)}[l+1-(ic/K),2l+2;-i2KR]\times L_l[l+1-(ic/K),2l+2;-i2KR], \quad R>R',$$

$$H_l^{(\pm)}=\begin{pmatrix}2iW_1\\-2iW_2\end{pmatrix}e^{(\pi c/2K)}\frac{|\Gamma[l+1-(ic/K)]|}{(2l+1)!}\times(2KR)^le^{iKR},$$

$$L_l=e^{(\pi c/2K)}\frac{|\Gamma[l+1-(ic/K)]|}{(2l+1)!}\times(2KR')^le^{iKR'}(W_1+W_2),$$

$$W_1+W_2={}_1F_1[l+1-(ic/K),2l+2;-i2KR'],$$

$$c=\mu/ma_0, \quad a_0=\hbar^2/me^2. \quad (1b)$$

In Eqs. (1a) and (1b), e=electron charge, μ=reduced mass, $\hbar$=Planck's constant divided by 2π, a_0=first Bohr radius, and $K^2=2\mu E/\hbar^2$, with E the energy. (The absolute-value signs which appear on m in the addition theorem are omitted in the subsequent work.) $W_{1,2}$ are the two linearly independent solutions of the confluent hypergeometric equation that have the asymptotic forms given in Eq. (1c).[6]

$$H_l^{(\pm)}\sim\exp(\pm i)[KR-(l\pi/2)+\eta_l+(c/K)\log(2KR)],$$
$$\eta_l=\arg\Gamma[l+1-(ic/K)]. \quad (1c)$$

[1] E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second Order Differential Equations* (Oxford University Press, London), Vol. I (1946); Vol. II (1958).
[2] Nathan Marcuwitz, Communs. Pure and Appl. Math. 4, 263 (1951).
[3] Bernard Friedman, *Principles and Techniques of Applied Mathematics* (John Wiley & Sons, Inc., New York, 1956). Chaps. 3 and 4; B. Friedman and E. Gerjuoy, *Scattering Problems in Nonrelativistic Quantum Mechanics*, Research Report CX-4, Cont. AF 19(122)-463, New York University.
[4] Reference 1, Vol. I, p. 87; Vol. II, p. 134.
[5] Robert A. Mapleton, Phys. Rev. 117, 479 (1960).
[6] W. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford University Press, New York, 1949), 2nd ed., p. 52.

In Eq. (1b), $R>R'$, and this will be assumed in the subsequent calculations. Since the eigenfunction expansion of G is symmetric in R and R', only this case has to be treated.

Make the substitution $K=+\lambda^{\frac{1}{2}}$ in Eq. (1b) and consider the integral on a circle of infinite radius in the complex λ plane as given by Eq. (2a).[2]

$$I^{(+)}=\frac{1}{2\pi i}\oint\frac{d\lambda'}{\lambda'-(\lambda+i\epsilon)}G^{(+)}(\lambda';\mathbf{R},\mathbf{R}'),\qquad\lambda=K^2,\quad\epsilon>0. \tag{2a}$$

(In the remaining discussion of the diverging wave form of G, the superscript $(+)$ is omitted.) On this circle,

$$A_l\sim(e^{i[\lambda'^{\frac{1}{2}}R-(l\pi/2)]}/R\lambda'^{\frac{1}{2}})\times\frac{\sin[\lambda'^{\frac{1}{2}}R'-(l\pi/2)]}{R'\lambda'^{\frac{1}{2}}},$$

and

$$\lim_{\lambda'\to\infty}\frac{\lambda'^{\frac{1}{2}}A_l d\lambda'}{\lambda'-(\lambda+i\epsilon)}\to 0,$$

provided that $0<\arg\lambda'<2\pi$. I of Eq. (2a) is zero; the integral is evaluated at $\lambda'=\lambda+i\epsilon$, and the resulting closed contour is deformed so that $\lambda+i\epsilon$ is excluded, but the branch cut on $(0\leq\lambda'\leq\infty)$ and the poles of G on the negative λ' axis are included. With this task accomplished, Eq. (2a) is written as Eq. (2b).

$$G(\lambda+i\epsilon)=-\frac{1}{2\pi i}\int_{C_1}\frac{d\lambda' G(\lambda')}{\lambda'-(\lambda+i\epsilon)}-\sum(\text{Residues of }[G(\lambda')/\lambda'-(\lambda+i\epsilon)]\text{ at the poles of }G(\lambda')). \tag{2b}$$

C_1 is a contour that encircles the branch cut in a counterclockwise direction. Call this contour integral I_c.

$$I_c=-\frac{1}{2\pi i}\left[\int_\infty^0\frac{d\lambda'}{\lambda'-(\lambda+i\epsilon)}G(\lambda')+\int_0^\infty\frac{d\lambda'}{\lambda'-(\lambda+i\epsilon)}G(\lambda' e^{2\pi i})\right]. \tag{2c}$$

At this stage of the development, it is easy to understand why $\lambda+i\epsilon$ is associated with the diverging wave form of G. In order to derive Eqs. (2b) and (2c), the principal value of $\arg\lambda'^{\frac{1}{2}}$ was restricted by the inequality $0<\arg\lambda'^{\frac{1}{2}}<\pi$. This specification of $\arg\lambda'^{\frac{1}{2}}$ is consistent with the relation $K=\lim_{\epsilon\to 0}(\lambda+i\epsilon)^{\frac{1}{2}}$ but is inconsistent with the form $(\lambda-i\epsilon)^{\frac{1}{2}}$, since the branch cut is approached from below, and thus, $\lim_{\epsilon\to 0}(\lambda-i\epsilon)^{\frac{1}{2}}=-K$.

In order to establish the relation between G on opposite sides of the branch cut, several integral representations are introduced.[6]

$$\begin{aligned}
{}_1F_1(a,c;z)&=\frac{(c-1)!}{2\pi i}\int^{(z+,0+)}dt e^t t^{(a-c)}(t-z)^{-a},\\
W_1(a,c;z)&=\frac{(c-1)!}{2\pi i}\int_{-\infty}^{(0+)}dt e^t t^{(a-c)}(t-z)^{-a},\\
W_2(a,c;z)&=\frac{(c-1)!}{2\pi i}e^z\int_{-\infty}^{(0+)}dt e^t t^{-a}(t+z)^{(a-c)},\\
-a&=-l-1+(ic/\lambda'^{\frac{1}{2}}),\\
a-c&=-l-1-(ic/\lambda'^{\frac{1}{2}}).
\end{aligned} \tag{3a}$$

In Eq. (3a), the notation $(z+,0+)$ signifies that the contour integral encircles the points $t=z$, 0 in a counterclockwise sense. (There is a branch cut connecting $t=z$, 0.) The contour which represents ${}_1F_1$ is replaced by two contours. Each of these contours proceeds from $(-\infty)$, encircles one of the points $t=z$, 0 in a counterclockwise sense, and returns to $(-\infty)$. The contour that encircles $t=0$ defines W_1, and the other contour is transformed to encircle $t=0$ by the substitution $u=t-z\to t$. $W_{1,2}$ are multivalued functions, but they are rendered single valued by the following choice of initial and terminal values of the phases at $t=-\infty$.

W_2: $\arg t$, $(-\pi$ to $\pi)$, $\arg(t+z)$, $(-\pi$ to $-\pi)$;

W_1: $\arg t$, $(-\pi$ to $\pi)$, $\arg(t-z)$, $(\pi$ to $\pi)$.

An examination of the contour that defines ${}_1F_1$ shows that this selection of phases is consistent with the position of $t=z=-i2\lambda^{\frac{1}{2}}R$ relative to $t=0$ in the t plane for λ real and positive. It is easy to derive a useful identity with the use of Eq. (3a). Thus,

$${}_1F_1(c-a,c;-z)=\frac{(c-1)!}{2\pi i}\int^{(0+,-z+)}dt e^t t^{-a}(t+z)^{(a-c)},$$

and with the change of variable, $t+z=\mu\to t$, this becomes

$${}_1F_1(c-a,c;-z)=\frac{(c-1)!}{2\pi i}e^z\int^{(z+,0+)}dt e^t t^{(a-c)}(t-z)^{-a},$$

from which relation the identity emerges.[7]

$$e^z{}_1F_1(a,c;-2z)=e^{-z}{}_1F_1(c-a,c;2z).$$

From Eq. (3a), one obtains

$$\begin{aligned}
&W_1(\lambda e^{2\pi i})\\
&\quad=W_1(l+1+(ic/\lambda^{\frac{1}{2}}),2l+2;i2\lambda^{\frac{1}{2}}R)\\
&\quad=\frac{(2l+1)!}{2\pi i}\int_{-\infty}^{(0+)}dt e^t t^{[-l-1+(ic/\lambda^{\frac{1}{2}})]}\qquad(-\pi\to\pi)\\
&\qquad\times(t-i2\lambda^{\frac{1}{2}}R)^{-[l+1+(ic/\lambda^{\frac{1}{2}})]}\qquad(\pi\to\pi)\\
&\quad=e^{i2\lambda^{\frac{1}{2}}R}e^{(2\pi c/\lambda^{\frac{1}{2}})}W_2(l+1-(ic/\lambda^{\frac{1}{2}}),2l+2;-i2\lambda^{\frac{1}{2}}R).
\end{aligned}$$

[7] W. Magnus and F. Oberhettinger, *Special Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1949), English edition, p. 87.

The factor $e^{(2\pi c/\lambda'^{\frac{1}{2}})}$ originates from the factor $e^{2\pi i}$ that is abstracted from $(t-i2\lambda^{\frac{1}{2}}R)^{-[l+1+(ic/\lambda^{\frac{1}{2}})]}$ so that the phase of this term varies from $-\pi$ to $-\pi$, which phase is required in the integral that defines W_2. Now that the relations of the quantities above and below the branch cut have been derived, Eq. (2c) can be written in the form given by Eq. (3b).

$$I_c=\frac{1}{4\pi^2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(2l+1)\frac{(l-m)!}{(l+m)!}P_l^m(\cos\theta)P_l^m(\cos\theta')e^{im(\phi-\phi')}\int_0^{\infty}\frac{d\lambda'\lambda'^{\frac{1}{2}}}{\lambda+i\epsilon-\lambda'}\frac{|\Gamma[l+1-(ic/\lambda'^{\frac{1}{2}})]|^2}{[(2l+1)!]^2}e^{(c\pi/\lambda'^{\frac{1}{2}})}(2\lambda'^{\frac{1}{2}}R)^l$$
$$\times(2\lambda'^{\frac{1}{2}}R')^l e^{i\lambda'^{\frac{1}{2}}(R+R')}{}_1F_1(l+1-(ic/\lambda'^{\frac{1}{2}}),\,2l+2;\,-i2\lambda'^{\frac{1}{2}}R)\,{}_1F_1(l+1-(ic/\lambda'^{\frac{1}{2}}),\,2l+2;\,-i2\lambda'^{\frac{1}{2}}R'). \quad (3b)$$

The change of variable $\lambda'=k^2$ is made; the relation,

$$e^{ikR'}{}_1F_1(l+1-(ic/k),\,2l+2;\,-i2kR')=e^{-ikR'}{}_1F_1(l+1+(ic/k),\,2l+2;\,i2kR'),$$

is used, and several definitions and identities are invoked.

$$N_{lm}=\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!},$$

$$d\mathbf{k}=dkk^2d\Omega_k=dkk^2d\theta_k\sin\theta_k d\phi_k,$$

$$N_{lm}P_l^m(\cos\theta)P_l^m(\cos\theta')e^{im(\phi-\phi')}$$
$$=N_{lm}{}^2\int d\Omega_k P_l^m(\cos\theta)P_l^m(\cos\theta_k)e^{im(\phi-\phi_k)}P_l^m(\cos\theta_k)P_l^m(\cos\theta')e^{im(\phi_k-\phi')},$$

$$\phi_{klm}(\mathbf{R})=(2/\pi)^{\frac{1}{2}}N_{lm}P_l^m(\cos\theta)P_l^m(\cos\theta_k)e^{im(\phi-\phi_k)}e^{(c\pi/2k)}\frac{\Gamma[l+1-(ic/k)]}{(2l+1)!}(2kR)^l e^{ikR}{}_1F_1(l+1-(ic/k),\,2l+2;\,-i2kR).$$

These facts are used to rewrite Eq. (3b) as Eq. (3c).

$$I_c=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\int\frac{d\mathbf{k}\phi_{klm}(\mathbf{R})\phi_{klm}{}^*(\mathbf{R}')}{\lambda+i\epsilon-k^2}. \quad (3c)$$

These continuum functions are normalized to the delta function; that is, they satisfy the following relation[8]:

$$\sum_{l,l'=0}^{\infty}\sum_{m'=-l'}^{l'}\sum_{m=-l}^{l}\int d\mathbf{R}\phi_{klm}(\mathbf{R})\phi_{k'l'm'}{}^*(\mathbf{R})=\frac{\delta(k-k')\delta(\theta_k-\theta_{k'})\delta(\phi_k-\phi_{k'})}{k^2\sin\theta_k}=\delta(\mathbf{k}-\mathbf{k}').$$

The next task to accomplish is the evaluation of G at its poles.

The factor $|\Gamma[l+1-(ic/\lambda^{\frac{1}{2}})]|^2$ of Eq. (1b) is written as $\Gamma[l+1+(ic/\lambda^{\frac{1}{2}})]\Gamma[l+1-(ic/\lambda^{\frac{1}{2}})]$. There are poles of the gamma function at $l+1-(ic/\lambda^{\frac{1}{2}})=-(n-l-1)$, and n assumes the values $l+1$, $l+2$, $\cdots$. Therefore, λ satisfies the equation $i\lambda^{\frac{1}{2}}=-c/n$. Since $\arg\lambda^{\frac{1}{2}}$ is restricted by $0<\arg\lambda^{\frac{1}{2}}<\pi$, this is the correct pole; for $l+1+(ic/\lambda^{\frac{1}{2}})=-(n-l-1)$ would require that $-\pi<\arg\lambda^{\frac{1}{2}}<0$ which is not satisfied. In Eq. (3a), we have $-a=n-l-1$, $a-c=-(n+l+1)$, $-i2\lambda^{\frac{1}{2}}R=2cR/n=x$, and

$${}_1F_1(a,c;\,x)=(2l+1)!/2\pi i\int^{(0+)}dte^t t^{-(n+l+1)}\times(t-x)^{(n-l-1)}.$$

Since $(-a)$ is a positive integer (or zero), the original requirement that the contour encircle the point $t=z$ is removed.[6] Consequently, one has the following relation:

$${}_1F_1(-n+l+1,\,2l+2;\,x)$$
$$=W_1(-n+l+1,\,2l+2;\,x)$$
$$=\frac{(2l+1)!}{(n+l)!}\left[\frac{d^{n+l}}{dt^{n+l}}e^t(t-x)^{n-l-1}\right]_{t=0}. \quad (4a)$$

The Leibnitz Rule for the differentiation of a product is used to show that Eq. (4a) is consistent with Eq. (4b).[9,10]

[8] A. Sommerfeld, Ann. Physik **11**, 257 (1931); Robert A. Mapleton, Phys. Rev. **109**, 1166 (1958).

[9] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, p. 785.

[10] L. Pauling and E. B. Wilson, *Introduction to Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1935), pp. 130–132; H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic Press, Inc., New York, 1957), p. 13. ($L_{n-l-1}{}^{2l+1}$ of reference 9 $=-L_{n+l}{}^{2l+1}$ of these references.)

$$ {}_1F_1 = W_1 = \frac{(2l+1)!(n-l-1)!}{[(n+l)!]^2} L_{n-l-1}^{2l+1}(x), $$

$$ \int_0^\infty dx x^{2l+2} e^{-x} [L_{n-l-1}^{2l+1}(x)]^2 = \frac{2n[(n+l)!]^3}{(n-l-1)!}. \tag{4b} $$

The substitution $\mu = l+1-(ic/\lambda^{\frac{1}{2}})$ is used to evaluate the residue of[11]

$$ \Gamma(\mu) \quad \text{at} \quad \mu = -(n-l-1). $$

$$ \lambda'^{\frac{1}{2}} d\lambda' = -\frac{i2c^3}{(l+1-\mu)^4} d\mu, $$

$$ \text{Residue } \Gamma(\mu)]_{\mu=-(n-l-1)} = \frac{(-1)^{n-l-1}}{(n-l-1)!}. $$

These facts are collected to obtain Eq. (4c).[12]

$$ -\text{Residue} \left[\frac{G(\lambda')}{\lambda'-\lambda-i\epsilon} \right]_{\lambda'=-c^2/n^2} $$

$$ = \frac{\phi_{nlm}(\mathbf{R})\phi_{nlm}^*(\mathbf{R}')}{\lambda+i\epsilon+c^2/n^2}, \quad \phi_{nlm}(\mathbf{R}) $$

$$ = \left[\left(\frac{2c}{n}\right)^3 \frac{2l+1}{8\pi n} \frac{(n-l-1)!}{[(n+l)!]^3} \frac{(l-m)!}{(l+m)!} \right]^{\frac{1}{2}} $$

$$ \times e^{-(cR/n)} \left(\frac{2cR'}{n}\right)^l L_{n-l-1}^{2l+1}\left(\frac{2cR}{n}\right) $$

$$ \times P_l^m(\cos\theta) e^{im\phi}. \tag{4c} $$

One observes that

$$ \sum_{l=0}^{\infty} \sum_{n=l+1}^{\infty} = \sum_{n=1}^{\infty} \sum_{l=0}^{n-1}, $$

so that with the aid of Eqs. (2b), (3c), and (4c), one obtains Eq. (4d) as the desired result.

$$ G^{(+)}(K;\, \mathbf{R},\mathbf{R}') = \lim_{\epsilon\to 0} \left[\sum_{n=1}^{\infty} \sum_{l=0}^{n-1} \sum_{m=-l}^{l} \frac{\phi_{nlm}(\mathbf{R})\phi_{nlm}^*(\mathbf{R}')}{\lambda+i\epsilon+c^2/n^2} \right. $$

$$ \left. + \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \int d\mathbf{k} \frac{\phi_{klm}(\mathbf{R})\phi_{klm}^*(\mathbf{R}')}{\lambda+i\epsilon-k^2} \right]. \tag{4d} $$

The derivation of the expansion for the converging wave Green's function, $G^{(-)}$, is sketched. In Eq. (1b), $2iW_1$ is replaced by $-2iW_2$. The integral over the large circle given by Eq. (2a) vanishes provided that $-2\pi < \arg\lambda' < 0$. (In Eq. (2a), ϵ is replaced by $-\epsilon$; this will be assumed throughout the discussion; the superscript on $G^{(-)}$ is omitted.) Thus, Eq. (2b) is transformed into Eq. (5a).

$$ G = -\frac{1}{2\pi i} \int_{C_1} \frac{d\lambda'}{\lambda'-(\lambda-i\epsilon)} G(\lambda') $$

$$ -\sum(\text{Residues of } [G(\lambda')/\lambda'-(\lambda-i\epsilon)] $$

$$ \text{at the poles of } G(\lambda')), \quad -2\pi < \arg\lambda' < 0. \tag{5a} $$

Equation (2c) is replaced by Eq. (5b).

$$ I_c = -\frac{1}{2\pi i} \left[\int_\infty^0 \frac{d\lambda'}{\lambda'-(\lambda-i\epsilon)} G(\lambda' e^{-2\pi i}) \right. $$

$$ \left. + \int_0^\infty \frac{d\lambda'}{\lambda'-(\lambda-i\epsilon)} G(\lambda') \right]. \tag{5b} $$

From the discussion following Eq. (2c), it is seen why $\lambda-i\epsilon$ is appropriate to this converging wave boundary condition. For the continuum part of the expansion, precisely as before, one derives that

$$ W_2(\lambda e^{-2\pi i}) = e^{i2\lambda^{\frac{1}{2}}R} e^{(2\pi c/\lambda^{\frac{1}{2}})} W_1(\lambda). $$

Therefore, I_c is given by Eq. (3c). The residues are obtained from the poles of $\Gamma[l+1+(ic/\lambda^{\frac{1}{2}})]$:

$$ l+1+(ic/\lambda^{\frac{1}{2}}) = -(n-l-1), \quad i\lambda^{\frac{1}{2}} = c/n. $$

This is consistent since $-\pi < \arg\lambda'^{\frac{1}{2}} < 0$. Several transformations are introduced, which relations are derived as under the diverging wave case.

$$ e^{i\lambda^{\frac{1}{2}}R} {}_1F_1(l+1-(ic/\lambda^{\frac{1}{2}}), 2l+2; -i2\lambda^{\frac{1}{2}}R) $$

$$ = e^{-i\lambda^{\frac{1}{2}}R} {}_1F_1(l+1+(ic/\lambda^{\frac{1}{2}}), 2l+2; i2\lambda^{\frac{1}{2}}R), $$

$$ W_2(l+1-(ic/\lambda^{\frac{1}{2}}), 2l+2; -i2\lambda^{\frac{1}{2}}R) $$

$$ = e^{-(2\pi c/\lambda^{\frac{1}{2}})} e^{-i2\lambda^{\frac{1}{2}}R} W_1(l+1+(ic/\lambda^{\frac{1}{2}}), 2l+2; i2\lambda^{\frac{1}{2}}R). $$

At the pole $\lambda = -c^2/n^2$, one obtains

$$ e^{-(2\pi c/\lambda^{\frac{1}{2}})} W_1(l+1+(ic/\lambda^{\frac{1}{2}}), 2l+2; i2\lambda^{\frac{1}{2}}R) $$

$$ = e^{-2\pi i n} {}_1F_1(-n+l+1, 2l+2; (2cR/n)). $$

Put $l+1+(ic/\lambda^{\frac{1}{2}}) = \mu$ to get

$$ d\lambda\lambda^{\frac{1}{2}} = \frac{i2c^3}{(l+1-\mu)^4} d\mu. $$

Thus, the residue at $\mu = -(n-l-1)$ leads to Eq. (4c), and this completes the calculation.

A remark is in order relevant to the Green's function for the repulsive Coulomb potential. This case is obtained by the replacement of c by $-c$, and it is easy to show that the residues of G vanish at the poles of G. Consequently, the spectrum of G is purely continuous for the repulsive potential.

There is an interesting application of Eq. (4d) for which only the continuum part of G contributes. The

[11] E. T. Copson, *An Introduction to the Theory of Functions of a Complex Variable* (Oxford University Press, London, 1957), p. 207.
[12] Reference 9, Vol. II, p. 1664.

solution of the scattering problem, $\Psi^{(\pm)}$, is given by

$$\Psi^{(\pm)}=\frac{\pm i\epsilon}{E\pm i\epsilon-H}\phi_i,$$

with ϕ_i the incident state $[H=H_0+V, (E-H_0)\phi_i=0, (E-H)\Psi=0]$, and the limit $\epsilon\to 0$ is taken after the calculation is performed.[13] However, one of the requirements for this method of calculation to give a unique solution is that the interaction potential V decrease faster than R^{-1} at infinity.[14] In spite of the fact that the V of this paper fails to meet this requirement, the quantity,

$$\lim_{\epsilon\to 0}\lim_{q\to K}\pm i\epsilon\int G^{(\pm)}(\mathbf{K};\mathbf{R},\mathbf{R}')d\mathbf{R}'e^{i\mathbf{q}\cdot\mathbf{R}'},$$

$$(\mathbf{q}q^{-1}=\mathbf{K}K^{-1})$$

does yield the correct coordinate dependence for the Coulomb scattering problem. As might be expected, the solution is not unique; not only is a divergent phase factor present, but the normalization depends upon the direction that q approaches K.[14,15]

[13] M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).

[14] S. Okubo and D. Feldman, Phys. Rev. **117**, 292 (1960).

[15] Robert A. Mapleton, J. Math. Phys. **2**, 482 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 4 JULY-AUGUST, 1961

Application of a Singular Wave-Function Operator in Scattering Theory

ROBERT A. MAPLETON
Geophysics Research Directorate, Air Force Research Division, Air Research and Development Command, L. G. Hanscom Field, Bedford, Massachusetts

(Received February 9, 1961)

A function which is defined in terms of the two-body singular wave-function matrix of Møller acting on a plane-wave state is calculated. In the paper by T. Pradhan, the function in question was obtained as the solution of a differential equation. Here, it is found that the function has a normalization which is different from the one assumed by Pradhan, and this behavior is attributed to the long-range character of the Coulomb potential that is in the Møller operator. An intermediate result of this paper agrees with the momentum-space solution of the nonrelativistic Coulomb two-body scattering problem which recently appeared in a paper published by S. Okubo and D. Feldman. As they show, the wave function must be renormalized to obtain the correct cross section, and this view is adopted in this paper. Consequently, the normalization assumed by Pradhan is reinstated. The Legendre expansion coefficients are calculated for the function,

$$(a-b\cos\theta)^{-c}.$$

IN a paper published by T. Pradhan,[1] an approximate method for calculating capture cross sections was developed, and this scheme was used to calculate the capture of electrons by protons passing through atomic hydrogen. As an intermediate step in P, it is necessary to calculate a wave function which is defined in terms of a Møller two-body operator acting on a plane wave. This was not done, but instead, appeal was made to a differential equation to obtain the wave function. If the wave function had been derived from the defining equation, it would have been found that the normalization was different from that assumed in P.

In the present paper, the Møller operator is used and the wave function is calculated in position space. In the course of the calculation, the quantity, as given by Eq. (A), is calculated.

$$F=\lim_{\epsilon\to 0} i\epsilon G(\epsilon,E)\phi_E. \qquad \text{(A)}$$

In Eq. (A), G is the Coulomb-Green's function, and ϕ_E is a plane wave. This part of the calculation is essentially the position-space equivalent of the momentum-space problem which was treated in the recent paper by S. Okubo and D. Feldman.[2] In OF, a Green's function is calculated in momentum space to obtain the wave function for the two-body, nonrelativistic, Coulomb-scattering problem. They show that the wave function must be renormalized in order to obtain the correct result for the cross section; moreover, they attribute this peculiar behavior to the long-range character of the Coulomb potential. The wave function obtained from Eq. (A) agrees with that of OF, and a transformation to momentum space shows that the two results must agree. With this information, it will become evident that some of the subsequent work could be replaced by the results of OF; nevertheless, in the author's opinion, these calculations in position space are a worthwhile complement to the momentum-space solutions of OF.

Some defining equations of P are reproduced here

[1] Trilochan Pradhan, Phys. Rev. **105**, 1250 (1957). (This paper will be called P.)

[2] S. Okubo and D. Feldman, Phys. Rev. **117**, 292 (1960). (This paper will be called OF.)

solution of the scattering problem, $\Psi^{(\pm)}$, is given by

$$\Psi^{(\pm)}=\frac{\pm i\epsilon}{E\pm i\epsilon-H}\phi_i,$$

with ϕ_i the incident state $[H=H_0+V, (E-H_0)\phi_i=0, (E-H)\Psi=0]$, and the limit $\epsilon\to 0$ is taken after the calculation is performed.[13] However, one of the requirements for this method of calculation to give a unique solution is that the interaction potential V decrease faster than R^{-1} at infinity.[14] In spite of the fact that the V of this paper fails to meet this requirement, the quantity,

$$\lim_{\epsilon\to 0}\lim_{q\to K}\pm i\epsilon\int G^{(\pm)}(\mathbf{K};\mathbf{R},\mathbf{R}')d\mathbf{R}'e^{i\mathbf{q}\cdot\mathbf{R}'},$$

$$(\mathbf{q}q^{-1}=\mathbf{K}K^{-1})$$

does yield the correct coordinate dependence for the Coulomb scattering problem. As might be expected, the solution is not unique; not only is a divergent phase factor present, but the normalization depends upon the direction that q approaches K.[14,15]

[13] M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).

[14] S. Okubo and D. Feldman, Phys. Rev. **117**, 292 (1960).

[15] Robert A. Mapleton, J. Math. Phys. **2**, 482 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 4 JULY-AUGUST, 1961

Application of a Singular Wave-Function Operator in Scattering Theory

ROBERT A. MAPLETON
Geophysics Research Directorate, Air Force Research Division, Air Research and Development Command, L. G. Hanscom Field, Bedford, Massachusetts

(Received February 9, 1961)

A function which is defined in terms of the two-body singular wave-function matrix of Møller acting on a plane-wave state is calculated. In the paper by T. Pradhan, the function in question was obtained as the solution of a differential equation. Here, it is found that the function has a normalization which is different from the one assumed by Pradhan, and this behavior is attributed to the long-range character of the Coulomb potential that is in the Møller operator. An intermediate result of this paper agrees with the momentum-space solution of the nonrelativistic Coulomb two-body scattering problem which recently appeared in a paper published by S. Okubo and D. Feldman. As they show, the wave function must be renormalized to obtain the correct cross section, and this view is adopted in this paper. Consequently, the normalization assumed by Pradhan is reinstated. The Legendre expansion coefficients are calculated for the function,

$$(a-b\cos\theta)^{-c}.$$

IN a paper published by T. Pradhan,[1] an approximate method for calculating capture cross sections was developed, and this scheme was used to calculate the capture of electrons by protons passing through atomic hydrogen. As an intermediate step in P, it is necessary to calculate a wave function which is defined in terms of a Møller two-body operator acting on a plane wave. This was not done, but instead, appeal was made to a differential equation to obtain the wave function. If the wave function had been derived from the defining equation, it would have been found that the normalization was different from that assumed in P.

In the present paper, the Møller operator is used and the wave function is calculated in position space. In the course of the calculation, the quantity, as given by Eq. (A), is calculated.

$$F=\lim_{\epsilon\to 0} i\epsilon G(\epsilon,E)\phi_E. \qquad \text{(A)}$$

In Eq. (A), G is the Coulomb-Green's function, and ϕ_E is a plane wave. This part of the calculation is essentially the position-space equivalent of the momentum-space problem which was treated in the recent paper by S. Okubo and D. Feldman.[2] In OF, a Green's function is calculated in momentum space to obtain the wave function for the two-body, nonrelativistic, Coulomb-scattering problem. They show that the wave function must be renormalized in order to obtain the correct result for the cross section; moreover, they attribute this peculiar behavior to the long-range character of the Coulomb potential. The wave function obtained from Eq. (A) agrees with that of OF, and a transformation to momentum space shows that the two results must agree. With this information, it will become evident that some of the subsequent work could be replaced by the results of OF; nevertheless, in the author's opinion, these calculations in position space are a worthwhile complement to the momentum-space solutions of OF.

Some defining equations of P are reproduced here

[1] Trilochan Pradhan, Phys. Rev. **105**, 1250 (1957). (This paper will be called P.)

[2] S. Okubo and D. Feldman, Phys. Rev. **117**, 292 (1960). (This paper will be called OF.)

for convenience. In position space, the center-of-mass, post coordinate system is used.[3] The cross section is obtained from the R matrix

$$R_{ba}=\langle\phi_b|\,U_{23}\,|\omega_{23}\mu_a\rangle, \tag{P 35}$$

$$\omega_{23}=1+\frac{1}{E_m-T-U_{23}+i\epsilon}U_{23}, \tag{P 23a}$$

$$\psi_m(23)=\chi_m+\frac{1}{E_m-T+i\epsilon}U_{23}\psi_m(23), \tag{P 24a}$$

$$\psi_m=N(K)\chi_m\,{}_1F_1\left(\frac{isa}{K},1;i[Kr_0-\mathbf{K}\cdot\mathbf{r}_0]\right), \tag{P 36d}$$

$$T\chi_m=E_m\chi_m,\quad (T+U_{23})\psi_m(23)=E_m\psi_m(23),$$

$$\mu_a=\textstyle\sum_l\chi_l\langle\chi_l|\mu_a\rangle,$$

$$\omega_{23}\mu_a=\textstyle\sum_l\omega_{23}\chi_l\langle\chi_l|\mu_a\rangle=\sum_l\psi_l(23)\langle\chi_l|\mu_a\rangle. \tag{P 37}$$

It will be shown that the asymptotic form of Eq. (P24a) is different from the asymptotic form of the assumed solution of Eq. (P24a) which is given by Eq. (P36d). In this paper, the function in question, ψ_m, is calculated from an expression which is equivalent to the relation which defines ψ_m. This is given by Eq. (B).

$$\psi_m=\omega_{23}\chi_m=\frac{i\epsilon}{E_m-T-U_{23}+i\epsilon}\chi_m. \tag{B}$$

This prescription for the calculation of ψ_m avoids the objectionable operator manipulation which is required in the derivation of Eq. (P24a). The operator acting on χ_m in Eq. (B), sometimes called the singular wave-function matrix of Møller, is discussed elsewhere.[4] The calculation of ψ_m is the main task of this paper.

Some change in notation is employed for convenience, and Fig. 2 of P should be consulted. m=electron mass,

$$M=\text{proton mass},\quad M_1=M+m,\quad a=\frac{M}{M_1},\quad \mu=\frac{MM_1}{M+M_1},$$

$$m_f=am\longrightarrow m,\quad \mathbf{x}=\mathbf{r}-a\mathbf{r}_0,\quad \mathbf{K}_{23,1}=\mathbf{K}_1,\quad \mathbf{K}=\mathbf{K},$$

$$\chi_l=(2\pi)^{-\frac{3}{2}}e^{i(-\mathbf{K}_1\cdot\mathbf{x}+\mathbf{K}\cdot\mathbf{r}_0)},\ \mathbf{k}_0=\mathbf{K}_a,$$

$$\mu_a=(2\pi)^{-\frac{3}{2}}\phi_0(\mathbf{r})e^{i\mathbf{K}_a\cdot(a\mathbf{r}-\mathbf{r}_0)},$$

$$\phi_b{}^*=(2\pi)^{-\frac{3}{2}}\phi_0(\mathbf{r}_0)e^{-i\mathbf{K}_b\cdot\mathbf{x}},\quad \mathbf{k}_s=\mathbf{K}_b,\quad \alpha=a/Ka_0,$$

$$a_0{}^{-1}=me^2/\hbar^2,\ N(K)=\Gamma(1-i\alpha)e^{\frac{1}{2}\pi\alpha},$$

$$\psi_m=N(K)\chi_m\,{}_1F_1(i\alpha,1;i[Kr_0-\mathbf{K}\cdot\mathbf{r}_0]),$$

$$T=-(\hbar^2/2\mu)\nabla_x{}^2-(\hbar^2/2m)\nabla_{r_0}{}^2,\quad U=U_{23}=-e^2/r_0,$$

$$(\hbar^2/2\mu)K_1{}^2+(\hbar^2/2m)K^2=(\hbar^2K_T{}^2/2\mu)=E_m.$$

[3] J. D. Jackson and H. Schiff, Phys. Rev. **89**, 359 (1953).
[4] M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).

Equation (P24a) is now written in position space. This is accomplished with the use of the Green's function[5] of Eq. (1a) which satisfies the differential equation (1b).

$$G=(2\pi)^{-6}\int\frac{d\mathbf{U}d\mathbf{V}\exp\{i[\mathbf{U}\cdot(\mathbf{x}-\mathbf{x}')+\mathbf{V}\cdot(\mathbf{r}_0-\mathbf{r}_0')]\}}{K_T{}^2-U^2-\mu V^2/m+i\epsilon}, \tag{1a}$$

$$(\nabla_x{}^2+(\mu/m)\nabla_{r_0}{}^2+K_T{}^2+i\epsilon)G=\delta(\mathbf{x}-\mathbf{x}')\delta(\mathbf{r}_0-\mathbf{r}_0'). \tag{1b}$$

Integrate over $\mathbf{U}$ and $\mathbf{x}'$, and use the standard form,

$$g=(2\pi)^{-3}\int\frac{d\mathbf{U}\exp[i\mathbf{U}\cdot(\mathbf{r}-\mathbf{r}')]}{K^2+i\epsilon-U^2}=-\frac{\exp[iK|\mathbf{r}-\mathbf{r}'|]}{4\pi|\mathbf{r}-\mathbf{r}'|},$$

to reduce Eq. (P24a) to the form given in Eq. (1c).

$$\psi_m=e^{-i\mathbf{K}_1\cdot\mathbf{x}}\left[e^{i\mathbf{K}\cdot\mathbf{r}_0}+\frac{N(K)}{2\pi a_0}\int g(\mathbf{r}_0,\mathbf{r})\frac{d\mathbf{r}}{r}e^{i\mathbf{K}\cdot\mathbf{r}}\times{}_1F_1(i\alpha,1;i[Kr-\mathbf{K}\cdot\mathbf{r}])\right]. \tag{1c}$$

The plane wave factor in K_1 which is common to both sides of Eq. (1c) is removed. The resulting integral equation for the Coulomb function has the incorrect asymptotic form, and in accordance with the remarks in OF relevant to Eq. (OF35), this equation has no solution. This result is an example of the errors that may incur in operator manipulations with a single Coulomb potential. (If there were two Coulomb potentials, each with equal and opposite charge, the resultant potential at large distances would decrease faster than the Coulomb potential.)

The calculation of ψ_m as given by Eq. (B) is presented. (The subscript on r_0 is omitted.) G satisfies the differential equation (2a) and is given by Eq. (2b).[6]

$$\left[\nabla_x{}^2+\frac{2\mu}{\hbar^2}\left(\frac{\hbar^2}{2m}\nabla_r{}^2+\frac{e^2}{r}\right)+K_T{}^2+i\epsilon'\right]G=\delta(\mathbf{x}-\mathbf{x}')\delta(\mathbf{r}-\mathbf{r}'), \tag{2a}$$

$$G=(2\pi)^{-3}\sum_n\int\frac{\phi_n(\mathbf{r})\phi_n{}^*(\mathbf{r}')d\mathbf{U}\exp[i\mathbf{U}\cdot(\mathbf{x}-\mathbf{x}')]}{K_T{}^2+(2\mu/\hbar^2)\epsilon_n-U^2+i\epsilon'}, \tag{2b}$$

($\epsilon'=2\mu\epsilon/\hbar^2$, ϵ_n=binding energy of hydrogen). In Eq.

[5] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. II, p. 1733.
[6] B. A. Lippman and Julian Schwinger, Phys. Rev. **79**, 469 (1950). (See Eq. 2.32.)

(2b), the integration over $\mathbf{U}$ and $\mathbf{x}'$ is performed to get Eq. (2c).

$$\psi_m=(2\pi)^{-\frac{1}{2}}e^{-i\mathbf{K}_1\cdot\mathbf{x}}F, \quad 2m\epsilon'/\hbar^2=\epsilon,$$
$$F=i\epsilon\sum_n\frac{\phi_n(\mathbf{r})\phi_n^*(\mathbf{r}')d\mathbf{r}'}{K^2+(2m\epsilon_n/\hbar^2)+i\epsilon}e^{i\mathbf{K}\cdot\mathbf{r}'}. \tag{2c}$$

In Eqs. (2b) and (2c), ϕ_n are the normalized hydrogen wave functions with the continuum part of the set normalized to the delta function.[7]

With the normalized plane wave factor removed from the right side of Eq. (2c), the quantity F can be written as

$$F=i\epsilon\int G_K(\mathbf{r},\mathbf{r}')d\mathbf{r}'e^{i\mathbf{K}\cdot\mathbf{r}'}.$$

In this expression, G_K is the complex Coulomb-Green's function. In Appendix I, it will be shown that an equivalent expression for F is

$$F=\int[\delta(\mathbf{r}-\mathbf{r}')+G_KV]d\mathbf{r}'e^{i\mathbf{K}\cdot\mathbf{r}'}, \quad (V=-2/a_0r'),$$

provided that a plane wave can be expanded in the hydrogenic set. This equivalence is commonly displayed, symbolically, as $i\epsilon G_E\phi_E=(1+G_EV)\phi_E$. As mentioned by others,[4] only the continuum part of the eigenfunction expansion for G_K contributes to F by reason of the presence of the factor ϵ. This fact is used, and the identity operator,

$$-\lim_{\lambda\to0}\frac{\partial}{\partial\lambda}\frac{e^{-\lambda r'}}{r'},$$

is introduced to facilitate the integrations. Thus, the new form for F is given below.

$$F=-\lim_{\epsilon\to0}\lim_{\lambda\to0}i\epsilon\frac{\partial}{\partial\lambda}\int\frac{d\mathbf{k}\phi_k(\mathbf{r})\phi_k^*(\mathbf{r}')e^{(-\lambda r'+i\mathbf{K}\cdot\mathbf{r}')}}{(K^2+i\epsilon-k^2)r'}.$$

The operator

$$-\lim_{\lambda\to0}\frac{\partial}{\partial\lambda}$$

is omitted until later in the calculation. A number of relevant quantities are defined below.[8]

$$\phi_k(\mathbf{r})=N(k)L_l(kr)P_l(\cos\Theta), \quad \alpha=(ka_0)^{-1},$$
$$N(k)=\frac{1}{2\pi\Gamma(1-i\alpha)}\left[\frac{1}{1-e^{-2\pi\alpha}}\right]^{\frac{1}{2}},$$
$$L_l(kr)=\frac{\Gamma(l+1-i\alpha)}{(2l)!}(i2kr)^le^{ikr}\times{}_1F_1(l+1-i\alpha,\,2l+2;\,-i2kr),$$
$$P_l(\cos\Theta)=\sum_{m=-l}^{l}N_{lm}P_l^{|m|}(\cos\theta_r)\times P_l^{|m|}(\cos\theta_k)e^{im(\phi_r-\phi_k)},$$
$$N_{lm}=\frac{(l-|m|)!}{(l+|m|)!},$$
$$\phi_k^*(\mathbf{r}')=[N(k)L_l(kr')P_l(\cos\Theta')]^*,$$
$$P_l(\cos\Theta')=\sum_{m=-l}^{l}N_{lm}P_l^{|m|}(\cos\theta_k)\times P_l^{|m|}(\cos\theta_{r'})e^{im(\phi_k-\phi_{r'})},$$
$$P_l(\cos\Theta_1)=\sum_{m=-l}^{l}N_{lm}P_l^{|m|}(\cos\theta_k)\times P_l^{|m|}(\cos\theta_K)e^{im(\phi_k-\phi_K)},$$
$$e^{i\mathbf{K}\cdot\mathbf{r}'}=\sum_{n=0}^{\infty}\sum_{p=-n}^{n}(2n+1)i^nN_{np}P_n^{|p|}(\cos\theta_{r'})\times P_n^{|p|}(\cos\theta_K)e^{ip(\phi_{r'}-\phi_K)}(\pi/2Kr')^{\frac{1}{2}}J_{n+\frac{1}{2}}(Kr').$$

The integration over the solid angle of $\mathbf{r}'$ space is performed to give Eq. (3a).

$$F=4\pi i\epsilon\int\frac{d\mathbf{k}|N(k)|^2}{K^2+i\epsilon-k^2}\sum_{l=0}^{\infty}L_l(kr)P_l(\cos\Theta)P_l(\cos\Theta_1)A_l, \tag{3a}$$
$$A_l=i^l\int_0^\infty dr'r'e^{-\lambda r'}L_l^*(kr')(\pi/2Kr')^\lambda J_{l+\frac{1}{2}}(Kr').$$

With the aid of an integral representation and a standard integral the r' integration can be performed.[9]

$${}_1F_1(l+1+i\alpha,\,2l+2;\,i2kr')=\frac{2^{-(2l+1)}(2l+1)!e^{ikr'}}{\Gamma(l+1-i\alpha)\Gamma(l+1+i\alpha)}\int_{-1}^{1}(1-t)^{l-i\alpha}\times(1+t)^{l+i\alpha}dte^{ikr't},$$
$$\frac{(2b)^\nu\Gamma(\nu+\frac{1}{2})}{\pi^{\frac{1}{2}}(a^2+b^2)^{\nu+\frac{1}{2}}}=\int_0^\infty r^\nu J_\nu(br)dre^{-ar}, \quad \nu=l+\tfrac{1}{2}, \quad a=\lambda-ikt, \quad b=K.$$

From these two relations, A_l becomes

$$A_l=\frac{(-1)^{l+1}(K/k)^l(2l+1)l!}{2k^2\Gamma(l+1-i\alpha)}\int_{-1}^{1}\frac{dt(1-t)^{l-i\alpha}(1+t)^{l+i\alpha}}{[(t-\beta_1)(t-\beta_2)]^{l+1}},$$

in which expression the identity,

$$[K^2+(\lambda-ikt)^2]^{-(l+1)}=[-k^2(t-\beta_1)(t-\beta_2)]^{-(l+1)},$$
$$\beta_{1,2}=-i\lambda/k+(-,+)K/k,$$

has been used. In order to evaluate A_l, a closed contour

[7] A. Sommerfeld, Ann. Physik **11**, 257 (1931).
[8] W. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford University Press, New York, 1949), 2nd ed.
[9] W. Magnus and F. Oberhettinger, *Special Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1949), pp. 33, 88.

in the t plane is used that excludes the branch cut, but which includes the poles: branch cut, $-1<t<1$; poles, $t=\beta_{1,2}$.

The phase of the quantity T is given below.

$$T=\left(\frac{1-t}{1+t}\right)^{-i\alpha}, \quad \arg T=\begin{cases}-2\pi, & \text{below cut}\\ 0, & \text{above cut.}\end{cases}$$

The contour consists of a circle at infinity which connects the real axis interval

$$-1<t<\infty$$

above the cut with

$$\infty>t>-1$$

below the cut. Small semi-circular arcs join the regions

$$t<1, \quad t>1$$

above and below the cut, and the contour is completed with a small circle about

$$t=-1.$$

With the contour traversed in the counter-clockwise direction, only the integrals on

$$-1<t<1$$

remain, and the Cauchy residue theorem leads to Eq. (3b) for A_l.

$$A_l=\frac{\pi i(-1)^{l+1}(2l+1)(K/k)^l}{k^2\Gamma(l+1-i\alpha)(1-e^{-2\pi\alpha})}\times\frac{d^l}{dt^l}\left[\frac{(1-t)^{l-i\alpha}(1+t)^{l+i\alpha}}{(t-\beta_{1,2})^{l+1}}\right]_{t=\beta_{2,1}}. \tag{3b}$$

The Leibnitz' rule for the successive differentiation of a product is applied and a standard formula is used.[10]

$$\frac{d^l}{dt^l}(fg)=\sum_{p=0}^{l}\frac{l!}{(l-p)!p!}\left(\frac{d^{(l-p)}}{dt^{(l-p)}}f\right)\frac{d^p}{dt^p}g,$$

$$f=(t-\beta)^{-(l+1)}, \quad g=(1-t)^{l-i\alpha}(1+t)^{l+i\alpha},$$

$$\frac{d^{(l-p)}}{dt^{(l-p)}}f=\frac{(-1)^{l-p}(2l-p)!}{l!(t-\beta)^{(2l+1-p)}},$$

$$\frac{d^p}{dt^p}g=(-2)^p(1-t)^{(l-p-i\alpha)}(1+t)^{(l-p+i\alpha)}\times\frac{\Gamma(l+1-i\alpha)}{\Gamma(l+1-p-i\alpha)}{}_2F_1\left(-p, 2l-p+1, l-p+1-i\alpha; \frac{1-t}{2}\right).$$

These forms are introduced into Eq. (3b), the values for $t=\beta_{1,2}$ are inserted, and some cancellation is effected to get Eq. (3c).

$$A_l=-\frac{2\pi i(2l+1)}{(4kK)^{l+1}(1-e^{-2\pi\alpha})}\sum_{p=0}^{l}\frac{(2l-p)!}{(l-p)!p!\Gamma(l-p+1-i\alpha)}\Bigg[(k+i\lambda-K)^{l-i\alpha}(k-i\lambda+K)^{l+i\alpha}\left(\frac{4kK}{k^2-[K-i\lambda]^2}\right)^p\times{}_2F_1\left(-p, 2l-p+1, l-p+1-i\alpha; \frac{k+i\lambda-K}{2k}\right)-(-1)^p(k+i\lambda+K)^{l-i\alpha}(k-i\lambda-K)^{l+i\alpha}\left(\frac{4kK}{k^2-[K+i\lambda]^2}\right)^p\times{}_2F_1\left(-p, 2l-p+1, l-p+1-i\alpha; \frac{k+i\lambda+K}{2k}\right)\Bigg]. \tag{3c}$$

Several definitions are introduced and identities noted.

$$A=\frac{(k+i\lambda-K)(k-i\lambda+K)}{4kK}=\frac{k^2-(K-i\lambda)^2}{4kK},$$

$$B=\frac{k+i\lambda-K}{2k}, \quad A-B=(k-K)^2+\lambda^2,$$

$$W_{\mp}=(k\mp K)^2+\lambda^2.$$

The series in Eq. (3c) are summable, and for this purpose all terms common to the factor

$$[\Gamma(l-N+1-i\alpha)]^{-1}$$

are found, which terms come from the following values of p:

$$p=N, N+1, \cdots l.$$

This group of terms leads to the series below which is summed.

$$\frac{(2l-N)!}{(l-N)!N!\Gamma(l-N+1-i\alpha)}\left[A^{l-N}-\frac{(l-N)}{1!}A^{(l-N-1)}B+\frac{(l-N)(l-N-1)}{2!}A^{(l-N-2)}B^2+\cdots+(-1)^{l-N}B^{l-N}\right]=\frac{(2l-N)!}{(l-N)!N!\Gamma(l-N+1-i\alpha)}(A-B)^{l-N}.$$

One now sums over the allowable values of N to obtain the sum of the original series.

$$\sum_{N=0}^{l}\frac{(2l-N)!}{(l-N)!N!\Gamma(l-N+1-i\alpha)}(A-B)^{l-N}$$

$$=\frac{1}{\Gamma(1-i\alpha)}\,{}_2F_1\left(-l,l+1,1-i\alpha;-\frac{W_-}{4kK}\right).$$

(This is proven in Appendix II by the use of an expansion theorem.)

The second series of Eq. (3c) is treated in a similar fashion, so that A_l can be written as in Eq. (3d).

$$A_l=-\frac{(2l+1)\pi i}{2kK\Gamma(1-i\alpha)(1-e^{-2\pi\alpha})}\left[\left(\frac{k+K-i\lambda}{k-K+i\lambda}\right)^{i\alpha}\right.$$

$$\times{}_2F_1\left(-l,l+1,1-i\alpha;-\frac{W_-}{4kK}\right)$$

$$-(-1)^l\left(\frac{k-K-i\lambda}{k+K+i\lambda}\right)^{i\alpha}$$

$$\left.\times{}_2F_1\left(-l,l+1,1-i\alpha;\frac{W_+}{4kK}\right)\right]. \quad (3d)$$

Several identities are noted, and the phase of the quantity T^{-1} (defined previously) is examined at the poles, $t=\beta_{1,2}$.

$$T^{-1}(\beta_2)=\left(\frac{k+K-i\lambda}{k-K+i\lambda}\right)^{i\alpha}=\left(\frac{M}{W_-}\right)^{i\alpha},$$

$$T^{-1}(\beta_1)=\left(\frac{k-K-i\lambda}{k+K+i\lambda}\right)^{i\alpha}=\left(\frac{M}{W_+}\right)^{i\alpha},$$

$$M=k^2-K^2-\lambda^2-i2\lambda k.$$

One should first observe that $W_\pm$ do not enter into the determination of the phase of T^{-1} since they are non-negative. The second point to note is that the poles are in the lower half of the t plane since λ is non-negative. With this information and the previous discussion of the phrase of T, it follows that the limiting value of the phase of M is given by

$$\lim_{\lambda\to 0}\arg M=\begin{cases}\pi, & k<K,\ |t|>1\\ 2\pi, & k>K,\ |t|<1\end{cases}.$$

F of Eq. (3a) is given by Eq. (3e).

$$F=2\pi^2\epsilon\int\frac{d\mathbf{k}|N(k)|^2}{K^2+i\epsilon-k^2}\sum_{l=0}^{\infty}L_l(kr)P_l(\cos\Theta)$$

$$\times P_l(\cos\Theta_1)\frac{(2l+1)M^{i\alpha}}{kK\Gamma(1-i\alpha)(1-e^{-2\pi\alpha})}$$

$$\times\left[W_-^{-i\alpha}\,{}_2F_1\left(-l,l+1,1-i\alpha;-\frac{W_-}{4kK}\right)\right.$$

$$\left.-(-1)^lW_+^{-i\alpha}\,{}_2F_1\left(-l,l+1,1-i\alpha;\frac{W_+}{4kK}\right)\right]. \quad (3e)$$

Use is made of the relation[9]

$$\Gamma(i\alpha)\Gamma(1-i\alpha)=\pi/\sin(\pi i\alpha)$$

and integration over the solid angle of $\mathbf{k}$ space is effected to obtain Eq. (4a) from Eq. (3e).

$$P_l(\cos\Theta)$$

$$=\sum_{m=-l}^{l}N_{lm}P_l^{|m|}(\cos\theta_r)P_l^{|m|}(\cos\theta_K)e^{im(\phi_r-\phi_K)},$$

$$F=\frac{i\epsilon}{2\pi K}\int_0^\infty\frac{dkk\Gamma(i\alpha)M^{i\alpha}e^{2\pi\alpha}}{K^2+i\epsilon-k^2}\sum_{l=0}^{\infty}L_l(kr)P_l(\cos\Theta)$$

$$\times\left[W_-^{-i\alpha}\,{}_2F_1\left(-l,l+1,1-i\alpha;-\frac{W_-}{4kK}\right)\right.$$

$$\left.-(-1)^lW_+^{-i\alpha}\,{}_2F_1\left(-l,l+1,1-i\alpha;\frac{W_+}{4kK}\right)\right]. \quad (4a)$$

The operator

$$-\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}$$

is applied to the relevant factor of Eq. (4a) to give

$$-\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}M^{i\alpha}=-\frac{2k\alpha(k^2-K^2)^{i\alpha}}{k^2-K^2}.$$

The resulting integral of Eq. (4a) can be made well-defined by the substitution

$$K=q,\quad q\lesseqgtr K$$

in all terms exclusive of the factor

$$(K^2+i\epsilon-k^2)^{-1}.$$

If the substitution

$$\exp(i\mathbf{K}\cdot\mathbf{r}')=\exp[i(q\mathbf{K}/K\cdot\mathbf{r}')]$$

is made, this result is obtained from Eq. (3a). After the integration is accomplished, the limit

$$q\to K$$

will be performed. For definiteness, q is chosen so that

$$q<K.$$

It is noted that the principal part of this integral is zero in the limit,

$$\epsilon\rightarrow 0.$$

This fact is exploited by taking ϵ sufficiently small. A closed contour is selected in the k plane that consists of a small segment of the k axis which is closed in the upper half plane so that the pole

$$k=(K^2+i\epsilon)^{\frac{1}{2}}$$

is included, but the branch point

$$k=q$$

is excluded. At this pole,

$$i\alpha=i/ka_0\approx(i/Ka_0)+(\epsilon/2K^3a_0),$$

and

$$\text{real part } (i\alpha)>0.$$

In Eq. (4a), the first series of terms has the factor

$$(k^2-q^2)^{i\alpha}W_-^{-i\alpha}=\left|\frac{(k+q)^2}{k^2-q^2}\right|^{i\alpha}e^{-2\pi\alpha},$$

which dominates the factor

$$(k^2-q^2)^{i\alpha}W_+^{-i\alpha}=\left|\frac{k^2-q^2}{(k+q)^2}\right|^{i\alpha}e^{-2\pi\alpha}$$

of the second series of terms for q near enough to k. Accordingly, the second term is set equal to zero. The contour integration is performed (counterclockwise sense); the limit is taken in the sequence

$$\lim_{\epsilon\rightarrow 0}\lim_{q\rightarrow K},$$

and the identity

$${}_2F_1(-l,l+1,1-i\alpha;0)=1$$

is used, all of which reduces Eq. (4a) to Eq. (4b).

$$F=i\alpha\Gamma(i\alpha)\left(\frac{4}{-\beta}\right)^{i\alpha}\sum_{l=0}^{\infty}L_l(Kr)P_l(\cos\Theta)$$

$$=\frac{\pi\alpha e^{i\mathbf{K}\cdot\mathbf{r}}}{\sinh(\pi\alpha)}\left(\frac{4}{-\beta}\right)^{i\alpha}{}_1F_1(i\alpha,1;i[Kr-\mathbf{K}\cdot\mathbf{r}]),$$

$$\alpha=(Ka_0)^{-1},\quad \beta=1-q^2/K^2. \tag{4b}$$

This is precisely Eq. (OF52). If the preceding integrations are effected for

$$K<q,$$

the result is obtained from the relation

$$\left(\frac{4}{-\beta}\right)^{i\alpha}\rightarrow\left(\frac{4}{-\beta'e^{-\pi i}}\right)^{i\alpha}=\left(\frac{4}{-\beta'}\right)^{i\alpha}e^{\pi\alpha},$$

$$\beta'=1-K^2/q^2.$$

This same result is given by Eq. (OF56). The function ψ_m is obtained from Eq. (2c), and it is seen that the normalization is different from the one used in P. From the relation, Eq. (P37),

$$\omega_{23}\mu_a=\textstyle\sum_m\psi_m\langle\chi_m|\mu_a\rangle,$$

it is easily established that

$$\omega_{23}\mu_a=\left(\frac{8a_0^3}{\pi^2a^6}\right)^{\frac{1}{2}}\int\frac{d\mathbf{K}F(q)\exp[i(\mathbf{K}_a+\mathbf{K})\cdot\mathbf{x}]/a}{[1+a_0^2|a\mathbf{K}_a-a^{-1}(\mathbf{K}_a+\mathbf{K})|^2]^2},$$

$$F(q)={}_1F_1(i\alpha,1;i[Kr_0-\mathbf{K}\cdot\mathbf{r}_0])\frac{\pi\alpha e^{i\mathbf{K}\cdot\mathbf{r}_0}}{\sinh(\pi\alpha)}$$

$$\times\begin{cases}(4/\beta)^{i\alpha}, & q<K,\\ (4/\beta')^{i\alpha}e^{\pi\alpha}, & q>K,\end{cases}\qquad\alpha=\frac{a}{Ka_0}.$$

With this wave function, R_{ba} of Eq. (P35) is calculated to be

$$R_{ba}\approx-\frac{32\pi^2a_0^2e^2}{[1+p^2a_0^2]^3}\exp\left[\frac{2}{pa_0}\tan^{-1}\left(\frac{1}{pa_0}\right)-\frac{\pi}{pa_0}\right]$$

$$\times\frac{\pi\alpha}{\sinh(\pi\alpha)}\begin{cases}(4/\beta)^{i\alpha}, & q<p\\ (4/\beta')^{i\alpha}e^{\pi\alpha}, & q>p\end{cases}.$$

The same approximations used in P,

$$\mathbf{p}=a\mathbf{K}_b-\mathbf{K}_a\approx a\mathbf{K}_a-\mathbf{K}_b,\quad\alpha\approx(pa_0)^{-1},$$

were used to calculate the preceding expression for R_{ba}. R_{ba} is brought into agreement with (P43) with the following renormalization scheme which is the same as in OF.

$$\frac{\sinh(\pi\alpha)}{\pi\alpha}\Gamma(1-i\alpha)e^{\frac{1}{2}\pi\alpha}\begin{cases}(4/\beta)^{-i\alpha}, & q<p\\ (4/\beta')^{-i\alpha}e^{-\pi\alpha}, & q>p\end{cases}.$$

Since α contains the scattering angle, this is a nontrivial renormalization. From the considerations of the analogous problem discussed in OF, it appears that the normalization adopted in P is the correct one.

It is of interest to repeat this calculation for the case that ϵ is replaced by $-\epsilon$. In this case, the contour in the k plane encloses the pole

$$k=(K^2-i\epsilon)^{\frac{1}{2}}$$

in the lower half plane. With the use of the preceding arguments it is seen that the second term in the bracket of Eq. (4a) dominates. Several identities are invoked so that the series may be recognized.[9,10]

$$(-1)^l{}_2F_1\left(-l,l+1,1-i\alpha;\frac{W_+}{4kK}\right)$$

$$=\frac{\Gamma(l+1+i\alpha)\Gamma(1-i\alpha)}{\Gamma(l+1-i\alpha)\Gamma(1+i\alpha)}{}_2F_1\left(-l,l+1,1+i\alpha;-\frac{W_-}{4kK}\right),$$

[10] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. II, pp. 169–170, Eqs. 10 and 16.

$${}_1F_1(l+1-i\alpha,\, 2l+2;\, -i2kr)e^{ikr}$$
$$={}_1F_1(l+1+i\alpha,\, 2l+2;\, i2kr)e^{-ikr},$$
$$(-1)^l P_l(\cos\theta)=P_l(\cos[\pi-\theta]).$$

The results of the calculation are given by Eq. (4c) for the case that

$$q<K.$$

$$F'=\frac{i\alpha\Gamma(i\alpha)\Gamma(1-i\alpha)}{\Gamma(1+i\alpha)}\left(\frac{4}{\beta}\right)^{-i\alpha}\sum_{l=0}^{\infty}L_l^*(Kr)P_l(\cos[\pi-\Theta])$$
$$=\frac{\pi\alpha}{\sinh(\pi\alpha)}\left(\frac{4}{\beta}\right)^{-i\alpha}\exp[-iKr\cos(\pi-\Theta)]$$
$$\times{}_1F_1(-i\alpha,\, 1;\, -iKr[1-\cos(\pi-\Theta)]). \quad (4c)$$

This is the familiar converging wave solution as it should be.

The correspondence of the position-space calculations of this paper with the momentum-space calculations is now established. The pertinent equations of OF are included for reference.

$$G'(\mathbf{p},\mathbf{q})=\delta(\mathbf{p}-\mathbf{q})-\frac{U_0}{2\pi^2(k^2+i\epsilon-p^2)}\int\frac{d\mathbf{p}'G'(\mathbf{p}',\mathbf{q})}{(\mathbf{p}-\mathbf{p}')^2},$$
$$(U_0=2/a_0),\quad (k^2=K^2,\ \mathbf{r}_0=\mathbf{x}). \quad \text{(OF 37)}$$

$$G'(\mathbf{p},\mathbf{q})$$
$$=\delta(\mathbf{p}-\mathbf{q})+\int_0^{\infty}\frac{dvH(v)}{[(p^2-k^2-i\epsilon)v+(\mathbf{p}-\mathbf{q})^2]^2}. \quad \text{(OF 38)}$$

$H(v)$ is given by (OF42).

$$\psi(\mathbf{x})=F(\mathbf{x})=\lim_{q\to k-0}\int d\mathbf{p}G'(\mathbf{p},\mathbf{q})e^{i\mathbf{p}\cdot\mathbf{x}}$$
$$=\frac{\pi\alpha}{\sinh(\pi\alpha)}\left(\frac{4}{\beta}\right)^{i\alpha}e^{i\mathbf{k}\cdot\mathbf{x}}\,{}_1F_1(i\alpha,\, 1;\, i[kx-\mathbf{k}\cdot\mathbf{x}]). \quad \text{(OF 52)}$$

The Coulomb-Green's function of this paper satisfies the integral equation

$$G(\mathbf{p},\mathbf{q})=\frac{1}{(k^2+i\epsilon-p^2)}\left[\delta(\mathbf{p}-\mathbf{q})-\frac{U_0}{2\pi^2}\int\frac{d\mathbf{p}'G(\mathbf{p}',\mathbf{q})}{(\mathbf{p}-\mathbf{p}')^2}\right],$$

with the solution

$$G(\mathbf{p},\mathbf{q})=[k^2+i\epsilon-q^2]^{-1}G'(\mathbf{p},\mathbf{q}).$$

Therefore, F of this paper is given by

$$F=\lim_{\epsilon\to 0}\lim_{q\to k-0}\frac{i\epsilon}{k^2+i\epsilon-q^2}\int d\mathbf{p}G'(\mathbf{p},\mathbf{q})e^{i\mathbf{p}\cdot\mathbf{x}},$$

which is seen to be identical to the solution in OF.

APPENDIX I

It is of interest to see what assumption is implicit in the relation

$$\int\left[\delta(\mathbf{r}-\mathbf{r}')\delta(\mathbf{x}-\mathbf{x}')+\frac{2\mu}{\hbar^2}GV(r')\right]\chi_m d\mathbf{r}'d\mathbf{x}'$$
$$=\frac{2\mu}{\hbar^2}i\epsilon\int Gd\mathbf{x}'d\mathbf{r}'\chi_m,$$

with G given by Eq. (2b). Integrate over $\mathbf{U}$ and $\mathbf{x}'$, [see Eq. (2c)], to reduce the above relation to

$$\int\left[\delta(\mathbf{r}-\mathbf{r}')+\sum_n\frac{\phi_n(\mathbf{r})\phi_n^*(\mathbf{r}')V(r')}{K^2+i\epsilon+(2m\epsilon_n/\hbar^2)}\right]d\mathbf{r}'e^{i\mathbf{K}\cdot\mathbf{r}'}$$
$$=i\epsilon\int\sum_n\frac{e^{i\mathbf{K}\cdot\mathbf{r}'}\phi_n(\mathbf{r})\phi_n^*(\mathbf{r}')d\mathbf{r}'}{K^2+i\epsilon+(2m\epsilon_n/\hbar^2)},$$

the common plane wave factor in K_1 having been removed. If the expansion theorem

$$\delta(\mathbf{r}-\mathbf{r}')=\textstyle\sum_n\phi_n(\mathbf{r})\phi_n^*(\mathbf{r}')$$

is valid for the function

$$e^{i\mathbf{K}\cdot\mathbf{r}},$$

one obtains

$$e^{i\mathbf{K}\cdot\mathbf{r}}=\sum_n\phi_n(\mathbf{r})\int\phi_n^*(\mathbf{r}')d\mathbf{r}'e^{i\mathbf{K}\cdot\mathbf{r}'}.$$

Operate on both sides of this relation with the operator

$$\nabla^2+K^2+i\epsilon,$$

and use the identity

$$(\nabla^2+K^2+i\epsilon)\phi_n=(K^2+i\epsilon+2m\epsilon_n/\hbar^2-U)\phi_n$$

to obtain the equation

$$(U+i\epsilon)e^{i\mathbf{K}\cdot\mathbf{r}}$$
$$=\sum_n[2m\epsilon_n/\hbar^2+K^2+i\epsilon]\phi_n(\mathbf{r})\int\phi_n^*(\mathbf{r}')d\mathbf{r}'e^{i\mathbf{K}\cdot\mathbf{r}'},$$
$$(U=-V=2/a_0r).$$

Multiply this last equation with

$$\phi_n(\mathbf{r}),$$

integrate over $\mathbf{r}$, and divide the result by

$$(K^2+i\epsilon+2m\epsilon_n/\hbar^2)$$

to get the relation

$$\int\phi_n^*(\mathbf{r})d\mathbf{r}e^{i\mathbf{K}\cdot\mathbf{r}}$$
$$=(K^2+i\epsilon+2m\epsilon_n/\hbar^2)^{-1}\int\phi_n^*(\mathbf{r})(U+i\epsilon)e^{i\mathbf{K}\cdot\mathbf{r}}d\mathbf{r}.$$

Multiply both sides of this expression by

$$\phi_n(\mathbf{r}),$$

sum over n, and rearrange terms to obtain the result

$$\int\left[\delta(\mathbf{r}-\mathbf{r}')+\sum_n \frac{\phi_n(\mathbf{r})\phi_n^*(\mathbf{r}')V(r')}{K^2+i\epsilon+2m\epsilon_n/\hbar^2}\right]d\mathbf{r}'e^{i\mathbf{K}\cdot\mathbf{r}'}$$

$$=i\epsilon\int\sum_n \frac{\phi_n(\mathbf{r})\phi_n^*(\mathbf{r}')d\mathbf{r}'e^{i\mathbf{K}\cdot\mathbf{r}'}}{K^2+i\epsilon+2m\epsilon_n/\hbar^2}.$$

Consequently, the results of this paper depend upon the validity of the expansion of a plane wave in the hydrogen set.

APPENDIX II

In this appendix a useful integral is evaluated. The result serves a check on certain calculations performed in this paper. The expansion of the quantity

$$(a-bw)^{-c},\quad (w=\cos\theta)$$

is derived. With the aid of the expansion theorem

$$\frac{\delta(\theta-\theta')}{\sin\theta}=\sum_{l=0}^{\infty}\left(\frac{2l+1}{2}\right)P_l(\cos\theta)P_l(\cos\theta'),$$

one gets

$$(a-bw)^{-c}=\sum_{l=0}^{\infty}\left(\frac{2l+1}{2}\right)A_lP_l(w),$$

with

$$A_l=\int_{-1}^{1}dwP_l(w)(a-bw)^{-c}.$$

Introduce Schläfli's[11] representation

$$P_l(w)=\frac{2^{-l}}{2\pi i}\oint\frac{(t^2-1)^ldt}{(t-w)^{l+1}}=\frac{2^{-l}}{l!}\frac{d^l}{dw^l}(w^2-1)^l,$$

and interchange the order of integration to obtain

$$A_l=\frac{2^{-l}}{2\pi i}\oint dt(t^2-1)^l\int_{-1}^{1}\frac{dw(a-bw)^{-c}}{(t-w)^{l+1}}.$$

The change of variable,

$$a-bw=x,$$

is made, and $(l+1)$-partial integrations are effected. With B_l defined by the relation

$$A_l=\frac{2^{-l}}{2\pi i}\oint dt(t^2-1)^lB_l,$$

B_l is given by

$$B_l=-\sum_{r=1}^{l+1}\frac{\Gamma(1-c)\Gamma(l+r-1)}{\Gamma(r+1-c)l!b^r}\left[\frac{(a-b)^{(r-c)}}{(t-1)^{(l+r)}}-\frac{(a+b)^{(r-c)}}{(t+1)^{(l+r)}}\right]$$

$$-\frac{\Gamma(1-c)(l+r)!b^l}{\Gamma(r-c+1)l!}\int_{(a+b)}^{(a-b)}\frac{dx\,x^{(l+1-c)}}{(x-a+bt)^{(2l+2)}}.$$

The remaining integral is easily eliminated by reverting to the original variable of integration and interchanging the order of integration. The relevant integral,

$$\oint\frac{dt(t^2-1)^l}{(t-w)^{2l+2}},$$

vanishes since

$$\frac{d^{(2l+1)}}{dw^{(2l+1)}}(w^2-1)^l=0.$$

The contour integration is now accomplished.

$$\frac{2^{-l}}{2\pi i}\oint\frac{(t^2-1)^ldt}{(t\mp1)^{(l+r)}}=\frac{2^{-l}}{(r-1)!}\frac{d^{(r-1)}}{dt^{(r-1)}}(t\pm1)^l$$

$$=\frac{(\pm1)^{(l-r+1)}l!2^{(1-r)}}{(l-r+1)!(r-1)!},$$

$$A_l=-2\sum_{r=1}^{l+1}\frac{\Gamma(1-c)(l+r-1)!(2b)^{-r}}{\Gamma(r+1-c)(l-r+1)!(r-1)!}$$

$$[(a-b)^{(r-c)}-(-1)^{(l-r+1)}(a+b)^{(r-c)}].$$

Set $r=s+1$ and rearrange to get

$$A_l=-\frac{1}{b}\sum_{s=0}^{l}\frac{\Gamma(1-c)(l+s)!}{\Gamma(s+2-c)(l-s)!s!}\left[\frac{1}{(a-b)^{(c-1)}}\left(\frac{a-b}{2b}\right)^s\right.$$

$$\left.-\frac{(-1)^l}{(a+b)^{(c-1)}}\left(-\frac{a+b}{2b}\right)^s\right].$$

This last result is recognized as

$$A_l=\frac{-1}{b(1-c)}\left[(a-b)^{(1-c)}\,{}_2F_1\left(-l,l+1,2-c;-\frac{a-b}{2b}\right)\right.$$

$$\left.-(-1)^l(a+b)^{(1-c)}\,{}_2F_1\left(-l,l+1,2-c;\frac{a+b}{2b}\right)\right].$$

For the case that c is a positive integer, this last formula valid only when

$$c\geq l+2.$$

An application of this expansion is found from the value

[11] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (The Macmillan Company, New York, 1946), American edition, Chap. XV.

of the following integral:

$$N^*(k)\Gamma(1+i\alpha)$$

$$\times\int \frac{d\mathbf{r}}{r}\,{}_1F_1(-i\alpha,1;-i[kr-\mathbf{k}\cdot\mathbf{r}])e^{-\lambda r+i[\mathbf{K}-\mathbf{k}]\cdot\mathbf{r}}.$$

This integral is given in the appendix of P; its value is

$$I=4\pi N^*(k)\Gamma(1+i\alpha)\frac{(\lambda^2+K^2+i2\lambda k-k^2)^{i\alpha}}{[\lambda^2+|\mathbf{K}-\mathbf{k}|^2]^{1+i\alpha}}.$$

The phase of the quantity M is given by the following scheme:

$$-M=\lambda^2+K^2+i2\lambda k-k^2,$$

$$\lim_{\lambda\to 0}\arg(-M)=\begin{cases}0, & k<K\\ \pi, & k>K\end{cases},$$

$$\lim_{\lambda\to 0}\arg(Me^{-\pi i})=\begin{cases}\pi, & k<K\\ 2\pi, & k>K\end{cases}.$$

Note that this choice of phase agrees with the scheme adopted for this term in the main body of the paper. Put

$$c=1+i\alpha,\quad a=\lambda^2+k^2+K^2,\quad b=2kK,$$

and

$$W_{\mp}=(k\mp K)^2+\lambda^2$$

in the expansion of

$$(a-b\cos\theta)^{-c}.$$

$$I=\frac{2\pi N^*(k)\Gamma(1+i\alpha)e^{\pi\alpha}M^{i\alpha}}{ikK\alpha}\sum_{l=0}^{\infty}\left(\frac{2l+1}{2}\right)$$

$$\times P_l(\cos\theta)\left[W_-^{-i\alpha}\,{}_2F_1\left(-l,l+1,1-i\alpha;-\frac{W_-}{4kK}\right)\right.$$

$$\left.-(-1)^lW_+^{-i\alpha}\,{}_2F_1\left(-l,l+1,1-i\alpha;\frac{W_+}{4kK}\right)\right].$$

In Eq. (3e), abstract the factor

$$\int\frac{i\epsilon d\mathbf{k}N(k)}{K^2+i\epsilon-k^2}L_l(kr)P_l(\cos\Theta)$$

from each term. The remaining factor (independent of r and θ and outside the bracket) is

$$\frac{\pi\Gamma(i\alpha)(2l+1)N^*(k)M^{i\alpha}e^{\pi\alpha}}{kK}.$$

Since

$$\Gamma(1+i\alpha)=i\alpha\Gamma(i\alpha),$$

it is observed that these two results agree. This serves as another method to accomplish the sum of the Jacobi polynomials that occur in Eq. (3c).

Note added in proof. It will be shown in a forthcoming publication that the results from Eq. (3a) through Eq. (4b) can be obtained in a more straightforward manner. This method utilizes the relation of the wave boundary condition to the branch cut in the continuous part of the spectrum. As a consequence of this relation, that series of Eq. (4a), which is not appropriate to the given boundary condition, does not occur.

Modification of Effective-Range Theory in the Presence of a Long-Range (r^{-4}) Potential*

THOMAS F. O'MALLEY, LARRY SPRUCH, AND LEONARD ROSENBERG
Physics Department, Washington Square College, and Institute of Mathematical Sciences, New York University, New York, New York

(Received March 3, 1961)

For short range potentials, there exists the effective-range theory expansion $k^{2L+1}\cot\eta(L) = -1/A(L)+\frac{1}{2}r_0(L)k^2+\cdots$, where $\eta(L)$ is the phase shift for angular momentum L. For long-range potentials, (potentials which vanish at large r only as some power of $1/r$), an expansion in k^2 does *not* exist. For $V(r) \to \text{const}\times r^{-n}$ for $r \to \infty$, the term at which the expansion breaks down depends upon L and upon n; $A(L)$ cannot be defined if $n \leqq 2L+3$, while for example for $L=0$ one cannot define r_0 in the usual way if $n \leqq 5$.

A detailed study is made of the case $n=4$. This case is of considerable interest since it arises, in the adiabatic approximation, in the scattering of a charged particle by a neutral polarizable system; the present analysis is concerned with the scattering by a static potential, but it can be readily generalized to include scattering by a compound (polarizable) system. The analysis is very much simplified by the existence of known mathematical solutions, Mathieu functions, of the Schrödinger equation with $V(r)$ *equal* to $\text{const}\times r^{-4}$. The expansion of $k\cot\eta(0)$ about $E=0$ contains a number of terms not present in the usual effective-range theory, including a term linear in k. The expansion about the energy of a weakly bound state does not contain these additional terms. It is rather of the usual form, but the correction will be of lower order than k^4. The leading terms in the expansion of $k^2\cot\eta(L)$ for $L\neq 0$ are also obtained.

1. INTRODUCTION

EFFECTIVE-RANGE theory,[1,2] which gives the leading terms in the expansion of $k^{2L+1}\cot\eta(L)$ as a function of the energy, where $\eta(L)$ is the scattering phase shift for angular momentum L, has been a very important tool in the analysis and interpretation of low-energy nucleon-nucleon scattering data. Restricting ourselves for the moment to the value $L=0$, it will be recalled that for neutron-proton (or neutron-neutron) scattering, one expects the shape-independent approximation[3]

$$k\cot\eta(0) \sim -1/A+\tfrac{1}{2}r_0k^2 \qquad (1.1)$$

to be a useful one for $k \lesssim k_{\max}$, where, roughly, $k_{\max}r_0=1$. In Eq. (1.1), $A\equiv A(0)$ is the scattering length and $r_0\equiv r_0(0)$ is the effective range. For neutron-proton scattering, a characteristic length is of the order of 10^{-13} cm, as is in particular r_0, and one therefore expects the shape-independent approximation to be useful for energies up to perhaps 10 Mev.

The range of validity of effective-range theory must to some extent be re-examined in each new application. Nevertheless, in atomic physics, one might well expect to have $r_0=ta_0$, with t a number somewhat greater than 1 and a_0 the Bohr radius; one might further expect that the range of validity of the theory would again be given by $k \lesssim 1/r_0$. Effective-range theory would then be quite useful.[4] For the scattering of an electron by an atom, for example, the theory would be useful from zero energy up to energies of the order of $\hbar^2/(2mr_0^2) = (13.6/t^2)$ ev; for t not too large compared to unity, this interval is one of the most interesting portions of the energy spectrum.

In point of fact, the situation in atomic scattering processes is rather more complicated than indicated so far. It is the purpose of the present paper to show that r_0 as normally defined is *infinite*, and that the very form of effective-range theory must be modified for the atomic case.[5] The origin of the deviation from the normal form lies in the fact that while it is not always explicitly stated, the usual effective-range theory assumes that the potentials are of short range, while the effective potentials of atomic scattering theory are of long range. We will refer to a potential as of short range if it falls off faster than any power of $1/r$. Thus, in effective-range theory for short-range potentials and for angular momentum L, one shows that there exists the expansion

$$k^{2L+1}\cot\eta(L) = -1/A(L)+\tfrac{1}{2}r_0(L)k^2+0(k^4). \qquad (1.2)$$

It is very well known, however, than an expansion of this form is not always valid. For a potential which asymptotically approaches a Coulomb potential, the very definition of phase shift must be altered; the

* This work was jointly supported by the U. S. Air Force Cambridge Research Laboratories, the Office of Ordnance Research, and the Office of Naval Research.

[1] J. M. Blatt and J. D. Jackson, Phys. Rev. **76**, 18 (1949). See also J. D. Jackson and J. M. Blatt, Revs. Modern Phys. **22**, 77 (1950); G. F. Chew and M. L. Goldberger, Phys. Rev. **75**, 1637 (1949).

[2] H. A. Bethe, Phys. Rev. **76**, 38 (1949).

[3] With the exception of the spherical Bessel functions $j_L(kr)$ and $n_L(kr)$, the angular momentum under consideration, when explicitly indicated, will appear in parentheses; subscripts will generally refer to the energy under consideration.

[4] The smallness of the numerical value of some characteristic length, which is after all a dimensional quantity, is not necessarily a measure of the usefulness of effective-range theory. Energies are in many regards "scaled" in relation to the characteristic length, and it is probably true that in general the range of validity of effective-range theory will cover a significant portion of the energy spectrum. It seems fair to say that effective-range theory was introduced in nuclear rather than in atomic theory not because of the disparity in the values of the characteristic lengths, but because of the lack of knowledge of the nuclear potential. To the extent that the nuclear potential can now be considered to be known, and to the extent that we restrict ourselves to scattering by a compound system—these are of course very severe restrictions—effective-range theory, appropriately modified, is potentially as valuable in atomic as in nuclear physics.

[5] A preliminary discussion of the results of the present paper has already appeared. [L. Spruch, T. F. O'Malley, and L. Rosenberg, Phys. Rev. Letters **5**, 375 (1960).] Equation (1) of this reference contains a misprint; in the last term on the right-hand side A^2 should be replaced by A^3.

extension of effective-range theory to include this case was given some time ago.[1,2] More significantly, for a repulsive potential which is everywhere proportional to $1/r^2$, a phase shift does exist but is energy independent so that for all positive k, and for all L,

$$k^{2L+1}\cot\eta(L)=C(L)k^{2L+1}. \tag{1.3}$$

Equation (1.3) is an identity, not the leading term of an expansion, and is clearly not of the form of Eq. (1.2).

Now for small energies the effective potentials of atomic physics do *not* fall off faster than any power of $1/r$. Thus, the interaction with a neutral molecule (we assume it doesn't have a quadrupole moment) of an ion of charge Ze and of sufficiently small energy behaves asymptotically as

$$V(r)\rightarrow -\tfrac{1}{2}Z^2e^2\alpha/r^4 \quad \text{for } r\rightarrow\infty, \tag{1.4}$$

where α is the molecular polarizability and r is the distance between the centers of the ion and of the molecule. For an electron or positron rather than an ion, one need merely replace Z^2e^2 by e^2. (It is rather more difficult to justify the adiabatic approximation in scattering of electrons or positrons by molecules than in the scattering of ions by molecules, but it is nevertheless, asymptotically,[6] a valid approximation at sufficiently low energies.) One then finds for all $L>0$ that $\cot\eta(L)$ is proportional to k^{-2} and hence that $k^{2L+1}\cot\eta(L)$ is proportional to k^{2L-1} for sufficiently small values of k. Thus, even the leading term in the above expansion, Eq. (1.2), is incorrect for $L>0$ when a long-range $1/r^4$ polarization potential is present. The leading term in $k\cot\eta(0)$ is independent of k for the polarization potential, as it is for short-range potentials. The next term, however, is *not* proportional to k^2, as we shall see, so that the shape-independent approximation is not valid even for $L=0$.

The problem which motivated the present research is the problem of the scattering of an electron or an ion by a neutral molecule which hasn't any permanent multipole moments. The asymptotic $1/r^4$ dependence of the effective interaction that is generally assumed for that case, at low energy, is a consequence of the adiabatic approximation. Though some doubt has recently been cast on the validity of this approximation, we believe it to be basically correct. (In the present paper we need only require that the approximation be valid beyond some value of r, which is certainly the case.) However, in the present paper that question will not be touched upon. In order to avoid the additional difficulties that are introduced in the analysis of scattering by compound systems, (the difficulties are complexities rather than questions of principle), we will restrict our attention to the mathematical question of the appropriate form of effective-range theory for scattering by a static central potential which obeys the usual requirements for small values of r and which approaches $-\frac{1}{2}Z^2e^2\alpha/r^4$ for large r. Applications to physically interesting problems will be considered in a subsequent paper.

Section 2 will consider the origin of the breakdown of the expansion of $k^{2L+1}\cot\eta(L)$ as a power series in k^2 for long-range potentials. The rest of the paper is primarily concerned with the polarization potential which goes asymptotically as a multiple of $1/r^4$. In Sec. 3, we consider in some detail the solutions of a Schrödinger equation with a potential which is everywhere proportional to $1/r^4$. These solutions play the same role as do the free solutions in the usual formulation of effective-range theory. The subsequent sections develop expansions of $k^{2L+1}\cot\eta(L)$ about $E=0$, and, for $L=0$, about the energy of a bound state.

2. BREAKDOWN OF POWER SERIES EXPANSION OF $k^{2L+1}\cot\eta(L)$ FOR LONG RANGE POTENTIALS

A. Conditions for Existence of $A(L)$

Before proceeding to an analysis of the long-range $1/r^4$ polarization potential which is our primary concern, we will attempt to obtain some insight into the general question of the origin of the breakdown of the expansion of $k^{2L+1}\cot\eta(L)$ as a power series in k^2.

The wave function $u(r)$ to be determined is defined by the Schrödinger equation

$$\left[\frac{d^2}{dr^2}+k^2-\frac{L(L+1)}{r^2}-\frac{2\mu}{\hbar^2}V(r)\right]u(r)=0, \tag{2.1}$$

and by the appropriate boundary conditions. For short-range potentials, $V(r)$ can be neglected entirely for large r and $u(r)$ must there approach a linear combination of the free solutions; we can, therefore, impose the boundary conditions

$$u(0)=0, \quad u(r)\rightarrow k^L[\cot\eta(L)krj_L(kr)-krn_L(kr)] \quad \text{for } r\rightarrow\infty. \tag{2.2}$$

For $k=0$, $u_0(r)$ should approach a linear combination of the $k=0$ free solutions, r^{L+1} and r^{-L}. If in Eq. (2.2) we let $k\rightarrow 0$, and if further we use

$$\lim_{k\rightarrow 0}k^{2L+1}\cot\eta(L)=-1/A(L), \tag{2.3}$$

we do in fact find that for $r\rightarrow\infty$ we have such a linear combination, namely,

$$u_0(r)\rightarrow(2L-1)!!r^{-L}-r^{L+1}/[(2L+1)!!A(L)], \tag{2.4}$$

where

$$(2L-1)!!\equiv 1\times 1\times 3\times\cdots(2L-1),$$
$$(2L+1)!!\equiv 1\times 3\times 5\times\cdots(2L+1).$$

We of course also have $u_0(0)=0$. The terms which were dropped in Eqs. (2.2) and (2.4) are terms which fall off with r faster than any power of $1/r$.

[6] Castillejo, Percival, and Seaton, Proc. Roy. Soc. (London) A254, 259 (1960).

For *long*-range potentials, however, $u_0(r)$ will *not* have the asymptotic form specified by Eq. (2.4); there are additional terms which fall off as powers of $1/r$ and which may in fact dominate over the r^{-L} term. These additional terms have their origin in the fact that one cannot entirely neglect a long-range potential no matter how large one chooses r. To prove this, we note that the Schrödinger equation for angular momentum L with

$$(2\mu/\hbar^2)V(r)=-\beta_n{}^2/r^n$$

has as its two independent solutions

$$r^{\frac{1}{2}}J_{(2L+1)/(n-2)}(x) \quad \text{and} \quad r^{\frac{1}{2}}N_{(2L+1)/(n-2)}(x),$$

where J and N are cylindrical Bessel and Neumann functions, respectively, and

$$x=2\beta_n r^{-\frac{1}{2}(n-2)}/(n-2).$$

From the asymptotic forms of these solutions, it follows that the leading terms in the asymptotic form of $u_0(r)$ are, for $n>2L+3$, r^{L+1} and r^{-L}, just as for a short-range potential; one then obtains the usual energy dependence $\eta(L)\to \text{const}k^{2L+1}$ for $k\to 0$. If, however, $n<2L+3$, the leading terms in the asymptotic form of $u_0(r)$ are r^{L+1} and r^{L-n+3}, which is not of the usual form, since the latter term dominates the usual r^{-L} term. One does not then have $\eta(L)\to$ const k^{2L+1} for $k\to 0$; that is, $A(L)$ as defined by Eq. (2.3) does not exist. For $n=2L+3$, the leading terms in the asymptotic form of $u_0(r)$ are r^{L+1} and $r^{-L}\ln r$, and once again $A(L)$ as defined by Eq. (2.3) does not exist. In this last case, incidentally, one has in the Born approximation that $\eta(L)\to \text{const}k^{2L+1}\ln k$, and not simply the k^{2L+1} dependence that has been given in the literature. In the above limiting forms for $\eta(L)$ we have ignored multiples of π.

A quite different proof of the fact that $A(L)$ as defined by Eq. (2.3) does not always exist follows from the variational expression for $A(L)$ which, in the present notation, takes the form

$$A(L)\approx A_t(L)-[(2L+1)!!]^2 \times\int u_t(r)\left[\frac{d^2}{dr^2}-\frac{L(L+1)}{r^2}-\frac{2\mu}{\hbar^2}V(r)\right]u_t(r)dr, \quad (2.5)$$

where the trial function u_t here satisfies the boundary conditions

$$u_t(0)=0, \quad u_t(r)\to -r^{L+1}+(2L+1)!!(2L-1)!!A_t(L)/r^L.$$

If $A(L)$ *does* exist, the variational expression, Eq. (2.5), should give a value arbitrarily close to $A(L)$ for sufficiently accurate trial functions u_t. However, for *any* u_t which has the above asymptotic form, the integrand in Eq. (2.5) approaches $(-2\mu/\hbar^2)V(r)r^{2L+2}$. It follows that if $-(2\mu/\hbar^2)V(r)\to\beta_n{}^2/r^n$ for large r, the integral is infinite if $2L+2-n\geqq -1$. For such values of n and L, then, $A(L)$, as defined by Eq. (2.3), does not exist, in agreement with the conclusion arrived at above.

Recently, a number of authors[7–9] have applied the Born approximation to determine the low-energy behavior of the phase shift for scattering by long-range potentials. To justify this procedure, it is usually stated that due to the presence of the centrifugal barrier, the low-energy scattering is determined only by the asymptotic form of $V(r)$, where $V(r)$ is very weak, the contribution to $\tan\eta$ from zero to any finite value of r going to zero as k^{2L+1}. One might then expect that, for L sufficiently large, the Born approximation will not only give the correct k dependence of the leading term in the expansion of $\tan\eta$ for small k, but will also give the correct coefficient of that term. One can go even further. It is almost certainly true that the leading terms in the Born expansion, where only the long-range part of $V(r)$ need be taken into account, will give exact results for all terms up to but not including the term in k^{2L+1}. (It should not be difficult to prove this but we have not attempted to do so.) The above expectation is strengthened by the results of the present paper which show by an entirely different approach that this expectation is verified for the particular case $n=4$. (Actually, the present paper includes the form of the energy dependence for all terms through k^{2L+3}.)

B. Conditions for Existence of r_0

If $A(L)$ does not exist, the question of the existence of $r_0(L)$ as defined by Eq. (1.2) does not arise. If $A(L)$ does exist, we must still examine the conditions under which $r_0(L)$ will also exist. We will restrict ourselves to the case $L=0$ and to $n\geqq 4$. (It should not be very difficult to do the same for $L>0$.) To analyze the existence conditions, it will be useful to very briefly review effective-range theory for short-range potentials. Of the derivations of the theory that have been given,[1,2] that of Bethe will be most convenient for our present purposes.

Since we are concerned with $L=0$, our boundary conditions, Eqs. (2.2) and (2.4), become

$$u(0)=0, \quad u(r)\to \cot\eta(0)\sin kr+\cos kr, \quad (2.2)'$$

$$u_0(0)=0, \quad u_0(r)\to 1-r/A. \quad (2.4)'$$

We also introduce the solutions of the free Schrödinger equation, with wave numbers k and 0, to be denoted by $v(r)$ and $v_0(r)$, respectively, with normalizations chosen so that they asymptotically approach $u(r)$ and $u_0(r)$, respectively. We then have, for all r,

$$v(r)=\cot\eta(0)\sin kr+\cos kr, \quad (2.6)$$

$$v_0(r)=1-r/A. \quad (2.7)$$

[7] E. Gerjuoy and S. Stein, Phys. Rev. **97**, 1671 (1955); B. H. Bransden, A. Dalgarno, T. L. John, and M. J. Seaton, Proc. Phys. Soc. (London) **71**, 877 (1958).
[8] R. M. Thaler, Phys. Rev. **114**, 827 (1959).
[9] L. M. Delves, Nuclear Phys. (to be published).

One can then readily derive the identity

$$k\cot\eta(0)=-(1/A)+k^2\int_0^\infty(vv_0-uu_0)dr. \quad (2.8)$$

At this stage, an approximation is made. It is argued that if $V(r)$ effectively vanishes for $r>R$, then v is effectively equal to u and v_0 to u_0 for $r>R$, so that the integrand is non-negligible only in the region 0 to R. It follows that for $\hbar^2k^2/(2\mu)$ small compared to some average value $\bar{V}$ of $V(r)$ in the region 0 to R, one can, with an error of the order of $k^2\hbar^2/(2\mu\bar{V})$, replace v by v_0 and u by u_0 in the integrand. We thereby obtain the shape-independent approximation, Eq. (1.2), with an error which is presumbaly proportional to k^4. The energy-independent effective range r_0 is then defined by

$$\tfrac{1}{2}r_0=\int_0^\infty(v_0^2-u_0^2)dr.$$

We have already seen that the shape-independent approximation is totally wrong for a repulsive potential equal to a multiple of $1/r^2$. The origin of the error is there clear; a solution $u_0(r)$ which satisfies the boundary conditions given by Eq. (2.4)′ does not exist, so that one cannot actually derive the indentity Eq. (2.8) for this case. For potentials which fall off at least as rapidly as $1/r^4$, the identity Eq. (2.8) *is* valid. For those cases for which r_0 does not exist, the origin of the difficulty must then lie in the subsequent approximations which were made in arriving at Eq. (1.2). The root of the trouble is that the asymptotic form of $u_0(r)$ contains terms in r, r^0, *and* $r^{-(n-3)}$ (we have $n\geqq4$), while $v_0(r)$ contains only the r and r^0 terms. Since $v_0^2-u_0^2$ will then contain a term of the form $1/r^{n-4}$, it follows that the integral which defines r_0 is infinite for $n=4$ or 5, that is, that r_0 does not exist for $n=4$ or 5.

For the case of primary interest, $V(r)\rightarrow-\frac{1}{2}Z^2e^2\alpha/r^4$, one has, for example,

$$u_0(r)\rightarrow(r/\beta)\sin(\beta/r)-(r/A)\cos(\beta/r)$$
$$=1-r/A+\tfrac{1}{2}\beta^2/(Ar)-\beta^2/(6r^2)+\cdots \quad (2.9)$$

for $r\rightarrow\infty$, where

$$\beta^2=(2\mu/\hbar^2)(\tfrac{1}{2}Z^2e^2\alpha). \quad (2.10)$$

Since $v_0(r)=1-r/A$ for all r, we have that while

$$\int_0^\infty(vv_0-uu_0)dr$$

is a finite number, the integral which presumably approximates it, namely,

$$\int_0^\infty(v_0^2-u_0^2)dr,$$

is infinite.

For those cases for which r_0 is infinite, the point then is that while $v_0(r)$ approaches $u_0(r)$ for $r\rightarrow\infty$ in the sense that the difference goes to zero, it does not approach it rapidly enough, and there is no point beyond which the effect of the long-range potential can be neglected. In fact, returning to the case of arbitrary L, there will for *any* finite n be *some* term at which the expansion of $k^{2L+1}\cot\eta(L)$ as a power series in k^2 will break down. On the other hand, if $V(r)\rightarrow0$ for $r\rightarrow\infty$ more rapidly than any power of $1/r$, then $v_0(r)-u_0(r)\rightarrow0$ for $r\rightarrow\infty$ more rapidly than any power of $1/r$ and the expansion will exist for all L.

3. MATHEMATICAL SOLUTIONS OF SCHRÖDINGER EQUATION WITH r^{-4} TERM

In preparation for the derivation of an effective-range formula for a potential which approaches a multiple of $1/r^4$ for large r, we seek two independent solutions of the differential equation

$$\left[\frac{d^2}{dr^2}-\frac{L(L+1)}{r^2}+\frac{\beta^2}{r^4}+k^2\right]M(r)=0. \quad (3.1)$$

Since we have allowed the term β^2/r^4 to extend all the way in to the origin, we cannot expect either of the two solutions to be a physically acceptable wave function; they will turn out to have a derivative at the origin whose value is not well defined. There cannot be any objection to such a procedure, however, since the solutions are not supposed to represent any physical situation but rather are simply useful mathematical constructs.

One could of course avoid the question by introducing an alternate differential equation in which the $1/r^4$ term is cut off at some small value of r. This would have the further advantage that this potential would correspond more closely to the true potential. On the other hand, it complicates the mathematics, and since the introduction of Eq. (3.1) does not lead to any difficulties, we prefer to use that approach.

Following Vogt and Wannier,[10] we introduce the new variable Y and the new function $\phi(Y)$ through

$$r=(\beta/k)^{\frac{1}{2}}e^Y \quad (3.2)$$

and

$$M(r)=r^{\frac{1}{2}}\phi(Y). \quad (3.3)$$

$\phi(Y)$ then satisfies the Mathieu equation,

$$[(d^2/dY^2)-(L+\tfrac{1}{2})^2+2\beta k\cosh(2Y)]\phi(Y)=0; \quad (3.4)$$

the properties of the solutions of this equation are

[10] E. Vogt and G. H. Wannier, Phys. Rev. **95**, 1190 (1954). Though we have made no use of their results beyond the basic result that one can convert the Schrödinger equation with a $1/r^4$ term into a Mathieu equation, it may be useful to record the connection between their parameters and ours, namely,

$$\beta=\nu-L,\quad \cos\pi\gamma=(1+m^2)/(2m),$$
$$e^\phi=(-1)^L(1-m^2)/(2m\sin\nu\pi).$$

rather well known.[11] Reverting to the coordinate r, the two independent solutions can be taken to be

$$M_{\pm\nu}{}^{(1)}=M_{\pm\nu}{}^{(1)}\{\ln[(\beta/k)^{-\frac{1}{2}}r]\}, \tag{3.5}$$

where, to lowest orders in βk, one has[12]

$$\nu\approx(L+\tfrac{1}{2})-(\beta k)^2/[4(L+\tfrac{3}{2})(L+\tfrac{1}{2})(L-\tfrac{1}{2})]. \tag{3.6}$$

It will also be convenient to introduce the quantities

$$m\equiv M_{\nu}{}^{(1)}(0)/M_{-\nu}{}^{(1)}(0) \tag{3.7}$$

and

$$\delta\equiv\tfrac{1}{2}\pi(\nu-L-\tfrac{1}{2})\approx-\pi\beta^2k^2/[8(L+\tfrac{3}{2})(L+\tfrac{1}{2})(L-\tfrac{1}{2})]. \tag{3.8}$$

m can be expanded as

$$m=(\tfrac{1}{4}\beta k)^{\nu}[\Gamma(1-\nu)/\Gamma(1+\nu)]\times[1+C_1\beta^2k^2+C_2\beta^4k^4+\cdots], \tag{3.9}$$

where C_1 and C_2 follow from an analysis of Eqs. (2.6–22) and (2.2–38) of reference 11, and where Γ is the gamma function. The leading term of Eq. (3.9) is given explicitly by

$$m=(-1)^L(2L+1)(\beta k)^{L+\frac{1}{2}}/[(2L+1)!!]^2+\cdots, \tag{3.10}$$

while the next term follows from the relationship

$$C_1=-\tfrac{1}{4}\nu/(1-\nu^2)^2. \tag{3.11}$$

For our purposes, it will be convenient to choose as the two independent solutions

$$v_{ps}(r)=Q[m\cos\delta M_{-\nu}{}^{(1)}+(-1)^L(1/m)\sin\delta M_{\nu}{}^{(1)}], \tag{3.12}$$

and

$$v_{pc}(r)=Q[m\sin\delta M_{-\nu}{}^{(1)}+(-1)^L(1/m)\cos\delta M_{\nu}{}^{(1)}], \tag{3.13}$$

where

$$Q\equiv(\tfrac{1}{2}\pi/\beta)^{\frac{1}{2}}r^{\frac{1}{2}}/\cos2\delta. \tag{3.14}$$

The subscript p refers to the fact that the differential equation under consideration contains the polarization potential. The subscripts s and c are suggested by the sine and cosine dependence of the functions for sufficiently small r. Thus, it follows from the known properties of the $M_{\pm\nu}{}^{(1)}$ that

$$v_{ps}(r)\sim\sin[(\beta/r)-\tfrac{1}{2}L\pi]/(\beta/r),\quad r\ll(\beta/k)^{\frac{1}{2}},$$
$$v_{pc}(r)\sim\cos[(\beta/r)-\tfrac{1}{2}L\pi]/(\beta/r),\quad r\ll(\beta/k)^{\frac{1}{2}}.$$

For large r, we have

$$v_{ps}(r)\sim\frac{(-1)^Lm}{(\beta k)^{\frac{1}{2}}\cos2\delta}\left[\cos(kr-\tfrac{1}{2}L\pi)\left(\cos^2\delta-\frac{\sin^2\delta}{m^2}\right)+\sin(kr-\tfrac{1}{2}L\pi)\sin\delta\cos\delta\left(\frac{1}{m^2}-1\right)\right],\quad r\gg(\beta/k)^{\frac{1}{2}} \tag{3.15}$$

$$v_{pc}(r)\sim\frac{(-1)^Lm}{(\beta k)^{\frac{1}{2}}\cos2\delta}\left[\cos(kr-\tfrac{1}{2}L\pi)\sin\delta\cos\delta\left(1-\frac{1}{m^2}\right)+\sin(kr-\tfrac{1}{2}L\pi)\left(\frac{\cos^2\delta}{m^2}-\sin^2\delta\right)\right],\quad r\gg(\beta/k)^{\frac{1}{2}}. \tag{3.16}$$

As for short-range potentials, the two independent solutions have been chosen in such a way as to make their behavior for small r relatively simple. There is no need to choose solutions which have a simple behavior at large r since the $u(r)$ and $v(r)$ are chosen to approach one another at large r and one never actually has to perform calculations with the functions at large r.

4. MODIFIED EFFECTIVE-RANGE EXPANSION

In our analysis of Eq. (2.1) for short-range potentials, we took $u(r)$ at large r to be a linear combination of the two solutions of that equation with $V(r)$ neglected, i.e., $krj_L(kr)$ and $krn_L(kr)$. The relative amplitude of these two terms, which is fixed by the requirement that the continuation of $u(r)$ must vanish at $r=0$, determines the phase shift. We are now however concerned with a long-range potential such that

$$\Delta V(r)\equiv V(r)+(\hbar^2/2\mu)\beta^2/r^4 \tag{4.1}$$

vanishes as $r\to\infty$ more rapidly than any power of $1/r$. A modified approach must then be used if we are to obtain the equivalent of an effective-range expansion, for we have seen that r_0 as normally defined does not exist. We rewrite our differential equation as

$$\left[\frac{d^2}{dr^2}-\frac{L(L+1)}{r^2}+\frac{\beta^2}{r^4}-\frac{2\mu}{\hbar^2}\Delta V(r)+k^2\right]u(r)=0.$$

Since $\Delta V(r)$ *is* a short-range potential, we can neglect it at large r, and write $u(r)$ as a linear combination of the solutions of the resultant equation, namely, v_{ps} and v_{pc}. The relative amplitude, which we will denote by B, can readily be related to $\tan\eta(L)$.

Thus, we solve the differential equation, Eq. (2.1), subject to the boundary conditions

$$u(0)=0,\quad u(r)\to v_{ps}(r)+Bv_{pc}(r). \tag{4.2}$$

Once B has been found, $\tan\eta(L)$ is determined, since inserting the asymptotic forms of v_{ps} and v_{pc}, and comparing with the usual asymptotic form $u(r)\to$ const $\times\sin[kr-\frac{1}{2}L\pi+\eta(L)]$, one finds

$$\tan\eta(L)=\frac{m^2-\tan^2\delta+B\tan\delta(m^2-1)}{\tan\delta(1-m^2)+B(1-m^2\tan^2\delta)}. \tag{4.3}$$

For the development of a modified effective-range

[11] J. Meixner and F. W. Schafke, *Die Grundlehren der Mathematischen Wissenschaften* (Springer-Verlag, Berlin, Germany, 1954), Vol. 71.

[12] The higher terms in the expansion of ν can be obtained from the expansion of $\cos\nu\pi$ given by Meixner and Schafke, reference 11, Eq. (2.2-44). These higher terms will not, however, be required for our present purposes.

theory, it is useful to define

$$v_p(r)=v_{ps}(r)+Bv_{pc}(r) \tag{4.4}$$

for all r. We then let $u_1(r)$ and $u_2(r)$ be solutions of Eq. (2.1) for energies E_1 and E_2, respectively, and $v_{p1}(r)$ and $v_{p2}(r)$ be the corresponding solutions of Eq. (3.1); all four solutions satisfy the appropriate boundary conditions. Proceeding then in a fashion almost identical with that used by Bethe,[2] we find

$$\begin{aligned}(u_1u_2'-u_2u_1')\Big|_a^b&=(k_1^2-k_2^2)\int_a^b u_1u_2dr,\\ (v_{p1}v_{p2}'-v_{p2}v_{p1}')\Big|_a^b&=(k_1^2-k_2^2)\int_a^b v_{p1}v_{p2}dr,\end{aligned} \tag{4.5}$$

where a and b are arbitrary. Subtracting, letting $a\to 0$ and $b\to\infty$, and using the specified forms of the various functions and their derivatives for very small and for very large values of r, we obtain

$$(B_2/\beta)-(B_1/\beta)=(k_2^2-k_1^2)\int_0^\infty (v_{p2}v_{p1}-u_2u_1)dr. \tag{4.6}$$

Eq. (4.6) is valid for all L. The situation in the present case is then in this regard simpler than for a short-range potential, where the derivation of an effective-range expansion is rather different for $L>0$ than for $L=0$, the difference arising due to the presence of the $L(L+1)/r^2$ terms which give rise to difficulties at $r=0$ for $L>0$. For our present long-range potential, the $1/r^4$ term dominates the $L(L+1)/r^2$ for $r\to 0$ for all L, so one need not distinguish between $L=0$ and $L>0$. Note that the identity Eq. (4.6) reduces for $L=0$ to that deduced by Bethe for $\beta\to 0$. That it does so follows from the observation that $B/\beta\to k\cot\eta(0)$ for $\beta\to 0$.

The long-range polarization effects having been accounted for, the difference between v_{p1} and u_1, or between v_{p2} and u_2, falls off faster than any power of $1/r$. It was precisely because the difference between v_1 and u_1 or between v_2 and u_2 did *not* fall off faster than any power of $1/r$, for the conventional form for v_1 or for v_2, that the effective-range expansion for $k^{2L+1}\times\cot\eta(L)$ broke down. We *do* however have a valid effective-range expansion for B/β.

5. EXPANSION ABOUT ZERO ENERGY

We specialize now at the case for which $E_1=0$. Denoting the corresponding quantities by the subscript 0, we then have

$$v_{p0}(r)=j_L(\beta/r)-B_0n_L(\beta/r). \tag{5.1}$$

When the subscript 2 is dropped, Eq. (4.6) becomes

$$(B/\beta)=(B_0/\beta)+\tfrac{1}{2}\rho_p(0,E)k^2, \tag{5.2}$$

where

$$\tfrac{1}{2}\rho_p(0,E)\equiv\int_0^\infty (v_{p0}v_p-u_0u)dr. \tag{5.3}$$

Thus far, all of our equations are exact. We now write[13]

$$(B/\beta)\approx(B_0/\beta)+\tfrac{1}{2}r_{p0}k^2, \tag{5.4}$$

where

$$\tfrac{1}{2}r_{p0}\equiv\tfrac{1}{2}\rho_p(0,0)=\int_0^\infty (v_{p0}^2-u_0^2)dr, \tag{5.5}$$

and where the error involves a term of order k^4, the relative error being of the order of $(\hbar^2k^2/2\mu)/\bar{V}$; here $\bar{V}$ is some average value of $V(r)$ taken over the range of the short-range potential.

From Eq. (4.3) we find that

$$\lim_{k\to 0}(\tan\eta(0)/k)\equiv -A=\beta/B_0. \tag{5.6}$$

By using Eqs. (2.3), (3.8), (3.9), and (4.3), Eq. (5.4) can be written, to terms in k^2, as

$$\begin{aligned}k\cot\eta(0)&=-\frac{1}{A}+\frac{\pi\beta^2}{3A^2}k+\frac{4\beta^2}{3A}k^2\ln\left(\frac{\beta k}{4}\right)\\&\quad+\left[\frac{1}{2}r_{p0}+\frac{\pi\beta}{3}+\frac{20\beta^2}{9A}-\frac{8\beta^2}{3A}\psi\left(\frac{3}{2}\right)-\frac{\pi\beta^3}{3A^2}-\frac{\pi^2\beta^4}{9A^3}\right]k^2\\&\quad+\ldots,\end{aligned} \tag{5.7}$$

where

$$\psi(\tfrac{3}{2})=\Gamma'(\tfrac{3}{2})/\Gamma(\tfrac{3}{2})=0.0365.$$

The expansion for $\tan\eta(0)$ is slightly simpler in form. We have given the expansion for $k\cot\eta(0)$ simply because that is the conventional form.

It is well known that $k\cot\eta(0)$ must be an even function of k for real k. This property does not depend upon any special properties of $V(r)$, contrary to the situation when one is attempting to prove the analyticity of $k\cot\eta(0)$. Despite the appearance of Eq. (5.7), $k\cot\eta(0)$ *is* in fact an even function of k. The point is that for k real and negative, one must utilize an alternate asymptotic expansion of the Mathieu functions,[14] and it is found that in Eq. (5.7) k must everywhere be replaced by $-k$.

For a potential $V(r)$ for which $\Delta V(r)$ as defined by Eq. (4.1) vanishes more rapidly than any power of $1/r$, as we have assumed, one could obtain the coefficient of k^3 in the expansion of $k\cot\eta(0)$ without introducing

[13] Equation (5.4) is identical in form with the usual expansion for $k^{2L+1}\cot\eta(L)$, v_p having been chosen to accomplish this. Equation (4.3) is then rather more complicated in form than it might otherwise be.

[14] The alternate asymptotic expansion of the Mathieu function follows from the realization (see reference 11) that $M_\nu^{(1)}$ approaches J_ν, and that there are different asymptotic expansions of J_ν depending upon which sector of the complex plane the argument of J_ν lies in.

any new parameters; the coefficient is in fact independent of r_0 and can be expressed in terms of A and β. In most cases of interest, however, $\Delta V(r)$ will have additional terms that vanish as d/r^6, and the coefficient will then depend upon d as well. We have, therefore, not calculated the k^3 term. We will not here go into this point any further.

The term linear in k in Eq. (5.7) has previously been found by Thaler[8] in a Born approximation analysis of the effect of the polarizability of the neutron on the scattering of neutrons by nuclei with large Z.

For $L>0$, Eq. (4.3) leads to

$$\tan\eta(L)=-\tan\delta+\frac{(2L+1)^2(\beta k)^{2L+1}}{[(2L+1)!!]^4 B_0}+\cdots. \quad (5.8)$$

[This equation is in fact valid for $L=0$ as well, but for that case we have gone well beyond this result in Eq. (5.7).] Higher-order terms in the expansion of δ, [see Eq. (3.8)], can be obtained. The expansion of $\tan\eta(L)$ up through terms not including k^{2L+1} can then be expressed in terms of β alone, while the expansion including the k^{2L+1} term involves in addition only B_0. The k^{2L+1} term is of course the leading term for short-range potentials.

It should be noted that the techniques previously[15] introduced for short-range potentials for the determination of rigorous upper bounds on $A(L)$ can readily be adapted to determine rigorous upper bounds on $-1/B_0$. One would use the normalization

$$u_0(r)\to n_L(\beta/r)-(1/B_0)j_L(\beta/r).$$

Returning to Eq. (5.8), the leading term in the expansion is given by

$$\tan\eta(L)=\{\pi\beta^2/[8(L+\tfrac{3}{2})(L+\tfrac{1}{2})(L-\tfrac{1}{2})]\}k^2+\cdots. \quad (5.9)$$

That $\eta(L)$ is proportional to k^2 and not to k^{2L+1} for k sufficiently small and $L>0$ has previously been shown to be true in the Born approximation.[7,8] Thaler in fact found the above coefficient for $L=1$ and $L=2$.[16]

It might be noted that to lowest order in k, the total cross section is now given by the expression

$$\sigma=4\pi A^2+(8/3)\pi^2\beta^2 A k+\cdots.$$

We have throughout the paper expressed all expansions in terms of β, B_0, and r_{p0}. In many ways, it would be preferable to determine the expansion in terms of B_0, r_{p0}, and the various quantities connected with the Mathieu solutions, such as ν, m, and δ, without expanding these latter quantities as power series in k^2, since these quantities are rather delicate functions of k. For a polarizability numerically equal to that of a hydrogen atom, the expansions break down in the neighborhood of 1.5 ev. For scattering by a compound system, where one is dealing with a partial differential equation, it would be perfectly reasonable to determine the quantities ν, m, and δ, which are defined by an ordinary differential equation, numerically, and to proceed from there. With this point of view, the potential which goes as $1/r^4$ is no simpler than other long-range potentials.

The limitation to energies below about 1.5 ev, for β equal to that of a hydrogen atom, might seem to be severe. On the other hand, this is easily the energy interval most difficult to study, both theoretically and experimentally. Furthermore, the adiabatic approximation on which the $1/r^4$ potential is based cannot be valid over a significant region of r at any but very low energies, though precisely how low is not really known. Thus, it does not seem worthwhile to seek alternate expansions of ν, m, and δ which are valid for higher energies, even though they must exist, or even to seek numerical values of ν, m, and δ in the higher-energy range.

6. EXPANSION ABOUT ENERGY OF BOUND STATE

If there exists a bound state of angular momentum L with energy $E_\gamma=-\hbar^2\gamma^2/(2\mu)$ which is very small and which is known, it is well known in the usual effective-range theory that it is often more convenient to expand about $E=E_\gamma$ rather than about $E=0$. The same is true for the present case of a long-range potential, though here we will be concerned initially with the expansion of B and only then with the expansion of $\cot\eta(L)$.

Thus, while Eq. (4.6) was derived for two non-negative energies, it remains valid when one of the energies is negative, and when both of the energies are negative. The point is that the expressions for the Mathieu functions that were used in the derivation of Eq. (4.6) are valid not simply for k real and positive but rather in the sector $-\pi<\arg k<\pi$. (We make specific note of the fact that the expressions are *not* valid for k real and negative. This point was critical in the proof that $\eta(L)$ is an odd function of k.) Equation (4.6) can then be written as

$$(B/\beta)=(B_\gamma/\beta)+(k^2+\gamma^2)\int_0^\infty (v_{p\gamma}v_p-u_\gamma u)dr. \quad (6.1)$$

Proceeding in a fashion identical to that used at zero energy, this identity can be approximated by

$$(B/\beta)\approx(B_\gamma/\beta)+\tfrac{1}{2}\rho_{p\gamma}(k^2+\gamma^2), \quad (6.2)$$

where the energy-independent effective range $\rho_{p\gamma}$ is defined by

$$\tfrac{1}{2}\rho_{p\gamma}\equiv\int_0^\infty (v_{p\gamma}{}^2-u_\gamma{}^2)dr. \quad (6.3)$$

One advantage of the present expansion about $E=E_\gamma$

[15] L. Spruch and L. Rosenberg, Phys. Rev. **116**, 1034 (1959); **117**, 1095 (1960) and L. Rosenberg, L. Spruch, and T. F. O'Malley, *ibid.* **118**, 184 (1960).

[16] Thaler's expression (reference 8) for arbitrary L is correct only for $L=1$ and for $L=2$.

over the expansion about $E=0$ is that whereas B_0 depended upon $V(r)$ for small r in a detailed way, B_γ follows immediately simply from a knowledge of E_γ, and E_γ will often be known experimentally. Thus the requirement that the bound state function behave asymptotically like $e^{-\gamma r}$ leads [see Eqs. (3.15), (3.16), and (4.2), replacing k by $i\gamma$] to the expression

$$B_\gamma=\frac{im_\gamma^2e^{2i\delta_\gamma}-\tan\delta_\gamma}{1+im_\gamma^2e^{2i\delta_\gamma}\tan\delta_\gamma}, \tag{6.4}$$

where m_γ and δ_γ are to be evaluated at $E=E_\gamma$.[17] Specializing now to the particularly important case of $L=0$, we have

$$\delta_\gamma=-\tfrac{1}{3}\pi\beta^2\gamma^2+\cdots \quad \text{for } L=0 \tag{6.5}$$

and

$$im_\gamma^2e^{2i\delta_\gamma}=\beta\gamma+0(\gamma^3\ln\gamma) \quad \text{for } L=0 \tag{6.6}$$

so that, to terms in γ^2,

$$B_\gamma\approx-\beta\gamma+\tfrac{1}{3}\pi\beta^2\gamma^2, \quad \text{for } L=0. \tag{6.7}$$

From Eqs. (3.8), (3.11), (4.3), (6.2), and (6.7), one then finds

$$k\cot\eta(0)\approx-\gamma+(\tfrac{1}{2}\rho_{p\gamma}+\tfrac{1}{3}\pi\beta)(\gamma^2+k^2). \tag{6.8}$$

Equation (6.8) differs from the usual expansion in that it contains $\frac{1}{2}\rho_{p\gamma}+\frac{1}{3}\pi\beta$ rather than $\frac{1}{2}\rho_\gamma$. We will now show that to the order to which we are working, these two expressions are the same.

Thus, consider the function $v_{p\gamma'}$ which is defined by the same differential equation and boundary conditions which determine $v_{p\gamma}$ except that γ is replaced by γ'. (We do not assume the existence of a bound state with energy $(\hbar^2/2\mu)\gamma'^2$.) By manipulating the two differential equations for $v_{p\gamma}$ and $v_{p\gamma'}$ we obtain, in the usual way,

$$(\gamma^2-\gamma'^2)\int_0^\infty v_{p\gamma}v_{p\gamma'}dr=(v_{p\gamma}{}'v_{p\gamma'}-v_{p\gamma}v_{p\gamma'}{}')\Big|_0^\infty$$
$$=-(1/\beta)(B_\gamma-B_{\gamma'}).$$

If we let γ' approach γ, the desired integral may be written

$$\int_0^\infty v_{p\gamma}{}^2dr=-\frac{1}{\beta}\frac{dB_\gamma}{d\gamma^2}=-\frac{1}{2\beta\gamma}\frac{dB_\gamma}{d\gamma}\approx\frac{1}{2\gamma}-\frac{1}{3}\pi\beta.$$

It follows that

$$\tfrac{1}{2}\rho_{p\gamma}\approx1/(2\gamma)-\tfrac{1}{3}\pi\beta-\int_0^\infty u_\gamma{}^2dr. \tag{6.9}$$

On the other hand, since $v_\gamma=\exp(-\gamma r)$, we have

$$\tfrac{1}{2}\rho_\gamma\equiv\int(v_\gamma{}^2-u_\gamma{}^2)dr=1/(2\gamma)-\int_0^\infty u_\gamma{}^2dr. \tag{6.10}$$

Equation (6.8) can therefore be written as

$$k\cot\eta(0)\approx-\gamma+\tfrac{1}{2}\rho_\gamma(\gamma^2+k^2). \tag{6.11}$$

Thus, to order γ^2, the formulation which explicitly takes into account the long-range character of the potential leads to precisely the same result as the usual formulation utilized in application to short-range potentials. This conclusion must still, of course, be verified for scattering by a compound system when identical particles are involved. Furthermore, even if the ordinary shape-independent formula is valid in this case, it is very unlikely that the error term is of order k^4 since in the one-body problem one does find lower-order corrections. With these qualifications, our results then justify the very nice use which has been made of Eq. (6.11) in the analysis of low-energy singlet scattering of electrons by hydrogen atoms,[18] and makes less surprising the good agreement obtained between the singlet scattering length thereby obtained and that determined on the basis of a rigorous minimum principle.[19] (On the other hand, the good agreement does not guarantee that the results are accurate.)

It is of interest to inquire into the absence of a term linear in k (as well as a number of other terms) in the present expansion about $E=E_\gamma$ as contrasted to the situation for the expansion about $E=0$. This can be discussed in a number of ways. On the one hand, we saw that the integral which defined ρ_0 did not even exist, due to the difference in the asymptotic behavior of v_0 and of u_0; while v_γ and u_γ also have a different asymptotic behavior, they are both now dominated by the factor $e^{-\gamma r}$, so that $\int v_\gamma{}^2dr$ and $\int u_\gamma{}^2dr$ exist separately, and hence so does ρ_γ. More physically, one can understand the difference in the following way. In the expansion about $E=0$, no assumption whatever was made about the magnitude of A and only terms in k^3 were neglected. The expansion about $E=E_\gamma$ assumes however that E_γ and hence γ is small, and terms in γ^3, γ^2k, and γk^2 were neglected, as well as terms in k^3. Now it follows from a comparison of the two expansions of $k\cot\eta(0)$, about $E=0$ and about $E=E_\gamma$, that $1/A$ is small if γ is small. (In fact, $1/A\approx\gamma$.) If in Eq. (5.7) we drop terms in $1/A^3$, in (k/A^2), and in (k^2/A), which are now on the same footing as k^3 terms, we remain with

$$k\cot\eta(0)\approx-1/A+(\tfrac{1}{2}r_{p0}+\tfrac{1}{3}\pi\beta)k^2$$

which is just an expansion of the usual form in that it contains a constant and a term in k^2. On the other hand, no matter how small $1/A$ is, there will always be an interval in the neighborhood of $k=0$, a very small interval to be sure, for which the term linear in k will be more important than the term in k^2.

[17] The replacement of k by $i\gamma$ is permissible because, as noted previously, the Mathieu functions that we have used are valid for $-\pi<\arg k<\pi$; the question of the analyticity of $k\cot\eta(L)$, for example, does not arise. Equation (6.4), incidentally, is formally equivalent to replacing $\tan\eta(L)$ by i in Eq. (4.3).

[18] T. Ohmura, Y. Hara, and T. Yamanouchi, Progr. Theoret. Phys. (Kyoto) **22**, 152 (1959); **20**, 80 (1958); T. Ohmura and H. Ohmura, Phys. Rev. **118**, 154 (1960).

[19] L. Rosenberg, L. Spruch, and T. F. O'Malley, Phys. Rev. **119**, 164 (1960).

Transition Matrix for Nucleon-Nucleon Scattering*

K. L. KOWALSKI AND D. FELDMAN
Department of Physics, Brown University, Providence, Rhode Island
(Received December 22, 1960)

As part of a study of the influence of off-the-energy-shell effects on the optical potential for nucleon-nucleus scattering, a method is presented for the calculation, via the reactance matrix, of the nucleon-nucleon transition matrix in terms of an internucleon potential and the scattering amplitude. The singular integral equations for the partial-wave amplitudes of the reactance matrix are reduced to a Fredholm form which contains the scattering amplitude parametrically. The iteration solution of these Fredholm equations is shown to be generally unreliable; however, the zeroth-order iteration approximates the exact solution quite well near the energy shell. The replacement of the kernels of these integral equations by separable functions is discussed; the validity of such an approximation is illustrated by a simple example. The requirement that the solutions of the (exact) Fredholm equations be consistent with the original singular integral equations yields a solution for the scattering amplitude in terms of the resolvent kernels of the Fredholm equations. The entire formalism is so constructed as to include the possibility of a hard core being present in the nucleon-nucleon interaction.

INTRODUCTION

MOST contemporary investigations of the scattering of high-energy nucleons by nuclei formulate the many-body scattering problem in terms of two-body processes[1]; that is, the complete many-body transition matrix is expressed in terms of the two-body transition matrices. Thus, the problem is viewed as a series of processes between the incident nucleon and each one of the bound nucleons. In this case, the usual two-body kinematics are modified by the fact that one particle is bound.

A customary approximation in these studies has been to neglect the struck particle's initial energy in the laboratory system as compared with the energy of the incident particle in the same reference frame. It is then usual to regard the two-body process as one occurring between two free nucleons.[2] Since the OES values[3] of the two-body transition matrices can, to some extent, be determined experimentally,[4] the many-body scattering problem can then be investigated phenomenologically in a well-defined way.

It may be expected that FES processes will have to be taken into account in order to extend the usefulness of the various scattering formalisms to larger scattering angles and a wider range of energies.[5,6] To accomplish this, a knowledge of the FES characteristics of the two-nucleon transition matrix is obviously needed. It is the intent of the present study to provide a relatively useful method for obtaining this information.

At the present time, the most convenient way to investigate the FES properties of the two-nucleon transition matrix is by means of an internucleon potential. Since at least two nucleon-nucleon potentials exist which reproduce the two-body data reasonably well up to about 300 Mev,[7,8] such an approach has a fairly secure foundation phenomenologically. Indeed, an investigation of FES effects in many-body scattering problems could conceivably produce another means of checking the validity of a particular potential.

Once a particular potential is assumed, all the properties of the nucleon-nucleon transition matrix can be calculated. Such a direct approach of course requires considerable effort. However, if the OES properties are regarded as known, the potential can be used merely to describe the virtual or FES properties of the two-body scattering. Thus the OES values of this matrix are simply regarded as numerical parameters. Then it may be possible to obtain the remaining information about the transition matrix in a less involved manner than in an analysis where nothing is assumed known except the potential.

In order to utilize the information for OES scatterings, it is convenient to consider the integral equation for the transition matrix rather than to proceed through the intermediate step of calculating a wave function. There exist, however, two formal difficulties in this approach. The first arises because of the possibility of the *potential* containing a hard-core singularity.[7,9] In this case, the usual form of the integral equation is meaningless.[10] In Sec. I, the integral equations appropriate to the case when the potential contains a hard-core singularity are developed for the various quantities which occur in scattering theory.

The second difficulty is a result of the fact that the pertinent *integral equation* is singular in general.

* This work was supported, in part, by the U. S. Atomic Energy Commission.

[1] A. Kerman, H. McManus, and R. M. Thaler, Ann. Phys. 8, 551 (1959), and works cited therein.

[2] This last step is exact only for forward scattering.

[3] A process between states of unequal energy will be referred to as off-the-energy shell (FES) and one between states equal in energy as on-the-energy shell (OES).

[4] Cf. H. P. Stapp, T. J. Ypsilantis, and N. Metropolis, Phys. Rev. **105**, 302 (1957); H. P. Stapp, U. S. Atomic Energy Commission Rept. UCRL 3098 (1955).

[5] T. Fulton and P. Schwed, Phys. Rev. **115**, 973 (1959).

[6] H. Feshbach, Ann. Rev. Nuclear Sci. 8, 49 (1958).

[7] J. L. Gammel and R. M. Thaler, Phys. Rev. **107**, 291, 1337 (1957).

[8] P. S. Signell and R. E. Marshak, Phys. Rev. **109**, 1229 (1958).

[9] P. S. Signell, R. Zinn, and R. E. Marshak, Phys. Rev. Letters **1**, 416 (1958); R. A. Bryan, Nuovo cimento **16**, 895 (1960).

[10] H. Feshbach, Ann. Phys. **5**, 357 (1958).

Because of this, it is difficult to formulate a well-defined approximation procedure. After a partial-wave reduction is performed in Sec. II, the angle-independent singular integral equations are reduced to Fredholm form in Sec. III. This is done in such a way as to incorporate parametrically the OES values of the transition matrix into the general solution.

The solution of the Fredholm equations is discussed in Sec. IV. Under certain conditions, it is found that the solutions of these equations can be approximated by finite expressions which involve the matrix elements of the potential and the OES values of the transition matrix. Finally, the consequences of the requirement of consistency between the solutions of the Fredholm equations and the original singular integral equations are deduced in Sec. V.

In this investigation, electromagnetic effects are completely neglected. In addition, all processes are considered to be nonrelativistic.

I. SCATTERING EQUATIONS WITH A HARD-CORE SINGULARITY

The boundary conditions on the wave function which are required by the presence of a hard core (h.c.) in the interaction can be replaced by a singular potential function.[11,12] When this replacement is made so as to have the singularity at the core radius, the procedure is exact[13] for the two-body problem.[12] This method of handling the h.c. will now be applied to the integral equations of scattering theory. The ultimate result will be the integral equations for the transition and reactance matrices, modified so as to include a h.c. singularity in the interaction; it is this which constitutes the extension of the previously noted investigations. All considerations will be limited to the case of two bodies.

Consider the following modification of the Lippmann-Schwinger integral equation[14] corresponding to outgoing waves at infinity and energy E_i:

$$\psi_i^{(+)}=\phi_i+G_i(V\psi_i^{(+)}+f), \tag{1.1}$$

where G_i is the *propagator* $(E_i-H_0+i\epsilon)^{-1}$, H_0 is the free two-particle Hamiltonian (excluding the center-of-mass motion), and ϕ_i is the normalized eigenstate of H_0 corresponding to an energy E_i. The interparticle potential is denoted by V, and f is an arbitrary vector (in Hilbert space).

Let $\mathbf{r}$ denote the relative position vector of the two particles. The coordinate representatives of V and f will be assumed to have the forms

$$V(\mathbf{r})=0, \quad |\mathbf{r}|<a, \tag{1.2}$$

$$f(\mathbf{r})=\delta(r-a)\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\lambda_{lm}Y_l^m(\mathbf{r},\mathbf{k}_i), \tag{1.3}$$

where λ_{lm} and a are constants, and $Y_l^m(\mathbf{r},\mathbf{k}_i)$ denotes the normalized spherical harmonic which is a function of the angles of $\mathbf{r}$ with respect to a set of coordinate axes whose polar axis is in the direction of $\mathbf{k}_i$. The incident wave vector $\mathbf{k}_i$ has a magnitude defined by $E_i=\hbar^2k_i^2/2M$, where M is the reduced two-particle mass. In the following, sums over l and m will always have the form indicated in Eq. (1.3), unless specified explicitly to the contrary. Henceforth, the limits of such sums will not be indicated.

It is clear, from Eqs. (1.1) and (1.3), that $\psi_i^{(+)}(\mathbf{r})$ satisfies the Schrödinger equation, with $V(\mathbf{r})$ as the interparticle potential, for $|\mathbf{r}|>a$. The constants λ_{lm} will be chosen so that

$$\psi_i^{(+)}(\mathbf{r})=0, \quad |\mathbf{r}|\leqslant a. \tag{1.4}$$

It then remains to determine the required values of the λ_{lm}. To simplify the discussion, the particles will be assumed to be nonidentical and spinless. The introduction of spin and statistics will be considered later and presents no special difficulty.

If use is made of Eq. (1.3), Eq. (1.1) can be written in the coordinate representation as

$$\psi_i^{(+)}(\mathbf{r})=\phi_i(\mathbf{r})+\int d\mathbf{r}'G_i(\mathbf{r}|\mathbf{r}')V(\mathbf{r}')\psi_i^{(+)}(\mathbf{r}') +a^2\sum_{l,m}\lambda_{lm}g_l(r|a)Y_l^m(\mathbf{r},\mathbf{k}_i). \tag{1.5}$$

Here

$$G_i(\mathbf{r}|\mathbf{r}')=\frac{2M}{(2\pi)^3\hbar^2}\int d\mathbf{k}_p\frac{\exp[i\mathbf{k}_p\cdot(\mathbf{r}-\mathbf{r}')]}{k_i^2-k_p^2+i\epsilon} =\sum_l\left(\frac{2l+1}{4\pi}\right)^{\frac{1}{2}}g_l(r|r')Y_l^0(\mathbf{r},\mathbf{r}'), \tag{1.6}$$

$$g_l(r|r')=-i\frac{2Mk_i}{\hbar^2}\times\begin{cases}j_l(k_ir')h_l^{(1)}(k_ir), & r>r',\\ j_l(k_ir)h_l^{(1)}(k_ir'), & r<r',\end{cases} \tag{1.7}$$

j_l is the spherical Bessel function, and $h_l^{(1)}$ is the spherical Hankel function of the first kind.

If $\psi_i^{(+)}(\mathbf{r})$ is expanded in spherical harmonics, it follows from Eq. (1.5) that the coefficient $\psi_{lm}^{(+)}(r)$ of $Y_l^m(\mathbf{r},\mathbf{k}_i)$ in this series has the form

$$\psi_{lm}^{(+)}(r)=\delta_{0,m}i^l\left(\frac{2l+1}{2\pi^2}\right)^{\frac{1}{2}}j_l(k_ir)+\int d\mathbf{r}'g_{lm}(r|\mathbf{r}') \times V(\mathbf{r}')\psi_i^{(+)}(\mathbf{r}')+a^2\lambda_{lm}g_l(a|a), \tag{1.8}$$

[11] K. Huang and C. N. Yang, Phys. Rev. **105**, 767 (1952); R. Abe, Progr. Theoret. Phys. (Kyoto) **19**, 1 (1958).

[12] R. Abe, Progr. Theoret. Phys. (Kyoto) **19**, 699 (1958); K. Brueckner and J. L. Gammel, Phys. Rev. **109**, 1023 (1958).

[13] That is to say, the boundary conditions are satisfied both inside the core and at the core radius. In contrast to this, the point pseudopotential of Huang and Yang (reference 11) satisfies only the conditions at the core radius.

[14] B. Lippmann and J. Schwinger, Phys. Rev. **79**, 469 (1950).

where

$$g_{lm}(r|\mathbf{r}')=g_l(r|r')Y_l{}^m(\mathbf{r}',\mathbf{k}_i)^*. \tag{1.9}$$

The condition (1.4) at $|\mathbf{r}|=a$, when imposed on Eq. (1.8), determines the constants λ_{lm} uniquely. Then Eq. (1.8) may be written in the form

$$\psi_{lm}{}^{(+)}(r)=\delta_{0,m}i^l\left(\frac{2l+1}{2\pi^2}\right)^{\frac{1}{2}}\left[j_l(k_ir)-\frac{g_l(r|a)}{g_l(a|a)}j_l(k_ia)\right]$$
$$+\int d\mathbf{r}'\left[g_{lm}(r|\mathbf{r}')-\frac{g_l(r|a)}{g_l(a|a)}g_{lm}(a|\mathbf{r}')\right]$$
$$\times V(\mathbf{r}')\psi_i{}^{(+)}(\mathbf{r}'). \tag{1.10}$$

It follows, from Eqs. (1.6), (1.7), (1.9), and the last relation, that the wave function $\psi_i{}^{(+)}(\mathbf{r})$ whose radial part is given by Eq. (1.10) satisfies the condition (1.4); it is also a solution of the Schrödinger equation for $|\mathbf{r}|>a$ with the potential $V(\mathbf{r})$. Hence, $\psi_i{}^{(+)}(\mathbf{r})$ is the wave function for the case when the interaction consists of a h.c. and an ordinary local potential outside the core.

Equation (1.5) can now be written in the form

$$\psi_i{}^{(+)}(\mathbf{r})=\phi_i(\mathbf{r})+\int d\mathbf{r}'G_i(\mathbf{r}|\mathbf{r}')\Big\{V(\mathbf{r}')\psi_i{}^{(+)}(\mathbf{r}')$$
$$+\int d\mathbf{r}''[\omega_i(\mathbf{r}'|\mathbf{r}'')V(\mathbf{r}'')\psi_i{}^{(+)}(\mathbf{r}'')$$
$$+\Gamma_i(\mathbf{r}'|\mathbf{r}'')\phi_i(\mathbf{r}'')]\Big\}, \tag{1.11}$$

where ω_i and Γ_i are operators which have the following coordinate representatives:

$$\omega_i(\mathbf{r}'|\mathbf{r}'')=-\frac{\delta(r'-a)}{a^2}\sum_{l,m}Y_l{}^m(\mathbf{r}',\mathbf{k}_i)\frac{g_{lm}(a|\mathbf{r}'')}{g_l(a|a)}, \tag{1.12}$$

$$\Gamma_i(\mathbf{r}'|\mathbf{r}'')=-\frac{\delta(r'-a)\delta(r''-a)}{a^4}$$
$$\times\sum_{l,m}\frac{Y_l{}^m(\mathbf{r}',\mathbf{k}_i)Y_l{}^m(\mathbf{r}'',\mathbf{k}_i)^*}{g_l(a|a)}. \tag{1.13}$$

It then follows from Eq. (1.11) that, when the interaction contains a h.c., Eq. (1.1) has the form

$$\psi_i{}^{(+)}=\phi_i+G_i(U_i\psi_i{}^{(+)}+\Gamma_i\phi_i), \tag{1.14}$$

where

$$U_i=(1+\omega_i)V. \tag{1.15}$$

Similar results hold for the case of the incoming-wave solution $\psi_i{}^{(-)}$.

The preceding method can also be carried out with the standing-wave equation. In the absence of a h.c. interaction, the standing-wave vector ψ_i satisfies the integral equation[15]

$$\psi_i=\phi_i+\bar{G}_iV\psi_i, \tag{1.16}$$

where $\bar{G}_i$ is simply $P(E_i-H_0)^{-1}$ and P denotes the Cauchy principal value. Now assume that ψ_i satisfies the conditions (1.4) and that V has the form (1.2). Then ψ_i satisfies the modified standing-wave equation

$$\psi_i=\phi_i+\bar{G}_i(\bar{U}_i\psi_i+\bar{\Gamma}_i\phi_i), \tag{1.17}$$

where

$$\bar{U}_i=(1+\bar{\omega}_i)V. \tag{1.18}$$

Here $\bar{\omega}_i$ and $\bar{\Gamma}_i$ are operators which have coordinate representatives identical to those of ω_i and Γ_i, respectively, except for the replacement of $g_l(r|r')$ by $\bar{g}_l(r|r')$, where

$$\bar{g}_l(r|r')=\frac{2M}{\hbar^2}k_i\times\begin{cases}j_l(k_ir')n_l(k_ir), & r>r',\\ j_l(k_ir)n_l(k_ir'), & r<r',\end{cases} \tag{1.19}$$

and n_l is the spherical Neumann function; also, in the derivation of (1.19), $G_i(\mathbf{r}|\mathbf{r}')$ is replaced by

$$\bar{G}_i(\mathbf{r}|\mathbf{r}')=\frac{2M}{\hbar^2(2\pi)^3}P\int d\mathbf{k}_p\frac{\exp[i\mathbf{k}_p\cdot(\mathbf{r}-\mathbf{r}')]}{k_i{}^2-k_p{}^2}. \tag{1.20}$$

Equations (1.14) and (1.17) will now be used to derive the integral equations for the transition (t) and reactance (K) matrices. It is evident from Eq. (1.14), in the case when the potential V is absent, that $\psi_i{}^{(+)}$ is simply the h.c. wave function $\chi_i{}^{(+)}$:

$$\chi_i{}^{(+)}=\phi_i+G_i\Gamma_i\phi_i. \tag{1.21}$$

Now define an operator γ_i such that

$$\gamma_i\chi_i{}^{(+)}=\Gamma_i\phi_i. \tag{1.22}$$

Also define a wave matrix Ω by

$$\psi_i{}^{(+)}=\Omega\phi_i. \tag{1.23}$$

Then with the aid of Eqs. (1.21) and (1.22), it follows from Eqs. (1.14) and (1.23) that

$$\Omega=1+G_i[U_i+\gamma_i(1-G_iU_i)]\Omega. \tag{1.24}$$

The t matrix[16] will now be defined as

$$t\equiv[U_i+\gamma_i(1-G_iU_i)]\Omega. \tag{1.25}$$

Then it follows from Eqs. (1.22) and (1.24) that t satisfies the integral equation

$$t=U_i+\Gamma_i+U_iG_it. \tag{1.26}$$

In addition, it can be deduced from definition (1.25) in conjunction with Eq. (1.14) that

$$\langle f|t|i\rangle=\langle f|U_i\psi_i{}^{(+)}\rangle+\langle f|\Gamma_i|i\rangle. \tag{1.27}$$

It is easily seen from Eq. (1.14) that the scattering

[15] M. L. Goldberger, Phys. Rev. 84, 929 (1951).

[16] The customary superscript (+) will be left off t since only the t matrix defined with respect to $\psi^{(+)}$ will be used.

amplitude will be proportional to the OES matrix elements of t. Finally, it should be observed that Ω and t take on their customary forms when the h.c. radius is taken to be zero.

In the usual treatments of scattering theory without a h.c.,[14,15] the matrix elements of t and K are simply the representatives of the vectors to the right of G_i in the $\psi_i^{(+)}$ equation and $\bar{G}_i$ in the ψ_i equation, respectively. It is clear that the preceding definition of t has this property. Thus, by analogy, the matrix elements of K will be defined as

$$\langle f|K|i\rangle \equiv \langle f|\bar{U}_i\psi_i\rangle + \langle f|\bar{\Gamma}_i|i\rangle. \tag{1.28}$$

Then it follows from Eq. (1.17) that

$$K=\bar{U}_i+\bar{\Gamma}_i+\bar{U}_i\bar{G}_iK. \tag{1.29}$$

In the case of an ordinary interaction, the connection between t and K is easily derived and is unique if the solution of the standing-wave equation is unique.[15] This relationship is

$$t=K-i\pi K\delta(E_i-H_0)t. \tag{1.30}$$

However, when a h.c. is present, the situation is not so transparent, as a result of the rather unsymmetrical structure of the new equations for t and K. Nevertheless, Eq. (1.30) is true in this case also and under the same conditions. A proof of this is given in the Appendix.

There now remain the questions of spin and statistics. These will be examined in a fairly explicit way for the sake of completeness. Specifically, the two-nucleon problem will be considered with the interaction assumed charge independent.

The assumptions of Fermi statistics and of parity conservation imply with charge independence that ordinary spin is a good quantum number. In this case the matrix elements of the interaction potential V have the form[17]

$$\langle f,\beta,\nu'|V|i,\alpha,\mu'\rangle=\delta_{\nu',\mu'}\delta_{S\beta,S\alpha}\langle f,\beta,\mu'|V|i,\alpha,\mu'\rangle. \tag{1.31}$$

Here S_α is the total ordinary-spin quantum number corresponding to the spin state α. The interaction can then be written in the form

$$V=\sum_{T,S=0}^{1} V^{(T,S)}\Lambda_{(T,S)}, \tag{1.32}$$

where $\Lambda_{(T,S)}$ is the projection operator for a state with total i-spin T and total ordinary-spin S. The i-spin dependence of $V^{(T,S)}$ is limited only to T.

It is consistent with the preceding assumptions to introduce, into the interaction, hard cores with core radii $a(T,S)$ which depend only on the total spin quantum numbers T and S. Let $\langle \mathbf{r},\alpha,\mu'|\psi_i^{(+)}\rangle$ denote the coordinate representative of the (α,μ') component of $\psi_i^{(+)}$. Then the boundary conditions (1.4) take the form

$$\langle \mathbf{r},\alpha,\mu'|\psi_i^{(+)}\rangle=0, \quad |\mathbf{r}|\leqslant a(T_{\mu'},S_\alpha). \tag{1.33}$$

The method of replacing these boundary conditions by a singular potential can then be applied to the equations corresponding to a total spin state characterized by T and S exactly as in the spin-independent case.

The spin-dependent operators ω_i, $\bar{\omega}_i$, Γ_i, and $\bar{\Gamma}_i$ will only depend upon the total spin state (T,S). The coordinate representatives of these operators for a given total spin state will have the same forms as Eqs. (1.12) and (1.13) except that a is replaced by the $a(T,S)$ appropriate to the spin state in question.

The integral equations (1.26) and (1.29) for the operators t and K, respectively, will have the same form as in the spin-independent case. However, now, V, ω_i, $\bar{\omega}_i$, Γ_i, and $\bar{\Gamma}_i$ will have the spin and isotopic-spin dependences described above.

The integral equations associated with the matrix elements of t and K will next be expressed in antisymmetrized forms. This will be done in an explicit way for the t equation; an identical procedure will lead to corresponding results for the K equation.

Let ζ be an operator which exchanges all the variables describing the two nucleons and let[18]

$$\xi=1-\zeta. \tag{1.34}$$

Then the antisymmetrized t matrix is given by

$$t=\xi t^0, \tag{1.35}$$

where the superscript zero denotes the unsymmetrized quantity.

It follows from the previous remarks concerning the spin and isotopic-spin dependence of ω and Γ that U and Γ can be written in the form of Eq. (1.32).[19] Thus,[20]

$$\begin{aligned}\langle f,\beta,\nu'|\xi U|q,\alpha,\mu'\rangle &= \delta_{\nu',\mu'}\delta_{S\beta,S\alpha}\langle f,\beta,\mu'|U^{(T_{\mu'},S_\alpha)}\epsilon(T_{\mu'},S_\alpha)|q,\alpha,\mu'\rangle \\ &\equiv \delta_{\nu',\mu'}\delta_{S\beta,S\alpha}U(\mathbf{k}_f,\beta|\mathbf{k}_q,\alpha),\end{aligned} \tag{1.36}$$

where

$$\epsilon(T_{\mu'},S_\alpha)=[1-(-1)^{T_{\mu'}+S_\alpha}\Pi], \tag{1.37}$$

and Π is the space-exchange operator. Similarly,

$$\langle f,\beta,\nu'|\xi\Gamma|q,\alpha,\mu'\rangle\equiv\delta_{\nu',\mu'}\delta_{\beta,\alpha}\Gamma(\mathbf{k}_f|\mathbf{k}_q). \tag{1.38}$$

Finally, the assumed form of the interaction implies that

$$\langle f,\beta,\nu'|t|q,\alpha,\mu'\rangle\equiv\delta_{\nu',\mu'}\delta_{S\beta,S\alpha}\,t(\mathbf{k}_f,\beta|\mathbf{k}_q,\alpha). \tag{1.39}$$

[17] Henceforth, the following notations will be adopted. Latin indexes refer to momentum states. Unprimed Greek indexes denote ordinary-spin states and primed Greek indexes the i-spin states. The spin indexes take on the values -1, 0, 1, and 0′, the index 0′ referring to singlet states.

[18] It follows from the assumed form of V that V and ξ commute. It is easily shown that ω_i, $\bar{\omega}_i$, Γ_i, and $\bar{\Gamma}_i$ also commute with ξ.

[19] The energy index i will now be deleted from the interaction terms. The dependence of these operators on the incident energy has been established and the value of this energy will always be clear from the context.

[20] For simplicity, the dependence on the total spin state (T,S) has not been indicated explicitly in the definitions (1.36), (1.38), and (1.39). If it is kept in mind that any equation in which these quantities occur pertains to a definite total spin state, no confusion is likely to arise.

If the unsymmetrized equation (1.26) for t^0 and Eq. (1.35) are combined and use is made of the results of the preceding paragraph, then integral equations for the matrix elements of the antisymmetrized t matrix can be written in the form[21]

$$t(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)=U(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)+\delta_{\beta,\alpha}\Gamma(\mathbf{k}_f|\mathbf{k}_i)+\tfrac{1}{2}\sum_\gamma\int d\mathbf{k}_p\frac{U(\mathbf{k}_f,\beta|\mathbf{k}_p,\gamma)t(\mathbf{k}_p,\gamma|\mathbf{k}_i,\alpha)}{E_i-E_p+i\epsilon}. \quad (1.40)$$

The equations for the matrix elements of K corresponding to Eqs. (1.40) are, with analogous definitions,

$$K(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)=\bar{U}(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)+\delta_{\beta,\alpha}\bar{\Gamma}(\mathbf{k}_f|\mathbf{k}_i)+\tfrac{1}{2}\sum_\gamma P\int d\mathbf{k}_p\frac{\bar{U}(\mathbf{k}_f,\beta|\mathbf{k}_p,\gamma)K(\mathbf{k}_p,\gamma|\mathbf{k}_i,\alpha)}{E_i-E_p}. \quad (1.41)$$

Finally, the equations which relate the matrix elements of t to those of K are

$$t(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)=K(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)-\frac{i\pi}{2}\sum_\gamma\int d\mathbf{k}_pK(\mathbf{k}_f,\beta|\mathbf{k}_p,\gamma)\times\delta(E_i-E_p)t(\mathbf{k}_p,\gamma|\mathbf{k}_i,\alpha). \quad (1.42)$$

Equations (1.41) constitute the set of relations which will be studied in the remainder of this paper. Direct consideration of the integral equations (1.40) for t is cumbersome as a result of the complex propagator. On the other hand, Eqs. (1.41) for K can easily be reduced to real form; the solutions of the K equations can then be related to the matrix elements of t by use of (1.42).

II. PARTIAL-WAVE ANALYSIS

The integral equations (1.41) are not in forms amenable to solution. As a first step in obtaining more convenient expressions, a partial-wave analysis of these equations will be performed. This type of reduction is useful only if relatively few partial waves are important in the scattering; however, this is the case for all energies at which phenomenological nucleon-nucleon interactions have proved successful.

Let $\mathbf{J}$ and $\mathbf{S}$ denote the total and spin angular momentum operators, respectively, for the two-nucleon system. Also, let A be an operator which conserves $\mathbf{J}^2$, $\mathbf{S}^2$, and parity, and which is independent of the orientation of $\mathbf{J}$. The matrix elements of A with respect to the free two-particle states will be written as

$$A(\mathbf{k}_f,\beta|\mathbf{k}_q,\alpha)\equiv\langle f,\beta|A|q,\alpha\rangle=\frac{1}{(2\pi)^3}\int d\mathbf{r}\exp[-i\mathbf{k}_f\cdot\mathbf{r}]\chi_\beta^\dagger A\exp[i\mathbf{k}_q\cdot\mathbf{r}]\chi_\alpha. \quad (2.1)$$

[21] The factor of $\frac{1}{2}$ arises from the fact that $\xi^2=2\xi$.

Here, χ_α is the two-nucleon spin function. The axis of quantization will always be taken in the direction of $\mathbf{k}_i$; the component of total angular momentum in this direction is denoted by J_i with eigenvalues $\hbar M$. Until mentioned to the contrary, only those matrix elements between triplet states will be considered; the singlet case will prove to be relatively simple. Finally, the isotopic spin is neglected since it enters into the equations of interest in a trivial manner.

Let F_{Jl}^M denote the eigenfunctions of $\mathbf{S}^2$, $\mathbf{l}^2$, $\mathbf{J}^2$, and J_i for which $S=1$. The orbital angular momentum operator is denoted by $\mathbf{l}$. Then the matrix elements, $A_{l'l}^J$, of A will be defined by the equation

$$\delta_{J,J'}\delta_{m'+\beta,m+\alpha}A_{l'l}^J(k_f|k_q)\equiv\int d\mathbf{r}j_{l'}(k_fr)(F_{J'l'}^{m'+\beta})^\dagger AF_{Jl}^{m+\alpha}j_l(k_qr). \quad (2.2)$$

Note that $A_{l'l}^J$ is independent of $m+\alpha$.

Equation (2.1) can be written in terms of the $A_{l'l}^J$ by expressing $e^{i\mathbf{k}\cdot\mathbf{r}}\chi$ in terms of the F_{Jl}^M and employing Eq. (2.2).[22] One finds

$$A(\mathbf{k}_f,\beta|\mathbf{k}_q,\alpha)=(2/\pi)\sum_{l,l',m}Y_l^m(\mathbf{k}_q,\mathbf{k}_i)^*Y_{l'}^{m+\alpha-\beta}(\mathbf{k}_f,\mathbf{k}_i)\delta_{l,l'}\times\sum_{J=|l-1|}^{l+1}\bar{C}_{l'l}(J,m;\beta,\alpha)A_{l'l}^J(k_f|k_q), \quad (2.3)$$

where

$$\bar{\delta}_{l,l'}=\delta_{l,l'}-\delta_{|l'-l|,2}, \quad (2.4)$$

$$\bar{C}_{l'l}(J,m;\beta,\alpha)=C_{l'1}(J,m+\alpha;m+\alpha-\beta,\beta)\times C_{l1}(J,m+\alpha;m,\alpha), \quad (2.5)$$

and $C_{jj'}(J,m;\beta,\alpha)$ is the Clebsch-Gordan coefficient. If $A(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)$ is expanded in partial waves, viz.,

$$A(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)=\sum_{l,m}A_{lm}^{\beta\alpha}(k_f|k_i)Y_l^m(\mathbf{k}_f,\mathbf{k}_i), \quad (2.6)$$

then it is apparent from Eq. (2.3) that

$$A_{lm}^{\beta\alpha}(k_f|k_i)=\delta_{m,\alpha-\beta}A_{l,\alpha-\beta}^{\beta\alpha}(k_f|k_i)\equiv\delta_{m,\alpha-\beta}At^{\beta\alpha}(k_f|k_i). \quad (2.7)$$

Consider an integral equation of the form

$$U(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)=V(\mathbf{k}_f,\beta|\mathbf{k}_i,\alpha)+\sum_\gamma\int d\mathbf{k}_pW(\mathbf{k}_f,\beta|\mathbf{k}_p,\gamma)\times U(\mathbf{k}_p,\gamma|\mathbf{k}_i,\alpha), \quad (2.8)$$

where U, V, and W are any operators which satisfy all the conditions imposed on the operator A. The partial-wave form of Eq. (2.8) is obtained immediately with the aid of the relevant forms of Eqs. (2.3), (2.6), and (2.7). Then, if the indicated angular integration is carried out and the necessary sums performed, it is

[22] J. Blatt and V. Weisskopf, *Theoretical Nuclear Physics* (John Wiley & Sons, Inc., New York, 1952), Appendix A.

found that

$$U_l^{\beta\alpha}(k_f|k_i) = V_l^{\beta\alpha}(k_f|k_i)+(2/\pi)\sum_{\gamma,l'}\bar{\delta}_{l,l'}\sum_{J=|l'-1|}^{l'+1}\bar{C}_{ll'}(J,\alpha-\gamma;\beta,\gamma) \times\int_0^\infty dk_p k_p^2 W_{ll'}^J(k_f|k_p)U_{l'}^{\gamma\alpha}(k_p|k_i). \quad (2.9)$$

Although Eqs. (2.9) are obviously angle independent, their form is still inconvenient since they do not constitute a closed set. A further reduction can be accomplished if both sides of Eq. (2.9) are multiplied by $C_{l''1}(J,\alpha;\ 0,\alpha)C_{l1}(J,\alpha;\ \alpha-\beta,\ \beta)$, summed over α and β, and use is made of Eqs. (2.3) and (2.7). The application of the symmetry properties of the Clebsch-Gordan coefficients[22] then yields the following angle-independent form of Eq. (2.8):

$$\bar{\delta}_{l'',l}[U_{ll''}^J(k_f|k_i)-V_{ll''}^J(k_f|k_i)] = (2/\pi)\sum_{l'}\bar{\delta}_{l,l'}\bar{\delta}_{l',l''}\int_0^\infty dk_p k_p^2 \times W_{ll'}^J(k_f|k_p)U_{l'l''}^J(k_p|k_i). \quad (2.10)$$

Equations (1.41) and (1.42) in the triplet case have the same form as Eq. (2.8) and the operators $\bar{U}$, $\bar{\Gamma}$, K, and t all satisfy the conditions which were imposed on the operator A. The latter fact is evident from the assumptions which were made on the two-nucleon potential. Therefore, the integral equation for the matrix elements of K and the equation which links the matrix elements of t and K can be put into the form (2.10). It is then found that[23]

$$K_{JJ}^J(k_f|k_i)=\bar{U}_{JJ}^J(k_f|k_i)+\bar{\Gamma}_{JJ}^J(k_f|k_i)+\frac{1}{\pi}P\int_0^\infty dk_p k_p^2\frac{\bar{U}_{JJ}^J(k_f|k_p)K_{JJ}^J(k_p|k_i)}{E_i-E_p}, \quad (2.11a)$$

$$K_{J\pm1,J+1}^J(k_f|k_i)=\bar{U}_{J\pm1,J+1}^J(k_f|k_i)+\delta_{J\pm1,J+1}\bar{\Gamma}_{J+1,J+1}^J(k_f|k_i)+\frac{1}{\pi}P\int_0^\infty\frac{dk_p k_p^2}{E_i-E_p} \times[\bar{U}_{J\pm1,J+1}^J(k_f|k_p)K_{J+1,J+1}^J(k_p|k_i)+\bar{U}_{J\pm1,J-1}^J(k_f|k_p)K_{J-1,J+1}^J(k_p|k_i)], \quad (2.11b)$$

$$K_{J\pm1,J-1}^J(k_f|k_i)=\bar{U}_{J\pm1,J-1}^J(k_f|k_i)+\delta_{J\pm1,J-1}\bar{\Gamma}_{J-1,J-1}^J(k_f|k_i)+\frac{1}{\pi}P\int_0^\infty\frac{dk_p k_p^2}{E_i-E_p} \times[\bar{U}_{J\pm1,J+1}^J(k_f|k_p)K_{J+1,J-1}^J(k_p|k_i)+\bar{U}_{J\pm1,J-1}^J(k_f|k_p)K_{J-1,J-1}^J(k_p|k_i)]. \quad (2.11c)$$

The corresponding forms for Eq. (1.42) can be obtained by comparison with Eqs. (1.41) and (2.11) and can be reduced to a set of algebraic equations if the integration over energy is carried out.

Now the matrix element $A_{0'0'}$ of the operator A between singlet states can be written as

$$A_{0'0'}(\mathbf{k}_f|\mathbf{k}_q)=\sum_l A_l(k_f|k_q)Y_l^0(\mathbf{k}_f,\mathbf{k}_q). \quad (2.12)$$

Consider the integral equation (2.8) in the singlet case. Then, since all the matrix elements have the form (2.12), the angular variables are easily eliminated and Eq. (2.8) reduces to

$$U_l(k_f|k_i)=V_l(k_f|k_i)+\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}\int_0^\infty dk_p k_p^2 \times W_l(k_f|k_p)U_l(k_p|k_i). \quad (2.13)$$

Thus, in the singlet case, Eq. (1.41) reduces to

$$K_l(k_f|k_i)=\bar{U}_l(k_f|k_i)+\bar{\Gamma}_l(k_f|k_i)+\left(\frac{\pi}{2l+1}\right)^{\frac{1}{2}} \times P\int_0^\infty dk_p k_p^2\frac{\bar{U}_l(k_f|k_p)K_l(k_p|k_i)}{E_i-E_p}. \quad (2.14)$$

Also, Eq. (1.42) becomes a single equation which may be written as an algebraic equation just as in the triplet case.

III. REDUCTION TO FREDHOLM FORM

The integral equations (2.11) and (2.14) are singular as a result of the pole in the kernel at k_i. It is difficult, in general, to prescribe a method of solution for such equations as they stand. A standard technique, however, for the treatment of singular integral equations is to attempt a reduction to a Fredholm form.[24] The usual methods of solution and approximation can then be employed.[25] The reduction of the integral equations for the K matrices to Fredholm form is the object of this section.

The basic assumptions which will be made is that the integral equations (2.11) and (2.14) have solutions for $k_f=k_i$; also, that these solutions are known.[26] In the latter case, it is possible to perform a reduction which is much simpler than if this assumption were not true. Moreover, the solutions for arbitrary k_f can then be

[23] It should be remarked that the relation $K_{J+1,J-1}^J=K_{J-1,J+1}^J$, which results from the symmetric character of the S matrix, holds only on the energy shell.

[24] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff, Ltd., Groningen, Holland, 1953).

[25] F. Smithies, *Integral Equations* (Cambridge University Press, New York, 1958).

[26] By Eq. (1.42), this is equivalent to assuming a knowledge of the scattering amplitude at the energy in question.

expressed in terms of those for $k_f=k_i$. It is just this form of reduction which is desired.

Consider an integral equation of the form[27]

$$R(x|y)=f(x|y)+P\int\frac{z^2\xi(x|z)}{z^2-y^2}R(z|y)dz, \quad (3.1)$$

where, in general, $R(x|y)$ and $f(x|y)$ will be regarded as column matrices and $\xi(x|z)$ as a square matrix.[28] Equations (2.11a) and (2.14) each correspond to an equation of the form (3.1) where the function matrices are one dimensional. Equations (2.11b) and (2.11c) are two independent sets of two coupled integral equations, each set corresponding to an equation of the type (3.1) where ξ is a 2×2 matrix and both R and f have two components.

It will be assumed that $f(x|y)$, $\xi(x|z)$, and $\partial\xi(x|z)/\partial z$ are continuous throughout the domain $0\leqslant(x,z)<\infty$; also, that a continuous solution, $R(x|y)$, of (3.1) exists which vanishes for $x\rightarrow\infty$.[29] This behavior for $x\rightarrow\infty$ will be regarded as sufficient to ensure the existence of the integral

$$P\int\frac{R(x|y)}{x-y}dx.$$

The determinant of $\xi(y|y)$ will be assumed nonzero and the function $R(y|y)$ will be regarded as known.

Let

$$T(x|z)\equiv z^2\xi(x|z)/(z+y), \quad (3.2)$$

and

$$h(x|z)\equiv[T(x|z)-T(x|y)]/(z-y). \quad (3.3)$$

Then Eq. (3.1) may be written as[30]

$$R(x|y)=f(x|y)+T(x|y)P\int\frac{R(z|y)}{z-y}dz+\int h(x|z)R(z|y)dz. \quad (3.4)$$

An expression for the principal-value integral can be found by setting $x=y$ in this equation. If this expression is then substituted into Eq. (3.4), the following Fredholm form of Eq. (3.1) is obtained:

$$R(x|y)=[f(x|y)-\tau(x|y)f(y|y)]+\tau(x|y)R(y|y)+\int\Lambda(x|z)R(z|y)dz, \quad (3.5)$$

where

$$\tau(x|z)=T(x|z)T^{-1}(y|z), \quad (3.6)$$

$$\Lambda(x|z)=h(x|z)-\tau(x|y)h(y|z). \quad (3.7)$$

The new kernel $\Lambda(x|z)$ contains no poles.[31]

The solution of Eq. (3.5) can be written in the form

$$R(x|y)=\varphi(x|y)R(y|y)+\theta(x|y), \quad (3.8)$$

where φ and θ satisfy the integral equations

$$\varphi(x|y)=\tau(x|y)+\int\Lambda(x|z)\varphi(z|y)dz, \quad (3.9)$$

$$\theta(x|y)=[f(x|y)-\tau(x|y)f(y|y)]+\int\Lambda(x|z)\theta(z|y)dz. \quad (3.10)$$

It is clear that

$$\varphi(y|y)=1, \quad (3.11a)$$

$$\theta(y|y)=0. \quad (3.11b)$$

The separation of the integral equation for $R(x|y)$ into two integral equations was done primarily for clarity. However, under some circumstances this separation may be convenient for obtaining an approximate solution. In typical physical situations, the inhomogeneous term in the θ equation is small or zero.[32] If this term vanishes, and if the solution $R(x|y)$ is unique, then θ must be identically zero. Otherwise, the fact that the inhomogeneous term is small can be useful if it is possible to solve Eq. (3.5) by iteration. The φ and θ equations can then be iterated separately with less iterations being needed for θ.

IV. ON THE SOLUTION OF THE INTEGRAL EQUATIONS

The results of the last section permit the reduction of the integral equations (2.11) and (2.14) to the Fredholm form (3.5). Once this is done, all matters of principle concerning the nature of the solutions of these equations as well as the applicability of various means of obtaining approximate solutions are taken care of by the usual Fredholm theory.[25] With this in mind, some techniques for obtaining solutions will now be investigated.

For the sake of clarity, only equations of the singlet (uncoupled) type will be considered. The functional properties which will be assumed or deduced for the various quantities appearing in the uncoupled equations are valid for the matrix counterparts of these quantities which occur in the coupled case. The extension to the

[27] Henceforth, the limits on all integrals will be assumed to run from 0 to ∞, unless indicated explicitly to the contrary.

[28] These matrices are assumed to be of finite dimensions, with matrix elements which have a functional dependence on x, y, and z.

[29] Most of the results to be derived hold for much weaker conditions on f and ξ and with R a member of a broader class of functions. The case considered, however, covers many physical situations.

[30] The principal-value symbol is omitted on the second integral because $h(x|z)$ is continuous at $z=y$.

[31] This particular reduction apparently does not hold if the determinant of $\xi(y|y)$ vanishes. However, it is not expected that this limitation on $\xi(y|y)$ will be unduly restrictive in physical applications. In any case, it can be shown that, if the OES matrix elements $R(y|y)$ are treated on the same footing as the FES matrix elements (i.e., instead of determining $R(y|y)$ experimentally, it too is derived from a potential), then no restriction on $\det\xi(y|y)$ need be imposed.

[32] This term is identically zero if no h.c. is present.

coupled case is then straightforward because of the *formal* identity of the uncoupled and coupled equations as was illustrated in Sec. III.

In the singlet state the interaction potential is central. Thus the partial-wave amplitudes of the matrix elements $\bar{U}(\mathbf{k}_f|\mathbf{k}_i)$ and $\bar{\Gamma}(\mathbf{k}_f|\mathbf{k}_i)$ can be written in the form[33]

$$\bar{U}_l(k_f|k_p)=2[(2l+1)\pi^{-3}]^{\frac{1}{2}}u_l(k_f|k_p), \tag{4.1}$$

and

$$\bar{\Gamma}_l(k_f|k_i)=-(2E_i/k_i^3)[(2l+1)\pi^{-3}]^{\frac{1}{2}} \times[j_l(k_fa)/n_l(k_ia)], \tag{4.2}$$

where

$$u_l(k_f|k_p)=\int_a^\infty j_l(k_fr)V(r)j_l(k_pr)r^2dr - \frac{j_l(k_fa)}{n_l(k_ia)}\int_a^\infty n_l(k_ir)V(r)j_l(k_pr)r^2dr, \tag{4.3}$$

and a is the h.c. radius. On comparing Eq. (2.14) with Eq. (3.1) and noting the definitions (3.6) and (3.7), it is found that

$$\tau_l(k_f|k_p)=u_l(k_f|k_p)/u_l(k_i|k_p), \tag{4.4}$$

$$f_l(k_f|k_i)-\tau_l(k_f|k_i)f_l(k_i|k_i) =\bar{\Gamma}_l(k_i|k_i)\{[j_l(k_fa)/j_l(k_ia)]-\tau_l(k_f|k_i)\}, \tag{4.5}$$

and

$$\Lambda_l(k_f|k_p)=-(4M/\pi\hbar^2)[\tau_l(k_f|k_p)-\tau_l(k_f|k_i)] \times u_l(k_i|k_p)k_p^2(k_p^2-k_i^2)^{-1}. \tag{4.6}$$

Equations (4.4)–(4.6) together with Eq. (3.5) define the Fredholm form of Eq. (2.14) for the function $K_l(k_f|k_i)$.

If no h.c. is contained in the interaction, the integral equation for $K_l(k_f|k_i)$ is somewhat easier to study. This restriction on the interaction will be made in most of the detailed considerations to follow. However, the effects of introducing a h.c. will be discussed when appropriate.

Without a h.c., it is sufficient to consider the function $\varphi_l(k_f|k_i)$ which is defined by Eqs. (3.8) and (3.9) with R replaced by K_l, and τ_l and Λ_l given by Eqs. (4.4) and (4.6), respectively. Since $\varphi_l(k_f|k_i)$ is a dimensionless quantity, it will prove convenient to write Eq. (3.9) in a manifestly dimensionless form. To accomplish this, first introduce the dimensionless variables

$$s=k/k_i, \quad \bar{r}=k_ir, \tag{4.7}$$

and, secondly, write

$$V(r)=V_0F(\bar{r}), \tag{4.8}$$

where V_0 is a constant with the dimensions of energy. Then, for this case, Eq. (3.9) can be rewritten in the form[34]

$$\varphi_l(s|1)=\tau_l(s|1)-\lambda\int\frac{s'^2}{s'^2-1}[\tau_l(s|s')-\tau_l(s|1)] \times\bar{u}_l(1|s')\varphi_l(s'|1)ds', \tag{4.9}$$

where

$$\bar{u}_l(s|s')=\int j_l(s\bar{r})F(\bar{r})j_l(s'\bar{r})\bar{r}^2d\bar{r}, \tag{4.10}$$

and

$$\lambda=(2V_0/\pi E_i). \tag{4.11}$$

The remainder of this section is devoted primarily to the discussion of two methods of obtaining approximate solutions of Eq. (4.9). First, the iteration or Neumann series solution is investigated. Second, a study is made of the possibility of replacing Eq. (4.9) by an integral equation which is exactly soluble and whose solution closely approximates the actual solution of Eq. (4.9).

Let us consider the iteration series derived from Eq. (4.9). Here, the essential problem is to determine when this series actually represents the solution of Eq. (4.9). A secondary question is the matter of the rapidity of the convergence of this series assuming that it does represent the solution. Associated with the latter problem is the determination of when the Born approximation provides a good representation of the actual solution.[35]

Instead of examining Eq. (4.9) directly, it is convenient to consider the integral equation satisfied by the Hankel transform, Φ_l, of φ_l, where

$$\Phi_l(w)=\int j_l(ws)\varphi_l(s|1)s^2ds, \tag{4.12a}$$

$$\varphi_l(s|1)=\frac{2}{\pi}\int j_l(ws)\Phi_l(w)w^2dw. \tag{4.12b}$$

If Eqs. (4.12a), (4.12b), and (4.9) are combined, it is found that

$$\Phi_l(w)=\frac{\pi F(w)j_l(w)}{2\bar{u}_l(1|1)}+\tfrac{1}{2}\pi\lambda\int z^2F(w) \times\left[g_l'(w|z)-\frac{j_l(w)}{\bar{u}_l(1|1)}\int j_l(t)g_l'(z|t)F(t)t^2dt\right] \times\Phi_l(z)dz, \tag{4.13}$$

where

$$g_l'(w|z)=\begin{cases} j_l(z)n_l(w), & w>z, \\ j_l(w)n_l(z), & w<z. \end{cases} \tag{4.14}$$

Let $k_l(w|z)$ denote the kernel of Eq. (4.13). Then an iteration solution of Eq. (4.13) exists, is unique, and

[33] The numerical factors resulting from the antisymmetrization of states are included. Along with this, l is assumed to be always even or odd.

[34] To simplify the notation, $\varphi_l(k_f|k_i)$, for example, has been written as $\varphi_l(s|1)$, etc.

[35] In this paper the Born approximation is taken to mean the zeroth order iteration of a given integral equation.

converges uniformly if[36]

$$B_l \equiv \int dw \int dz\, |k_l(w|z)|^2 \leqslant \left(\frac{E_i}{V_0}\right)^2. \quad (4.15)$$

The uniform convergence permits the integration of the iteration series for Φ_l term-by-term to yield the corresponding series for φ_l.

A straightforward way to gain some insight into the behavior of B_l as a function of energy is to assume a form for the interaction $F(w)$ and then carry out the integration in (4.15) for a significant range of energies. For a square well, i.e.,

$$F(w)=\begin{cases}1, & w<\rho,\\ 0, & w>\rho,\end{cases} \quad (4.16)$$

where $\rho=Rk_i$ and R is the range of the well, $k_l(w|z)$ is simply

$$k_l(w|z) = z^2\left[g_l'(w|z)-\frac{j_l(w)}{\bar{u}_l(1|1)}\int_0^\rho j_l(t)g_l'(z|t)t^2dt\right], \quad (4.17a)$$

for $(w,z)<\rho$ and

$$k_l(w|z)=0 \quad (4.17b)$$

otherwise. In particular, for $l=0$, Eq. (4.17a) becomes

$$k_0(w|z)=(z/2w)\{[\sin|z-w|-\sin(z+w)] -[2\sin w/\bar{u}_l(1|1)] \times[z-(\cos\rho+\rho\sin\rho)\sin z]\}. \quad (4.17a')$$

Now, for the interaction (4.16), Eq. (4.15) may be rewritten as

$$B_l/\rho^4 \leqslant (\hbar^2/2MR^2V_0)^2. \quad (4.18)$$

If the kernel (4.17a′) is used to compute B_0, it is found that

$$\lim_{\rho\to 0}(B_0/\rho^4)\approx 0.23, \quad (4.19a)$$

and

$$B_0/\rho^4 \sim (5\pi/12\rho)(5-4\sin^2\rho), \quad (\rho\to\infty). \quad (4.19b)$$

From Eqs. (4.19a) and (4.18) it is clear that the iteration solution can exist at zero incident energy for nonzero V_0; indeed, with $R=2.6$ f the iteration solution exists provided $|V_0|$ is less than about 12 Mev. Also, from Eq. (4.19b) it follows that for high enough energy the iteration solution will always exist; however, for typical potentials this energy may be so high as to be of no practical interest.

In order to obtain more detailed information on the iteration solution, the quantity B_0 was computed for laboratory energies from 0 to 500 Mev, assuming a square-well interaction. The well parameters were chosen to be those which fit the low-energy proton-proton data ($|V_0|=13.3$ Mev, $R=2.6$ f). Since Eq. (4.15) is only a sufficient condition for the existence of the iteration solution, just those energies where the series is definitely a solution of Eq. (4.13) were established. The iteration series was found to be a valid solution in the two (approximate) intervals 0–118 and 200–425 Mev; the most rapid rates of convergence occurred in the lower-energy interval.

It can be concluded from the preceding results that the iteration method will most probably be unjustified for the interactions and energies of interest in the nucleon-nucleon scattering problem. The presence of a h.c. in the interaction is unlikely to alter this conclusion.

The statements made above concerning the validity and convergence of the iteration series refer to the entire domain of the independent variable. However, it is evident from Eq. (4.9) that, in the neighborhood of $s=1$, the solution consists predominantly of the inhomogeneous term $\tau_l(s|1)$. Hence, there exists a neighborhood of $s=1$ where the Born approximation is valid independently of whether or not the iteration solution is valid in the whole domain of s. It is then of interest to investigate in how large a neighborhood of $s=1$ will the solution be given accurately by the Born approximation.

The validity of the Born approximation can be easily checked if the exact solution φ_l is known for a particular problem. In the case of a square-well interaction, φ_l is known, e.g., for $l=0$,

$$\varphi_0(s|1)=(1/s)\{j_0[(s-\bar{\alpha})\rho]-j_0[(s+\bar{\alpha})\rho]\} \times\{j_0[(1-\bar{\alpha})\rho]-j_0[(1+\bar{\alpha})\rho]\}^{-1}, \quad (4.20)$$

where

$$\bar{\alpha}=[(E_i+V_0)/E_i]^{\frac{1}{2}}. \quad (4.21)$$

The inhomogeneous term $\tau_0(s|1)$ may be obtained from (4.20) by setting $\bar{\alpha}$ equal to unity.

The exact solution $\varphi_0(s|1)$ and the Born approximation $\tau_0(s|1)$ were calculated using the same well parameters as before at laboratory energies of 40, 118, and 150 Mev in the interval $0\leqslant s\leqslant 2.0$. At the first energy, the iteration series definitely represents the solution of the integral equation in question. The second energy represents a transition point between regions of energy for which the iteration series is a solution and where it is possibly not. At 150 Mev nothing is known concerning the validity of the iteration series.

The intervals about $s=1$ for which

$$|\varphi_0-\tau_0|/|\varphi_0| \leqslant 0.10$$

are (0–1.9), (0.7–1.3), (0.8–1.2) for the energies 40, 118, and 150 Mev, respectively. The agreement between φ_0 and τ_0 is poorest at the end points of the intervals considered; however, for $s\geqslant 2.0$ the magnitudes of both φ_0 and τ_0 are small compared to unity so that the difference between them is unlikely to be significant in physical applications.

[36] See reference 25, p. 29.

For $l>0$, the agreement between φ_l and τ_l is likely to be better than in the $l=0$ case. This is a result of the fact[37] that both φ_l and τ_l are zero for $s=0$ when $l\neq 0$. Thus, there are two points in the interval (0–2.0) where the Born approximation is strictly valid.

It can be concluded, at least for a square-well potential, that the Born approximation is valid in a fairly large energy interval (on the order of E_i) about the energy shell. There is little in the structure of the integral equation (4.9) to suggest that this conclusion will not hold for any of the typical nucleon-nucleon potentials.

If a h.c. is present in the interaction, it is necessary to consider, in addition to φ_l, the function θ_l which is defined by Eqs. (3.8) and (3.10). Also, $\bar{u}_l(s|s')$ will now contain an additional term corresponding to the last term of Eq. (4.3). However, from Eqs. (3.11a) and (3.11b) it is clear that there exists a neighborhood of $s=1$ where the Born approximation is valid for each of the integral equations satisfied by φ_l and θ_l. From the form of these integral equations, there is again no reason to expect much deviation from the results quoted above for the case that a h.c. is absent. The only significant difference from the behavior of φ_l without a h.c. is that φ_l and θ_l will drop off more slowly for large s. This will be discussed later.

A second method of obtaining an approximate solution of Eq. (4.9) will now be investigated. Consider the integral, $H_l(s)$, in Eq. (4.9),

$$H_l(s)\equiv\int\frac{s'^2}{s'^2-1}[\tau_l(s|s')-\tau_l(s|1)]\times\bar{u}_l(1|s')\varphi_l(s'|1)ds'. \quad (4.22)$$

The bracketed term under the integral sign in Eq. (4.22) is, in general, an inseparable function of s and s'. If this term could be replaced by a separable function, the integral equation (4.9) would be exactly soluble. The possibility of making such a replacement will now be examined.

Now it is much simpler to approximate a function on a finite rather than on an infinite domain. Accordingly, make the change of variable

$$s=(1-\zeta)/\zeta. \quad (4.23)$$

Then Eq. (4.22) can be written as

$$H_l[s(\zeta)]=-\tfrac{1}{2}\int_0^1\frac{(1-\zeta')^2}{(\zeta'-\tfrac{1}{2})}\{\tau_l[s(\zeta)|s'(\zeta')]-\tau_l[s(\zeta)|s(\tfrac{1}{2})]\}\times\bar{u}_l[s(\tfrac{1}{2})|s'(\zeta')]\varphi_l[s'(\zeta')|s(\tfrac{1}{2})]\frac{d\zeta'}{\zeta'^2}. \quad (4.24)$$

Assume for the moment that $\tau_l[s(\zeta)|s'(\zeta')]$ admits of a power series expansion in ζ' about the point $\zeta'=\frac{1}{2}$. The lowest-order approximation to $H_l[s(\zeta)]$ is then obtained by retaining the first two terms of this series, viz.,

$$H_l[s(\zeta)]\approx-\tfrac{1}{2}\int_0^1(1-\zeta')^2\left\{\frac{\partial}{\partial\zeta'}\tau_l[s(\zeta)|s'(\zeta')]\right\}_{\zeta'=\frac{1}{2}}\times\bar{u}_l[s(\tfrac{1}{2})|s'(\zeta')]\varphi_l[s'(\zeta')|s(\tfrac{1}{2})]\frac{d\zeta'}{\zeta'^2}. \quad (4.25a)$$

The retention of the first two terms in the series for τ_l is consistent with the assumptions given in the paragraph immediately following Eq. (3.1) on the functions entering into the integral equations of interest.

If Eq. (4.23) is employed, Eq. (4.25a) may be transformed back to the infinite domain with the result that

$$H_l(s)\approx\int\frac{2s'^2}{(s'+1)^2}F_l(s|1)\bar{u}_l(1|s')\varphi_l(s'|1)ds', \quad (4.25b)$$

where

$$F_l(s|s')=[\tau_l(s|s')-\tau_l(s|1)](s'-1)^{-1}. \quad (4.26)$$

With the approximation (4.25b) for $H_l(s)$, the integral equation (4.9) is exactly soluble [see Eq. (4.34) below].

A justification for making approximation (4.25b) will now be given. First, introduce the quantities

$$\bar{F}_l(s|s')=\tfrac{1}{2}(s'+1)[F_l(s|s')-2(s'+1)^{-1}F_l(s|1)], \quad (4.27a)$$

and

$$\bar{\varphi}_l(s|1)=2s^2(s+1)^{-2}\bar{u}_l(1|s)\varphi_l(s|1). \quad (4.27b)$$

Equation (4.22) becomes, with the use of Eqs. (4.26) and (4.27),

$$H_l(s)=\int[\bar{F}_l(s|s')+F_l(s|1)]\bar{\varphi}_l(s'|1)ds'. \quad (4.28)$$

The approximation (4.25b) consists in neglecting $\bar{F}_l(s|s')$ in this last equation.

Some functional properties of the terms under the integral sign in Eq. (4.28) will now be noted. For all s', $\bar{F}_l(s|s')$ is bounded; also,

$$\bar{F}_l(s|1)=0. \quad (4.29)$$

For typical short-range potentials it is easy to deduce that

$$\bar{u}_l(1|s)=O(1/s^2), \quad (s\to\infty), \quad (4.30)$$

and

$$\varphi_l(s|1),\quad \bar{u}_l(1|s)=O(s^l), \quad (s\to 0). \quad (4.31)$$

Equations (4.29)–(4.31) are valid when a h.c. is present; in this case $\bar{u}_l(1|s)$ contains terms corresponding to both the terms in Eq. (4.3).

When the interaction does not contain a h.c., $\varphi_l(s|1)=K_l(k_f|k_i)/K_l(k_i|k_i)$; it is then simple to deduce that, with a typical short-range potential,

$$\varphi_l(s|1)=O(1/s^2), \quad (s\to\infty). \quad (4.32)$$

[37] This is valid for any potential whether or not a h.c. is present.

From the asymptotic properties (4.30)–(4.32) and Eq. (3.11a) it follows that

$$\bar{\varphi}_l(s|1)=O(1/s^4), \quad (s\to\infty), \tag{4.33a}$$

$$\bar{\varphi}_l(s|1)=O(s^{2l+2}), \quad (s\to 0), \tag{4.33b}$$

and

$$\bar{\varphi}_l(s|1)\approx\tfrac{1}{2}\bar{u}_l(1|s), \quad (s\approx 1). \tag{4.33c}$$

Finally, it will be assumed that the difference between the maximum value of $\bar{u}_l(1|s)$ and $\bar{u}_l(1|1)$ is small. This assumption can be shown to be fairly good for square and Yukawa wells with or without a h.c.

It seems reasonable to conclude from the properties (4.33), along with the assumption made on $\bar{u}_l(1|s')$, that $\bar{\varphi}_l(s'|1)$ will have its maximum value in a neighborhood of $s'=1$. However, it is in this region where $\bar{F}_l(s|s')/F_l(s|1)$ is smallest. Therefore, the value of $H_l(s)$ will probably not be altered much if $\bar{F}_l(s|s')$ is dropped from the integrand. If this approximation is made, the solution, $\varphi_l{}^a$, of the approximate integral equation is

$$\varphi_l{}^a(s|1)=\tau_l(s|1)-\frac{2\lambda F_l(s|1)}{\bar{u}_l(1|1)}\left\{\int\left[\frac{s'}{s'+1}\bar{u}_l(1|s')\right]^2 ds'\right\}$$
$$\times\left[1+2\lambda\int\frac{s'^2}{(s'+1)^2}F_l(s'|1)ds'\right]^{-1}. \tag{4.34}$$

The approximate solution $\varphi_l{}^a$ was computed using the square-well potential discussed previously for the case $l=0$. The quantity

$$\Delta\equiv|\varphi_0-\varphi_0{}^a|/|\varphi_0|$$

was found to be less than 0.01, 0.03, and 0.02 for the energies 40, 118, and 150 Mev, respectively, in the interval (0–1.0). In the region $s>1.0$, Δ is less than 0.05, 0.07, and 0.08 in the interval (1.0–1.8) for the energies 40, 118, and 150 Mev, respectively. In the interval (0.8–1.2), φ_0 and $\varphi_0{}^a$ are indistinguishable. For $s\gtrsim 1.8$, both φ_0 and $\varphi_0{}^a$ are small compared to unity.

It can be concluded that $\varphi_0{}^a$ is quite a good approximation to φ_0 when $0\lesssim s\lesssim 1.8$ for the case considered. The agreement between φ_l and $\varphi_l{}^a$ will be better for larger l because of the property (4.33b).

Now the validity of the approximation (4.25b) rests essentially on two points. The first of these is that $\varphi_l(s|1)$ be peaked in the neighborhood of $s=1$. The position of this peak will probably not vary much among the class of typical nucleon-nucleon potentials. Secondly, it is required that in the neighborhood of $s'=1$ the function $\tau_l(s|s')$ be predominantly a linear function of $(s'-1)/(s'+1)$ in order that $\bar{F}_l(s|s')$ be slowly varying in this region. This condition on $\tau_l(s|s')$ is true for any potential in a small enough neighborhood of $s'=1$. It is unlikely that such a property of τ_l would be strongly dependent upon the detailed shape of the potential. Hence, the results for any other typical short-range potential should not differ much from the results obtained for the square well.

When the interaction contains a h.c., the situation is complicated by the fact that $K_l(k_f|k_i)$ will drop off only as fast as $1/k_f$ for large k_f instead of exhibiting the behavior (4.32). If the potential V is absent, then $K_l=\bar{\Gamma}_l$ and the $1/k_f$ behavior is evident from Eq. (4.2). The origin of this property can be understood since the h.c. may be viewed as being equivalent, in some respects, to a delta-function potential.[11,12]

Apart from the asymptotic properties of $K_l(k_f|k_i)$ for large k_f, there is essentially no difference in the functional behavior of $u_l(k_i|k_p)$ and $K_l(k_f|k_i)$ from the case without a h.c. [cf. Eqs. (4.30), (4.31), (4.33b), (4.33c), and the associated remarks]. Since the asymptotic behavior for large k_f is a result of the h.c., it seems reasonable to expect that, if the effects of the h.c. could somehow be "subtracted" out of the K matrix, the resultant "reduced" K matrix would have properties similar to K when no h.c. is present. If this is the case, one can, as before, replace the kernel of the "reduced" integral equation by a separable function.

A precise formulation of the subtraction procedure, discussed in the preceding paragraph, can be obtained as follows. Let

$$K^0(\mathbf{k}_f|\mathbf{k}_i)\equiv\langle f|V\psi_i\rangle; \tag{4.35}$$

when the h.c. radius is zero, K^0 is identical with K. It is a simple matter to prove that $K^0(\mathbf{k}_f|\mathbf{k}_i)$ will behave as $1/k_f{}^2$ for large k_f when V is a typical short-range potential. The behavior of K^0 for small k_f is the same as in (4.31). As usual, the potential $V(\mathbf{r})$ is assumed to be central.

From definition (1.28) it is observed that

$$K(\mathbf{k}_f|\mathbf{k}_i)=K^0(\mathbf{k}_f|\mathbf{k}_i)+\int d\mathbf{k}_p\bar{\omega}(\mathbf{k}_f|\mathbf{k}_p)K^0(\mathbf{k}_p|\mathbf{k}_i)$$
$$+\bar{\Gamma}(\mathbf{k}_f|\mathbf{k}_i). \tag{4.36}$$

Equation (4.36) can be rewritten in terms of the partial-wave amplitudes as

$$K_l(k_f|k_i)=K_l{}^0(k_f|k_i)-j_l(k_fa)\int\bar{g}_l(k_p)K_l{}^0(k_p|k_i)dk_p$$
$$+\bar{\Gamma}_l(k_f|k_i), \tag{4.37}$$

where

$$\bar{g}_l(k_p)=\frac{2}{\pi}k_p{}^2\int j_l(k_pr)\frac{\bar{g}_l(a|r)}{\bar{g}_l(a|a)}r^2dr. \tag{4.38}$$

Now $K_l(k_f|k_i)$ satisfies the integral equation

$$K_l(k_f|k_i)=[f_l(k_f|k_i)-\tau_l(k_f|k_i)f_l(k_i|k_i)]$$
$$+\tau_l(k_f|k_i)K_l(k_i|k_i)$$
$$+\int\Lambda_l(k_f|k_p)K_l(k_p|k_i)dk_p; \tag{4.39}$$

the kernel and the inhomogeneous terms of this equation are given in Eqs. (4.4)–(4.6). By inserting Eq. (4.37) into Eq. (4.39), it is found that $K_l^0(k_f|k_i)$ satisfies the integral equation

$$K_l^0(k_f|k_i)=f_l^{(1)}(k_f|k_i)+f_l^{(2)}(k_f)\int \bar{g}_l(k_p)\times K_l^0(k_p|k_i)dk_p+\int \Lambda_l(k_f|k_p)\times K_l^0(k_p|k_i)dk_p, \quad (4.40)$$

where

$$f_l^{(1)}(k_f|k_i)=[f_l(k_f|k_i)-\tau_l(k_f|k_i)f_l(k_i|k_i)]+\tau_l(k_f|k_i)K_l(k_i|k_i)-\bar{\Gamma}_l(k_f|k_i)+\int \Lambda_l(k_f|k_p)\bar{\Gamma}_l(k_p|k_i)dk_p, \quad (4.41a)$$

and

$$f_l^{(2)}(k_f)=j_l(k_fa)-\int \Lambda_l(k_f|k_p)j_l(k_pa)dk_p. \quad (4.41b)$$

Now, K_l^0 has all the properties which were attributed to the function φ_l in the case of an ordinary interaction. Also, it follows from the previous discussion of the properties of $u_l(k_i|k_p)$ that the kernel Λ_l has all the properties of the kernel of the integral equation for φ_l in the case without the h.c. Hence, the kernel Λ_l can be replaced by a separable kernel in the same manner as before and with the same justification. Equation (4.40) is then exactly soluble.

The solution K_l^0 of Eq. (4.40), and Eq. (4.37), determine $K_l(k_f|k_i)$ in terms of $K_l(k_i|k_i)$. It should be remarked that the integral in Eq. (4.37) need not be evaluated, since it is clearly one of the constants to be determined in the solving of Eq. (4.40).

V. CONSISTENCY REQUIREMENT AND THE ON-THE-ENERGY-SHELL MATRIX ELEMENTS

In Sec. III a singular integral equation (3.1) was reduced to the Fredholm form (3.5) with the assumption that the solution $R(x|y)$ of (3.1) was known for $x=y$. It is essential that the solution of the Fredholm equation be consistent with the original singular integral equation. In this section the condition of consistency will be derived and some of its consequences discussed. For the sake of generality, the notation of Sec. III will be used throughout.

Since Eq. (3.5) is a Fredholm equation, its solution may be written in the form

$$R(x|y)=f'(x|y)+\tau(x|y)R(y|y)+\int \mathfrak{R}_y(x|z)f'(z|y)dz+\left[\int \mathfrak{R}_y(x|z)\tau(z|y)dz\right]R(y|y), \quad (5.1)$$

where

$$f'(x|y)=f(x|y)-\tau(x|y)f(y|y), \quad (5.2)$$

and the *resolvent kernel* $\mathfrak{R}_y(x|z)$ is independent of $R(y|y)$ and satisfies the condition

$$\mathfrak{R}_y(y|z)=0. \quad (5.3)$$

A consistency requirement on the solution (5.1) can now be derived by substituting Eq. (5.1) into the integral of Eq. (3.1) and setting $x=y$, viz.,

$$R(y|y)=\left\{1-P\int T'(y|z)\left[\tau(z|y)+\int \mathfrak{R}_y(z|t)\tau(t|y)dt\right]dz\right\}^{-1}\times\left\{f(y|y)+P\int T'(y|z)\left[f'(z|y)+\int \mathfrak{R}_y(z|t)f'(t|y)dt\right]dz\right\}, \quad (5.4)$$

where

$$T'(y|z)=z^2\xi(y|z)/(z^2-y^2). \quad (5.5)$$

In order that the solutions of Eqs. (3.1) and (3.5) be consistent with one another, $\mathfrak{R}_y(x|z)$ must satisfy Eq. (5.4) with $R(y|y)$ regarded as known.

On the other hand, suppose that $R(y|y)$ is not known. Then the consistency requirement (5.4) is, in fact, the expression of the solution for $R(y|y)$ in terms of $\mathfrak{R}_y(x|z)$. Any possible usefulness of this observation depends on how precisely $\mathfrak{R}$ must be known in order to obtain a good approximation for $R(y|y)$. If $\mathfrak{R}$ has to be known very accurately, then it is certainly more convenient to calculate $R(y|y)$ (for physical problems) using conventional techniques.

If the major contribution to the integral in Eq. (3.1) comes from a rather small neighborhood of $z=y$, then the solution (5.4) will be quite useful. The point is that, when $R(x|y)$ as given by Eq. (5.1) is substituted into the integral of Eq. (3.1), it is mainly the inhomogeneous term in Eq. (5.1) which will contribute in the region near $z=y$ because of (5.3). Thus, if this neighborhood of $z=y$ is really the only important region in the domain of integration, any inaccuracies in $\mathfrak{R}$ will be minimized.

Apart from any questions of validity, it is interesting to examine the structure of a simple approximate solution for $R(y|y)$ which can be derived from (5.4). Consider the case when the resolvent kernel is taken to be identically zero. Then, if T' and f are replaced by $\lambda T'$ and λf, respectively, where λ is a constant, it

follows from Eq. (5.4) that

$$R(y|y)=\lambda f(y|y)+\lambda^2\left[1-\lambda P\int T'(y|z)\tau(z|y)dz\right]^{-1}\times P\int T'(y|z)f(z|y)dz. \quad (5.6)$$

Approximation (5.6) has a resemblance to the first iteration or the "second Born approximation" for $R(y|y)$. However, Eq. (5.6) would appear to be a better approximation than this.[38] The second Born approximation is simply the replacement of $R(z|y)$ in the integral of Eq. (3.1) by $f(z|y)$, whereas approximation (5.6) is equivalent to the replacement of $R(z|y)$ by $\tau(z|y)R(y|y)+f'(z|y)$. But the latter expression is exactly equal to $R(z|y)$ for $z=y$; also, it has roughly the same asymptotic properties as $f(z|y)$ for z much different from y (at least in physical problems). On the other hand, $f(z|y)$ is, in general, not a good approximation to $R(z|y)$ for any value of z.

It should be noted that the expression on the right-hand side of Eq. (5.6) involves all orders of the *coupling constant* λ. To this extent, the approximation procedure leading to (5.6) can be called nonperturbative.

[38] When no h.c. is present, approximation (5.6) for $R(y|y)$ can be shown to be equivalent to the use of a plane wave trial function in a variational expression for $R(y|y)$ of the form derived by Schwinger.

APPENDIX

A proof of Eq. (1.30) when the interaction contains a h.c. will now be given. If Eq. (1.17) is used to eliminate the free ϕ_i term in Eq. (1.14), it is found that

$$\Psi_i=\bar{G}_iV\Psi_i+\bar{G}_i[(\omega_i-\bar{\omega}_i)V\psi_i^{(+)}+i\pi\textstyle\sum_j\bar{\Gamma}_j\phi_j\delta(E_i-E_j)t_{ji}+(\Gamma_i-\bar{\Gamma}_i)\phi_i], \quad (A1)$$

where

$$\Psi_i=\psi_i^{(+)}-\psi_i+i\pi\textstyle\sum_j\psi_j\delta(E_i-E_j)t_{ji}. \quad (A2)$$

A summation sign has been used for the sake of brevity.

If the coordinate representatives [cf. Eqs. (1.12) and (1.13)] for ω_i, $\bar{\omega}_i$, Γ_i, and $\bar{\Gamma}_i$ are employed, the following identities may be established by direct verification:

$$\Gamma_i=\bar{\Gamma}_i-i\pi\bar{\Gamma}_i\delta(E_i-H_0)\Gamma_i, \quad (A3)$$

$$\omega_i=\bar{\omega}_i-i\pi\bar{\Gamma}_i\delta(E_i-H_0)(1+\omega_i), \quad (A4)$$

where ω_i and $\bar{\omega}_i$ are assumed to operate only on vectors whose coordinate representatives vanish for $|\mathbf{r}|\leqslant a$. Equation (A3) is simply Eq. (1.30) when the interaction consists only of a h.c. With these relations, Eq. (A1) becomes simply

$$\Psi_i=\bar{G}_iV\Psi_i. \quad (A5)$$

If the solution of Eq. (1.17) is unique, then it is easily shown that the operator $\bar{G}_iV$ has no eigenvectors corresponding to the eigenvalue unity. In this case, it is evident from Eq. (A5) that Ψ_i must vanish identically. Then the combination Eqs. (A2)–(A4) yields Eq. (1.30). This proves the statement made in the text.

Multiple Scattering of Waves*

P. C. WATERMAN
Research and Advanced Development Division, Avco Corporation, Wilmington, Massachusetts

AND

ROHN TRUELL
Metals Research Laboratory, Division of Applied Mathematics, Brown University, Providence, Rhode Island

(Received October 14, 1960)

Multiple scattering effects due to a random array of obstacles are considered. Employing a "configurational averaging" procedure, a criterion is obtained for the validity of approximate integral equations describing the various field quantities of interest. The *extinction theorem* is obtained and shown to give rise to the *forward-amplitude theorem of multiple scattering*. In the limit of vanishing correlations in position, the complex propagation constant κ of the *scattering medium* is obtained. Under appropriate restrictions, the expression for κ is shown to include both the square-root law of isotropic scatterers and the additive rule for cross sections valid for sufficiently low densities of anisotropic obstacles. Some specific examples from acoustics and electromagnetic theory then indicate that at least in the simplest cases the results remain valid for physically allowable densities of obstacles.

1. INTRODUCTION

THE problem of scattering of an incident wave by a single obstacle has been considered rather thoroughly in the literature, beginning with Rayleigh's work in fluids, and continuing to the present day with various quantum-mechanical, electromagnetic, and elastic cases. The literature on multiple scattering, on the other hand, is not so extensive. Regular arrays of scatterers were treated by Huygen's principle or various perturbation schemes, mainly with an eye to obtaining their "strong" filtering properties arising due to periodicity. Out of this work came x-ray diffraction theory and the band theory of solids. Random distributions of scatterers first drew interest in connection with the problem of dispersion in metals, and theoretical analysis gave results in reasonable agreement with experiment over the range where classical concepts should be expected to apply.

Such treatments were concerned with problems in which the linear dimensions of obstacles were very small in comparison with the wavelength of incident radiation, so that the scattered waves from individual obstacles were either isotropic or dipole in nature. Recent investigators have been concerned with the more difficult case where the ratio of wavelength to scatterer size can be arbitrary, thus requiring that higher-order multipoles be included in the scattered radiation. Interest in such problems has been stimulated by improvements in the technique of generating and observing various kinds of radiation. Ultrasonic techniques, for example, permit one to generate and observe the behavior of nearly plane elastic waves in solid specimens containing a volume distribution of scattering regions which might be individual grains of a polycrystalline specimen.[1] Microwave techniques allow one to investigate the propagation behavior of artificial dielectrics composed of an array of infinite conducting cylinders[2] or metal or dielectric spheres embedded in a supporting matrix.[3]

In 1945 Foldy introduced the concept of "configurational" averaging by utilizing the joint probability distribution for the occurrence of a given configuration of (isotropic point) scatterers to average the resulting wave over all configurations.[4] This procedure was later generalized by Lax to include point scatterers with quite general scattering properties, using a quantum-mechanical formulation.[5] Twersky has used the same procedure to treat scattering and reflection of sound waves by a rough surface.[6] Extensive references to other related work are given by the last two authors.

With the advent of Foldy's paper, the fundamental difficulty inherent in multiple scattering computations involving random arrays of obstacles was brought out in *bas-relief*. In simple terms, it is the question of what relation exists between the exciting field acting on a scatterer at a point, and the total field which would exist at that point if the scatterer were not there. The answer to this question is extremely important because without it one cannot obtain a governing equation for the desired average field quantity. Foldy assumed *a priori* that the two fields in question were equal, while Lax related them by an undetermined constant

* This work is an extension of earlier work reported by P. C. Waterman and Rohn Truell, Tech. Rept. WAL 143/14-49, Metals Research Laboratory, Brown University (October, 1957). The more recent work was supported by the U. S. Air Force Ballistic Missile Division of the Air Research and Development Command.

[1] W. P. Mason and H. J. McSkimin, J. Acoust. Soc. Am. **19**, 464 (1947); J. Appl. Phys. **19**, 940 (1948); also, W. Roth, J. Appl. Phys. **19**, 901 (1948).
[2] Z. A. Kaprielian, J. Appl. Phys. **27**, 1491 (1956).
[3] R. W. Corkum, Proc. Inst. Radio Engr. **40**, 574 (1952).
[4] L. L. Foldy, Phys. Rev. **67**, 107 (1945).
[5] M. Lax, Revs. Modern Phys. **23**, 287 (1951); Phys. Rev. **85**, 621 (1952).
[6] V. Twersky, I.R.E. Trans. on Antennas and Propagation **AP-5**, 81 (1957); J. Acoust. Soc. Am. **29**, 209 (1957). Reflection by rough surfaces has also been considered by M. A. Biot, J. Appl. Phys. **28**, 1455 (1957); **29**, 998 (1958). A recent review of multiple scattering has been given by V. Twersky, J. Research Natl. Bur. Standards **64D**, 715 (1960).

of proportionality which was assumed to be close to unity.

The purpose of this paper is to derive the equations governing wave motion in a medium containing an array of finite scattering regions. It will be seen that within the limitations of the specific case considered, the treatment is equally applicable to classical and quantum mechanical problems. The program is as follows. Using the statistical approach of Foldy and Lax, the first partial average of the exciting field is computed; this is simply the field "incident" on a scatterer known to be at a given position. By "incident" field is meant the entire field in the external neighborhood of a scatterer, less the outgoing scattered wave from the scatterer. This first partial average is obtained in terms of the second partial average of the exciting field. The second partial average is that of the field incident on a scatterer known to be at a given position where in addition the position of a second scatterer is known.

On continuing this procedure, one obtains each time the field incident on a scatterer with n scatterer positions known in terms of the corresponding field with one additional scatterer position known; thus, with a total of N scatterers there results a system of N equations in N unknowns. Next the averaged total field is computed. This is the field one would actually measure if no explicit information were available concerning scatterer positions, and is obtained in terms of the first partial average of the exciting field.

Although the resulting system of $N+1$ equations may in principle be solved, the labor is prohibitive, and an approximation procedure must be employed. In what follows, the conditions have been indicated under which the exciting field may be approximated by the total field, in the process obtaining governing equations for the total coherent field and the exciting fields when one, two, or more scatterers are known to be at given positions. Under the same conditions, a governing equation may also be obtained for the total energy density, including both coherent and incoherent effects.

Because of the extreme complexity of these equations, only the averaged exciting field with one scatterer fixed will be considered explicitly. A solution is obtained for the simplest geometry in which a plane wave is incident normally on an infinite half-space containing a uniform random distribution of identical scatterers. The central result of this computation is an expression for the complex propagation constant describing propagation in the "scattering medium." With appropriate restrictions on the density and scattering amplitude of obstacles, this result appears to be valid over the whole frequency spectrum, thus bridging the gap between the low-frequency limit of point scatterers, where Foldy's results are obtained for isotropic scatterers, and the high-frequency limit where results are in agreement with the picture given by geometrical optics. By consideration of some specific examples in Rayleigh (low frequency) and geometric optics (high frequency) limits, it is suggested that the result is also applicable to all physically allowable fractional volumes of obstacles. In particular, the description of propagation behavior is in many cases correct at the limit when the fractional volume of obstacles is unity, corresponding to complete replacement of the supporting medium by the material of which the scatterers are made up.

For the sake of clarity and compactness, we will confine the main discussion to only the scalar problem where the vector nature of pertinent field quantities is secondary, so that the computations need consider only a single scalar potential. The formal extension of the governing equations to include more involved solenoidal wave motion and mode conversion is described briefly.

2. GOVERNING EQUATIONS OF MULTIPLE SCATTERING

2.1 Statistical Preliminaries

Problems of physical interest within the framework of multiple scattering range from the case of a regular array of obstacles with spacings comparable with the wavelength of incident radiation to the case in which scatterers are stationed nearly at random. Attention here is focused mainly on the latter situation, utilizing the methods developed by Foldy[4] and Lax[5] to average over all configurations of obstacles consistent with the statistical information available. In many physical applications such an averaging process occurs naturally in the course of experiment, either due to the measuring device averaging over a region large compared with any of the lengths involved, or where the configurations are changing with time rapidly in comparison with the time scale of measurement. An extensive discussion of measurements and applications has been given by Lax.

Consider a function $f(\mathbf{r}|\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{r}_N)$ of space coordinates $\mathbf{r}$ and N position vectors $\mathbf{r}_1$, $\mathbf{r}_2$, $\cdots$, $\mathbf{r}_N$ drawn to the center of each of the scattering regions. If the specific configuration of scatterers is not of special interest in the problem, the configuration may be regarded as one state in an ensemble and an average over all states may be taken.

In accordance with Foldy,[4] let

$$p(\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{r}_N)d\tau_1\cdots d\tau_N \tag{2.1}$$

be the probability of finding the first scattering region with center in the volume element $d\tau_1$ at $\mathbf{r}_1$, while at the same time the second scatterer is in $d\tau_2$ at $\mathbf{r}_2$, and so forth. This joint probability distribution $p(\mathbf{r}_1,\cdots\mathbf{r}_N)$ is to be restricted at the outset by the following conditions: The integral of the joint probability distribution over all configurations of scatterers is normalized to unity; the N scatterers are indistinguishable, which allows position vectors to be interchanged at will in Eq. (2.1) without changing the numerical value of the expression: finally, overlapping of scattering regions is

excluded, hence $p(\mathbf{r}_1,\cdots,\mathbf{r}_N)$ is made to vanish for all such situations.

The probability of finding a particular scatterer in $d\tau_1$ at $\mathbf{r}_1$ if no other information is available, is

$$p(\mathbf{r}_1)d\tau_1=d\tau_1\int\cdots\int d\tau_2\cdots d\tau_N p(\mathbf{r}_1,\cdots,\mathbf{r}_N). \quad (2.2)$$

Since each of the N scatterers has equal likelihood of occupying $d\tau_1$, the *density* $n(\mathbf{r}_1)$ of scatterers at $\mathbf{r}_1$ is then given by

$$n(\mathbf{r}_1)=Np(\mathbf{r}_1). \quad (2.3)$$

The joint probability of finding the first scatterer in $d\tau_1$ and the second in $d\tau_2$ is given by

$$p(\mathbf{r}_1,\mathbf{r}_2)d\tau_1 d\tau_2$$

$$=d\tau_1 d\tau_2\int\cdots\int d\tau_3\cdots d\tau_N p(\mathbf{r}_1,\cdots,\mathbf{r}_N), \quad (2.4)$$

where again in terms of densities

$$p(\mathbf{r}_1,\mathbf{r}_2)=p(\mathbf{r}_1)p(\mathbf{r}_2|\mathbf{r}_1)=\frac{n(\mathbf{r}_1)n(\mathbf{r}_2|\mathbf{r}_1)}{N(N-1)}. \quad (2.5)$$

Here $p(\mathbf{r}_2|\mathbf{r}_1)$ and $n(\mathbf{r}_2|\mathbf{r}_1)$ are *conditional* probability and density, respectively, at $\mathbf{r}_2$ if a scatterer is known to be at $\mathbf{r}_1$.[7] The first equality constitutes a definition of conditional probability, while the second follows from Eq. (2.3) and the observation that any one of the $N-1$ remaining scatterers has equal probability of lying in $d\tau_2$ if one scatterer is in $d\tau_1$.

The configurational average of $f(\mathbf{r}|\mathbf{r}_1,\cdots\mathbf{r}_N)$ is defined by

$$\langle f(\mathbf{r})\rangle\equiv\int\cdots\int d\tau_1\cdots d\tau_N \times p(\mathbf{r}_1,\cdots,\mathbf{r}_N)f(\mathbf{r}|\mathbf{r}_1,\cdots,\mathbf{r}_N). \quad (2.6)$$

The partial average with one or more scatterers held fixed is obtained by averaging over the appropriate conditional probability: Thus $f(\mathbf{r}|\mathbf{r}_1)$, the average of $f(\mathbf{r}|\mathbf{r}_1,\cdots,\mathbf{r}_N)$ over all configurations for which the first scatterer is at $\mathbf{r}_1$, is given by

$$f(\mathbf{r}|\mathbf{r}_1)\equiv\int\cdots\int d\tau_2\cdots d\tau_N \times p(\mathbf{r}_2,\cdots,\mathbf{r}_N|\mathbf{r}_1)f(\mathbf{r}|\mathbf{r}_1,\cdots,\mathbf{r}_N) \quad (2.7)$$

and so on.

2.2 Averaged Exciting Field

Consider a homogeneous isotropic medium capable of sustaining wave motion according to the scalar Helmholtz equation

$$(\Delta+k^2)\psi(\mathbf{r})=0, \quad (2.8)$$

and consider only problems for which the time dependence is simple harmonic. The time-dependent scalar potential is obtained by multiplying $\psi(\mathbf{r})$ by $\exp(-iwt)$. The (irrotational) vector field quantity of interest is given by ∇ Real Part$[\psi(\mathbf{r})\exp(-iwt)]$. The propagation constant k may be real or complex; the complex case corresponds to dissipation in the (matrix) medium.

Embedded in this medium are N identical scattering regions located at $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N$. Under the influence of an incident wave $\psi^{\text{inc}}(\mathbf{r})$ and the scattering from other scatterers, a scattered wave $\psi^S(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)$ is generated by the jth scatterer. Here the first coordinate $\mathbf{r}$ specifies the field point of evaluation, $\mathbf{r}_j$ gives the location of the scatterer originating the radiation, and the $\mathbf{r}_1, \cdots, \mathbf{r}_N$ indicate the dependence of the scattered wave on the specific configuration chosen. The properties of a single scatterer are assumed known, so that a rule is available relating the scattered wave and the *exciting field* $\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)$ acting on the jth scatterer to produce scattering. This rule defines a linear *scattering operator* $T(\mathbf{r}_j)$ by the relation

$$\psi^S(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)\equiv T(\mathbf{r}_j)\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N). \quad (2.9)$$

The exciting field is assumed to be a regular solution (in $\mathbf{r}$) of the unperturbed Helmholtz equation, Eq. (2.8), in the neighborhood of $\mathbf{r}_j$ in order that the single-scatterer computation be applicable.[8] Employing the radiation condition, the scattered wave has the form of outgoing radiation and is a regular solution of Eq. (2.8) everywhere but at $\mathbf{r}=\mathbf{r}_j$, where a singularity is present.

Now the exciting field acting on the jth scatterer is just the sum of the original incident wave and the scattered waves from all scatterers *other than the* jth, giving the implicit relations

$$\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N) =\psi^{\text{inc}}(\mathbf{r})+\sum_{k\neq j}^{N}T(\mathbf{r}_k)\psi^E(\mathbf{r}|\mathbf{r}_k;\mathbf{r}_1,\cdots,\mathbf{r}_N); \quad j=1,2,\cdots,N. \quad (2.10)$$

These relations account completely for the effect on each scatterer due to the presence of other scatterers. In terms of a *multiple orders of scattering* approach,[9] where *primary* scattering is due to the incident wave alone, *secondary* scattering represents one rescattering of the primary waves, and so on, all orders are included in Eqs. (2.10). This is easily seen by employing repeated

[7] The concept of *conditional* probability has been defined lucidly by W. Feller, *Probability Theory and Its Applications* (John Wiley & Sons, Inc., New York, 1950), Vol. I, p. 78.

[8] This assumption on the regular, unperturbed nature of the exciting field is tenable because of exclusion of interpenetration, which specifies a distance of closest approach of singular points.

[9] Discussed by V. Twersky, J. Acoust. Soc. Am. 24, 42 (1952). See also K. Fukiwara, J. Phys. Soc. Japan 14, 1513 (1959).

iteration to obtain the infinite series

$$\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)=\psi^{\text{inc}}(\mathbf{r})+\sum_{k\neq j}^{N}T(\mathbf{r}_k)\psi^{\text{inc}}(\mathbf{r})+\sum_{m\neq j}^{N}T(\mathbf{r}_m)\sum_{k\neq m}^{N}T(\mathbf{r}_k)\psi^{\text{inc}}(\mathbf{r})+\cdots. \quad (2.11)$$

Here the single summation gives the primary scattered terms, the double summation the secondary terms, and so on. Although we may be dealing with finite size scattering regions, behavior inside these regions need not enter in a discussion of the exciting field.

Because of the complicated nature of Eqs. (2.10), it does not appear feasible to attempt to invert them to obtain explicit expressions for the exciting fields. Instead, the equations will be averaged as they stand. Consider the partial average according to Eq. (2.7) of the first of Eqs. (2.10) over those configurations for which the first scatterer is fixed at $\mathbf{r}_1$. The incident wave may be taken outside of all integrations as it is independent of scatterer positions, leaving an integral normalized to unity. Also, since the scattering operator $T(\mathbf{r}_k)$ is linear, it may be taken outside all but the integration over $\mathbf{r}_k$, which is deferred until last. At this point one has

$$\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle=\psi^{\text{inc}}(\mathbf{r})+\sum_{k=2}^{N}\int d\tau_k T(\mathbf{r}_k)\int\cdots\int d\tau_2\cdot!\cdot d\tau_N \times p(\mathbf{r}_2,\cdots,\mathbf{r}_N|\mathbf{r}_1)\psi^E(\mathbf{r}|\mathbf{r}_k;\mathbf{r}_1,\cdots,\mathbf{r}_N).$$

The conditional probability may be rewritten in the form

$$p(\mathbf{r}_2,\cdots,\mathbf{r}_N|\mathbf{r}_1)=[n(\mathbf{r}_k|\mathbf{r}_1)/(N-1)]p(\mathbf{r}_2,\cdots,\mathbf{r}_N|\mathbf{r}_1,\mathbf{r}_k),$$

and upon insertion in the multiple integral the term in square brackets may be factored out of all but the $\mathbf{r}_k$ integration. The $(N-2)$-fold integrals are now recognized as partial averages of the exciting field with two scatterers held fixed. Further, the summation may be replaced by a multiplying factor $(N-1)$ due to indistinguishability, so that one finally obtains

$$\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle=\psi^{\text{inc}}(\mathbf{r})+\int d\tau' n(\mathbf{r}'|\mathbf{r}_1)T(\mathbf{r}')\times\langle\psi^E(\mathbf{r}|\mathbf{r}';\mathbf{r}_1)\rangle. \quad (2.12)$$

This lack of completeness, the fact that the exciting field with one scatterer fixed is given in terms of the field with two scatterers fixed, is the basic difficulty encountered in the implicit approach to multiple scattering. Because of the exclusion from the range of integration of points in the vicinity of $\mathbf{r}_1$, $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$ is a solution of the unperturbed medium equation in $\mathbf{r}$, as are the higher partial averaged fields given below.

In computing higher partial averages of the exciting field, one new feature enters. Consider the partial average at $\mathbf{r}_1$ with $\mathbf{r}_1$ and $\mathbf{r}_2$ fixed, for example. Since one term in the summation is a scattered wave from $\mathbf{r}_2$, after the averaging is performed this term will stand apart from the others, outside of the integral sign. Additional terms of this form appear as more scatterers are held fixed. This effect serves the immediate purpose of removing just the right number of terms from the summation to permit exact cancellation of any factors containing N. Aside from these additional terms, it should be clear at this point that each partially averaged exciting field will involve the integral of the exciting field with one additional scatterer held fixed. Thus one obtains the hierarchy of equations

$$\langle\psi^E(\mathbf{r}|\mathbf{r}_1;\mathbf{r}_1,\cdots,\mathbf{r}_n)\rangle=\psi^{\text{inc}}(\mathbf{r})+\int d\tau' n(\mathbf{r}'|\mathbf{r}_1,\cdots,\mathbf{r}_n)T(\mathbf{r}')\langle\psi^E(\mathbf{r}|\mathbf{r}';\mathbf{r}_1,\cdots,\mathbf{r}_n)\rangle+\sum_{j=2}^{n}T(\mathbf{r}_j)\langle\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_n)\rangle;$$
$$n=1,2,\cdots,N-1, \quad (2.13)$$

where $n(\mathbf{r}'|\mathbf{r}_1,\mathbf{r}_2)$ is the density of scatterers at $\mathbf{r}'$ if scatterers are known to be at $\mathbf{r}_1$ and $\mathbf{r}_2$, and so forth.

Three methods can be seen of treating this hierarchy. First, one can in principle obtain an exact solution by solving the implicit initial representation Eq. (2.10) for $\psi^E(\mathbf{r}|\mathbf{r}_N;\mathbf{r}_1,\cdots,\mathbf{r}_N)$, inserting the result in the last equation of the system, and working up from the bottom. As before, this scheme is discarded because the labor is prohibitive. Second, one might iterate the system, replacing the exciting field in the integrand of one equation by the right-hand side of the following equation, and continuing to replace the exciting field wherever it occurs, ultimately obtaining an infinite series representation for each of the exciting fields which involves only operations on $\psi^{\text{inc}}(\mathbf{r})$. This result expresses the exciting field in terms of multiple orders of scattering, and could have been obtained more directly by taking configurational averages of Eq. (2.11), the multiple orders of scattering starting representation of the exciting field. Such series rapidly become cumbersome to treat exactly when more than one or two terms are considered, so that this scheme is only feasible when multiple scattering effects are very weak. Finally, Lax has suggested breaking off the hierarchy at some stage by arbitrarily replacing the exciting field in an integrand by the corresponding field with one less scatterer held fixed.[5] The resulting equation is then solved, and each of the preceding equations solved by quadrature. More will be said about this procedure later. Further discussion of the functional behavior of the exciting fields is also required, but first it is necessary to consider the total field.

2.3 Averaged Total Field

The expression for the total field is somewhat more involved than for the exciting field, as it will depend on whether the point of evaluation falls inside or outside

of a scattering region. The computation of the averaged exciting fields did not depend on the existence of potentials within the scattering regions. However, these potentials are essential to the computation of the averaged total field. The viewpoint used is as follows: If a wave equation exists for the interior of scattering regions, then a potential is defined in these regions and the procedure is straightforward. If the interiors of the scattering regions cannot sustain wave motion, but may be considered as limiting cases of media which can, then the computation is carried through with internal potentials, with the understanding that the appropriate limiting process be employed subsequently.

If neither of these conditions is fulfilled, it is meaningless to speak of the total field. This situation is satisfactory for our purposes, because any measurement which did not disturb the ensemble of scatterers would necessarily be made on a partially averaged exciting field. That is, in order to guarantee that a measurement be possible at any particular point, we would have the alternative of either modifying the ensemble so that the point in question was never in the interior of a scatterer, which seems rather unsatisfactory, or agreeing to measure, say, the first partial average of the exciting field on a scatterer fixed at the point.

In exact analogy with Eq. (2.9) defining $T(\mathbf{r}_j)$, the boundary conditions of the single-scatterer problem may be used to define an internal scattering operator $T^I(\mathbf{r}_j)$ generating the internal potential $\psi^I(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)$ from the exciting field according to

$$\psi^I(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)\equiv T^I(\mathbf{r}_j)\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N). \quad (2.14)$$

Now the total field at any point is the sum of the incident and all scattered waves, if the point of interest lies outside of all scatterers. Inside a scatterer the appropriate internal field must be employed. In order to formulate this statement mathematically, the following device is used: let

$$\alpha(\boldsymbol{\varrho})=\begin{cases}0 & \boldsymbol{\varrho}\text{ ``inside'' }0\\ 1 & \boldsymbol{\varrho}\text{ ``outside'' }0,\end{cases} \quad (2.15)$$

where "$\boldsymbol{\varrho}$ inside 0" means that the point $\boldsymbol{\varrho}$ is in the interior (or on the boundary) of a scattering region with "center" at the origin, and "$\boldsymbol{\varrho}$ outside 0" is the complementary statement. The total field may now be written

$$\psi(\mathbf{r}|\mathbf{r}_1,\cdots,\mathbf{r}_N)=\prod_{k=1}^{N}\alpha(\mathbf{r}-\mathbf{r}_k)$$
$$\times\{\psi^{\mathrm{inc}}(\mathbf{r})+\sum_{j=1}^{N}T(\mathbf{r}_j)\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)\}$$
$$+\sum_{j=1}^{N}[1-\alpha(\mathbf{r}-\mathbf{r}_j)]T^I(\mathbf{r}_j)\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N), \quad (2.16)$$

and it is readily verified using the definition of $\alpha(\boldsymbol{\varrho})$ that this expression has the required form.[10] Before averaging, some simplification is possible. Consider the identity

$$1\equiv\prod_{k=1}^{N}[\alpha(\mathbf{r}-\mathbf{r}_k)+1-\alpha(\mathbf{r}-\mathbf{r}_k)]$$
$$=\prod_{k=1}^{N}\alpha(\mathbf{r}-\mathbf{r}_k)+\sum_{j=1}^{N}[1-\alpha(\mathbf{r}-\mathbf{r}_j)]\prod_{k\neq j}^{N}\alpha(\mathbf{r}-\mathbf{r}_k)$$
$$+\sum_{j,k\neq j}^{N}[1-\alpha(\mathbf{r}-\mathbf{r}_j)][1-\alpha(\mathbf{r}-\mathbf{r}_k)]\prod_{m\neq j,k}^{N}\alpha(\mathbf{r}-\mathbf{r}_m)$$
$$+\cdots+\prod_{k=1}^{N}[1-\alpha(\mathbf{r}-\mathbf{r}_k)].$$

Multiplying through by $p(\mathbf{r}_1,\cdots,\mathbf{r}_N)$, it is seen that on the right all terms after the second vanish identically, since products of the form $[1-\alpha(\mathbf{r}-\mathbf{r}_j)]\,[1-\alpha(\mathbf{r}-\mathbf{r}_k)]$ are zero except when interpenetration occurs, at which time the joint probability distribution vanishes. Consequently,

$$p(\mathbf{r}_1,\cdots,\mathbf{r}_N)\prod_{k=1}^{N}\alpha(\mathbf{r}-\mathbf{r}_k)$$
$$=p(\mathbf{r}_1,\cdots,\mathbf{r}_N)\{1-\sum_{j=1}^{N}[1-\alpha(\mathbf{r}-\mathbf{r}_j)]\prod_{k\neq j}^{N}\alpha(\mathbf{r}-\mathbf{r}_k)\}$$
$$=p(\mathbf{r}_1,\cdots,\mathbf{r}_N)\{1-\sum_{j=1}^{N}[1-\alpha(\mathbf{r}-\mathbf{r}_j)]\}, \quad (2.17)$$

where the second equality follows from the fact that the $(N-1)$-termed product contributes no additional information (seen by employing the identity expansion again). On placing this last result in Eq. (2.16), the configurational averaging process is straightforward, and for the averaged total field there results

$$\langle\psi(\mathbf{r})\rangle=\psi^{\mathrm{inc}}(\mathbf{r})$$
$$+\int_{\mathbf{r}\text{ ``outside'' }\mathbf{r}'}d\tau' n(\mathbf{r}')T(\mathbf{r}')\langle\psi^E(\mathbf{r}|\mathbf{r}')\rangle$$
$$+\int_{\mathbf{r}\text{ ``inside'' }\mathbf{r}'}d\tau' n(\mathbf{r}')[T^I(\mathbf{r}')-1]\langle\psi^E(\mathbf{r}|\mathbf{r}')\rangle, \quad (2.18)$$

where the statement "$\mathbf{r}$ outside $\mathbf{r}'$" means, as before, that integration is to be carried over all points $\mathbf{r}'$ such that $\mathbf{r}$ is outside the scatterer having center $\mathbf{r}'$. The interpretation of this result is as follows: The first two terms represent the "outside point" contributions to

[10] In many situations the fundamental physical quantity is *proportional* to ψ, e.g., in acoustics pressure is given by the product of ψ with density, which of course may differ for scatterer and matrix material. We suppose that the *total* field in Eqs. (2.16) and (2.18) has been normalized in this sense, according to the physics of the specific problem of interest. Such normalization of the *exciting* field is unnecessary as only the unperturbed matrix medium is involved.

the total field arising from those configurations for which the evaluation point $\mathbf{r}$ is external to all scatterers. The first term in the second integral gives the "inside point" contributions, and the final term serves to cancel out the "outside point" contributions as the likelihood of $\mathbf{r}$ being an outside point diminishes.

2.4 Criterion for the Validity of Approximate Equations

The system of exact equations governing multiple scattering is now complete, with the averaged total field given in terms of the exciting field by Eq. (2.18). Unfortunately, the equations are not at all tractable in present form, as has been pointed out in Sec. 2.2, so approximations must be used. One can begin by expressing the exciting field $\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)$ incident on the jth scatterer as the total field $\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N)$ that would exist in the neighborhood if the jth scatterer were not present. The prime signifies that $\mathbf{r}_j$ and its effects are omitted in this expression. When the scatterer is then "inserted" in position, additional terms must be added to the total field, representing rescattering from all other scatterers of the radiation from the "inserted" source $T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N)$, rescattering of the "primary" scattered waves, and so on. These terms are included in the present starting representation in the form of an infinite series of multiple orders of scattering. It should be stressed that this representation, while not in closed form, is exact.

Step two consists of interpreting the infinite series as a sum of singly scattered waves from all other scatterers, with the role of incident wave played by the "source" $T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N)$, propagating in a scattering medium with complex index of refraction. To support this interpretation, a crude summation of the infinite series is given in Appendix I for the simple case of isotropic point scatterers, using a Fresnel half-period zone approach in three dimensions. Statistical fluctuations are neglected in this estimate, as are correlations in scatterer positions. In the course of the estimate there is obtained as a by-product the interesting result that the role of the nth-order term in the *multiple orders* approach is simply to generate the nth term in the power series expansion of an exponential characterizing the modified properties of the *scattering medium*. This provides a useful criterion for determining the feasibility of the multiple orders approach in specific cases.

Next, returning to the starting representation, the first term in the infinite series is estimated and shown to be negligible under a wide range of conditions. Finally, in the light of the above interpretation of the infinite series, it is pointed out that the criterion so obtained cannot be appreciably influenced when the additional terms of the series are included (equivalent to saying that if primary scattering of the "inserted" field is negligible, secondary and higher orders must be even more so).

If we assume that the criterion is satisfied, then we may replace the exciting field in a frozen configuration of scatterers by the total field which would exist if the scatterer in question (and its effects) were removed. With the aid of this replacement the configurational averaging process is then readily employed to yield linear integro-operational equations for the various field quantities of interest.

On the basis of the above discussion and Appendix I, the starting (exact) representation for the exciting field is given by

$$\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)=\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N)+\sum_{k\neq j}T(\mathbf{r}_k)$$
$$\times\{1+\sum_{m\neq k,j}T(\mathbf{r}_m)+\sum_{m\neq k}T(\mathbf{r}_m)\sum_{n\neq m,j}T(\mathbf{r}_n)$$
$$+\sum_{m\neq k}T(\mathbf{r}_m)\sum_{n\neq m}T(\mathbf{r}_n)\sum_{p\neq n,j}T(\mathbf{r}_p)+\cdots\}$$
$$\times T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N). \quad (2.19)$$

Each summation extends over all scatterers except those noted. In the summation over $\mathbf{k}$ (which has been factored out of the brackets for reasons apparent shortly), the value $\mathbf{j}$ is excluded because we are calculating the exciting field on the $\mathbf{j}$th scatterer. All other omitted indices are determined by the simple rule that the product $T(\mathbf{r}_m)T(\mathbf{r}_m)$ never occurs; physically, the $(n+1)$st-order scattered wave is generated by an exciting field made up of the nth-order scattered waves from all scatterers *but* the one in question. The reader may find it interesting to compare this equation with Eq. (2.11), the pure multiple orders of scattering representation. It is not trivial to collapse the series above to an equivalent implicit form because of the intermediate role played by the jth scatterer, which (1) serves as source, (2) does not participate in primary scattering, and thereupon (3) serves as a scatterer for secondary and higher orders.

Now for fixed k the series in brackets operating on $\chi_0\equiv T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N)$ is formally just the exciting field on the kth scatterer generated by an incident wave χ_0 and all orders of scattering of this incident wave. If the concept of a *scattering medium* with modified index of refraction is to be valid, then the addition of succeeding terms in the series must effect this modification in propagation properties. The manner in which this transition to a *scattering medium* occurs is shown explicitly (if approximately) for isotropic point scatterers in the Appendix.

Assuming the interpretation valid for the moment, consider the first term, say φ_0 of the series, given by

$$\varphi_0(\mathbf{r})\equiv\sum_{k\neq j}^{N}T(\mathbf{r}_k)T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N).$$

Each term in the summation represents rescattering

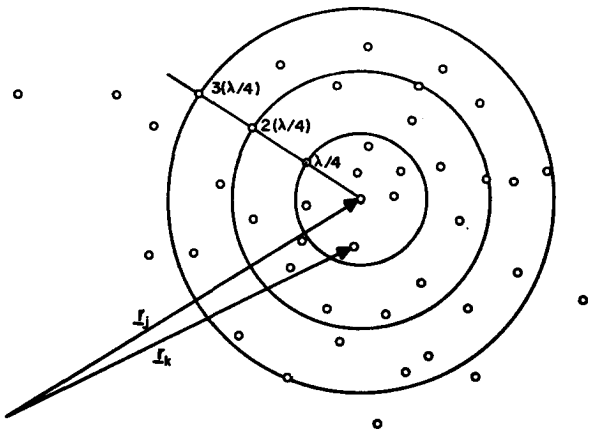


FIG. 1. The half-period zones for an isotropic point source at $\mathbf{r}_j$ are concentric spherical shells of thickness $\lambda/4$. The rescattering to $\mathbf{r}_j$ from all scatterers in a given zone will be on the average equal in magnitude but opposite in sign from the contribution of the preceding zone.

from one of the obstacles of an outgoing scattered wave from $\mathbf{r}_j$. Furthermore, since the exciting field is only required in the immediate neighborhood of the scatterer in question, in this case the jth according to Eq. (2.19), we need only estimate $\varphi_0(\mathbf{r})$ in this neighborhood; for simplicity consider $\varphi_0(\mathbf{r}_j)$.

Surfaces of equal phase of the outgoing wave will be more or less spherical, depending on the degree of anisotropy of single scattering, so that the phase of rescattered waves returning to $\mathbf{r}_j$ will retard roughly according to the round trip distance $2|\mathbf{r}_k-\mathbf{r}_j|$ from $\mathbf{r}_j$ to the individual scatterer. A set of concentric half-period zones may be constructed about $\mathbf{r}_j$, each zone defined by the requirement that its rescattered waves are no more than half a period out of phase at $\mathbf{r}_j$. These zones are illustrated in Fig. 1 for the limiting case of isotropic point scatterers. Suppose there is a uniform random distribution of density n_0; further supposing the amplitude of the scattered wave to be proportional to that of the exciting field, one may write

$$T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots,\mathbf{r}_N)=f\psi_j\frac{\exp[ik|\mathbf{r}-\mathbf{r}_j|]}{|\mathbf{r}-\mathbf{r}_j|},$$

using the abbreviation $\psi_j\equiv\psi(\mathbf{r}_j|\mathbf{r}_1,\cdots',\mathbf{r}_N)$. $\varphi_0(\mathbf{r}_j)$ now becomes

$$\varphi_0(\mathbf{r}_j)=\sum_{k\neq j}^{N} f^2\psi_j\frac{\exp[2ik|\mathbf{r}_k-\mathbf{r}_j|]}{|\mathbf{r}_k-\mathbf{r}_j|^2},$$

and it is clear that the contributions from succeeding zones are the same order of magnitude but alternating in sign, the square law increase of number of scatterers with distance being precisely compensated by the inverse behavior of both scattered and rescattered waves. One concludes that summation over the first zone should suffice to give an order of magnitude estimate of $\varphi_0(\mathbf{r}_j)$, and further approximating the sum by an integral we obtain

$$\varphi_0(\mathbf{r}_j)\approx f^2\psi_j\sum_{k\neq j}^{\text{1st zone}}\frac{\exp[2ik|\mathbf{r}_k-\mathbf{r}_j|]}{|\mathbf{r}_k-\mathbf{r}_j|^2}$$

$$\approx n_0 f^2\psi_j\int_{\text{1st zone}} d\tau'\frac{\exp[2ik|\mathbf{r}'-\mathbf{r}_j|]}{|\mathbf{r}'-\mathbf{r}_j|^2}$$

$$=(4\pi n_0/ik)f^2\psi_j.$$

Thus $\varphi_0(\mathbf{r}_j)$ may be neglected in comparison with ψ_j in Eq. (2.19), provided

$$|(4\pi n_0/ik)f^2|=(4\pi n_0/k)|f|^2\ll 1.$$

Further, since $4\pi|f|^2$ can be shown to be just the scattering cross section Q_s of a single scatterer, the above criterion simplifies to

$$(n_0Q_s/k)\ll 1. \tag{2.20}$$

This criterion is always satisfied for Rayleigh-type scattering,[11] where the scatterers are spherical obstacles of radius a with $ka\ll 1$. Here the scattering cross section has order of magnitude $Q_s\approx k^4a^6$, and the maximum allowable density $n_0\approx a^{-3}$ (at which point the scatterers fill the entire volume), so the criterion becomes $(ka)^3\ll 1$, which is automatically satisfied.

Relaxation type behavior is encountered in the scattering of electromagnetic waves by free electrons whose motion is damped due to collisions. On employing results given by Stratton,[12] one can show that the single electron scattering cross section (neglecting radiation reaction) is given by

$$Q_s=\frac{8\pi e^4}{3m^2c^4}\frac{1}{[1+(w_d/w)^2]},$$

where w_d is an empirical collision frequency, and the other quantities have their usual meaning. By inspection, one sees that the criterion of Eq. (2.20) will be most demanding at a signal frequency $w=w_d$, and making use also of the expression for static conductivity $\sigma_0=n_0e^2/mw_d$, one obtains

$$(n_0Q_s/k)_{\max}=(4\pi/3)(e^2\sigma_0/mc^3)\approx 3.5\times 10^{-13}\sigma_0(\text{mhos/m}).$$

For silver, with $\sigma_0\approx 6\times 10^7$ mhos/m,[13] (n_0Q_s/k) max $\approx 2\times 10^{-5}$, while for a Lorentz gas in general under the extreme conditions $T\approx 10^8$ °K, $n_0\approx 10^{24}$ cm^{-3}, we

[11] See, for example, the discussion of the Neumann problem in the low-frequency limit, given by P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 1484.

[12] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 326; the required expression for power radiated by an electric dipole may be found on p. 437. The scattered wave is represented by a Hertz vector having a single cartesian component whose magnitude is the isotropic scalar potential appropriate to the present discussion (see p. 432).

[13] Reference 12, p. 605.

obtain[14] $\sigma_0 \approx 4\times 10^9$ mhos/m, resulting in $(n_0 Q_s/k)_{\max} \approx 10^{-3}$. From these results, it appears that the criterion is satisfied for those problems involving scattering by collision damped free electrons, i.e., dispersion in metals, and electromagnetic wave propagation in partially and fully ionized gases.

As an example of resonance phenomena, consider the scattering of sound waves by air bubbles in water. At the resonance frequency, characterized by an effective mass of water adjacent to the bubble with restoring force related to bubble stiffness, the scattering cross section becomes[15] $Q_s = 4\pi a^2/\epsilon^2$, where **a** is bubble radius and ϵ a dimensionless damping constant, in general frequency dependent, accounting for losses due to viscosity and nonadiabatic behavior. Defining the fractional volume of air bubbles by $\delta = 4\pi a^3 n_0/3$, and noting that $k_{\text{res}} a = [3(\rho v^2)\ \text{air}/(\rho v^2)\ \text{water}]^{\frac{1}{2}}$, the criterion becomes, at resonance,

$$(n_0 Q_s/k)_{\text{resonance}} = (\delta/\epsilon^2)[3(\rho v^2)\ \text{water}/(\rho v^2)\ \text{air}]^{\frac{1}{2}} \approx 2.6\times 10^2 (\delta/\epsilon^2).$$

There is some discrepancy in the literature as to the magnitude of ϵ at resonance. Morse and Feshbach suggest that this value is somewhat larger than unity,[15] while MacPherson, for example, gives an experimental value of order 8×10^{-2}, based on transmission and reflection measurements on a bubble screen (two-dimensional array) in the kilocycle range.[16] At any event, it appears that the criterion may not be satisfied here even at fractional volumes below 10^{-2}, and further investigation is strongly indicated in this case.

The above estimates have only been concerned with primary order of the multiple orders of scattering correction to the exciting field, i.e., the first term of the infinite series appearing on the right side of Eq. (2.19). One now infers that the estimate remains valid when the entire series is considered, because of the following observations. The infinite series has been interpreted as generating a modified "scattering medium" propagation behavior. Physically, it is necessary to restrict the modified propagation constant to the first quadrant in the complex plane; from the usual interpretation in terms of phase velocity and attenuation, this restriction simply requires that the modified scattered wave remain outgoing and have nonnegative attenuation, i.e., not increase in energy content as it travels. In the preceding integration over the first zone, the integral is then modified to extend over a sphere of radius $\pi[\text{Re}(\kappa)+k]^{-1}$, and k in the exponent is replaced by $(\kappa+k)/2$. The reader may easily verify that under the physical limitations imposed on κ, the integral can be increased by at most a factor of two, and this only

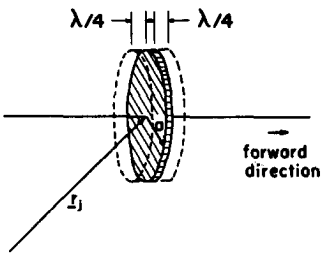


FIG. 2. The first half-period zone for rescattering of scattered waves from the large diameter disk at $\mathbf{r}_j$ consists of the two dashed cylindrical regions of thickness $\lambda/4$ and radius a.

in the extreme limit $\kappa \to 0$. In simpler terms, $\text{Im}(\kappa) > 0$ corresponds to attenuation of the outgoing wave, which can only decrease the collective return; $\text{Re}(\kappa) \to 0$, which corresponds to very large outgoing phase velocity, can at most increase the radius of the first zone by a factor of two, because of the fixed phase velocity of the return signals.

Consider next the situation near the geometrical optics limit. In order to obtain an order of magnitude estimate in this case, it is expedient to neglect diffraction effects altogether. Suppose the obstacles to be thin circular disks of radius a very large in comparison with wavelength, with disk j, shown in Fig. 2, illuminated normally by a plane sound wave of amplitude ψ_j. The scattered wave consists of plane waves of amplitude $g_{\text{forward}}\psi_j$, $g_{\text{back}}\psi_j$, propagating away from the disk in the forward and back directions, respectively, and confined to a cylindrical volume with cross section defined by the disk.

The first zone for rescattering of this wave (to the center point of disk j) consists of the cylinders of radius a extending a distance $\lambda/4$ upstream and downstream from the disk, shown dotted in Fig. 2. Consider first the upstream portion, which is slightly simpler to analyze. By the methods employed above, with back-scattered wave given by

$$T_{\text{back}}(\mathbf{r}_j)\psi_j e^{ikz} = g_{\text{back}}\psi_j e^{ik|z-z_j|},$$

we obtain a contribution after one rescattering of

$$\varphi_0(\mathbf{r}_j) \approx n_0 g^2_{\text{back}}\psi_j \int_{\text{1st (back) zone}} d\tau' e^{ik|z'-z_j|} e^{ik|z_j-z'|} = -(\pi a^2 n_0 g^2_{\text{back}}/ik)\psi_j.$$

This quantity is negligible in comparison with ψ_j provided

$$|\pi a^2 n_0 g^2_{\text{back}}/ik| = (\pi a^2 n_0 R/k) \ll 1,$$

where the reflection coefficient $R = |g_{\text{back}}|^2$ has been introduced in accordance with the usual definition. Noting that $\pi a^2 R$ is just power scattered in the back direction, we must have scattering cross section $Q_s \approx \pi a^2 R$ (which neglects a factor of perhaps two because forward-scattered power is not included). Making this replacement, one finally obtains precisely the criterion that resulted in the small scatterer limit, namely,

$$(n_0 Q_s/k) \ll 1. \tag{2.20}$$

[14] L. Spitzer, *Physics of Fully Ionized Gases* (Interscience Publishers, Inc., New York, 1956), p. 83. The values of Table 5.1 for an electron-proton gas have been employed.

[15] Reference 11, p. 1499.

[16] J. D. MacPherson, Proc. Phys. Soc. (London) **B70**, 85 (1957).

We have up to this point only tacitly restricted the thickness z_0 of the disks to be much less than cylinder diameter so that a simple cylindrical zone picture could be employed. Suppose now that the disk is many wavelengths thick, so that $kz_0 \gg 1$. The scattering cross section is at most of order πa^2, and $n_0 \leqslant (\pi a^2 z_0)^{-1}$ from volume considerations, so

$$(n_0 Q_s/k) \leqslant (1/kz_0) \ll 1$$

is automatically satisfied. At the other extreme of disks very thin in comparison with wavelength, one can readily show that $R \xrightarrow[z_0 \to 0]{} 0(kz_0)^2$ provided the disks are not completely opaque,[17] and thus

$$(n_0 Q_s/k) \leqslant kz_0 \ll 1$$

is again automatically satisfied. In intermediate cases, the criterion may be satisfied by suitable combinations of the three parameters R, n_0, and k. The calculations on the shadow side of the disk proceed in similar fashion except that the return from the first zone must be rescattered from $\mathbf{r}_j$. When this result is compared with the primary scattering from $\mathbf{r}_j$, one again obtains precisely the criterion of Eq. (2.20). The higher orders of scattering may be treated in the same manner as previously. Details of the calculation, not given here, are somewhat simpler for the cylindrical zone structure.

One is tempted to infer that Eq. (2.20) is a generally valid criterion for the approximation we desire to make, although we have obviously not come close to proving such a statement. Certainly it is appealing in its simplicity, and physically reasonable, requiring that the fraction of area "blacked out" per unit wavelength due to scattering cross section be much less than unity. The practical significance of the criterion lies in the fact that no limitations are put on the magnitude of absorption losses associated with the obstacles, so the theory should be capable of handling problems involving "lossy" obstacles even when the resulting coherent field is highly attenuated.

2.5 Approximate Integral Equations

Using the result of the previous section, that the exciting field on a scatterer may be approximated by the total field existent in the neighborhood if the scatterer were not there, it is possible to obtain equations for the individual quantities of interest. The simplest case is that of statistically independent isotropic point scatterers, the problem considered by Foldy. Since it is not necessary to include the effects of the interior of the scatterers, the starting representation, Eq. (2.16), reduces to

$$\psi(\mathbf{r}|\mathbf{r}_1,\cdots,\mathbf{r}_N) = \psi^{\text{inc}}(\mathbf{r}) + \sum_{j=1}^{N} T(\mathbf{r}_j)\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)$$

$$\approx \psi^{\text{inc}}(\mathbf{r}) + \sum_{j=1}^{N} T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots,'\mathbf{r}_N),$$

where the exciting field in each term in the summation has been replaced by the total field with that scatterer absent.

Taking the configurational average of this equation, employing the factorization $p(\mathbf{r}_1,\cdots,\mathbf{r}_N) = p(\mathbf{r}_1)p(\mathbf{r}_2)\cdots p(\mathbf{r}_N)$ necessary for statistical independence, one has

$$\langle\psi(\mathbf{r})\rangle \approx \psi^{\text{inc}}(\mathbf{r}) + \sum_{j=1}^{N} \int d\tau_j p(\mathbf{r}_j) T(\mathbf{r}_j)$$

$$\times \int d\tau_1 \cdots' d\tau_N p(\mathbf{r}_1,\cdots,'\mathbf{r}_N)\psi(\mathbf{r}|\mathbf{r}_1,\cdots,'\mathbf{r}_N)$$

$$= \psi^{\text{inc}}(\mathbf{r}) + \int d\tau_1 n(\mathbf{r}_1) T(\mathbf{r}_1)$$

$$\times \int d\tau_2 \cdots d\tau_N p(\mathbf{r}_2,\cdots,\mathbf{r}_N)\psi(\mathbf{r}|\mathbf{r}_2,\cdots,\mathbf{r}_N).$$

The $(N-1)$-fold integral in the last line is independent of $\mathbf{r}_1$ and, in fact, is precisely the averaged total field for the same problem with one less scatterer involved. As the number of scatterers increases, this quantity must differ from the total field with N scatterers by terms of order $1/N$, and by making this replacement one obtains the integral equation

$$\langle\psi(\mathbf{r})\rangle \approx \psi^{\text{inc}}(\mathbf{r}) + \int d\tau' n(\mathbf{r}') T(\mathbf{r}') \langle\psi(\mathbf{r})\rangle$$

$$= \psi^{\text{inc}}(\mathbf{r}) + f \int d\tau' n(\mathbf{r}') \langle\psi(\mathbf{r}')\rangle$$

$$(e^{ik|\mathbf{r}-\mathbf{r}'|}/|\mathbf{r}-\mathbf{r}'|), \quad (2.21a)$$

where in the second line the explicit form for the isotropic scattered wave given earlier has been substituted. Following Foldy, one may operate on Eq. (2.21a) with the unperturbed Helmholtz operator $\Delta + k^2$, obtaining because of the Green's function nature of the kernel the equivalent differential equation

$$[\Delta + \kappa^2(\mathbf{r})]\langle\psi(\mathbf{r})\rangle = 0, \quad (2.21b)$$

where the complex propagation factor $\kappa(\mathbf{r})$ is given by

$$\kappa^2(\mathbf{r}) = k^2\left[1 + \frac{4\pi f n(\mathbf{r})}{k^2}\right]. \quad (2.22)$$

This last equation defines the properties of the *scattering medium*, in which the average field is propagated with attenuation and modified phase velocity specified by a complex propagation factor $\kappa(\mathbf{r})$.

By modifying the starting representation at the beginning of this section so as to represent the exciting field and following the same procedure, one readily obtains the simple result

$$\langle\psi^E(\mathbf{r}|\mathbf{r})\rangle = \langle\psi(\mathbf{r})\rangle.$$

[17] For an example in the electromagnetic case, see Sec. 3.3.

That is, the average exciting field on an isotropic point scatterer *at the scatterer* is just equal to the average total field at the same point, a result that could of course be obtained more directly from the exciting field approximation.

Further discussion of the preceding results is deferred until the following section, and the more difficult problem of anisotropic finite scattering regions with correlations in position is considered. We do not attempt to obtain an integral equation for $\langle\psi(\mathbf{r})\rangle$ this time, instead working with the exciting field, from which $\langle\psi(\mathbf{r})\rangle$ may be obtained by quadrature according to Eq. (2.18). The exciting field on scatterer 1 is given by

$$\psi^E(\mathbf{r}|\mathbf{r}_1;\mathbf{r}_1,\cdots,\mathbf{r}_N)\approx\psi(\mathbf{r}|\mathbf{r}_2,\cdots,\mathbf{r}_N)$$
$$=\psi^{\text{inc}}(\mathbf{r})+\sum_{j=2}^{N}T(\mathbf{r}_j)\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_2,\cdots,\mathbf{r}_N),$$

where in the second line the internal fields do not appear because of the exclusion of interpenetration.[18] One proceeds by observing that none of the scattered waves in the summation depend on $\mathbf{r}_1$. By splitting the joint probability distribution into two terms, the first of which involves $\mathbf{r}_1$ only in a simple manner, and averaging, an integral will result containing $\langle\psi^E(\mathbf{r}|\mathbf{r}')\rangle_{N-1}$, the averaged exciting field for the same problem with $N-1$ scatterers, plus a correction term $R(\mathbf{r}|\mathbf{r}_1)$ due purely to statistical correlations in scatterer positions. As before, it is assumed that $\langle\psi^E(\mathbf{r}|\mathbf{r}')\rangle_{N-1}\approx\langle\psi^E(\mathbf{r}|\mathbf{r}')\rangle$ for a sufficiently large number of scatterers.

The probability distribution required here is $p(\mathbf{r}_2,\cdots,\mathbf{r}_N|\mathbf{r}_1)$, and this may be written

$$p(\mathbf{r}_2,\cdots,\mathbf{r}_N|\mathbf{r}_1)=p(\mathbf{r}_2|\mathbf{r}_1)p(\mathbf{r}_3,\cdots,\mathbf{r}_N|\mathbf{r}_2)$$
$$-p(\mathbf{r}_2|\mathbf{r}_1)[p(\mathbf{r}_3,\cdots,\mathbf{r}_N|\mathbf{r}_2)-p(\mathbf{r}_3,\cdots,\mathbf{r}_N|\mathbf{r}_1,\mathbf{r}_2)].$$

Employing this to average the above expression for the exciting field, one obtains with no difficulty

$$\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle\approx\psi^{\text{inc}}(\mathbf{r})+\int d\tau' n(\mathbf{r}'|\mathbf{r}_1)T(\mathbf{r}')\langle\psi^E(\mathbf{r}|\mathbf{r}')\rangle$$
$$-R(\mathbf{r}|\mathbf{r}_1),\quad(2.24a)$$

where

$$R(\mathbf{r}|\mathbf{r}_1)\equiv\int d\tau_2 n(\mathbf{r}_2|\mathbf{r}_1)T(\mathbf{r}_2)\int d\tau_3\cdots d\tau_N$$
$$\times[p(\mathbf{r}_3,\cdots,\mathbf{r}_N|\mathbf{r}_2)-p(\mathbf{r}_3,\cdots,\mathbf{r}_N|\mathbf{r}_1,\mathbf{r}_2)]$$
$$\times\psi^E(\mathbf{r}|\mathbf{r}_2;\mathbf{r}_2,\cdots,\mathbf{r}_N).\quad(2.24b)$$

The remainder term $R(\mathbf{r}|\mathbf{r}_1)$ is expected to be small for several reasons. That the term is due to statistical correlations in position is apparent; if the scatterers are statistically independent, the two probability distributions become identical and the integrand vanishes. $R(\mathbf{r}|\mathbf{r}_1)$ also relies on fluctuations in the exciting field for its existence. If we approximated $\psi^E(\mathbf{r}|\mathbf{r}_2;\mathbf{r}_2,\cdots,\mathbf{r}_N)$ by its average value $\langle\psi^E(\mathbf{r}|\mathbf{r}_2)\rangle$, this latter quantity could be taken outside of all integrations except that over $\mathbf{r}_2$, and because of normalization of the probability distributions, the remaining $(N-2)$-fold integral vanishes.

Ultimately, the limit of statistical independence will be employed in order to obtain explicit results for those physical situations where position correlations are unimportant; nevertheless, it seems appropriate to comment on some approximate procedures for treating $R(\mathbf{r}|\mathbf{r}_1)$, to provide a starting point for handling some of the problems where correlations are an essential ingredient. The key to simplifying $R(\mathbf{r}|\mathbf{r}_1)$ lies in rewriting the exciting field in a form where only a few scatterer positions appear simultaneously, thus making most of the integrations trivial. One immediately apparent scheme is the multiple orders of scattering representation, given in Eq. (2.11). Making this substitution in Eq. (2.24b), one obtains a series representation for $R(\mathbf{r}|\mathbf{r}_1)$, each succeeding term involving one more integration and higher joint probability distributions. Evaluation of these terms may then be feasible, depending of course on the complexity of the statistics and the rapidity of convergence.

A somewhat simpler method, requiring, however, an additional assumption, is this: It has already been observed that only fluctuations in ψ^E are significant in computing $R(\mathbf{r}|\mathbf{r}_1)$. If one assumes that these fluctuations are predominately due to the variations in position of individual scatterers relative to the phase of the *averaged* exciting field acting on them (thus neglecting fluctuations in the latter), one is led to try the self-consistent approximation

$$\psi^E(\mathbf{r}|\mathbf{r}_2;\mathbf{r}_2,\cdots,\mathbf{r}_N)\approx\psi^{\text{inc}}(\mathbf{r})+\sum_{j=2}^{N}T(\mathbf{r}_j)\langle\psi^E(\mathbf{r}|\mathbf{r}_j)\rangle.$$

On inserting this expression in Eq. (2.24b) there results

$$R(\mathbf{r}|\mathbf{r}_1)\approx\int d\tau' n(\mathbf{r}'|\mathbf{r}_1)T(\mathbf{r}')\int d\tau''[n(\mathbf{r}''|\mathbf{r}')$$
$$-n(\mathbf{r}''|\mathbf{r}',\mathbf{r}_1)]T(\mathbf{r}'')\langle\psi^E(\mathbf{r}|\mathbf{r}'')\rangle.\quad(2.25)$$

If correlations that are short range relative to wavelength are being dealt with, so that the difference in densities vanishes outside of a region of small extent in comparison with wavelength, then $T(\mathbf{r}'')\langle\psi^E(\mathbf{r}|\mathbf{r}'')\rangle$ may be set equal to $T(\mathbf{r}_1)\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$ and taken outside the $\mathbf{r}''$ integration. Defining the *specific correlation volume* $v(\mathbf{r}_1)$ by

$$n(\mathbf{r}_1)v(\mathbf{r}_1)\equiv\int d\tau''[n(\mathbf{r}''|\mathbf{r}')-n(\mathbf{r}''|\mathbf{r}',\mathbf{r}_1)],$$

Eq. (2.25) becomes

$$R(\mathbf{r}|\mathbf{r}_1)\approx n(\mathbf{r}_1)v(\mathbf{r}_1)\int d\tau' n(\mathbf{r}'|\mathbf{r}_1)T(\mathbf{r}')T(\mathbf{r}_1)\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle.$$

[18] That is, the configuration is such that space is available for scatterer 1 without interpenetration.

$R(\mathbf{r}|\mathbf{r}_1)$ is thus negligible in this case. Notice that the product $n(\mathbf{r}_1)v(\mathbf{r}_1)\leqslant 1$ since at the upper limit the packing has reached the maximum value allowable by correlations. The integral is of the form evaluated earlier and shown to be negligible in the course of obtaining our criterion for approximating the exciting field. If correlations are not of this short-range nature, then $R(\mathbf{r}|\mathbf{r}_1)$ must be retained in some form in the integral equation Eq. (2.24a).

The higher partially averaged exciting fields may be treated in similar fashion to yield integral equations analogous to the one obtained for $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$. With reference to the system of equations (2.13), inspection of the above procedure indicates the following modifications: (1) the exciting field within the integral is replaced by the corresponding field with $\mathbf{r}_1$ absent; (2) a remainder term is added containing the appropriate higher conditional probabilities; (3) the exciting fields employed to obtain the summation of fixed scattered waves are calculated with $\mathbf{r}_1$ absent, so that the summation becomes an inhomogeneous term, involving the next lower partial average of the exciting field.

In closing this discussion, some comments on energy and measurement seem appropriate. Quantities such as energy density and energy flux, which are quadratic in the field amplitudes, must be computed separately, because the process of averaging cannot be expected to commute with the nonlinear operation of squaring the absolute magnitude of a field quantity. The two alternative propagation behaviors which can, at least in principle, be computed to obtain $|\langle\psi(\mathbf{r})\rangle|^2$ and $\langle|\psi(\mathbf{r})|^2\rangle$ correspond to the values that would be obtained experimentally employing amplitude-sensitive and energy-sensitive measuring devices, respectively.

The fundamental approximation $\psi^E(\mathbf{r}|\mathbf{r}_j;\mathbf{r}_1,\cdots,\mathbf{r}_N)\approx\psi(\mathbf{r}|\mathbf{r}_1,\cdots,\mathbf{r}_{j-1},\mathbf{r}_{j+1},\cdots,\mathbf{r}_N)$ derived in Sec. 2.4 is the relation required for treating the more involved field quantities related to energy. For example, forming $\psi(\mathbf{r})\psi^*(\mathbf{r}_0)$ from the first equation of the present section and averaging, invoking also Eq. (2.21), one obtains

$$\langle\psi(\mathbf{r})\psi^*(\mathbf{r}_0)\rangle-\langle\psi(\mathbf{r})\rangle\langle\psi^*(\mathbf{r}_0)\rangle$$
$$=-f^2\int d\tau' n(\mathbf{r}')\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}e^{-ik|\mathbf{r}_0-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|\,|\mathbf{r}_0-\mathbf{r}'|}\langle|\psi(\mathbf{r}')|^2\rangle$$
$$-f^2\int d\tau'\int d\tau'' n(\mathbf{r}')n(\mathbf{r}'')\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}e^{-ik|\mathbf{r}_0-\mathbf{r}''|}}{|\mathbf{r}-\mathbf{r}'|\,|\mathbf{r}_0-\mathbf{r}'|}$$
$$\times[\langle\psi(\mathbf{r}')\psi^*(\mathbf{r}'')\rangle-\langle\psi(\mathbf{r}')\rangle\langle\psi^*(\mathbf{r}'')\rangle]. \quad (2.26)$$

This is Foldy's result for identical isotropic point scatterers, and by interation techniques is convertible directly into an integral equation for $\langle|\psi(\mathbf{r})|^2\rangle$.[4] To obtain Eq. (2.26), Foldy assumed the validity of two approximations

$$\langle\psi^E(\mathbf{r}_j|\mathbf{r}_j)\psi^{E^*}(\mathbf{r}_j|\mathbf{r}_j)\rangle\approx\langle\psi(\mathbf{r}_j)\psi^*(\mathbf{r}_j)\rangle,$$
$$\langle\psi^E(\mathbf{r}_j|\mathbf{r}_j;\mathbf{r}_j,\mathbf{r}_k)\psi^{E^*}(\mathbf{r}_k|\mathbf{r}_k;\mathbf{r}_j,\mathbf{r}_k)\rangle\approx\langle\psi(\mathbf{r}_j)\psi^*(\mathbf{r}_k)\rangle,$$

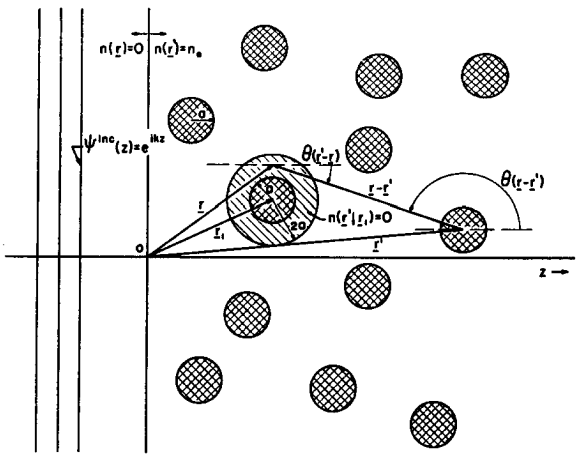


FIG. 3. The geometry of a semi-infinite array of identical spherical obstacles of radius a. Because interpenetration is excluded, the exciting field acting on the scatterer at $\mathbf{r}_1$ can have no singularities in the shaded region $|\mathbf{r}-\mathbf{r}_1|<2a$.

in addition to a previous assumption $\langle\psi^E(\mathbf{r}_j|\mathbf{r}_j)\rangle\approx\langle\psi(\mathbf{r}_j)\rangle$ needed to obtain an integral equation for $\langle\psi(\mathbf{r})\rangle$. All three of these approximations follow immediately by forming the appropriate products from the approximation above, and averaging.

The "quasi-crystalline" approximation

$$\langle\psi^E(\mathbf{r}|\mathbf{r}_1;\mathbf{r}_1,\mathbf{r}_2)\rangle\approx\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$$

employed by Lax[5] is not compatible with the present formulation. An objection may be raised; from inspection of the first two equations, (2.13), it can be seen that the exciting field with two scatterers fixed has a singularity located at the second scatterer, whereas $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$ does not. This would seem to determine a different behavior for the partial wave expansion coefficients. Perhaps a more significant objection is this: upon substituting the quasi-crystalline approximation in the first equation of the hierarchy, Lax obtains the approximate integral equation Eq. (2.24a) of this paper with no remainder term. From this it appears that the quasi-crystalline approximation is equivalent with the present formulation whenever $R(\mathbf{r}|\mathbf{r}_1)$ may be neglected.

3. HOMOGENEOUS SCATTERING MEDIUM

3.1 Solution of the Integral Equation for $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$

The goal of the following discussion is to obtain expressions for the exciting field and propagation characteristics of a homogeneous scattering medium wherein averaged field quantities have planar symmetry, and scatterers are distributed with constant density.

An important requirement from the computational viewpoint is statistical independence. The method of approaching this limit has mathematical difficulties,

but the results are physically reasonable; the anisotropic point scatterer is a mathematical fiction, and one must define its behavior carefully. A much less fundamental simplification is the employment of spherical obstacles in a geometry in which both exciting and scattered fields have axial symmetry. Full anisotropy with respect to the polar direction is retained, however, and the extension to full azimuthal anisotropy is immediate. Finally, a discussion of the results is given, in particular as they relate to earlier theories.

Consider a plane wave normally incident on a semi-infinite uniform array of spherical obstacles of radius a, as shown in Fig. 3. Choosing the plane $z=0$ for the boundary of the scattering region, the number of scatterers per unit volume is given by

$$n(\mathbf{r})=\begin{cases} n_0 & \text{(constant) for } z\gtrless 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.1a}$$

Restricting the conditional density for the time being to exclusion of interpenetration, one has

$$n(\mathbf{r}|\mathbf{r}_1)\begin{cases} \to n(\mathbf{r}) & \text{for } |\mathbf{r}-\mathbf{r}_1|\to\infty \\ =0 & \text{for } |\mathbf{r}-\mathbf{r}_1|\leqslant 2a. \end{cases} \tag{3.1b}$$

In any physical application the difference $n(\mathbf{r}|\mathbf{r}_1)-n(\mathbf{r})$ must be cut off at some finite range.

Fixing attention on the scatterer with center at $\mathbf{r}_1$, shown in Fig. 3, the exciting field is given by Eq. (2.24a),

$$\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle=\psi^{\text{inc}}(\mathbf{r})+\int d\tau' n(\mathbf{r}'|\mathbf{r}_1)T(\mathbf{r}')\langle\psi^E(\mathbf{r}|\mathbf{r}')\rangle - R(\mathbf{r}|\mathbf{r}_1). \tag{3.2}$$

Using partial wave expansions about origin $\mathbf{r}_1$ to represent the field quantities, the incident plane wave $\exp(ikz)$ has the well-known representation[19]

$$\psi^{\text{inc}}(\mathbf{r})=e^{ikz_1}e^{ik(z-z_1)} =e^{ikz_1}\sum_{n=0}^{\infty} a_n j_n(k|\mathbf{r}-\mathbf{r}_1|)P_n[\cos\theta(\mathbf{r}-\mathbf{r}_1)], \tag{3.3}$$

with expansion coefficients $a_n=(2n+1)(i)^n$. The $j_n(x)$ and $P_n(\mu)$ are, respectively, spherical Bessel functions and Legendre polynomials of nth order.[20] $\theta(\mathbf{r}-\mathbf{r}_1)$, shown in the figure, is the polar angle formed by the vector $\mathbf{r}-\mathbf{r}_1$ and the positive z direction.

Because of exclusion of interpenetration, the exciting field must be a regular solution of the unperturbed Helmholtz equation for all values of $\mathbf{r}$ in the excluded volume $|\mathbf{r}-\mathbf{r}_1|\leqslant 2a$, shown shaded in Fig. 3. From the planar symmetry of the problem, $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$ must depend on scatterer position only through z_1, the z component of $\mathbf{r}_1$, and, further, must be axially symmetric about the z direction in r. The most general form of the exciting field meeting these requirements is

$$\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle=\sum_{n=0}^{\infty} A_n(z_1)j_n(k|\mathbf{r}-\mathbf{r}_1|) \times P_n[\cos\theta(\mathbf{r}-\mathbf{r}_1)], \tag{3.4}$$

where the undetermined expansion coefficients $A_n(z_1)$ give the intrinsic dependence of $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$ on scatterer position.

In the partial wave formalism, the scattering operator may be defined by the relation

$$T(\mathbf{r}_1)j_n(k|\mathbf{r}-\mathbf{r}_1|)P_n[\cos\theta(\mathbf{r}-\mathbf{r}_1)] =B_n h_n(k|\mathbf{r}-\mathbf{r}_1|)P_n[\cos\theta(\mathbf{r}-\mathbf{r}_1)], \tag{3.5}$$

where by virtue of the radiation condition the scattered wave represents outgoing radiation.[21] The $h_n(x)$ are spherical Hankel functions of the first kind,[20] with superscript omitted because no ambiguity will arise. The expansion coefficients B_n are determined by the boundary conditions of the specific problem considered, and are assumed known.

More specifically, in the present notation the scattered radiation generated when a plane wave impinges on a single scatterer at the origin may be found by invoking the linearity of $T(\mathbf{r}_1)$ to obtain

$$T(0)e^{ikz}=\sum_{n=0}^{\infty} a_n B_n h_n(kr)P_n(\cos\theta)\underset{r\to\infty}{\longrightarrow} f(\theta)\frac{e^{ikr}}{r}, \tag{3.6a}$$

where

$$f(\theta)=\frac{1}{ik}\sum_{n=0}^{\infty}(2n+1)B_n P_n(\cos\theta) \tag{3.6b}$$

and comparison of Eq. (3.6a) with single-scatterer computations immediately gives the B_n. The far-field scattering amplitude $f(\theta)$ is the essential parameter required to specify behavior of the *scattering medium*, as will be seen.

Scattering caused by the exciting field is computed with the scatterer held fixed, so that $T(\mathbf{r}')$ commutes with intrinsic dependence on position, giving for a scattering center at field point $\mathbf{r}'$

$$T(\mathbf{r}')\langle\psi^E(\mathbf{r}|\mathbf{r}')\rangle=\sum_{j=0}^{\infty} A_j(z')B_j h_j(k|\mathbf{r}-\mathbf{r}'|) \times P_j[\cos\theta(\mathbf{r}-\mathbf{r}')]. \tag{3.7}$$

[19] See, for example, reference 11, p. 1574.

[20] We use the notation adopted by Morse and Feshbach, reference 11, throughout this Section. All the requisite formulas may be found in the tabulations following Chaps. 10 and 11 of the reference.

[21] Two modifications of Eq. (3.5) should be noted. One can allow azimuthal dependence of the scattered wave, thus employing the full spherical harmonics on the right-hand side. Because of the axial symmetry of the problem, however, these additional terms vanish identically upon performing the azimuthal integration in Eq. (3.2). Also, for obstacles which do not have spherical symmetry, an infinite series of partial waves are in general required on the right side. This modification again causes no trouble; one simply carries along another summation index in the computation. Physical interpretation of the results in terms of far-field amplitude of the single obstacle is unchanged.

Inserting the expressions (3.4), (3.3), (3.7) for the exciting, incident, and scattered fields, respectively, in integral equation (3.2), there results

$$\sum_{n=0}^{\infty}[A_n(z_1)-a_ne^{ikz_1}]j_n(k|\mathbf{r}-\mathbf{r}_1|)P_n[\cos\theta(\mathbf{r}-\mathbf{r}_1)]$$
$$=\sum_{j=0}^{\infty}B_j\int d\tau' n(\mathbf{r}'|\mathbf{r}_1)A_j(z')h_j(k|\mathbf{r}-\mathbf{r}'|)$$
$$\times P_j[\cos\theta(\mathbf{r}-\mathbf{r}')]. \quad (3.8)$$

This equation may be simplified by re-expanding the summation on the right side about the fixed origin $\mathbf{r}_1$. To this end, noting from Fig. 3 that $\cos\theta(\mathbf{r}-\mathbf{r}')=-\cos\theta(\mathbf{r}'-\mathbf{r})$, we may write

$$h_j(k|\mathbf{r}-\mathbf{r}'|)P_j[\cos\theta(\mathbf{r}-\mathbf{r}')]$$
$$=(-1)^jh_j(k|\mathbf{r}'-\mathbf{r}|)P_j[\cos\theta(\mathbf{r}'-\mathbf{r})]$$
$$=(i)^jP_j\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)h_0(k|\mathbf{r}'-\mathbf{r}|). \quad (3.9)$$

Here the first equality corresponds to rewriting the partial wave about origin $\mathbf{r}$ and follows from the relation $P_j(-\mu)=(-1)^jP_j(\mu)$. The second equality consists of a multipole expansion about origin $\mathbf{r}$ according to the expansion rule

$$h_j(kr)P_j(\cos\theta)=(-i)^jP_j\left(\frac{1}{ik}\frac{\partial}{\partial z}\right)h_0(kr) \quad (3.10)$$

for spherical partial waves.[22]

The expansion of the zero-order spherical Hankel function now appearing in Eq. (3.9) in terms of partial waves about the origin $\mathbf{r}_1$ is known to be given by[23]

$$h_0(k|\mathbf{r}'-\mathbf{r}|)=\sum_{n=0}^{\infty}(2n+1)\sum_{m=0}^{n}\epsilon_m\frac{(n-m)!}{(n+m)!}$$
$$\times\cos\{m[\varphi(\mathbf{r}-\mathbf{r}_1)-\varphi(\mathbf{r}'-\mathbf{r}_1)]\}$$
$$\times P_n^m[\cos\theta(\mathbf{r}-\mathbf{r}_1)]P_n^m[\cos\theta(\mathbf{r}'-\mathbf{r}_1)]$$
$$\times j_n(k|\mathbf{r}-\mathbf{r}_1|)h_n(k|\mathbf{r}'-\mathbf{r}_1|) \quad (3.11)$$

for $|\mathbf{r}-\mathbf{r}_1|<|\mathbf{r}'-\mathbf{r}_1|$, where $\epsilon_0=1$, $\epsilon_m=2$ for $m=1, 2, \cdots$. The $P_n^m(\mu)$ are the associated Legendre functions, and the notation on angles follows that employed previously. Because of the exclusion of the shaded region in Fig. 3 from the range of integration in Eq. (3.8), this expansion is valid within the integral provided $\mathbf{r}$ is in the neighborhood of the scatterer under examination, i.e., provided $|\mathbf{r}-\mathbf{r}_1|<2a$.

The advantage of restricting the problem to axial symmetry now becomes apparent. When Eq. (3.11) is inserted in the volume integral of Eq. (3.8) and the integration over azimuthal angle $\varphi(\mathbf{r}'-\mathbf{r}_1)$ performed, all the terms in $\cos[m\varphi(\mathbf{r}'-\mathbf{r}_1)]$ will vanish identically except those for which $m=0$. Thus, substituting Eqs. (3.9) and (3.11) in (3.8) and interchanging orders of summation and integration, the last may be rewritten in the considerably simpler and more compact form

$$\sum_{n=0}^{\infty}j_n(k|\mathbf{r}-\mathbf{r}_1|)P_n[\cos\theta(\mathbf{r}-\mathbf{r}_1)]\Big\{A_n(z_1)-a_ne^{ikz_1}$$
$$-a_n^*\sum_{j=0}^{\infty}(i)^jB_j\int d\tau' n(\mathbf{r}'|\mathbf{r}_1)A_j(z')P_j\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)$$
$$\times P_n\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)h_0(k|\mathbf{r}'-\mathbf{r}_1|)\Big\}=-R(\mathbf{r}|\mathbf{r}_1). \quad (3.12)$$

Here we have employed the multipole expansion (3.10) again, this time about $\mathbf{r}_1$ as origin, and noted that the complex conjugate plane wave expansion coefficients are given by $a_n^*=(-i)^n(2n+1)$.

Equation (3.12) may be developed further by the assumption of reasonably simple statistical correlations and an explicit form for the remainder term and expansion functions $A_j(z')$. The results appear hopelessly difficult to untangle, however, and so the alternative procedure is suggested of going to the limit of statistically independent point scatterers. This approach may be regarded as yielding the zero-order result in an expansion in which statistical correlations provide the perturbation. The advantages of this approach are twofold: first, $R(\mathbf{r}|\mathbf{r}_1)$ vanishes in the limit; second, from the resulting integrated equations the expansion functions $A_n(z')$ may be determined exactly.

This limit is taken in two steps. First, imagine finite scattering regions replaced by point scatterers with, however, unchanged scattering behavior. The scattered wave in what was formerly the interior of a scattering region is readily defined by analytic continuation; it is thus a regular solution of the unperturbed Helmholtz equation everywhere but at the point where the scatterer is located, where it has a high but finite order singularity. Keeping the conditional density unchanged, the length $2a$ appearing in Eq. (3.1b), formerly a quantity required to exclude interpenetration, now plays the role of a "distance of closest approach" for point scatterers.

[22] This expansion has been suggested by B. Van der Pol, Physica 3, 393 (1936), and a proof by induction has been given in an earlier report upon which the present work is in part based (see title reference). A somewhat simpler derivation is the following: J. W. Strutt, Baron Rayleigh, *The Theory of Sound* (Dover Publications, New York, 1945), Vol. II, p. 259, has shown that

$$h_n(kr)=(-i)^nP_n\left(\frac{1}{ik}\frac{d}{dr}\right)h_0(kr).$$

Replacing spherical radius r by cartesian z, this may be rewritten

$$h_n(kz)=\left[(-i)^nP_n\left(\frac{1}{ik}\frac{\partial}{\partial z}\right)h_0(kr)\right]_{r=z},$$

and now is interpreted as a boundary condition giving the value of the function in brackets along the z axis. Noting that the function in question is a solution of the wave equation with axial symmetry, and employing the boundary condition, the multipole expansion (3.10) follows immediately.

[23] Reference 11, p. 1466.

The procedure up to this point really only amounts to renaming a few quantities, since Eq. (3.12) is unchanged. Now, however, it is desired to let the distance of closest approach go to zero. In order to do this, one first extends the definition of the exciting field. $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$ has been defined so as to be a regular solution of the unperturbed Helmholtz equation with respect to $\mathbf{r}$ in the annular region $a\leqslant|\mathbf{r}-\mathbf{r}_1|\leqslant 2a$. In order to let $a\to 0$ without losing the dual functional nature of $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$, simply extend the definition into all space by analytic continuation according to the unperturbed Helmholtz equation. All the equations are formally unchanged by this process. Now consider the limit $n(\mathbf{r}'|\mathbf{r}_1)\to n(\mathbf{r}')$ in Eq. (3.12). $R(\mathbf{r}|\mathbf{r}_1)$ vanishes identically. The volume integrations are extended over the entire half-space $z'\geqslant 0$, with the exception of points in the immediate neighborhood of singularities at $\mathbf{r}_1$, which must be temporarily excluded in order that the integrals be properly defined. The operation of excluding a volume (to be specified shortly) containing the singular point $\mathbf{r}_1$ until after integration, then letting it go to zero, is denoted by script $\mathcal{P}(\mathbf{r}_1)$ in front of the integral sign.

With these considerations in mind, Eq. (3.12) now goes in the limit to

$$\sum_{n=0}^{\infty} j_n(k|\mathbf{r}-\mathbf{r}_1|)P_n[\cos\theta(\mathbf{r}-\mathbf{r}_1)]\Big\{A_n(z_1)-a_n e^{ikz_1}$$
$$-a_n^*\sum_{j=0}^{\infty}(i)^j B_j\mathcal{P}(\mathbf{r}_1)\int d\tau' n(\mathbf{r}')A_j(z')P_j\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)$$
$$\times P_n\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)h_0(k|\mathbf{r}'-\mathbf{r}_1|)\Big\}=0. \quad (3.13)$$

Notice that the $\mathbf{r}$ dependence of the original equation has been separated out into partial wave solutions of the unperturbed medium equation. If Eq. (3.13) is to be valid for all $\mathbf{r}$, then by virtue of the orthogonality of the $P_n[\cos\theta(\mathbf{r}-\mathbf{r}_1)]$ the term in brackets must vanish for each value of n, giving the set of reduced equations

$$A_n(z_1)=a_n e^{ikz_1}$$
$$+a_n^*\sum_{j=0}^{\infty}(i)^j B_j\mathcal{P}(\mathbf{r}_1)\int d\tau' n(\mathbf{r}')A_j(z')P_j\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)$$
$$\times P_n\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)h_0(k|\mathbf{r}'-\mathbf{r}_1|);\ n=0,\ 1,\ 2,\ \cdots. \quad (3.14)$$

At this point the functional behavior of the exciting field with respect to $\mathbf{r}$ is assumed to be unchanged as the scatterer is moved further into the medium, the intrinsic behavior of each of the partial waves with position being characterized by the common function $\exp(i\kappa z_1)$. κ, at present undetermined, will serve as a propagation constant for the *scattering medium*, its real and imaginary parts related to modified phase velocity and attenuation, respectively, in the usual manner.[24]

Thus, incorporating the plane wave expansion coefficients a_n into redefined $A_n(z_1)$ for convenience, consider the trial functions

$$A_n(z_1)=a_n A_n^0 e^{i\kappa z_1}. \quad (3.15)$$

Substitution of these expressions into the preceding equations gives a coupled set of integral equations for $\exp(i\kappa z_1)$ and the A_n^0,

$$A_n^0 e^{i\kappa z_1}=e^{ikz_1}+\sum_{j=0}^{\infty}(-1)^{j+n}(2j+1)B_j A_j^0 I_{jn}(z_1), \quad (3.16)$$

where for brevity we have written the integrals

$$I_{jn}(z_1)\equiv\mathcal{P}(\mathbf{r}_1)\int d\tau' n(\mathbf{r}')e^{i\kappa z'}P_j\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)$$
$$\times P_n\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)h_0(k|\mathbf{r}'-\mathbf{r}_1|). \quad (3.17)$$

The generalized principal value indicated by the operator $\mathcal{P}(\mathbf{r}_1)$ must now be specified. Notice first that the kernel $P_j(\partial/ik\partial z')P_n(\partial/ik\partial z')h_0(k|\mathbf{r}'-\mathbf{r}_1|)$ has a singularity of order $(j+n+1)$ at $\mathbf{r}_1$. The kernel consists of a sum of axial derivatives up to order $j+n$ of the source function appropriate to the unperturbed medium. Because of the behavior in the neighborhood of $\mathbf{r}_1$, this point must be approached with caution. There are several possible excluded volumes for which the integrations indicated in Eq. (3.17) may be carried out without difficulty (e.g., sphere of radius $|\mathbf{r}'-\mathbf{r}_1|=\epsilon\to 0$, thin circular disk of arbitrarily small but fixed radius with thickness $|z'-z_1|=\epsilon\to 0$, thin circular rod parallel to the z_1 axis with fixed length and radius $\to 0$). Each of these choices gives rise to a different solution of Eq. (3.16), as the reader may verify by straightforward computations (in the deleted sphere case particularly it is convenient to convert to surface integrals by means of Green's theorem). We are thus confronted with the problem of determining the "correct" volume to exclude.

Fortunately there is one additional physical requirement that must be met by $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$, that of differentiability with respect to $\mathbf{r}$, which will specify the appropriate manner for taking the principal value. Before passing to the limit of statistical independence, the exciting field was a well-behaved solution of the unperturbed medium equation. In particular, any desired number of differentiations with respect to the $\mathbf{r}$ coordinates could be taken inside the integral sign, and this commutation with the integration is just the requirement for commutation of the corresponding operation with the averaging process. Thus for example,

[24] The physical interpretation of κ will be established later.

applying the gradient operator to the exciting field as given by Eq. (2.24a) or (2.12), it is seen immediately that

$$\nabla\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle=\langle\nabla\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle. \tag{3.18}$$

Here scatterer position $\mathbf{r}_1$ is held fixed, and $\langle\nabla\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$ is defined by computing the gradient of the exciting field before averaging, starting from Eq. (2.10).

The commutation of gradient and averaging process stated in Eq. (3.18) is an important requirement, since the response of a scatterer will depend in general on the gradient of the exciting field as well as the field itself. Does Eq. (3.18) hold in the present case, where the exciting field has been analytically continued outside of the region of no singularities? One finds that it does, provided the principal value is taken by excluding a thin circular disk of thickness $|z'-z_1|=\epsilon\to 0$. If the principal value is taken by deleting a small sphere centered at $\mathbf{r}_1$, then the gradient of the averaged exciting field and the average of the gradient of the exciting field are in general unequal, i.e., Eq. (3.18) is not valid, except in the special case treated earlier of isotropic scatterers, where there is actually no need to employ principal values (because the $1/r$ singularity is adequately restrained by a factor r^2 in the volume element). Computations on which the above remarks are based are straightforward but extremely lengthy, and for this reason they are not presented here.

Equations (3.17) may now be integrated. Choosing cylindrical coordinates $(\rho')^2\equiv|\mathbf{r}'-\mathbf{r}_1|^2-(z'-z_1)^2$, $\varphi'\equiv\varphi(\mathbf{r}'-\mathbf{r}_1)$, $z'\equiv z'$ for the integration, the φ' integration only introduces a multiplying factor of 2π. Upon further noticing that the entire infinite slab $|z'-z_1|\leqslant\epsilon$ may be omitted from the range of integration, since the portion beyond the disk $\rho'=$constant will contribute nothing in the limit $\epsilon\to 0$, we have

$$\begin{aligned}
I_{jn}(z_1)&=2\pi n_0\lim_{\epsilon\to 0}\left\{\int_0^{z_1-\epsilon}dz'+\int_{z_1+\epsilon}^{\infty}dz'\right\}e^{i\kappa z'}P_j\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)P_n\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)\int_0^{\infty}d\rho'\rho'\frac{\exp\{ik[(\rho')^2+(z_1-z')^2]^{\frac{1}{2}}\}}{ik[(\rho')^2+(z_1-z')^2]^{\frac{1}{2}}}\\
&=2\pi n_0\lim_{\epsilon\to 0}\left\{\int_0^{z_1-\epsilon}dz'+\int_{z_1+\epsilon}^{\infty}dz'\right\}e^{i\kappa z'}P_j\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)P_n\left(\frac{1}{ik}\frac{\partial}{\partial z'}\right)\frac{e^{ik|z'-z_1|}}{k^2}\\
&=\frac{2\pi n_0}{k^2}\int_0^{z_1}dz'e^{i\kappa z'}(-1)^{j+n}e^{ik(z_1-z')}+\frac{2\pi n_0}{k^2}\int_{z_1}^{\infty}dz'e^{i\kappa z'}(+1)^{j+n}e^{-ik(z_1-z')}\\
&=\frac{-2\pi n_0(-1)^{j+n}}{i(\kappa-k)k^2}e^{ikz_1}+\frac{2\pi n_0}{k^2}\left\{\frac{(-1)^{j+n}}{i(\kappa-k)}-\frac{1}{i(\kappa+k)}\right\}e^{i\kappa z_1}.
\end{aligned}\tag{3.19}$$

Here in the second step the finitely oscillating portion of the ρ' integration has been discarded, invoking the usual arguments.[25] The following step comes from the observation that $P_n(\partial/ik\partial z')\exp(\pm ikz')=P_n(\pm 1)\times\exp(\pm ikz')=(\pm 1)^n\exp(\pm ikz')$. Convergence of the second z' integration is ensured if κ has a positive imaginary part, corresponding to attenuation in the *scattering medium*.

Substituting the final expression (3.19) for $I_{jn}(z_1)$ in Eq. (3.16) results in an equation containing four terms, two representing waves propagating according to the unperturbed medium equation, two propagating in the modified *scattering medium*; mathematically each pair must cancel separately if the equation is to be satisfied.

Consider first the pair corresponding to wave motion in the unperturbed medium. The vanishing of the coefficient of $\exp(ikz_1)$ requires physically that the original incident wave is exactly cancelled by waves generated by scatterers situated at the boundary plane $z=0$ of the scattering region. That these are the only scatterers involved is immediately apparent if one rewrites Eq. (3.17) in terms of surface integrals by means of a Green's theorem. The statement that the incident wave is cancelled by waves generated at the boundary is known as the *extinction theorem*; it was first given by Ewald in 1912 in connection with optical dispersion.[26] As a consequence of the theorem, one obtains in the present case the interesting result

$$\kappa=k+(2\pi n_0/k)F(0), \tag{3.20}$$

where $F(0)$ is the forward-scattered amplitude *with multiple scattering effects included*, given by

$$F(0)=\frac{1}{ik}\sum_{j=0}^{\infty}(2j+1)A_j{}^0B_j. \tag{3.21}$$

The intrinsic dependence on scatterer position has been normalized out in the definition of $F(\theta)$; from Eqs. (3.7), (3.15), utilizing the asymptotic form of the Hankel functions for large argument, one has

[25] Endowing the matrix medium with a small attenuation which one then lets go to zero after integration.

[26] P. P. Ewald (dissertation, München, 1912); Ann. Physik **49**, 1 (1916), considered the crystalline case. Isotropic media were considered by C. W. Oseen, Ann. Physik **48**, 1 (1915). For a modern account, see M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, New York, 1959), p. 99.

$$e^{-i\kappa z'}T(\mathbf{r}')\langle\psi^E(\mathbf{r}|\mathbf{r}')\rangle \xrightarrow[|\mathbf{r}-\mathbf{r}'|\to\infty]{} \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{ik|\mathbf{r}-\mathbf{r}'|}$$

$$\times\sum_{j=0}^{\infty}(2j+1)A_j{}^0B_jP_j(\cos\theta)=F(\theta)\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}. \quad (3.22)$$

The physical interpretation of Eq. (3.20), giving the modified propagation behavior in terms of forward-scattered amplitude, is as follows: The mechanism by which the scatterer modifies the behavior of the exciting field is simply interference of the scattered wave with the main beam, producing a shadow. Because the relative phase of the two waves is continuously changing in all other directions, this interference is only significant in the forward direction.[27]

Equation (3.20) remains valid when statistical correlations are considered. Referring back to Eq. (3.12), the conditional density $n(\mathbf{r}'|\mathbf{r}_1)$ may be rewritten as the sum of the ordinary density $n(\mathbf{r}')$ and the correlation density $n(\mathbf{r}'|\mathbf{r}_1)-n(\mathbf{r}')$. If the field point $\mathbf{r}_1$ is situated sufficiently far within the medium that the correlation density is negligible at the boundary plane $z'=0$, then the integrals involving $n(\mathbf{r}'|\mathbf{r}_1)-n(\mathbf{r}')$ will generate modified fields propagating according to the scattering medium behavior. Similar remarks apply to the remainder term [see, for example, Eq. (2.25)]. Now the integrals containing $n(\mathbf{r}')$ may be converted by Green's theorem, using the trial functions (3.15), and application of the extinction theorem immediately gives Eq. (3.20).

It is perhaps curious that this fundamental relation between forward-scattering amplitude and modified medium behavior was not noticed in some of the earlier work on dispersion of light. The omission was probably due to the fact that the modified dielectric constant could be obtained more directly by virtue of the basically isotropic nature of the scattering. Neglecting the additional complications required for the Lorentz-Lorenz corrections,[5] the situation is exactly analogous to the procedure followed by Foldy and mentioned earlier for the special case of isotropic point scatterers, where we obtained the propagation factor $\kappa(\mathbf{r})$ in Eq. (2.22). It was not necessary to consider either the exciting field or the far-field amplitude $F(\theta)$ explicitly, simply because it was possible to formulate an integral equation for the total field $\langle\psi(\mathbf{r})\rangle$ directly.

On arriving at the extinction theorem, pause for a moment to consider the total field $\langle\psi(\mathbf{r})\rangle$. It has been established that $\langle\psi(\mathbf{r})\rangle$ is obtained from the exciting field by quadrature, according to Eq. (2.18). Comparing the first two terms on the right-hand side of Eq. (2.18) with the corresponding terms in the integral equation (3.2) for $\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle$, it may be seen that the extinction theorem will serve to exactly cancel the contribution of $\psi^{\text{inc}}(\mathbf{r})$ to the total field. A factor $\exp(i\kappa z)$ may be taken out of the remaining surface and volume integrals, reducing the integrals to constants independent of position. We conclude that the total field must have the form

$$\langle\psi(z)\rangle=\text{const}\times e^{i\kappa z},$$

and based on this comment the interpretation of κ as the propagation constant appropriate to the *scattering medium* is established. It should be stressed that bulk properties are being dealt with here; the interpretation of κ and the forward-amplitude theorem (3.20) break down as soon as one gets sufficiently close to the surface that $n(\mathbf{r}'|\mathbf{r}_1)-n(\mathbf{r}')$ is appreciable there.

Returning to the computation in progress, the exciting field and modified propagation constant finally are obtained in their dependence on the known single scatterer behavior by equating the coefficient of $\exp(i\kappa z_1)$, the remaining term in Eq. (3.16), to zero. Physically this corresponds to the statement that the exciting field is made up entirely of contributions from scatterers in the immediate neighborhood of the field point (again in light of transformation from volume to surface integrals by Green's formula). Utilizing the forward-scattered amplitude theorem (3.20) to effect some simplification, there results

$$A_n{}^0=1-\frac{2\pi n_0(-1)^n}{ik^2(\kappa+k)}\sum_{j=0}^{\infty}(-1)^j(2j+1)B_jA_j{}^0;$$
$$n=0, 1, 2, \cdots. \quad (3.23)$$

The implicit nature of these relations, giving the $A_n{}^0$ in terms of themselves, is of course just the manifestation of the implicitness inherent in multiple scattering computations; we have observed this twice before, once in the implicit starting representation (2.10) for a frozen configuration and again in the integral equation (2.24a) for the averaged exciting field. There is an important advantage, however. Equations (3.23) may be inverted immediately to obtain the expansion coefficients in terms of κ and the known far-field amplitude $f(\theta)$ appropriate to the single-scatterer problem. One may iterate Eqs. (3.23) indefinitely and collapse the resulting series, or alternatively and more simply, iterate once and eliminate the unknown summation, which reappears intact, between the resulting and original equations.

Following the latter procedure and noting that the residual series are simply the forward and backscattered amplitudes $f(0)$ and $f(\pi)$, respectively, of the single-scatterer problem, one obtains

$$A_n{}^0=1-\frac{2\pi n_0(-1)^nf(\pi)}{k(\kappa+k)+2\pi n_0f(0)}; \quad n=0, 1, 2, \cdots, \quad (3.24a)$$

where, in accordance with Eq. (3.6b), the single-

[27] This is precisely the physical basis for the theorem relating forward-scattered amplitude to total scattering cross section in the single scatterer case, as has been pointed out by M. Lax, Phys. Rev. 78, 306 (1950).

scatterer amplitudes are given by

$$f(0)=(1/ik)\sum_{n=0}^{\infty}(2n+1)B_n$$

$$f(\pi)=(1/ik)\sum_{n=0}^{\infty}(-1)^n(2n+1)B_n. \quad (3.24b)$$

Substitution of Eqs. (3.24a, b) in the forward amplitude theorem (3.20) and rearrangement of some terms yields the expression

$$\left(\frac{\kappa}{k}\right)^2=\left[1+\frac{2\pi n_0 f(0)}{k^2}\right]^2-\left[\frac{2\pi n_0 f(\pi)}{k^2}\right]^2. \quad (3.25)$$

This is the central result of the computation: *The behavior of the scattering medium, characterized by the complex propagation constant κ, may be specified explicitly in terms of the number of scatterers per unit volume and the far field amplitude $f(\theta)$ obtained for a single scatterer.*

The exciting field may be obtained by substituting Eqs. (3.24a) and (3.15) in the partial wave expansion (3.4). By comparison with the plane wave expansion (3.3), note that the first term on the right-hand side of Eq. (3.24a) generates the expansion coefficients of a forward-traveling plane wave, while the second term corresponds to the coefficients of a back-traveling plane wave. Thus one obtains for the exciting field the expression

$$\langle\psi^E(\mathbf{r}|\mathbf{r}_1)\rangle=\langle\psi^E(z|z_1)\rangle=e^{ikz_1}\left\{e^{ik(z-z_1)}-\frac{2\pi n_0 f(\pi)}{k(\kappa+k)+2\pi n_0 f(0)}e^{-ik(z-z_1)}\right\}. \quad (3.26)$$

The exciting field $\langle\psi^E(z|z_1)\rangle$ acting on a scatterer fixed at z_1 is the sum of forward- and back-traveling plane waves, generated by scatterers situated in front of and behind z_1, respectively. Both these waves propagate as if in the unperturbed medium, as expected. The intrinsic dependence of the waves on scatterer position is given by the term $\exp(ikz_1)$; as the scatterer is moved further into the medium, both plane waves exhibit periodic variation in phase and are, in general, attenuated. The situation is indicated schematically in Fig. 4. The field exciting a scatterer at z_1 consists of the two plane waves shown, while at the later station z_1' both plane waves are still present in the same proportion and same relative phase, but with decreased amplitude.

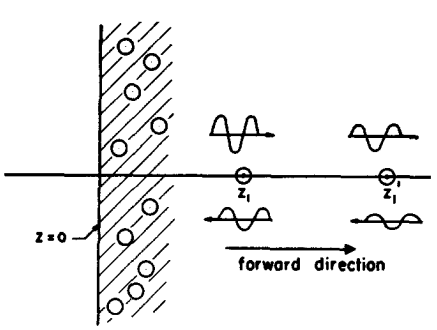


FIG. 4. The exciting field acting on a scatterer at z_1 consists of two plane waves propagating in the forward and back direction. At a station z_1' further in the medium both waves are still present in the same proportion and same relative phase, but with diminished amplitude.

A detailed physical picture of the role played by the individual scatterer in generating the intrinsic behavior of the exciting field is now apparent. The scattered radiation in the forward direction combines with the forward wave of the exciting field so as to produce on the average a slight modification in phase and amplitude. The net result of these modifications is described by the propagation constant κ. A similar situation occurs in the back direction, the major difference being that the amplitude of the back wave is *increased*, as is readily seen from Fig. 4.

In terms of energy considerations, one can say roughly the following: *The scatterer serves to remove some energy from the forward wave; part of this energy is dissipated within the scatterer itself, depending on the absorption mechanisms present, and another portion is transferred to the back wave.* This comment, while not a rigorous statement of energy balance,[28] may help to provide an explanation of the apparently paradoxical result that the averaged field may suffer attenuation even though there are no true loss mechanisms present.

The remainder of this section will be devoted to a discussion of propagation behavior in the scattering medium under certain restrictions. Consider first the case of *weak scattering density*, defined by the requirement $|n_0 f(\theta)/k^2|\ll 1$. This situation arises physically whenever the number of scatterers per unit volume is sufficiently small, or the individual scatterer is a weak perturbation on the matrix medium. Neglecting terms in Eq. (3.25) that are quadratic in the scattering density $n_0 f(\theta)/k^2$, there results

$$(\kappa/k)\approx 1+[2\pi n_0 f(0)/k^2]. \quad (3.27)$$

This result may be obtained by the "thin slab" approximation discussed by Fermi[29] and later extended to anisotropic scattering by Lax.[5] Equation (3.27) is also equivalent to Eq. (A4), the result obtained by summation of multiple orders of scattering, for the special case of isotropic scattering.

Making use of the forward amplitude theorem relating *extinction* (i.e., scattering plus absorption) cross

[28] The rigorous statement of energy balance is obtained exactly as in the case of a single scatterer by computing the net flux due to the exciting and scattered fields through a closed surface containing one scatterer.

[29] E. Fermi, *Nuclear Physics* (University of Chicago Press, Chicago, Illinois, 1950), p. 201, revised edition (notes by J. Orear, A. Rosefeld, and R. Schluter).

section Q_{ex} to the imaginary part of the forward scattered amplitude by[27]

$$Q_{ex} \equiv Q_{scat} + Q_{abs} = (4\pi/k)\ \mathrm{Im} f(0),$$

we may rewrite the preceding equation in the form

$$\frac{\kappa}{k} \approx 1 + \frac{2\pi n_0}{k^2}\mathrm{Re} f(0) + \frac{i n_0 Q_{ex}}{2k}. \tag{3.28}$$

Thus, in the approximation of weak scattering density, modified phase velocity is determined by the real part of the forward-scattered amplitude of the single-scatterer problem; fractional energy loss per unit volume is given simply by single-scatterer total cross section per unit volume. This establishes the rule of additive behavior of cross sections.[30]

If, on the other hand, one is concerned with isotropic scatterers, Eq. (3.25) becomes, noting that $f(0) = f(\pi) \equiv f$, say,

$$(\kappa/k) = [1 + (4\pi n_0 f/k^2]^{\frac{1}{2}}. \tag{3.29}$$

This is just the result obtained by Foldy,[4] given previously in Eq. (2.22), for the homogeneous case. It has been employed by Carstensen and Foldy[31] and several other workers[32] to describe scattering of sound by resonant air bubbles in water, and experimental results are for the most part in good agreement with those predicted. As mentioned previously, Eq. (3.29) also governs dispersion of transverse electromagnetic radiation in metals and homogeneous plasmas (although not dielectrics because of the complications involved in Lorentz-Lorenz corrections).

In addition to providing a marriage between the two formerly separate bodies of theory expressed in Eqs. (3.28) and (29) and indicating their range of applicability, our present results have penetrated more deeply into the detailed nature of multiple scattering interactions to provide a rational basis for the discussion of questions of energy balance—or alternatively in quantum-mechanical problems conservation of number of particles. Ekstein, for example, has criticized the radiation condition on the basis of apparent nonconservation of particles.[33] While the problem of outgoing *versus* standing-wave boundary conditions has not been resolved here, the dual plane-wave nature of the exciting field may provide a quite reasonable physical explanation of any apparent energy inbalance or nonconservation of particles associated with the total field.

It seems appropriate to comment on Lax's results.[5] He obtains for the complex propagation constant, using our notation insofar as possible,

$$\left(\frac{\kappa}{k}\right)^2 = 1 + \frac{4\pi n_0 c f(\boldsymbol{\kappa},\boldsymbol{\kappa})}{k^2}. \tag{3.30}$$

Here $f(\boldsymbol{\kappa},\boldsymbol{\kappa})$ is the forward-scattered amplitude associated with a single scatterer *embedded in the scattering medium*; c is an undetermined constant, assumed near unity, expressing the ratio of exciting field to total field. Notice that Lax's expression will reduce to Eq. (3.27) in the limit of weak scattering density, taking $c=1$. It could also be made compatible with the isotropic result of Eq. (3.27) by the device of choosing $c = f(0)/f(\boldsymbol{\kappa},\boldsymbol{\kappa})$. It appears inadequate, however, in comparison with the present result (3.25), in failing to indicate the important role played by back-scattered amplitude. We would criticize his assumption of proportionality between the exciting and total fields, which amounts to discarding the dual functional nature of the exciting field entirely, with consequent loss of essential information.

3.2 Multiple Scattering in the Rayleigh Limit

It is worthwhile to consider some simple examples at high- and low-frequency limits, where one can obtain analytical results, as opposed to involved series expressions, without difficulty. As a result of these considerations, it will be seen that whenever the single scatterer is equivalent to a *proper* perturbation on the bulk properties of the medium, in a sense to be discussed, the description of propagation in the scattering medium by Eq. (3.25) is adequate even in the limit where scatterers are close packed so as to completely fill the available volume. In addition to building further confidence in the theory, we demonstrate that it can reasonably be expected to apply in a variety of problems where such close packing is a requirement. The polycrystalline solid represents an important example of such problems. Because of elastic anisotropy, individual single-crystal grains scatter acoustic waves; the problem of acoustic wave propagation in polycrystals is thus a case of multiple scattering where by definition *the scatterers occupy the entire volume.* Similar remarks apply to electromagnetic propagation in polycrystalline dielectrics whose structure has lower than cubic symmetry.

In the following discussion some examples from acoustics are considered in the low-frequency limit. The scatterers might be oil bubbles or (nonresonant) air bubbles in water, for example. We represent the single scatterer by a sphere of radius a, shown in Fig. 5, having density ρ' and propagation constant k'

[30] Employed in discussing neutron diffraction in crystals by O. Halpern, M. Hamermesh, and M. H. Johnson, Phys. Rev. **59**, 981 (1941). For an application in acoustics, see P. S. Epstein and R. R. Carhart, J. Acoust. Soc. Am. **25**, 553 (1953).

[31] E. L. Carstensen and L. L. Foldy, J. Acoust. Soc. Am. **19**, 481 (1947).

[32] See reference 16. Also, E. Silberman, J. Acoust. Soc. Am. **29**, 925 (1957); F. E. Fox, S. R. Curley and G. S. Larson, *ibid.* **27**, 534 (1955); E. Meyer and E. Skudrzyk, Acustica **3**, 434 (1953). The case of particle suspensions in water has been investigated by R. J. Urick and W. S. Ament, J. Acoust. Soc. Am. **21**, 115 (1949). They present an interesting alternate derivation of Eq. (3.25) above.

[33] H. Ekstein, Phys. Rev. **89**, 490 (1953).

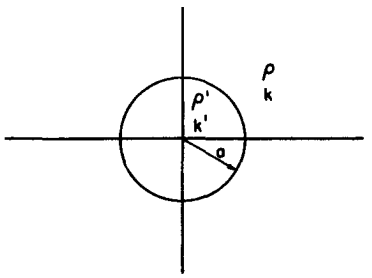


FIG. 5. The single spherical scatterer in the acoustic example. Density and propagation constant appropriate to obstacle and supporting matrix are shown.

distinct from the parameters ρ, k of the matrix medium alone. From physical considerations, one obtains the boundary conditions that pressure and normal component of particle motion are continuous across the bubble surface.

We further suppose the frequency is low enough that individual bubbles are very small in comparison with wavelength, so that the product ka is negligible compared with unity. Under these conditions Rayleigh has shown that only the first two terms in the series (3.24b) for the far-field amplitude are important. The expansion coefficients for isotropic and dipole radiation, respectively, become[34]

$$B_0=\frac{i(ka)^3}{3}\left[\frac{(k'/k)^2}{\rho'/\rho}-1\right]$$
$$B_1=\frac{i(ka)^3}{3}\left[\frac{\rho'-\rho}{\rho+2\rho'}\right]. \tag{3.31}$$

It is convenient to introduce the fractional volume of bubbles δ, defined as

$$\delta=\tfrac{4}{3}\pi a^3 n_0. \tag{3.32}$$

Inserting the preceding relations in Eqs. (3.24b) and (3.25), the propagation constant κ for the "bubbly" medium is obtained in the form

$$\left(\frac{\kappa}{k}\right)^2=\left\{1+\delta\left[\frac{(k'/k)^2}{\rho'/\rho}-1\right]\right\}\left\{1+3\delta\left[\frac{\rho'-\rho}{\rho+2\rho'}\right]\right\}. \tag{3.33}$$

Suppose first that phase velocities and densities are equal inside and outside the bubbles, but that the presence of some dissipation mechanism gives rise to an attenuation α' in the bubble material. In this case, substituting $\rho'=\rho$, $k'=k+i\alpha'$ in Eq. (3.33), the complex index of refraction of the bubbly medium becomes

$$\kappa/k=[1-\delta(\alpha'/k)^2+2i\delta\alpha'/k]^{\frac{1}{2}}. \tag{3.34}$$

The phase velocity and attenuation associated with the bubbly medium are given in the usual manner by the real and imaginary parts of the expression on the right-hand side of Eq. (3.34); it is interesting to note that in general the phase velocity in the bubbly medium is altered even though the phase velocity was everywhere constant before averaging over configurations of bubbles. Note also that attenuation in the bubbly medium is not simply proportional to α'.

Both effects are of course due to multiple scattering interactions and disappear at sufficiently low fractional volumes; if δ is neglected in comparison with unity in the real and imaginary parts of Eq. (3.34) separately, there results

$$\kappa\approx k+i\delta\alpha';\quad \delta\ll 1. \tag{3.34a}$$

In this limit phase velocity is unchanged, while attenuation in the bubbly medium is simply the attenuation within the bubbles weighted by the fractional volume of bubbles. The latter statement is merely a rephrasing of the additive rule for total cross sections mentioned following Eq. (3.28).

Equation (3.34) should also be examined at high fractional volumes of bubbles. In the light of observations made by Rayleigh[35] one can infer that obstacles of small extent in comparison with wavelength generate a response proportional to their volume and independent of their shape. Thus in the present instance we can replace spherical bubbles by cubes of the same volume, for example, without changing their scattering response; having done this, it is then permissible to close pack the bubbles up to the limit $\delta=1$, which corresponds to complete replacement of the unperturbed medium by bubbles. Equation (3.34) becomes

$$\kappa=k+i\alpha'=k';\quad \delta=1, \tag{3.34b}$$

and the bubbly medium behaves in this limit like a homogeneous medium composed of the bubble material. Equation (3.34b) by itself is not a surprising result; what is interesting is the inference that the original Eq. (3.34) is valid in this example for all physically allowable fractional volumes.

Now consider the situation with no intrinsic losses. Introducing the elastic moduli or reciprocal compressibilities M, M' related to phase velocities by $v=(M/\rho)^{\frac{1}{2}}$, $v'=(M'/\rho')^{\frac{1}{2}}$, Eq. (3.33) may be written

$$\left(\frac{1}{\bar{v}}\right)^2=\left\{\frac{1-\delta}{M}+\frac{\delta}{M'}\right\}\left\{\rho+3\delta\rho\left[\frac{\rho'-\rho}{\rho+2\rho'}\right]\right\}. \tag{3.35}$$

Here $\bar{v}$ is phase velocity in the bubbly medium, defined by the relation $\kappa=\omega/\bar{v}$ whenever κ is real.[36] The terms in the first bracket represent simply the volume averaged compressibility. The terms in the second bracket are not so readily interpretable; it will be helpful to first consider the approximate behavior as $\rho'\to\rho$. If the reader neglects the second-order terms in the density difference, he may verify that Eq. (3.35) yields the remarkably simple result

$$\bar{v}\underset{\rho'\to\rho}{\longrightarrow}(\bar{M}/\bar{\rho})^{\frac{1}{2}}, \tag{3.35a}$$

[34] See reference 22, p. 283.

[35] See reference 22, p. 149.

[36] We note that for certain values of densities κ may become imaginary, in which case the right-hand side of Eq. (3.35) is related instead to attenuation.

where the volume averaged compressibilities and densities are given by

$$1/\bar{M}=(1-\delta)/M+\delta/M'$$
$$\bar{\rho}=(1-\delta)\rho+\delta\rho'. \qquad (3.35b)$$

In other words, in this limit *phase velocity in the bubbly medium is expressed in the usual manner provided one employs the composition-averaged values of the appropriate parameters.* This result was suggested by Wood[37] to describe the case of air bubbles in sea water, an example where Eq. (3.35a) is not valid according to the present theory, incidentally, because of the large disparity in densities.

The present concern is with showing the adequacy of the general description (3.25) of the scattering medium in the limit of large fractional volume of scatterers. The result (3.35) is not very helpful in this regard because of the involved manner in which it depends on densities. It turns out that consideration of a more difficult example will yield simpler results.

The failure of Eq. (3.35) to yield a simple description of propagation in the bubbly medium may be seen by examination of the single-scatterer boundary conditions. Recall that the normal components of particle displacement across the bubble surface are required to be continuous; no analogous requirement is imposed on the tangential components, with the result that they become discontinuous as soon as density disparities are introduced [if one then lets $\rho' \to \rho$, discontinuities become small in comparison with the displacement themselves, permitting the simpler behavior (3.35a)]. In short, Eq. (3.35) describes a wave motion riddled with discontinuous particle displacements.

This difficulty may be eliminated by considering instead the problem of scattering by an array of solid elastic spheres embedded in a homogeneous isotropic elastic solid. Because the solid media can support shearing stresses, one must include in addition to the previous parameters the shear moduli G, G' of matrix and scatterer material, respectively. The single scatterer is illustrated in Fig. 6, which also indicates the bulk parameters appropriate to each medium.

The single-scatterer problem has been considered in detail by Ying and Truell.[38] Briefly, the boundary conditions require the *vector* surface tractions and particle displacements to be continuous across the interface between the two media, thus eliminating the difficulty encountered in the fluid case. To satisfy these conditions in the presence of an incident plane longitudinal wave, four additional waves must be introduced; scattered compressional (irrotational) and transverse

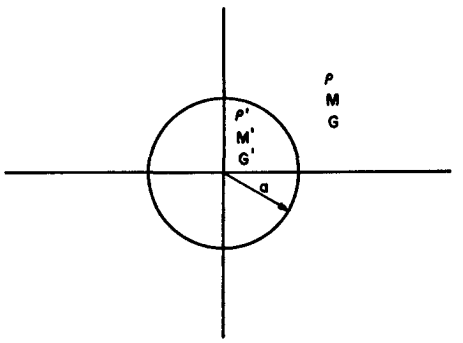


FIG. 6. The single elastic obstacle of radius a, showing the densities and bulk moduli appropriate to obstacle and matrix material.

(solenoidal) waves in the region $r \geqslant a$ outside the scatter, and a corresponding pair inside the scatterer.

Fortunately the phenomenon of *mode conversion*, the generation of transverse scattered waves by a longitudinal incident wave, will not materially affect the present discussion, as we may see by a simple symmetry argument. In the limiting case of point scatterers which we have employed, the exciting field has planar symmetry. Any transverse components of the exciting field must thus be plane waves propagating in the forward and back directions, each with an associated polarization normal to the forward direction, defining the direction of particle displacement. But the problem has complete axial symmetry about the forward direction, so there can be no preferred polarization direction. We conclude there can be no transverse components of the exciting field. The transverse modes are self-extinguishing in the sense that they never appear in the exciting field, even though they are being generated at all scattering sites. For this reason it is permissible to represent the exciting field by a single scalar potential. Similarly a single scalar potential is sufficient to represent that portion of the scattered wave which will not vanish identically upon integrating over scattering sites, so the formal procedure employed to obtain $\langle\psi^E(z|z_1)\rangle$ carries through without requiring extensions.

By inspection of the results obtained by Ying and Truell, we find that the first three expansion coefficients are significant in the low-frequency limit, and these are given by, dropping terms of order $(ka)^2$ in comparison with unity,[38]

$$B_0=\frac{i(ka)^3}{3}\left[\frac{4(G'-G)-3(M'-M)}{3M'-4(G'-G)}\right]$$

$$B_1=\frac{i(ka)^3}{9}\left[\frac{\rho'-\rho}{\rho}\right]$$

$$B_2=-\frac{4i(ka)^3}{3}\left[\frac{G(G'-G)}{3M(2G'+3G)+4G(G'-G)}\right]. \qquad (3.36)$$

[37] A. B. Wood, *A Textbook of Sound* (G. Bell and Sons, London, England, 1957), p. 361.

[38] C. F. Ying and R. Truell, J. Appl. Phys. **27**, 1086 (1956). Because of differences in notation, the expansion coefficients A_{en} obtained in this reference must be converted for present use by the relation $B_n=(-i)^{n+1}kA_{en}^*/(2n+1)$. We have also corrected the sign of A_{e1} in the reference.

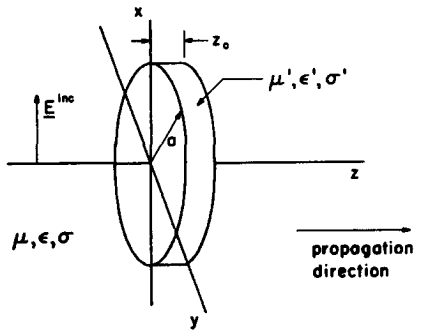


FIG. 7. The single scatterer for the electromagnetic case is a thin disk of thickness z_0, and radius a much greater than wavelength. Only normal incidence is considered, with electric field polarized in the x direction.

Substitution in Eqs. (3.24b) and (3.25) gives the propagation behavior in the scattering medium. Note that the form of the dependence on elastic moduli is extremely involved. This complexity (and for that matter even the presence of shear moduli G, G' in an expression describing longitudinal wave motion) may be attributed to the fact that in the neighborhood of a scatterer both media are undergoing some lateral contractions, so that the moduli M, M' appropriate to infinite plane waves are no longer adequate.

Fortunately we may prevent the shear moduli from intruding in the equations and demonstrate a simple result by taking the special case $G'=G$. The phase velocity $\bar{v}$ in the scattering medium is then given by

$$\bar{v}=(\bar{M}/\bar{\rho})^{\frac{1}{2}}, \tag{3.37}$$

where the volume averaged parameters $\bar{M}$, $\bar{\rho}$ are those defined earlier in Eqs. (3.35b) and the restriction to small disparities in density required in the fluid bubble case (3.35a) is no longer necessary.

The preceding examples indicate that at least in the Rayleigh limit the theory is not restricted to small fractional volumes of obstacles. In certain cases the propagation constant behaves correctly even at the limit $\delta=1$ where scatterers completely fill the available volume. In cases where this limiting behavior does not result, the difficulty can be attributed to aberrations in single-scatterer behavior, in the light of which simple results are not expected.

3.3 Multiple Scattering in the Geometrical Optics Limit

It is also instructive to investigate the region of very high frequencies or, what is equivalent, scatterers very large in comparison with wavelength. In this case one must consider a highly idealized example, an array of thin disks of large diameter. With the exception of the small spheres treated earlier, this is the only other geometrical shape, in three dimensions, meeting two requirements of present interest: (1) simple analytic expressions may be obtained for the far field of the single scatterer; (2) fractional volumes of unity are physically allowable.

A Kirchhoff approximation[39] is used for the case of plane electromagnetic waves normally incident on a uniform array of disks having electrical parameters distinct from those of the medium in which they are embedded. The single scatterer is shown in Fig. 7; the intrinsic parameters permeability μ, dielectric constant ϵ, and conductivity σ for the surrounding medium, and the corresponding primed values for the disk material itself, are indicated.

If the disk is many wavelengths in diameter, i.e., if we assume $\mathrm{Re}(k)a\gg 1$, then to good accuracy the scattered waves in the neighborhood of the disk will be simply plane reflected and transmitted waves. Suppose that corrections due to the cylindrical boundary strip at the periphery of the disk may be made as small as desired by making the disk thin in comparison with diameter; actually for simplicity, the somewhat more severe restriction $\mathrm{Re}(k)z_0\ll 1$ is introduced, i.e., the disk is thin in comparison with (unperturbed) wavelength.

The reasons for choosing such a highly specialized example are several: First, the electric field vectors in the neighborhood of the disk have only one cartesian component, and thus are describable by a single scalar potential; second, these vectors are readily obtained from a one-dimensional treatment of transmission and reflection from a plane sheet; finally, analytic expressions for the far fields may be obtained by means of a simple integral representation.

The one-dimensional calculation of transmission and reflection from a plane sheet of material has been discussed by Stratton.[40] A unit incident wave $\mathbf{E}^{inc}$ polarized in the x direction, as indicated in Fig. 7, may be written

$$\mathbf{E}^{inc}=(-1/ik)\nabla\times\mathbf{y}_1\psi^{inc}(z), \tag{3.38}$$

in terms of the scalar potential $\psi^{inc}(z)=e^{ikz}$, where $\mathbf{y}_1$ is the unit vector in the y direction. The propagation constant k is given by

$$k^2=\mu\epsilon\omega^2+i\mu\sigma\omega, \tag{3.39}$$

the imaginary part representing losses due to nonzero conductivity. The propagation constant k' within the sheet is given by Eq. (3.39) with primed values of the electrical parameters employed on the right-hand side.

Assuming reflected and transmitted waves of amplitude E^-, $1+E^+$, respectively, and plane waves in both forward and back directions within the sheet $0\leqslant z\leqslant z_0$, E^+ and E^- are obtained by requiring continuity of the tangential components of electric and magnetic field vectors at the two interfaces between sheet and surrounding medium.

Employing a Green's function technique, the scattered

[39] Reference 12, p. 463.
[40] Reference 12, p. 511 ff.

potential $\psi^+(\mathbf{r})$ in the forward hemisphere $z>0$ may now be written[41]

$$\psi^+(\mathbf{r}) = -(1/2\pi) \times \int_\sigma d\sigma \psi^+(\sigma) \cos(n,R)(\partial/\partial R)(e^{ikR}/R), \quad (3.40)$$

where the integration is over the plane surface σ of the disk, as shown in Fig. 8. R is distance from surface element to field point $\mathbf{r}$, and (n,R) is the angle between R and the normal $\mathbf{n}$ to the surface. For the value $\psi^+(\sigma)$ of the scattered potential on the surface we use E^+, the difference between the transmitted wave $1+E^+$ and the (unit amplitude) incident wave.

In the limit of large distances from the origin in the forward direction, R may be replaced by spherical radius r in the integrand. The entire integrand may then be taken outside the integral sign, resulting in

$$\psi^+(\mathbf{r}) \underset{r\to\infty}{\longrightarrow} -\frac{ika^2E^+}{2}\frac{e^{ikr}}{r}; \quad \theta=0. \quad (3.41)$$

The forward-scattered far field amplitude is immediately obtained from Eq. (3.41) as

$$f(0) = (-ika^2E^+/2). \quad (3.42a)$$

In analogous fashion, employing the amplitude E^- of the reflected wave, the back-scattered far field amplitude is given by

$$f(\pi) = (-ika^2E^-/2), \quad (3.42b)$$

and substitution of $f(0)$ and $f(\pi)$ in the general expression (3.25) gives the propagation constant κ for the scattering medium in the form

$$\left(\frac{\kappa}{k}\right)^2 = \left[1-\frac{i\pi a^2 n_0 E^+}{k}\right]^2 + \left[\frac{\pi a^2 n_0 E^-}{k}\right]^2. \quad (3.43)$$

Consider the two limiting cases of opaque and transparent obstacles. The opaque case arises when the conductivity σ' of the disks becomes very large and they appear essentially as perfect conductors. In terms

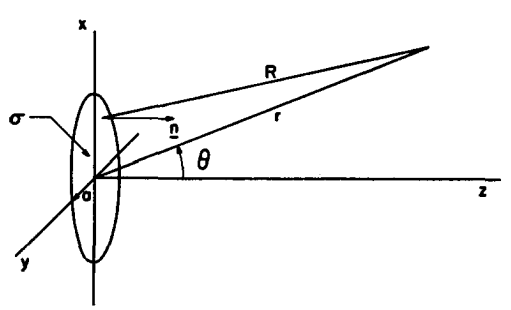


FIG. 8. The field in the forward hemisphere $z>0$ may be obtained by integration over the (shadow side) surface σ of the disk. The geometry is shown for the integration according to Eq. (3.40).

[41] A. Sommerfeld, *Optics* (Academic Press, Inc., New York, 1954) p. 199 ff.

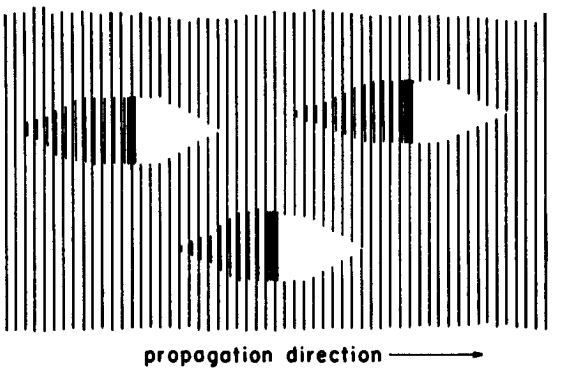


FIG. 9. The electromagnetic disks of high conductivity behave as perfect reflectors: in the immediate neighborhood of each disk on the illuminated side standing waves are formed, while on the shadow side total shadow is created.

of *skin depth*[42] $d' = (2/\omega\mu'\sigma')^{\frac{1}{2}}$, the requirement is simply that the disks be much thicker than skin depth, i.e., $z_0/d' \gg 1$. Imposing this condition, one obtains from Stratton's results[40]

$$E^+ = -1 \qquad E^- = -1. \quad (3.44)$$

As illustrated in Fig. 9, individual disks behave as perfect reflectors in the case; near the disks in the back direction standing waves are set up so that the resulting electric vector vanishes at the disk surface, while on the forward side a total shadow has been formed.

If we substitute Eqs. (3.44) in the preceding equation and note the scattering cross section $Q_s \approx 2\pi a^2$ to the degree of accuracy implied in the one-dimensional calculation, the result may be written

$$\kappa \approx k + i(n_0 Q_s/2). \quad (3.45)$$

In extracting the square root, it was necessary to drop terms of order $(n_0Q_s/k)^2$ compared with unity, in accordance with the fundamental criterion (2.20) governing all the present work. Whether Eq. (3.43) is more generally valid in this particular example is a matter requiring separate investigation.

Note from Eq. (3.45) that waves in the scattering medium are attenuated just according to the additive rule (3.28) for cross sections (because of the limitation mentioned, the fraction of power lost per unit unperturbed wavelength must be small). This result appears reasonable, even when one observes that no energy is being *dissipated* by the obstacles. Phase velocity, on the other hand, is unchanged. Both these results may be seen from Fig. 9: Because of the shadow formed behind each scatterer, the net amplitude in the forward direction is decreased per unit path length by a fraction equal to the area of each shadow times the number of shadows per unit volume, or just $n_0Q_s/2$; surfaces of equal phase, on the other hand, are everywhere spaced as in the unperturbed medium, so there is no modification in phase velocity.

[42] Reference 12, p. 504.

Next we consider the case of nearly transparent disks. This situation arises when there are arbitrary disparities in electrical parameters between the disk and surrounding medium, but the path length z_0 within the disks is so short that the internal wave functions do not change appreciably in the course of traversing a disk. This is achieved by requiring that $|k'z_0|\ll 1$, or equivalently that the disks be thin in comparison with both (internal) wavelength and skin depth.

If we neglect second-order terms in $|k'z_0|$, the scattered amplitudes according to Stratton are given by[40]

$$E^+=(ikz_0/2\mu\mu'k^2)[2\mu\mu'k(k'-k)+(\mu k'-\mu' k)^2]$$
$$E^-=(ikz_0/2\mu\mu'k^2)[(\mu k')^2-(\mu' k)^2]. \tag{3.46}$$

Substituting these expressions in Eq. (3.43), replacing k and k' according to Eq. (3.39), and collecting terms yields

$$\kappa^2=\bar{\mu}\bar{\epsilon}\omega^2+i\bar{\mu}\bar{\sigma}\omega, \tag{3.47a}$$

where the electrical parameters are the volume-averaged parameters defined in terms of fractional volume $\delta=\pi a^2 z_0 n_0$:

$$\bar{\mu}=(1-\delta)\mu+\delta\mu'$$
$$\bar{\epsilon}=(1-\delta)\epsilon+\delta\epsilon' \tag{3.47b}$$
$$\bar{\sigma}=(1-\delta)\sigma+\delta\sigma'.$$

Thus again the propagation constant is related to the bulk parameters of the medium exactly as if the medium were homogeneous, provided one employs composition averaged values. Furthermore, by inspection of the preceding calculations, in particular Eq. (3.40), it is clear that the *shape* of the disk is not important, but rather the *area*. We could just as well have employed disks of square cross section, for example; having done this and then proceeding to close pack the disks, it appears that Eq. (3.47a) is correct at the limit $\delta=1$, where the scattering medium has become a homogeneous medium made up of disk material.

As a final comment, observe that the medium containing similarly oriented disks is certainly highly anisotropic. We were restricted to the case of normal incidence as this is the only situation allowing description of the vector field quantities by a single scalar potential. The extensions of the theory required to handle the vector wave equation, and hence such problems as the present example, in full, will be sketched in the closing discussion.

4. DISCUSSION

The intent of this paper has been to establish the machinery needed in problems involving multiple scattering of waves. Beginning with the concept of "configurational" averaging introduced by Foldy,[4] a criterion (2.20) was found enabling one to obtain integral equations for the averaged field quantities. It was verified that Foldy's equations for isotropic scatterers could all be based on this single criterion.

The situation became more complex with the introduction of anisotropic scatterers and statistical correlations in position; in order to obtain explicit results, it was necessary to pass to the limit of statistical independence. The one result essentially independent of statistical correlations was the forward-scattered amplitude theorem (3.20) of multiple scattering.

Perhaps the most important single result was the description (3.25) of propagation in the scattering medium in terms of the far-field amplitude of the single scatterer. By appropriate specialization, this formula was seen to contain both the square-root law (3.29) for isotropic scatterers, and the addition rule (3.28) for total cross sections valid for anisotropic scatterers in the limit of weak scattering density. By then considering some specific examples in the following sections, it was demonstrated in both high- and low-frequency limits that the theory is apparently adequate for treating all physically allowable densities n_0 of obstacles satisfying the fundamental criterion (2.20), provided that effects of correlation in position can be neglected.

In addition, the double plane wave nature of the exciting field was discussed, and it was pointed out that interplay between these waves in the course of scattering may provide a satisfactory answer to questions relating to energy conservation.

In closing, extensions are listed, some of which are immediate, the others posing yet unanswered questions which are in need of further investigation.

1. Removing the restriction of identical scatterers. This extension is straightforward, and has been considered by Foldy[4] and Lax.[5] One employs an additional distribution, say $p(\alpha)$, in any number of distributed parameters α of the scatterers. The approximation required in this case must read: "The exciting field may be replaced by the total field that would act if the scatterer were not present, *regardless of the kind of scatterer present.*" This statement certainly holds if one requires the fundamental criterion to hold in the extreme case, say $n_0 Q_s^{\max}/k\ll 1$, where $Q_s^{\max}$ is the largest scattering cross section encountered in the allowed range. Presumably one could develop a weaker criterion if necessary. The result (3.25) now becomes

$$\left(\frac{\kappa}{k}\right)^2=\left[1+\frac{2\pi n_0}{k^2}\langle f(0)\rangle\right]^2-\left[\frac{2\pi n_0}{k^2}\langle f(\pi)\rangle\right]^2, \tag{3.25a}$$

where

$$\langle f(\theta)\rangle=\int d\alpha p(\alpha) f(\theta;\alpha)$$

is the far-field amplitude averaged over α.

2. Correlations in position. These effects will become important whenever there are significant interaction forces between scatterers. Also, the exclusion of interpenetration is a reasonable physical requirement, and the consequences should be investigated.

A perturbation approach for the exciting field has been hinted at earlier, and appears straightforward in principle. Assuming an explicit form for the remainder term $R(\mathbf{r}|\mathbf{r}_1)$ in Eq. (3.2), the equation may be carried to the stage analogous to Eq. (3.14) and integrated without much difficulty for reasonably simple forms of the conditional density. This will result in a set of implicit equations somewhat more involved than Eqs. (3.23) for the expansion coefficients, which together with the forward amplitude theorem (3.20) specifies the solution of the problem completely.

The only difficulty lies in disentangling these equations to obtain an explicit result, and an iteration procedure appears to be the best bet. In the right side of the new Eqs. (3.23), the old A_n^0, κ are substituted. The equation then yields a new set of expansion coefficients on the left side, and substituting these in the forward amplitude theorem gives a corrected κ. This procedure may then in principle be repeated until convergence is obtained, with κ and the A_n^0 taking on final values. Notice that this process is analogous to a Born iteration technique where instead of neglecting scattering in the trial functions, one neglects *corrections to scattering* due to statistical correlations in position.

3. Averaged field quantities related to energy. The equations describing energy density $\langle|\psi|^2\rangle$ and energy flux $\langle\psi^*\nabla\psi-\psi\nabla\psi^*\rangle$ merit further investigation. The first of these has been discussed by Foldy only briefly.[4] Apparently additional approximations are required to obtain a governing equation which is at all tractable. Consideration of averaged energy flux should lead to the phenomenological equations of radiative transfer discussed in great detail by Chandrasekhar,[43] together with criteria for their applicability.

The quantity $\langle|\psi|^2\rangle$ is equally as important as $\langle\psi\rangle$ in experimental application, and what is really needed is more detailed investigation along the lines of Sec. 3 so as to facilitate direct comparison of behavior of the two field quantities and provide a more unified picture of multiple scattering processes.

4. Cylindrical obstacles. Extension of the theory to cover scattering by a parallel array of infinite cylinders offers interesting possibilities. It is well known that separation[44] of the vector wave equation in any of the cylindrical coordinate systems can be accomplished relatively simply. At the minor cost of restricting the discussion essentially to two dimensions, once could retain the description of various vector fields of electromagnetic and elasticity theory by a single scalar potential, and at the same time increase the scope of the theory to cover physical mechanisms which cannot be treated by the present methods, e.g., polarization-induced charge density on the surface of obstacles. In addition, cylinder arrays offer certain practical advantages over spheres in the construction of artificial media. In elastic solids, for example, specimens may be simply drilled and filled with the desired scattering material. With artificial acoustic and dielectric lenses, cylinders only require support at their endpoints, and thus need not be embedded in a supporting matrix.

5. The vector wave equation. Finally, in order to encompass the full vector range of problems of interest in elasticity and electromagnetic theory, the present theory must be extended. The clue to performing this extension is provided by the single-scatterer problem. It has been shown that solutions of the vector wave equation are completely describable by three scalar potentials; two of these are related to solenoidal wave motion, the third generates irrotational waves.[44]

Instead of a single scalar potential representing the incident wave, the latter may be thought of as a column vector having as components the three scalar potentials in question. The boundary conditions will specify three new scalar potentials representing the scattered radiation; in general, each of these is determined in part by all three of the incident potentials. The problem is still linear, however, and the above comments immediately suggest introducing a 3×3 scattering matrix which when operating on the triplet of incident potentials, or incident "vector," yields the scattered "vector." Notice that this *mode conversion* or interaction between *different* components of incident and scattered wave, resulting mathematically in the scattering matrix being nondiagonal, is just what prevents the vector problem from being treated as three individual scalar problems.

At this point the theoretical discussion of the present paper may be reformulated, and it appears that the resulting integral equations for various field quantities will essentially be modified only by the extensions described, i.e., the scalar potentials replaced by 3-vectors, the scattering operator by a 3×3 scattering matrix. Analytically, one will have the problem of finding eigenvectors of these equations, in addition to considering their functional properties.

ACKNOWLEDGMENTS

The authors are indebted to several people for their helpful comment and criticism during the course of this work. In particular, they take pleasure in thanking Professor G. F. Newell of Brown University, Professor H. A. Bethe of Cornell University, T. Einwohner and Dr. J. Yos of Avco. Dr. R. E. Watson of Avco was kind enough to read the manuscript and to make many valuable suggestions.

APPENDIX I. SUMMATION OF MULTIPLE ORDERS OF SCATTERING

It is worthwhile to consider for a simple case the infinite series consisting of the bracket in Eq. (2.19)

[43] S. Chandrasekhar, *Radiative Transfer* (Clarendon Press, Oxford, England, 1950).

[44] Reference 11, Chap. 13. We use the term "separation" in the sense employed by Morse and Feshbach, of splitting into distinct vector solutions.

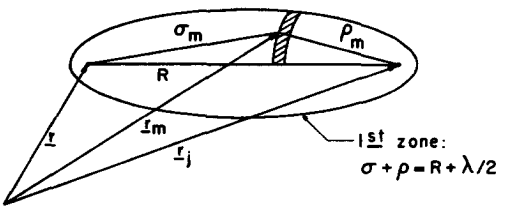


FIG. 10. For the isotropic point source at $\mathbf{r}_j$, waves rescattered once from $\mathbf{r}_m$ in the course of getting to the evaluation point $\mathbf{r}$ will be no more than π out of phase for all $\mathbf{r}_m$ lying in the first ellipsoidal half-period zone, defined by $\sigma_m+\rho_m \leqslant R+\lambda/2$.

operating on $T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N)$, in order to show roughly at least how a modified index of refraction characterizing the *scattering medium* is generated from multiple orders of scattering.

We have

$$\chi=\{1+\sum_{m\neq k,j} T(\mathbf{r}_m)+\sum_{m\neq k} T(\mathbf{r}_m)\sum_{n\neq m,j} T(\mathbf{r}_n)+\cdots\} \times T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N). \quad \text{(A1)}$$

Take a random array of isotropic point scatterers distributed with constant density n_0, and scattering operator defined by

$$\chi_0\equiv T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N)\equiv f\psi_j\frac{e^{ik|\mathbf{r}-\mathbf{r}_j|}}{|\mathbf{r}-\mathbf{r}_j|},$$

where we have written $\psi_j\equiv\psi(\mathbf{r}_j|\mathbf{r}_1,\cdots',\mathbf{r}_N)$ for brevity. The far field amplitude f is determined by the physics of the scattering process, and will in general be complex. The second term in brackets in Eq. (A1) now becomes (at field point $\mathbf{r}$)

$$\sum_{m\neq k,j} T(\mathbf{r}_m)T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N) = \sum_{m\neq k,j} f\cdot f\psi_j\frac{e^{ik|\mathbf{r}_m-\mathbf{r}_j|}}{|\mathbf{r}_m-\mathbf{r}_j|}\frac{e^{ik|\mathbf{r}-\mathbf{r}_m|}}{|\mathbf{r}-\mathbf{r}_m|} = f^2\psi_j\sum_{m\neq k,j}\frac{e^{ik(\rho_m+\sigma_m)}}{\rho_m\sigma_m},$$

where $\rho_m\equiv|\mathbf{r}_j-\mathbf{r}_m|$ and $\sigma_m\equiv|\mathbf{r}-\mathbf{r}_m|$, shown in Fig. (10), are the distances from "source" $\mathbf{r}_j$ to scatterer m and scatterer m to field point $\mathbf{r}$, respectively, and the distance from source to field point is given by $R\equiv|\mathbf{r}-\mathbf{r}_j|$.

Now consider those scatterers in the ellipsoid defined by $\rho_m+\sigma_m\leqslant R+\lambda/2$, shown in Fig. 10. The relative phase of rescattered waves is determined by the total path length $\rho_m+\sigma_m$, hence contributions from those scatterers on the connecting line between $\mathbf{r}_j$ and $\mathbf{r}$ are exactly in phase, the waves from other scatterers gradually falling further behind and the worst laggards, on the surface of the ellipsoid, being just π behind in phase. Thus waves rescattered once within the ellipsoid all add *with the same sign*, and the region serves as a first "Fresnel half-period zone." The higher zones are defined by $R+n\lambda/2\leqslant\rho_m+\sigma_m\leqslant R+(n+1)\lambda/2$, $n=1, 2, 3, \cdots$, and it is easy to verify that the contribution from each zone has alternate sign from that of the preceding one.

We can readily estimate the contribution χ_1 from the first zone. Neglecting fluctuations, since we seek *averaged* field quantities, the sum may be replaced by an integral, giving

$$\chi_1\approx f^2\psi_j\sum_{m\neq k,j}^{\text{1st zone}}\frac{e^{ik(\rho_m+\sigma_m)}}{\rho_m\sigma_m} \approx n_0f^2\psi_j\int_{\text{1st zone}} d\tau\frac{e^{ik(\rho+\sigma)}}{\rho\sigma}.$$

Choosing bipolar coordinates ρ, σ, θ for the integration, where θ is the azimuthal angle of rotation about the connecting line between $\mathbf{r}$ and $\mathbf{r}_j$, it has been shown that the volume element is given by $d\tau=(\rho\sigma/R)d\rho d\sigma d\theta$.[45] Taking R large so that terms of order λ/R can be neglected in comparison with unity, χ_1 becomes, integrating first over the strip of constant ρ shown in Fig. 10,

$$\chi_1\approx 2\pi n_0f^2\psi_j\cdot\frac{1}{R}\int_0^R d\rho\int_{R-\rho}^{R-\rho+\lambda/2} d\sigma e^{ik(\rho+\sigma)} =\frac{4\pi in_0fR}{k}f\psi_j\frac{e^{ikR}}{R}.$$

Noting that the spatial variation of the phase of χ_1 is identical to that of χ_0, we may calculate the first zone contribution χ_2 to secondary scattering, the next term in the series of Eq. (A1), in precisely the same manner:

$$\chi_2\equiv\sum_{m\neq k}^{\text{1st zone}} T(\mathbf{r}_m)\sum_{n\neq m,j} T(\mathbf{r}_n)T(\mathbf{r}_j)\psi(\mathbf{r}|\mathbf{r}_1,\cdots',\mathbf{r}_N) \approx\sum_{m\neq k}^{\text{1st zone}} T(\mathbf{r}_m)\chi_1 \approx(1/2!)(4\pi in_0fR/k)^2f\psi_j(e^{ikR}/R).$$

At this point it is quite easy to show by induction, following the procedure outlined above, that the first zone contribution to the nth order of scattering is just

$$\chi_n\approx\frac{1}{n!}\left(\frac{4\pi in_0fR}{k}\right)^n f\psi_j\frac{e^{ikR}}{R}. \quad \text{(A2)}$$

Further consideration of the effect of the remaining zones indicates that the total contribution of all zones to each order is just half the contribution from the first zone, a not surprising result. Thus, modifying each χ_n by a factor 2^{-n} because of the cumulative nature of the change, the series of Eq. (A1) may be summed,

[45] This (nonorthogonal) coordinate system has been discussed by T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956), p. 203.

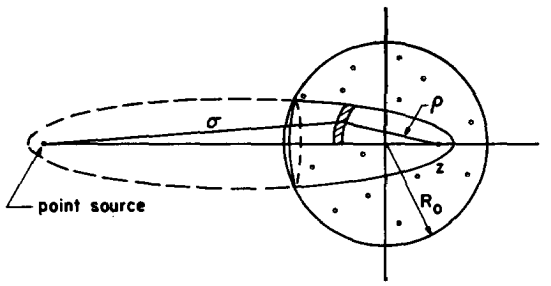


FIG. 11. A specified array of scatterers, more or less uniformly distributed throughout the spherical region of radius R_0, is illuminated by a point source. The first half-period zone for rescattering to the point z, constructed exactly as in Fig. 10, is now only partially filled, the dashed portion being unoccupied.

obtaining

$$\chi \approx \sum_{n=0}^{\infty} \frac{\chi_n}{2^n}$$

$$\approx \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{2\pi i n_0 f R}{k}\right)^n f\psi_j \frac{e^{ikR}}{R}$$

$$= f\psi_j(e^{ik'R}/R), \quad \text{(A3)}$$

which describes an outgoing spherical wave propagating in a *scattering medium* with complex index of refraction given by

$$k'/k \approx 1+(2\pi n_0 f/k^2). \quad \text{(A4)}$$

This expression is only correct in the limit of *weak scattering density*, i.e., for $|k'/k| \to 1$, as shown in the text. Note that, since f is just the *forward scattered amplitude* for a single scatterer,[46] and by virtue of the theorem[47] relating the imaginary part of this quantity to *extinction* (i.e., scattering plus absorption) cross section Q_{ex} we may rewrite Eq. (A4) as

$$\frac{k'}{k} \approx 1+\frac{2\pi n_0}{k^2}\mathrm{Re}f+\frac{in_0 Q_{ex}}{2k}. \quad \text{(A5)}$$

Thus in this approximation propagation in the scattering medium is characterized by a modified phase velocity dependent upon the real part of forward scattered amplitude, and fractional energy loss per unit length equal to the extinction cross section per unit volume $n_0 Q$.

As mentioned in the text, we presented the above estimate simply as heuristic support for the interpretation of the series as a field with modified propagation behavior. It is interesting to note, however, an important by-product of the discussion. For problems in which the field behavior is desired for either an explicit "frozen" or a random configuration of weak scattering strength, it may be practical to use the multiple orders of scattering approach if good convergence can be obtained with only the first few orders. In view of the role shown to be played by succeeding orders, it appears that one can now readily estimate the number of orders required for convergence in a given problem.

For example, consider the behavior of a specified configuration of isotropic point scatterers, distributed more or less uniformly throughout a sphere of radius R_0, when illuminated by an isotropic point source, as shown in Fig. 11. The first zone is now only partially occupied with scatterers, being nearly empty near the illuminated surface of the sphere and containing the maximum number near the shadowed surface. This indicates that scatterers on the shadow side will be more subject to multiple scattering effects, i.e., "shielded" from the incident wave by those scatterers on the illuminated side.

It is reasonable that the fluctuations encountered at each order will be comparable with the average field of that order, provided of course that no strong periodicities in scatterer positions occur. The average effect of the nth order of scattering may be estimated by integrating over the occupied part of the first zone. The calculation runs exactly as above except that the ρ integration only extends to R_0+z (for scatterers lying on the z axis). Multiple scattering effects are strongest at the point in deepest shadow, where $z=R_0$, and at this point the ratio of incident field due to nth-order scattering, say $\psi^{\mathrm{inc}(n)}$, to original incident wave is given by

$$\frac{\psi^{\mathrm{inc}(n)}}{\psi^{\mathrm{inc}}} \approx \frac{1}{n!}\left(\frac{4\pi i n_0 f R_0}{k}\right)^n. \quad \text{(A6)}$$

The number of orders that must be employed for good convergence is given by the value of n which makes the above ratio small in absolute value compared to one. If $|4\pi n_0 f R_0/k| \ll 1$, then multiple scattering can be neglected and the scatterers treated as independent obstacles each illuminated by the original incident wave alone. If $|4\pi n_0 f R/k| \gg 1$, the scattering strength is no longer weak and the simple zone treatment given here breaks down. The multiple orders of scattering treatment is no longer practical in this latter case, however, as too many orders are required for convergence. It is in the intermediate range, $|4\pi n_0 f R_0/k| \approx 1$, where only two or three orders are required, that the multiple orders of scattering approach may prove fruitful in a variety of physical situations.

Note also that the criterion given in Eq. (A6) involves both real and imaginary parts of the forward scattered amplitude, so that even in the limit of vanishing extinction cross-section multiple scattering effects can still be significant. The reason for this is that individual scattered waves can add in such a manner as to build up a field on the shadowed scatterers which, while undiminished in amplitude, may be appreciably modified in phase velocity and hence wavelength.

Finally, since the criterion is independent of the source location, by moving the source off to infinity in the usual manner we see that the results are equally valid for incident plane waves.

[46] See Eqs. (3.6).
[47] Preceding Eq. (3.28).

Exact Solution of a Critical Problem for a Slab

ROMAN ŻELAZNY
Institute of Nuclear Research of Polish Academy of Science, Institute of Theoretical Physics of Warsaw University, Warsaw, Poland
(Received October 25, 1960)

By using the Case method of expansion of the angular neutron distribution into series with respect to eigenfunctions of the plane Boltzmann equation, the critical problem of a slab has been formulated. By means of symmetry considerations, the problem of boundary condition has been reduced to one singular integral equation, which has been treated by classical methods. This treatment has given an integral equation for expansion coefficients, which by means of a simple transformation can be reduced to a Fredholm type with a regular kernel, and an additional equation, which plays the role of an exact critical condition. The methods and results of numerical calculations will be published soon.

I. INTRODUCTION

THE purpose of this paper is to present an application of method, developed by Case[1] for the one-velocity Boltzmann equation of the neutron transport theory, to a new problem, not yet solved by means of classical exact methods. Case's method was inspired by Van Kampen's work[2] on problems of plasma oscillations. In the latter work it was stated that in problems involving transport equations, the eigenfunctions may be distributions in Schwartz's sense. Observables in such cases are obtained by integration of eigenfunctions and, therefore, are the "proper" functions.

In the above quoted work, Case has constructed in explicit form the eigenfunctions of the Boltzmann equation; he has proved their completeness and applied these methods to elementary problems of the neutron transport theory, as for example to Milne problem, to Albedo problem. He has also obtained Green functions in an infinite and semi-infinite medium.

It should be mentioned that a similar method has been used previously by Lafore and Millot[3] but with a somewhat different aspect, without completeness theorem and in a less effective form.

In this paper the problem of critical thickness of a slab is treated by means of the eigenfunction expansion, and an integral equation reducible to Fredholm type for expansion coefficients with an exact critical condition is derived. Numerical calculations of these equations are now being done and will be published soon.

II. EIGENFUNCTIONS OF THE ONE-VELOCITY BOLTZMANN EQUATION

The one-velocity Boltzmann equation in the neutron transport theory in the case of plane symmetry has the form

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

Following Case's treatment, we are looking for the eigenfunctions in the form

$$\psi(x,\mu)=e^{-x/\nu}\phi_\nu(\mu).$$

The equation for $\phi_\nu(\mu)$ which results from Eq. (1) is

$$(1-\mu/\nu)\phi_\nu(\mu)=(c/2)\int_{-1}^{1}\phi_\nu(\mu')d\mu'=c/2, \tag{2}$$

where it has been taken into account that the right side can be arbitrarily normalized.

Now, the full solution of this equation gives

$$\phi_\nu(\mu)=(c/2)P(\nu/\nu-\mu)+\lambda(\nu)\delta(\mu-\nu).$$

The unknown function $\lambda(\nu)$ can be determined by normalization condition, i.e.,

$$\int_{-1}^{1}\phi_\nu(\mu)d\mu=1=(c/2)\nu P\int_{-1}^{1}\frac{d\mu}{\nu-\mu}+\lambda(\nu)\int_{-1}^{1}\delta(\mu-\nu)d\nu. \tag{3}$$

Two cases must be taken into consideration: (a) $\nu\notin(-1, 1)$. Equation (3) takes the form

$$1=(c/2)\nu\int_{-1}^{1}(d\mu/\nu-\mu)=c\nu\tanh^{-1}(1/\nu)\nu,$$

which gives two discrete eigenvalues $\pm\nu_0$. The corresponding eigenfunctions have the form

$$\psi_{0\pm}(x,\mu)=\phi_{0\pm}(\mu)e^{\mp x/\nu_0}, \tag{4a}$$

where

$$\phi_{0\pm}=(c/2)(\nu_0/\nu_0\mp\mu). \tag{4b}$$

(b) $\nu\in(-1, 1)$. Equation (3) has now the form

$$1=(c/2)\nu P\int_{-1}^{1}(d\mu/\nu-\mu)+\lambda(\nu).$$

It is easily seen that

$$\lambda(\nu)=1-c\nu\tanh^{-1}\nu \tag{5a}$$

and

$$\psi_\nu(x,\mu)=\phi_\nu(\mu)e^{-x/\nu}, \tag{5b}$$

where

$$\phi_\nu(\mu)=(c\nu/2)P(1/\nu-\mu)+\lambda(\nu)\delta(\mu-\nu). \tag{5c}$$

[1] K. M. Case, Ann. Phys. 9, 1 (1960).
[2] N. G. Van Kampen, Physica 21, 949 (1955).
[3] P. Lafore and J. P. Millot, Rept. CEA No. 1072, Saclay, 1958.

Case has proved that the functions $\phi_{0\pm}$ and ϕ_ν $(-1 \leqslant \nu \leqslant 1)$ are complete for all functions $\psi(\mu)$ of physical interest defined in the interval $(-1, 1)$.

III. EIGENFUNCTION EXPANSION FOR A SLAB

Solutions of the Boltzmann equation (1) may be now written in the form

$$\psi(x,\mu)=a_{0+}\phi_{0+}(\mu)e^{-x/\nu_0}+a_{0-}\phi_{0-}(\mu)e^{x/\nu_0}+\int_{-1}^{1}a(\nu)\phi_\nu(\mu)e^{-x/\nu}d\nu, \quad (6)$$

where the constants a_{0+}, a_{0-}, $a(\nu)$ should be determined by suitable boundary conditions.

Let us consider a problem of the critical thickness of a slab, which is infinite in all directions perpendicular to the x axis. Let the center of the coordinate system be placed in the middle of the slab. Let us denote by d the half-thickness of the slab. The boundary conditions are as follows:

$$\psi(-d,\mu)=0 \quad \text{for} \quad \mu>0, \quad (7a)$$

$$\psi(d,\mu)=0 \quad \text{for} \quad \mu<0, \quad (7b)$$

where μ is a cosine of an angle between the direction of the neutron velocity and the x axis. Inserting the general form (6) of the solution into Eqs. (7a) and (7b) gives two integral equations for the expansion coefficients. This formulation of a problem is inconvenient because we have two integral equations for $a(\nu)$ in two separate intervals of μ. Let us take into account symmetry properties of the neutron distribution in a slab:

$$\psi(x,\mu)=\psi(-x,-\mu). \quad (8)$$

This fact enables us to take only a symmetric part of the general expansion (6) as a general solution for a slab. It has the form:

$$\psi_s(x,\mu)=a_{0+}[\phi_{0+}(\mu)e^{-x/\nu_0}+\phi_{0+}(-\mu)e^{x/\nu_0}]+a_{0-}[\phi_{0-}(\mu)e^{x/\nu_0}+\phi_{0-}(-\mu)e^{-x/\nu_0}]+\int_{-1}^{1}a(\nu)[\phi_\nu(\mu)e^{-x/\nu}+\phi_\nu(-\mu)e^{x/\nu}]d\nu. \quad (9)$$

Owing to the shape of the discrete eigenfunctions $\phi_{0+}(\mu)$ and $\phi_{0-}(\mu)$, one can easily prove that the following identity takes place:

$$\phi_{0+}(\mu)e^{-x/\nu_0}+\phi_{0+}(-\mu)e^{x/\nu_0} \equiv \phi_{0-}(\mu)e^{x/\nu_0}+\phi_{0-}(-\mu)e^{-x/\nu_0} \equiv \frac{c}{2}\left[\frac{\nu_0}{\nu_0-\mu}e^{-x/\nu_0}+\frac{\nu_0}{\nu_0+\mu}e^{x/\nu_0}\right]. \quad (10)$$

So the general form of a solution of the Boltzmann equation (1) with a symmetry (8) has the form:

$$\psi_s(x,\mu)=(a_{0+}+a_{0-})[\phi_{0+}(\mu)e^{-x/\nu_0}+\phi_{0+}(-\mu)e^{x/\nu_0}]+\int_{-1}^{1}a(\nu)[\phi_\nu(\mu)e^{-x/\nu}+\phi_\nu(-\mu)e^{x/\nu}]d\nu, \quad (11)$$

which can be further simplified. The integral part of the expansion (11) may be written as follows:

$$\begin{aligned}
&\int_{-1}^{1}a(\nu)[\phi_\nu(\mu)e^{-x/\nu}+\phi_\nu(-\mu)e^{x/\nu}]d\nu \\
&=\int_{-1}^{0}a(\nu)[\phi_\nu(\mu)e^{-x/\nu}+\phi_\nu(-\mu)e^{x/\nu}]d\nu \\
&\quad+\int_{0}^{1}a(\nu)[\phi_\nu(\mu)e^{-x/\nu}+\phi_\nu(-\mu)e^{x/\nu}]d\nu \\
&=-\int_{1}^{0}a(-\nu)[\phi_{-\nu}(\mu)e^{x/\nu}+\phi_{-\nu}(-\mu)e^{-x/\nu}]d\nu \\
&\quad+\int_{0}^{1}a(\nu)[\phi_\nu(\mu)e^{-x/\nu}+\phi_\nu(-\mu)e^{x/\nu}]d\nu \\
&=\int_{0}^{1}[a(\nu)+a(-\nu)][\phi_\nu(\mu)e^{-x/\nu}+\phi_\nu(-\mu)e^{x/\nu}]d\nu,
\end{aligned}$$

where in one integral the change of integration variable $\nu \to -\nu$ has been performed and the following property of $\phi_\nu(\mu)$ has been used:

$$\phi_{-\nu}(\mu)\equiv\phi_\nu(-\mu).$$

The ultimate form of the solution of the Boltzmann equation (1) with the above mentioned symmetry is

$$\psi_s(x,\mu)=a_0[\phi_{0+}(\mu)e^{-x/\nu_0}+\phi_{0+}(-\mu)e^{x/\nu_0}]-a_0\int_{0}^{1}\tilde{a}(\nu)[\phi_\nu(\mu)e^{-x/\nu}+\phi_\nu(-\mu)e^{x/\nu}]d\nu, \quad (12)$$

where a_0 is a new constant and $\tilde{a}(\nu)$ is a new expansion function. This form of the expansion has a very useful property. Boundary conditions (7a) and (7b) now have an identical form:

$$\int_{0}^{1}\tilde{a}(\nu)[\phi_\nu(\mu)e^{d/\nu}+\phi_\nu(-\mu)e^{-d/\nu}]d\nu = \phi_{0+}(\mu)e^{d/\nu_0}+\phi_{0+}(-\mu)e^{-d/\nu_0} \quad \text{for} \quad 0\leqslant\mu\leqslant 1 \quad (13)$$

(we can divide without any loss of generality both sides of the boundary condition by a_0).

On inserting the functions $\phi_\nu(\mu)$ into Eq. (13), we obtain

$$\lambda(\mu)\tilde{a}(\mu)e^{d/\mu} + (c/2)P\int_0^1 \nu\tilde{a}(\nu)[(e^{d/\nu}/\nu-\mu)+(e^{-d/\nu}/\nu+\mu)]d\nu = \phi_{0+}(\mu)e^{d/\nu_0}+\phi_{0+}(-\mu)e^{-d/\nu_0}. \quad (14)$$

Let us use a new expansion function

$$b(\nu)=\tilde{a}(\nu)e^{d/\nu}. \quad (15)$$

Equation (14) should be written now in the form

$$\lambda(\mu)b(\mu)+(c/2)P\int_0^1 \nu b(\nu)\left[\frac{1}{\nu-\mu}+\frac{e^{-2d/\nu}}{\nu+\mu}\right]d\nu = \phi_{0+}(\mu)e^{d/\nu_0}+\phi_{0+}(-\mu)e^{-d/\nu_0}, \quad (16)$$

and the solution of the Boltzmann equation (1) for a slab is

$$\psi_s(x,\mu)=a_0[\phi_{0+}(\mu)e^{-x/\nu_0}+\phi_{0+}(-\mu)e^{x/\nu_0}] - a_0\int_0^1 b(\nu)[\phi_\nu(\mu)e^{-(x+d)/\nu}+\phi_\nu(-\mu)e^{x-d/\nu}]d\nu. \quad (17)$$

IV. REDUCTION OF THE INTEGRAL EQUATION (16) TO THE FREDHOLM TYPE

Equation (16) is a singular integral equation. An elegant theory of these equations is given in Muskhelishvili's monography.[4] Using his concepts, we decompose our equation into two parts: a singular and a regular one. We treat the regular part as a nonhomogenity of an equation, which now takes the form of a dominant equation:

$$\lambda(\mu)b(\mu)+(c/2)P\int_0^1 \frac{\nu b(\nu)}{\nu-\mu}d\nu=g(\mu), \quad (18)$$

where

$$g(\mu)=\phi_{0+}(\mu)e^{d/\nu_0}+\phi_{0+}(-\mu)e^{-d/\nu_0} - (c/2)\int_0^1 \frac{\nu e^{-2d/\nu}}{\nu+\mu}b(\nu)d\nu. \quad (19)$$

The integrability of the second term in Eq. (17), which gives us the observable quantity, namely the angular neutron distribution, requires $b(\nu)$ function to satisfy the H^* condition, i.e., $b(\nu)$ should satisfy Hölder's condition $H(\gamma)$ on every closed part of (0.1) interval, and near any end c_k of this interval ($c_1=0$, $c_l=1$) it must be of the form

$$b(\nu)=b^*(\nu)/(\nu-c_k)^\alpha, \quad 0\leqslant\alpha<1, \quad (20)$$

where $b^*(\nu)$ belongs to the class $H(\gamma)$.

[4] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff, Ltd., Groningen, Holland, 1953).

On introducing the sectionally holomorfic function

$$N(z)\equiv(1/2\pi i)\int_0^1 \frac{c}{2}\frac{\nu b(\nu)}{\nu-z}d\nu, \quad (21)$$

the problem of solution of the dominant equation (18) can be reduced to the Hilbert problem.

According to Plemelj formulas

$$N^+(\mu)+N^-(\mu)=\frac{1}{\pi i}P\int_0^1 \frac{c}{2}\frac{\nu b(\nu)}{\nu-\mu}d\nu, \quad (22)$$

$$N^+(\mu)-N^-(\mu)=(c/2)\mu b(\mu), \quad (23)$$

(where N^+ and N^- denote the limits of N as z goes to the cut from above and from below, respectively) and the integral equation (18) takes the form:

$$\left[\lambda(\mu)+\frac{i\pi c\mu}{2}\right]N^+(\mu)-\left[\lambda(\mu)-\frac{i\pi c\mu}{2}\right]N^-(\mu)=\frac{c}{2}\mu g(\mu) \quad (24)$$

or

$$G(\mu)N^+(\mu)-N^-(\mu)=\frac{c}{2}\frac{\mu g(\mu)}{\lambda(\mu)-(i\pi c\mu/2)}, \quad (25)$$

where

$$G(\mu)=\frac{\lambda(\mu)+(i\pi c\mu/2)}{\lambda(\mu)-(i\pi c\mu/2)}. \quad (26)$$

Function $N(z)$ has from the definition (21) the following properties:

1. It is analytical in the complex plane with a cut from 0 to 1 along the real axis.
2. It vanishes in infinity as $1/z$.
3. It is bounded at the ends of (0,1) interval by

$$c_1/z^\alpha \quad \text{and} \quad [c_2/(1-z)^\alpha], \quad \alpha<1$$

as $z\to 0$ and $z\to 1$, respectively.

Now the problem is to find such a sectionally holomorfic function $N(z)$ which satisfies Eq. (25) with $G(\mu)$ defined by Eq. (26) and which has the properties listed above. This is a nonhomogeneous Hilbert problem in the case of arc.

Let $\ln G(t)$ be an arbitrary branch of this multivalued function, and let us define a function $\Gamma(z)$:

$$\Gamma(z)=\frac{1}{2\pi i}\int_0^1 \frac{\ln G(t)}{t-z}dt; \quad (27)$$

then the function $X_0^{-1}(z)$, where $X_0(z)$ is defined as

$$X_0(z)=e^{\Gamma(z)}, \quad (28)$$

satisfies the homogeneous part of Eq. (25) (to prove this one must use the Plemelj formulas). However, this function does not satisfy all conditions which were formulated above for function $N(z)$. It is easily seen

that condition 3 is violated. Let us examine the behavior of $X_0(z)$ near $z=c_k$:

$$e^{\Gamma(z)}=(z-c_k)^{\alpha_k+i\beta_k}\Omega(z), \tag{29}$$

where α_k and β_k are real constants given by

$$\alpha_k+i\beta_k=(-1)^k\frac{\ln G(c_k)}{2\pi i} \tag{30}$$

and $\Omega(z)$ is a bounded, nonvanishing function at the point c_k. In our case

$$(1/2\pi i)\ln G(\mu)=(1/\pi)\theta(\mu), \tag{31}$$

where

$$\begin{aligned}\theta(\mu)&=\arg[\lambda(\mu)+(i\pi c\mu/2)]\\&=\arg[1-c\mu\tanh^{-1}\mu+(i\pi c\mu/2)],\\ \theta(0)&=0;\qquad \theta(1)=\pi;\\ \alpha_1+i\beta_1&=0,\quad \alpha_2+i\beta_2=1.\end{aligned} \tag{32}$$

So, $X_0(z)$ can be written in the form

$$X_0(z)=(z-1)\Omega(z), \tag{33}$$

and it is easily seen that X_0^{-1} is not bounded at the end $z=1$. Now we may construct in a unique fashion a function $X(z)$, which has a proper behavior at $z=1$,

$$X(z)=(z-1)^{-1}X_0(z), \tag{34}$$

and which also satisfies a homogeneous part of Eq. (25). Using again Plemelj formulas, we obtain

$$X^+(\mu)=e^{\frac{1}{2}\ln G(\mu)}X'(\mu), \tag{35a}$$

$$X^-(\mu)=e^{-\frac{1}{2}\ln G(\mu)}X'(\mu), \tag{35b}$$

where

$$X'(\mu)=(z-1)^{-1}\exp\left[\frac{1}{2\pi i}P\int_0^1\frac{\ln G(\mu')}{\mu'-\mu}d\mu'\right]$$

and

$$G(\mu)=\frac{X^+(\mu)}{X^-(\mu)}. \tag{36}$$

Now we are able to solve the nonhomogeneous Eq. (25). On inserting (36) into (25), we obtain:

$$N^+(\mu)X^+(\mu)-N^-(\mu)X^-(\mu)=\frac{c}{2}\frac{\mu g(\mu)X^-(\mu)}{\lambda(\mu)-(i\pi c\mu/2)}. \tag{37}$$

At present the problem can be formulated as follows: we are looking for a function $M(z)=N(z)X(z)$, which should be sectionally holomorfic and should be a solution of the nonhomogeneous Hilbert problem (37). This solution follows immediately from the Plemelj formulas:

$$N(z)X(z)=\frac{1}{2\pi i}\int_0^1\frac{c}{2}\frac{\mu' g(\mu')X^-(\mu')d\mu'}{[\lambda(\mu')-(i\pi c\mu'/2)](\mu'-z)}+P_k(z), \tag{38}$$

where $P_k(z)$ is an arbitrary polynomial of degree k. So we may write

$$N(z)=\frac{1}{2\pi i X(z)}\int_0^1\frac{c}{2}\frac{\mu' g(\mu')X^-(\mu')d\mu'}{[\lambda(\mu')-(i\pi c\mu/2)](\mu'-z)}+\frac{P_k(z)}{X(z)}. \tag{39}$$

The last term is evidently a solution of the homogeneous Hilbert problem.

Now we must verify that the function of this form satisfies the condition at infinity. First of all, the polynomial $P_k(z)$ must vanish. Secondly, as the function $X(z)$ in the denominator causes the function $N(z)$ in the form (39) to tend to a constant in infinity, we must add an additional condition:

$$\int_0^1\frac{\mu' g(\mu')X^-(\mu')d\mu'}{\lambda(\mu')-(i\pi c\mu'/2)}=0. \tag{40}$$

Knowing already the $N(z)$ function, we can calculate the coefficients of our eigenfunction expansion $b(\mu)$ by means of formula (23). This calculation must take into account the Plemelj formulas for the integral in formula (39) and the relation (36). The ultimate formula for $b(\mu)$ function is

$$b(\mu)=\frac{\lambda(\mu)g(\mu)}{\lambda^2(\mu)+(\pi c\mu/2)^2}-\frac{1}{X^-(\mu)[\lambda(\mu)+(i\pi c\mu/2)]}\times\frac{c}{2}P\int_0^1\frac{\mu' g(\mu')X^-(\mu')d\mu'}{[\lambda(\mu')-(i\pi c\mu'/2)](\mu'-\mu)}, \tag{41}$$

and it is to be remembered that the condition (40) must be satisfied.

Now we replace in (40) and (41) $g(\mu)$ by the right side of the formula (19) and we obtain an integral equation for $b(\mu)$:

$$\begin{aligned}b(\mu)&=A(\mu)\phi_d(\mu)\\&\quad-\frac{1}{B(\mu)}\frac{c}{2}P\int_0^1\frac{\mu'\phi_d(\mu')X^-(\mu')d\mu'}{[\lambda(\mu')-(i\pi c\mu'/2)](\mu'-\mu)}\\&\quad+\frac{1}{B(\mu)}\frac{c}{2}\int_0^1\nu e^{-2d/\nu}b(\nu)\\&\quad\times\left[P\frac{c}{2}\int_0^1\frac{\mu' X^-(\mu')d\mu'}{[\lambda(\mu')-(i\pi c\mu'/2)](\mu'-\mu)(\nu+\mu')}-\frac{B(\mu)A(\mu)}{\nu+\mu}\right]d\nu,\end{aligned} \tag{42}$$

where

$$\phi_d(\mu)=\phi_{0+}(\mu)e^{d/\nu_0}+\phi_{0+}(-\mu)e^{-d/\nu_0},$$

$$A(\mu)=\frac{\lambda(\mu)}{\lambda^2(\mu)+(\pi c\mu/2)^2},$$

$$B(\mu)=X^-(\mu)[\lambda(\mu)+(i\pi c\mu/2)]$$
$$=X^+(\mu)[\lambda(\mu)-(i\pi c\mu/2)],$$

and an additional condition

$$\int_0^1 \frac{\mu'\phi_d(\mu')X^-(\mu')d\mu'}{\lambda(\mu')-(i\pi c\mu'/2)}=\frac{c}{2}\int_0^1 \frac{\mu'X^-(\mu')}{\lambda(\mu')-(i\pi c\mu'/2)}$$
$$\times\left[\int_0^1 \frac{\nu e^{-2d/\nu}}{\nu+\mu'}b(\nu)d\nu\right]d\mu'. \quad (43)$$

Function $X^-(\mu)$ can be evaluated according to Appendix B of Case's paper:

$$X^-(\mu)=\frac{e^{\Gamma^-(\mu)}}{\mu-1},$$

but

$$\Gamma^-(\mu)=\frac{1}{2\pi i}\left[P\int_0^1 \frac{\ln G(\mu')d\mu'}{\mu'-\mu}-\pi i\ln G(\mu)\right]$$
$$=\frac{1}{\pi}P\int_0^1 \frac{\theta(\mu')d\mu'}{\mu'-\mu}-\tfrac{1}{2}\ln G(\mu),$$

therefore,

$$X^-(\mu)=\frac{G^{-\frac{1}{2}}(\mu)}{\mu-1}\exp\frac{P}{\pi}\int_0^1 \frac{\theta(\mu')}{\mu'-\mu}d\mu'$$
$$=\left[\frac{\lambda(\mu)-(i\pi c\mu/2)}{\lambda(\mu)+(i\pi c\mu/2)}\right]^{\frac{1}{2}}\frac{1}{\mu-1}\exp\frac{P}{\pi}\int_0^1 \frac{\theta(\mu')d\mu'}{\mu'-\mu}. \quad (44)$$

V. CONCLUSION

The problem of the critical thickness of a slab has been reduced to the system of two equations (42) and (43) for the expansion coefficient $b(\nu)$ of the continuous eigenfunctions of the Boltzmann equation and for the critical thickness d of a slab. Both equations are coupled by means of $b(\nu)$ function. Treating Eq. (42) as an integral equation (which by means of simple transformations can be reduced to the Fredholm type with a regular kernel) determining $b(\nu)$ function, Eq. (43) can be considered an exact critical condition for a slab. The functions appearing in these two equations are rather complicated, especially the $X^-(\mu)$ function, which cannot be expressed by elementary functions and must be evaluated numerically and tabulated. The numerical calculations are being done now, and the methods of calculations and the results will be published soon.

Tschebyscheff Polynomial Approximation Method of the Neutron-Transport Equation

SHIN YABUSHITA
Department of Nuclear Engineering, Kyoto University, Yoshida, Kyoto, Japan
(Received November 9, 1960)

The Tschebyscheff polynomial approximation method of the neutron-transport equation is developed. The relations between the relaxation constants in this approximation and the positive roots of $T_{N+1}(\mu)=0$ are derived. Using these relations, the necessary condition for reactor criticality is discussed. Application to Milne's problem leads to an explicit expression for extrapolated end point which is formally the same in the spherical-harmonics method. Numerical comparison of this method with the spherical-harmonics method verifies Conkie's conclusion [W. R. Conkie, Nuclear Sci. and Eng. 6, 260 (1959)] that, for weak absorbers, the spherical-harmonics method gives the values for extrapolated end point which is closer to the exact value than the T_N method does while, for strong absorbers, the T_N method gives closer value than the P_L method does.

1. INTRODUCTION

THE purpose of this paper is to develop the Tschebyscheff polynomial approximation method of the neutron-transport equation. The spherical-harmonics method has been used widely,[1-3] but there is no *a priori* superiority of the P_L method to other expansion methods. Any set of orthogonal functions can be used and the criterion for the choice of a set of functions are the rapidity of convergence and the simplicity of treatment. The P_L method surely satisfies the second requirement, but so long as the rapidity of convergence is concerned, there is some evidence that the P_L method is not a good approximation. For instance, the neutron angular distribution in Milne's problem obtained by the P_L method approaches the true distribution, at best, very slowly. The Tschebyscheff polynomial method was first proposed by Aspelund[4] and by Conkie.[5] The first author, however, considered only the T_1 and T_3 approximations and applied the method to Milne's problem with Marshak's boundary condition.[6] Conkie tried to obtain the general expression of the method and has given the tables for the reciprocals of relaxation constants in the T_5, T_7, and T_9 approximations and the values of the extrapolated end point for Milne's problem in the T_1, T_3, and T_5 approximations for $\Sigma_s/\Sigma=0.1$, 0.5, and 1.0, where Σ_s and Σ are scattering and total cross sections, respectively. However, Conkie did not give explicit expression for the extrapolated end point, and his treatment is not suitable to multilayer problem. Now, as will be seen in the following, we can develop the T_N method almost as completely as the P_L method. However, the problem of how to apply the T_N method to spherical or cylindrical geometry has not been solved, and the inapplicability to such geometry may remain an unremovable fault of the T_N method. In Sec. 3, mathematical properties of relaxation constants are discussed; and in Sec. 4, the necessary condition for the existence of a nonzero solution to the homogeneous Boltzman equation is discussed.

2. T_N APPROXIMATION METHOD

We consider the one-dimensional Boltzmann equation

$$\mu\partial f(x,\mu)/\partial x+\Sigma f(x,\mu)=\tfrac{1}{2}\Sigma_s\int_{-1}^{1}f(x,\mu)d\mu \quad (2.1)$$

where the symbols are: μ cosine of angle with positive x axis; x coordinate; $f(x,\mu)$ flux density of neutrons at position x traveling in direction μ; Σ and Σ_s are total and scattering cross sections, respectively. In the monoenergetic case, the absorptive and multiplicative systems are characterized by $\Sigma>\Sigma_s$ and by $\Sigma<\Sigma_s$, respectively. We expand $f(x,\mu)$ into the Tschebyscheff polynomial[7] $T_n(\mu)$ according to Aspelund

$$f(x,\mu)=\pi^{-1}\psi_0(x)T_0(\mu)+2\pi^{-1}\sum_{n=1}^{N}\psi_n(x)T_n(\mu). \quad (2.2)$$

The orthogonal relations between $T_n(\mu)$ are

$$\int_{-1}^{1}T_n(\mu)T_m(\mu)(1-\mu^2)^{-\frac{1}{2}}d\mu=\begin{cases}\pi & m=n=0\\ \frac{1}{2}\pi & m=n\neq 0.\\ 0 & m\neq n\end{cases}$$

From (2.2), (2.1), and the above orthogonality relations, we get the set of equations

$$\Sigma\psi_n(x)+\tfrac{1}{2}\psi_{n-1}'(x)+\tfrac{1}{2}\psi_{n+1}'(x)=0, \quad (2.3a)$$

$$\Sigma\psi_0(x)+\psi_1'(x)=\Sigma_s\psi_0(x)-2\Sigma_s\sum_{n=1}\frac{\psi_{2n}(x)}{4n^2-1}. \quad (2.3b)$$

If we put

$$\psi_n(x)=g_n(\lambda)e^{-\Sigma x/\lambda}, \quad (2.4)$$

[1] B. Davison, *Neutron Transport Theory* (Oxford University Press, New York, 1957).
[2] A. M. Weinberg and E. P. Wigner, *The Physical Theory of Neutron Chain Reactors* (The University of Chicago Press, Chicago, Illinois, 1958).
[3] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, New York, 1950).
[4] O. Aspelund, PICG 16, 530 (1958).
[5] W. R. Conkie, Nuclear Sci. and Eng. 6, 260 (1959).
[6] See p. 533 of reference 4.
[7] J. Arving and N. Mullinieux, *Mathematics in Physics and Engineering* (Academic Press, Inc., New York, 1959).

we get a system of simultaneous equations

$$(\Sigma-\Sigma_s)g_0(\lambda)-\frac{\Sigma}{\lambda}g_1(\lambda)=-2\Sigma_s\sum_{n=1}\frac{g_{2n}(\lambda)}{4n^2-1}, \quad (2.5a)$$

$$2\lambda g_n(\lambda)=g_{n-1}(\lambda)+g_{n+1}(\lambda), \quad n\geq 1. \quad (2.5b)$$

Aspelund and Conkie did not give an explicit expression for $g(\lambda)$. Conkie assumes a solution of the form

$$f(x,\mu)=\sum_k a_k e^{\Sigma x/\lambda_k}\varphi_k(\mu),$$

and derives the equation for $\varphi_k(\mu)$

$$(\mu/\lambda_k+1)\varphi_k(\mu)-\tfrac{1}{2}(\Sigma_s/\Sigma)\int_{-1}^{1}\varphi_k(\mu)d\mu=\tau_k W_{N+1}(\mu).$$

If $\varphi_k(\mu)$ is an exact solution, the right-hand side of the above equation vanishes. In the T_N approximation, Conkie puts $W_{N+1}(\mu)=T_{N+1}(\mu)$ and obtains

$$\varphi_k(\mu)=\frac{\Sigma_s\cdot\lambda_k}{2\Sigma T_{N+1}(\lambda_k)}\cdot\left\{\frac{T_{N+1}(\mu)-T_{N+1}(\lambda_k)}{\mu-\lambda_k}\right\},$$

which satisfies the normalization condition

$$N_k=\int_{-1}^{1}\varphi_k(\mu)d\mu=1.$$

It is possible to determine $g_n(\lambda)$ only when the determinant, of which the coefficients of $g(\lambda)$ in (2.5) are the elements, vanishes. Aspelund took this method. However, it is possible to determine $g(\lambda)$ in another way that is essentially the same as the P_L approximation.[8] There are two sets of polynomials that satisfy the following recurrence relation

$$T_{n+1}(\lambda)=2\lambda T_n(\lambda)-T_{n-1}(\lambda), \quad (2.6a)$$

$$U_{n+1}(\lambda)=2\lambda U_n(\lambda)-U_{n-1}(\lambda), \quad (2.6b)$$

where the $T(\lambda)$ and $U(\lambda)$ are the Tschebyscheff polynomials of the first and second kind, respectively. $T_n(\lambda)$ and $U_n(\lambda)$ are defined by

$$T_n(\lambda)=\cos(n \text{ arc } \cos\lambda),$$
$$U_n(\lambda)=\sin(n \text{ arc } \cos\lambda)/(1-\lambda^2)^{\frac{1}{2}}.$$

The explicit form of the first few T's and U's are

$$T_0=1, \quad T_1=\lambda, \quad T_2=2\lambda^2-1,$$
$$U_0=0, \quad U_1=1, \quad U_2=2\lambda.$$

$T_n(\lambda)$ is a polynomial in λ of order n, and $U_n(\lambda)$ is a polynomial in λ of order $n-1$. It is possible to put

$$g_n(\lambda)=AT_n(\lambda)-BU_n(\lambda), \quad n\geq 0. \quad (2.7)$$

The last equation of (2.5b) corresponding to $n=N$ is

$$g_{N-1}(\lambda)-2\lambda g_N(\lambda)=0, \quad (2.8)$$

which we can write

$$AT_{N+1}(\lambda)-BU_{N+1}(\lambda)=0. \quad (2.9)$$

Equation (2.9) is an equation which determines the permissible value of λ when A and B are properly determined. The solution of (2.3) corresponding to a particular λ is

$$\psi_n(x)=g_n(\lambda_k)e^{-\Sigma x/\lambda_k}.$$

The general solution of (2.1) in the T_N approximation is then

$$f(x,\mu)=\pi^{-1}\sum_k \alpha_k g_0(\lambda_k)e^{-\Sigma x/\lambda_k}$$
$$+2\pi^{-1}\sum_k\sum_{n=1}^{N}\alpha_k g_n(\lambda_k)e^{-\Sigma x/\lambda_k}, \quad (2.10)$$

where α_k are arbitrary constants and λ's are the roots of Eq. (2.9). Now, we can show the following summation formulas:

$$1+2\sum_{n=1}^{N}T_n(\mu)T_n(\lambda)=\frac{T_{N+1}(\mu)T_N(\lambda)-T_N(\mu)T_{N+1}(\lambda)}{\mu-\lambda},$$
$$2\sum_{n=1}^{N}T_n(\mu)U_n(\lambda)=\frac{T_{N+1}(\mu)U_N(\lambda)+1-T_N(\lambda)U_{N+1}(\lambda)}{\mu-\lambda}. \quad (2.11)$$

When we use these summation formulas, we get

$$f(x,\mu)=\sum_k \alpha_k e^{-\Sigma x/\lambda_k}\varphi_k(\mu), \quad (2.12)$$

where

$$\varphi_k(\mu)=\frac{1}{\pi}\left\{\frac{AT_{N+1}(\lambda_k)-BU_{N+1}(\lambda_k)}{\mu-\lambda_k}-\frac{B}{\mu-\lambda_k}\right\}. \quad (2.13)$$

From (2.5), it is possible to determine only the ratio A/B, and it is possible to take $A=1$. Here, we take the normalization condition[9] rather than $A=1$.

$$N_k=\int_{-1}^{1}\varphi_k(\mu)d\mu=1. \quad (2.14)$$

It is easily found from (2.10) and (2.5a) that

$$N_k=2B\Sigma/\pi\lambda_k\Sigma_s,$$

from which we get

$$B_k=\pi\lambda_k\Sigma_s/2\Sigma,$$

where, in deriving N_k, we used the relation

$$U_L T_M-T_L U_M=U_{L-M}, \quad L\geq M\geq 0. \quad (2.15)$$

[8] W. Kofink, ORNL-2334 (1957).

[9] The referee pointed out that our solution is equivalent to Conkie's solution when we take (2.14) rather than $A=1$. Equation (2.14) leads to a simple expression for B, while A is more complicated.

Thus, we get

$$\varphi_k(\mu)=\frac{\Sigma_s\lambda_k}{2\Sigma T_{N+1}(\lambda_k)}\frac{T_{N+1}(\mu)-T_{N+1}(\lambda_k)}{\mu-\lambda_k}, \quad (2.16)$$

which is the same as Conkie's solution.[10]

The equation which determines the permissible value of λ is obtained from Eqs. (2.9) and (2.5a). Thus, we get[11]

$$T_{N+1}(\lambda)+(\Sigma_s/\Sigma)\lambda L_{N+1}(\lambda)=0, \quad (2.17)$$

where

$$L_{N+1}(\lambda)=2\sum_{n=1}\left\{\frac{U_{N+1-2n}}{4n^2-1}\right\}-U_{N+1}(\lambda). \quad (2.18)$$

Now, from Eq. (2.17) it is clear that if N is even, one of the λ's is equal to zero. But infinitely short relaxation constants are inadmissible, so we always choose N to be an odd number. If N is odd, the left-hand side of (2.17) is an even polynomial in λ of degree $(N+1)$, therefore, if λ is a root of (2.17), then $-\lambda$ is also a root of (2.17). But if $\Sigma=\Sigma_s$, then (2.17) reduces to the polynomial of degree $(N-1)$. In this case, the preceding discussions hold for λ that are not infinite. It is sufficient to obtain another solution. We start from (2.3) and easily obtain

$$\psi_0(x)=Cx+D, \quad \psi_1(x)=-C/2\Sigma,$$
$$\psi_n(x)=0, \quad n\geq 2.$$

Thus, we obtain the general solution

$$f(x,\mu)=(1/\pi)(Cx+D)-(C/\pi\Sigma)T_1(\mu)+\sum_{k=1}^{\frac{1}{2}(N-1)}\alpha_k e^{-\Sigma x/\lambda_k}\varphi_k(\mu)+\sum_{k=1}^{\frac{1}{2}(N-1)}\alpha_{-k}e^{\Sigma x/\lambda_k}\varphi_{-k}(\mu). \quad (2.19)$$

Equation (2.19) is equivalent to Conkie's solution[12] if we put $C=\pi$.

3. PROPERTIES OF RELAXATION CONSTANTS

In this section we consider the mathematical properties of relaxation constants. The results of this section will be used in the later discussion on the necessary condition for reactor criticality.

In order to see the properties of relaxation constants, we consider L defined by Eq. (2.18). We denote the positive roots of $T_{N+1}(\mu)=0$ by $\mu_1, \mu_2, \cdots, \mu_{\frac{1}{2}(N+1)}$ in the order of magnitude, that is, $\mu_1>\mu_2>\cdots\mu_{\frac{1}{2}(N+1)}>0$. From the definition of $T_{N+1}(\mu)$,

$$\mu_l=\cos\frac{\pi(2l-1)}{2(N+1)}, \quad l=1, 2, \cdots, \tfrac{1}{2}(N+1).$$

Now it is clear that

$$\sin(N+1)\arccos\mu_l=(-1)^{l+1}.$$

Moreover, for $-1<\lambda<+1$, we have

$$(1-\lambda^2)^{\frac{1}{2}}\left|2\sum_{n=1}^{\frac{1}{2}(N-1)}\frac{U_{N+1-2n}(\lambda)}{4n^2-1}\right|\leqq\frac{2}{3}+\frac{2}{15}+\frac{2}{35}+\cdots\frac{2}{N^2-2N},$$
$$=\tfrac{2}{3}+\tfrac{1}{3}-\tfrac{1}{5}+\tfrac{1}{5}\cdots\frac{1}{N-2}-\frac{1}{N},$$
$$=1-(1/N)<1.$$

Therefore, so long as N is a finite number, we can conclude that

$$\text{sign of } L_{N+1}(\mu_l)=\text{sign of } (-1)^l. \quad (3.1)$$

Now we consider a function of λ defined by

$$F(\lambda)=T_{N+1}(\lambda)+(\Sigma_s/\Sigma)\lambda L_{N+1}(\lambda).$$

$F(\lambda)$ is an even function of λ, so we consider $F(\lambda)$ for positive λ only. From (3.1), we can show

$$\text{sign of } F(\mu_l)=\text{sign of } (-1)^l.$$

Therefore, $F(\lambda)$ will cross the λ axis in the interval $\mu_{l-1}<\lambda<\mu_l$. Therefore, we get the following inequality:

$$1>\mu_1>\lambda_1>\mu_2>\lambda_2\cdots\lambda_{\frac{1}{2}(N-1)}>\mu_{\frac{1}{2}(N+1)}>0. \quad (3.2)$$

The property of the remaining relaxation constants depends on whether $\Sigma>\Sigma_s$ or $\Sigma<\Sigma_s$. The coefficients of λ^N in $L_{N+1}(\lambda)$ is equal to the negative value of the coefficient of λ^{N+1} in $T_{N+1}(\lambda)$. Therefore, the function defined above will cross the λ axis for λ greater than μ_1 if $\Sigma_s<\Sigma$. Moreover, we can write

$$F(\lambda)=2^N\left[\left(1-\frac{\Sigma_s}{\Sigma}\right)\lambda^2-\frac{1}{2^N\lambda_1^2\cdot\lambda_2^2\cdots\lambda_{\frac{1}{2}(N-1)}^2}\right]\times(\lambda-\lambda_1)\cdots(\lambda+\lambda_{\frac{1}{2}(N-1)}).$$

Therefore, if $\Sigma<\Sigma_s$, the remaining roots are λ_0 and $-\lambda_0$, where λ_0 is a purely imaginary number.

$$\lambda_0=\frac{i\Sigma^{\frac{1}{2}}}{(\Sigma_s-\Sigma)^{\frac{1}{2}}\lambda_1\cdot\lambda_2\cdots\lambda_{\frac{1}{2}(N-1)}\cdot 2^{\frac{1}{2}N}}.$$

Thus, we get the following inequality connecting the zeros of $T_{N+1}(\mu)$ and the relaxation constants:

$$\Sigma>\Sigma_s, \quad \lambda_0>\mu_1>\lambda_1\cdots, \quad \lambda_{\frac{1}{2}(N-1)}>\mu_{\frac{1}{2}(N+1)}, \quad (3.3)$$

$$\Sigma<\Sigma_s, \quad \mu_1>\lambda_1>\mu_2\cdots, \quad \lambda_{\frac{1}{2}(N-1)}>\mu_{\frac{1}{2}(N+1)}, \quad (3.4)$$

and the remaining roots are a pair of imaginary number.

$$\Sigma=\Sigma_s, \quad \mu_1>\lambda_1>\mu_2\cdots, \quad \lambda_{\frac{1}{2}(N-1)}>\mu_{\frac{1}{2}(N+1)}, \quad (3.5)$$

and the remaining roots are infinitely large. The roots of $T_{N+1}(\mu)=0$ are all in the interval $-1<\mu<+1$. Therefore, the distribution of λ will be dense as the

[10] See Eq. (14) of reference 5.

[11] Equation (2.17) is equivalent to Eq. (15) of reference 5. However, our Eq. (2.17) is more transparent than Conkie's treatment.

[12] See Eq. (16) of reference 5.

order of approximation increases. From the above discussion, it is clear that there are two distinct classes among the relaxation constants. λ_0 belongs to one class and $\lambda_1, \lambda_2, \cdots \lambda_{\frac{1}{2}(N-1)}$ belong to the other. As N increases, $\lambda_1, \cdots \lambda_{\frac{1}{2}(N-1)}$ will be uniformly distributed in the interval $0<\lambda<1$; and in the limit $N\to\infty$, these λ's will form a continuous spectrum. On the other hand, λ_0 forms a discrete spectrum. Wigner[13] has shown that the rigorous solution of (2.1) can be expressed as a superposition of functions belonging to continuous λ plus solutions which belong to discrete λ. Our results seem to be consistent with Wigner's results.

4. ON THE CRITICALITY CONDITION IN THE T APPROXIMATION

The criticality condition in transport theory is the condition for the existence of a nonzero solution of a homogeneous equation that satisfies the given boundary condition. In the criticality problem, the boundary condition is that there is no inward current at the surface where the system is surrounded by a vacuum. In the T_N approximation, it is impossible to satisfy the given boundary conditions completely. Therefore, there exist, as in the P_L approximation, some arbitrariness in the choice of boundary conditions. Here we use the modified Mark's condition[14] for the sake of analytical simplicity. In the following, it will be shown that the absorptive system cannot be critical. The physical meaning of this fact is clear. But so long as the transport equation describes the physical phenomenon correctly, the theory can show this fact only by the transport equation and by mathematics.

We consider a system which is surrounded by vacuum at $x=a$ and is symmetric about the plane $x=0$. Then the boundary conditions are

$$f(0,\mu)=f(0,-\mu), \tag{4.1}$$

$$f(a,\mu)=0, \quad \mu=-\mu_1, -\mu_2, \cdots, -\mu_{\frac{1}{2}(N+1)}, \tag{4.2}$$

where $\mu_1, \mu_2, \mu_{\frac{1}{2}(N+1)}$ are, as before, the positive roots of $T_{N+1}(\mu)=0$. With (2.12), the condition (4.1) gives

$$\begin{aligned}\pi f(a,\mu) &= \sum_{k=0}^{\frac{1}{2}(N-1)} \alpha_k e^{-\Sigma x/\lambda_k} \frac{AT_N(\lambda_k)-BU_N(\lambda_k)}{\mu-\lambda_k}\cdot T_{N+1}(\mu) \\ &\quad - \sum_{k=0}^{\frac{1}{2}(N-1)} \alpha_k e^{-\Sigma x/\lambda_k}\cdot\frac{B(\lambda_k)}{\mu-\lambda_k} \\ &\quad - \sum_{k=0}^{\frac{1}{2}(N-1)} \alpha_k e^{\Sigma x/\lambda_k}\cdot\frac{AT_N(\lambda_k)-BU_N(\lambda_k)}{\mu+\lambda_k}\cdot T_{N+1}(\mu) \\ &\quad + \sum_{k=0}^{\frac{1}{2}(N-1)} \alpha_k e^{\Sigma x/\lambda_k}\cdot\frac{B(\lambda_k)}{\mu+\lambda_k}.\end{aligned} \tag{4.3}$$

Clearly, the first and the third parts of the above equation vanish for $\mu=-\mu_1, \cdots, -\mu_{\frac{1}{2}(N+1)}$. Therefore, when the second and fourth parts of the above equation are brought to common denominator, the numerator must vanish for $\mu=-\mu_l$. Here, we define $r(\mu)$ and $p(\mu)$ according to Weinberg and Wigner.[15]

$$\begin{aligned}p(\mu)&=(\mu-\mu_1)(\mu-\mu_2)\cdots(\mu-\mu_{\frac{1}{2}(N+1)}),\\ r(\mu)&=(\mu-\lambda_0)(\mu-\lambda_1)\cdots(\mu-\lambda_{\frac{1}{2}(N-1)}).\end{aligned} \tag{4.4}$$

Then we can write

$$T_{N+1}(\mu)=2^N p(\mu)p(-\mu)(-1)^{\frac{1}{2}(N+1)}. \tag{4.5}$$

Then we can write (4.3) in the form

$$\pi f(a,\mu)=\sum_{k=0}^{\frac{1}{2}(N-1)}\cdot\frac{\gamma_k T_{N+1}(\mu)}{\mu-\lambda_k}+\sum_{k=0}^{\frac{1}{2}(N-1)}\frac{\delta_k T_{N+1}(\mu)}{\mu+\lambda_k} - \frac{p(-\mu)Q(\mu)}{r(\mu)r(-\mu)}, \tag{4.6}$$

where

$$\begin{aligned}\gamma_k&=\alpha_k\{AT_N(\lambda_k)-BU_N(\lambda_k)\}e^{-\Sigma a/\lambda_k},\\ \delta_k&=-\alpha_k\{AT_N(\lambda_k)-BU_N(\lambda_k)\}e^{\Sigma a/\lambda_k},\end{aligned} \tag{4.7}$$

and $Q(\mu)$ is a polynomial in μ of degree $\frac{1}{2}(N-1)$. $Q(\mu)$ is a polynomial in μ of degree $\frac{1}{2}(N-1)$, and we have $\frac{1}{2}(N+1)$ constants to be determined. However, the condition (4.2) is homogeneous, therefore, we chose $Q(\lambda_0)$ to be unity. In this case, we have to determine $\frac{1}{2}(N-1)$ constants. The constants γ_k and δ_k are determined by the requirements that $f(a,\mu)$ have no poles at $\mu=\lambda_k$ and at $\mu=-\lambda_k$. The requirements that the poles at λ_k cancel is

$$\gamma_k T_{N+1}(\lambda_k)=\frac{p(-\lambda_k)Q(\lambda_k)}{r'(\lambda_k)r(-\lambda_k)}. \tag{4.8a}$$

Similarly, the condition at $\mu=-\lambda_k$ is

$$\delta_k T_{N+1}(-\lambda_k)=-\frac{p(\lambda_k)Q(-\lambda_k)}{r'(\lambda_k)r(\lambda_k)}, \tag{4.8b}$$

where

$$r'(\lambda_k)=(\lambda_k-\lambda_0)(\lambda_k-\lambda_1)\cdots(\lambda_k-\lambda_{\frac{1}{2}(N-1)}),$$

in which the factor $\lambda_k-\lambda_k$ is omitted. If we divide (4.8a) by (4.8b), and remembering the identity

$$T_{N+1}(\lambda_k)=T_{N+1}(-\lambda_k),$$

we get

$$e^{-2\Sigma a/\lambda_k}=\frac{p(-\lambda_k)Q(\lambda_k)}{p(\lambda_k)Q(-\lambda_k)}, \quad k=0, 1, \cdots, \tfrac{1}{2}(N-1). \tag{4.9}$$

This is a set of equations that determine the coefficients of $Q(\mu)$. But, as has been said above, we have to determine $\frac{1}{2}(N-1)$ constants. The condition (4.9) gives $\frac{1}{2}(N+1)$ conditions; therefore, it is possible to

[13] E. P. Wigner's speech at the meeting held in New York for the American Mathematical Society, April 23–24, 1959.

[14] For Mark's condition in the P_L approximation, see p. 130 of reference 1.

[15] See p. 258 of reference 2.

determine the coefficients of $Q(\mu)$ only when a is equal to some particular value. From (4.9), we get

$$\frac{Q(\lambda_k)}{Q(-\lambda_k)}=(-1)^{\frac{1}{2}(N+1)} \times\frac{(\lambda_k-\mu_1)(\lambda_k-\mu_2)\cdots[\lambda_k-\mu_{\frac{1}{2}(N-1)}]}{(\lambda_k+\mu_1)(\lambda_k+\mu_2)\cdots[\lambda_k+\mu_{\frac{1}{2}(N+1)}]}e^{-2\Sigma a/\lambda_k}. \tag{4.10}$$

First, we consider the case $\Sigma>\Sigma_s$. In this case, as was pointed out in Sec. 3, all the relaxation constants are real and satisfy the inequality (3.3). From the inequality and from (4.10), $Q(\mu)$ must change sign $\frac{1}{2}(N+1)$ times in the interval $-\lambda_0<\mu<\lambda_0$. But, as has been said, $Q(\mu)$ is a polynomial of degree $\frac{1}{2}(N-1)$ and, therefore, $Q(\mu)$ can change its sign at most $\frac{1}{2}(N-1)$ times. Therefore, in this case, we cannot find a positive number a that is consistent with (4.10). The above discussions hold for arbitrary N so long as N is an odd number. This is the mathematical expression, at least in the T_N approximation, of the physical fact that the absorptive system cannot be critical. But, if the system is multiplicative, or $\Sigma<\Sigma_s$, the conflictions mentioned above do not occur. For instance, in the T_1 approximation, (4.10) reduces to

$$1=\frac{Q(\lambda_0)}{Q(-\lambda_0)}=-\frac{(\lambda_0-\mu_1)}{(\lambda_0+\mu_1)}e^{-2\Sigma a/\lambda_0},$$

where

$$\lambda_0=i\left(\frac{\Sigma}{2(\Sigma_s-\Sigma)}\right)^{\frac{1}{2}},\quad \mu_1=\frac{1}{\sqrt{2}}.$$

In the T_1 approximation, we can surely obtain a positive number a which is consistent with the above equation and, therefore, the condition is the necessary and sufficient condition for the existence of a nonzero solution. However, for arbitrary N, it is difficult to show sufficient condition for the existence of a nonzero solution.

The importance of the above discussion is that, at present, there is no practical criterion to distinguish a multiplicative system from an absorptive system in multigroup transport theory. Davison[16] says that, if at least a pair of relaxation constants are purely imaginary, the system is multiplicative. However, he does not give any explanation mathematically. It is true that if a pair of relaxation constants are purely imaginary, there can be a periodic solution. But if we wish to treat the problem rigorously, we must consider the necessary and sufficient condition for the existence of a nonzero solution of the homogeneous multigroup Boltzmann equation. Thus our approach might show the way of approach to the problem.

[16] See p. 252 of reference 1.

5. APPLICATION TO MILNE'S PROBLEM

Milne's problem[17] is a one-dimensional two region problem. A half-plane $x<0$ bounded by the plane $x=0$ is filled by a noncapturing medium, which scatters neutrons isotropically without changing their velocity. No source is present in a finite region, and only the outward current is applicable at the surface $x=0$. There is no inward current at the surface. To obtain the neutron flux in the medium, and the angular distribution at the surface, Milne's problem for linear anisotropic scattering in the P_L method has been studied by Kofink.[18] Milne's problem in the T_N method was considered by Conkie using Mark's[19] condition. However, Conkie did not give explicit expression for the angular distribution and extrapolated end point. As will be seen in the following, it is possible to solve the problem analytically in the T_N approximation.

Before we discuss Milne's problem, we discuss the more general Milne's problem in which the medium is not necessarily noncapturing. In this case, all the decaying modes have died out except the one with the smallest decaying constant. Therefore, for the medium $\Sigma>\Sigma_s$, the neutron distribution is given by (2.10) with $\alpha_k=0$; $k=1,\cdots,\frac{1}{2}(N-1)$.

$$\pi f(0,\mu)=\left\{\frac{\gamma_0}{\mu-\lambda_0}+\sum_{k=0}^{\frac{1}{2}(N-1)}\frac{\delta_k}{\mu+\lambda_k}\right\}T_{N+1}(\mu) -\frac{\alpha_0 B(\lambda_0)}{\mu-\lambda_0}+\sum_{k=0}^{\frac{1}{2}(N-1)}\frac{\beta_k B(\lambda)}{\mu_k+\lambda_k}, \tag{5.1}$$

where

$$\begin{aligned}\gamma_0&=\alpha_0\{A(\lambda_0)T_N(\lambda_0)-B(\lambda_0)U_N(\lambda_0)\},\\ \delta_k&=-\beta_k\{A(\lambda_k)T_N(\lambda_k)-B(\lambda_k)U_N(\lambda_k)\}.\end{aligned} \tag{5.2}$$

Now, when we apply Mark's condition

$$f(0,-\mu_l)=0,$$

the first part of the right-hand side of (5.1) is zero for $\mu=-\mu_1,\cdots-\mu_{\frac{1}{2}(N+1)}$. Therefore, when the second and third parts of the right-hand side of (5.1) are brought to common denominator, the numerator must vanish for $\mu=-\mu_1$. That is, we can write

$$\frac{\alpha_0 B(\lambda_0)}{\mu-\lambda_0}-\sum_{k=0}^{\frac{1}{2}(N-1)}\frac{\beta_k B(\lambda_k)}{\mu+\lambda_k}=\frac{c\phi(-\mu)}{(\mu-\lambda_0)r(-\mu)},$$

where c is a constant. The constants γ_0 and δ_k are determined by the requirement[20] that $f(0,\mu)$ have no

[17] G. Placzek and W. Seidel, Phys. Rev. **72**, 550 (1947).
[18] See reference 8.
[19] See reference 14.
[20] This method has been used in the P_L method. See, for instance, p. 258 of reference 2.

TABLE I.[a] Values of $d\Sigma_s$ in various order of approximation.

Σ_s/Σ	T_1	T_3	T_5	T_7	T_9
0.1	0.1355	0.2285*	0.2719*	0.3001	0.3212
0.2	0.2283	0.3841	0.4551	0.5005	0.5337
0.3	0.3068	0.5046	0.5868	0.6353	0.6676
0.4	0.3767	0.5969	0.6721	0.7074	0.7245
0.5	0.4407*	0.6635	0.7151*	0.7272	0.7273
0.6	0.5009	0.7057	0.7253	0.7213	0.7175
0.7	0.5559	0.7252	0.7191	0.7134	0.7121
0.8	0.6087	0.7256	0.7107	0.7094	0.7098
0.9	0.6590	0.7127	0.7052	0.7075	0.7086
1.0	0.707107	0.694190	0.702066	0.706452	0.707963
	P_1	P_3	P_5	P_7	Exact value
0.1	0.1107	0.2025	0.2470	0.2759	0.8539
0.2	0.1864	0.3398	0.4126	0.4594	0.7851
0.3	0.2505	0.4475	0.5342	0.5863	0.7491
0.4	0.3076	0.5324	0.6188	0.6637	0.7305
0.5	0.3598	0.5976	0.6708	0.6999	0.7207
0.6	0.4083	0.6448	0.6968	0.7103	0.7155
0.7	0.4539	0.6756	0.7059	0.7106	0.7127
0.8	0.4970	0.6921	0.7069	0.7091	0.7113
0.9	0.5381	0.6970	0.7056	0.7078	0.7106
1.0	0.577350	0.694025	0.703899	0.706920	0.710446

[a] P_L values are reproduced from J. C. Mark, National Research Council of Canada, Atomic Energy Project, Report MT 97, 1945. Exact values are taken from K. M. Case, F. de Hoffmann, and G. Placzek, *Introduction to the Theory of Neutron Diffusion* (Government Printing Office, Washington, D. C., 1953), where the exact value is calculated by Wiener-Hoph method.

poles at $\mu=\lambda_0$ and at $\mu=-\lambda_k$. The condition at $\mu=-\lambda_0$ gives

$$\gamma_0 T_{N+1}(\lambda_0)=\frac{cp(-\lambda_0)}{r(-\lambda_0)}.$$

By using a similar condition at $\mu=-\lambda_k$, we get

$$\delta_k=\frac{c}{\lambda_k+\lambda_0}\frac{p(\lambda_k)}{r'(\lambda_k)}-\frac{1}{T_{N+1}(\lambda_k)}, \tag{5.3}$$

$$c=2^N p(\lambda_0)r(-\lambda_0)\times[A(\lambda_0)T_N(\lambda_0)-B(\lambda_0)U_N(\lambda_0)](-1)^{\frac{1}{2}(N+1)}. \tag{5.4}$$

Next we consider the case of a noncapturing medium. In this case, we easily obtain the flux distribution in the same way as in the preceding discussion.

$$\pi f(0,\mu)=D-\frac{C}{\Sigma}T_1(\mu)+\sum_{k=1}^{\frac{1}{2}(N-1)}\beta_k\times\left\{\frac{A(-\lambda_k)T_N(-\lambda_k)-B(-\lambda_k)U_N(-\lambda_k)}{\mu+\lambda_k}\right\}\times T_{N+1}(\mu)-\sum_{k=1}^{\frac{1}{2}(N-1)}\frac{\beta_k B(-\lambda_k)}{\mu+\lambda_k}, \tag{5.5}$$

TABLE II. Values of relaxation constants.

Σ_s/Σ	T_1 λ_0	T_3 λ_0	T_3 λ_1
0	0.7071068	0.9238795	0.3826834
0.1	0.745356	0.937569	0.397494
0.2	0.790569	0.954991	0.413915
0.3	0.845154	0.977795	0.432174
0.4	0.912871	1.008695	0.452501
0.5	1.000000	1.052438	0.475088
0.6	1.118034	1.118034	0.500000
0.7	1.290994	1.224745	0.527046
0.8	1.581139	1.422897	0.555606
0.9	2.236068	1.912672	0.584541
1.0			0.6123724

where

$$D/C=-(1/\Sigma)\{\mu_1+\mu_2+\cdots+\mu_{\frac{1}{2}(N+1)}-\lambda_1\cdots-\lambda_{\frac{1}{2}(N-1)}\},$$

$$\beta_k=-\frac{C}{\Sigma}\frac{(-1)^{\frac{1}{2}(N+1)}}{2^N S'(\lambda_k)p(-\lambda_k)}\times\frac{1}{A(\lambda_k)T_N(\lambda_k)-B(\lambda_k)U_N(\lambda_k)}, \tag{5.6}$$

$$S'(\lambda_k)=(\lambda_k-\lambda_1)(\lambda_k-\lambda_2)\cdots[\lambda_k-\lambda_{\frac{1}{2}(N-1)}].$$

Thus, the neutron distributions are determined in the T_N approximation. A quantity which is of practical interest in a reactor and which also gives a measure for the accuracy of the approximation is the extrapolated end-point value.[21] This quantity is easily obtained.

$$\Sigma>\Sigma_s,\quad d=\frac{\lambda_0}{2\Sigma}\ln\frac{(\lambda_0+\mu_1)(\lambda_0+\mu_2)\cdots}{(\lambda_0-\mu_1)(\lambda_0-\mu_2)\cdots}\times\frac{(\lambda_0-\lambda_1)(\lambda_0-\lambda_2)\cdots}{(\lambda_0+\lambda_1)(\lambda_0+\lambda_2)\cdots}, \tag{5.8a}$$

$$\Sigma=\Sigma_s,\quad d=(1/\Sigma)\{\mu_1+\mu_2+\cdots-\lambda_1-\lambda_2\cdots-\lambda_{\frac{1}{2}(N-1)}\}. \tag{5.8b}$$

The expression (5.8) is the same as in the P_L approximation[22,23] if the relaxation constants and the μ's were replaced by corresponding quantities in the P_L approximation. Conkie did not give the explicit expression (5.8). Table I shows the values of d for various values of Σ_s/Σ. Conkie gives values only for $\Sigma_s/\Sigma=0.1$, 0.5, and 1.0 in the T_1, T_3, and T_5 approximations. It is possible to calculate d from the formula (5.8) for the T_7 and T_9 approximations. The relaxation constants for T_1 and T_3 approximations were calculated from (2.17) by KDC-I electronic computer of Kyoto

[21] See p. 260 of reference 2.

[22] See Eq. (9.46b) and Eq. (9.46c) of the Weinberg and Wigner text.

[23] After the revised manuscript for Sec. 5, including Eq. (5.8a), was submitted, the referee pointed out that it is easy to obtain (5.8a).

University and given in Table II. However, the values of d with an asterisk are not in agreement with Conkie's results. Relaxation constants for the T_5, T_7, and T_9 approximations were also calculated by KDC-I and found to be in complete agreement with Conkie's values.

From Table I we can conclude that, for weak absorbers, the P_L method gives a value closer to the exact value than the T_N method does. However, for the values of Σ_s/Σ smaller than 0.8 or 0.7, the T_N method gives a better value for d, In spite of the good approximation of the T_N method for Σ_s/Σ between 0.8 or 0.7 and 0.5; for very strong absorbers, both the T_N and P_L methods are no longer good approximations. For such strong absorbers, another method might be desirable.

APPENDIX 1. PROOF OF EQ. (2.11)

We prove the summation formulas (2.11) by mathematical induction. For $N=1$, Eq. (2.11) surely holds. If (2.11) is correct for N, then

$$1+2\sum_{n=1}^{N+1} T_n(\mu)T_n(\lambda)=\frac{T_{N+1}(\mu)T_N(\lambda)-T_N(\mu)T_{N+1}(\lambda)}{\mu-\lambda}+2T_{N+1}(\mu)T_{N+1}(\lambda),$$

$$=\frac{\{2\mu T_{N+1}(\mu)-T_N(\mu)\}T_{N+1}(\lambda)-T_{N+1}(\mu)\{2\lambda T_{N+1}(\lambda)-T_N(\lambda)\}}{\mu-\lambda}.$$

By virtue of the recurrence relation (2.6a),

$$1+2\sum_{n=1}^{N+1} T_n(\mu)T_n(\lambda)=\frac{T_{N+2}(\mu)T_{N+1}(\lambda)-T_{N+1}(\mu)T_{N+2}(\lambda)}{\mu-\lambda}.$$

Therefore, by mathematical induction, the formula (2.11) is proven.

APPENDIX 2. PROOF OF EQ. (2.15)

First we prove the relation

$$U_l(\lambda)T_1(\lambda)-T_l(\lambda)U_1(\lambda)=U_{l-1}(\lambda).$$

The above relation surely holds for $l=1$ and for $l=2$. If the above relation holds for $l=L$ and for $l=L-1$, then it holds for $l=L+1$ by the recurrence relation (2.6). Therefore, the above relation is right for arbitrary l by mathematical induction. Moreover, we have the relation

$$U_l(\lambda)T_0(\lambda)-T_l(\lambda)U_0(\lambda)=U_l(\lambda).$$

If the relation

$$U_l(\lambda)T_m(\lambda)-T_l(\lambda)U_m(\lambda)=U_{l-m}(\lambda)$$

holds for $m=M$ and for $m=M-1$, then, by (2.6), the above equation is right for $m=M+1$. However, the above relation surely holds for $m=0$ and for $m=1$. Therefore, by mathematical induction, (2.15) holds.

Theory of the Dipole Antenna and the Two-Wire Transmission Line*

TAI TSUN WU
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts
(Received September 19, 1960)

The properties of the dipole antenna are studied by an approximate procedure that makes use of the Wiener-Hopf integral equation. In particular, the input admittance and the radiation pattern are found. The present results thus supplement the existing theories, which are concerned mostly with short dipoles. The same procedure is then applied to several related problems. First, the back-scattering cross section of a dipole antenna is found approximately for normal incidence. Secondly, the two-wire transmission line is studied in detail by considering it to be two coupled dipole antennas. The capacitive end correction for an open end is evaluated, and the radiated power and the radiation resistance are found for a resonant section of transmission line with both ends open. Finally, the dielectric-coated antenna is considered briefly.

INTRODUCTION

THE problem of the center-driven cylindrical antenna has been investigated by numerous authors.[1] There exist now principally three kinds of attacks: iterative procedures, variational methods, and Fourier series expansions. Recently, Duncan and Hinchey[2] used the last method to get some extremely interesting results. To employ this method, it is essential to carry out the calculations on a high-speed digital computer. The other two kinds of methods are described in detail in the monumental book of King,[3] which will be designated by K in subsequent references. For convenience, numerous references will be made to this book instead of the original papers.

For thin dipole antennas of length not much more than one wavelength, the King-Middleton iterative procedure [K, p. 101ff] yields current distributions in good agreement with the experimental results. So far as the input impedance is concerned, the various iterative and variational methods seem to give comparable results for thin antennas not much more than two wavelengths in total length [K, p. 843]. If h is the half-length of the thin antenna, it is reasonable to think that cases where $h \lesssim \lambda$ are fairly well understood.

The situation is much less favorable for $h > \lambda$. Both theoretical and experimental results are very scarce in this vast range of antenna lengths. So far as the author is aware, the following three pieces of information are available: (1) first-order King-Middleton distributions of currents for antennas with $h = 6\lambda$ [K, p. 115], (2) Robert's measurement of current distribution for an antenna with $h \sim 11\lambda$ [K, p. 140], and (3) Altshuler's[4] recent measurements of current distributions for antennas with $h \sim 2\lambda$. No general conclusions can be drawn from these results, except that there is essentially no evidence of any agreement between theory and experiment. A more direct comparison is possible in the related problem of the determination of the back-scattering cross sections of a dipole receiving antenna at normal incidence [K, pp. 508 and 516]. Here the theoretical and experimental results are certainly in disagreement for $h \gtrsim 0.8\lambda$.

If these few pieces of experimental data are not dismissed as incorrectly recorded, then one is forced to consider the possibility that the existing theories of the dipole antenna may be inapplicable when $h \gtrsim \lambda$. An iterative procedure is one which gives correction terms to an initial, rough approximation of the current distribution. Since usually only a small number of iterations can be carried out, the accuracy of the results depends critically on the accuracy of the initial approximation. Since the iterative procedure itself can hardly be questioned, an explanation of the discrepancy between theory and experiment may be sought in the inadequacy of the initial approximation.

In view of this shortcoming of both the variational methods and the iterative procedures, in this paper a completely different approach is to be adopted where there is no necessity of guessing a first approximation. In other words, the purpose here is to find a procedure that starts from the integral-equation formulation of the dipole antenna problem and gives for the various properties of the antenna approximate expressions that are not too complicated and yet are reasonably accurate. A possible procedure that satisfies this criterion is given in Part I. There, only the symmetrical, center-driven, dipole antenna is treated for definiteness; the generalization to the asymmetrical case seems to offer no difficulty in principle. In Part II, the same procedure is applied to a few somewhat more complicated cases. In Part III, some of the results are tabulated and

* Supported in part by NSF Grant 9721-7750 and Contract Nonr-1866(32).

[1] A small sample of the vast literature on the dipole antenna and related topics is: L. V. King, Trans. Roy. Soc. (London) **A236**, 381 (1937); E. Hallen, Nova Acta Regiae Soc. Sci. Upsaliensis 4–11, 1 (1938); S. A. Schelkunoff, Proc. IRE **29**, 493 (1941); R. King and F. G. Blake, Jr., Proc. IRE **30**, 335 (1942); C. J. Bouwkamp, Physica **9**, 609 (1942); M. C. Gray, J. Appl. Phys. **15**, 61 (1944); R. King and D. Middleton, Quart. Appl. Math. **3**, 302 (1946); J. H. Van Vleck, F. Bloch, and M. Hamermesh, J. Appl. Phys. **18**, 274 (1947); C. T. Tai, Technical Report No. 12, Project No. 188, Stanford Research Institute (1950); J. E. Storer, doctoral dissertation, Harvard University (1951). Variational methods are used in the last two references.

[2] R. H. Duncan and F. A. Hinchey, Sandia Corp., Rept. No. SCTM367-59-(14), 1960.

[3] R. W. P. King, *The Theory of Linear Antennas* (Harvard University Press, Cambridge, Massachusetts, 1956).

[4] E. Altshuler, Ph.D. dissertation, Harvard University, 1960.

compared with existing experimental data when possible.

In the procedure of Part I, it is assumed that the radius of the dipole antenna is very small compared with either the wavelength or the length of the dipole. With the antenna assumed to be perfectly conducting, the total tangential electric field is zero on the surface of the antenna except at the driving point, and hence by consideration of symmetry, the tangential component of the vector potential is proportional to

$$C\cos kz+\tfrac{1}{2}\sin k|z|,$$

where k is the wave number while z is the distance along the dipole antenna. By expressing the vector potential as an integral over the current on the antenna, an integral equation is obtained and may be used to determine the current. The constant C is fixed by the condition that the current vanishes at the ends of the antenna, and, roughly speaking, large values of C indicate resonance while small values indicate antiresonance. The determination of C is a major step in this procedure. The content of the above-mentioned integral equation may be qualitatively described by saying that the current on the antenna is determined by the conditions that (1) the vector potential is as prescribed on the antenna, say $-h\leq z\leq h$, and (2) the current vanishes outside the antenna, i.e., $z>h$ or $z>-h$. In order to determine C, it is necessary to express the current at $z=h$ in terms of C. When the radius of the antenna is small, the kernel of the integral equation is relatively large for small values of the argument, and relatively small and oscillatory for large values of the argument. Accordingly, the vector potential is comparatively small for $z>h$ or $z<-h$. Furthermore, the values of the current at $z=h$ is not sensitive to the condition at $z<-h$. Consequently, the current at $z=h$ may be found approximately by assuming instead the conditions that (1′) the vector potential is as prescribed on the antenna and vanishes for $z<-h$, and (2′) the current vanishes for $z>h$. For this modified problem, the current can be found explicitly by the method of Wiener and Hopf, and this gives an approximate determination of C. Once C is known, some of the properties of the antenna, such as the input impedance and the radiation pattern, may be found by assuming the somewhat less accurate but considerably simpler condition that the vector potential is as prescribed on the antenna and vanishes for both $z<-h$ and $z>h$. This in principle gives formulas for the desired quantities. However, as is well-known, the solutions of Wiener-Hopf equations are in general extremely complicated, being integrals of exponentials of contour integrals. In order to compare with experimental data, the theoretical results have to be evaluated numerically. Since a direct evaluation of these integrals seems extemely difficult, it is a necessary task to obtain simple approximations to these rather involved expressions. For this purpose, the smallness of the radius of the antenna is used repeatedly. It should be emphasized that, in view of the magnitudes of the various quantities involved, the approximation cannot be simply described as an expansion in some small parameter. For this reason, the procedure is very complicated in detail although the final result is simple. The particular procedure adopted yields quite accurate answers over large ranges of antenna dimensions relative to the wavelength. After the completion of the present theory. Prasad measured, with a wavelength of about 37 cm, the input impedances of dipole antennas of diameter $\frac{1}{4}$ in. and various lengths up to about 12 ft.[5] Except for a small discrepency for which the theory may not be responsible, the theoretical and experimental results are in unexpectedly good agreement. The detailed comparisons are being carried out.

PART I. CENTER-DRIVEN DIPOLE ANTENNA

1. Formulation of the Problem

The dipole antenna is assumed to be a symmetric tubular antenna of zero thickness and infinite conductivity defined by $r=a$, $|z|\leqslant h$, as shown in Fig. 1. If $I(z)$ is the total z component of the current at z, including both the current on the outside of the tube and that on the inside, then

$$I(h)=0, \tag{1.1}$$

and the z component of the vector potential on the cylinder $r=a$ is given by

$$A(z)=\frac{\mu_0}{4\pi}\int_{-h}^{h}dz'I(z')K(z-z'), \tag{1.2}$$

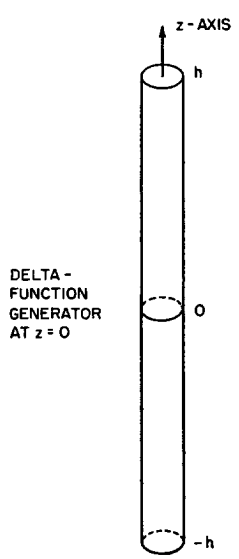


FIG. 1. The center-driven dipole antenna of half-length h.

[5] I am indebted to Dr. Sheila Prasad for keeping me informed on the progress of the experiment.

where μ_0 is the free-space magnetic permeability and the kernel K is given by

$$K(z)=(2\pi)^{-1}\int_{-\pi}^{\pi}d\theta[z^2+(2a\sin\theta/2)^2]^{-\frac{1}{2}} \times\exp\{ik[z^2+(2a\sin\theta/2)^2]^{\frac{1}{2}}\}. \quad (1.3)$$

Here k is the wave number. As usual in antenna theory, the term "vector potential" is used to denote the vector potential in the Lorentz gauge satisfying the Sommerfeld radiation condition. On the other hand, if the strength of the δ-function generator is taken to be -1, the z component of the vector potential is of the form

$$A(z)=\frac{\mu_0 i}{\zeta_0}[C\cos kz+\tfrac{1}{2}\sin k|z|] \quad (1.4)$$

for $|z|<h$. Here ζ_0 is the characteristic impedance of free space. The combination of (1.2) and (1.4) gives the integral equation for $I(z)$, where C is to be determined by the boundary condition (1.1).

In the King-Middleton iterative solution, the vector potential is assumed to be proportional to the current at the same point in order to get the initial rough approximation. This assumption is reasonable except near the ends of the dipole antenna, where the current changes rapidly with z. Unfortunately, the value of the constant C is determined at $z=h$, precisely where this approximation is poor. It is here proposed to find C by a different procedure, making use of the observation that $A(z)$ is relatively small for $|z|>h$ [K, pp. 429 and 527]. This observation is useful because then the antenna may be approximated by a semi-infinite one driven by a vector potential distribution which is of the form (1.4) for $|z|<h$ and is zero for $z>h$. Consequently, the problem of the semi-infinite antenna is to be studied first.

2. Semi-Infinite Antenna

The semi-infinite antenna is described by an integral equation of the Wiener-Hopf type:

$$\int_0^\infty dz' I(z')K(z-z')=F(z), \quad (2.1)$$

where $F(z)$ is known for $z>0$. It is assumed that $F(0+)$ and $F'(0+)$ both exist, and $F(z)$ approaches zero sufficiently rapidly as $z\to\infty$. Under these circumstances, $I(z)$ is in general unbounded near $z=0$. If, however, $F(z)$ satisfies a certain integral condition, $I(z)$ becomes bounded near $z=0$, and furthermore $\lim_{z\to 0} I(z)=0$. It is desired to find this integral condition. For this purpose, the usual Wiener-Hopf procedure may be used.

Consider k to have a small positive imaginary part which is eventually allowed to approach zero:

$$\mathrm{Im}k=\epsilon>0, \quad (2.2)$$

and define the relevant Fourier transforms by

$$\bar{I}(\zeta)=\int_0^\infty dz I(z)\exp(-i\zeta z), \quad (2.3a)$$

$$\bar{K}(\zeta)=\int_{-\infty}^{\infty}dzK(z)\exp(-i\zeta z) =\pi i J_0[a(k^2-\zeta^2)^{\frac{1}{2}}]H_0^{(1)}[a(k^2-\zeta^2)^{\frac{1}{2}}], \quad (2.3b)$$

$$\bar{F}_-(\zeta)=\int_0^\infty dzF(z)\exp(-i\zeta z), \quad (2.3c)$$

and

$$\bar{F}_+(\zeta)=\int_{-\infty}^0 dzF(z)\exp(-i\zeta z). \quad (2.3d)$$

Then (2.1) leads to

$$\bar{I}(\zeta)\bar{K}(\zeta)=\bar{F}_-(\zeta)+\bar{F}_+(\zeta). \quad (2.4)$$

For any function f of ζ analytic in the strip $|\mathrm{Im}\zeta|<\epsilon$, define

$$[f(\zeta)]_-=-(2\pi i)^{-1}\int_{-\infty+i\epsilon/2}^{\infty+i\epsilon/2}d\zeta'(\zeta'-\zeta)^{-1}f(\zeta'), \quad (2.5a)$$

and

$$[f(\zeta)]_+=(2\pi i)^{-1}\int_{-\infty-i\epsilon/2}^{\infty-i\epsilon/2}d\zeta'(\zeta'-\zeta)^{-1}f(\zeta'), \quad (2.5b)$$

where the Cauchy principal values are taken at ∞. Then $[f(\zeta)]_-$ is analytic for $\mathrm{Im}\zeta<\epsilon/2$, and $[f(\zeta)]_+$ is analytic for $\mathrm{Im}\zeta>-\epsilon/2$. Furthermore, in the strip $|\mathrm{Im}\zeta|<\epsilon/2$,

$$f(\zeta)=[f(\zeta)]_-+[f(\zeta)]_+. \quad (2.6)$$

In terms of the function $\bar{K}(\zeta)$, define

$$\bar{L}_\pm(\zeta)=\exp[\mp\ln\bar{K}(\zeta)]_\pm, \quad (2.7)$$

then

$$\bar{K}(\zeta)=\bar{L}_-(\zeta)/\bar{L}_+(\zeta) \quad (2.8)$$

for $|\mathrm{Im}\zeta|<\epsilon/2$, and furthermore,

$$\bar{L}_-(\zeta)=[\bar{L}_+(-\zeta)]^{-1}. \quad (2.9)$$

From (2.4) and (2.8) it follows that

$$\bar{I}(\zeta)\bar{L}_-(\zeta)-[\bar{F}_-(\zeta)\bar{L}_+(\zeta)]_- =[\bar{F}_-(\zeta)\bar{L}_+(\zeta)]_+ +\bar{F}_+(\zeta)\bar{L}_+(\zeta). \quad (2.10)$$

This defines an entire function, which must be zero because of the behavior at infinity. Therefore,

$$\bar{I}(\zeta)\bar{L}_-(\zeta)=[\bar{F}_-(\zeta)\bar{L}_+(\zeta)]_-. \quad (2.11)$$

It is assumed that $I(z)$ has no singularity at $z=0$. Thus, as $|\zeta|\to\infty$ in the half-plane $|\mathrm{Im}\zeta|<\epsilon/2$,

$$\bar{I}(\zeta)=O(|\zeta|^{-1}), \quad (2.12)$$

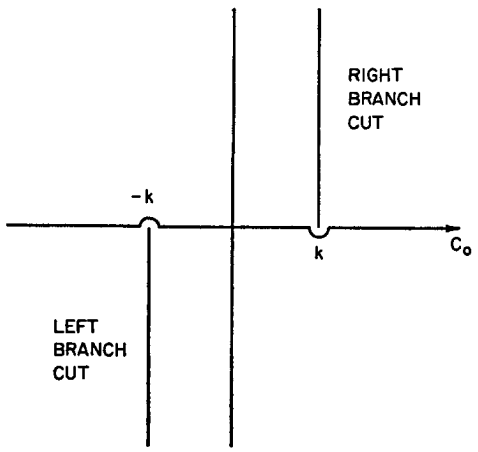


FIG. 2. The ζ plane and the contour C_0.

and furthermore

$$\bar{L}_-(\zeta)=O(|\zeta|^{-\frac{1}{2}}). \tag{2.13}$$

It, therefore, follows from (2.11) that

$$[\bar{F}_-(\zeta)\bar{L}_+(\zeta)]_-=O(|\zeta|^{-1}), \tag{2.14}$$

as $|\zeta|\to\infty$. This is the required condition on $F(z)$.

In order to put (2.14) in a simpler form, let $\epsilon\to 0+$ and note that $\bar{K}(\zeta)$ is analytic in the cut plane as shown in Fig. 2. Furthermore, the contours used in defining $[f(z)]_\pm$ both become C_0, also shown in Fig. 2, and $\bar{L}_+(\zeta)$ and $\bar{L}_-(\zeta)$ are analytic in the entire complex plane except, respectively, the left- and right-branch cuts of $\bar{K}(\zeta)$. Define

$$\bar{F}_-{}^0(\zeta)=\bar{F}_-(\zeta)-F(0+)/[i(\zeta+k)], \tag{2.15}$$

then $\bar{F}_-{}^0(\zeta)\bar{L}_+(\zeta)$ is integrable along C_0, at least in the sense of Euler summability. Therefore, the left-hand side of (2.14) is explicitly

$$\begin{aligned}&[\bar{F}_-(\zeta)\bar{L}_+(\zeta)]_-\\&=F(0+)\left[\frac{1}{i(\zeta+k)}\bar{L}_+(\zeta)\right]_-+[\bar{F}_-{}^0(\zeta)\bar{L}_+(\zeta)]_-\\&=-(2\pi i)^{-1}\int_{C_0}d\zeta'(\zeta'-\zeta)^{-1}\bar{F}_-{}^0(\zeta)\bar{L}_+(\zeta).\end{aligned} \tag{2.16}$$

The condition (2.14) is thus explicitly

$$\int_{C_0}d\zeta\bar{F}_-{}^0(\zeta)\bar{L}_+(\zeta)=0. \tag{2.17}$$

3. Approximations for Thin Antennas

The condition (2.17) for the vanishing of the current at the end of the semi-infinite antenna is exact. This rather complicated condition can be greatly simplified if it is assumed (1) that the antenna is thin in the sense that $a/\lambda\ll 1$ and (2) that the characteristic distance for the variation of $F(z)$ is much larger than a. Under the second assumption, which is satisfied if the right-hand side of (1.4) is taken to be $F(h-z)$, the behavior of $\bar{L}_+(\zeta)$ for $\zeta\sim a^{-1}$ is unimportant, and thus it is permissible to use the following approximation of $\bar{K}(\zeta)$:

$$\bar{K}(\zeta)=2\Omega_1-\ln[(k^2-\zeta^2)/k^2], \tag{3.1}$$

where

$$\Omega_1=\Omega_0+i\pi/2, \tag{3.2}$$

and

$$\Omega_0=\ln(2/ka)-\gamma. \tag{3.3}$$

In (3.3), γ is Euler's constant, numerically about 0.57722. In view of a previous discussion[6] on the meaning of the input admittance of a linear antenna driven by a delta function, the approximation (3.1) does not introduce any further error beyond those inherent in the model of the delta-function generator. Once (3.1) is used, it is possible to define

$$\bar{M}(\zeta)=[\bar{K}(\zeta)]^{-1}, \tag{3.4}$$

$$\bar{M}(\zeta)=\int_{-\infty}^{\infty}dzM(z)\exp(-i\zeta z), \tag{3.5}$$

and

$$\bar{L}_+(\zeta)=\int_{-\infty}^{\infty}dzL_+(z)\exp(-i\zeta z). \tag{3.6}$$

It should be emphasized that $M(z)$ and $L_+(z)$ are meaningful only when an approximation of the type (3.1) is used. The exact $\bar{M}(\zeta)$ and $\bar{L}_+(\zeta)$ are not the Fourier transforms of integrable functions.

First, an approximation is to be found for $M(z)$, which is given by

$$\begin{aligned}M(z)=ik(2\pi)^{-1}e^{ikz}&\int_0^\infty d\xi e^{-kz\xi}\\&\times(\{2\Omega_0-\ln[\xi(2+i\xi)]-\tfrac{1}{2}i\pi\}^{-1}\\&\quad-\{2\Omega_0-\ln[\xi(2+i\xi)]+\tfrac{3}{2}i\pi\}^{-1})\end{aligned} \tag{3.7}$$

for $z>0$. When z is not too small, it will be a reasonably good approximation to replace $\ln(2+\xi)$ in the integrand by $\ln 2$, and replace $\ln\xi$ by the average

$$\left[\int_0^\infty d\xi e^{-kz\xi}\right]^{-1}\int_0^\infty d\xi e^{-kz\xi}\ln\xi=-\ln(kz)-\gamma. \tag{3.8}$$

With these replacements, (3.7) becomes

$$\begin{aligned}M(z)=i(2\pi z)^{-1}e^{ikz}[&(2\Omega_0-\ln(2/kz)+\gamma-\tfrac{1}{2}i\pi)^{-1}\\&-(2\Omega_0-\ln(2/kz)+\gamma+\tfrac{3}{2}i\pi)^{-1}].\end{aligned} \tag{3.9}$$

Therefore, $\int_1^\infty|M(z)|dz$ exists. The same statement is true of $L_+(-z)$. If the integral equation had been written in terms of the electric field instead of the vector potential, the corresponding M and L_+ functions would not have this property, i.e., they would not

[6] T. T. Wu and R. W. P. King, J. Appl. Phys. **30**, 76 (1959).

approach zero sufficiently rapidly. The rapid decrease of M and L_+ implies that the behavior of A for $z>h$ is relatively unimportant in the present calculation. As will be seen in Sec. 8, the condition that M and L_+ should decrease sufficiently rapidly dictates the choice of the differential operator that gives E from A.

4. Application to the Dipole Antenna

To apply the results of the last two sections to the dipole antenna, it is convenient to define for $Z>0$.

$$S(Z)=\int_{C_0} d\zeta \bar{M}(\zeta) \int_0^\infty dz \exp(-i\zeta z) \times \sin kz H(Z-z), \quad (4.1)$$

$$T(Z)=\int_{C_0} d\zeta \bar{M}(\zeta) \int_0^\infty dz \exp(-i\zeta z) \times [\cos kz H(Z-z)-\exp(-ikz)], \quad (4.2)$$

$$S'(Z)=\int_{C_0} d\zeta \bar{L}_+(\zeta) \int_0^\infty dz \exp(-i\zeta z) \times \sin kz H(Z-z), \quad (4.3)$$

and

$$T'(Z)=\int_{C_0} d\zeta \bar{L}_+(\zeta) \int_0^\infty dz \exp(-i\zeta z) \times [\cos kz H(Z-z)-\exp(-ikz)], \quad (4.4)$$

where H is the Heaviside function

$$H(x)=\begin{cases}1 & x>0\\ 0 & x<0.\end{cases}$$

By comparing (1.2) and (1.4) with (2.1), it is seen that for the dipole antenna and for $z>0$

$$F(z)=4\pi i\zeta_0^{-1}[C\cos k(h-z) + \tfrac{1}{2}\sin k|h-z|]H(2h-z), \quad (4.5)$$

which may be alternatively expressed as

$$F(z)=4\pi i\zeta_0^{-1}[\sin kh \cos kz H(h-z) - \cos kh \sin kz H(h-z) + (C\cos kh - \tfrac{1}{2}\sin kh) \times \cos kz H(2h-z) + (C\sin kh + \tfrac{1}{2}\cos kh) \times \sin khH(2h-z)]. \quad (4.6)$$

When this is substituted into (2.17), the result is, using (4.3–4.4),

$$\sin khT'(h) - \cos khS'(h) + (C\cos kh - \tfrac{1}{2}\sin kh)T'(2h) + (C\sin kh + \tfrac{1}{2}\cos kh)S'(2h) = 0. \quad (4.7)$$

The constant C is thus explicitly given by

$$C=-\tfrac{1}{2}[\cos khT'(2h) + \sin khS'(2h)]^{-1}\{\sin kh[2T'(h)-T'(2h)] - \cos kh[2S'(h)-S'(2h)]\}. \quad (4.8)$$

On the other hand, it follows from (4.5) that the input admittance of the antenna is

$$Y=2i\zeta_0^{-1}[S(h)+CU(h)], \quad (4.9)$$

where

$$U(Z)=\int_{C_0} d\zeta \bar{M}(\zeta) \int_{-Z}^{Z} dz \exp(-i\zeta z)\cos kz. \quad (4.10)$$

In order to put (4.8) and (4.10) in forms that are easily computed, many approximations have to be made. In (4.1–4), the contour C_0 can be deformed so that it is wrapped around the left-branch cut. When kZ is not too small, the contributions to the four functions, insofar as the ζ integral is concerned, come mainly from the region $|\zeta+k|Z\lesssim 1$. On the other hand, from (2.8), $\bar{L}_+(\zeta)$ is the same as $\bar{M}(\zeta)\bar{L}_-(\zeta)$, where $\bar{L}_-(\zeta)$ is analytic in the vicinity of $\zeta=-k$. Therefore, approximately

$$S'(Z)=\bar{L}_-(-k)S(Z), \quad (4.11)$$

and

$$T'(Z)=\bar{L}_-(-k)T(Z). \quad (4.12)$$

Since dipole antennas with $kh\lesssim\pi$ are well understood in terms of the King-Middleton theory, the task here is thus to calculate $S(Z)$, $T(Z)$, and $U(Z)$ for kZ large. This is carried out in Appendix A, with the results

$$2S(Z) = -\ln[1+\pi i(\Omega_0-\ln 2)^{-1}] - (\pi^2/12)[(\Omega_0-2\ln 2)^{-2}-(\Omega_0-2\ln 2+\pi i)^{-2}] + \ln\{[\Omega_2(Z)]^{-1}\Omega_3(Z)\} + \tfrac{1}{2}\gamma'\{[\Omega_2(Z)]^{-2}-[\Omega_3(Z)]^{-2}\} - i(2kZ)^{-1} \times \exp(2ikZ)\{[\Omega_2(Z)]^{-1}-[\Omega_3(Z)]^{-1}\}, \quad (4.13)$$

$$-2iT(Z) = -\ln[1+\pi i(\Omega_0-\ln 2)^{-1}] - (\pi^2/12)[(\Omega_0-2\ln 2)^{-2}-(\Omega_0-2\ln 2+\pi i)^{-2}] - \ln\{[\Omega_2(Z)]^{-1}\Omega_3(Z)\} - \tfrac{1}{2}\gamma'\{[\Omega_2(Z)]^{-2}-[\Omega_3(Z)]^{-2}\} - i(2kZ)^{-1} \times \exp(2ikZ)\{[\Omega_2(Z)]^{-1}-[\Omega_3(Z)]^{-1}\}, \quad (4.14)$$

and

$$iU(Z) = \ln\{[\Omega_2(Z)]^{-1}\Omega_3(Z)\} + \tfrac{1}{2}\gamma'\{[\Omega_2(Z)]^{-2}-[\Omega_3(Z)]^{-2}\} + i(2kZ)^{-1} \times \exp(2ikZ)\{[\Omega_2(Z)]^{-1}-[\Omega_3(Z)]^{-1}\}. \quad (4.15)$$

In (4.13–4.15), the following symbols have been used:

$$\Omega_2(Z)=2(\Omega_0-\ln 2)+\ln(2kZ)+\gamma-i\pi/2, \quad (4.16)$$

$$\Omega_3(Z)=2(\Omega_0-\ln 2)+\ln(2kZ)+\gamma+3i\pi/2, \quad (4.17)$$

and

$$\gamma'=\Gamma''(1)-\gamma^2. \quad (4.18)$$

γ' is numerically about 1.6449.

5. The Radiation Field

The procedure of Sec. 4 may be used to get the current distribution not close to either the generator

or the ends. However, this point is not studied further because excessive numerical computation seems to be necessary.

Let a spherical coordinate system (r,θ,ϕ) be set up such that the ends of the antenna are at $(h,0,\phi)$ and (h,π,ϕ). All field quantities are independent of ϕ because of rotational symmetry. Define the field pattern by

$$F(\theta)=-\lim_{r\to\infty} E_\theta(r,\theta)r\exp(-ikr). \tag{5.1}$$

By (1.2), this is

$$F(\theta)=i\omega\mu_0(4\pi)^{-1}\sin\theta\int_{-h}^{h}dzI(z)(2\pi)^{-1}\int_{-\pi}^{\pi}d\theta' \times\exp[-ik(z\cos\theta-a\cos\theta'\sin\theta)]. \tag{5.2}$$

When the small term proportional to a is neglected, this simplifies to

$$F(\theta)=i\omega\mu_0(4\pi)^{-1}\bar{I}(k\cos\theta)\sin\theta. \tag{5.3}$$

Within the framework of the present approximation, this is simply

$$F(\theta)=-\tfrac{1}{2}\sin\theta[\Omega_1-\ln\sin\theta]^{-1}\Big\{C\frac{\sin[kh(1-\cos\theta)]}{1-\cos\theta} +C\frac{\sin[kh(1-\cos\theta)]}{1+\cos\theta}+\frac{1}{2}\frac{1-\cos[kh(1-\cos\theta)]}{1-\cos\theta} +\frac{1}{2}\frac{1-\cos[kh(1+\cos\theta)]}{1+\cos\theta}\Big\}. \tag{5.4}$$

This differs from the usual zeroth-order field pattern only in the appearance of the factor $[\Omega_1-\ln\sin\theta]^{-1}$ when C is chosen in a sufficiently simple way. For a long dipole, this factor has the effect of reducing the end-firing major lobes.

PART II. GENERALIZATIONS

6. Back-Scattering Cross Section

In this and the two following sections, the procedure of Part I is to be applied to three situations mathematically similar to the one already treated. The first problem is the determination of the back-scattering cross section of an unloaded dipole antenna at normal incidence. The geometry is shown in Fig. 3. Without loss of generality, the incident electric field is taken to be 1 at $x=0$. Since the radius a of the dipole antenna is assumed to be very small compared with the wavelength, the scattered field is considered to be rotationally symmetrical. Under this approximation, the current induced on the antenna satisfies the integral equation:

$$\int_{-h}^{h}dz'I_s(z')K(z-z')=4\pi i(\mu_0\omega)^{-1}[1+C_s\cos kz]. \tag{6.1}$$

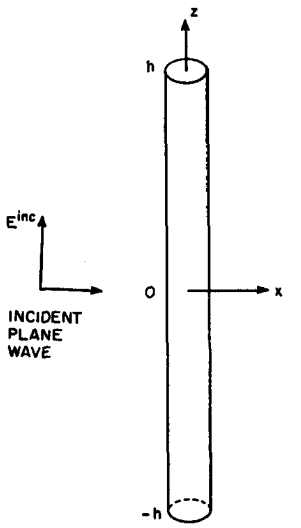


FIG. 3. The receiving dipole of half-length h.

Here the subscript s is used to distinguish the present scattering problem, and the constant C_s is to be determined from the boundary condition

$$I_s(h)=0. \tag{6.2}$$

In terms of I_s, the back-scattering cross section is

$$\sigma_B=(4\pi)^{-1}\omega^2\mu_0^2\left|\int_{-h}^{h}dzI_s(z)\right|^2. \tag{6.3}$$

Within the framework of the approximations used in Part I, this is given by

$$\sigma_B=4\pi|\Omega_1|^{-2}|h+C_sk^{-1}\sin kh|^2. \tag{6.4}$$

Without the C_s term, this is just the nonresonant formula of Chu.[7]

As in the case of the driven antenna, the value of C_s is to be determined from (2.17) with the following form for F,

$$F_s(z)=4\pi i(\mu_0\omega)^{-1}[1+C_s\cos k(h-z)]H(2h-z). \tag{6.5}$$

It is thus useful to define

$$V'(Z)=\int_{C_0}d\zeta\bar{L}_+(\zeta)\int_0^\infty dz\exp(-i\zeta z) \times[H(Z-z)-\exp(-ikz)]. \tag{6.6}$$

Thus, C_s is given by

$$C_s=-V'(2h)/[\cos khT'(2h)+\sin khS'(2h)], \tag{6.7}$$

or approximately with (4.11–4.12)

$$C_s=-V'(2h)\bar{L}_+(k)/[\cos khT(2h)+\sin khS(2h)]. \tag{6.8}$$

[7] L. J. Chu (private communication). See also J. H. Van Vleck, F. Block, and M. Hamermesh, J. Appl. Phys. 18, 274 (1947).

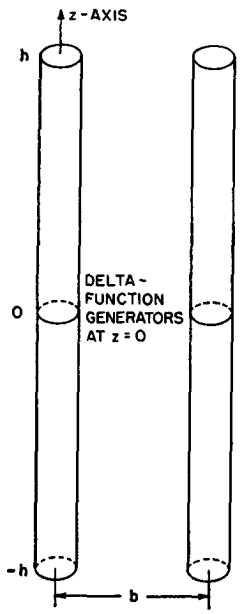


FIG. 4. Two identical, parallel, nonstaggered antennas of half-length h.

From (6.6), V' is found to be

$$V'(Z)=2\pi\bar{L}_+(0)+i[\bar{L}_+(k)]^{-1}\int_{-\infty}^{-1}d\zeta\zeta^{-1}\times\exp(-i\zeta kZ)\{[2\Omega_0-\ln(\zeta^2-1)]^{-1}-[2\Omega_0-\ln(\zeta^2-1)+2\pi i]^{-1}\}. \quad (6.9)$$

For kZ large, this is approximately

$$V'(Z)\bar{L}_+(k)=2\pi\bar{L}_+(0)\bar{L}_+(k)+2e^{ikZ}\{[\Omega_2(Z)]^{-1}-[\Omega_3(Z)]^{-1}\}. \quad (6.10)$$

To avoid numerical integration, it is convenient to use

$$\bar{L}_+(0)\bar{L}_+(k)\sim(2\Omega_1)^{-1}. \quad (6.11)$$

The explicit formula for the back-scattering cross section is given by (6.4), with (6.8) and (6.10–6.11).

7. The Two-Wire Line

In principle, the method of Part I is applicable to the case of a system of two identical, nonstaggered, parallel dipole antennas as shown in Fig. 4 provided that it is admissible to assume that the current distribution is rotationally symmetrical on each dipole. This assumption is reasonable if the antennas are thin and if the separation b between the antennas is not too small. If the symmetrical and antisymmetrical parts of the currents are used to set up integral equations, then the present case differs from the case of a single dipole only in the appearance of a more complicated kernel. For example, the kernel for the antisymmetrical part of the current is

$$K_a(z)=K(z)-(z^2+b^2)^{-\frac{1}{2}}\exp[ik(z^2+b^2)^{\frac{1}{2}}], \quad (7.1)$$

with the Fourier transform

$$\bar{K}_a(\zeta)=\bar{K}(\zeta)-\pi iH_0^{(1)}[b(k^2-\zeta^2)^{\frac{1}{2}}]. \quad (7.2)$$

The kernel K_s for the symmetric part of the current differs only in a sign.

Besides a, this problem is characterized by three lengths: λ, h, and b. In the general case, it seems difficult to get easily computable formulas from these kernels. When $kb\ll1$, the $H_0^{(1)}$ in the Fourier transforms of the kernels may be replaced by a logarithm, analogous to (3.1). In particular,

$$\bar{K}_s(\zeta)=2\{2\Omega_{1s}-\ln[(k^2-\zeta^2)/k^2]\}, \quad (7.3)$$

where

$$\Omega_{1s}=\ln[2/(k^2ab)^{\frac{1}{2}}]-\gamma. \quad (7.4)$$

A well-known result follows immediately from (7.3–7.4), namely, that the equivalent radius of this antenna system is $(ab)^{\frac{1}{2}}$ [K, p. 275]. In more general cases, the present point of view reproduces all the results of Harrison and King[8] on effective radii.

In the remainder of this section, the following special situation of the antisymmetrical case is to be considered

$$kb^2/\lambda\ll1. \quad (7.5)$$

Furthermore, attention is to be restricted entirely to the approximate determination of the total power radiated and the capacitive end correction, both topics outside of the realm of conventional transmission-line theory. An important difference between the present case and that of a single dipole antenna is that $\bar{K}_a(\zeta)$ is bounded in the vicinity of $\zeta=-k$ while $\bar{K}(\zeta)$ is not. Consequently,

$$\bar{L}_+(-k)=0, \quad \text{but} \quad \bar{L}_{+a}(-k)\neq0. \quad (7.6)$$

Note that $\bar{L}_{\pm a}(\zeta)$ etc. are defined analogous to $\bar{L}_\pm(\zeta)$ etc. except that $\bar{K}_a(\zeta)$ is used instead of $\bar{K}(\zeta)$. Because of (7.6), (4.4) cannot be generalized to the present case without modification. Instead, define

$$W_\pm'(Z)=\int_{C_0}d\zeta\bar{L}_{+a}(\zeta)\left[\int_0^Z dz\exp(-i\zeta z\pm ikz)+\int_{-\infty}^0 dz\exp(-i\zeta z-ikz)\right]. \quad (7.7)$$

If for large kZ only terms of the orders Z^0 and Z^{-1} are kept, W' may be evaluated approximately to be

$$W_+'(Z)=2\pi\bar{L}_{+a}(k), \quad (7.8)$$

and

$$W_-'(Z)=2\pi\bar{L}_{+a}(-k)\left[1-\frac{ikb^2}{4Z\ln(b/a)}\right]. \quad (7.9)$$

[8] C. W. Harrison and R. W. P. King, Trans. IRE **AP-9**, 171 (1961). See also, C. W. Harrison, Ph.D. dissertation, Harvard University, 1954.

The constant C_a in the integral equation for the current,

$$\int_{-h}^{h} dz' I_a(z') K_a(z-z') = 4\pi i \zeta_0^{-1}[C_a \cos kz + \tfrac{1}{2} \sin k|z|], \quad (7.10)$$

is determined from the boundary condition $I_a(h)=0$. It follows from (4.6) and (7.7) that

$$C_a = \tfrac{1}{2} i [e^{ikh} W_-'(2h) + e^{-ikh} W_+'(2h)]^{-1} \times \{e^{ikh}[2W_-'(h) - W_-'(2h)] - e^{-ikh}[2W_+'(h) - W_+'(2h)]\}. \quad (7.11)$$

In this formula, insofar as $\bar{L}_{+a}(\zeta)$ is concerned, only the ratio

$$\bar{L}_{+a}(-k)/\bar{L}_{+a}(k) = \Gamma \exp(2ikh_c) \quad (7.12)$$

enters. Equation (7.12) defines the real numbers Γ and h_c. The quantity h_c is to be interpreted as the apparent change in line length due to the capacitive end correction. In the language of transmission-line theory, h_c is equal to the apparent terminal capacitance divided by the capacitance per unit length of the infinite line. The substitution of (7.12) into (7.11) gives

$$C_a = -\frac{1}{2} i \left[1 + \left(1 - \frac{i\alpha}{2kh}\right)\Lambda\right]^{-1} \left[1 - \left(1 - \frac{3i\alpha}{2kh}\right)\Lambda\right], \quad (7.13)$$

where

$$\alpha = \tfrac{1}{4} k^2 b^2 / \ln(b/a), \quad (7.14)$$

and

$$\Lambda = \Gamma \exp[2ik(h+h_c)]. \quad (7.15)$$

In terms of C_a, the input admittance is[9]

$$Y_a = 2i\zeta_0^{-1}[S_a(h) + C_a U_a(h)], \quad (7.16)$$

where S_a and U_a are analogous to (4.1) and (4.10), respectively:

$$S_a(Z) = \int_{C_0} d\zeta [\bar{K}_a(\zeta)]^{-1} \int_0^Z dz \sin kz \times \exp(-i\zeta Z), \quad (7.17)$$

and

$$U_a(Z) = \int_{C_0} d\zeta [\bar{K}_a(\zeta)]^{-1} \int_{-Z}^{Z} dz \cos kz \times \exp(-i\zeta z). \quad (7.18)$$

It may be noted that (7.16) is correct only to $(kh)^{-1}$. The reason is that a term of the order $(kh)^{-2}$ must be added to correct for the fact that $A(z)$ does not vanish for $|z|>h$. When $kZ \gg 1$, the function U_a of (7.18) may be evaluated by the procedure used to derive (7.8–7.9):

$$U_a(Z) = \pi[\ln(b/a)]^{-1}[1 - i\alpha/(kZ)]. \quad (7.19)$$

The radiation conductance G^e is just 2 ReY_a, since the driving voltage has been taken to be 1. The factor 2 here comes from the fact that there are two wires. It follows from (7.16) that

$$G^e = -4\zeta_0^{-1}[\mathrm{Im}S_a(h) + \mathrm{Im}C_a U_a(h)]. \quad (7.20)$$

This is fortunate since the imaginary part of S_a is simpler than the real part.

So far, only the simplifying assumption (7.5) has been used. In order to get explicit and useful answers, the further assumption that $kb \ll 1$ is to be made. In this limit, the first few terms for Γ, h_c and Im$S_a(h)$ are found in Appendix B to be

$$\ln\Gamma = \frac{-k^2b^2}{4\ln(b/a)}\left\{1 - \frac{k^2b^2}{6}\left[\frac{1}{4} - \frac{\ln(kb)+\gamma-11/6}{\ln(b/a)}\right]\right\}, \quad (7.21)$$

$$h_c = -\frac{b}{\pi}\int_0^\infty \frac{d\xi}{\xi^2} \ln \frac{I_0(a\xi/b)K_0(a\xi/b) - K_0(\xi)}{\ln(b/a)}, \quad (7.22)$$

and

$$\mathrm{Im}S_a(h) = -\frac{\pi}{2}\frac{k^2b^2}{4[\ln(b/a)]^2} \times \left\{1 - \frac{k^2b^2}{6}\left[\frac{1}{4} - 2\frac{\ln(kb)+\gamma-11/6}{\ln(b/a)}\right]\right\}. \quad (7.23)$$

The substitution of (7.13), (7.19), and (7.23) into (7.20) gives

$$G^e = \frac{2\pi}{\zeta_0 \ln(b/a)}\left\{-\frac{2}{\pi}\ln\frac{b}{a}\,\mathrm{Im}S_a(h) + \frac{1-\Gamma^2-3\alpha\Gamma \sin 2k(h+h_c)/kh}{1+\Gamma^2+\Gamma[2\cos 2k(h+h_c)+\alpha \sin 2k(h+h_c)/kh]}\right\}. \quad (7.24)$$

The term "near resonance" shall be used to refer to the situation where $|C_a|$ in (7.20) is of the order of magnitude $(kb)^{-2}$. Near resonance, the first term in the braces of (7.24) is smaller than the second by a factor of the order $(kb)^{-4}$, and hence may be neglected. As a function of h, the second term shows sharp maxima in the vicinity of $k(h+h_c) \sim (n+\frac{1}{2})\pi$. From the point of view of carrying out experiments on the power radiation from a two-wire line, the widths of these resonances are of interest. Let δ be the total half-power width, i.e., the interval on the h axis where G^e is larger than half of the maximum value of G^e, then

$$k\delta = -\ln\Gamma, \quad (7.25)$$

independent of n. This is accurate to the order $(kh)^{-1}$ but not $(kh)^{-2}$.

In (7.22), h_c is expressed in terms of an integral with the relative error $(kb)^2$. This integral remains to be evaluated numerically. When $\ln(b/a) \gg 1$, a condition almost never fulfilled in practice, an approximate

[9] This S_a has nothing to do with that of reference 3.

evaluation is possible:

$$h_c \sim (b/\pi)[\ln(b/a)]^{-1} \times \int_0^\infty d\xi \xi^{-2}[K_0(\xi) - \ln(2/\xi) + \gamma]. \quad (7.26)$$

This integral can be evaluated by shifting the contour of integration and applying the Weber-Schafheitlin integral:

$$h_c \sim \tfrac{1}{2} b/\ln(b/a). \quad (7.27)$$

This formula is due to King,[10] whose derivation is much simpler.

Equations (7.22) and (7.24) give the required answers on the two-wire line. However, it is desirable to calculate the so-called radiation resistance for comparison with the theory of Storer and King.[11] Attention will be restricted to the case $kb \ll 1$ and $kh \gg 1$. The radiation resistance R^e is defined as G^e divided by the square of a "maximum" current. In the theory of Storer and King, a sinusoidal current is assumed on the two-wire line, and thus this definition is meaningful. From the present point of view, the various current "maxima" are of slightly different size and thus R^e is not precisely defined. Let

$$I_{\max} = 2\zeta_0^{-1}(|C|^2 + \tfrac{1}{4})^{\frac{1}{2}} \pi/\ln(b/a). \quad (7.28)$$

Up to $(kh)^{-1}$, this approximates the maximum currents near the driving point with an error of the order $(kb)^{-2}$ in general, but of the order of $(kb)^{-4}$ near resonance. With the definition

$$R^e = I_{\max}^{-2} G^e, \quad (7.29)$$

R^e is given by

$$R^e = \pi^{-1}\zeta_0 \ln(b/a)[1+\Gamma^2 - \alpha\Gamma \sin 2k(h+h_c)/(kh)]^{-1} \times \{[-(2/\pi)\ln(b/a)\ \mathrm{Im} S_a(h)][1+\Gamma^2 + 2\Gamma \cos 2k(h+h_c) + \alpha\Gamma \sin 2k(h+h_c)/(kh)] + 1 - \Gamma^2 - 3\alpha\Gamma \sin 2k(h+h_c)/(kh)\}. \quad (7.30)$$

The leading term of this is

$$R^e = (4\pi)^{-1}\zeta_0 k^2 b^2 \times \left[2 + \cos 2k(h+h_c) - \frac{3 \sin 2k(h+h_c)}{2kh}\right]. \quad (7.31)$$

Except for the end correction h_c, this is derived by the method of Storer and King in Appendix C. Near resonance, however, the present procedure gives more information. Here, the first term in the braces of (7.30) may be neglected. Since $\Gamma^{-1} - \Gamma \sim -2\ln\Gamma$ and $\Gamma^{-1} + \Gamma \sim 2$, (7.30) reduces near resonance to

$$R^e = \pi^{-1}\zeta_0 \ln(b/a)\left[1 + \alpha \frac{\sin 2k(h+h_c)}{2kh}\right] \times \left[-\ln\Gamma - 3\alpha \frac{\sin 2k(h+h_c)}{2kh}\right]. \quad (7.32)$$

Furthermore, near resonance, the quantity $\sin 2k(h+h_c)$ is of the order of $(kb)^2$, so that the first bracket may be neglected. The substitution of (7.14) and (7.21) into (7.32) gives finally

$$R^e = (4\pi)^{-1}\zeta_0 (kb)^2 \left\{1 - \frac{k^2b^2}{6}\left[\frac{1}{4} - \frac{\ln(kb) + \gamma - 11/6}{\ln(b/a)}\right] - \frac{3 \sin 2k(h+h_c)}{2kh}\right\}. \quad (7.33)$$

Right at the point of maximum power radiation, the last term is negligible with the result

$$R^e = -\pi^{-1}\zeta_0 \ln(b/a) \ln\Gamma = (4\pi)^{-1}\zeta_0 (kb)^2 \times \left\{1 - \frac{k^2b^2}{6}\left[\frac{1}{4} - \frac{\ln(kb) + \gamma - 11/6}{\ln(b/a)}\right]\right\}. \quad (7.34)$$

Note that, from (7.25), (7.34) may be written as

$$R^e = k\delta R_c, \quad (7.35)$$

where R_c is the characteristic resistance of the infinite two-wire line. Equation (7.34) has been previously reported.[12]

It should be emphasized that all results in this section depend on the initial assumption of a rotationally symmetrical current distribution on each dipole. In the language of transmission-line theory, the proximity effect of the two wires has been neglected. This leads to a relative error of the order of $(a/b)^2$ in all the physical quantities studied.

8. The Dielectric-Coated Antenna

A dipole antenna with a thin layer of dielectric material on the outside has many interesting properties. When the antenna is relatively short, its behavior does not differ much from the dipole without the dielectric. When it is relatively long, it behaves more like a transmission line than an antenna. In this section, this dielectric-coated antenna is to be studied only from the point of view of illuminating certain essential points of the procedure used in this paper.

[10] R. W. P. King, *Transmission-Line Theory* (McGraw-Hill Book Company, Inc., New York, 1955), p. 367.

[11] J. E. Storer and R. W. P. King, Proc. Inst. Radio Engrs. 39, 1408 (1951).

[12] R. W. P. King, "Quasi-stationary and nonstationary currents in electric circuits," in S. Flügge, *Encyclopedia of Physics*, Vol. 16 (Springer-Verlag, Berlin, 1958), p. 232. The footnote reference 3 mentioned in this article has never been published.

Although this problem is an extremely interesting one, no quantitative calculations will be made since no systematic experimental data seem to be available for this type of antenna.

The present procedure depends on the solution of a Wiener-Hopf integral equation. Therefore, it is essential that the geometry of the problem be translationally invariant in the direction of the antenna after the removal of the perfectly conducting dipole antenna. Accordingly, in studying the dielectric-coated antenna, the dielectric layer is assumed to extend to infinity in the $\pm z$ directions. When the dielectric tube is sufficiently small in cross section to support no "mode," these extensions may be expected to make no significant difference. The geometry of the assumed model is shown in Fig. 5. To avoid unnecessary complications, the dielectric material is assumed to be nonmagnetic with a dielectric constant $\epsilon=\epsilon_0\epsilon_r>\epsilon_0$.

The problem of setting up an integral equation analogous to (1.2) is not entirely straightforward. It follows from the time-independent Maxwell's equations that, for this geometry, a current distribution

$$\mathbf{J}=\hat{z}\delta(r-a)\exp(-i\zeta z) \tag{8.1}$$

leads to the following electric field at $r=a$:

$$E_z(a,z)=a(i\omega\epsilon_0)^{-1}G(\zeta)\exp(-i\zeta z), \tag{8.2}$$

where

$$G(\zeta)=(k^2-\zeta^2)\{\ln[(k^2-\zeta^2)^{\frac{1}{2}}b/2]+\gamma-\pi i/2\} -\epsilon_r^{-1}(k'^2-\zeta^2)\ln(b/a). \tag{8.3}$$

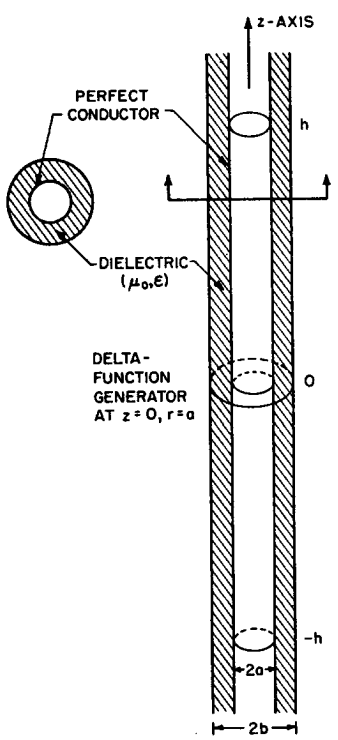


FIG. 5. The dielectric-coated antenna of half-length h.

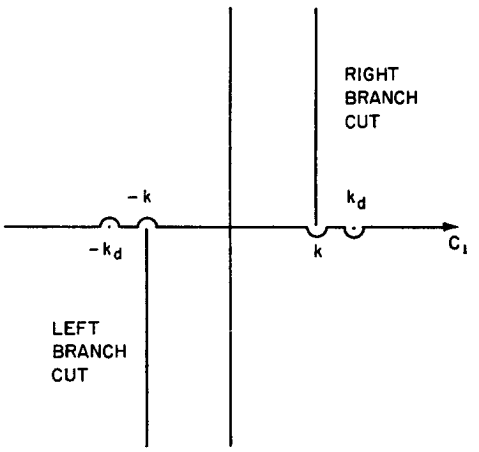


FIG. 6. The contour C_1.

This is approximately valid, in the same sense as (3.1), when

$$k'b\ll 1, \tag{8.4}$$

where

$$k'=\epsilon_r^{\frac{1}{2}}k. \tag{8.5}$$

So far as branch cuts are concerned, $G(\zeta)$ has the same structure as $\bar{K}(\zeta)$. On the real axis, when $\zeta>k$, (8.3) gives

$$G(\zeta)=-(\zeta^2-k^2)\{\ln[(\zeta^2-k^2)^{\frac{1}{2}}b/2]+\gamma\} -\epsilon_r^{-1}(k'^2-\zeta^2)\ln(b/a). \tag{8.6}$$

It follows that on the real axis when $\zeta>k$, $G(\zeta)$ is real with one zero between k and k'. Call this zero $\zeta=k_d$. A more accurate calculation indicates that this zero is actually located slightly above the real axis. Thus, within the present approximation, all integrations along the real axis in the ζ plane should be carried out with the contour C_1 shown in Fig. 6.

Formally, this zero of $G(\zeta)$ at $\zeta=k_d$ leads to

$$\lim_{z\to\infty}\left|\int_{C_1}d\zeta\exp(i\zeta z)[G(\zeta)]^{-1}\right|\neq 0. \tag{8.7}$$

However, for the procedure of Part I to work, the inverse Fourier transform of $\bar{M}$ must approach zero sufficiently rapidly as $z\to\infty$. In order to get the kernel $\bar{K}$, it is thus necessary to remove from $G(\zeta)$ the zero at $\zeta=k_d$. The simplest way to remove this zero is to define

$$\bar{K}_d(\zeta)=2(\zeta^2-k_d^2)^{-1}G(\zeta). \tag{8.8}$$

With this definition, it follows from (8.2) that, in the present case, the integral equation for the current is

$$A_d(z)=(4\pi)^{-1}\mu_0\int_{-h}^{h}dz'I_d(z')K_d(z-z'), \tag{8.9}$$

where K_d is the inverse Fourier transform of $\bar{K}_d$. and

A_d satisfies

$$[(d^2/dz^2)+k_d^2]A_d(z)=i\omega\mu_0\epsilon_0\delta(z), \tag{8.10}$$

for $|z|<h$. By symmetry and for $|z|<h$, $A_d(z)$ is given by,

$$A_d(z)=\mu_0 i\zeta_d^{-1}(C_d\cos k_d z+\tfrac{1}{2}\sin k_d|z|), \tag{8.11}$$

where

$$\zeta_d=(\omega\epsilon_0)^{-1}k_d. \tag{8.12}$$

Equations (8.9) and (8.11) are analogous to (1.2) and (1.4), respectively. But it should be emphasized that A_d is *not* the vector potential for this problem. From this point of view, the use of the vector potential in Part I may be considered to be coincidental.

As before, the constant C_d in (8.11) is to be determined from the boundary condition

$$I_d(h)=0. \tag{8.13}$$

Thus, the present problem is formulated in terms of an integral equation entirely analogous to that of the dipole antenna without dielectric coating, the only difference being in the kernel. In principle, the procedure of Part I may be applied here, but the details are somewhat more complicated. Without going into the details, however, several qualitative statements may be made about this antenna by considering the kernel. First, unless b/a is very large, which is not feasible practically, k_d is quite close to k. Thus, the thin dielectric coating makes only a slight modification on the behavior of the dipole antenna unless the antenna is at least several wavelengths long. On the other hand, when z is large, the behavior of $K_d(z)$ is determined almost entirely by that of $\bar{K}_d(\zeta)$ in the vicinity of the singularities at $\zeta=\pm k$. In this vicinity, $\bar{K}_d(\zeta)$ is qualitatively very similar to $\bar{K}_a(\zeta)$ of the last section. Accordingly, a long dielectric-coated antenna behaves like a transmission line. In particular, $\lim_{h\to\infty}C_d$ in (8.11) does not exist. That is, the end of the antenna has a profound effect on the current distribution near the driving point no matter how long the antenna is. This is typical of a transmission line. This is also in agreement with the experimental observation that any abrupt bend in the antenna causes significant radiation. Moreover, the present point of view makes it possible to make a semi-quantitative statement: a bend causes appreciable radiation unless the radius of curvature of the bend is much larger than $\lambda(k_d/k-1)^{-1}$.

In order to make a quantitative comparison with the approach making use of a surface impedance [K, p. 28], it remains to define the equivalent radius and the equivalent surface reactance. This is done by writing the $G(\zeta)$ of (8.3) in the form

$$G(\zeta)=(k^2-\zeta^2)\{\ln[(k^2-\zeta^2)^{\frac{1}{2}}b/2] -\epsilon_r^{-1}\ln(b/a)+\gamma-\pi i/2\} -\epsilon_r^{-1}(k'^2-k^2)\ln(b/a). \tag{8.14}$$

The second term of (8.14) gives immediately the surface impedance per unit length

$$z^i=-i\omega L, \tag{8.15}$$

where the inductance per unit length L is

$$L=\mu_0(2\pi)^{-1}\ln(b/a)(1-\epsilon_r^{-1}). \tag{8.16}$$

On the other hand, the first term on the right hand of (8.14) gives the equivalent radius a_d in the form

$$a_d=b(a/b)^{1/\epsilon_r}. \tag{8.17}$$

A dipole antenna of a with a dielectric coating of relative dielectric constant ϵ_r and thickness $b-a$ is equivalent to a dipole antenna of radius a_d and surface impedance per unit length $-i\omega L$.

PART III. NUMERICAL RESULTS AND DISCUSSIONS

In the process of getting results that are sufficiently simple to be useful, many approximations have been made. Since a mathematical estimate of the error seems impossibly difficult, several numerical calculations have been carried out to give some idea of the accuracy and the range of validity of the present theory. The following results seem to be particularly useful in this respect.

First, the back-scattering cross section of an unloaded dipole receiving antenna has been obtained for one value of a/λ, using the formulas of Sec. 6. In Fig. 7, the theoretical results are shown together with the

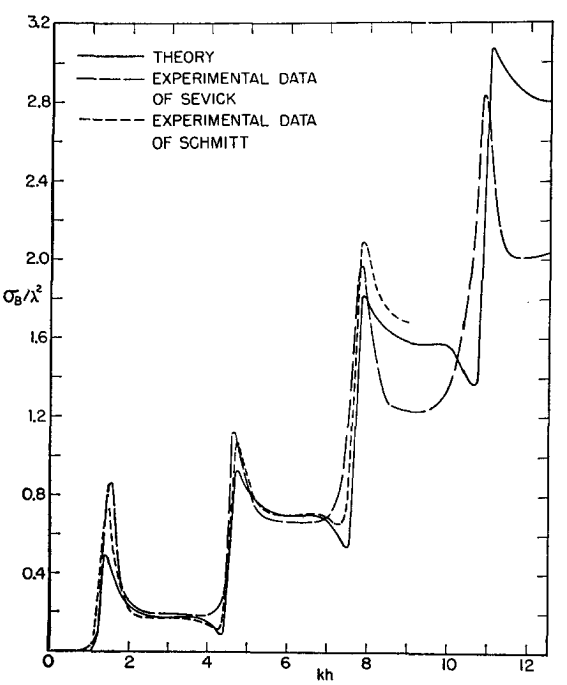


FIG. 7. The back-scattering cross section of an unloaded dipole receiving antenna.

experimental data of Sevick [K, p. 516] and Schmitt.[13] Both the theoretical curve and the experimental curve of Sevick pertain to the parameter $a/\lambda=0.0035$, but that of Schmitt is for $a/\lambda=0.0041$. When the dipole is not too short, the theoretical results are in fair agreement with the experimental data. When there is strong interference between the resonant currents and the nonresonant currents as described by Chu,[7] some discrepancy occurs, both between the theoretical results and the experimental data and between the two sets of experimental data. The large discrepancy with Sevick's data for relatively large values of kh may be traced to an inadequacy of his experimental setup. Since the transmitter was not sufficiently far away from the scattering dipole, the wave front was not planar along the length of the antenna. This effect alone can be estimated to make an error of about 30%

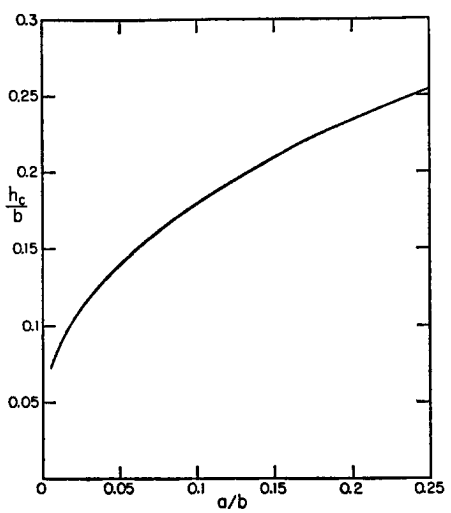


FIG. 8. The end correction for an open-end, two-wire transmission line.

in the back-scattering cross section for $kh=12$. For $kh>2$, the present results are in more satisfactory agreement with the experimental data than any previous theory that yields sufficient numerical results for comparison.

Secondly, the end correction for an open-end two-wire transmission line has been computed from (7.22). The results are shown in Fig. 8. The author has found just one piece of experimental data which was obtained by Tomiyasu[10] for $a/b=0.0794$. The experimental result is larger by a factor of 2. It is not clear how accurate this experimental point is, but the author is of the opinion that it may be in error. Further experiment seems very desirable on this problem. An attempt

[13] R. W. P. King and T. T. Wu, *The Scattering and Diffraction of Waves* (Harvard University Press, Cambridge, Massachusetts, 1959), Fig. 49, p. 161. In this figure the curves for $a=0.132$ and $a=0.026$ should be interchanged.

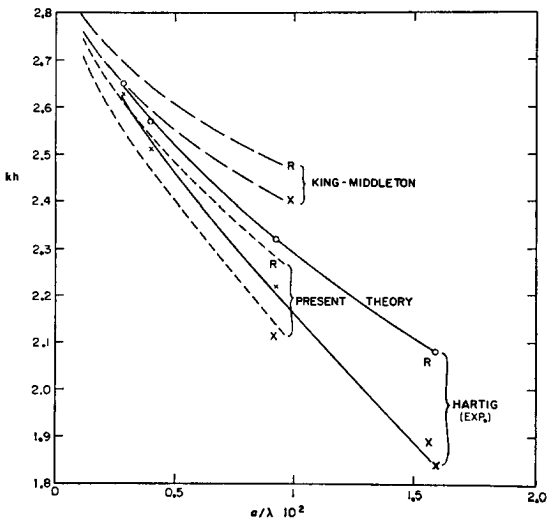


FIG. 9. Positions of antiresonance (X) and maximum resistance (R) as functions of the radius of the antenna.

has also been made to calculate the end correction from the alternative formula

$$h_c'=-\frac{b}{\pi}\int_0^\infty \frac{d\xi}{\xi^2}\ln\frac{I_0(a\xi/b)[K_0(a\xi/b)-K_0(\xi)]}{\ln(b/a)}.$$

The results are changed by no more than 2.5%.

By far the most extensive calculation is that for the input impedance of the center-driven dipole antenna. This was carried out on an IBM 650 on the basis of the results of Part I. The numerical results are shown in Table I. No systematic analysis of these extensive data has been carried out yet. Up to about one wavelength, where comparison is possible, the results are in excellent agreement with the King-Middleton second-order values[3] with one exception. This one exception is that the first antiresonance occurs at a significantly shorter length of the antenna, particularly when the antenna is fat. However, when the results are compared with the experimental data of Hartig [K, pp. 232–238], it is seen that the results of the present theory is closer to the measurement, as shown in Fig. 9. It may also be noted the present result takes the form of an algebraic expression involving only elementary functions, while the second-order formula of King and Middleton involves integrals of integrated sine and cosine functions.

It may thus be concluded that the present theory gives results that are in excellent agreement with what is already known unless the antenna is quite short, and of course gives new results in many cases, particularly when the antenna is long. Furthermore, although so far the discussion has been restricted to nondissipative media, the entire theory may be applied to

TABLE I. Input impedances of the center-driven dipole antenna in ohms.

kh \ a/λ	0.001191	0.001588	0.002381	0.003175
1.0	17.42 + 270.05	17.52 + 247.59	17.83 + 216.24	18.30 + 194.25
1.1	27.73 + 212.18	27.97 + 194.33	28.51 + 169.42	29.15 + 151.96
1.2	38.54 + 156.79	38.98 + 143.10	39.87 + 124.01	40.81 + 110.67
1.3	50.52 + 102.45	51.24 + 92.60	52.62 + 78.90	54.00 + 69.38
1.4	64.35 + 47.97	65.46 + 41.74	67.51 + 33.14	69.51 + 27.25
1.5	80.86 - 7.75	82.49 - 10.48	85.47 - 14.14	88.33 - 16.47
1.6	101.07 - 65.74	103.42 - 65.01	107.71 - 63.76	111.78 - 62.50
1.7	126.38 - 127.09	129.76 - 122.84	135.91 - 116.51	141.74 - 111.48
1.8	158.80 - 192.92	163.65 - 184.94	172.49 - 173.11	180.86 - 163.81
1.9	201.25 - 264.37	208.26 - 252.20	221.04 - 233.96	233.11 - 219.40
2.0	258.17 - 342.40	268.35 - 325.13	286.88 - 298.64	304.28 - 276.80
2.1	336.36 - 427.25	351.19 - 403.09	377.97 - 364.66	402.64 - 331.47
2.2	446.42 - 517.04	467.84 - 482.37	505.58 - 424.58	538.72 - 372.07
2.3	604.44 - 604.15	634.07 - 551.66	683.11 - 459.99	721.04 - 373.84
2.4	832.15 - 666.78	867.89 - 582.41	916.95 - 432.01	940.33 - 293.05
2.5	1147.33 - 652.37	1172.11 - 514.52	1177.53 - 279.37	1138.69 - 84.24
2.6	1520.90 - 464.39	1484.03 - 263.10	1360.89 + 34.15	1209.18 + 231.65
2.7	1798.30 - 25.79	1635.53 + 182.03	1341.45 + 420.41	1100.18 + 530.98
2.8	1763.87 + 533.20	1501.33 + 637.65	1133.39 + 709.06	888.09 + 707.87
2.9	1443.79 + 928.16	1189.24 + 905.29	870.06 + 835.10	674.35 + 763.86
3.0	1065.43 +1071.56	875.86 + 980.11	643.85 + 845.90	503.34 + 748.48
3.1	761.55 +1057.73	634.02 + 948.34	475.59 + 801.18	377.90 + 701.47
3.2	547.34 + 979.35	463.36 + 874.74	356.11 + 736.30	288.14 + 643.91
3.3	401.15 + 883.22	345.25 + 790.44	271.77 + 667.38	223.87 + 585.15
3.4	300.93 + 787.75	262.97 + 707.69	211.71 + 600.71	177.42 + 528.77
3.5	231.14 + 698.83	204.83 + 630.41	168.44 + 538.24	143.54 + 475.80
3.6	181.78 + 617.55	163.21 + 559.26	137.00 + 480.17	118.71 + 426.20
3.7	146.49 + 543.24	133.21 + 493.68	114.15 + 425.99	100.62 + 379.53
3.8	121.26 + 474.78	111.69 + 432.74	97.75 + 375.03	87.74 + 335.21
3.9	103.50 + 410.95	96.59 + 375.46	86.43 + 326.55	79.07 + 292.66
4.0	91.56 + 350.64	86.62 + 320.94	79.30 + 279.85	73.98 + 251.31
4.1	84.39 + 292.88	80.95 + 268.33	75.85 + 234.31	72.17 + 210.64
4.2	81.39 + 236.75	79.13 + 216.89	75.83 + 189.33	73.51 + 170.16
4.3	82.28 + 181.48	81.00 + 165.91	79.23 + 144.35	78.12 + 129.39
4.4	87.05 + 126.29	86.65 + 114.76	86.30 + 98.84	86.35 + 87.90
4.5	95.97 + 70.50	96.43 + 62.80	97.50 + 52.31	98.75 + 45.28
4.6	109.59 + 13.41	110.97 + 9.45	113.61 + 4.31	116.25 + 1.19
4.7	128.82 - 45.65	131.29 - 45.85	135.82 - 45.54	140.13 - 44.56
4.8	155.04 - 107.27	158.87 - 103.56	165.80 - 97.42	172.26 - 91.92
4.9	190.27 - 171.94	195.89 - 163.91	205.95 - 151.16	215.22 - 140.26
5.0	237.43 - 239.80	245.43 - 226.70	259.61 - 205.80	272.45 - 187.89
5.1	300.71 - 310.27	311.83 - 290.69	331.23 - 258.88	348.29 - 231.13
5.2	386.00 - 381.20	400.97 - 352.64	426.28 - 305.09	447.28 - 262.70
5.3	501.14 - 447.21	520.13 - 405.33	550.08 - 334.06	571.87 - 269.82
5.4	655.31 - 496.61	676.24 - 434.40	704.00 - 327.83	716.90 - 233.36
5.5	855.11 - 506.69	870.01 - 414.74	876.86 - 261.22	861.00 - 133.27
5.6	1093.16 - 440.40	1082.75 - 311.95	1034.03 - 112.87	963.60 + 34.34
5.7	1327.40 - 256.59	1261.35 - 101.87	1120.74 + 108.51	984.22 + 238.47
5.8	1472.73 + 48.56	1331.34 + 189.51	1098.41 + 348.53	916.59 + 424.02
5.9	1453.39 + 397.84	1257.73 + 476.81	981.90 + 540.32	793.70 + 551.27
6.0	1283.47 + 676.88	1082.72 + 679.86	822.61 + 653.24	656.86 + 615.12
6.1	1049.00 + 831.49	878.82 + 779.91	665.09 + 696.54	531.91 + 631.15
6.2	823.91 + 880.12	693.81 + 802.22	530.33 + 693.71	428.06 + 617.61
6.3	638.67 + 863.74	543.73 + 778.82	422.49 + 665.03	345.39 + 587.98
6.4	495.92 + 814.60	427.68 + 732.56	338.61 + 623.58	280.73 + 550.48
6.5	388.52 + 751.61	339.47 + 676.59	273.96 + 576.73	230.43 + 509.72
6.6	308.09 + 684.41	272.62 + 617.65	224.19 + 528.28	191.33 + 468.03
6.7	247.71 + 617.45	221.87 + 558.92	185.87 + 480.04	160.95 + 426.55
6.8	202.24 + 552.54	183.29 + 501.70	156.41 + 432.74	137.48 + 385.70
6.9	168.02 + 490.18	154.07 + 446.35	133.97 + 386.55	119.61 + 345.54

TABLE I.—*Continued.*

kh \ a/λ	0.001191	0.001588	0.002381	0.003175
7.0	142.50 + 430.27	132.23 + 392.77	117.25 + 341.37	106.40 + 305.95
7.1	123.94 + 372.41	116.43 + 340.64	105.35 + 296.92	97.28 + 266.70
7.2	111.15 + 316.09	105.75 + 289.53	97.74 + 252.89	91.88 + 227.50
7.3	103.42 + 260.76	99.66 + 238.99	94.11 + 208.90	90.09 + 188.05
7.4	100.33 + 205.87	97.92 + 188.53	94.42 + 164.58	91.98 + 148.03
7.5	101.79 + 150.86	100.54 + 137.68	98.85 + 119.56	97.85 + 107.15
7.6	108.00 + 95.17	107.81 + 85.98	107.84 + 73.49	108.26 + 65.13
7.7	119.42 + 38.30	120.30 + 32.99	122.12 + 26.07	124.05 + 21.81
7.8	136.89 − 20.26	138.95 − 21.65	142.78 − 22.87	146.45 − 22.84
7.9	161.70 − 80.88	165.16 − 78.19	171.38 − 73.30	177.14 − 68.49
8.0	195.69 − 143.77	200.88 − 136.57	210.08 − 124.73	218.40 − 114.24
8.1	241.48 − 208.70	248.86 − 196.21	261.73 − 175.85	273.09 − 158.09
8.2	302.70 − 274.64	312.76 − 255.45	329.91 − 223.90	344.45 − 196.19
8.3	384.14 − 338.98	397.22 − 310.72	418.54 − 263.54	435.29 − 221.74
8.4	491.73 − 396.19	507.33 − 355.02	530.59 − 285.52	545.86 − 224.02
8.5	631.42 − 435.63	646.84 − 375.93	664.82 − 275.51	669.76 − 188.86
8.6	805.70 − 438.93	813.41 − 354.13	810.14 − 215.84	789.28 − 103.60
8.7	1005.64 − 379.49	990.34 − 266.27	940.20 − 93.94	875.68 + 31.96
8.8	1199.26 − 231.10	1139.64 − 98.42	1017.90 + 82.41	901.94 + 196.16
8.9	1329.62 + 7.56	1211.74 + 131.97	1016.54 + 277.77	861.83 + 352.44
9.0	1344.08 + 289.48	1179.07 + 370.99	940.89 + 446.77	773.84 + 471.16
9.1	1239.71 + 538.90	1060.08 + 561.04	821.55 + 562.08	665.67 + 542.94
9.2	1064.73 + 705.30	900.73 + 676.87	691.35 + 622.15	558.69 + 574.23
9.3	874.97 + 785.26	740.69 + 725.97	570.79 + 639.69	463.68 + 576.76
9.4	703.82 + 801.38	600.21 + 728.57	468.01 + 629.41	383.86 + 561.12
9.5	562.51 + 778.55	484.77 + 703.24	383.89 + 602.70	318.60 + 534.84
9.6	450.49 + 734.80	392.78 + 662.80	316.35 + 567.03	265.92 + 502.71
9.7	363.18 + 681.04	320.39 + 614.98	262.57 + 526.85	223.63 + 467.56
9.8	295.48 + 623.28	263.70 + 564.08	219.88 + 484.65	189.82 + 430.99
9.9	243.05 + 564.61	219.38 + 512.33	186.14 + 441.76	162.92 + 393.83
10.0	202.55 + 506.46	184.88 + 460.79	159.67 + 398.79	141.77 + 356.45
10.1	171.47 + 449.35	158.29 + 409.85	139.22 + 355.94	125.51 + 318.96
10.2	148.03 + 393.35	138.26 + 359.55	123.94 + 313.22	113.53 + 281.31
10.3	130.99 + 338.26	123.83 + 309.71	113.24 + 270.46	105.50 + 243.35
10.4	119.52 + 283.74	114.40 + 260.06	106.81 + 227.43	101.26 + 204.89
10.5	113.16 + 229.40	109.68 + 210.26	104.55 + 183.88	100.86 + 165.69
10.6	111.71 + 174.80	109.61 + 159.94	106.61 + 139.53	104.57 + 125.55
10.7	115.30 + 119.52	114.42 + 108.75	113.36 + 94.10	112.89 + 84.27
10.8	124.32 + 63.13	124.61 + 56.34	125.47 + 47.40	126.59 + 41.74
10.9	139.50 + 5.24	141.03 + 2.45	143.93 − .68	146.78 − 1.97
11.0	162.00 − 54.43	164.90 − 53.06	170.14 − 49.99	174.99 − 46.46
11.1	193.46 − 115.93	198.01 − 110.03	206.03 − 99.98	213.22 − 90.80
11.2	236.21 − 178.94	242.76 − 167.77	254.06 − 149.32	263.91 − 133.04
11.3	293.38 − 242.35	302.28 − 224.63	317.24 − 195.37	329.72 − 169.64
11.4	369.00 − 303.62	380.40 − 277.26	398.70 − 233.32	412.75 − 194.60
11.5	467.84 − 357.76	481.14 − 319.45	500.54 − 255.21	512.81 − 198.92
11.6	594.44 − 395.64	607.12 − 340.76	621.31 − 249.27	624.36 − 171.11
11.7	750.29 − 402.31	756.01 − 325.63	751.89 − 201.46	733.23 − 101.08
11.8	928.13 − 356.87	914.67 − 255.74	871.82 − 101.57	816.95 + 12.15
11.9	1103.98 − 238.61	1054.39 − 119.27	952.19 + 46.18	853.24 + 153.51
12.0	1235.17 − 43.28	1136.58 + 73.96	969.58 + 216.99	833.65 + 295.32
12.1	1277.73 + 199.46	1134.46 + 286.41	921.96 + 375.34	768.63 + 411.78
12.2	1218.22 + 433.03	1053.24 + 471.14	828.96 + 494.73	678.96 + 490.72
12.3	1084.26 + 609.08	924.33 + 599.33	716.97 + 567.56	583.61 + 533.42
12.4	920.12 + 712.14	782.01 + 668.32	606.03 + 600.19	494.31 + 547.77
12.5	759.80 + 753.03	648.56 + 690.64	506.59 + 603.71	416.15 + 542.52
12.6	619.73 + 751.00	533.48 + 681.63	422.04 + 588.28	350.15 + 524.68
12.7	504.05 + 722.87	438.40 + 653.60	352.09 + 561.31	295.46 + 499.16
12.8	411.06 + 680.26	361.46 + 614.82	295.00 + 527.68	250.58 + 469.09
12.9	337.21 + 630.26	299.82 + 570.39	248.75 + 490.36	213.97 + 436.41

TABLE I.—*Continued.*

kh \ a/λ	0.001191	0.001588	0.002381	0.003175
13.0	278.87 + 576.93	250.68 + 523.25	211.46 + 451.08	184.28 + 402.23
13.1	232.96 + 522.43	211.70 + 474.97	181.62 + 410.81	160.43 + 367.16
13.2	197.04 + 467.83	181.03 + 426.36	158.04 + 370.00	141.60 + 331.47
13.3	169.32 + 413.54	157.32 + 377.73	139.85 + 328.82	127.21 + 295.24
13.4	148.48 + 359.64	139.56 + 329.11	126.45 + 287.26	116.88 + 258.44
13.5	133.62 + 305.96	127.12 + 280.37	117.50 + 245.20	110.45 + 220.95
13.6	124.21 + 252.26	119.64 + 231.29	112.87 + 202.46	107.94 + 182.60
13.7	119.97 + 198.20	116.99 + 181.60	112.64 + 158.83	109.55 + 143.23
13.8	120.94 + 143.42	119.34 + 131.01	117.13 + 114.10	115.74 + 102.70
13.9	127.44 + 87.58	127.11 + 79.24	126.94 + 68.12	127.20 + 60.93
14.0	140.10 + 30.37	141.03 + 26.07	142.94 + 20.83	144.94 + 18.03
14.1	159.92 − 28.44	162.22 − 28.56	166.41 − 27.61	170.33 − 25.63
14.2	188.37 − 88.87	192.24 − 84.48	199.04 − 76.65	205.13 − 69.16
14.3	227.51 − 150.56	233.21 − 141.02	243.02 − 125.06	251.52 − 110.79
14.4	280.07 − 212.47	287.89 − 196.67	300.97 − 170.46	311.81 − 147.37
14.5	349.52 − 272.32	359.56 − 248.42	375.58 − 208.65	387.82 − 173.71
14.6	439.94 − 325.76	451.63 − 290.94	468.66 − 232.79	479.48 − 182.08
14.7	555.13 − 365.06	566.37 − 315.40	579.17 − 232.95	582.43 − 162.68
14.8	696.54 − 377.82	702.14 − 308.83	700.01 − 197.15	685.38 − 106.47
14.9	858.93 − 346.79	849.00 − 255.77	815.03 − 115.90	769.92 − 11.05
15.0	1024.28 − 253.97	984.63 − 144.74	900.38 + 9.85	816.27 + 113.43
15.1	1159.16 − 92.14	1077.20 + 20.29	934.11 + 162.67	813.83 + 245.19
15.2	1224.92 + 120.76	1100.04 + 212.77	909.25 + 313.58	767.39 + 360.84
15.3	1201.37 + 341.48	1049.41 + 393.35	837.68 + 436.74	692.70 + 446.16
15.4	1101.68 + 524.47	946.03 + 531.41	740.89 + 520.32	606.91 + 498.62
15.5	960.38 + 646.39	819.37 + 617.05	638.23 + 565.83	522.33 + 523.12
15.6	810.60 + 708.70	692.60 + 656.52	541.77 + 581.37	445.50 + 526.86
15.7	672.42 + 725.36	578.22 + 662.10	456.89 + 575.88	378.80 + 516.38
15.8	553.88 + 711.66	480.59 + 645.36	384.83 + 556.69	322.31 + 496.66
15.9	455.95 + 679.52	399.66 + 614.94	324.81 + 528.96	275.11 + 471.10
16.0	376.55 + 636.89	333.57 + 576.55	275.35 + 496.07	236.00 + 441.90
16.1	312.78 + 588.68	280.03 + 533.74	234.87 + 460.12	203.81 + 410.41
16.2	261.84 + 537.73	236.93 + 488.57	201.97 + 422.35	177.53 + 377.44
16.3	221.40 + 485.58	202.49 + 442.21	175.52 + 383.46	156.38 + 343.43
16.4	189.64 + 433.02	175.33 + 395.22	154.65 + 343.77	139.78 + 308.56
16.5	165.19 + 380.33	154.46 + 347.83	138.75 + 303.40	127.34 + 272.89
16.6	147.09 + 327.56	139.15 + 300.05	127.43 + 262.32	118.86 + 236.36
16.7	134.72 + 274.56	129.00 + 251.75	120.53 + 220.42	114.32 + 198.87
16.8	127.73 + 221.09	123.83 + 202.74	118.06 + 177.55	113.90 + 160.29
16.9	126.09 + 166.88	123.72 + 152.79	120.31 + 133.55	117.97 + 120.52
17.0	130.02 + 111.64	129.01 + 101.67	127.78 + 88.29	127.17 + 79.50
17.1	140.04 + 55.08	140.35 + 49.21	141.24 + 41.74	142.38 + 37.32
17.2	157.04 − 2.96	158.71 − 4.65	161.84 − 5.96	164.85 − 5.67
17.3	182.31 − 62.50	185.49 − 59.73	191.10 − 54.30	196.17 − 48.68
17.4	217.66 − 123.20	222.56 − 115.44	230.98 − 102.19	238.28 − 90.14
17.5	265.51 − 184.16	272.36 − 170.44	283.79 − 147.56	293.27 − 127.32
17.6	328.91 − 243.38	337.79 − 222.13	351.98 − 186.76	362.88 − 155.72
17.7	411.43 − 297.14	421.90 − 265.93	437.28 − 213.92	447.32 − 168.64
17.8	516.50 − 338.92	526.86 − 294.35	539.22 − 220.50	543.37 − 157.50
17.9	645.77 − 358.29	651.87 − 296.39	652.47 − 195.97	642.12 − 113.81
18.0	795.81 − 340.64	789.59 − 258.52	764.25 − 131.06	728.22 − 33.80
18.1	953.22 − 269.77	922.60 − 169.29	854.49 − 24.08	783.71 + 76.29
18.2	1091.40 − 136.21	1024.11 − 28.25	902.40 + 113.23	796.35 + 199.32
18.3	1176.16 + 50.71	1068.00 + 146.44	897.38 + 257.21	766.40 + 314.06
18.4	1183.22 + 258.41	1044.10 + 322.03	845.09 + 383.08	705.36 + 404.87
18.5	1114.59 + 445.27	964.28 + 467.56	762.54 + 475.89	628.53 + 466.07
18.6	995.66 + 582.77	853.05 + 567.60	668.02 + 533.03	548.54 + 499.87
18.7	857.35 + 664.44	733.53 + 622.81	574.77 + 559.72	473.13 + 511.97
18.8	722.23 + 698.85	620.58 + 642.24	489.89 + 563.65	405.88 + 508.43
18.9	601.77 + 699.29	521.02 + 636.48	416.03 + 551.90	347.77 + 494.22

TABLE I.—*Continued.*

kh \ a/λ	0.001191	0.001588	0.002381	0.003175
19.0	499.54 + 677.59	436.55 + 614.28	353.35 + 529.87	298.44 + 472.94
19.1	415.02 + 642.42	366.35 + 581.91	300.92 + 501.29	257.01 + 447.05
19.2	346.10 + 599.43	308.66 + 543.43	257.45 + 468.59	222.48 + 418.15
19.3	290.34 + 552.12	261.63 + 501.40	221.67 + 433.29	193.93 + 387.21
19.4	245.52 + 502.52	223.56 + 457.32	192.50 + 396.29	170.60 + 354.83
19.5	209.83 + 451.76	193.10 + 412.02	169.08 + 358.10	151.91 + 321.29
19.6	181.86 + 400.38	169.20 + 365.91	150.78 + 318.93	137.46 + 286.74
19.7	160.58 + 348.57	151.11 + 319.13	137.18 + 278.85	127.02 + 251.19
19.8	145.29 + 296.34	138.34 + 271.66	128.07 + 237.83	120.53 + 214.58
19.9	135.57 + 243.52	130.67 + 223.37	123.42 + 195.75	118.13 + 176.81
20.0	131.31 + 189.92	128.12 + 174.10	123.45 + 152.50	120.13 + 137.80
20.1	132.64 + 135.29	130.94 + 123.66	128.58 + 107.96	127.08 + 97.52
20.2	140.00 + 79.39	139.67 + 71.89	139.51 + 62.12	139.78 + 56.03
20.3	154.13 + 22.07	155.17 + 18.76	157.24 + 15.10	159.35 + 13.64
20.4	176.17 − 36.69	178.67 − 35.61	183.15 − 32.66	187.23 − 28.97
20.5	207.73 − 96.60	211.86 − 90.67	218.99 − 80.23	225.19 − 70.44
20.6	250.95 − 156.89	256.90 − 145.30	266.86 − 125.80	275.15 − 108.42
20.7	308.54 − 215.86	316.40 − 197.25	329.02 − 166.25	338.82 − 139.02
20.8	383.70 − 270.34	393.15 − 242.64	407.24 − 196.54	416.74 − 156.42
20.9	479.61 − 314.84	489.34 − 275.10	501.55 − 209.27	506.70 − 152.99
21.0	598.18 − 340.55	604.89 − 285.18	608.15 − 195.04	601.76 − 120.67
21.1	737.48 − 334.91	734.62 − 260.88	717.01 − 144.77	689.15 − 54.58
21.2	887.71 − 283.22	864.99 − 190.86	811.23 − 54.69	752.54 + 42.23
21.3	1027.80 − 174.54	973.50 − 71.39	871.27 + 68.05	778.63 + 156.65
21.4	1127.77 − 11.74	1035.17 + 86.09	884.00 + 204.57	763.95 + 269.72
21.5	1161.28 + 181.72	1035.26 + 255.05	849.85 + 331.70	716.08 + 364.95
21.6	1121.66 + 369.13	978.10 + 405.51	781.42 + 432.29	648.28 + 434.03
21.7	1025.50 + 519.10	882.85 + 517.95	695.50 + 500.08	573.18 + 476.59
21.8	900.21 + 618.37	771.68 + 587.82	606.05 + 537.42	499.52 + 496.78
21.9	769.69 + 669.99	661.17 + 620.78	521.66 + 550.59	432.00 + 500.03
22.0	648.43 + 684.65	560.41 + 626.03	446.37 + 546.26	372.48 + 491.25
22.1	542.61 + 673.73	472.87 + 612.23	381.30 + 529.96	321.18 + 474.22
22.2	453.37 + 646.35	398.84 + 586.05	326.08 + 505.72	277.56 + 451.63
22.3	379.52 + 608.88	337.20 + 552.08	279.75 + 476.32	240.79 + 425.31
22.4	319.06 + 565.44	286.36 + 513.33	241.20 + 443.53	210.07 + 396.40
22.5	269.94 + 518.57	244.76 + 471.64	209.40 + 408.47	184.64 + 365.65
22.6	230.36 + 469.74	211.05 + 428.12	183.51 + 371.79	163.95 + 333.46
22.7	198.89 + 419.76	184.17 + 383.36	162.88 + 333.86	147.56 + 300.04
22.8	174.44 + 369.01	163.32 + 337.64	147.05 + 294.83	135.23 + 265.48
22.9	156.23 + 317.61	147.97 + 291.05	135.78 + 254.71	126.85 + 229.76
23.0	143.80 + 265.51	137.83 + 243.54	128.99 + 213.45	122.52 + 192.82
23.1	136.92 + 212.55	132.84 + 194.98	126.84 + 170.96	122.49 + 154.59
23.2	135.65 + 158.55	133.19 + 145.23	129.67 + 127.17	127.25 + 115.04
23.3	140.33 + 103.31	139.33 + 94.15	138.10 + 82.03	137.50 + 74.24
23.4	151.60 + 46.68	151.99 + 41.71	153.02 + 35.67	154.24 + 32.43
23.5	170.44 − 11.36	172.28 − 11.98	175.65 − 11.56	178.79 − 9.81
23.6	198.28 − 70.59	201.68 − 66.50	207.58 − 58.86	212.76 − 51.33
23.7	237.02 − 130.34	242.14 − 120.86	250.75 − 104.70	257.95 − 90.10
23.8	289.08 − 189.21	296.02 − 173.18	307.24 − 146.41	316.06 − 122.80
23.9	357.37 − 244.51	365.92 − 220.17	378.87 − 179.65	387.90 − 144.35
24.0	444.88 − 291.63	454.05 − 256.39	466.12 − 198.01	472.14 − 147.95
24.1	553.73 − 323.09	560.94 − 273.71	566.48 − 193.06	563.51 − 125.96
24.2	683.19 − 328.07	683.16 − 261.47	672.19 − 156.00	651.48 − 72.69
24.3	826.32 − 293.21	810.34 − 208.58	769.14 − 81.51	721.40 + 11.41
24.4	966.65 − 206.81	923.79 − 109.00	839.32 + 27.01	759.42 + 116.85
24.5	1078.29 − 66.94	1000.21 + 31.31	867.97 + 155.11	759.09 + 227.12
24.6	1134.52 + 110.79	1021.92 + 191.73	851.13 + 281.78	724.26 + 325.55
24.7	1122.14 + 295.34	986.77 + 344.43	797.07 + 388.59	665.91 + 401.62
24.8	1049.32 + 454.55	908.36 + 467.17	720.53 + 466.08	596.27 + 452.48
24.9	938.86 + 569.58	806.93 + 550.68	635.70 + 513.71	524.86 + 480.62

TABLE I.—*Continued.*

kh \ a/λ	0.001191	0.001588	0.002381	0.003175
25.0	814.80 + 637.90	700.11 + 596.94	552.45 + 596.08	457.43 + 490.67
25.1	694.09 + 667.01	599.02 + 613.39	476.17 + 539.29	396.75 + 487.39
25.2	585.49 + 667.37	508.96 + 608.33	408.99 + 528.88	343.65 + 474.69
25.3	491.98 + 648.29	431.44 + 588.67	351.17 + 509.15	297.95 + 455.48
25.4	413.43 + 616.76	366.01 + 559.48	302.09 + 483.16	259.05 + 431.79
25.5	348.36 + 577.52	311.45 + 524.21	260.85 + 452.98	226.21 + 404.96
25.6	294.96 + 533.62	266.35 + 485.08	226.48 + 419.93	198.75 + 375.87
25.7	251.50 + 486.90	229.42 + 443.46	198.17 + 384.85	176.10 + 345.05
25.8	216.53 + 438.46	199.58 + 400.15	175.24 + 348.21	157.83 + 312.80
25.9	188.90 + 388.87	175.99 + 355.58	157.23 + 310.25	143.65 + 279.25
26.0	167.77 + 338.38	158.08 + 309.93	143.84 + 271.07	133.44 + 244.44
26.1	152.60 + 287.04	145.48 + 263.24	134.95 + 230.66	127.23 + 208.34
26.2	143.09 + 234.77	138.07 + 215.42	130.65 + 188.97	125.23 + 170.91
26.3	139.23 + 181.43	135.98 + 166.38	131.23 + 145.93	127.85 + 132.12
26.4	141.25 + 126.85	139.55 + 116.01	137.21 + 101.51	135.72 + 92.01
26.5	149.68 + 70.90	149.43 + 64.26	149.39 + 55.80	149.74 + 50.78
26.6	165.39 + 13.55	166.57 + 11.21	168.85 + 9.10	171.09 + 8.92
26.7	189.61 − 45.04	192.30 − 42.78	197.04 − 37.93	201.26 − 32.60
26.8	224.04 − 104.32	228.38 − 96.91	235.72 − 83.99	241.93 − 72.08
26.9	270.88 − 169.12	276.96 − 149.58	286.87 − 126.83	294.77 − 106.64
27.0	332.74 − 219.20	340.45 − 198.01	352.31 − 162.73	360.85 − 131.91
27.1	412.47 − 268.66	421.08 − 237.59	432.91 − 186.10	439.55 − 141.80
27.2	512.34 − 305.19	519.87 − 261.32	527.17 − 189.43	527.01 − 129.15
27.3	632.53 − 319.44	634.76 − 259.70	629.28 − 164.29	614.71 − 87.75
27.4	768.42 − 299.35	758.04 − 222.15	727.64 − 104.31	689.71 − 15.98
27.5	907.31 − 232.99	874.39 − 141.03	805.99 − 9.84	738.15 + 79.92
27.6	1027.36 − 115.02	962.69 − 17.96	848.76 + 108.76	751.26 + 186.11
27.7	1102.72 + 45.62	1003.66 + 132.04	848.44 + 233.14	729.45 + 286.35
27.8	1115.61 + 224.02	989.83 + 284.23	809.06 + 344.46	681.11 + 368.42
27.9	1066.55 + 389.04	929.10 + 414.97	742.82 + 430.59	617.66 + 427.11
28.0	972.78 + 517.71	838.95 + 510.92	663.60 + 488.11	549.14 + 463.07
28.1	857.24 + 602.05	737.26 + 570.22	582.30 + 519.70	482.28 + 480.01
28.2	738.65 + 645.84	636.90 + 598.12	505.57 + 530.65	420.75 + 482.39
28.3	628.27 + 658.05	544.97 + 602.20	436.60 + 526.38	366.04 + 474.17
28.4	531.03 + 647.88	464.31 + 589.49	376.37 + 511.38	318.38 + 458.46
28.5	448.03 + 622.82	395.28 + 565.44	324.67 + 488.99	277.41 + 437.50
28.6	378.46 + 588.19	337.09 + 533.94	280.79 + 461.56	242.51 + 412.83
28.7	320.81 + 547.55	288.54 + 497.57	243.90 + 430.65	213.06 + 385.47
28.8	273.45 + 503.18	248.40 + 458.01	213.19 + 397.24	188.50 + 356.07
28.9	234.94 + 456.44	215.60 + 416.27	188.01 + 361.96	168.38 + 325.01
29.0	204.09 + 408.13	189.27 + 372.94	167.84 + 325.13	152.39 + 292.50
29.1	180.02 + 358.64	168.78 + 328.30	152.35 + 286.93	140.39 + 258.63
29.2	162.10 + 308.12	153.73 + 282.46	141.39 + 247.41	132.35 + 223.39
29.3	149.97 + 256.58	143.93 + 235.43	134.99 + 206.53	128.43 + 186.78
29.4	143.51 + 203.93	139.41 + 187.13	133.36 + 164.26	128.99 + 148.75
29.5	142.90 + 150.02	140.46 + 137.48	136.97 + 120.58	134.57 + 109.35
29.6	148.54 + 94.76	147.60 + 86.42	146.48 + 75.54	145.97 + 68.74
29.7	161.17 + 38.08	161.69 + 34.01	162.90 + 29.38	164.26 + 27.30
29.8	181.89 − 19.88	183.90 − 19.45	187.52 − 17.34	190.80 − 14.16
29.9	212.21 − 78.70	215.81 − 73.32	221.95 − 63.54	227.19 − 54.18
30.0	254.09 − 137.41	259.38 − 126.26	268.05 − 107.32	275.05 − 90.34

TABLE I.—*Continued.*

kh \ a/λ	0.003969	0.004763	0.006350	0.009525
1.0	18.90 + 177.42	19.63 + 163.85	21.42 + 142.92	26.01 + 114.82
1.1	29.91 + 138.62	30.77 + 127.91	32.79 + 111.46	37.79 + 89.90
1.2	41.84 + 100.51	42.96 + 92.40	45.45 + 80.11	51.39 + 64.79
1.3	55.43 + 62.20	56.94 + 56.53	60.20 + 48.21	67.71 + 39.07
1.4	71.52 + 22.92	73.60 + 19.63	78.00 + 15.21	87.90 + 12.56
1.5	91.17 − 17.99	94.07 − 18.90	100.13 − 19.28	113.53 − 14.67
1.6	115.82 − 61.12	119.90 − 59.54	128.37 − 55.49	146.72 − 42.01
1.7	147.49 − 106.95	153.28 − 102.55	165.22 − 93.20	190.29 − 67.77
1.8	189.10 − 155.58	197.37 − 147.69	214.19 − 131.30	247.62 − 88.23
1.9	244.93 − 206.31	256.69 − 193.60	280.08 − 166.75	321.57 − 96.02
2.0	321.10 − 256.55	337.51 − 236.35	368.59 − 192.59	411.18 − 78.62
2.1	425.80 − 299.51	447.40 − 266.76	483.93 − 194.78	504.76 − 20.25
2.2	567.63 − 319.90	591.74 − 265.87	620.96 − 149.92	574.17 + 86.23
2.3	747.80 − 288.15	762.33 − 201.87	750.96 − 32.91	586.73 + 220.71
2.4	939.73 − 162.76	916.98 − 42.99	818.38 + 152.83	536.28 + 340.42
2.5	1068.92 + 74.11	980.02 + 197.19	782.92 + 350.21	450.14 + 415.59
2.6	1054.60 + 358.00	909.91 + 433.83	668.02 + 491.19	360.20 + 446.14
2.7	907.28 + 576.00	753.50 + 585.82	531.05 + 556.61	283.17 + 446.66
2.8	714.19 + 680.83	585.42 + 644.47	409.88 + 566.84	222.72 + 430.98
2.9	541.29 + 699.50	444.90 + 642.52	315.20 + 547.23	176.92 + 407.92
3.0	407.97 + 672.64	338.66 + 610.87	244.67 + 514.48	142.57 + 382.20
3.1	310.59 + 626.85	261.05 + 567.62	192.81 + 477.29	116.80 + 356.07
3.2	240.37 + 575.45	204.66 + 521.48	154.60 + 439.68	97.41 + 330.52
3.3	189.52 + 524.20	163.45 + 476.14	126.28 + 403.27	82.83 + 305.87
3.4	152.37 + 475.24	133.10 + 432.91	105.20 + 368.55	71.93 + 282.14
3.5	125.05 + 429.11	110.65 + 392.05	89.55 + 335.50	63.97 + 259.21
3.6	104.95 + 385.65	94.12 + 353.36	78.09 + 303.90	58.43 + 236.87
3.7	90.34 + 344.46	82.18 + 316.45	70.01 + 273.42	55.01 + 214.91
3.8	80.06 + 305.05	73.94 + 280.90	64.79 + 243.72	53.55 + 193.10
3.9	73.39 + 266.93	68.87 + 246.29	62.13 + 214.47	54.01 + 171.24
4.0	69.90 + 229.61	66.66 + 212.19	61.90 + 185.35	56.49 + 149.11
4.1	69.37 + 192.64	67.21 + 178.21	64.17 + 156.04	61.21 + 126.53
4.2	71.83 + 155.61	70.61 + 143.97	69.13 + 126.25	68.56 + 103.35
4.3	77.46 + 118.10	77.13 + 109.14	77.17 + 95.73	79.10 + 79.50
4.4	86.69 + 79.75	87.26 + 73.40	88.92 + 64.27	93.63 + 55.03
4.5	100.18 + 40.23	101.75 + 36.50	105.26 + 31.80	113.23 + 30.27
4.6	118.93 − .69	121.69 − 1.68	127.44 − 1.58	139.32 + 5.98
4.7	144.38 − 43.04	148.63 − 41.00	157.16 − 35.32	173.60 − 16.24
4.8	178.52 − 86.49	184.65 − 80.83	196.64 − 68.12	217.75 − 33.49
4.9	224.04 − 129.95	232.54 − 119.59	248.53 − 97.33	272.67 − 40.89
5.0	284.41 − 170.97	295.60 − 154.03	315.33 − 118.01	336.66 − 31.48
5.1	363.55 − 204.60	377.07 − 177.98	397.90 − 122.01	402.87 + 2.32
5.2	464.57 − 221.88	478.07 − 181.09	492.14 − 97.91	457.91 + 63.96
5.3	586.27 − 208.49	593.30 − 148.76	584.65 − 34.68	485.94 + 146.58
5.4	716.88 − 146.69	705.15 − 67.01	652.07 + 69.05	478.59 + 232.77
5.5	828.24 − 25.24	783.15 + 64.88	671.63 + 195.97	441.18 + 304.23
5.6	882.84 + 143.41	798.90 + 222.69	638.21 + 315.62	387.58 + 351.84
5.7	859.57 + 318.76	748.89 + 366.60	568.02 + 403.96	330.81 + 376.41
5.8	772.53 + 457.51	656.62 + 467.97	484.16 + 454.72	278.55 + 383.53
5.9	656.41 + 541.25	551.76 + 522.09	403.33 + 474.43	233.78 + 379.16
6.0	540.53 + 576.21	453.87 + 539.24	333.15 + 473.09	196.78 + 367.83
6.1	439.38 + 577.65	370.77 + 532.46	275.27 + 459.07	166.71 + 352.55
6.2	356.63 + 559.22	303.37 + 511.91	228.66 + 438.02	142.45 + 335.15
6.3	290.86 + 530.19	249.79 + 484.13	191.49 + 413.36	122.95 + 316.68
6.4	239.18 + 496.05	207.51 + 452.92	161.94 + 387.04	107.32 + 297.71
6.5	198.69 + 459.83	174.20 + 420.33	138.50 + 360.08	94.89 + 278.52
6.6	166.99 + 423.08	148.01 + 387.42	119.98 + 332.98	85.14 + 259.22
6.7	142.25 + 386.49	127.51 + 354.65	105.50 + 305.93	77.73 + 239.80
6.8	123.10 + 350.34	111.67 + 322.16	94.43 + 278.95	72.42 + 220.20
6.9	108.58 + 314.61	99.75 + 289.90	86.35 + 251.95	69.11 + 200.32

TABLE I.—*Continued.*

kh \ a/λ	0.003969	0.004763	0.006350	0.009525
7.0	98.02 + 279.17	91.27 + 257.74	81.00 + 224.78	67.78 + 180.05
7.1	91.01 + 243.80	85.96 + 225.46	78.29 + 197.28	68.55 + 159.26
7.2	87.35 + 208.25	83.71 + 192.85	78.27 + 169.26	71.65 + 137.85
7.3	87.02 + 172.27	84.61 + 159.68	81.16 + 140.53	77.44 + 115.73
7.4	90.20 + 135.57	88.91 + 125.70	87.32 + 110.93	86.47 + 92.92
7.5	97.30 + 97.92	97.08 + 90.74	97.35 + 80.39	99.49 + 69.57
7.6	108.95 + 59.13	109.84 + 54.67	112.10 + 48.93	117.49 + 46.09
7.7	126.09 + 19.12	128.23 + 17.53	132.73 + 16.83	141.71 + 23.38
7.8	150.06 − 21.94	153.65 − 20.33	160.77 − 15.18	173.52 + 3.10
7.9	182.65 − 63.47	187.97 − 58.05	198.08 − 45.60	214.08 − 11.91
8.0	226.17 − 104.11	233.49 − 93.82	246.66 − 71.68	263.52 − 17.32
8.1	283.37 − 141.17	292.67 − 124.25	308.04 − 88.75	319.55 − 7.48
8.2	356.93 − 169.80	367.42 − 143.59	381.76 − 89.80	375.78 + 22.89
8.3	448.15 − 182.05	457.19 − 143.12	463.01 − 66.19	421.57 + 75.15
8.4	554.18 − 166.61	555.88 − 111.99	540.23 − 10.99	445.66 + 143.17
8.5	663.92 − 111.35	648.78 − 41.63	596.16 + 74.82	442.57 + 214.19
8.6	755.99 − 10.58	713.98 + 66.22	615.52 + 177.95	415.84 + 275.17
8.7	804.85 + 125.28	732.59 + 193.90	594.82 + 277.37	374.86 + 318.92
8.8	796.15 + 268.83	701.51 + 314.46	544.05 + 355.79	329.06 + 344.88
8.9	737.02 + 390.19	634.75 + 406.83	478.74 + 406.73	284.78 + 356.25
9.0	649.19 + 473.19	552.32 + 464.34	411.68 + 432.70	245.11 + 357.12
9.1	554.15 + 517.83	469.78 + 491.32	350.09 + 439.87	211.00 + 350.97
9.2	465.33 + 532.60	395.38 + 496.05	296.76 + 434.35	182.30 + 340.37
9.3	388.37 + 527.18	331.88 + 486.21	251.98 + 420.82	158.40 + 327.00
9.4	324.23 + 509.25	279.19 + 467.47	214.95 + 402.47	138.63 + 311.93
9.5	271.77 + 484.02	236.06 + 443.54	184.56 + 381.27	122.34 + 295.82
9.6	229.24 + 454.76	200.97 + 416.70	159.73 + 358.45	109.02 + 279.02
9.7	194.92 + 423.36	172.56 + 388.29	139.53 + 334.67	98.24 + 261.75
9.8	167.35 + 390.89	149.66 + 359.04	123.26 + 310.30	89.71 + 244.02
9.9	145.36 + 357.90	131.40 + 329.30	110.36 + 285.49	83.25 + 225.88
10.0	128.09 + 324.61	117.13 + 299.22	100.46 + 260.27	78.75 + 207.28
10.1	114.92 + 291.07	106.38 + 268.79	93.33 + 234.58	76.21 + 188.14
10.2	105.45 + 257.19	98.91 + 237.92	88.88 + 208.32	75.73 + 168.36
10.3	99.47 + 222.85	94.60 + 206.46	87.15 + 181.37	77.54 + 147.88
10.4	96.96 + 187.85	93.51 + 174.27	88.35 + 153.59	81.97 + 126.65
10.5	98.07 + 152.00	95.89 + 141.16	92.81 + 124.89	89.53 + 104.69
10.6	103.15 + 115.14	102.17 + 107.02	101.10 + 95.20	100.91 + 82.18
10.7	112.81 + 77.14	113.01 + 71.78	113.99 + 64.61	117.01 + 59.52
10.8	127.90 + 37.99	129.35 + 35.54	132.55 + 33.43	138.93 + 37.58
10.9	149.63 − 2.09	152.49 − 1.32	158.13 + 2.41	167.87 + 17.88
11.0	179.62 − 42.45	184.07 − 37.90	192.38 − 27.04	204.77 + 2.99
11.1	219.87 − 81.75	226.06 − 72.46	236.97 − 52.34	249.67 − 3.32
11.2	272.70 − 117.47	280.53 − 101.89	293.05 − 69.36	300.53 + 3.79
11.3	340.22 − 145.21	348.85 − 121.08	360.02 − 72.07	351.95 + 28.76
11.4	423.23 − 158.12	430.28 − 122.63	433.66 − 53.29	395.15 + 72.90
11.5	519.00 − 146.85	519.52 − 97.69	504.43 − 7.35	420.68 + 131.62
11.6	618.35 − 101.63	604.60 − 39.29	558.38 + 65.29	423.25 + 194.87
11.7	704.30 − 17.69	668.04 + 51.68	582.65 + 154.75	404.62 + 251.68
11.8	756.40 + 97.87	694.01 + 162.37	572.83 + 244.54	371.73 + 295.06
11.9	761.35 + 224.79	677.69 + 271.83	535.09 + 319.69	332.45 + 323.29
12.0	721.45 + 338.81	627.70 + 361.62	481.28 + 372.83	292.70 + 338.22
12.1	651.73 + 423.90	559.31 + 423.34	422.47 + 404.09	255.81 + 342.95
12.2	569.85 + 476.42	486.24 + 457.96	365.90 + 417.56	223.16 + 340.47
12.3	488.77 + 501.03	417.11 + 471.09	315.11 + 418.12	195.03 + 333.12
12.4	415.34 + 505.09	355.83 + 468.99	271.19 + 409.95	171.12 + 322.55
12.5	351.96 + 495.30	303.41 + 456.81	233.97 + 396.10	150.98 + 309.86
12.6	298.65 + 476.70	259.40 + 438.23	202.76 + 378.61	134.09 + 295.76
12.7	254.41 + 452.67	222.82 + 415.68	176.76 + 358.81	120.02 + 280.68
12.8	217.96 + 425.38	192.61 + 390.68	155.22 + 337.52	108.42 + 264.87
12.9	188.10 + 396.13	167.79 + 364.15	137.52 + 315.19	99.01 + 248.44

TABLE I.—*Continued.*

kh \ a/λ	0.003969	0.004763	0.006350	0.009525
13.0	163.81 + 365.67	147.59 + 336.61	123.15 + 292.09	91.62 + 231.44
13.1	144.29 + 334.39	131.39 + 308.31	111.78 + 268.31	86.16 + 213.86
13.2	128.96 + 302.47	118.78 + 279.34	103.19 + 243.86	82.61 + 195.65
13.3	117.41 + 269.92	109.47 + 249.70	97.27 + 218.68	81.07 + 176.74
13.4	109.42 + 236.68	103.37 + 219.31	94.06 + 192.70	81.73 + 157.08
13.5	104.96 + 202.64	100.51 + 188.05	93.72 + 165.82	84.90 + 136.62
13.6	104.13 + 167.65	101.10 + 155.80	96.58 + 137.94	91.05 + 115.41
13.7	107.26 + 131.59	105.52 + 122.46	103.14 + 109.05	100.79 + 93.58
13.8	114.87 + 94.36	114.37 + 88.01	114.11 + 79.23	114.96 + 71.52
13.9	127.76 + 55.99	128.53 + 52.54	130.45 + 48.78	134.56 + 49.99
14.0	147.02 + 16.70	149.14 + 16.44	153.41 + 18.36	160.65 + 30.34
14.1	174.08 − 22.89	177.70 − 19.48	184.45 − 10.71	194.15 + 14.82
14.2	210.76 − 61.57	215.97 − 53.63	225.07 − 36.18	235.18 + 6.76
14.3	259.07 − 97.04	265.75 − 83.24	276.32 − 54.39	282.15 + 10.42
14.4	320.88 − 125.46	328.29 − 103.86	337.81 − 60.10	330.65 + 29.90
14.5	396.93 − 140.89	403.06 − 109.08	406.14 − 47.01	373.26 + 66.92
14.6	485.07 − 135.31	485.81 − 91.19	473.48 − 9.84	401.49 + 118.40
14.7	577.95 − 100.16	566.81 − 43.79	528.06 + 52.00	409.72 + 176.29
14.8	661.74 − 30.46	631.42 + 33.63	558.11 + 131.39	398.07 + 230.82
14.9	718.90 + 69.64	665.21 + 131.90	557.98 + 214.84	371.70 + 274.82
15.0	736.14 + 184.80	661.57 + 233.99	530.92 + 288.61	337.44 + 305.54
15.1	712.01 + 294.31	625.15 + 322.91	486.18 + 344.42	301.03 + 323.77
15.2	656.82 + 381.98	567.91 + 388.81	433.73 + 380.53	266.06 + 331.89
15.3	585.48 + 441.47	502.30 + 430.15	380.94 + 399.47	234.33 + 332.52
15.4	510.61 + 474.54	437.26 + 450.49	331.95 + 405.22	206.44 + 327.87
15.5	439.96 + 486.72	377.61 + 454.99	288.50 + 401.57	182.36 + 319.60
15.6	377.08 + 483.93	325.24 + 448.42	250.94 + 391.50	161.78 + 308.86
15.7	322.93 + 471.05	280.36 + 434.42	218.91 + 377.13	144.30 + 296.41
15.8	277.12 + 451.58	242.41 + 415.55	191.84 + 359.89	129.55 + 282.74
15.9	238.76 + 427.88	210.58 + 393.50	169.09 + 340.70	117.21 + 268.16
16.0	206.86 + 401.46	184.05 + 369.35	150.12 + 320.13	107.02 + 252.81
16.1	180.52 + 373.26	162.12 + 343.73	134.47 + 298.51	98.81 + 236.78
16.2	158.99 + 343.79	144.22 + 317.02	121.80 + 276.00	92.50 + 220.09
16.3	141.71 + 313.35	129.93 + 289.40	111.90 + 252.66	88.05 + 202.70
16.4	128.27 + 282.05	118.97 + 260.91	104.65 + 228.48	85.55 + 184.57
16.5	118.45 + 249.88	111.23 + 231.52	100.08 + 203.42	85.16 + 165.65
16.6	112.16 + 216.78	106.71 + 201.17	98.32 + 177.38	87.16 + 145.91
16.7	109.49 + 182.65	105.59 + 169.77	99.66 + 150.31	91.96 + 125.36
16.8	110.72 + 147.39	108.21 + 137.23	104.54 + 122.18	100.14 + 104.15
16.9	116.31 + 110.92	115.10 + 103.54	113.61 + 93.07	112.44 + 82.58
17.0	126.97 + 73.29	127.05 + 68.79	127.76 + 63.24	129.79 + 61.32
17.1	143.69 + 34.69	145.11 + 33.31	148.09 + 33.28	153.18 + 41.53
17.2	167.78 − 4.33	170.62 − 2.16	175.96 + 4.32	183.51 + 25.18
17.3	200.85 − 42.69	205.19 − 36.23	212.77 − 21.65	221.06 + 15.17
17.4	244.75 − 78.39	250.49 − 66.49	259.58 − 41.45	264.72 + 15.32
17.5	301.22 − 108.07	307.74 − 89.06	316.27 − 50.48	310.98 + 29.56
17.6	371.08 − 126.58	376.75 − 98.33	380.27 − 43.07	353.55 + 60.09
17.7	452.88 − 126.89	454.31 − 87.42	445.26 − 14.05	384.67 + 105.21
17.8	540.86 − 101.21	532.54 − 50.09	501.11 + 38.23	398.30 + 158.56
17.9	623.67 − 44.13	598.99 + 15.50	536.82 + 108.97	393.10 + 211.35
18.0	685.93 + 42.69	640.15 + 103.16	545.56 + 187.14	372.64 + 256.21
18.1	714.21 + 147.83	647.88 + 199.04	528.11 + 259.99	342.86 + 289.45
18.2	704.07 + 253.47	623.59 + 287.35	491.51 + 318.41	309.34 + 310.85
18.3	661.92 + 343.44	576.39 + 357.10	444.77 + 359.07	275.98 + 322.15
18.4	600.22 + 409.16	517.57 + 404.58	395.32 + 383.10	244.94 + 325.66
18.5	531.11 + 449.95	456.27 + 431.56	347.87 + 393.64	217.15 + 323.49
18.6	463.09 + 469.63	398.11 + 442.17	304.77 + 394.16	192.81 + 317.30
18.7	400.73 + 473.36	345.77 + 440.80	266.82 + 387.56	171.75 + 308.30
18.8	345.81 + 465.82	300.05 + 431.05	233.99 + 376.04	153.69 + 297.33
18.9	298.56 + 450.62	260.81 + 415.59	205.89 + 361.14	138.28 + 284.92

TABLE I.—*Continued.*

kh \ a/λ	0.003969	0.004763	0.006350	0.009525
19.0	258.42 + 430.29	227.47 + 396.29	182.01 + 343.87	125.24 + 271.42
19.1	224.63 + 406.54	199.36 + 374.34	161.86 + 324.91	114.33 + 257.04
19.2	196.38 + 380.47	175.83 + 350.53	145.02 + 304.65	105.38 + 241.89
19.3	172.98 + 352.74	156.34 + 325.33	131.15 + 283.32	98.28 + 225.99
19.4	153.88 + 323.73	140.49 + 298.97	120.04 + 261.03	93.01 + 209.36
19.5	138.67 + 293.63	127.99 + 271.59	111.56 + 237.80	89.62 + 191.95
19.6	127.11 + 262.51	118.71 + 243.19	105.72 + 213.60	88.25 + 173.72
19.7	119.08 + 230.35	112.62 + 213.74	102.62 + 188.38	89.15 + 154.65
19.8	114.65 + 197.08	109.87 + 183.17	102.51 + 162.10	92.69 + 134.74
19.9	114.03 + 162.62	110.74 + 151.43	105.78 + 134.71	99.39 + 114.10
20.0	117.64 + 126.93	115.72 + 118.48	113.02 + 106.28	109.92 + 92.99
20.1	126.11 + 90.01	125.51 + 84.42	125.03 + 77.04	125.13 + 71.95
20.2	140.33 + 52.06	141.06 + 49.52	142.83 + 47.48	145.97 + 52.00
20.3	161.48 + 13.53	163.60 + 14.42	167.65 + 18.56	173.33 + 34.80
20.4	191.04 − 24.63	194.60 − 19.69	200.84 − 7.99	207.66 + 22.91
20.5	230.71 − 60.71	235.62 − 50.73	243.47 − 29.42	248.26 + 19.76
20.6	282.15 − 91.80	287.94 − 75.31	295.75 − 41.69	292.46 + 29.10
20.7	346.34 − 113.42	351.71 − 88.54	355.85 − 39.63	334.98 + 53.60
20.8	422.44 − 119.36	424.62 − 84.18	418.74 − 18.16	368.74 + 92.73
20.9	506.11 − 102.43	500.43 − 56.12	475.76 + 25.29	387.40 + 141.71
21.0	588.05 − 56.83	568.51 − 1.37	516.60 + 87.95	388.32 + 192.75
21.1	654.76 + 17.83	616.25 + 76.43	533.52 + 161.01	373.55 + 238.37
21.2	692.88 + 113.49	634.34 + 166.24	525.12 + 232.75	348.12 + 274.00
21.3	695.58 + 215.01	621.40 + 253.57	496.34 + 293.47	317.42 + 298.49
21.4	665.79 + 306.55	583.87 + 326.60	455.17 + 338.38	285.60 + 312.89
21.5	613.52 + 377.80	531.77 + 379.73	409.03 + 367.26	255.19 + 319.20
21.6	550.27 + 425.77	474.28 + 413.00	363.13 + 382.43	227.45 + 319.41
21.7	485.06 + 452.61	417.70 + 429.50	320.39 + 386.97	202.81 + 315.22
21.8	423.40 + 462.66	365.48 + 433.16	282.07 + 383.72	181.27 + 307.88
21.9	367.88 + 460.35	319.03 + 427.55	248.46 + 374.95	162.61 + 298.29
22.0	319.31 + 449.32	278.58 + 415.42	219.38 + 362.28	146.56 + 287.05
22.1	277.51 + 432.31	243.81 + 398.78	194.41 + 346.86	132.85 + 274.57
22.2	241.92 + 411.19	214.18 + 378.98	173.14 + 329.43	121.25 + 261.08
22.3	211.85 + 387.21	189.13 + 356.93	155.16 + 310.47	111.59 + 246.74
22.4	186.67 + 361.18	168.13 + 333.18	140.16 + 290.27	103.76 + 231.59
22.5	165.83 + 333.58	150.80 + 308.07	127.91 + 268.98	97.71 + 215.67
22.6	148.93 + 304.68	136.84 + 281.77	118.27 + 246.66	93.50 + 198.94
22.7	135.70 + 274.61	126.09 + 254.35	111.23 + 223.31	91.22 + 181.37
22.8	126.00 + 243.39	118.50 + 225.79	106.86 + 198.90	91.09 + 162.94
22.9	119.87 + 210.98	114.18 + 196.06	105.37 + 173.38	93.45 + 143.64
23.0	117.47 + 177.34	113.38 + 165.11	107.11 + 146.72	98.75 + 123.55
23.1	119.16 + 142.42	116.52 + 132.91	112.60 + 118.97	107.62 + 102.89
23.2	125.50 + 106.23	124.22 + 99.53	122.57 + 90.31	120.82 + 82.09
23.3	137.29 + 68.91	137.35 + 65.21	137.94 + 61.14	139.26 + 62.01
23.4	155.60 + 30.86	157.03 + 30.46	159.87 + 32.29	163.83 + 44.07
23.5	181.78 − 7.13	184.60 − 3.69	189.61 + 5.20	195.10 + 30.48
23.6	217.41 − 43.58	221.57 − 35.46	228.31 − 17.74	232.78 + 24.33
23.7	264.09 − 76.01	269.24 − 61.94	276.40 − 32.97	274.88 + 29.18
23.8	322.96 − 100.53	328.07 − 78.82	332.75 − 35.83	317.08 + 48.01
23.9	393.72 − 111.65	396.55 − 80.47	393.45 − 21.33	353.03 + 81.35
24.0	473.20 − 102.70	469.84 − 60.91	451.19 + 13.82	376.26 + 125.86
24.1	553.95 − 67.67	538.99 − 16.28	496.44 + 68.62	383.01 + 174.84
24.2	624.19 − 4.47	592.34 + 51.95	520.85 + 136.33	373.88 + 220.89
24.3	670.91 + 81.72	619.85 + 135.37	521.10 + 206.45	352.87 + 258.70
24.4	685.46 + 178.44	617.71 + 220.98	500.10 + 268.95	325.10 + 286.16
24.5	667.65 + 270.59	589.78 + 296.55	464.64 + 317.76	294.89 + 303.64
24.6	624.98 + 346.54	544.62 + 354.82	422.00 + 351.31	265.15 + 312.73
24.7	567.97 + 401.19	491.24 + 394.10	377.81 + 371.02	237.48 + 315.33
24.8	505.98 + 434.98	436.52 + 416.38	335.53 + 379.55	212.59 + 313.11
24.9	445.32 + 451.29	384.62 + 425.05	296.92 + 379.65	190.52 + 307.39

TABLE I.—*Continued.*

kh \ a/λ	0.003969	0.004763	0.006350	0.009525
25.0	389.42 + 454.21	337.56 + 423.57	262.59 + 373.60	171.27 + 299.15
25.1	339.68 + 447.40	296.00 + 414.78	232.56 + 363.16	154.58 + 289.05
25.2	296.32 + 433.73	259.88 + 400.81	206.55 + 349.55	140.21 + 277.54
25.3	259.02 + 415.26	228.80 + 383.16	184.19 + 333.62	127.95 + 264.91
25.4	227.20 + 393.40	202.27 + 362.86	165.12 + 315.94	117.61 + 251.34
25.5	200.29 + 369.09	179.82 + 340.57	149.03 + 296.84	109.08 + 236.91
25.6	177.77 + 342.91	161.06 + 316.70	135.68 + 276.53	102.31 + 221.65
25.7	159.23 + 315.22	145.69 + 291.48	124.94 + 255.10	97.31 + 205.57
25.8	144.39 + 286.20	133.52 + 265.02	116.76 + 232.59	94.18 + 188.64
25.9	133.08 + 255.93	124.50 + 237.36	111.18 + 208.96	93.10 + 170.83
26.0	125.31 + 224.40	118.68 + 208.47	108.38 + 184.20	94.35 + 152.12
26.1	121.19 + 191.58	116.27 + 178.31	108.66 + 158.27	98.35 + 132.58
26.2	121.02 + 157.44	117.64 + 146.86	112.49 + 131.20	105.65 + 112.37
26.3	125.31 + 121.98	123.34 + 114.18	120.52 + 103.12	116.98 + 91.85
26.4	134.76 + 85.31	134.15 + 80.44	133.59 + 74.38	133.17 + 71.71
26.5	150.34 + 47.75	151.09 + 46.08	152.77 + 45.64	155.10 + 53.17
26.6	173.29 + 9.97	175.41 + 11.94	179.26 + 18.13	183.47 + 38.12
26.7	205.08 − 26.78	208.54 − 20.44	214.22 − 6.11	218.27 + 29.31
26.8	247.26 − 60.39	251.79 − 48.59	258.31 − 23.97	258.16 + 30.10
26.9	301.07 − 87.50	305.89 − 68.73	310.94 − 31.25	299.68 + 43.63
27.0	366.69 − 103.25	370.02 − 75.78	369.21 − 23.13	337.25 + 71.31
27.1	441.95 − 101.51	440.52 − 64.00	427.08 + 4.21	364.48 + 111.14
27.2	520.98 − 76.20	510.00 − 28.88	475.89 + 51.22	376.74 + 157.61
27.3	593.69 − 23.94	567.90 + 29.92	507.01 + 113.18	373.24 + 203.62
27.4	647.73 + 52.63	603.81 + 106.45	515.54 + 180.97	356.84 + 243.29
27.5	673.15 + 143.69	611.96 + 189.43	502.36 + 244.57	332.21 + 273.58
27.6	667.03 + 235.30	593.68 + 266.63	472.91 + 296.83	303.80 + 294.10
27.7	634.23 + 315.00	555.85 + 329.42	434.11 + 334.84	274.85 + 306.00
27.8	584.01 + 375.78	507.06 + 374.44	391.90 + 359.05	247.31 + 311.01
27.9	525.78 + 416.31	454.58 + 402.42	350.29 + 371.62	222.12 + 310.80
28.0	466.55 + 438.88	403.28 + 416.16	311.51 + 375.10	199.62 + 306.73
28.1	410.56 + 447.12	355.80 + 418.89	276.53 + 371.83	179.79 + 299.83
28.2	359.84 + 444.64	313.24 + 413.50	245.59 + 363.63	162.47 + 290.86
28.3	315.04 + 434.41	275.84 + 402.26	218.56 + 351.85	147.46 + 280.31
28.4	276.10 + 418.66	243.36 + 386.81	195.15 + 337.44	134.54 + 268.51
28.5	242.58 + 398.98	215.40 + 368.29	175.01 + 321.02	123.55 + 255.69
28.6	213.99 + 376.42	191.54 + 347.47	157.87 + 303.02	114.35 + 241.94
28.7	189.83 + 351.70	171.39 + 324.85	143.47 + 283.68	106.89 + 227.34
28.8	169.70 + 325.24	154.65 + 300.72	131.67 + 263.14	101.15 + 211.89
28.9	153.29 + 297.29	141.12 + 275.24	122.39 + 241.45	97.21 + 195.57
29.0	140.42 + 267.98	130.71 + 248.47	115.65 + 218.61	95.22 + 178.35
29.1	131.05 + 237.34	123.44 + 220.41	111.61 + 194.60	95.44 + 160.23
29.2	125.27 + 205.36	119.49 + 191.06	110.51 + 169.39	98.22 + 141.23
29.3	123.32 + 172.01	119.17 + 160.37	112.77 + 143.00	104.07 + 121.49
29.4	125.63 + 137.29	122.96 + 128.40	118.96 + 115.53	113.65 + 101.28
29.5	132.84 + 101.29	131.56 + 95.27	129.86 + 87.24	127.74 + 81.17
29.6	145.82 + 64.26	145.90 + 61.33	146.45 + 58.69	147.20 + 62.17
29.7	165.69 + 26.75	167.14 + 27.30	169.87 + 30.89	172.79 + 45.92
29.8	193.84 − 10.19	196.62 − 5.55	201.28 + 5.59	204.77 + 34.82
29.9	231.75 − 44.78	235.68 − 35.10	241.50 − 14.54	242.30 + 31.97
30.0	280.75 − 74.11	285.23 − 58.07	290.42 − 25.70	282.75 + 40.61

dissipative media by simply letting k be complex. This is also an advantage over many existing theories.

ACKNOWLEDGMENTS

I am greatly indebted to Professor Ronold W. P. King for introducing me to this subject and for his patient guidance and numerous discussions throughout many years. For several very helpful discussions I am deeply grateful to Professor T. D. Lee. I also wish to thank Mrs. Joan G. Beaton for carrying out the calculation for Fig. 8 and Table I, and to Mrs. Betty Koerner for the calculation of Fig. 7.

APPENDIX A

This appendix is devoted to the detailed derivation of (4.13–4.15). It follows from (4.1) that $S(Z)$ has the integral representation

$$S(Z)=-\tfrac{1}{2}\int_0^\infty \{[d\xi[2(\Omega_0-\ln 2)-\ln\xi(\xi+1)]^{-1}-[2(\Omega_0-\ln 2)+2\pi i-\ln\xi(\xi+1)]^{-1}\} \times\{(1+\xi)^{-1}\exp[2ikZ(1+\xi)]-\xi^{-1}\exp(2ikZ\xi)+[\xi(1+\xi)]^{-1}\}. \quad \text{(A1)}$$

Let Ξ be a large number, then

$$\int_0^\Xi d\xi\{[2(\Omega_0-\ln 2)-\ln\xi(\xi+1)]^{-1}-[2(\Omega_0-\ln 2)+2\pi i-\ln\xi(\xi+1)]^{-1}\}[\xi^{-1}+(\xi+1)^{-1}] =\ln\{1+\pi i[\Omega_0-\ln 2-\ln\Xi]^{-1}\}, \quad \text{(A2)}$$

and

$$2\int_0^\infty d\xi\{[2(\Omega_0-\ln 2)-2\ln(\xi+1)]^{-1}-[2(\Omega_0-\ln 2)+2\pi i-2\ln(\xi+1)]^{-1}\}(\xi+1)^{-1} =\ln\{1+\pi i[\Omega_0-\ln 2-\ln\Xi]^{-1}\}-\ln\{1+\pi i[\Omega_0-\ln 2]^{-1}\}. \quad \text{(A3)}$$

On the other hand, approximately

$$\int_0^\infty d\xi\{[2(\Omega_0-\ln 2)-\ln\xi(\xi+1)]^{-1}-[2(\Omega_0-\ln 2)-2\ln(\xi+1)]^{-1}\}(\xi+1)^{-1}$$

$$\sim-\int_0^\infty d\xi[2(\Omega_0-\ln 2)-2\ln(\xi+1)]^{-2}\ln[(\xi+1)/\xi](\xi+1)^{-1}$$

$$\sim-\int_0^\infty d\xi[2(\Omega_0-\ln 2)-2\ln 2]^{-2}\ln[(\xi+1)/\xi](\xi+1)^{-1}=-\tfrac{1}{6}\pi^2[2(\Omega_0-2\ln 2)]^{-2}. \quad \text{(A4)}$$

Similarly,

$$\int_0^\infty d\xi\{[2(\Omega_0-\ln 2)+2\pi i-\ln\xi(\xi+1)]^{-1}-[2(\Omega_0-\ln 2)+2\pi i-2\ln(1+\xi)]^{-1}\}(\xi+1)^{-1} \sim-\tfrac{1}{6}\pi^2[2(\Omega_0-2\ln 2)+2\pi i]^{-2}. \quad \text{(A5)}$$

The combination of (A2–A5) yields

$$\int_0^\infty d\xi\{[2(\Omega_0-\ln 2)-\ln\xi(\xi+1)]^{-1}-[2(\Omega_0-\ln 2)+2\pi i-\ln\xi(\xi+1)]^{-1}\}[\xi(1+\xi)]^{-1} =\ln\{1+\pi i[\Omega_0-\ln 2]^{-1}\}+\tfrac{1}{3}\pi^2\{[2(\Omega_0-2\ln 2)]^{-2}-[2(\Omega_0-2\ln 2)+2\pi i]^{-2}\}. \quad \text{(A6)}$$

It remains to study the other two terms of (A1). It follows from (4.16–4.18) that

$$\int_0^\infty d\xi\{[2(\Omega_0-\ln 2)-\ln\xi(\xi+1)]^{-1}-[2(\Omega_0-\ln 2)+2\pi i-\ln\xi(\xi+1)]^{-1}\}\xi^{-1}\exp(2ikZ\xi)$$

$$\sim\int_0^\infty d\xi\xi^{-1}e^{-\xi}\{[\Omega_2(Z)-(\ln\xi+\gamma)]^{-1}-[\Omega_3(Z)-(\ln\xi+\gamma)]^{-1}\}$$

$$=\int_0^\infty d\xi e^{-\xi}\ln\{[\Omega_2(Z)-(\ln\xi+\gamma)]^{-1}[\Omega_3(Z)-(\ln\xi+\gamma)]\}\sim\ln\{[\Omega_2(Z)]^{-1}\Omega_3(Z)\}+\tfrac{1}{2}\gamma'\{[\Omega_2(Z)]^{-2}-[\Omega_3(Z)]^{-2}\}. \quad \text{(A7)}$$

Also,

$$\int_0^\infty d\xi\{[2(\Omega_0-\ln 2)-\ln\xi(\xi+1)]^{-1}-[2(\Omega_0-\ln 2)+2\pi i-\ln\xi(\xi+1)]^{-1}\}(1+\xi)^{-1}\exp[2ikZ(1+\xi)]$$
$$\sim i(2kZ)^{-1}\exp(2ikZ)\{[\Omega_2(Z)^{-1}-[\Omega_3(Z)]^{-1}\}. \quad \text{(A8)}$$

The substitution of (A6–A8) into (A1) yields (4.13).

Next, it follows from (4.1–4.2) that

$$T(Z)-iS(Z)=i\int_{C_0} d\zeta \bar{M}(\zeta)(\zeta+k)^{-1}\exp[-i(\zeta+k)Z], \quad \text{(A9)}$$

or

$$T(Z)-iS(Z)=-i\int_0^\infty d\xi\xi^{-1}\exp(2ikZ\xi)\{[2(\Omega_0-\ln 2)-\ln\xi(1+\xi)]^{-1}-[2(\Omega_0-\ln 2)+2\pi i-\ln\xi(1+\xi)]^{-1}\}. \quad \text{(A10)}$$

Equation (4.14) follows from (A7) and (A10).

Finally, it follows from (4.10) that

$$U(Z)=-i\int_0^\infty d\xi\{[2(\Omega_0-\ln 2)-\ln\xi(1+\xi)]^{-1}-[2(\Omega_0-\ln 2)+2\pi i-\ln\xi(1+\xi)]^{-1}\}$$
$$\times\{(1+\xi)^{-1}\exp[i2kZ(1+\xi)]+\xi^{-1}\exp(i2kZ\xi)\}. \quad \text{(A11)}$$

Equation (4.15) then follows from (A7), (A8), and (A11).

APPENDIX B

In this appendix (7.21–7.23) are to be derived. For this purpose, it is convenient to define, with (7.12),

$$\Delta=-\ln\Gamma-2ikh_c=\ln[\bar{L}_{+a}(k)/\bar{L}_{+a}(-k)]. \quad \text{(B1)}$$

Since

$$\bar{K}_a(k)=\bar{K}_a(-k)=2\ln(b/a), \quad \text{(B2)}$$

it is also convenient to introduce

$$\bar{K}_a'(\zeta)=\bar{K}_a(\zeta)/[2\ln(b/a)]. \quad \text{(B3)}$$

Then,

$$\ln\bar{L}_{+a}(\zeta)=-[\ln\bar{K}_a'(\zeta)]_+-\tfrac{1}{2}\ln[2\ln(b/a)]. \quad \text{(B4)}$$

It follows from (B1) and (B4) that

$$\Delta=-(\pi i)^{-1}k\int_{C_0} d\zeta(\zeta^2-k^2)^{-1}\ln\bar{K}_a'(\zeta). \quad \text{(B5)}$$

But the integrand here is bounded in the neighborhoods of $\zeta=\pm k$. Thus, the contour C_0 may be replaced by the real axis:

$$\Delta=\pi^{-1}2ik\int_0^\infty d\zeta(\zeta^2-k^2)^{-1}\ln\bar{K}_a'(\zeta). \quad \text{(B6)}$$

Since the integrand is real for $\zeta>k$, only the part $0<\zeta<k$ contributes to Γ. Write

$$\Delta=\Delta_1+\Delta_2, \quad \text{(B7)}$$

where

$$\Delta_1=\frac{2ik}{\pi}\int_0^k d\zeta(\zeta^2-k^2)^{-1}\ln\frac{\pi i\{J_0[a(k^2-\zeta^2)^{\frac{1}{2}}]H_0^{(1)}[a(k^2-\zeta^2)^{\frac{1}{2}}]-H_0^{(1)}[b(k^2-\zeta^2)^{\frac{1}{2}}]\}}{2\ln(b/a)}, \quad \text{(B8)}$$

and

$$\Delta_2=\frac{2ik}{\pi}\int_k^\infty d\zeta(\zeta^2-k^2)^{-1}\ln\frac{I_0[a(\zeta^2-k^2)^{\frac{1}{2}}]K_0[a(\zeta^2-k^2)^{\frac{1}{2}}]-K_0[b(\zeta^2-k^2)^{\frac{1}{2}}]}{\ln(b/a)}. \quad \text{(B9)}$$

First Δ_2 is to be calculated to the accuracy $(kb)^2$. The change of variable

$$\xi=b(\zeta^2-k^2)^{\frac{1}{2}} \quad \text{(B10)}$$

gives

$$\Delta_2=\pi^{-1}2ikb(S_1+S_2), \quad \text{(B11)}$$

where

$$S_1=\int_0^\infty d\xi\xi^{-2}$$
$$\times\ln\{[\ln b/a]^{-1}[I_0(a\xi/b)K_0(a\xi/b)-K_0(\xi)]\} \quad \text{(B12)}$$

depends only on the ratio b/a, and

$$S_2=\int_0^\infty d\xi\xi^{-1}[(\xi^2+k^2b^2)^{-\frac{1}{2}}-\xi^{-1}]$$
$$\times\ln\{[\ln b/a]^{-1}[I_0(a\xi/b)K_0(a\xi/b)-K_0(\xi)]\}. \quad \text{(B13)}$$

S_2 may be determined approximately by expanding the logarithmic factor in powers of ξ. If $\xi'=\xi(kb)^{-1}$, then

$$S_2=kb[4\ln(b/a)]^{-1}\int_0^\infty d\xi'\xi'[(\xi'^2+1)^{-\frac{1}{2}}-\xi'^{-1}]$$
$$\times[\gamma-1+\ln(kb\xi'/2)]. \quad \text{(B14)}$$

This gives, finally,

$$S_2=kb[4\ln(b/a)]^{-1}[2-\gamma-\ln(kb)]. \quad \text{(B15)}$$

In order to get any new results on the radiation resistance of the two-wire line, it is necessary to calculate Γ to the accuracy $(kb)^4$. Therefore, the real part of Δ_1 should be calculated to the accuracy $(kb)^4$, the imaginary part only to $(kb)^2$. For this purpose, write Δ_1 in terms of the new variable of integration $\xi=(k^2-\zeta^2)^{\frac{1}{2}}$:

$$\Delta_1=-\pi^{-1}2ik \times\int_0^k d\xi\xi^{-1}(k^2-\xi^2)^{-\frac{1}{2}}\ln\{\pi i[2\ln(b/a)]^{-1} \times[J_0(a\xi)H_0^{(1)}(a\xi)-H_0^{(1)}(b\xi)]\}. \quad \text{(B16)}$$

Now, the logarithm may be expanded, keeping terms up to b^4, but neglecting a^2. This gives the result

$$\Delta_1=\frac{-2ik}{\pi}\int_0^k\frac{d\xi}{\xi}(k^2-\xi^2)^{-\frac{1}{2}} \times\left\{\left(\ln\frac{b}{a}\right)^{-1}\left[-\left(\gamma-1+\ln\frac{b\xi}{2}-\frac{\pi i}{2}\right)\times\left(\frac{b^2\xi^2}{4}-\frac{b^4\xi^4}{64}\right)-\frac{b^4\xi^4}{128}\right] -\frac{1}{2}\left(\ln\frac{b}{a}\right)^{-2}\left(\gamma-1+\ln\frac{b\xi}{2}-\frac{\pi i}{2}\right)^2\frac{b^4\xi^4}{16}\right\}. \quad \text{(B17)}$$

The rest of the calculation is tedious but straightforward. Because of the difference of accuracy required in the real and imaginary parts, it is advantageous to write them separately. The real part gives (7.21) and the imaginary part turns out to be simply

$$\mathrm{Im}\Delta_1=-\pi^{-1}2kbS_2. \quad \text{(B18)}$$

Equation (B11) and (B18) may be combined to yield

$$\mathrm{Im}\Delta=2kbS_1/\pi, \quad \text{(B19)}$$

which is the same as (7.22).

In order to get (7.23), write the $S_a(Z)$ of (7.17) in the form

$$S_a(Z)=S_{a1}(Z)+S_{a2}(Z), \quad \text{(B20)}$$

where

$$S_{a1}(Z)=\tfrac{1}{2}\int_{C0}d\zeta\{[\bar{K}_a(\zeta)]^{-1}-[\bar{K}_a(k)]^{-1}\} \times[(k-\zeta)^{-1}+(k+\zeta)^{-1}], \quad \text{(B21)}$$

and

$$S_{a2}(Z)=-\tfrac{1}{2}\int_{C0}d\zeta\{[\bar{K}_a(\zeta)]^{-1}-[\bar{K}_a(k)]^{-1}\} \times\{(k-\zeta)^{-1}\exp[i(k-\zeta)Z] +(k+\zeta)^{-1}\exp[-i(k+\zeta)Z]\}. \quad \text{(B22)}$$

To the order $(kZ)^{-1}$, $S_{a2}(Z)$ is real. Therefore, approximately,

$$\mathrm{Im}S_a(Z)=\mathrm{Im}S_{a1}(Z). \quad \text{(B23)}$$

Similar to the case of $\ln\Gamma$, the imaginary part of $S_{a1}(Z)$ comes only from the range $0<\zeta<k$. By series expansion, it follows from (B21) and (B23) that

$$\mathrm{Im}S_a(Z)=-k\left(\ln\frac{b}{a}\right)^{-1}\mathrm{Im}\int_0^k\frac{d\xi}{\xi}(k^2-\xi^2)^{-\frac{1}{2}} \times\left\{\left(\ln\frac{b}{a}\right)^{-1}\left[-\left(\gamma-1+\ln\frac{b\xi}{2}-\frac{\pi i}{2}\right)\times\left(\frac{b^2\xi^2}{4}-\frac{b^4\xi^4}{64}\right)-\frac{b^4\xi^4}{128}\right] -\left(\ln\frac{b}{a}\right)^{-2}\left(\gamma-1+\ln\frac{b\xi}{2}-\frac{\pi i}{2}\right)^2\frac{b^4\xi^4}{16}\right\}. \quad \text{(B24)}$$

This integrand differs from that of (B17) only in the absence of $\frac{1}{2}$. Thus, (7.23) follows.

APPENDIX C

In order to derive (7.31) by the method used by Storer and King, the current distribution is assumed to be

$$I(z)=\sin k(h-|z|). \quad \text{(C1)}$$

The tangential component of the vector potential at large distance is immediately verified to be

$$A_{\tan}(\theta,\phi)=\mathrm{const}[\cos(kh\cos\theta)-\cos kh]\cos\phi, \quad \text{(C2)}$$

where the constant is independent of k, h, θ, and ϕ. Integration over θ and ϕ leads to

$$R^e=\mathrm{const}[1+2\cos^2kh-3\sin 2kh/(2kh)]. \quad \text{(C3)}$$

The constant here may be determined by the observation that, at resonance [i.e., $kh=(n+\frac{1}{2})\pi$], the present case is identical to that treated by Storer and King. Except for h_c, this gives (7.31).

Boundary-Layer Behavior in the Superconductor Transition Problem

HIRSH COHEN AND W. L. MIRANKER
International Business Machines Corporation, Research Center, Yorktown Heights, New York
(Received October 25, 1960)

A superconducting material may be isothermally transformed to a normal conductor by raising the magnetic field to a value greater than a certain critical field. When this is done, the transition takes place along an interface, determined by the critical field value and by a magnetic-flux condition. In the present paper, the effect on the transition rate of magnetic-field penetration into the superconductor is studied. This involves the solution of a free-boundary problem in which the free boundary divides two regions, each governed by a different parabolic differential equation. The problem is solved by using the asymptotic techniques of singular perturbation theory. A boundary layer is shown to exist along the moving interface on the superconducting side. The presence of the boundary layer slows the motion of the free boundary. Also included in the solution is a study of the effect of the magnitude of the initial field on the starting motion of the free boundary. Finally, some numerical results are presented.

1. INTRODUCTION

RECENT technical interest in the kinetics of transition in superconductors has produced a series of studies on certain aspects of this problem. These studies have dealt with the propagation of a magnetic field into a superconductor and the resultant change to a normal conductor either as an isothermal process[1,2] or, acknowledging the thermodynamics more precisely, by taking into account latent heat and eddy currents.[3-5] In all of these studies in which time dependent solutions are obtained, it has been assumed that penetration of the magnetic field into the superconducting region is small enough so that it has negligible effect on the rate of transition. The present work considers the nature of this assumption and shows the effect of both penetration and initial external field on propagation rate. This means that the London equation[6] for the magnetic field H

$$\nabla^2 H=\frac{4\pi}{\Lambda c^2}H+\frac{4\pi\sigma^s}{c^2}H_\tau+\frac{1}{c^2}H_{\tau\tau} \tag{1}$$

(Gaussian units are used. c is the velocity of light, τ is the real time variable, σ^s is the normal conductivity in the superconductor, Λ is the London characteristic constant and subscripts denote partial differentiation) is applied to the superconducting region instead of assuming that the magnetic field is identically zero in that region. The transition is assumed to be isothermal, although our methods may certainly be extended to take into account eddy currents and latent heat. The London equation (1) which includes the penetration effect in a linear form is known to give only a limiting representation of the penetration effect. Several nonlocal penetration terms have been discussed and studied for the static case.[7,8] We believe that the techniques and results found here may be useful in transition problems with nonlocal penetration effects.

From the mathematical point of view, this paper presents a study of a free-boundary problem in which the free boundary separates two regions each governed by a different parabolic partial-differential equation. A novel feature is the form which the conditions on the free boundary take. In typical transition problems,[9] the local condition at the free boundary relates the flux across the interface to some functional of the interface motion. In the present case, the flux is not related in a local manner to the interface. It will be seen that, at each instant, the entire distribution of magnetic field in the superconducting region affects the flux across the interface. We shall obtain approximate solutions to the boundary-value problem that describe this situation. The approximation methods used include boundary-layer techniques and series expansions. These approximations, their relation to each other and to certain limiting cases, will be discussed in the sections that follow. We will neglect displacement-current effects. The discussion is limited to one-dimensional transitions. In particular, the case considered is that of a half-space of superconducting material being switched to a normal conductor in the presence of an external field H_e greater than the critical field H_c and perpendicular to the direction of the moving boundary. Then H is a function of x only (see Fig. 1).

Particular attention must be given to the boundary conditions that hold on the free boundary. As has been indicated above, the field attains a value H_c everywhere along this moving boundary. Since the boundary's position is not specified, another condition must be given to determine it. This second condition is obtained in the following manner[10]:

[1] A. B. Pippard, Phil. Mag. **41**, 243 (1950); E. M. Lifshitz, J. Exptl. Theoret. Phys. (U.S.S.R.) **9**, 843 (1950).
[2] J. B. Keller, Phys. Rev. **111**, 6, 1497 (1958).
[3] W. B. Ittner, Phys. Rev. **111**, 148s (1958).
[4] A. W. Duijvestijn, IBM J. Research and Develop. **3**, 2, 132 (1959).
[5] T. E. Faber, Proc. Roy. Soc. (London) **A219**, 75 (1953).
[6] F. London, *Superfluids* (John Wiley & Sons, Inc., New York, 1950), Vol. 1, p. 27 *et seq.*
[7] J. R. Schrieffer, Phys. Rev. **106**, 1, 47 (1957).
[8] A. B. Pippard, Proc. Roy. Soc. (London) **A216**, 547 (1953).
[9] W. L. Miranker and J. B. Keller, J. Math. and Anal. **9**, 67 (1960).
[10] We would like to express our indebtedness to Dr. J. Swihart, IBM Research, Poughkeepsie, for his help in deriving this condition and to the referee of this paper for his comments on the derivation.

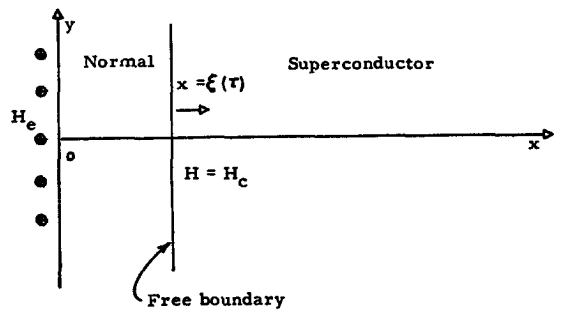


FIG. 1.

Because the magnetic field is perpendicular to the $x-y$ plane, there is a current component and an electric-field component, only in the y direction. The electric field is continuous at the boundary

$$E^-|_{x=\xi(\tau)}=E^+|_{x=\xi(\tau)} \tag{2}$$

($-$ and $+$ represent normal and superconducting regions, respectively, and $x=\xi(\tau)$ is the free boundary). The current on the normal side is related to the electric field by Ohm's law. The total current in the superconducting region is the sum of normal and superconducting components

$$j^+=j_n{}^+ + j_s{}^+. \tag{3}$$

The superconducting component is related to the magnetic field by one of London's phenomenological equations

$$\Lambda \operatorname{curl} j_s{}^+ = -H^+/c. \tag{4}$$

If this is integrated, we obtain, for our particular geometry,

$$j_s{}^+\Big|_{x=\xi} = (c\Lambda)^{-1}\int_{\xi(\tau)}^{\infty} H^+(x,\tau)dx. \tag{5}$$

$j_n{}^+$ is given by Ohm's law:

$$j_n{}^+ = \sigma^s E^+, \tag{6}$$

where σ^s is the normal conductivity in the superconducting region. Now using (2), (3), the Maxwell relation

$$\partial H/\partial x = -(c/4\pi)j \tag{7}$$

applied to j^- and j^+, and also (6) applied to j^-, we obtain the equation

$$\left(H_x{}^+ - \frac{\sigma^s}{\sigma^N}H_x{}^-\right)\Big|_{x=\xi(\tau)} = -\frac{4\pi}{\Lambda c^2}\int_{\xi(\tau)}^{\infty} H^+(x,\tau)dx. \tag{8}$$

(Here σ^N is the conductivity in the normal region.) This is the novel free-boundary condition referred to earlier. It relates the loss in magnetic flux across the boundary to the boundary motion and to the amount of magnetic field that penetrates into the superconducting region.

Other conditions that need be satisfied are:

$$H(0,\tau)=H_e, \quad \tau>0, \tag{9}$$

$$H(x,0)=H_0(x), \quad x\geq 0, \tag{10}$$

$$H(x,\tau)\to 0, \quad (x-\xi(\tau))\to\infty. \tag{11}$$

Condition (9) is the initially imposed field which switches the superconducting material to the normal conducting state. Equation (10) refers to the solution of the static penetration problem. In the present problem, it will be assumed that for time $\tau<0$ the applied field has constant magnitude H_0 and is applied in the z direction. Then, $H_0(x)=H_0\exp[-(4\pi/\Lambda c^2)^{\frac{1}{2}}x]$ is the desired solution of (1) when H is independent of τ. Condition (11) is the statement that, far in front of the moving boundary, no field has penetrated.

Equations (1) [for $x>\xi(\tau)$], (8)–(11), and the magnetic-field equation in the normal region

$$H_{xx}{}^- = (4\pi\sigma^N/c^2)H_\tau{}^-, \quad x<\xi(\tau), \tag{12}$$

together with the condition

$$H(\xi(\tau),\tau)=H_c, \tag{13}$$

define the free-boundary problem to be solved. In Sec. 2, the case of $\Lambda\to 0$ (no penetration) is set out for ease of reference. This case is thoroughly discussed in references 1 and 2. It is mathematically related to the Stefan problem known in heating and melting problems. In Sec. 3, the approximate solution for large times is obtained. This is useful nearly everywhere except for the region where τ is very small, and x is near to $\xi(\tau)$. In Sec. 4, an approximation is obtained for this last-mentioned exceptional region. Section 5 summarizes the results and offers some comment on the method.

Before going on, however, it is convenient to rewrite the equations with some changes of variables and symbols. Let $\tau=T\cdot t$ where $T=(x_0{}^2/c^2)4\pi\sigma^N$, and let $z=x/x_0$. Here, x_0 is any convenient distance scale; it may be, for instance, the width in the x direction of the superconducting material. T has the dimension of time, so that t and z are now dimensionless variables. If we further define

$$x_0{}^2 4\pi/(\Lambda c^2)=\alpha$$

[$(\Lambda c^2/4\pi)^{\frac{1}{2}}$ is the London penetration distance], then the two differential equations become

$$H_{zz}{}^- = H_t{}^-, \quad z<\zeta(t), \quad t>0, \tag{14}$$

$$H_{zz}{}^+ = \alpha H^+ + \beta H_t, \quad z>\zeta(t), \quad t>0; \tag{15}$$

β is the ratio σ^S/σ^N, $\zeta(t)$ is the free boundary in the dimensionless coordinate system. The boundary conditions are now

$$H^-(\zeta(t),t)=H^+(\zeta(t),t)=H_c, \tag{16}$$

$$H^-(0,t)=H_e \quad t>0, \tag{17}$$

$$H^+(z,0)=H_0\exp(-\alpha^{\frac{1}{2}}z) \quad z\geq 0, \tag{18}$$

$$H^+(z,t)\to 0, \quad (z-\zeta(t))\to\infty, \tag{19}$$

$$(H_z^+-\beta H_z^-)_{z=\zeta(t)}=-\alpha\int_{\zeta(t)}^{\infty}H^+(z,t)dz. \tag{20}$$

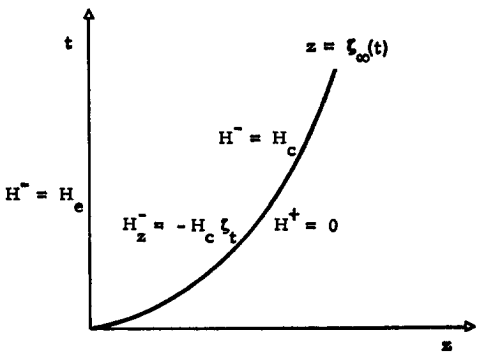


FIG. 2.

The ratio of conductivities β which enters Eq. (15) and condition (20) is a function of temperature. For this isothermal analysis it will be constant, but it has the range $0\leq\beta\leq 1$. $\beta=1$ corresponds to a temperature θ_c at which the normal conductor becomes superconducting in the presence of a magnetic field of zero strength. $\beta=0$ corresponds to the temperature $\theta=0$, and expresses the fact that *all* of the electrons in the superconducting regions are in the superconducting state. The number α is also temperature dependent. The dependence is often given as

$$\alpha=\alpha_{\theta=0}(1-(\theta/\theta_c)^4)^{\frac{1}{2}},$$

where $\alpha_{\theta=0}$ is the value at zero temperature. For large superconducting regions, so-called bulk specimens, say where x_0 would be of the order of one centimeter and for all values of the temperature except those close to θ_c, α is a large number of the order of 10^{10}. This is the situation in which the London equation (1) holds and to which our analysis applies.

2. THE SPECIAL CASE OF NO PENETRATION, $\alpha\to\infty$

When $\alpha\to\infty$, denote the resulting $\zeta(t)$ by $\zeta_\infty(t)$. In this case, we see that Eqs. (15) and (18) formally go into

$$H=0 \quad z>\zeta_\infty(t), \tag{21}$$

$$H(z,0)=0 \quad z>0. \tag{22}$$

(see Fig. 2). In the language of singular-perturbation theory, Eq. (21) is the formal reduced equation. The problem is now similar to the Stefan problem (a problem of heating or melting)[9] except for the fact that the magnetic field H is discontinuous along $z=\zeta_\infty(t)$. There is clearly a jump of magnitude H_c across the curve. In the Stefan problem, a temperature is attained along the free boundary and maintained at this constant value for $z>\zeta_\infty(t)$. In this latter situation, no heat flows into the region $z>\zeta_\infty$ since all the heat coming to the boundary from the left is used for melting. In the present problem, the physical meaning of $\alpha\to\infty$ $(\Lambda=0)$ is that the magnetic field is completely blocked from entering the region $z>\zeta_\infty(t)$. Thus, all the flux delivered to ζ_∞ is used *at the boundary* to change the superconducting material to normal.

It is important to remark that Eq. (20) does not give, for $\alpha\to\infty$, a useful condition. It is possible to derive independently a condition on the transfer of the energy of the magnetic field at the boundary into motion of the boundary when there is no field penetration. In the normal conductor, according to the law of induction, the change of the magnetic flux at the boundary generates a local current

$$H_c\frac{dx}{d\tau}\bigg|_{x=\xi(\tau)}=\frac{c}{\sigma^N}j^-\bigg|_{x=\xi(\tau)}. \tag{23}$$

This is related by one of the Maxwell equations to H^-,

$$\frac{c}{\sigma^N}j^-=-\frac{c^2}{4\pi\sigma^N}H_x^-, \quad x=\xi(\tau). \tag{24}$$

In the dimensionless coordinates, we then obtain

$$H_z^-=-H_c\zeta_t \quad z=\zeta(t). \tag{25}$$

This condition is then the one which appears in the classical Stefan problem.

It is not at all apparent at this point how condition (25) derives from condition (20), but this will be shown in the next section. Using the variable $s=zt^{-\frac{1}{2}}$, it is easy to show that Eq. (10), together with conditions (16), (17), and (25), has the solution

$$H^-(s)=H_e-H_ck_0\exp(k_0^2/4)\int_0^{s/2}\exp(-p^2)dp, \tag{26}$$

with $\zeta_\infty(t)=k_0t^{\frac{1}{2}}$, where the constant k_0 is given by

$$\frac{H_e-H_c}{H_c}=k_0\exp(k_0^2/4)\int_0^{k_0/2}\exp(-p^2)dp. \tag{27}$$

This is the solution of references 1, 2, etc.

3. ASYMPTOTIC BEHAVIOR FOR αt LARGE

For α large but finite, it may be expected that the solution in the superconducting region will again be zero far from the free boundary and from the origin in the z, t plane (for $z \gg \zeta(t)$ and $t>0$). But in this case, since H is not identically zero, it must satisfy the conditions *on* the free boundary and on $t=0$ (see Fig. 3).

From the behavior of the static solution Eq. (18), we expect a sharp exponential drop from H_c to nearly zero in the neighborhood of $z=\zeta(t)$. This behavior suggests a boundary layer or singular-perturbation treatment. We will first consider the part of the problem in the superconducting region. We will temporarily drop the free-boundary condition (20) and obtain an expression for H^+ which satisfies only the two remaining boundary conditions, $H^+(z=\zeta(t))=H_c$ and $H^+(z,0)=H_0\exp(-\alpha^{\frac{1}{2}}z)$. This expression will be an asymptotic expansion for large α, and the expansion will be a functional of ζ. When this is then inserted into the free-boundary condition (20), we will then be able to obtain an asymptotic approximation for ζ itself.

The justification for this procedure is based on some results of Aronson.[11] He shows that, for a class of equations similar to (15), there exist asymptotically-valid solutions in any open domain ($z>\zeta(t)$, $t>t_0>0$) which consist of a formal limit solution plus a boundary layer. The formal limit solution satisfies the reduced equation. For the case at hand, the solution to the reduced equation is, of course, $H^+\equiv 0$. For $t=0$, $H_0(z)$ approaches zero for z large, so that, clearly, the only area of concern for $z>\zeta(t)$ is near the free boundary and the origin. Thus, the asymptotic representation of H^+ will consist entirely of the boundary layer. Aronson's results do not exactly cover our case, but it is clear that a simple extension of his work will give the same results for the present case.

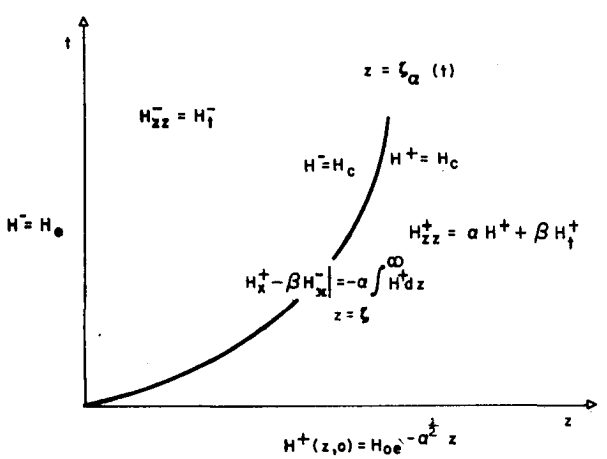


FIG. 3.

[11] D. G. Aronson, J. of Math. and Anal. 5, 1003 (1956).

It is convenient, in treating Eq. (15), to make the temporary transformation $t=\beta\theta$. Then the equation becomes

$$H_{zz}{}^{+}=H_{\theta}{}^{+}+\alpha H^{+}. \tag{28}$$

Let $H^+(z,\theta)$ be represented by

$$H^+(z,\theta)=a(z,\theta)\exp[\alpha^{\frac{1}{2}}\varphi(z,\theta)]+H_0\exp(-\alpha^{\frac{1}{2}}z). \tag{29}$$

The first term on the right-hand side is of boundary-layer type. The second term is a correction chosen so that the condition at $t=\theta=0$ may be satisfied. Let us consider first the term $a(z,\theta)\exp[\alpha^{\frac{1}{2}}\varphi(z,\theta)]$, and let

$$a(z,\theta)=a_0(z,\theta)+a_1(z,\theta)/\alpha^{\frac{1}{2}}+0(1/\alpha), \tag{30}$$

$$\zeta(\theta)=\zeta_0(\theta)+\zeta_1(\theta)/\alpha^{\frac{1}{2}}+0(1/\alpha). \tag{31}$$

To satisfy the condition on the free boundary, $H^+(\zeta(\theta),\theta)=H_c$, we choose for a_0 and a_1

$$a_0(\zeta(\theta),\theta)=H_c-H_0\exp[-\alpha^{\frac{1}{2}}\zeta(\theta)], \tag{32}$$

$$a_1(\zeta(\theta),\theta)=0, \tag{33}$$

and $\varphi(\zeta(\theta),\theta)=0$. With these choices the condition will be exactly satisfied.

Now we insert (29) into the partial-differential equation (28) and formally expand the result in inverse powers of $\alpha^{\frac{1}{2}}$. Equating coefficients of each power of $\alpha^{\frac{1}{2}}$ to zero then gives a sequence of differential equations. The first of these may be solved for φ.

$$\varphi(z,\theta)=-(z-\zeta(\theta)). \tag{34}$$

The second and third give a_0 and a_1, respectively.

$$a_0(z,\theta)=(H_c-H_0\exp[-\alpha^{\frac{1}{2}}\zeta_0(\theta)]) \times\exp[-\tfrac{1}{2}\dot{\zeta}_0(z-\zeta_0)]. \tag{35}$$

$$a_1(z,\theta)=e^{-\frac{1}{2}\dot{\zeta}_0(z-\zeta_0)}\{-\tfrac{1}{2}H_c(\dot{\zeta}_1+\tfrac{1}{4}\dot{\zeta}_0{}^2)(z-\zeta_0) +\tfrac{1}{4}H_c\ddot{\zeta}_0[(z-\zeta_0)/2]^2+0(1/\alpha^{\frac{1}{2}})\} \tag{36}$$

(dots represent differentiation with respect to θ). The procedure may, of course, be continued to determine a_2, a_3, ζ_2, ζ_3, etc. We may, therefore, write for the boundary layer solution

$$a(z,t)=\exp[-\tfrac{1}{2}\dot{\zeta}_0(z-\zeta_0)]\{(H_c-H_0\exp[-\alpha^{\frac{1}{2}}\zeta_0]) +\alpha^{-\frac{1}{2}}[-\tfrac{1}{2}H_c(\dot{\zeta}_1+\tfrac{1}{4}\dot{\zeta}_0{}^2)(z-\zeta_0) +\tfrac{1}{4}H_c\ddot{\zeta}_0[(z-\zeta_0)/2]^2]\}+0(1/\alpha). \tag{37}$$

With this choice of a and φ, H^+ as represented in (29) satisfies:

(a) $H^+(\zeta(t),t)=H_c+0(1/\alpha)$,

(b) $H^+(z,0)=H_0\exp(-\alpha^{\frac{1}{2}}z)$, provided $\dot{\zeta}_0(\theta)\to\infty$ for $\theta\to 0$,

(c) the differential equation (28), up to but not including order $1/\alpha$.

We may now write H^+ correct to $0(1/\alpha^{\frac{1}{2}})$. (We revert to the variable t instead of θ, but we will let $\zeta_0(\theta)\to\zeta_0(t)$. Dots now refer to differentiation with respect to t.)

$$H^+(z,t)=\exp[-(\alpha^{\frac{1}{2}}+\tfrac{1}{2}\dot{\zeta}_0\beta)(z-\zeta_0)]\times\{(H_c-H_0\exp[-\alpha^{\frac{1}{2}}\zeta_0])+1/\alpha^{\frac{1}{2}}\times[-\tfrac{1}{2}H_c\beta(\dot{\zeta}_1+\tfrac{1}{4}\dot{\zeta}_0^2\beta)\cdot(z-\zeta_0)+\tfrac{1}{4}H_c\ddot{\zeta}_0\beta^2\tfrac{1}{2}(z-\zeta_0)^2]\}+H_0\exp(-\alpha^{\frac{1}{2}}z). \quad (38)$$

It is now possible to compute $H_z^+|_{z=\zeta}$ and $\int_\zeta^\infty H^+dz$ so that condition (20) may be written correctly up to terms of order $1/\alpha$ and $(1/\alpha)\exp(-\alpha^{\frac{1}{2}}\zeta_0)$ as

$$-H_z^-\Big|_{z=\zeta}=\frac{1}{\beta}\left\{H_c\left(\beta\dot{\zeta}_0+\frac{1}{\alpha^{\frac{1}{2}}}\dot{\zeta}_1\beta\right)-H_0\beta\dot{\zeta}_0\exp(-\alpha^{\frac{1}{2}}\zeta_0)(1-\tfrac{1}{4}\beta\dot{\zeta}_0\alpha^{-\frac{1}{2}})\right\}. \quad (39)$$

Note that, for $\alpha\to\infty$, (39) goes over exactly to the Stefan type condition (25). This would be difficult to ascertain without the specific representation of H^+ given in (38).

Equations (14), (16), (17), and (39) now define a new boundary value problem for the determination of $H^-(z,t)$, $\zeta_0(t)$, $\zeta_1(t)$, etc., correct up to order $1/\alpha$. By using the elegant method of Kolodner,[12] a functional equation for $\zeta(t)=\zeta_0+\zeta_1\alpha^{-\frac{1}{2}}+\cdots$ may be derived:

$$\tfrac{1}{2}\dot{\zeta}(t)\{H_c-H_0\exp[-\alpha^{\frac{1}{2}}\zeta(t)]\}+\frac{H_c-H_e}{(\pi t)^{\frac{1}{2}}}\exp\left[-\frac{\zeta^2(t)}{4t}\right]=\int_0^t\frac{\dot{\zeta}(\tau)\{-H_c+H_0\exp[-\alpha^{\frac{1}{2}}\zeta(\tau)]\}}{4(\pi(t-\tau))^{\frac{1}{2}}}\frac{\zeta(t)-\zeta(\tau)}{t-\tau}\times\exp\left\{-\frac{[\zeta(t)-\zeta(\tau)]^2}{4(t-\tau)}\right\}d\tau-\int_0^t\frac{\zeta(\tau)\{-H_c+H_0\exp[-\alpha^{\frac{1}{2}}\zeta(\tau)]\}}{4(\pi(t-\tau))^{\frac{1}{2}}}\frac{\zeta(t)+\zeta(\tau)}{t-\tau}\times\exp\left\{-\frac{[\zeta(t)+\zeta(\tau)]^2}{4(t-\tau)}\right\}d\tau. \quad (40)$$

This equation may be analysed to obtain a solution to the asymptotic problem. It is possible, however, to make use of the properties of the solution for large α in a somewhat simpler fashion. The right-hand side of (39) contains a term which will be exponentially small for $\alpha^{\frac{1}{2}}$ large, except for $\zeta_0(t)$ small. Thus we neglect this term compared to the remaining members of (39). Now for αt large, if a new function $\zeta_\alpha(t)$ is defined in such a way that $\zeta_\alpha(t)=\zeta_0(t)=\zeta_1(t)=\cdots\zeta_n(t)\cdots$, the boundary condition may be written as

$$-H_z^-|_{z=\zeta(t)}=H_c\dot{\zeta}_\alpha/(1-\alpha^{-\frac{1}{2}}). \quad (41)$$

This choice of representation of the boundary condition implies that a solution will be sought that takes this form. It has the advantage of exhibiting explicitly the dependence of the solution on α. The fact that a solution of this type does exist is a justification for the step. The problem is now the same as that solved in Sec. 2 for the case of $\alpha\to\infty$ with H_c in (25) replaced by $H_c/(1-\alpha^{-\frac{1}{2}})$. The solution may thus be written

$$H^-(s)=H_e-\frac{H_c}{1-\alpha^{-\frac{1}{2}}}k_\alpha\exp(k_\alpha^2/4)\int_0^{s/2}\exp(-p^2)dp, \quad (42)$$

where k_α is a constant determined by

$$\frac{H_e-H_c}{H_c/(1-\alpha^{-\frac{1}{2}})}=k_\alpha\exp(k_\alpha^2/4)\int_0^{k_\alpha/2}\exp(-p^2)dp, \quad (43)$$

with

$$\zeta_\alpha(t)=k_\alpha t^{\frac{1}{2}}. \quad (44)$$

A comparison of (43) and (27) shows that magnetic-field penetration ($\alpha\neq\infty$) slows up the motion of the interface. The result, which is obvious qualitatively, may be quantitatively evaluated by comparing the solutions k_0 and k_α of (43) and (27) numerically. This result is valid only for αt large enough so that the exponential term of (39) may be neglected. Bearing in mind that α is of the order 10^{10}, this does not put a heavy restriction on t, nor on real time $\tau=t\cdot(1/c^2)4\pi\sigma^N$. We now turn, however, to the task of finding a solution for small t, or really for small αt. This is carried out in the next section.

4. SOLUTION FOR αt SMALL

It is convenient to reformulate the boundary value problem in a slightly different manner than for the boundary layer solution of Sec. 3. To effect this, let $H^+(z,t)=e^{-\alpha t/\beta}u^+(z,t)$. Then (15) becomes

$$u_{zz}^+=\beta u_t^+. \quad (45)$$

Also

$$u^+(\zeta(t),t)=e^{\alpha t/\beta}H_c, \quad (46)$$

and

$$u^+(z,0)=H_0\exp(-\alpha^{\frac{1}{2}}z). \quad (47)$$

Additionally it is useful to write

$$u^-(z,t)=H^-(z,t)-H_e, \quad (48)$$

so that

$$u_{zz}^-=u_t^- \quad (49)$$

$$u^-(0,t)=0, \quad (50)$$

$$u^-(\zeta(t),t)=H_c-H_e. \quad (51)$$

[12] I. I. Kolodner, Communs. Pure and Appl. Math. IX, 1, 1 (1956).

The jump condition on the derivatives Eq. (20) becomes

$$(e^{-\alpha t/\beta}u_z^+ - u_z^-)_{z=\zeta(t)} = -\alpha\int_{\zeta(t)}^{\infty} e^{-\alpha t/\beta}u^+(z,t)dz. \quad (52)$$

The equations in both regions, normal and superconducting, are now heat equations. The statement of the problem requires u^- and u^+ to exist in certain subdomains of $t>0$. Nevertheless, we find both u^+ and u^- by assuming that they are continuable as solutions of the heat equation everywhere in $t>0$. (Of course this assumption will be justified as soon as u^+ and u^- are produced.)

It is well known that a solution to the heat equation for $t>0$ may be represented as a distribution of heat sources along the z axis. Thus, for Eq. (45) we write

$$u^+(z,t) = \int_{-\infty}^{+\infty} \frac{\omega(\sigma)\exp[-(z-\sigma)^2/(4t/\beta)]}{2(\pi t/\beta)^{\frac{1}{2}}}d\sigma, \quad t>0, \quad (53)$$

and for (49)

$$u^-(z,t) = \int_{-\infty}^{+\infty} \frac{f(\sigma)\exp-[(z-\sigma)^2/4t]}{2(\pi t)^{\frac{1}{2}}}d\sigma, \quad t>0. \quad (54)$$

These define both u^+ and u^- for the entire upper half plane, providing ω and f are known for the whole real axis. The problem thus becomes one of choosing ω and f so that the conditions imposed on u^+ and u^-, Eqs. (46), (47), (50), and (51) are satisfied. This process will also yield the free-boundary curve, $\zeta(t)$.

We now make two observations concerning the function $f(z)$ in (54). First, (50) shows that $f(z)$ must be an odd function. Second, consider the function $f_0(z)$ corresponding to $f(z)$ when $\alpha=\infty$. In Sec. 2 where this case was discussed, we could have deduced that $f_0(z)$ is the following step function:

$$f_\infty(z) = \frac{(\pi)^{\frac{1}{2}}(H_c-H_e)}{2\int_0^{\frac{1}{2}k_0}\exp(-p^2)dp} \quad z>0 \quad (55)$$

$$= -\frac{(\pi)^{\frac{1}{2}}}{2}\frac{(H_c-H_e)}{\int_0^{\frac{1}{2}k_0}\exp(-p^2)dp} \quad z<0, \quad (56)$$

where k_0 is determined again by Eq. (27). With this in mind, $f(z)$ for the present case is chosen to be the sum of a step function and an odd function $g(z)$, i.e.,

$$\begin{aligned} f(z) &= \tfrac{1}{2}(g_{-1})h(z)+g(z), \\ h(z) &= -1, \quad z<0, \\ &= +1, \quad z>0 \end{aligned} \quad (57)$$

g_{-1}, a constant.

The function ω in Eq. (53) is known for $z>0$. It is the known initial value. Then $\omega(z)$ is written

$$\omega(z) = H_0\exp(-\alpha^{\frac{1}{2}}z), \quad z>0 \quad (58)$$

$$= \omega^*(z), \quad z<0. \quad (59)$$

Since the region not yet covered is that for t small, $\zeta(t)$ is now to be taken an analytic function of $t^{\frac{1}{2}}$,

$$\zeta(t) = (t/\beta)^{\frac{1}{2}}\sum_0^\infty \zeta_n(t/\beta)^{n/2} = \zeta_0(t/\beta)^{\frac{1}{2}}+\zeta_1(t/\beta)+\zeta_2(t/\beta)^{\frac{3}{2}}+\cdots, \quad (60)$$

while for $g(z)$ and $\omega^*(z)$ we choose the respective representations

$$g(z) = \sum_{n=1}^{\infty} g_{2n-1}z^{2n-1}, \quad (61)$$

and

$$\omega^*(z) = \sum_{n=0}^{\infty}\omega_n z^n. \quad (62)$$

These series are substituted into the equations for u^+ and u^- [(53) and (54)] and then used with conditions (46), (47), and (51). A set of equations is produced by equating coefficients of like powers of $t^{\frac{1}{2}}$. In this way, for instance, the coefficients g_1, $g_3\cdots\omega_0$, ω_1 and finally ζ_0 and ζ_1, etc. may be determined. Without entering into the details the results are as follows:

ζ_0 is determined from the relation

$$\frac{H_e-H_c}{H_c-H_0} = \frac{E(\frac{1}{2}\zeta_0/\beta^{\frac{1}{2}})}{(\pi\beta)^{\frac{1}{2}}(\frac{1}{2}-E(\frac{1}{2}\zeta_0/\pi^{\frac{1}{2}}))}, \quad (63)$$

where

$$E(\tfrac{1}{2}\zeta_0) = \int_0^{\frac{1}{2}\zeta_0}\exp(-p^2)dp. \quad (64)$$

For $\beta=1$, this becomes

$$(H_e-H_c)/2(H_c-H_0) = E(\tfrac{1}{2}\zeta_0)/\pi^{\frac{1}{2}}. \quad (65)$$

Then the expression for ζ_1 can be shown to be (when $\beta=1$)

$$\zeta_1 = \frac{-(\alpha)^{\frac{1}{2}}\left(\frac{Q-P}{2Q}\right)\left[\frac{\zeta_0}{\zeta_0[(Q-P)/2Q]-\exp(-\zeta_0^2/4/\pi^{\frac{1}{2}})}\right]}{\left(\frac{2Q}{P+1-Q}\right)\exp(-\zeta_0^2/4)\left\{-\frac{1}{2\pi^{\frac{1}{2}}}\left(\frac{2+\zeta_0^2}{2\zeta_0}\right)+\left[\frac{\zeta_0}{4\pi^{\frac{1}{2}}}+\frac{[(Q-P)/2Q]}{2(\zeta_0[(Q-P)/2Q]-\exp(-\zeta_0^2/4/\pi^{\frac{1}{2}}))}\right]\right\}}, \quad (66)$$

where $P=(H_e-H_c)/H_c$ and $Q=(H_e-H_0)/H_c$. There is a similar expression obtainable for β other than unity.

5. SUMMARY OF RESULTS

The results of Secs. 3 and 4 for αt either large or small may be summarized by exhibiting the expressions for the free boundary. For the first case (using the original dimensional variables now but taking $x_0=1$ cm)

$$\xi_\alpha(\tau)=k_\alpha\frac{c}{(4\pi\sigma^N)^{\frac{1}{2}}}\tau^{\frac{1}{2}}, \tag{67}$$

with k_α being determined by Eq. (43). This includes the limiting case $\alpha\to\infty$, $k_\alpha\to k_0$ (Eq. 27). The quantity $c/(4\pi\sigma^N)^{\frac{1}{2}}$ will be very close to unity for some of the common superconductors at a temperature not too close to the critical temperature θ_c, and we shall take it to be equal to unity for these calculations. It appears as a common factor and could instead be incorporated into the time variable. It is clear that one must only inspect the size of the parameter α to gain a knowledge of how k_α differs from k_0. From its definition

$$\alpha=x_0{}^2 4\pi/\Lambda c^2,$$

the dependence of α on the particular superconducting material depends only on Λ which is a number of order 10^{-30}. Thus, α is of order 10^{10} when x_0 is of order unity. It is therefore evident that such a large number cannot give values for k_α much different for k_0. We have actually computed k_α for $\alpha^{\frac{1}{2}}=10^4$ and it differs for k_0 for the same $P=(H_e-H_c)/H_c$ only in the fifth decimal place. Plotted curves for $\xi_\alpha(\tau)$ in a useful range for τ do not differ at all from those for $\xi_\infty(\tau)$ and this is true for any P in the range of interest. Thus, for large superconductor length scales ($x_0=1$ cm), α is large, and the effect of the penetration parameter on transition rate is negligible.

On the other hand, the analysis of Sec. 4 indicates that the influence of the initial external field H_0 may not be negligible. The motion of the boundary is, in this case,

$$\xi(\tau)=\tau^{\frac{1}{2}}\zeta_0\frac{c}{(4\pi\sigma^N)^{\frac{1}{2}}}\frac{1}{\beta^{\frac{1}{2}}}+\zeta_1\tau\frac{c^2}{4\pi\sigma^N}\frac{1}{\beta^2}+\cdots \tag{68}$$

(x_0 has been again set equal to 1 cm.) It is convenient for the discussion to take $\beta=1$ and $c/(4\pi\sigma^N)^{\frac{1}{2}}=1$. ζ_0 and ζ_1 are given by Eqs. (63) or (65) and (66) in terms of the two parameters P and Q. From the definition of P and Q we have

$$P\leq Q\leq P+1. \tag{69}$$

Q assumes its lower bound when $H_0=H_c$ and its upper

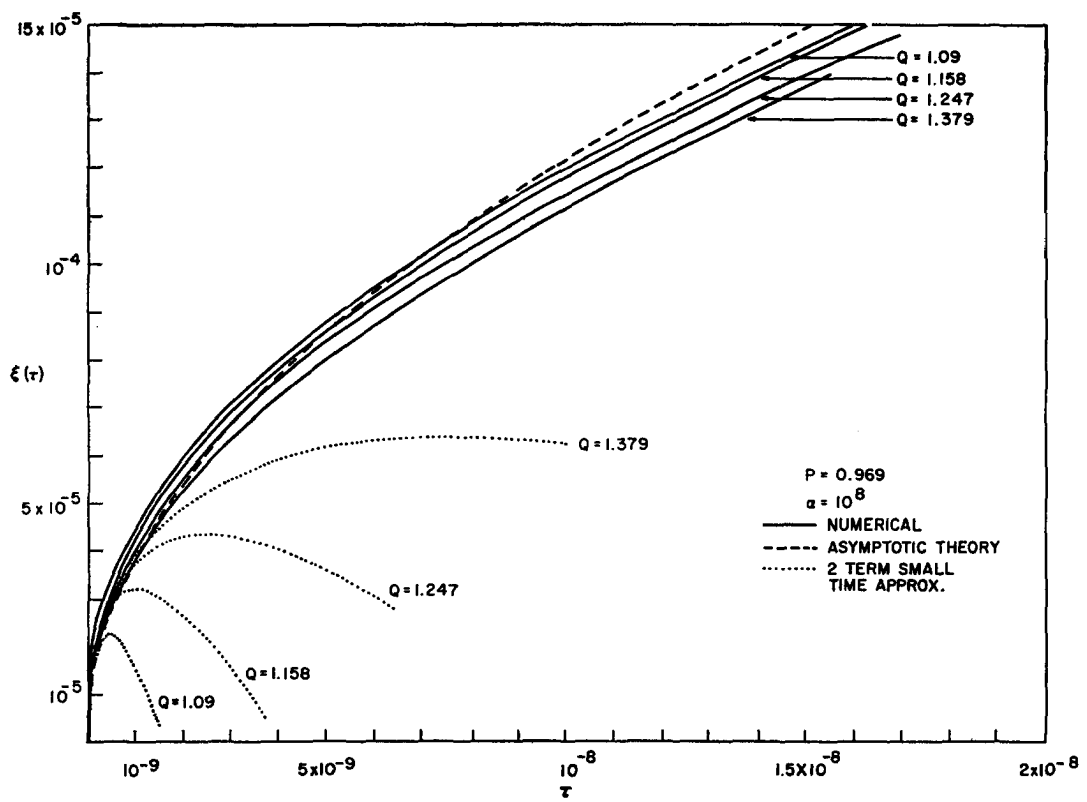


FIG. 4. Normal front motion $\xi(\tau)$ for various values of Q. $Q=(H_e-H_0)/H_c$, $P=(H_e-H_c)/H_c$.

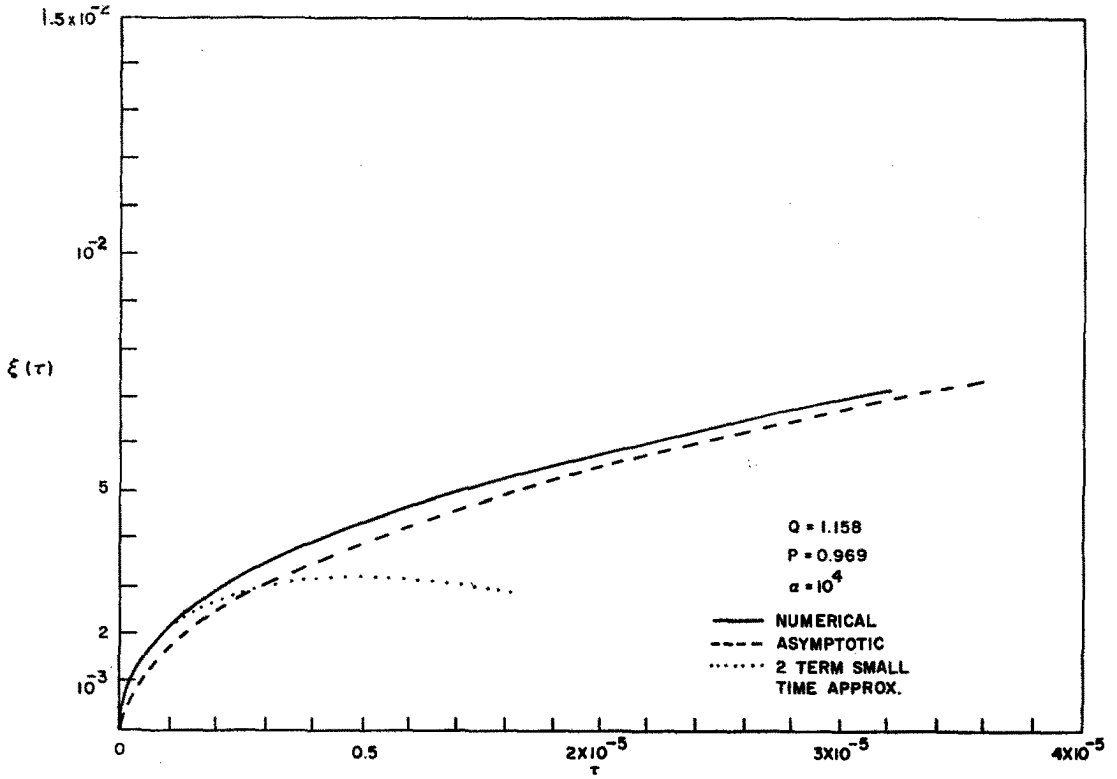


FIG. 5. Normal front motion $\xi(\tau)$ for $\alpha=10^4$.

bound when $H_0=0$. From (65), when $H_0=H_c$, $\zeta_0\to\infty$, and, of course, the series expansion indicated in (68) loses its validity.

Some curves have been prepared showing the free boundary dependence on Q. In Fig. 4, $P=0.969$. The dashed curve is the free boundary as a function of time for the asymptotic theory of Sec. 3 with $\alpha^{\frac{1}{2}}=10^4$, $k_\alpha=0.1225$. It is valid, of course, only in its upper

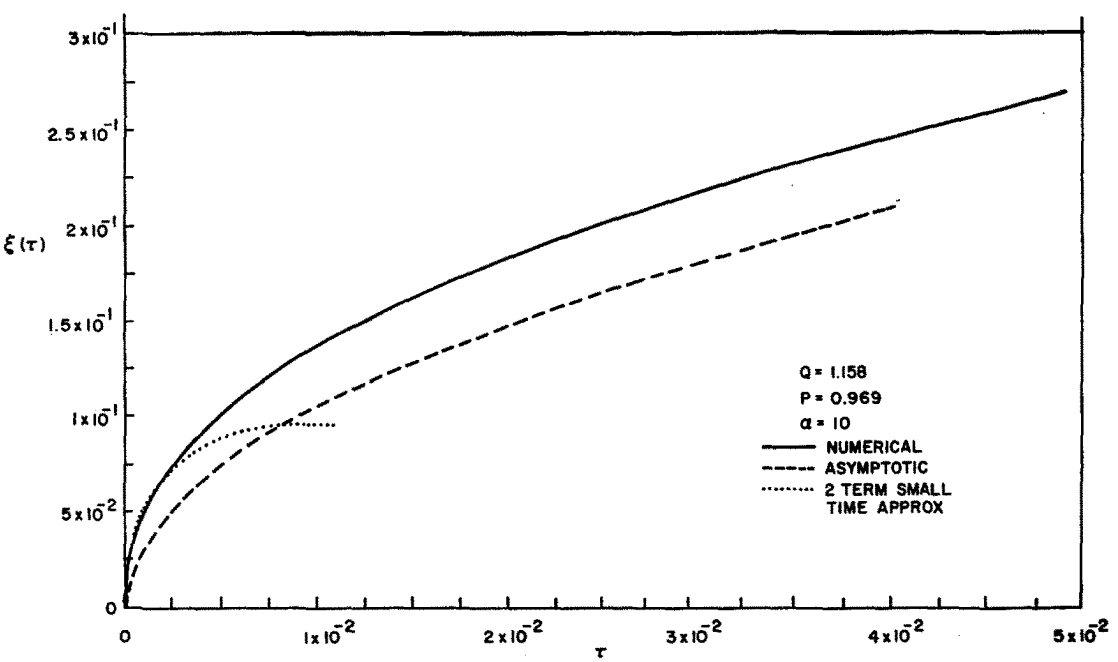


FIG. 6. Normal front motion $\xi(\tau)$ for $\alpha=10$.

region, for $\alpha\tau$ large. The dotted curves show the theory of Sec. 4 in the two-term approximation. The solid curves are the result of a numerical solution of the exact boundary value problem. This numerical study has been carried out separately by Werner Liniger of IBM Research and will be reported on in connection with a study of the transition through a finite strip which Liniger is now preparing. The numerical solution presents special difficulties for such a free-boundary problem; these will be discussed in Liniger's report. We are very grateful to him to be able to use these results.

In general, one can see that the asymptotic theory is very good for this large value of α. For small values of time, the slope of the curve is incorrectly given by this theory but correctly given by the two-term approximation. For the larger values of Q, the two-term approximation shows its best agreement.

For smaller values of α, one would expect the asymptotic theory to be at greater variance with the exact results. This is shown on Fig. 5. Here $\alpha=10^4$ and $P=0.969$. The case $Q=1.158$ is exhibited and may be compared to the similar case in Fig. 4. The two-term, small time expansion would be useful up to times of the order of 3×10^{-6} sec in this case. An even smaller value of α is shown in Fig. 6. The slope of the asymptotic curve is clearly in error at $t=0$. The small time expansion is good here out to $t=3\times10^{-3}$ sec.

It should be remarked that the small values of α may already refer to cases in which the London theory is inapplicable. Nonlocal effects probably have to be accounted for in these situations.

Frequency Spectrum of a Disordered One-Dimensional Lattice

J. S. LANGER
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received January 26, 1961)

The frequency spectrum of a disordered one-dimensional lattice is calculated via an investigation of the phonon propagator. The spectrum is evaluated in detail for a low concentration of light impurities inserted at random along a linear chain. It is found that an impurity band occurs near the frequency of the local mode. Higher-order effects resulting from clusters of impurities are calculated and discussed.

I. INTRODUCTION

THE purpose of this paper is to present a new and fairly simple technique for calculation of the frequency spectra of disordered lattices. The particular case considered is a one-dimensional chain of point masses connected to each other by springs. The disorder is introduced by choosing fixed concentrations of two different masses and arranging these masses at random along the chain. This one-dimensional problem recently has attracted a considerable amount of interest. The dynamical properties of the model may be qualitatively similar to those of a great many physical systems ranging from real crystals containing impurities, through electrons in alloys, and perhaps even to long molecular chains occurring in organic materials. Furthermore, the disordered chain provides a relatively simple model for the theoretical investigation of transport phenomena.

Of the theoretical techniques which have been applied to this problem,[1] one of the most successful in terms of actual results seems to be the moment-trace method originally devised by Montroll for calculation of the vibration spectra of ordered crystals.[2] It turns out that, with some work, one can compute the even moments of the disordered spectrum exactly. Domb, Maradudin, Montroll, and Weiss have performed this calculation up through the twentieth moment; and their polynomial fit to the spectrum agrees roughly with the results given in this paper.[3] Somewhat more recently, Dean has succeeded in devising a machine technique for computing these spectral functions.[4,5] Especially in the case of large mass ratio, Dean finds much more complicated spectra than those given by Domb *et al.* As we shall see, the present calculation lends support to Dean's results, although we shall not find it convenient to make a really direct comparison. A definite disadvantage of both of these methods, however, is that they offer little or no insight into the actual dynamics of the lattice. For example, they tell nothing about phonon lifetimes or the nature of energy propagation in the impurity bands. The present method does seem to give information of this sort.

Qualitatively, the dynamical properties of the disordered chain are not difficult to predict. We know that the introduction of a small number of impurities along the chain will cause the ordinary phonon modes to decay. Equivalently, an excitation with a given wave number will not propagate with a well-defined frequency, but will have associated with it a width. Furthermore, we know that if the impurities are heavy, they will cause a shift of the spectrum towards the lower frequencies. The more interesting situation occurs when the impurities are light. A single light impurity inserted into a long chain gives rise to a high-frequency "local mode" in which the displacements of the individual mass points fall off exponentially away from the position of the light mass.[6,7] (The mathematical description of this mode is reviewed at the beginning of Sec. V.) Thus, a chain containing a small number of light impurities should propagate signals at frequencies near that of the local mode. That is, we expect to find an impurity band near the upper end of the spectrum.

In this paper these impurity effects will be evaluated via an investigation of the phonon propagator. The formal connection between the spectrum and this propagation function is established in Sec. II. In Sec. III it is shown how the perturbation expansion for the propagator may be resumed after performing an average over configurations of the masses. The actual evaluation of the spectrum is reproduced in Secs. IV and V; and these results are further discussed and interpreted in Sec. VI.

II. FORMULATION OF THE PROBLEM

We consider a very long chain containing two kinds of mass points M and M'. Let the less common mass be M', and let q be the frequency with which M' occurs. That is, the probability of any particular mass being M' is q. In most of the following work we shall assume $q \ll 1$. Thus, we shall consider the chain in which all points have mass M as the zeroth approximation to our system and refer to it as the unperturbed lattice.

Start by considering a particular distribution of

[1] For a review of previous approaches to this problem, see A. A. Maradudin, P. Mazur, E. W. Montroll, and G. H. Weiss, Revs. Modern Phys. **30**, 175 (1958).
[2] E. W. Montroll, J. Chem. Phys. **10**, 218 (1942); **11**, 481 (1943).
[3] C. Domb, A. A. Maradudin, E. W. Montroll, and G. H. Weiss, Phys. Rev. **115**, 18, 24 (1959).
[4] P. Dean, Proc. Phys. Soc. (London) **73**, 413 (1959).
[5] P. Dean, Proc. Roy. Soc. (London) **A254**, 507 (1960).
[6] E. W. Montroll and R. B. Potts, Phys. Rev. **100**, 525 (1955); **102**, 72 (1956).
[7] L. M. Lifshitz, Nuovo cimento Suppl. **3**, 716 (1956).

masses M and M' along the chain. For this configuration the displacement x_l of the lth mass point obeys the equation of motion

$$m_l\ddot{x}_l=\gamma(x_{l-1}-2x_l+x_{l+1}), \tag{1}$$

where m_l is the mass at the lth point and γ is the force constant, assumed the same for all links in the chain. We rewrite Eq. (1) in the form

$$\ddot{x}_l-(\gamma/M)(x_{l+1}-2x_l+x_{l-1}) =\gamma[(1/m_l)-(1/M)](x_{l+1}-2x_l+x_{l-1}). \tag{2}$$

The transformation to the normal coordinates Q_k of the unperturbed lattice is

$$x_l=\frac{1}{(NM)^{\frac{1}{2}}}\sum_{k=-N/2}^{N/2}Q_k\exp[2\pi ikl/N-i\omega t], \tag{3}$$

where N is the number of mass points in the chain. On substituting (3) into (2), we have

$$(\omega_k^2-\omega^2)Q_k=-\sum_{k'}\Phi_{k,k'}Q_{k'}, \tag{4}$$

where

$$\omega_k=\omega_M|\sin(\pi k/N)|, \quad \omega_M=(4\gamma/M)^{\frac{1}{2}}, \tag{5}$$

and

$$\Phi_{k,k'}=\frac{\omega_{k'}^2}{N}\sum_l\left(\frac{M}{m_l}-1\right)\exp\left[\frac{2\pi i(k'-k)l}{N}\right]. \tag{6}$$

Next we introduce the notation

$$[D^{-1}(\omega^2)]_{k,k'}=(\omega_k^2-\omega^2)\delta_{k,k'}+\Phi_{k,k'}, \tag{7}$$

so that

$$\textstyle\sum_{k'}(D^{-1}{}_{k,k'})Q_{k'}=0. \tag{8}$$

The exact eigenfrequencies Ω_n for this particular configuration of masses are the roots of the equation

$$\det D^{-1}(\omega^2)=\prod_{n=1}^{N}(\Omega_n^2-\omega^2)=0. \tag{9}$$

If we invert the matrix D^{-1} after diagonalization, we find

$$\mathrm{Tr}D(\omega^2)=\textstyle\sum_n(\Omega_n^2-\omega^2)^{-1}. \tag{10}$$

Finally, we define the spectral distribution function

$$g(\omega)=\lim_{N\to\infty}\frac{2\omega}{N}\sum_n\delta(\Omega_n^2-\omega^2), \tag{11}$$

where it is understood that the delta function which appears here is the limit of a function whose width is much greater than the spacing between the Ω_n's. That is,

$$\delta(\Omega_n^2-\omega^2)=-\frac{1}{\pi}\lim_{\epsilon\to0}\mathrm{Im}\frac{1}{\Omega_n^2-\omega^2-i\epsilon}, \tag{12}$$

and we keep $\epsilon\gg(\omega_M/N)$ in taking the limit. Thus,

$$g(\omega)=\frac{2\omega}{\pi}\lim_{\substack{N\to\infty\\ \epsilon\to0}}\frac{1}{N}\,\mathrm{Im}\,\mathrm{Tr}D(\omega^2+i\epsilon). \tag{13}$$

Since the trace is independent of representation, we may evaluate (13) in the k representation as defined by Eq. (7). Furthermore, the process of averaging over mass configurations has the effect of restoring momentum conservation in D, because the "average" chain possesses translational invariance. We write

$$\bar{D}_{k,k'}(\omega^2)=D_k(\omega^2)\delta_{k,k'}. \tag{14}$$

Then, since the operation of taking the configuration average commutes with the other operations in (13), we have

$$\bar{g}(\omega)=\frac{2\omega}{\pi}\lim_{\substack{N\to\infty\\ \epsilon\to0}}\frac{1}{N}\,\mathrm{Im}\sum_k D_k(\omega^2+i\epsilon). \tag{15}$$

III. CONFIGURATION AVERAGE

According to Eq. (7), $D_{k,k'}(\omega^2)$ satisfies

$$D_{k,k'}(\omega^2)=\frac{\delta_{k,k'}}{\omega_k^2-\omega^2}-\frac{1}{\omega_k^2-\omega^2}\sum_{k''}\Phi_{k,k''}D_{k'',k'}(\omega^2), \tag{16}$$

which may be iterated in the usual way:

$$D_{k,k'}(\omega^2)=\frac{\delta_{k,k'}}{\omega_k^2-\omega^2}-\frac{1}{\omega_k^2-\omega^2}\Phi_{k,k'}\frac{1}{\omega_{k'}^2-\omega^2}+\frac{1}{\omega_k^2-\omega^2}\times\sum_{k_1}\frac{\Phi_{k,k_1}\Phi_{k_1,k'}}{\omega_{k_1}^2-\omega^2}\frac{1}{\omega_{k'}^2-\omega^2}+\cdots. \tag{17}$$

We achieve a considerable simplification of this formula by introducing the configuration average at this point.

Before drawing general conclusions, let us consider the two simplest terms in $\bar{D}$. First look at

$$\langle\Phi_{k,k'}\rangle_{\mathrm{av}} =\left\langle\frac{\omega_{k'}^2}{N}\sum_l\left(\frac{M}{m_l}-1\right)\exp\left[\frac{2\pi i(k'-k)l}{N}\right]\right\rangle_{\mathrm{av}}. \tag{18}$$

This expression is evaluated by keeping l fixed while summing over configurations. Obviously, we get zero contribution from any configuration in which $m_l=M$. Since $m_l=M'$ with frequency q, we have

$$\langle\Phi_{k,k'}\rangle_{\mathrm{av}}=\frac{q\lambda\omega_{k'}^2}{N}\sum_l\exp\left[\frac{2\pi i(k'-k)l}{N}\right]=q\lambda\omega_{k'}^2\delta_{k,k'}, \tag{19}$$

where we have used the notation

$$\lambda=(M/M')-1. \tag{20}$$

Next consider the second-order term,

$$\langle\Phi_{k,k'}\Phi_{k'',k'}\rangle_{\rm av}$$
$$=\left\langle\frac{\omega_{k''}{}^2\omega_{k'}{}^2}{N^2}\sum_{l_1,l_2}\left(\frac{M}{m_{l_1}}-1\right)\left(\frac{M}{m_{l_2}}-1\right)\right.$$
$$\left.\times\exp\left[\frac{2\pi i(k''-k)l_1}{N}\right]\exp\left[\frac{2\pi i(k'-k'')l_2}{N}\right]\right\rangle_{\rm av}. \quad (21)$$

If we distinguish the cases $l_1=l_2$ and $l_1\neq l_2$, we may write

$$\langle\Phi_{k,k'}\Phi_{k'',k'}\rangle_{\rm av}$$
$$=\frac{q\lambda^2\omega_{k''}{}^2\omega_{k'}{}^2}{N^2}\sum_l\exp\left[\frac{2\pi i(k'-k)l}{N}\right]$$
$$+\frac{q^2\lambda^2\omega_{k''}{}^2\omega_{k'}{}^2}{N^2}\sum_{\substack{l_1,l_2\\ l_1\neq l_2}}\exp\left[\frac{2\pi i(k''-k)l_1}{N}\right]$$
$$\times\exp\left[\frac{2\pi i(k'-k'')l_2}{N}\right]$$
$$=(\omega_{k''}{}^2\omega_{k'}{}^2\lambda^2/N)(q-q^2)\delta_{k,k'}$$
$$+\omega_{k''}{}^2\omega_{k'}{}^2q^2\lambda^2\delta_{k,k''}\delta_{k'',k'}. \quad (22)$$

Note that, in performing the second sum, we added and then subtracted the term in which $l_1=l_2$, thus finding a correction of order q^2 to the coefficient of $\delta_{k,k'}$.

Now we are ready to find a prescription for the calculation of the configuration average of a product of any number of Φ's. The expression we want to evaluate is

$$\langle\Phi_{k_1,k_2}\cdots\Phi_{k_n,k_{n+1}}\rangle_{\rm av}$$
$$=\frac{\omega_{k_2}{}^2\cdots\omega_{k_{n+1}}{}^2}{N^n}\left\langle\sum_{l_1,\cdots l_n}\left(\frac{M}{m_{l_1}}-1\right)\cdots\left(\frac{M}{m_{l_n}}-1\right)\right.$$
$$\left.\times\exp\left[\frac{2\pi i}{N}(p_1l_1+\cdots+p_il_i+\cdots)\right]\right\rangle_{\rm av}, \quad (23)$$

where p_i is the momentum transfer $(k_{i+1}-k_i)$ in the ith scattering event. The first step is to write down all possible partitions of the l_i's and to set the l's equal in each group. That is,

$$\langle\Phi_{k_1,k_2}\cdots\rangle_{\rm av}$$
$$=\frac{\omega_{k_1}{}^2\cdots}{N^n}\left\langle\sum_l\left(\frac{M}{m_l}-1\right)^n\right.$$
$$\left.\times\exp\left[\frac{2\pi il}{N}(p_1+p_2+\cdots+p_n)\right]\right\rangle_{\rm av}$$
$$+\frac{\omega_{k_1}{}^2\cdots}{N^n}\sum_{\substack{s_1,s_2\\ s_1+s_2=n}}{}^{(2)}\left\langle\sum_{\substack{l_1,l_2\\ l_1\neq l_2}}\left(\frac{M}{m_{l_1}}-1\right)^{s_1}\left(\frac{M}{m_{l_2}}-1\right)^{s_2}\right.$$
$$\left.\times\exp\left[\frac{2\pi il_1}{N}(p_1+\cdots)\right]\exp\left[\frac{2\pi il_2}{N}(p_1+\cdots)\right]\right\rangle_{\rm av}$$
$$+\frac{\omega_{k_1}{}^2\cdots}{N^n}\sum_{\substack{s_1,s_2,s_3\\ s_1+s_2+s_3=n}}{}^{(3)}\left\langle\sum_{\substack{l_1,l_2,l_3\\ l_1\neq l_2\neq l_3}}\left(\frac{M}{m_{l_1}}-1\right)^{s_1}\right.$$
$$\times\left(\frac{M}{m_{l_2}}-1\right)^{s_2}\left(\frac{M}{m_{l_3}}-1\right)^{s_3}$$
$$\left.\times\exp\left[\frac{2\pi il_1}{N}(p_1+\cdots)\right]\cdots\right\rangle_{\rm av}+\cdots, \quad (24)$$

where $\sum^{(2)}$ sums all partitions of the l's into two groups, $\sum^{(3)}$ sums three groups, and so on. On taking the average and summing over the l's, we find an expression of the form

$$\frac{\omega_{k_1}{}^2\cdots}{N^{n-1}}\lambda^nP_n(q)\delta_{p_1+\cdots+p_n,0}$$
$$+\frac{\omega_{k_1}{}^2\cdots}{N^{n-2}}\lambda^n\sum_{s_1,s_2}{}^{(2)}P_{s_1}(q)P_{s_2}(q)\delta_{p_1+\cdots,0}\delta_{p_i+\cdots,0}$$
$$+\frac{\omega_{k_1}{}^2\cdots}{N^{n-3}}\lambda^n\sum_{s_1,s_2,s_3}{}^{(3)}P_{s_1}(q)P_{s_2}(q)P_{s_3}(q)\delta\ldots\delta\ldots\delta\ldots$$
$$+\cdots, \quad (25)$$

where $P_s(q)$ is an sth-degree polynomial in q. For small q, the leading terms in P_s are

$$P_s=q-(2^{s-1}-1)q^2+\cdots. \quad (26)$$

Note that, whenever a sequence of momentum transfers occurs in a Kroenecker delta symbol, e.g., $\delta_{p_1+\cdots+p_n,0}$, then all of the associated scattering events have occurred at the same lattice point l. This fact suggests that it will be useful to adopt a diagrammatic representation of the various terms occurring in the perturbation expansion of $\bar{D}$. In Fig. 1 a horizontal line represents the phonon of momentum k and frequency ω whose propagation is described by $D_k(\omega^2)$.

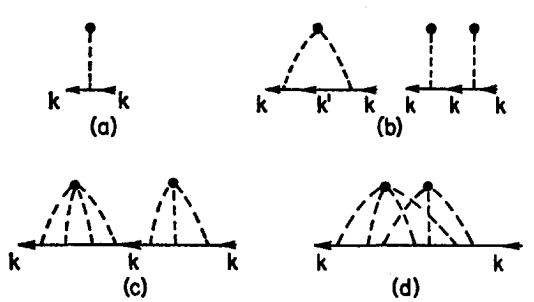


FIG. 1. Some graphs occurring in the perturbation expansion of $\bar{D}$.

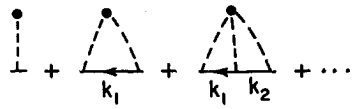


FIG. 2. The graphs which contribute to $G^{(1)}$.

The interactions are denoted by dashed lines which start at dots representing the lattice points where the interactions occur and connect to the phonon line in the order in which they occur in the perturbation expansion (17). Figures 1(a) and 1(b) are the graphs associated with Eqs. (19) and (22), respectively. Figures 1(c) and 1(d) depict typical graphs associated with the $\sum^{(2)}$ term in Eq. (25).

These graphs have the general form of a single phonon line with a series of self-energy parts. In the usual manner, we define a proper self-energy part as one which cannot be broken into two such parts simply by cutting the phonon line once. For example, Fig. 1(c) contains two proper parts, whereas Fig. 1(d) contains only one. Since the total momentum transfer associated with a single lattice point is always zero, the proper self-energy parts are always diagonal in the k index. Let us denote the sum of all proper parts by $G_k(\omega^2)$. Then

$$D_k(\omega^2)=\frac{1}{\omega_k^2-\omega^2}\,\frac{1}{\omega_k^2-\omega^2}G_k(\omega^2)D_k(\omega^2)$$
$$=\frac{1}{\omega_k^2-\omega^2+G_k(\omega^2)}. \quad (27)$$

IV. CALCULATION OF FIRST-ORDER TERMS

In this section we shall compute $G_k(\omega^2)$ to first order in q and use the resulting expression for calculation of the spectrum.

According to Eqs. (25) and (26), the only self-energy graphs which contain contributions linear in q are those in which all the scattering events occur at the same point in the lattice. Thus the leading contribution to G is the sum of graphs depicted in Fig. 2. To lowest order in q, this sum is

$$G_k^{(1)}(\omega^2)=q\omega_k^2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}\lambda^n}{N^{n-1}}$$
$$\times\sum_{k_1,\cdots k_{n-1}}\left(\frac{\omega_{k_1}^2}{\omega_{k_1}^2-\omega^2}\right)\cdots\left(\frac{\omega_{k_{n-1}}^2}{\omega_{k_{n-1}}^2-\omega^2}\right)$$
$$=\frac{q\lambda\omega_k^2}{1+\frac{\lambda}{N}\sum_{k'}\frac{\omega_{k'}^2}{\omega_{k'}^2-\omega^2}}=\frac{q\kappa\omega_k^2}{1+\frac{\kappa\omega^2}{N}\sum_{k'}\frac{1}{\omega_{k'}^2-\omega^2}}, \quad (28)$$

where

$$\kappa=\lambda/(1+\lambda)=1-(M'/M). \quad (29)$$

Before taking the limit $N\to\infty$ in (28), let us examine some qualitative features of these results. The poles of $G_k^{(1)}(\omega^2)$ are determined by

$$1+\frac{\kappa\omega^2}{N}\sum_{k'}\frac{1}{\omega_{k'}^2-\omega^2}=0, \quad (30)$$

which is the dispersion relation for a single mass M' inserted into a chain of masses M. For $\omega^2<\omega_M^2$, the solutions of (30) lie in one-to-one correspondence with the unperturbed spectrum of ω_k's, the frequency shifts being of order N^{-1}. In the case $M'<M$, $0<\kappa<1$, there exists a single solution of (30) for $\omega^2>\omega_M^2$. This is the local mode mentioned in the Introduction. A graph of $G_k^{(1)}(\omega^2)$ is shown in Fig. 3.

The poles of $D_k(\omega^2)$ lie at the solutions of

$$\omega_k^2-\omega_{k,i}^2+G_k(\omega_{k,i}^2)=0, \quad (31)$$

which are illustrated graphically in Fig. 3 for some particular value of k. Notice that for each value of k there are many solutions $\omega_{k,i}$ of (31). In other words, in the "average chain," a phonon of wave number k does not propogate with a single well-defined frequency. The contribution to $\bar{g}(\omega)$ from any one of the frequencies associated with k is proportional to the residue at the corresponding pole of $D_k(\omega^2)$, which is

$$\operatorname*{Res}_{\omega=\omega_{k,i}}\{D_k(\omega^2)\}$$
$$=[-2\omega_{k,i}+(\partial/\partial\omega)G_k(\omega^2)|_{\omega=\omega_{k,i}}]^{-1}. \quad (32)$$

Again referring to Fig. 3, we see that this residue is largest for $\omega_{k,i}$ in the neighborhood of ω_k and diminishes rapidly on either side of this point as $\partial G/\partial\omega|_{\omega=\omega_{k,i}}$ becomes large. Thus there is a width associated with each phonon of wave number k; i.e., the phonon has a finite lifetime as a result of the lattice disorder. Also notice that, when $M'<M$, each phonon is coupled to the local mode, the coupling being strongest for large k.

If we now take the limit $N\to\infty$ as prescribed by Eq. (13), the line of poles in the region $\omega^2<\omega_M^2$ turns

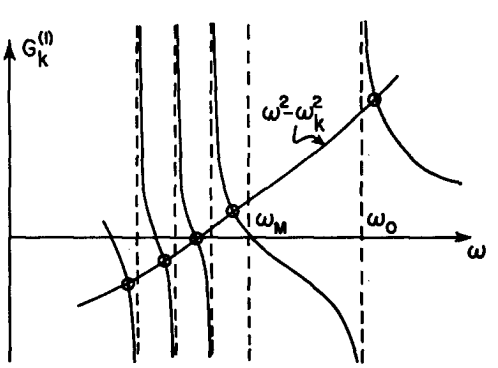


FIG. 3. The function $G^{(1)}$. The small circles indicate the solutions of Eq. (31).

into a branch cut. We have

$$\lim_{\substack{N\to\infty \\ \epsilon\gg\omega_M/N\to 0}} \frac{1}{N}\sum_k \frac{1}{\omega_k^2-\omega^2-i\epsilon}=\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\frac{dx}{\omega_M^2\sin^2 x-\omega^2}$$

$$=\frac{i}{\omega(\omega_M^2-\omega^2)^{\frac{1}{2}}}. \quad (33)$$

Thus,

$$D_k(\omega^2)=\left[\omega_k^2-\omega^2+\frac{q\kappa\omega_k^2(\omega_M^2-\omega^2)^{\frac{1}{2}}}{(\omega_M^2-\omega^2)^{\frac{1}{2}}+i\kappa\omega}\right]^{-1}. \quad (34)$$

The cut is defined by (33) so that

$$D_k(\omega^2)=\left[\omega_k^2-\omega^2+\frac{q\kappa\omega_k^2(\omega^2-\omega_M^2)^{\frac{1}{2}}}{(\omega^2-\omega_M^2)^{\frac{1}{2}}-\kappa\omega}\right]^{-1} \text{ for } \omega^2>\omega_M^2. \quad (35)$$

If we analytically continue from above onto the unphysical Riemann sheet in (34), we find that D has a single pole for each value of k. To first order in q, this pole occurs at

$$\omega=\omega_k+\Delta_k-i\Gamma_k;$$

$$\Delta_k=\frac{\omega_k}{2}\frac{q\kappa(\omega_M^2-\omega_k^2)}{\omega_M^2-\omega_k^2(1-\kappa^2)};\quad \Gamma_k=\frac{\omega_k}{2}\frac{q\kappa^2\omega_k(\omega_M^2-\omega_k^2)^{\frac{1}{2}}}{\omega_M^2-\omega_k^2(1-\kappa^2)}, \quad (36)$$

Γ_k being a measure of the width described in the preceding paragraph. In the region $\omega^2>\omega_M^2$, according to Eq. (35), $D_k(\omega^2)$ may have a pole on the real axis just above the local mode frequency.

The computation of $\bar{g}(\omega)$ now may be performed explicitly:

$$\bar{g}(\omega)=\lim_{N\to\infty}\frac{2\omega}{N\pi}\operatorname{Im}\sum_k D_k(\omega^2+i\epsilon)$$

$$=\frac{2\omega}{\pi^2}\operatorname{Im}\int_{-\pi/2}^{\pi/2}dx$$

$$\times\left[\omega_M^2\sin^2 x\left(1+\frac{q\kappa(\omega_M^2-\omega^2)^{\frac{1}{2}}}{(\omega_M^2-\omega^2)^{\frac{1}{2}}+i\kappa\omega}\right)-\omega^2\right]^{-1}$$

$$=\frac{2}{\pi}\operatorname{Re}\left[\omega_M^2-\omega^2+\frac{\omega_M^2 q\kappa(\omega_M^2-\omega^2)^{\frac{1}{2}}}{(\omega_M^2-\omega^2)^{\frac{1}{2}}+i\kappa\omega}\right]^{-\frac{1}{2}}. \quad (37)$$

For frequencies $\omega^2\ll\omega_M^2$, Eq. (37) yields just the unperturbed spectrum for the ordered chain all of whose masses are equal to the mean mass of our disordered system. That is,

$$\bar{g}(\omega)\cong(2/\pi)[1/(\omega_{\bar{M}}^2-\omega^2)^{\frac{1}{2}}],\quad \omega^2\ll\omega_M^2, \quad (38)$$

where

$$\bar{M}=(1-q)M+qM'$$

and

$$\omega_{\bar{M}}^2=4\gamma/\bar{M}\cong\omega_M^2(1+q\kappa) \quad (39)$$

to first order in q. At the upper end of the spectrum, on the other hand, the correction term predominates in (37). A singularity still occurs at $\omega^2=\omega_M^2$, but its strength is reduced from inverse square root to inverse fourth root. Thus, the high-frequency impurity modes may be said to be taken from the upper end of the unperturbed phonon spectrum. It turns out that the inverse fourth-root singularity is retained in the second-order corrections [as may be seen from examination of Eq. (59)]. However, it is likely that this singularity will disappear in a self-consistent calculation to all orders in the concentration q.

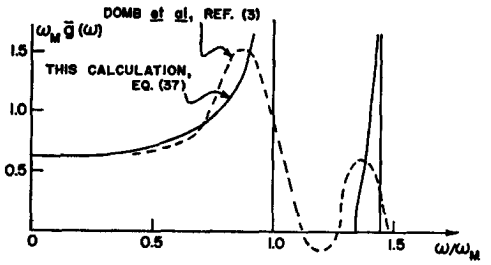


FIG. 4. The spectral function $\bar{g}$ as expressed in Eqs. (37) and (40) plotted for $q=\frac{1}{10}$ and $M=3M'$.

For $\omega^2>\omega_M^2$, we may write (37) in the form

$$\bar{g}(\omega)=\frac{2}{\pi}\operatorname{Re}\left[\omega_M^2-\omega^2+\frac{\omega_M^2 q\kappa(\omega^2-\omega_M^2)^{\frac{1}{2}}}{(\omega^2-\omega_M^2)^{\frac{1}{2}}-\kappa\omega}\right]^{-\frac{1}{2}}. \quad (40)$$

For values of ω between ω_M and ω_0, the frequency of the local mode, the quantity under the square root in (40) is negative and $\bar{g}$ is zero. For a small range of frequencies just above ω_0, however, this quantity becomes positive. Thus, for $M'<M$, the local mode singularity in $G_k^{(1)}(\omega^2)$ gives rise to an impurity band in $\bar{g}(\omega)$.

The spectrum as expressed in Eqs. (37) and (40) is plotted in Fig. 4 for the case $q=\frac{1}{10}$ and $M=3M'$.

V. SECOND-ORDER CORRECTIONS

As we have seen in the previous section, our "average" lattice will propagate phonons whose frequencies lie in a narrow band near the frequency of the local mode. The local mode frequency itself appears as a singular point in the function $G_k^{(1)}(\omega^2)$, the term in the phonon self-energy which is linear in q. Now, if we insert any finite number of light mass points into an infinitely long chain, we always find a number of new high-frequency normal modes of the chain. In the disordered lattice we then expect to find narrow impurity bands at each of these new frequencies. Furthermore, we expect the eigenfrequencies associated with a cluster of, say, n inserted light masses to appear as singular points of that term in the self-energy which is proportional to q^n, the probability of finding the cluster in the chain.

In this section we shall calculate the contribution to $\bar{g}(\omega)$ from clusters of two light masses. Before looking

into the details of this calculation, however, let us look at some of the properties of the isolated clusters. It will be worthwhile to look first at the case of a single light mass M' in a chain of masses M. If M' occurs at the position $l=0$, Eq. (4) becomes

$$(\omega_k^2-\omega^2)Q_k=-\sum_{k'}(\omega_{k'}^2/N)\lambda Q_{k'}. \tag{41}$$

On dividing (41) by $\omega_k^2-\omega^2$ and solving for $\sum_k \omega_k^2 Q_k$, we arrive at the dispersion relation

$$1+(\lambda/N)\sum_k[\omega_k^2/(\omega_k^2-\omega^2)]=0, \tag{42}$$

which is exactly the same as Eq. (30). The individual displacements associated with the local mode are

$$\begin{aligned} x_l &\sim \sum_k \frac{1}{\omega_k^2-\omega_0^2}\exp\left[\frac{2\pi ikl}{N}\right]\sim\int_{-\pi/2}^{\pi/2}\frac{e^{2ilx}}{\omega_M^2\sin^2x-\omega_0^2}dx \\ &\sim\frac{1}{\omega_0(\omega_0^2-\omega_M^2)^{\frac{1}{2}}}[\alpha(\omega_0/\omega_M)]^{|l|}, \quad l\neq 0, \end{aligned} \tag{43}$$

where

$$\alpha(t)=1-2t^2+2t(t^2-1)^{\frac{1}{2}}, \quad t=\omega_0/\omega_M, \tag{44}$$

and ω_0 is the high-frequency solution of (42). The integration involved in (43) is performed in the Appendix, where it is apparent that the magnitude of α always must be less than unity. Thus, we may express (43) in the form

$$x_l\sim\exp[-(|l|/l_0)], \quad l_0=-[\log|\alpha|]^{-1}. \tag{45}$$

For $M=3M'$, $\omega_0/\omega_M=1.34$, and $l_0=0.64$; i.e., the mode is highly localized for this mass ratio.

Next consider the case of two light masses inserted at, say, positions 0 and ν in an otherwise uniform chain. Equation (4) is now

$$(\omega_k^2-\omega^2)Q_k=-\sum_{k'}\frac{\lambda\omega_{k'}^2}{N}\left\{1+\exp\left[\frac{2\pi i(k'-k)\nu}{N}\right]\right\}Q_{k'}. \tag{46}$$

The new eigenfrequencies may be obtained from (46) by solving for

$$A=(1/N)\sum_k\omega_k^2Q_k$$

and

$$B=(1/N)\sum_k\omega_k^2\exp[2\pi ik\nu/N]Q_k.$$

We have

$$Q_k=-[\lambda/(\omega_k^2-\omega^2)][A+\exp(-2\pi ik\nu/N)B]. \tag{47}$$

Thus,

$$\begin{aligned} A&=-\lambda f(\omega,0)A-\lambda f(\omega,-\nu)B \\ B&=-\lambda f(\omega,\nu)A-\lambda f(\omega,0)B, \end{aligned} \tag{48}$$

where

$$\begin{aligned} f(\omega,\nu)&=\frac{1}{N}\sum_k\frac{\omega_k^2\exp(2\pi ik\nu/N)}{\omega_k^2-\omega^2} \\ &=\delta_{\nu,0}-\frac{\omega}{(\omega^2-\omega_M^2)^{\frac{1}{2}}}[\alpha(\omega/\omega_M)]^{|\nu|}. \end{aligned} \tag{49}$$

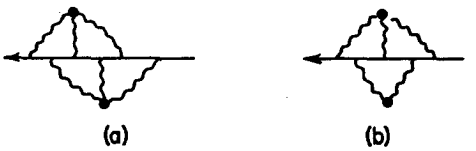


FIG. 5. The two kinds of reduced graphs which contribute to $G^{(2,2)}$.

The resulting secular equation is

$$[1+\lambda f(\omega,0)]^2-\lambda^2 f(\omega,\nu)f(\omega,-\nu)=0. \tag{50}$$

For $\nu=1$ (the two light masses next to each other) and $M=3M'$, (50) has roots at $\omega/\omega_M=1.06$ and 1.53. Since the single modes are so strongly localized, we may expect the terms in which the two light mass points are separated by one or more heavy masses ($|\nu|>1$) to contribute very little to the second-order spectrum.

On having tentatively identified the singularities in the second-order self-energy $G_k^{(2)}(\omega^2)$, we turn now to its detailed calculation. There are two different contributions to this second-order term. First, there are corrections of order q^2 to the graphs considered in the last section in which all the scattering events occurred at the same point in the lattice. That is, we now must include the q^2 term in $P_s(q)$ as given in Eq. (26). Second, there are the graphs, evaluated to lowest order in q, in which two different lattice points take part in the scattering. Let us denote these two corrections to the self-energy by $G^{(2,1)}$ and $G^{(2,2)}$, respectively.

$G^{(2,1)}$ is evaluated easily by using Eq. (26) and the analysis of Sec. IV. We have

$$\begin{aligned} G_k^{(2,1)}(\omega^2)&=q^2\omega_k^2\sum_{n=2}^{\infty}(-\lambda)^n(2^{n-1}-1)[f(\omega,0)]^{n-1} \\ &=q^2\omega_k^2\frac{2\lambda^2f(\omega,0)}{1+2\lambda f(\omega,0)}-q^2\omega_k^2\frac{\lambda^2f(\omega,0)}{1+\lambda f(\omega,0)}. \end{aligned} \tag{51}$$

[Compare this expression with Eq. (28).]

In order to evaluate $G^{(2,2)}$, we must sum all graphs of the sort shown in Fig. 1(d), remembering that the total momentum transfer at each of the two lattice points must be zero. The first step in evaluating this sum is to reduce each graph to a form similar to one or the other of the two graphs shown in Fig. 5. In this figure, each wavy line represents any number of consecutive interactions at a single lattice point. Mathematically, each wavy line indicates a factor

$$\begin{aligned} l_{k,k'}(\omega)&=\frac{\lambda\omega_{k'}^2}{N}-\frac{\lambda\omega_{k'}^2}{N^2}\sum_{k1}\frac{\omega_{k1}^2}{\omega_{k1}^2-\omega^2}+\cdots \\ &=\frac{\lambda\omega_{k'}^2}{N}\frac{1}{1+\lambda f(\omega,0)}\equiv\frac{\omega_{k'}^2}{N}t(\omega). \end{aligned} \tag{52}$$

The sum of all diagrams like Fig. 5(a) in which the same number of wavy-line interactions occur at each lattice point is given by

$$-N^2q^2\sum_{n=2}^{\infty}\sum_{k_1,\cdots k_{2n-1}}\frac{t_{k,k_1}(\omega)\cdots t_{k_{2n-1},k}(\omega)}{(\omega k_1{}^2-\omega^2)\cdots(\omega k_{2n-1}{}^2-\omega^2)}\times\delta_{k_1-k_2+k_3-\cdots+k_{2n-1}-k,\,0}. \quad (53)$$

Similarly, the sum of all diagrams like Fig. 5(b) is

$$+N^2q^2\sum_{n=1}^{\infty}\sum_{k_1,\cdots k_{2n}}\frac{t_{k,k_1}(\omega)\cdots t_{k_{2n},k}(\omega)}{(\omega k_1{}^2-\omega^2)\cdots(\omega k_{2n}{}^2-\omega^2)}\times\delta_{k_1-k_2+k_3-\cdots+k_{2n-1}-k_{2n},\,0}. \quad (54)$$

If we write the delta function in the form

$$\delta_{k,0}=\frac{1}{N}\sum_{\nu=-N/2}^{N/2}\exp\left(\frac{2\pi i\nu k}{N}\right), \quad (55)$$

then the sums over k_1, k_2, etc., factor and we have

$$\begin{aligned}G_k{}^{(2,2)}(\omega^2)&=-q^2\omega_k{}^2\sum_{n=2}^{\infty}\sum_{\nu}\exp\left(\frac{-2\pi i\nu k}{N}\right)\\&\quad\times[t(\omega)]^{2n}[f(\omega,\nu)]^n[f(\omega,-\nu)]^{n-1}\\&\quad+q^2\omega_k{}^2\sum_{n=1}^{\infty}\sum_{\nu}[t(\omega)]^{2n+1}[f(\omega,\nu)]^n[f(\omega,-\nu)]^n \quad (56)\\&=q^2\omega_k{}^2[t(\omega)]^3\sum_{\nu}\\&\quad\times\frac{f(\omega,\nu)f(\omega,-\nu)\{1-\exp(-2\pi i\nu k/N)t(\omega)f(\omega,\nu)\}}{1-[t(\omega)]^2f(\omega,\nu)f(\omega,-\nu)}.\end{aligned}$$

By referring to the definition of $t(\omega)$ in (52), we see that the poles of $G^{(2,2)}$ do, in fact, occur at the solutions of (50), and that the integer ν plays the same role in both equations. In (56), however, we have a contribution for $\nu=0$,

$$\begin{aligned}G^{(2,2)}(\omega^2)\,|_{\nu=0}&=q^2\omega_k{}^2[t(\omega)]^3\frac{[f(\omega,0)]^2}{1+t(\omega)f(\omega,0)}\\&=q^2\omega_k{}^2\frac{\lambda^3[f(\omega,0)]^2}{[1+\lambda f(\omega,0)]^2[1+2\lambda f(\omega,0)]}, \quad (57)\end{aligned}$$

which occurs because we summed over the two lattice points independently of each other in deriving this term. This mistake is corrected by including $G^{(2,1)}$ as given by (51).

$$\begin{aligned}G^{(2,1)}+G^{(2,2)}\,|_{\nu=0}&=q^2\omega_k{}^2\frac{\lambda^2f(\omega,0)}{[1+\lambda f(\omega,0)]^2}\\&=q^2\omega_k{}^2f(\omega,0)[t(\omega)]^2. \quad (58)\end{aligned}$$

Notice that the only remaining singularity in (58) is a second-order pole at the frequency of the isolated local mode. In fact, each nth-order correction to $G_k(\omega^2)$ will contain an nth-order pole at the local mode frequency; thus, the expansion in powers of q cannot converge very near $\omega=\omega_0$. This divergence implies a shift of the position of the singularity in $G_k(\omega^2)$ and a resulting displacement of the impurity band. This shift might be calculated by summing the most singular terms in $G_k(\omega^2)$ to all orders in q; but we shall not do this here. In any case, the shift must be small for small q and well-separated impurity bands.

On returning to (56), we see that the residue of any pole in $G^{(2,2)}$ contains a factor $[\alpha(\omega/\omega_M)]^{2\nu}$ arising from the factor $f(\omega,\nu)f(\omega,-\nu)$. As mentioned above, α is always less then unity, and is, in fact, of the order of 0.2 for the example calculated here. Thus, the pole strengths decrease rapidly as the separation of the light mass points increases. This justifies our earlier guess that well-separated impurities will not contribute much to the spectrum.

For reference, we write out a complete expression for the second-order self-energy:

$$\begin{aligned}G_k{}^{(2)}(\omega^2)&=q^2\omega_k{}^2\Big\{f(\omega,0)[t(\omega)]^2+2[t(\omega)]^3\\&\quad\times\sum_{\nu=1}^{N/2}\frac{[f(\omega,\nu)]^2[1-t(\omega)f(\omega,\nu)\cos(2\pi\nu k/N)]}{1-[t(\omega)f(\omega,\nu)]^2}\Big\}. \quad (59)\end{aligned}$$

VI. DISCUSSION

As shown in Fig. 4, the spectrum calculated here is in good qualitative agreement with the results of reference (3) for a case in which there is only a small concentration of light impurities and the mass ratio M/M' is relatively large. It should be mentioned that, in their published paper, Domb *et al.* indicate a second impurity band at $\omega/\omega_M\cong1.7$. An examination of their calculation, however, indicates that the moment inversion was converging very badly near the upper end of the spectrum; so we have omitted the second peak in Fig. 4.

There is good reason to believe that the calculation described in this paper is essentially exact for sufficiently small impurity concentrations q. In the case of light impurities, although the nth-order correction to $G_k(\omega^2)$ introduces a large number of infinite spikes in $\bar{g}(\omega)$, the area under each such spike is only of order q^n. In this sense, our series of approximations in which $G_k(\omega^2)$ was expanded in powers of q yields a convergent expression for $\bar{g}$, at least for small q.

On the other hand, it is apparent that our procedure does not converge for values of q near unity. In such a case it would be appropriate to expand $G_k(\omega^2)$ in powers of $1-q$, i.e., to think of inserting heavy impurities into a light lattice. In this case, the spectrum would be given by Eq. (37) with a negative value of

κ. The absence of spikes in this expression indicates that, at some value of q, there must be a radical change in the nature of $\bar{g}(\omega)$, and that this change must be associated with a divergence in our procedure. As q increases, the higher order spikes become more and more important, and the spectrum acquires a very complicated structure. The correct nth-order term in our expansion is roughly equal to the number of modes associated with all possible nth-order clusters of impurities multiplied by the average area under an nth-order spike ($\sim q^n$). When this sum of areas diverges, q is no longer a good expansion parameter. Note that the critical value of q depends upon the mass ratio. The number of possible nth-order clusters depends upon the number of heavy masses we can insert between two light ones before the coupling between the high-frequency modes becomes so small that the cluster is effectively broken. Thus, the contribution to $\bar{g}$ from the higher order clusters is larger when M/M' is close to unity and the local modes are not very well localized. Conversely, we expect the spikey structure of $\bar{g}$ to persist to higher impurity concentrations when the mass ratio is high. For example, Dean[5] finds an extremely complicated $\bar{g}$ in the case $M/M'=3$ and $q=0.5$. For actual calculation of spectra in such situations, Dean's technique is probably preferable to ours.

On returning to the simpler case of a small concentration of light impurities, it looks as if energy propagation in the impurity bands may be similar to a quantum-mechanical tunneling process. As we have seen in Sec. IV, the local modes are coupled weakly to the phonons. Via this coupling, the energy of excitation of one impurity must be able to "leak" across a region of heavy masses to the next impurity position. The average rate of energy transfer will be smaller the larger the separation between light masses. Such processes may have an effect on the transport properties of disordered lattices; it is hoped that the techniques devised in this paper may shed some light on this question.

Finally, it should be noted that there is no reason in principle why the present calculation cannot be generalized to apply to real three-dimensional crystals.

ACKNOWLEDGMENT

The author is greatly indebted to Dr. A. A. Maradudin for bringing this problem to his attention and for participating actively in many stages of the work.

APPENDIX. EVALUATION OF $f(\omega,\nu)$

The function $f(\omega,\nu)$ as defined in (49) and used at various places in this paper is

$$f(\omega,\nu)=\frac{1}{N}\sum_{k=-N/2}^{N/2}\frac{\omega_k^2\exp(2\pi ik\nu/N)}{\omega_k^2-\omega^2}. \tag{A1}$$

In the limit $N\to\infty$, this sum may be evaluated as in integral in the form

$$\begin{aligned} f(\omega,\nu)&=\delta_{\nu,0}+\frac{\omega^2}{N}\sum_k\frac{\exp(2\pi ik\nu/N)}{\omega_M^2\sin^2(\pi k/N)-\omega^2}\\ &=\delta_{\nu,0}+\frac{\omega^2}{\pi\omega_M^2}\int_{-\pi/2}^{\pi/2}\frac{e^{2i\nu x}}{\sin^2x-t^2}dx;\quad t=\omega/\omega_M. \end{aligned} \tag{A2}$$

With the transformation $z=e^{ix}$, the integral in (A2) becomes

$$\begin{aligned} h(t,\nu)&\equiv\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\frac{e^{2i\nu x}}{\sin^2x-t^2}dx\\ &=-\frac{1}{2\pi i}\oint_{\text{unit circle}}\frac{z^{2\nu-1}dz}{\frac{1}{4}[z-(1/z)]^2+t^2}\\ &=-\frac{2}{\pi i}\oint\frac{z^{2\nu+1}dz}{[z^2-\alpha_+(t)][z^2-\alpha_-(t)]}, \end{aligned} \tag{A3}$$

where

$$\alpha_\pm(t)=1-2t^2\pm2t(t^2-1)^{\frac{1}{2}}. \tag{A4}$$

In (A3) we may consider only positive integer values of ν since $f(\omega,\nu)$ is a symmetric function of ν; thus, the only singularities of the integrand in (A3) occur at $z^2=\alpha_\pm(t)$. The integral is nonvanishing only if the two functions $\alpha_\pm$ lie on opposite sides of the unit circle. Now $h(t,\nu)$ has a branch cut along the real axis for $-1<t<+1$. For $|t|>1$, it is apparent that $|\alpha_+|<1$, $|\alpha_-|>1$. Thus, $h(t,\nu)$ is given by the residues at $z^2=\alpha_+(t)$; and, by analytic continuation, the resulting expression for $h(t,\nu)$ must be valid throughout the physical sheet of the t plane. In this way we find

$$h(t,\nu)=-[1/t(t^2-1)^{\frac{1}{2}}][\alpha_+(t)]^{|\nu|}; \tag{A5}$$

and Eq. (48) follows immediately.

Exact Calculation of the Fluctuation Spectrum for a Nonlinear Model System

N. G. VAN KAMPEN
*Physics Section, Convair San Diego, San Diego, California**

(Received January 23, 1961)

Fluctuations in a circuit consisting of a diode and a condenser have been treated by means of approximate methods [C. T. J. Alkemade, Physica **24**, 1029 (1958); N. G. van Kampen, *ibid.* **26**, 585 (1960)]. In the present paper, exact eigenfunctions and eigenvalues of the master equation for this system are obtained, and the spectral density of the equilibrium fluctuations is calculated. The most striking result is that the spectrum of the eigenvalues (which are reciprocal relaxation times) has an accumulation point, corresponding to the average time interval between two successive electron transitions.

1. INTRODUCTION

IN this paper an exact calculation is presented of the spectral density of the equilibrium fluctuations in a simple model of a nonlinear system. The system is an electrical circuit consisting of a condenser and a diode in contact with a heat reservoir. It has been shown[1] that the electrical fluctuations in this system are governed by the master equation

$$\partial P(N)/\partial t = P(N+1)-P(N) + \zeta\{e^{-\epsilon(N-1)}P(N-1)-e^{-\epsilon N}P(N)\}. \quad (1.1)$$

$P(N)$ is the probability of having N electrons on the left-hand condenser plate (see Fig. 1); $\epsilon = e^2/kTC$, e being the electron charge, C the capacity of the condenser, T the temperature of the whole system, k Boltzmann's constant;

$$\zeta = \exp\frac{W_1 - W_2 - e^2/2C}{kT},$$

where W_1 and W_2 are the work functions of the two electrodes; t is time measured in appropriate units.

The reasons for this investigation are twofold.

Firstly, the problem of fluctuations in nonlinear systems has been studied by several authors,[1-10] but no agreement has yet been reached. It therefore seems worthwhile to obtain an explicit and rigorous result for a nontrivial example. In particular, it will be shown that there are terms in the spectral density that cannot be obtained by the usual expansion methods.

Secondly, the model illustrates the distinction between slow fluctuations, which involve large numbers of particles, and rapid fluctuations connected with individual particles. The relaxation times of the slow fluctuations are mainly determined by the RC time of the circuit; they are the ones that are found by the usual approximate methods. The relaxation times of the rapid fluctuations are determined by the average time between two successive individual electron transitions. The two kinds of fluctuations are separated by an accumulation point in the spectrum of relaxation times. It seems likely that such a distinction is a general feature of fluctuations in nonlinear systems. In linear systems, of course, there is just one relaxation time.

The mathematical problem consists of finding the eigenvalues and eigenfunctions of the difference equation (1.1), together with their relevant properties. Although the mathematics is closely related to the theory of "q-difference equations",[11] some of the results appear to be new and may have some interest by themselves.

As much of the mathematical work as seemed possible has been removed from the text to a series of appendixes.

2. PRELIMINARIES

Equation (1.1) can be written more simply

$$\partial P(N)/\partial t = (\mathbf{E}-1)P(N)+\zeta(\mathbf{E}^{-1}-1)e^{-\epsilon N}P(N),$$

where $\mathbf{E}$ is the operator defined by $\mathbf{E}f(N)=f(N+1)$. This can be further simplified by replacing N with a new variable s

$$N = s+g, \quad P(N) \to P(s),$$

and choosing the constant g such that $\zeta e^{-\epsilon g}=1$. Then,

$$\partial P(s)/\partial t = (\mathbf{E}-1)P(s)+(\mathbf{E}^{-1}-1)e^{-\epsilon s}P(s). \quad (2.1)$$

Unless g happens to be an integer, s runs over a set of noninteger values. Let γ denote the distance of g to the

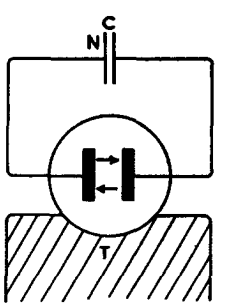


FIG. 1.

* On leave of absence from the University at Utrecht, Netherlands.

[1] N. G. van Kampen, Physica **26**, 585 (1960).
[2] D. K. C. MacDonald, Phil. Mag. **45**, 63 (1954); D. Polder, *ibid.* **45**, 69 (1954); D. K. C. MacDonald, Phys. Rev. **108**, 541 (1957).
[3] N. G. van Kampen, Phys. Rev. **110**, 319 (1958).
[4] C. T. J. Alkemade, Physica **24**, 1029 (1958).
[5] R. O. Davies, Physica **24**, 1055 (1958).
[6] W. Bernard and H. B. Callen, Revs. Modern Phys. **31**, 1017 (1959); Phys. Rev. **118**, 1466 (1960).
[7] A. Marek, Physica **25**, 1358 (1959).
[8] M. Lax, Revs. Modern Phys. **32**, 25 (1960).
[9] A. Siegel, J. Math. Phys. **1**, 378 (1960).
[10] N. G. van Kampen, Can. J. Phys. **39**, 551 (1961).
[11] W. Hahn, Mathematische Nachrichten **2**, 4 (1949).

next higher integer[12]:

$$g+\gamma=\text{integer}, \quad 0\leqslant\gamma<1.$$

Then the set of values available to s is

$$\mathcal{S}_\gamma=(\cdots,-2+\gamma,-1+\gamma,\gamma,1+\gamma,2+\gamma,3+\gamma,\cdots).$$

Summation over all values $s\in\mathcal{S}_\gamma$ will be denoted by $\mathbf{S}$.

One readily verifies that one solution of (2.1) is

$$\Psi_0(s)=\exp(-\tfrac{1}{2}\epsilon s^2+\tfrac{1}{2}\epsilon s). \tag{2.2}$$

For future use, we note the property

$$\mathbf{E}\Psi_0=e^{-\epsilon s}\Psi_0. \tag{2.3}$$

We introduce a Hilbert space $\mathfrak{H}_\gamma$ of real functions $\psi(s)$ with $s\in\mathcal{S}_\gamma$ by defining the scalar product,

$$(\psi_1,\psi_2)=\mathbf{S}\frac{\psi_1(s)\psi_2(s)}{\Psi_0(s)}.$$

Ψ_0 itself belongs to $\mathfrak{H}_\gamma$, the square of its norm being

$$\begin{aligned}&\mathbf{S}\exp(-\tfrac{1}{2}\epsilon s^2+\tfrac{1}{2}\epsilon s)\\&=\exp(-\tfrac{1}{2}\epsilon\gamma^2+\tfrac{1}{2}\epsilon\gamma)\sum_{-\infty}^{+\infty}\exp[-\tfrac{1}{2}\epsilon N^2+(\tfrac{1}{2}-\gamma)\epsilon N]\\&=\exp(-\tfrac{1}{2}\epsilon\gamma^2+\tfrac{1}{2}\epsilon\gamma)\vartheta_3(\tfrac{1}{2}i\epsilon(\gamma-\tfrac{1}{2})).\end{aligned} \tag{2.4}$$

Here ϑ_3 denotes a theta function.[13] The quantity (2.4) is the partition function of the equilibrium distribution and will be denoted by Z.

It is convenient to introduce also the dual space $\tilde{\mathfrak{H}}_\gamma$ consisting of the functions

$$\phi(s)=\psi(s)/\Psi_0(s), \quad \psi(s)\in\mathfrak{H}_\gamma,$$

with the scalar product

$$(\phi_1,\phi_2)=(\psi_1,\psi_2)=\mathbf{S}\Psi_0\phi_1\phi_2.$$

In particular,

$$\Phi_0(s)=1.$$

The operator on the right of (2.1) will be denoted by $\mathbf{F}$,

$$\mathbf{F}=\mathbf{E}-1+(\mathbf{E}^{-1}-1)e^{-\epsilon s}.$$

It has an eigenvalue zero with eigenfunction Ψ_0. We show that it is negative definite for all other functions ψ (i.e., in the subspace orthogonal to Ψ_0). By using (2.3) and some obvious properties of the operator E, one finds

$$\begin{aligned}\mathbf{S}\Psi_0^{-1}\psi\mathbf{F}\psi&=\mathbf{S}\phi(\mathbf{E}-1)\Psi_0\phi+\mathbf{S}\phi(\mathbf{E}^{-1}-1)e^{-\epsilon s}\Psi_0\phi\\&=\mathbf{S}\Psi_0\phi(\mathbf{E}^{-1}-1)\phi+\mathbf{S}\phi(\mathbf{E}^{-1}-1)(\mathbf{E}\Psi_0)\phi\\&=\mathbf{S}\Psi_0\phi(\mathbf{E}^{-1}-1)\phi+\mathbf{S}\phi(1-\mathbf{E})\Psi_0\mathbf{E}^{-1}\phi\\&=-\mathbf{S}\Psi_0[(1-\mathbf{E}^{-1})\phi]^2.\end{aligned} \tag{2.5}$$

This is clearly negative, unless ϕ is a constant and hence proportional to Φ_0.

We also shall have to consider the operator

$$\tilde{\mathbf{F}}=\mathbf{E}^{-1}-1+e^{-\epsilon s}(\mathbf{E}-1),$$

with the properties

$$\mathbf{S}\phi\mathbf{F}\psi=\mathbf{S}\psi\tilde{\mathbf{F}}\phi,$$

$$\mathbf{F}\Psi_0\phi=\Psi_0\tilde{\mathbf{F}}\phi.$$

These relations show that $\tilde{\mathbf{F}}$ is the operator that corresponds to $\mathbf{F}$ in the dual space.

The solution of the time-dependent equation (2.1) is equivalent to solving the eigenvalue problem

$$\mathbf{F}\psi=-\lambda\psi,$$

or alternatively,

$$\tilde{\mathbf{F}}\phi=-\lambda\phi. \tag{2.6}$$

We shall indeed solve this last equation, but first an auxiliary function has to be introduced.

3. AUXILIARY FUNCTION $\pi(z)$

Definition:

$$\pi(z)\equiv\prod_{k=1}^{\infty}\frac{1-e^{-\epsilon k}}{1-e^{-\epsilon(k+z)}}.$$

Fundamental property:

$$\pi(z)=(1-e^{-\epsilon z})\pi(z-1). \tag{3.1}$$

From this follows, as $\pi(0)=1$,

$$\pi(m)=\prod_{k=1}^{m}(1-e^{-\epsilon k}) \quad \text{for} \quad m=1,2,\cdots.$$

$\pi(z)$ is periodic with period $2\pi i/\epsilon$. It has simple poles at $z=-m$ $(m=1,2,\cdots)$, the residues being given by

$$\lim_{z\to m}(z+m)\pi(z)=\frac{(-1)^{m+1}\exp(-\tfrac{1}{2}\epsilon m^2+\tfrac{1}{2}\epsilon m)}{\epsilon\pi(m-1)}.$$

There are of course additional poles at $z=-m+2\pi im'/\epsilon$, where $m'=\pm1,\pm2,\cdots$. Otherwise $\pi(z)$ is regular and it has no zeros. For all z and all integral m,

$$\frac{\pi(z-m)\pi(-z-1+m)}{\pi(z)\pi(-z-1)}=(-1)^m\exp[-\tfrac{1}{2}\epsilon m^2+\epsilon(z+\tfrac{1}{2})m]. \tag{3.2}$$

As $\operatorname{Re} z\to\infty$, $\pi(z)$ tends to a constant G,

$$\lim\pi(z)=\prod_{k=1}^{\infty}(1-e^{-\epsilon k})\equiv G. \tag{3.3}$$

[12] The introduction of γ as distinct from g is convenient but not strictly necessary.

[13] For the theta functions, we follow the notation of E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, New York, 1946), Chap. 21. The parameter q of all theta functions in this paper has the value $e^{-\frac{1}{2}\epsilon}$ (and will, therefore, not be written explicitly). Accordingly, the quasi-period is $\frac{1}{2}i\epsilon$.

This is the same constant G that occurs in the theory of theta functions.[13] As $\mathrm{Re}z \to -\infty$ avoiding the poles, $\pi(z)$ tends to zero. More specifically, if $m \to \infty$ through integral values,

$$\pi(z-m) \simeq \frac{\pi(z)\pi(-z-1)}{G}(-1)^m \times \exp[-\tfrac{1}{2}\epsilon m^2 + \epsilon(z+\tfrac{1}{2})m]. \quad (3.4)$$

Finally,

$$\pi(z)\pi(-1-z) = G^3/\vartheta_4(\tfrac{1}{2}i\epsilon z + \tfrac{1}{4}i\epsilon). \quad (3.5)$$

It is convenient to define as a separate function

$$\pi_1(z) = \pi(z + \pi i/\epsilon),$$

with the fundamental property

$$\pi_1(z) = (1+e^{-\epsilon z})\pi_1(z-1). \quad (3.6)$$

$\pi_1(z)$ has no singularities on the real axis. Equation (3.3) yields

$$\frac{\pi_1(z-m)\pi_1(-z-1+m)}{\pi_1(z)\pi_1(-z-1)} = \exp[-\tfrac{1}{2}\epsilon m^2 + \epsilon(z+\tfrac{1}{2})m], \quad (3.7)$$

while (3.4) becomes

$$\pi_1(z-m) \simeq \frac{\pi_1(z)\pi_1(-z-1)}{G} \times \exp[-\tfrac{1}{2}\epsilon m^2 + \epsilon(z+\tfrac{1}{2})m]. \quad (m \to \infty). \quad (3.8)$$

The so-called q-binomial coefficients[14,15] can be defined in terms of $\pi(z)$ by

$$\begin{bmatrix} z \\ y \end{bmatrix} \equiv \frac{\pi(z)}{\pi(y)\pi(z-y)} = \begin{bmatrix} z \\ z-y \end{bmatrix}. \quad (3.9)$$

This is an entire analytic function of y, which equals 1 for $y=0$, and vanishes for $y=-1, -2, \cdots$. From the fundamental property (3.1) of $\pi(z)$, the following three identities are obtained.

$$\begin{bmatrix} z \\ y \end{bmatrix} = \frac{1-e^{-\epsilon z}}{1-e^{-\epsilon y}} \begin{bmatrix} z-1 \\ y-1 \end{bmatrix}, \quad (3.10)$$

$$\begin{bmatrix} z \\ y \end{bmatrix} = \frac{1-e^{-\epsilon z}}{1-e^{-\epsilon z+\epsilon y}} \begin{bmatrix} z-1 \\ y \end{bmatrix}, \quad (3.11)$$

$$\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} z-1 \\ y \end{bmatrix} + e^{-\epsilon z+\epsilon y} \begin{bmatrix} z-1 \\ y-1 \end{bmatrix}. \quad (3.12)$$

From (3.2) follows

$$\begin{bmatrix} z \\ y+m \end{bmatrix} \begin{bmatrix} y+m \\ z+1 \end{bmatrix} = \begin{bmatrix} z \\ y \end{bmatrix} \begin{bmatrix} y \\ z+1 \end{bmatrix} (-1)^m \times \exp[\tfrac{1}{2}\epsilon m^2 - \epsilon(z-y+\tfrac{1}{2})m]. \quad (3.13)$$

From (3.4) one finds for $m \to +\infty$

$$\begin{bmatrix} z \\ y+m \end{bmatrix} \simeq \frac{\pi(z)}{\pi(z-y)\pi(y-z-1)}(-1)^m \times \exp[\tfrac{1}{2}\epsilon m^2 - \epsilon(z-y+\tfrac{1}{2})m]; \quad (3.14)$$

and for $m \to -\infty$

$$\begin{bmatrix} z \\ y+m \end{bmatrix} \simeq \frac{\pi(z)}{\pi(y)\pi(-y-1)}(-1)^m \times \exp[\tfrac{1}{2}\epsilon m^2 + \epsilon(y+\tfrac{1}{2})m]. \quad (3.15)$$

It is convenient to define separately

$$\begin{bmatrix} z \\ y \end{bmatrix}_1 \equiv \frac{\pi_1(z)}{\pi(y)\pi_1(z-y)} = \begin{bmatrix} z+\pi i/\epsilon \\ y \end{bmatrix}.$$

4. SOLUTION OF EQUATION (2.6)

In order to find solutions of the difference equation (2.6),

$$0 = (\tilde{\mathbf{F}} + \lambda)\phi = \{\mathbf{E}^{-1} - 1 + \lambda + e^{-\epsilon s}(\mathbf{E}-1)\}\phi, \quad (4.1)$$

we put

$$\phi(s) = \sum_{-\infty}^{+\infty} a_\nu e^{\epsilon(\nu+\sigma)s}.$$

a_ν and σ are yet to be determined. Clearly σ may be restricted by

$$0 \leqslant \mathrm{Re}\sigma < 1, \quad 0 \leqslant \mathrm{Im}\sigma < 2\pi/\epsilon. \quad (4.2)$$

Furthermore, we put[16]

$$\lambda = 1 - e^{-\epsilon\tau}, \quad 0 \leqslant \mathrm{Im}\tau < 2\pi/\epsilon. \quad (4.3)$$

Substituting this $\phi(s)$ into the equation, one obtains a two-term recursion formula for the a_ν,

$$e^{-\epsilon\tau}\{e^{-\epsilon(\nu+\sigma-\tau)} - 1\}a_\nu + e^{\epsilon(\nu+\sigma+1)}\{1 - e^{-\epsilon(\nu+\sigma+1)}\}a_{\nu+1} = 0.$$

This may be written in the form

$$-\frac{\pi(\nu+\sigma)}{\pi(\nu+\sigma-\tau-1)}a_\nu + e^{\epsilon(\nu+\sigma+\tau+1)}\frac{\pi(\nu+\sigma+1)}{\pi(\nu+\sigma-\tau)}a_{\nu+1} = 0.$$

It then follows immediately that

$$a_\nu = c\frac{\pi(\nu+\sigma-\tau-1)}{\pi(\nu+\sigma)} \exp[-\tfrac{1}{2}\epsilon\nu^2 - \epsilon(\sigma+\tau+\tfrac{1}{2})\nu],$$

[14] C. F. Gauss, Werke II (Göttingen, 1863) 11, in particular p. 16.

[15] G. Szegö, *Orthogonal Polynomials* (American Mathematical Society, New York, 1959), p. 33.

[16] The value $\lambda = 1$ would have to be studied separately. However, in Sec. 8 is shown that it cannot contribute to the density spectrum of the fluctuations.

with c independent of ν. On account of (3.2), one also has

$$a_\nu = c'(-1)^\nu \frac{\pi(\tau)}{\pi(\nu+\sigma)\pi(\tau-\sigma-\nu)} \exp(-\epsilon\nu^2-2\epsilon\sigma\nu),$$

$$= c'(-1)^\nu \begin{bmatrix} \tau \\ \nu+\sigma \end{bmatrix} \exp(-\epsilon\nu^2-2\epsilon\sigma\nu).$$

We thus have obtained the following general solution of (4.1):

$$\Phi(s;\sigma,\tau) \equiv \sum_{\nu=-\infty}^{+\infty} (-1)^\nu \begin{bmatrix} \tau \\ \nu+\sigma \end{bmatrix} \times \exp[-\epsilon\nu^2-2\epsilon\sigma\nu+\epsilon(\nu+\sigma)s], \quad (4.4)$$

τ being related to λ by (4.3). It follows from (3.3) and (3.4) that this series converges for all σ and τ, unless τ is a negative integer. If σ is an integer, all terms with $\nu<\sigma$ are zero; if $\tau-\sigma$ is an integer, all terms with $\nu>\tau-\sigma$ are zero.

$\Phi(s;\sigma,\tau)$ has a number of obvious periodicity properties with respect to σ and τ, but they are immaterial because of the restrictions (4.2) and (4.3). However, there is also a quasi-periodicity in τ, which is expressed by the following important recursion relations. Firstly, using (3.11) one finds

$$(1-e^{-\epsilon\tau}\mathbf{E})\Phi(s;\sigma,\tau) = \frac{\pi(\tau)}{\pi(\tau-1)}\Phi(s;\sigma,\tau-1). \quad (4.5)$$

Secondly, introducing the dual function

$$\Psi(s;\sigma,\tau) = \Psi_0(s)\Phi(s;\sigma,\tau),$$

one finds with the use of (3.12)

$$(1-e^{-\epsilon\tau}\mathbf{E}^{-1})\Psi(s;\sigma,\tau) = \Psi(s;\sigma,\tau+1). \quad (4.6)$$

Thirdly, either by using (3.10) or by using (4.5) and the difference equation (4.1) itself,

$$(1-\mathbf{E}^{-1})\Phi(s;\sigma,\tau) = -(1-e^{-\epsilon\tau})e^{\epsilon(s-1)}\mathbf{E}^{-2}\Phi(s;\sigma,\tau-1). \quad (4.7)$$

5. NORMALIZATION CONDITION

In order to determine the eigenfunctions of $\tilde{\mathbf{F}}$, those values of σ and τ have to be selected for which the normalization condition

$$\mathbf{S}\Psi_0(s)[\Phi(s;\sigma,\tau)]^2 < \infty \quad (5.1)$$

is satisfied. If σ is an integer (necessarily zero), and $\tau=n$ $(=0, 1, 2, \cdots)$, the series (4.4) breaks off and reduces to a polynomial in $e^{\epsilon s}$ of degree n. Hence, (5.1) is satisfied. This gives us a first type of eigenfunctions:

$$\Phi_n^{(1)}(s) \equiv \Phi(s;0,n) = \sum_{\nu=0}^{n} (-1)^\nu \begin{bmatrix} n \\ \nu \end{bmatrix} \exp(-\epsilon\nu^2+\epsilon\nu s). \quad (5.2)$$

These polynomials have been studied in the mathematical literature.[15,17]

In order to find the other types of eigenfunctions, the behavior of $\Phi(s;\sigma,\tau)$ for $s\to\pm\infty$ has now to be investigated.

Behavior for $s \to \infty$

If $\sigma\neq 0$, the series does not break off for negative ν. Thus, for $s\to-\infty$, the terms with large negative ν will predominate, so that the asymptotic expression (3.15) may be inserted. Thus,

$$\Phi(s\to-\infty;\sigma,\tau) \sim \sum_{-\infty}^{+\infty} \exp[-\tfrac{1}{2}\epsilon\nu^2-\epsilon(\sigma-\tfrac{1}{2}-s)\nu+\epsilon\sigma s].$$

The sum on the right is

$$e^{\epsilon\sigma s}\vartheta_3(\tfrac{1}{2}i\epsilon(\sigma-\tfrac{1}{2}-s)).$$

On account of the quasi-periodicity of the theta function, this is equal to

$$e^{\sigma s}\vartheta_3(\tfrac{1}{2}i\epsilon(\sigma-\tfrac{1}{2}-s+1))e^{-\epsilon(\sigma-s)}$$

$$= e^{\sigma s}\vartheta_3(\tfrac{1}{2}i\epsilon(\sigma-\tfrac{1}{2}-\gamma)) \prod_{k=0}^{s-\gamma-1} e^{-\epsilon(\sigma-s+k)}$$

$$= \vartheta_3(\tfrac{1}{2}i\epsilon(\sigma-\tfrac{1}{2}-\gamma)) \exp[\tfrac{1}{2}\epsilon s^2+\tfrac{1}{2}\epsilon s-\tfrac{1}{2}\epsilon\gamma^2+\epsilon\gamma(\sigma-\tfrac{1}{2})].$$

Because of the factor $\exp(\frac{1}{2}\epsilon s^2)$, this asymptotic behavior is irreconcilable with the normalization condition (5.1), *unless the theta function is zero.* Hence, $\frac{1}{2}i\epsilon(\sigma-\frac{1}{2}-\gamma)$ has to coincide with one of the zeros of ϑ_3,

$$\tfrac{1}{2}i\epsilon(\sigma-\tfrac{1}{2}-\gamma) = (m'+\tfrac{1}{2})\pi+(n'+\tfrac{1}{2})(i\epsilon/2),$$

with any integers m', n'. The only solution within the restriction (4.2) is

$$\sigma = \gamma+\pi i/\epsilon.$$

Behavior for $s \to +\infty$

If $\tau-\sigma$ is not an integer, the series does not break off for $\nu\to+\infty$. Then, for $s\to+\infty$, the terms with large positive ν will predominate, so that the asymptotic expression (3.14) may be used. Hence

$$\Phi(s\to+\infty;\sigma,\tau) \sim \sum_{-\infty}^{+\infty} \exp[-\tfrac{1}{2}\epsilon\nu^2-\epsilon(\sigma+\tau+\tfrac{1}{2}-s)\nu+\epsilon\sigma s].$$

This can be worked out like the previous example, with the result that the normalization condition can only be

[17] S. Wigert, Arkiv Mat. Astron. Fysik **17**, No. 18 (1923); G. Szegö, Sitzber. Preuss. Akad. Wiss. Physik.-Math. Kl. (1926) 242; L. Carlitz, Ann. Mat. Pura Appl. (4) **41**, 359 (1956). Szegö's polynomials $K_n(\xi,q)$ are related to our $\Phi_n^{(1)}$ by

$$\Phi_n^{(1)}(s) = K_n(-e^{\epsilon s+\epsilon}, e^{-\epsilon}).$$

It should be noted that his orthogonality relation involves an integral, whereas our Eq. (7.5) involves a summation over S_γ.

satisfied if

$$\tfrac{1}{2}i\epsilon(\sigma+\tau+\tfrac{1}{2}-\gamma)=(m''+\tfrac{1}{2})\pi+(n''+\tfrac{1}{2})(i\epsilon/2).$$

This yields a denumerable set of solutions, viz.,

$$\sigma+\tau=n''+\gamma+\pi i/\epsilon, \quad (n''=0, \pm1, \pm2, \cdots).$$

6. VARIOUS TYPES OF EIGENFUNCTIONS

Summarizing the results, one finds four possible sets of values for σ and τ.

(1) $\sigma=0, \quad \tau=n \quad (n=0, 1, 2, \cdots);$

(2) $\sigma=0, \quad \tau=n+\gamma+\pi i/\epsilon \quad (n=\cdots, -2, -1, 0, 1, 2, \cdots);$

(3) $\sigma=\gamma+\pi i/\epsilon, \quad \tau=n+\gamma+\pi i/\epsilon \quad (n=\cdots, -2, -1, 0, 1, 2, \cdots);$

(4) $\sigma=\gamma+\pi i/\epsilon, \quad \tau=n \quad (n=\cdots, -2, -1, 0, 1, 2, \cdots).$

They give rise to the following four possible types of eigenfunctions.

First type: polynomials in $e^{\epsilon s}$, see (5.2).

$$\Phi_n^{(1)}(s)\equiv\Phi(s; 0, n)=\sum_{\nu=0}^{n}(-1)^\nu\begin{bmatrix}n\\ \nu\end{bmatrix}\exp(-\epsilon\nu^2+\epsilon\nu s);$$

$$\lambda_n^{(1)}=1-e^{-\epsilon n} \quad (n=0, 1, 2, \cdots).$$

Second type:

$$\Phi_n^{(2)}(s)\equiv\Phi(s; 0, n+\gamma+\pi i/\epsilon)$$

$$=\sum_{\nu=0}^{\infty}(-1)^\nu\begin{bmatrix}n+\gamma\\ \nu\end{bmatrix}_1\exp(-\epsilon\nu^2+\epsilon\nu s); \quad (6.1)$$

$$\lambda_n^{(2)}=1+e^{-\epsilon(n+\gamma)} \quad (n=\cdots, -2, -1, 0, 1, 2, \cdots).$$

Third type:

$$\Phi_n^{(3)}(s)\equiv\Phi(s; \gamma+\pi i/\epsilon, n+\gamma+\pi i/\epsilon)$$

$$=\exp[(\epsilon\gamma+\pi i)s]\sum_{-\infty}^{n}(-1)^\nu\frac{\pi_1(n+\gamma)}{\pi_1(\gamma+\nu)\pi(n-\nu)}$$

$$\times\exp(-\epsilon\nu^2-2\epsilon\gamma\nu+\epsilon\nu s); \quad (6.2)$$

$$\lambda_n^{(3)}=1+e^{-\epsilon(n+\gamma)} \quad (n=\cdots, -2, -1, 0, 1, 2, \cdots).$$

An alternative form is

$$\Phi_n^{(3)}(s)=(-1)^n\exp(-\epsilon n^2-2\epsilon\gamma n)$$

$$\times\exp[(\epsilon n+\epsilon\gamma+\pi i)s]\sum_{\nu=0}^{\infty}(-1)^\nu\begin{bmatrix}n+\gamma\\ \nu\end{bmatrix}_1$$

$$\times\exp[-\epsilon\nu^2+2\epsilon(n+\gamma)\nu-\epsilon\nu s]. \quad (6.3)$$

Fourth type:

$$\Phi_n^{(4)}(s)\equiv\Phi(s; \gamma+\pi i/\epsilon, n)$$

$$=\exp[(\epsilon\gamma+\pi i)s]\sum_{-\infty}^{+\infty}(-1)^\nu\frac{\pi(n)}{\pi_1(\nu+\gamma)\pi_1(n-\nu-\gamma)}$$

$$\times\exp(-\epsilon\nu^2-2\epsilon\gamma\nu+\epsilon\nu s). \quad (6.4)$$

$$\lambda_n^{(4)}=1-e^{-\epsilon n} \quad (n=0, 1, 2, \cdots).$$

For negative n this expression is meaningless because $\pi(n)$ is infinite. We therefore define for $n<0$

$$\Phi_n^{(4)}(s)\equiv\lim_{\tau\to n}\frac{\Phi(s; \gamma+\pi i/\epsilon, \tau)}{\pi(\tau)}$$

$$=\exp[(\epsilon\gamma+\pi i)s]\sum_{-\infty}^{+\infty}\frac{(-1)^\nu}{\pi_1(\nu+\gamma)\pi_1(n-\nu-\gamma)}$$

$$\times\exp(-\epsilon\nu^2-2\epsilon\gamma\nu+\epsilon\nu s). \quad (6.5)$$

$$\lambda_n^{(4)}=1-e^{-\epsilon n}, \quad (n=\cdots, -2, -1, 0).$$

For $n=0$, either (6.4) or (6.5) may be used.

In Appendix A, it is proved that $\Phi_n^{(4)}$ for $n<0$ vanishes identically (i.e., for all $s\in\mathcal{S}_\gamma$), and that $\Phi_n^{(4)}$ for $n\geqslant 0$ is proportional to $\Phi_n^{(1)}$. This shows that the eigenvalues $\lambda_n^{(4)}$ for negative n are spurious. Moreover, it is proved in Appendix B that the $\Phi_n^{(3)}$ are proportional to the $\Phi_n^{(2)}$. Hence, only the first two types remain, and all eigenvalues are nondegenerate.

It still has to be verified that $\Phi_n^{(2)}$ actually does satisfy the normalization condition. For this, it is sufficient to note that from (6.1) follows

$$\Phi_n^{(2)}(s\to-\infty)=1+O(e^{\epsilon s}),$$

and from (6.3), with the aid of (B.3),

$$\Phi_n^{(2)}(s\to+\infty)=O(e^{\epsilon(n+\gamma)s}).$$

7. NORMALIZATION

In Appendix C, the following identities are proved

$$\mathbf{S}\Psi_0[\Phi_n^{(1)}]^2=\pi(n)e^{\epsilon n}Z, \quad (7.1)$$

$$\mathbf{S}\Psi_0[\Phi_n^{(2)}]^2=\frac{\pi_1(n+\gamma)}{\pi_1(\gamma)}e^{\epsilon n}Z'. \quad (7.2)$$

Here Z is the quantity defined by (2.4) and Z' is an abbreviation for

$$Z'\equiv\mathbf{S}\Psi_0(\Phi_0^{(2)})^2, \quad (7.3)$$

which is computed in Appendix E to be

$$Z'=\exp(-\tfrac{1}{2}\epsilon\gamma^2+\tfrac{1}{2}\epsilon\gamma)\frac{G^3}{\pi_1(-\gamma-1)}. \quad (7.4)$$

It is convenient to define normalized eigenfunctions

Ψ_0 and Φ_n in the following way:

$$\bar{\Psi}_0 \equiv Z^{-1}\Psi_0,$$

$$\bar{\Phi}_n^{(1)} \equiv [\pi(n)e^{\epsilon n}]^{-\frac{1}{2}}\Phi_n^{(1)},$$

$$\bar{\Phi}_n^{(2)} \equiv \frac{Z'}{Z}\left[\frac{\pi_1(n+\gamma)}{\pi_1(\gamma)}e^{\epsilon n}\right]^{-\frac{1}{2}}\Phi_n^{(2)}.$$

The general orthonormality relation is

$$\mathbf{S}\bar{\Psi}_0\bar{\Phi}_n^{(i)}\bar{\Phi}_n^{(j)} = \delta_{ij}\delta_{nm}. \tag{7.5}$$

Accordingly, we also put

$$\bar{\Psi}_n^{(i)} \equiv \bar{\Psi}_0\bar{\Phi}_n^{(i)}.$$

A special example of (7.5) is

$$\mathbf{S}\bar{\Psi}_n^{(i)} = \delta_{i1}\delta_{n0}. \tag{7.6}$$

We here mention a result, which is obtained by similar methods in Appendix D:

$$\mathbf{S}s\Psi_n^{(1)} = -\pi(n-1)Z, \quad (n \geqslant 1) \tag{7.7}$$

$$\mathbf{S}s\Psi_n^{(2)} = \frac{\pi_1(n+\gamma-1)}{\pi_1(\gamma-1)}Z''. \tag{7.8}$$

Here Z'' is a third constant defined by

$$Z'' \equiv \mathbf{S}s\Psi_0^{(2)}. \tag{7.9}$$

It is shown in Appendix F that

$$Z'' = -\exp(-\tfrac{1}{2}\epsilon\gamma^2+\tfrac{1}{2}\epsilon\gamma)\frac{G^3}{\pi_1(-\gamma)} = -\frac{Z'}{1-e^{\epsilon\gamma}}. \tag{7.10}$$

It should be noted that, on account of (7.6), the relations (7.7) and (7.8) remain valid if the factor s on the left is replaced with $N=s+g$.

8. FLUCTUATION SPECTRUM

The spectral density of the fluctuations in the number N of electrons is given by[3]

$$S_N(\omega) = \frac{2}{\pi}\sum\frac{\lambda}{\lambda^2+\omega^2}\{\mathbf{S}N\bar{\Psi}\}^2, \tag{8.1}$$

the summation extending over all normalized eigenfunctions $\bar{\Psi}$ with their eigenvalues λ. Inserting the results of the previous section, one gets

$$S_N(\omega) = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1-e^{-\epsilon n}}{(1-e^{-\epsilon n})^2+\omega^2}\frac{\{\pi(n-1)\}^2}{\pi(n)e^{\epsilon n}}$$
$$+\frac{2}{\pi}\sum_{-\infty}^{+\infty}\frac{1+e^{-\epsilon(n+\gamma)}}{(1+e^{-\epsilon(n+\gamma)})^2+\omega^2}\left\{\frac{\pi_1(n+\gamma-1)}{\pi_1(\gamma-1)}\right\}^2$$
$$\times\frac{Z''^2}{ZZ'}\frac{\pi_1(\gamma)}{\pi_1(n+\gamma)}e^{-\epsilon n},$$

or, with the aid of (3.1), (3.6), and (7.10),

$$S_N(\omega) = \frac{2}{\pi}\left[\sum_{n=1}^{\infty}\frac{e^{-\epsilon n}\pi(n-1)}{(1-e^{-\epsilon n})^2+\omega^2}\right.$$
$$\left.+\frac{Z'}{Z}\frac{e^{-2\epsilon\gamma}}{\pi_1(\gamma)}\sum_{-\infty}^{+\infty}\frac{e^{-\epsilon n}\pi_1(n+\gamma-1)}{(1+e^{-\epsilon(n+\gamma)})^2+\omega^2}\right].$$

In order to compare this with the result of reference 1, one has to replace ω by $\epsilon\omega$ and multiply the whole expression with ϵ, because of the difference in time scale. The first term then agrees with Eq. (44) of reference 1. It involves the relaxation times

$$\frac{\epsilon}{1-e^{-\epsilon n}}R_0C = \frac{1}{n}R_0C+O(\epsilon),$$

where R_0C is the RC time of the circuit in the linear region. The second term is new; it involves relaxation times of order

$$\frac{\epsilon}{1+e^{-\epsilon(n+\gamma)}}R_0C = \frac{e^2R_0}{kT}+\cdots.$$

The physical meaning has been discussed in the introduction, and in reference 1. Of course, these relaxation times loose their physical meaning when they become comparable with the time of flight of the electrons between both electrodes.

The expansion of $S_N(\omega)$ for large ω is

$$S_N(\omega) \simeq (2/\pi\omega^2)\sum\lambda\{\mathbf{S}N\bar{\Psi}\}^2,$$
$$\simeq\frac{2}{\pi\omega^2}\left[\sum_{n=1}^{\infty}e^{-\epsilon n}\pi(n-1)\right.$$
$$\left.+\frac{Z'}{Z}\frac{e^{-2\epsilon\gamma}}{\pi_1(\gamma)}\sum_{-\infty}^{+\infty}e^{-\epsilon n}\pi_1(n+\gamma-1)\right]. \tag{8.2}$$

It has been shown before[1,4] that this should be independent of the nonlinearity, which implies that the quantity [ ] should be unity. It can readily be verified (by comparing powers of $e^{-\epsilon}$ on both sides) that

$$\sum_{n=1}^{\infty}e^{-\epsilon n}\pi(n-1) = 1-G.$$

Moreover, it is shown in Appendix G that

$$\sum_{-\infty}^{+\infty}e^{-\epsilon n}\pi_1(n+\gamma-1) = Ge^{\epsilon\gamma}. \tag{8.3}$$

Substituting this in (8.2) and using (2.4), (7.4), and (3.5), one finds the desired result.

This check is important for the following reason. Equation (8.1) has been derived for a *complete* set of normalized eigenfunctions. Unfortunately, we have not been able to prove that the set consisting of the functions $\bar{\Psi}_n^{(1)}$ and $\bar{\Psi}_n^{(2)}$ is complete. However, it is clear that completeness is not actually necessary for (8.1); it is sufficient that no eigenfunction has been skipped for which

$$\lambda\{\mathbf{S}N\bar{\Psi}\}^2 > 0.$$

The fact that the quantity [ ] in (8.2) is actually

unity guarantees that this condition is met. In particular, it follows that eigenfunctions belonging to $\lambda=1$, if any, cannot contribute.

APPENDIX A. PROOF THAT $\Phi_n^{(4)}$ IS PROPORTIONAL TO $\Phi_n^{(1)}$

According to (6.5) and (3.7),

$$\Phi_{-1}^{(4)}(s)=\frac{\exp[(\epsilon\gamma+\pi i)s]}{\pi_1(-1-\gamma)\pi_1(\gamma)} \times\sum_{-\infty}^{+\infty}(-1)^\nu\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(s-\gamma+\tfrac{1}{2})\nu].$$

The infinite sum is

$$\vartheta_4(-\tfrac{1}{2}i\epsilon(s-\gamma+\tfrac{1}{2}))=0.$$

Thus, $\Phi_{-1}^{(4)}=0$. Moreover, the relation (4.5) when applied to $\Phi_{-m}^{(4)}$ with $m\geqslant 0$ takes the form

$$(1-e^{\epsilon m}\mathbf{E})\Phi_{-m}^{(4)}=\Phi_{-m-1}^{(4)}.$$

Substituting successively $m=1, 2, \cdots$, one finds that all $\Phi_{-m}^{(4)}$ are zero. Substituting $m=0$, one finds

$$(1-\mathbf{E})\Phi_0^{(4)}=0;\quad \text{hence, } \Phi_0^{(4)}=\text{constant}.$$

Thus $\Phi_0^{(4)}$ is proportional to $\Phi_0^{(1)}$. It then follows immediately from (4.6) that each $\Phi_n^{(4)}$ for $n\geqslant 0$ is proportional to $\Phi_n^{(1)}$ (with the same proportionality constant).

APPENDIX B. PROOF THAT $\Phi_n^{(3)}$ IS PROPORTIONAL TO $\Phi_n^{(2)}$

First one obtains from (6.2) directly

$$\Phi_n^{(3)}(2n+2\gamma-s)=(-1)^n\Phi_n^{(2)}(s) \times\exp[-\epsilon(n+\gamma+\pi i/\epsilon)s+\epsilon n^2+2\epsilon\gamma n+2\epsilon\gamma^2+2\pi i\gamma].$$

Substitute $n=0$, $s=\gamma$,

$$\Phi_0^{(3)}(\gamma)=\Phi_0^{(2)}(\gamma)\exp(\epsilon\gamma^2+\pi i\gamma). \tag{B.1}$$

Substitute $s=\gamma+1$,

$$\Phi_0^{(3)}(\gamma-1)=-\Phi_0^{(2)}(\gamma+1)\exp(\epsilon\gamma^2+\pi i\gamma-\epsilon\gamma).$$

On the other hand, from the difference equation for $\Phi_0^{(2)}$,

$$\Phi_0^{(2)}(\gamma-1)+e^{-\epsilon\gamma}\Phi_0^{(2)}(\gamma+1)=0.$$

Hence,

$$\Phi_0^{(3)}(\gamma-1)=\Phi_0^{(2)}(\gamma-1)\exp(\epsilon\gamma^2+\pi i\gamma). \tag{B.2}$$

From (B.1) and (B.2) and the fact that $\Phi_0^{(2)}(s)$ and $\Phi_0^{(3)}(s)$ satisfy the same difference equation (4.1), it follows that they must be proportional to each other. Finally, using the relations (4.5) and (4.6), one finds generally

$$\Phi_n^{(3)}(s)=\Phi_n^{(2)}(s)\exp(\epsilon\gamma^2+\pi i\gamma). \tag{B.3}$$

APPENDIX C. PROOF OF THE NORMALIZATION FORMULAS (7.1) AND (7.2)

If σ and τ have such values that the norm exists, one finds using the result (2.5)

$$\mathbf{S}\Psi_0(s)[\Phi(s;\sigma,\tau)]^2=\lambda^{-1}\mathbf{S}\Psi_0(s)[(1-\mathbf{E}^{-1})\Phi(s;\sigma,\tau)]^2.$$

From (4.7), this may be written

$$(1-e^{-\epsilon\tau})\mathbf{S}\Psi_0(s)e^{2\epsilon(s-1)}[\mathbf{E}^{-2}\Phi(s;\sigma,\tau-1)]^2,$$

which according to (2.3) is equal to

$$(1-e^{-\epsilon\tau})e^{\epsilon}\mathbf{S}[\mathbf{E}^{-2}\Psi_0(s)][\mathbf{E}^{-2}\Phi(s;\sigma,\tau-1)]^2 =(1-e^{-\epsilon\tau})e^{\epsilon}\mathbf{S}\Psi_0(s)[\Phi(s;\sigma,\tau-1)]^2.$$

Thus, we have obtained a recursion formula for the norm. When applied to the $\Phi_n^{(1)}$, it yields

$$\mathbf{S}\Psi_0[\phi_n^{(1)}]^2=(1-e^{-\epsilon n})(1-e^{-\epsilon(n-1)})\cdots \times(1-e^{-\epsilon})e^{\epsilon n}\mathbf{S}\Psi_0[\Phi_0^{(1)}]^2,$$

which is just (7.1). When applied to $\Phi_n^{(2)}$, it yields in a similar way (7.2).

APPENDIX D. PROOF OF (7.7) AND (7.8)

We first derive a recursion formula with the aid of (4.6).

$$\begin{aligned}\mathbf{S}s\Psi(s;\sigma,\tau)&=\mathbf{S}s(1-e^{-\epsilon(\tau-1)}\mathbf{E}^{-1})\Psi(s;\sigma,\tau-1)\\&=\mathbf{S}\Psi(s;\sigma,\tau-1)(1-e^{-\epsilon(\tau-1)}\mathbf{E})s\\&=(1-e^{-\epsilon(\tau-1)})\mathbf{S}s\Psi(s;\sigma,\tau-1)\\&\quad-e^{-\epsilon(\tau-1)}\mathbf{S}\Psi(s;\sigma,\tau-1).\end{aligned}$$

According to (7.6), the last term vanishes unless $\tau=1$ and $\sigma=0$. By iterating this recursion formula,

$$\mathbf{S}s\Psi_n^{(1)}(s)=-\pi(n-1)Z,$$

$$\mathbf{S}s\Psi_n^{(2)}(s)/\pi_1(n+\gamma-1)=\text{independent of } n.$$

These are the relations (7.7) and (7.8), respectively.

It should be noted that these equations do not give $\mathbf{S}s\Psi_0^{(1)}$, but this quantity does not occur in the fluctuation spectrum either. Actually, it is easy to compute

$$\mathbf{S}s\Psi_0^{(1)}(s)=\gamma Z+(2i)^{-1}\vartheta_3'(\tfrac{1}{2}i\epsilon(\gamma-\tfrac{1}{2}))\exp(-\tfrac{1}{2}\epsilon\gamma^2+\tfrac{1}{2}\epsilon\gamma),$$

where the prime denotes differentiation with respect to the argument of the theta function.

APPENDIX E. COMPUTATION OF Z'

In the definition (7.3) of Z', we substitute one factor $\Phi_0^{(2)}$ from (6.1) and the other from (6.3), using (B.3).

$$\begin{aligned}Z'=\mathbf{S}&\exp(-\tfrac{1}{2}\epsilon s^2+\tfrac{1}{2}\epsilon s)\\&\times\left\{\sum_{\nu=0}^{\infty}(-1)^\nu\begin{bmatrix}\gamma\\ \nu\end{bmatrix}_1\exp(-\epsilon\nu^2+\epsilon\nu s)\right\}\\&\times\Big\{\exp(-\epsilon\gamma^2-\pi i\gamma)\exp[(\epsilon\gamma+\pi i)s]\\&\quad\times\sum_{\mu=0}^{\infty}(-1)^\mu\begin{bmatrix}\gamma\\ \mu\end{bmatrix}_1\exp(-\epsilon\mu^2+2\epsilon\gamma\mu-\epsilon\mu s)\Big\}.\end{aligned}$$

Put $s=\gamma+N$, so that the summation variable N takes integral values,

$$Z'=\exp(-\tfrac{1}{2}\epsilon\gamma^2+\tfrac{1}{2}\epsilon\gamma)\times\lim_{N_1\to\infty}\sum_{-N_1}^{+N_1}\exp(-\tfrac{1}{2}\epsilon N^2+\tfrac{1}{2}\epsilon N+\pi iN)\times\sum_{\nu=0}^{\infty}(-1)^\nu\begin{bmatrix}\gamma\\ \nu\end{bmatrix}_1\exp(-\epsilon\nu^2+\epsilon\gamma\nu+\epsilon\nu N)\times\sum_{\mu=0}^{\infty}(-1)^\mu\begin{bmatrix}\gamma\\ \mu\end{bmatrix}_1\exp(-\epsilon\mu^2+\epsilon\gamma\mu-\epsilon\mu N).$$

Since we have cut off the range of N, the triple sum is absolutely convergent, and is therefore equal to

$$\sum_{\mu=0}^{\infty}(-1)^\mu\begin{bmatrix}\gamma\\ \mu\end{bmatrix}_1\exp(-\epsilon\mu^2+\epsilon\gamma\mu)\times\sum_{\nu=0}^{\infty}(-1)^\nu\begin{bmatrix}\gamma\\ \nu\end{bmatrix}\exp(-\epsilon\nu^2+\epsilon\gamma\nu)\times\sum_{-N_1}^{+N_1}(-1)^N\exp[-\tfrac{1}{2}\epsilon N^2+\epsilon(\nu-\mu+\tfrac{1}{2})N]. \quad (E.1)$$

It is readily checked that for integral $\nu-\mu$,

$$\sum_{-\infty}^{+\infty}(-1)^N\exp[-\tfrac{1}{2}\epsilon N^2+\epsilon(\nu-\mu+\tfrac{1}{2})N]=0. \quad (E.2)$$

Hence, the summation over N in (E.1) may be replaced,

$$\sum_{-N_1}^{+N_1}\to-\sum_{N_1}^{\infty}-\sum_{-\infty}^{-N_1}. \quad (E.3)$$

We consider the contribution of the first of these two terms to the sum (E.1).

First put $N-\nu+\mu=N'$ in (E.1);

$$-\sum_{\mu=0}^{\infty}\begin{bmatrix}\gamma\\ \mu\end{bmatrix}_1\exp[-\tfrac{1}{2}\epsilon\mu^2+\epsilon(\gamma-\tfrac{1}{2})\mu]\times\sum_{\nu=0}^{\infty}\begin{bmatrix}\gamma\\ \nu\end{bmatrix}_1\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(\gamma+\tfrac{1}{2})\nu-\epsilon\nu\mu]\times\sum_{N_1-\nu+\mu}^{\infty}(-1)^{N'}\exp[-\tfrac{1}{2}\epsilon N'^2+\tfrac{1}{2}\epsilon N'].$$

The last sum clearly tends to zero as $N_1\to\infty$ for any finite value of ν (regardless of μ). Hence, one makes no error by using the asymptotic value (3.14),

$$\begin{bmatrix}\gamma\\ \nu\end{bmatrix}_1\simeq\frac{1}{\pi_1(-\gamma-1)}\exp[\tfrac{1}{2}\epsilon\nu^2-\epsilon(\gamma+\tfrac{1}{2})\nu].$$

Moreover, put $\nu=N_1+\mu+\lambda$; the result is

$$\frac{-1}{\pi_1(-\gamma-1)}\sum_{\mu=0}^{\infty}\begin{bmatrix}\gamma\\ \mu\end{bmatrix}_1\exp[-\tfrac{3}{2}\epsilon\mu^2+\epsilon(\gamma-\tfrac{1}{2}-N_1)\mu]\times\sum_{\lambda=-N_1-\mu}^{\infty}\exp(-\epsilon\mu\lambda)\times\sum_{N'=-\lambda}^{\infty}(-1)^{N'}\exp(-\tfrac{1}{2}\epsilon N'^2+\tfrac{1}{2}\epsilon N'). \quad (E.4)$$

Consider separately the terms with $\mu\gtrsim 1$. Summation by parts yields for the repeated sum over λ and N

$$\sum_{-N_1-\mu}^{\infty}\frac{e^{-\epsilon\mu\lambda}}{1-e^{-\epsilon\mu}}(-1)^\lambda\exp(-\tfrac{1}{2}\epsilon\lambda^2-\tfrac{1}{2}\epsilon\lambda)+\frac{\exp[\epsilon\mu(N_1+\mu)]}{1-e^{-\epsilon\mu}}\sum_{N_1+\mu+1}^{\infty}(-1)^{N'}\times\exp(-\tfrac{1}{2}\epsilon N'^2+\tfrac{1}{2}\epsilon N'). \quad (E.5)$$

The second term, when inserted in the sum over μ, gives rise to a sum which in absolute value is less than

$$\sum_{\mu=1}^{\infty}\begin{bmatrix}\gamma\\ \mu\end{bmatrix}_1\frac{\exp[-\tfrac{1}{2}\epsilon\mu^2+\epsilon(\gamma-\tfrac{1}{2})\mu]}{1-e^{-\epsilon}}\times\exp[-\tfrac{1}{2}\epsilon(N_1+\mu+1)^2+\tfrac{1}{2}\epsilon(N_1+\mu+1)].$$

Obviously, this vanishes in the limit $N_1\to\infty$. The first term in (E.5) is majorized by

$$\frac{1}{1-e^{-\epsilon}}\sum_{-N_1-\mu}^{\infty}\exp[-\tfrac{1}{2}\epsilon(\lambda+\mu)^2-\tfrac{1}{2}\epsilon(\lambda+\mu)]\times\exp(\tfrac{1}{2}\epsilon\mu^2+\tfrac{1}{2}\epsilon\mu)<\text{const. }\exp(\tfrac{1}{2}\epsilon\mu^2+\tfrac{1}{2}\epsilon\mu).$$

When inserted in the sum (E.4), the result is less than

$$\sum_{\mu=1}^{\infty}\begin{bmatrix}\gamma\\ \mu\end{bmatrix}_1\exp[-\epsilon\mu^2+\epsilon(\gamma-N_1)\mu]\to 0.$$

Hence, only the term with $\mu=0$ survives in the limit $N_1\to\infty$. Its value is found from (E.4)

$$\frac{-1}{\pi_1(-\gamma-1)}\lim_{N_1\to\infty}\sum_{\lambda=-N_1}^{\infty}\sum_{-\lambda}^{\infty}(-1)^{N'}\exp(-\tfrac{1}{2}\epsilon N'^2+\tfrac{1}{2}\epsilon N').$$

Again, summing by parts, one finds for the sum

$$-\sum_{-N_1}^{\infty}(\lambda+N_1)(-1)^{-(\lambda+1)}\exp[-\tfrac{1}{2}\epsilon(\lambda+1)^2-\tfrac{1}{2}\epsilon(\lambda+1)]+\sum_{N_1}^{\infty}(-1)^{N'}\exp(-\tfrac{1}{2}\epsilon N'^2+\tfrac{1}{2}\epsilon N').$$

The last term vanishes in the limit and so does

$$-(N_1-1)\sum_{-N_1+1}^{\infty}(-1)^\lambda \exp(-\tfrac{1}{2}\epsilon\lambda^2-\tfrac{1}{2}\epsilon\lambda)$$
$$=(N_1-1)\sum_{-\infty}^{-N_1}(-1)^\lambda \exp(-\tfrac{1}{2}\epsilon\lambda^2-\tfrac{1}{2}\epsilon\lambda)$$

[using the identity (E.2) for $\nu-\mu=-1$]. The remaining term is

$$-\sum_{-\infty}^{+\infty}\lambda(-1)^\lambda \exp(-\tfrac{1}{2}\epsilon\lambda^2-\tfrac{1}{2}\epsilon\lambda)$$
$$=-(2i)^{-1}\vartheta_4'(\tfrac{1}{4}i\epsilon)=-G^3.$$

The second term in (E.3) may be shown, in a very similar way, to vanish. Thus, collecting results, one finds (7.4).

APPENDIX F. COMPUTATION OF Z''

According to the definition (7.9) and (7.6),

$$Z''=\sum_{-\infty}^{+\infty}(\gamma+N)\Psi_0{}^{(2)}(\gamma+N)$$
$$=\exp(-\epsilon\gamma^2-\pi i\gamma)\sum_{-\infty}^{+\infty}N\Psi_0\Phi_0{}^{(3)}.$$

Hence, $Z''\exp(\frac{1}{2}\epsilon\gamma^2-\frac{1}{2}\epsilon\gamma)$ is equal to

$$\sum_{N=-\infty}^{+\infty}(-1)^N N\exp(-\tfrac{1}{2}\epsilon N^2+\tfrac{1}{2}\epsilon N)$$
$$\times\sum_{\nu=0}^{\infty}(-1)^\nu\begin{bmatrix}\gamma\\ \nu\end{bmatrix}_1\exp(-\epsilon\nu^2+\epsilon\gamma\nu-\epsilon\nu N).$$

If N is cut off at the lower end, the summations may be interchanged so that one gets

$$\lim_{N_1\to-\infty}\sum_{\nu=0}^{\infty}(-1)^\nu\begin{bmatrix}\gamma\\ \nu\end{bmatrix}_1\exp(-\epsilon\nu^2+\epsilon\gamma\nu)$$
$$\times\sum_{-N_1}^{\infty}(-1)^N N\exp(-\tfrac{1}{2}\epsilon N^2+\tfrac{1}{2}\epsilon N-\epsilon\nu N). \quad (F.1)$$

It can be shown, by methods similar to those in Appendix E, that

$$\lim_{N_1\to\infty}\sum_{\nu=0}^{\infty}\cdots\sum_{N=-\infty}^{-N_1-1}\cdots=0.$$

Hence, in (F.1) one may simply replace N_1 with ∞, so that the summation over N becomes

$$\sum_{-\infty}^{+\infty}(-1)^N N\exp[-\tfrac{1}{2}\epsilon N^2-\epsilon(\nu-\tfrac{1}{2})N]$$
$$=(2i)^{-1}\vartheta_4'(\tfrac{1}{2}i\epsilon(\nu-\tfrac{1}{2}))$$
$$=(2i)^{-1}(-1)^{\nu-1}\exp(\tfrac{1}{2}\epsilon\nu^2-\tfrac{1}{2}\epsilon\nu)\vartheta_4'(\tfrac{1}{4}i\epsilon)$$
$$=(-1)^{\nu-1}\exp(\tfrac{1}{2}\epsilon\nu^2-\tfrac{1}{2}\epsilon\nu)G^3.$$

It remains to evaluate the sum

$$Q_0\equiv\sum_{\nu=0}^{\infty}\begin{bmatrix}\gamma\\ \nu\end{bmatrix}_1\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(\gamma-\tfrac{1}{2})\nu].$$

We shall first show

$$Q_0=\frac{1}{G}\sum_{\nu=0}^{\infty}\frac{1}{\pi(\nu)}\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(\gamma-\tfrac{1}{2})\nu]. \quad (F.2)$$

Define for $p=0, 1, 2, \cdots$

$$Q_p\equiv\sum_{\nu=0}^{\infty}\begin{bmatrix}\gamma+p\\ \nu\end{bmatrix}_1\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(\gamma-\tfrac{1}{2})\nu].$$

By use of (3.12), one finds the recurrence relation

$$Q_p=(1-e^{-\epsilon p})Q_{p-1}.$$

From this follows by iteration

$$Q_0=Q_p/\pi(p).$$

Hence, it suffices to find Q_p for $p\to\infty$. For fixed ν_1,

$$Q_p=\sum_{\nu=0}^{\nu_1}+\sum_{\nu_1+1}^{p}+\sum_{p+1}^{\infty}. \quad (F.3)$$

For the second term, one has

$$\sum_{\nu_1+1}^{p}\frac{\pi_1(\gamma+p)}{\pi_1(\gamma+p-\nu)\pi(\nu)}\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(\gamma-\tfrac{1}{2})\nu]$$
$$\leqslant\text{const.}\sum_{\nu_1+1}^{p}\frac{1}{\pi(\nu)}\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(\gamma-\tfrac{1}{2})\nu],$$
$$\leqslant\text{const.}\sum_{\nu_1+1}^{\infty}\exp(-\tfrac{1}{2}\epsilon\nu^2).$$

The third term in (F.3) approaches, for large p,

$$\sum_{p+1}^{\infty}\frac{1}{\pi_1(\gamma+p-\nu)}\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(\gamma-\tfrac{1}{2})\nu]$$
$$=G\frac{\exp[-\frac{1}{2}\epsilon p^2+\epsilon(\gamma-\frac{1}{2})p]}{\pi_1(\gamma)\pi_1(-\gamma-1)}\sum_{\nu=1}^{\infty}\pi_1(-\gamma-1+\nu)$$
$$\times\exp(-\epsilon p\nu-\epsilon\nu),$$

which obviously vanishes as $p\to\infty$. The first term, for $p\to\infty$, is

$$\sum_{\nu=0}^{\nu_1}\frac{1}{\pi(\nu)}\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(\gamma-\tfrac{1}{2})\nu].$$

Now let ν_1 also go to infinity; the result is (F.2).

In order to evaluate the series (F.2), define

$$R_p\equiv\sum_{\nu=0}^{\infty}\frac{1}{\pi(\nu)}\exp[-\tfrac{1}{2}\epsilon\nu^2+\epsilon(\gamma-\tfrac{1}{2})\nu-\epsilon p\nu],$$

so that $R_0=GQ_0$. With the aid of (3.1), one finds the recurrence relation

$$R_p=\exp(-\epsilon\gamma+\epsilon p)(R_{p-1}-R_p).$$

From this follows by iteration

$$R_0=\pi_1(p-\gamma)/\pi_1(-\gamma).$$

As $R_\infty=1$, one finally has $R_0=G/\pi_1(-\gamma)$. On collecting results, one finds (7.10).

APPENDIX G. COMPUTATION OF

$$P(z)=\sum_{-\infty}^{+\infty} e^{-\epsilon n}\pi(z+n-1)$$

Define

$$P_p\equiv\sum_{-\infty}^{+\infty} e^{-\epsilon p n}\pi(z+n-p).$$

Inserting for $\pi(z+n-p)$ the identity (3.1), one finds the recurrence relation

$$P_p=e^{-\epsilon p}P_p-e^{-\epsilon(z-p)}P_{p+1}.$$

From this follows by iteration

$$P_p=(-1)^r\frac{\pi(p-1)}{\pi(p+r-1)}\times\exp[\tfrac{1}{2}\epsilon r^2+\epsilon(p-z-\tfrac{1}{2})r]P_{p+r}. \quad \text{(G.1)}$$

Hence, it suffices to find the asymptotic value of P_p for large p.

For this purpose, write

$$P_p=\exp(-\epsilon p^2)\sum_{-\infty}^{+\infty}\pi(z+n)\exp(-\epsilon pn)=\exp(-\epsilon p^2)\{\sum_{-n_1+1}^{\infty}+\sum_{-\infty}^{-n_1}\}.$$

The first term is in absolute value not greater than

$$\exp(-\epsilon p^2+\epsilon n_1 p)\sum_{-n_1+1}^{\infty}|\pi(z+n)|\exp(-\epsilon n)=O(\exp[-\epsilon p^2])\quad\text{for fixed } n_1.$$

In the second term, the asymptotic value (3.4) for $\pi(z+n)$ may be used because n_1 may be chosen arbitrarily large.

$$\exp(-\epsilon p^2)\sum_{-\infty}^{-n_1}\pi(z+n)\exp(-\epsilon np)$$

$$\simeq\exp(-\epsilon p^2)\frac{\pi(z)\pi(-z-1)}{G}$$

$$\times\sum_{-\infty}^{-n_1}(-1)^n\exp[-\tfrac{1}{2}\epsilon n^2-\epsilon(p+\tfrac{1}{2}+z)n]$$

$$\simeq\frac{\pi(z)\pi(-z-1)}{G}(-1)^p\vartheta_4(\tfrac{1}{2}i\epsilon z+\tfrac{1}{4}i\epsilon)\times\exp[-\tfrac{1}{2}\epsilon p^2+\epsilon p(z+\tfrac{1}{2})].$$

On account of (3.5),

$$P_p\simeq(-1)^pG^2\exp[-\tfrac{1}{2}\epsilon p^2+\epsilon p(z+\tfrac{1}{2})].$$

By inserting this result in (G.1) for P_{p+r}, one obtains

$$P_p=(-1)^p\pi(p-1)G\exp[-\tfrac{1}{2}\epsilon p^2+\epsilon p(z+\tfrac{1}{2})].$$

In particular for $p=1$,

$$P(z)=-Ge^{\epsilon z}.$$

On substituting $z=\gamma+\pi i/\epsilon$, one obtains (8.3).

Exact Wave Functions in Superconductivity

DANIEL MATTIS AND ELLIOTT LIEB
IBM Research Center, Yorktown Heights, New York
(Received February 9, 1961)

The ground-state wave function and some of the excited states of the BCS reduced Hamiltonian are found. In the limit of large volume, the boundary and continuity conditions on the exact wave function lead directly to the equations which Bardeen, Cooper, and Schrieffer found by a variational technique. It is also shown in what sense the BCS trial wave function may be considered asymptotically exact in this limit. Finite-volume corrections are included in an appendix, and explicit calculations are carried out for a one-step model of the kinetic energy which has possible applications to the problem of the finite nucleus.

I. INTRODUCTION

WE wish to find the ground-state wave function and some of the elementary excited states of

$$H=\sum_{k,s}\epsilon(k,s)C_{k,s}{}^{*}C_{k,s}-v\sum_{k}\sum_{k'\neq k}C_{k\uparrow}{}^{*}C_{-k\downarrow}{}^{*}C_{-k'\downarrow}C_{k'\uparrow}. \quad (1)$$

The operators C and C^* are the usual Fermi operators and anti-commute. The sums are restricted to an immediate neighborhood of the Fermi surface, which includes $4n$ distinct states of momentum (k) and spin ($s=\uparrow$ or $\downarrow$), and which are populated by $2n$ electrons. In other words, our eigenfunctions must be simultaneously eigenfunctions of the number operator η

$$\eta=\sum_{k,s}C_{k,s}{}^{*}C_{k,s}, \quad (2)$$

with eigenvalue $2n$.

Our Hamiltonian is the famous "reduced Hamiltonian" of the BCS theory; and for an introduction to the present work, we refer the reader to Sec. II of the BCS paper.[1] In their notation, $n=N(0)\hbar\omega$, where $N(0)=$density of states at the Fermi surface and $\hbar\omega=$typical phonon energy. As has been stated, we wish to investigate the nature of the exact solutions to this problem, and we shall see that they are very similar to what BCS found by a variational calculation.

For the purposes of finding the ground state, it is convenient to think in terms of a pseudo-Hamiltonian $\tilde{H}$ which has the same ground state as (1). First, by time-reversal symmetry, we may assume that $\epsilon(k,\uparrow)=\epsilon(-k,\downarrow)$. Second, it is clear that, in the ground state, all electrons must be paired, as in $C_{k\uparrow}{}^{*}C_{-k\downarrow}{}^{*}$, because unpaired electrons do not benefit from the attractive interaction. Following BCS, then, we define

$$b_k{}^{*}=C_{k\uparrow}{}^{*}C_{-k\downarrow}{}^{*}, \quad b_k=C_{-k\downarrow}C_{k\uparrow}, \quad (3)$$

and consequently, the ground state of the pseudo-Hamiltonian

$$\tilde{H}=2\sum\epsilon_k b_k{}^{*}b_k-v\sum_{k}\sum_{k'\neq k}b_k{}^{*}b_{k'} \quad (4)$$

coincides with the ground state of H. [We have set $\epsilon_k=\epsilon(k,\uparrow)$.] Indeed, every eigenstate of $\tilde{H}$ is a state of H, but the converse is not true. The b_k operators have mixed commutation properties and may not be regarded as Bosons for which the diagonalization of (4) would be trivial. In fact, they are a set of Pauli operators.[1]

A complete set of states for our problem consists of all possible configurations of n pairs, of which a typical member is

$$\phi_i=[\prod_{k\subseteq\{k\}^i}b_k{}^{*}]|0\rangle \quad (5)$$

where $\{k\}_i$ is a set of n different k's chosen from the $2n$ permissible values. There are $(2n)!/(n!)^2\cong 2^{2n}/(\pi n)^{\frac{1}{2}}$ different ϕ_i's. For comparison, the totality of configurations (allowing an arbitrary occupation number) is 2^{2n}. For every ϕ_i there is a corresponding amplitude, which we may write as $f_i\equiv f[S(k_1)\cdots S(k_{2n})]$, where $S(k_j)=1$ or 0 according to whether k_j is in the set $\{k\}_i$ or not. It is to be understood that f is not defined for all possible values of its arguments (of which there are 2^{2n}) but only for those values such that $\sum_{k_j}S(k_j)\equiv n$. The general eigenfunction of $\tilde{H}$ is therefore

$$\psi=\sum_{\text{config.}}f_i\phi_i. \quad (6)$$

The problem now consists of finding the ground-state amplitudes f_i and the corresponding energy. For some insight into the general problem, we first turn to the strong-coupling limit which is well understood.

II. STRONG-COUPLING LIMIT[2]

We set $\epsilon_k=0$, and the Hamiltonian is simply

$$H_{\text{s.c.}}=-v\sum_{k}\sum_{k'\neq k}b_k{}^{*}b_{k'}. \quad (7)$$

As this is purely attractive, we may safely assume that the ground-state wave function possesses all the symmetry of the Hamiltonian. The outstanding symmetry property is invariance under the interchange of any two momenta k and k'. Therefore, one may pre-

[1] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).

[2] The strong-coupling limit is generally well understood. An exhaustive treatment of this limit, including a perturbation-theoretic approach to weak coupling, is given by Wada and Fukuda, Progr. Theoret. Phys. (Kyoto) **22**, 775 (1959).

sume that

$$f(\cdots S(k)\cdots S(k')\cdots)=f(\cdots S(k')\cdots S(k)\cdots), \quad (8)$$

i.e., that f is a symmetric function of its arguments. Now we make use of the property that $S(k)=0$ or 1, which is expressible as

$$S(k)=S^2(k). \quad (9)$$

Consequently,[3] the most general function which obeys Eq. (8) can be written as

$$f(S(k_1)\cdots S(k_{2n}))=f(\sum_k S(k))=f(n). \quad (10)$$

But as n is a constant, f must be constant and hence all amplitudes are equal in the strong-coupling ground state. We can check this directly:

$$(-v\sum_k\sum_{k'\neq k} b_k^*b_{k'})\cdot f(n)\sum_i\prod_{k\subseteq\{k\}_i} b_k^*|0\rangle$$
$$=E_{\text{s.c.}}f(n)\sum_i\prod b_k^*|0\rangle, \quad (11)$$

with $f(n)=n!(2n!)^{-\frac{1}{2}}\cong 2^{-n}(\pi n)^{\frac{1}{4}}$ for normalization. This is the Schrodinger equation, and each complexion is connected to n^2 other complexions. Therefore,

$$E_{\text{s.c.}}=-vn^2, \quad (12)$$

a well-known result. It may be useful to recall that n and v^{-1} are both proportional to the volume (for fixed density), so that E is an extensive property of the system. Eq. (12) is in perfect agreement with the BCS result taken in the strong-coupling limit, but is in slight disagreement with the calculation of Wada and Fukuda,[2] who include a diagonal term $-v\sum b_k^*b_k$ in their interaction. There is no particular significance in their discrepancy.

III. ONE-STEP MODEL

The number of sign changes (or nodes) in the amplitudes f is a good quantum number, and by the adiabatic theorem, its value persists as ϵ_k is changed from a constant value to some arbitrary function. We make use of this to solve for the ground state of a model which is not quite so trivial as the strong-coupling limit, and which may be of interest in the nuclear problem where energy levels are discrete. We shall assume that ϵ_k is a step function—zero over half of the states and equal to a positive constant (ϵ) over the remaining states.

The ground-state amplitudes must be nodeless functions which are symmetric under the interchange of any two pairs within the same half-space. Let the occupation numbers over each half-space be,

$$n_0=\sum_{\substack{k \text{ such that}\\ \epsilon_k=0}} S_k \quad \text{and} \quad n_\epsilon=\sum_{\substack{k \text{ such that}\\ \epsilon_k=\epsilon}} S_k. \quad (13)$$

[3] This theorem was kindly pointed out to us by Dr. D. Jepsen and Dr. T. D. Schultz of this laboratory.

We can eliminate n_0 by the relation

$$n=n_0+n_\epsilon=\text{const}, \quad (14)$$

and therefore the ground-state amplitudes are a function of n_ϵ alone, and are denoted $f(n_\epsilon)$. The equations for the amplitudes are simply

$$[2\epsilon n_\epsilon-E-2vn_\epsilon(n-n_\epsilon)]f(n_\epsilon)$$
$$=v(n-n_\epsilon)^2f(n_\epsilon+1)+vn_\epsilon^2f(n_\epsilon-1), \quad (15)$$

where n_ϵ assumes integer values from zero to a maximum of n. These equations are easily soluble when n is a small integer. For example, if $n=1$, there are only two amplitudes, $f(0)$ and $f(1)$, and the eigenvalue equation is the usual determinantal condition

$$\text{Det}\begin{vmatrix}-E & -v\\ -v & 2\epsilon-E\end{vmatrix}=0, \quad (16)$$

which has the solutions

$$E_\pm=\epsilon\pm(\epsilon^2+v^2)^{\frac{1}{2}}. \quad (17)$$

The lower of these E_- is the ground-state energy and belongs to the nodeless solution $f(1)/f(0)>0$, as expected.

For large n, the determinantal equation is impractical, and we now use a method for isolating the ground-state energy from all the other solutions in the limit of large volume, $n\to\infty$. Corrections in the form of an expansion in n^{-1} are discussed in the Appendix, and may be of value already for $n\geq 3$, when the determinantal method is cumbersome.

Because the amplitudes can be chosen real and positive in the ground state, we write

$$f(n_\epsilon)=\text{const}e^{nS(x)}, \quad (18)$$

where $x\equiv n_\epsilon/n$, and S is a real function. Next, we divide both sides of Eq. (15) by $nf(n_\epsilon)$ and find

$$2\epsilon x-W-2\lambda x(1-x)$$
$$=\lambda[(1-x)^2p(x+1/n)+x^2/p(x)], \quad (19)$$

where

$$p(x)=\exp\{n[S(x)-S(x-1/n)]\}, \quad (20)$$

$$W\equiv E/n, \quad \text{and} \quad \lambda\equiv vn. \quad (21)$$

The variable x goes from 0 to 1 in steps of $1/n$. One can now proceed to the limit $n\to\infty$, but first one notes that

$$\lim_{n\to\infty}\exp\{n[S(x+1/n)-S(x)]\}$$
$$=\lim_{n\to\infty}\exp\{+n[S(x)-S(x-1/n)]\}$$
$$=\exp[\partial/\partial x S(x)], \quad (22)$$

provided $S(x)$ is a sufficiently smooth function. Therefore, to order $1/n$ if $S(x)$ is sufficiently smooth, $p(x)=p[x+(1/n)]$, and Eq. (19) turns into an algebraic

equation

$$2\epsilon x - W - 2\lambda x(1-x) = \lambda\{(1-x)^2 p(x) + x^2/p(x)\}, \quad (23)$$

which is subject to the requirement that $p(x)$ be real, positive, and continuous. The conjecture that $S(x)$ approaches a continuous limit function as $n \to \infty$, which implies that $p(x) \cong p(x+1/n)$ and satisfies Eq. (23), which in turn implies that $S(x)$ has a limit function is certainly self-consistent. But it need not be true. Equation (19) is a nonlinear difference equation, and in order to get from the point $x=0$ to the point $x=1/4$ say, we must iterate it $n/4$ times. The assertion that $p(x+1/n)$ may be replaced by $p(x)$ will result in an error of order $1/n$. But since it takes $n/4$ steps to get to $x=1/4$, we may accumulate an error or order 1, in which case $p(1/4)$ will not satisfy Eq. (23). Once $p(x)$ ceases to satisfy the quadratic equation, we see from Eq. (19) that $p(x)$ will oscillate wildly. In the Appendix we prove that the errors do not in fact accumulate in the regions $(0,m)$ and $(n,1)$ where m and n are the least and greatest points, respectively, at which the discriminant of Eq. (23) vanishes. For the ground state, the discriminant vanishes at only one point and, hence, in this case, our smoothness assumption is justified everywhere except in a small neighborhood about the vanishing point. There are three critical points: at $x=0$ and 1, and at the turning point where the discriminant vanishes.

The "boundary conditions" are as follows: at $x=0$,

$$p(0) = -W/\lambda, \quad (24)$$

which follows from Eq. (19) at $x=0$. Obviously, W will have to be negative or zero. At $x=1$,

$$p(1) = \lambda/(2\epsilon - W), \quad (25)$$

which follows from Eq. (19) at $x=1$. At intermediate points, the quadratic equation possesses two solutions

$$p(x) = \frac{2\epsilon x - W - 2\lambda x(1-x)}{2\lambda(1-x)^2} \pm \frac{1}{1-x}\left[\left(\frac{2\epsilon x - W - 2\lambda x(1-x)}{2\lambda(1-x)}\right)^2 - x^2\right]^{\frac{1}{2}}. \quad (26)$$

The boundary conditions impose the positive root near $x=0$ and the negative root near $x=1$. Therefore, at one intermediate point, the discriminant must vanish so that the transition from positive to negative root may be continuous. The reality condition is translated into the requirement that the discriminant have a minimum at this "turning point" where it vanishes. Thus, simultaneously, we require

$$D = \left[\left(\frac{2\epsilon h_\epsilon - W - 2\lambda h_\epsilon(1-h_\epsilon)}{2\lambda(1-h_\epsilon)}\right)^2 - h_\epsilon^2\right] = 0, \qquad p(h_\epsilon) = \frac{h_\epsilon}{1-h_\epsilon}, \quad (27)$$

where $x=h_\epsilon$ is the turning point (by analogy with the BCS notation) and

$$\left.\frac{\partial D}{\partial x}\right|_{h_\epsilon} = 0. \quad (28)$$

It does *not* follow, however, that

$$\left.\frac{dp(x)}{dx}\right|_{x=h_\epsilon} = \left.\frac{d}{dx}\left(\frac{x}{1-x}\right)\right|_{x=h_\epsilon}, \quad \text{and indeed} \quad \frac{dp(x)}{dx}$$

is discontinuous at the point $x=h_\epsilon$, although it always remains finite. Equations (27) and (28) possess a solution provided $\lambda \geq \epsilon/2$,

$$h_\epsilon = \frac{1}{2}\left(1 - \frac{\epsilon}{2\lambda}\right), \quad p(h_\epsilon) = \frac{1-\epsilon/2\lambda}{1+\epsilon/2\lambda}, \quad (29)$$

and

$$W = -\lambda[1-\epsilon/(2\lambda)]^2. \quad (30)$$

Recalling that $\lambda = vn$ and $E = nW$, we find for the ground energy in the one-step model:

$$E_{\text{o.s.}} = -(vn)(n)\left(1 - \frac{\epsilon}{2(vn)}\right)^2, \quad \text{for} \quad (vn) \geq \frac{\epsilon}{2}. \quad (31)$$

For $(vn) = \lambda < \frac{1}{2}\epsilon$, the turning point sticks at $h_\epsilon = 0$, and one finds that only the negative solution is required for reality and continuity, provided $W=0$. Therefore,

$$E_{\text{o.s.}} = 0 \quad \text{for} \quad (vn) \leq \tfrac{1}{2}\epsilon. \quad (32)$$

Had we used the BCS trial function, the results would have been identical. As we shall see in Secs. IV and V, this is no coincidence, even though the BCS trial function is not an eigenfunction and does not conserve particles. It may also be easily verified that these results agree with the strong-coupling theory if we set $\epsilon=0$, even as to the constancy of the amplitudes $f(n_\epsilon)$ in that limit. For the excited states, we turn back to the Hamiltonian in its original form given in Eq. (1).

The low-lying excited states are relatively easy to find in the one-step model. We break up a pair, putting one electron in an $\epsilon_k = 0$ state, and the other in an $\epsilon_k = \epsilon$ state. There are $(n-1)$ remaining pairs for which $(n-1)$ $\epsilon_k = 0$ states are accessible, and an equal number of $\epsilon_k = \epsilon$ states. The energy of the "singles" is

$$E \text{ singles} = 0 + \epsilon = \epsilon, \quad (33)$$

and the lowest possible energy for the remaining pairs is [substituting $(n-1)$ for n in our previous result]

$$E^{(n-1)} = -(v)(n-1)^2\left[1 - \frac{\epsilon}{2v(n-1)}\right]^2, \quad (34)$$

provided $v(n-1) \geq \frac{1}{2}\epsilon$, and zero otherwise. Thus, the excitation energy Δ associated with such excited states

is (calculated to leading order in the volume)

$$\Delta = E \text{ singles} + E^{(n-1)} - E^{(n)} = 2(vn), \text{ if } (vn) \geq \tfrac{1}{2}\epsilon, \quad (35)$$

and

$$\Delta = \epsilon, \text{ if } (vn) \leq \tfrac{1}{2}\epsilon. \quad (36)$$

It is interesting to note from Eq. (35) that unless ϵ exceeds a critical value, this energy of excitation is independent of ϵ, hence, is the same as for the strong-coupling limit. This shows an amazing rigidity in the ground-state wave function.

IV. SOLUTION FOR ARBITRARY FUNCTION ϵ_k

Proceeding with a knowledge of the one-step model, we can now derive the BCS equations for an arbitrary function ϵ_k. We do this by approximating ϵ_k as closely as we please by a staircase function. If we call the number of states in the step about some discrete ϵ, N_ϵ, then as $n \to \infty$, $N_\epsilon \to \infty$. Thus, no matter how "fine" the staircase, each step will always have an infinite number of states associated with it. The limit to a continuous function $\epsilon(k)$ is taken *after* the limit $n \to \infty$, but always the number of steps on the staircase is regarded as large. We shall assume that ϵ varies from a minimum value $E_F - \hbar\omega$ to a maximum of $E_F + \hbar\omega$, where E_F is the unperturbed Fermi level and $\hbar\omega$ is the energy of the typical phonons responsible for the attractive interaction v. We define the population of the portion of phase space belonging to ϵ_k in a given complexion by

$$n_\epsilon = \sum_{\substack{k \text{ such that} \\ \epsilon_k = \epsilon}} S_k. \quad (37)$$

As before, n_ϵ can vary by integer steps from zero to a maximum value N_ϵ. If we denote a sum over distinct energy shells, (i.e., a sum over the steps in the staircase) by the usual summation symbol with superscript ϵ, we recall that

$$\sum^\epsilon n_\epsilon = n, \quad (38)$$

and

$$\sum^\epsilon N_\epsilon = 2n. \quad (39)$$

The algebraic equations for the amplitudes are

$$[2\sum^\epsilon \epsilon n_\epsilon - E - v\sum^\epsilon n_\epsilon(N_\epsilon - n_\epsilon)] f(\cdots n_\epsilon \cdots n_{\epsilon'} \cdots) = V \sum^\epsilon \sum^{\epsilon' \neq \epsilon} (N_\epsilon - n_\epsilon) n_{\epsilon'} \times f(\cdots (n_\epsilon + 1) \cdots (n_{\epsilon'} - 1) \cdots). \quad (40)$$

We let

$$f(\cdots n_\epsilon \cdots n_{\epsilon'} \cdots) = \exp n S(\cdots x_\epsilon \cdots x_{\epsilon'} \cdots), \quad (41)$$

where $x_\epsilon = n_\epsilon / N_\epsilon$, and again divide both sides of the equation by the amplitude $f(\cdots n_\epsilon \cdots n_{\epsilon'} \cdots)$. One defines

$$p_{\epsilon,\epsilon'} = \frac{f(\cdots (n_\epsilon + 1) \cdots (n_\epsilon - 1) \cdots)}{f(\cdots n_\epsilon \cdots n_\epsilon \cdots)}, \quad (42)$$

from which it follows that in the limit $n \to \infty$,

$$p_{\epsilon',\epsilon} = 1/p_{\epsilon,\epsilon'}, \quad \epsilon \neq \epsilon'. \quad (43)$$

Once again we have assumed that $p_{\epsilon,\epsilon'}$ approaches a continuous limit function. If we extend this definition to include the special case $\epsilon' = \epsilon$, $p_{\epsilon,\epsilon} = (p_{\epsilon,\epsilon})^{-1} = 1$, then in our limit, Eq. (40) simplifies to

$$2\sum^\epsilon N_\epsilon \epsilon x_\epsilon - E = \tfrac{1}{2} v \sum^\epsilon N_\epsilon \sum^{\epsilon'} N_{\epsilon'} \times \{(1 - x_\epsilon) x_{\epsilon'} p_{\epsilon,\epsilon'} + (1 - x_{\epsilon'}) x_\epsilon / p_{\epsilon,\epsilon'}\}. \quad (44)$$

Each $p_{\epsilon,\epsilon'}$ is required to be real and continuous in the ground state, with respect to variations in any of the independent variables x_ϵ, or of the parameters ϵ and ϵ'. For example, we must find

$$\lim_{\epsilon' = \epsilon} p_{\epsilon,\epsilon'} = 1, \quad \text{and} \quad p_{\epsilon,\epsilon''} p_{\epsilon'',\epsilon'} = p_{\epsilon,\epsilon'}, \quad (45)$$

but these conditions will be trivially satisfied by our solution. To investigate the continuity with respect to the independent variables, we isolate an arbitrary term on the right-hand side of Eq. (44), and combine all the other terms with the left-hand side. Thus,

$$\alpha = \beta p_{\epsilon,\epsilon'} + (\gamma / p_{\epsilon,\epsilon'}), \quad (46)$$

where

$$\alpha = 2\sum^\epsilon N_\epsilon \epsilon x_\epsilon - E - \tfrac{1}{2} v \sum^{\epsilon''} N_{\epsilon'} \sum^{\epsilon'''} N_{\epsilon''} \times \left\{ (1 - x'') x''' p_{\epsilon'',\epsilon'''} + \frac{(1 - x''') x''}{p_{\epsilon'',\epsilon''}} \right\}, \quad (\epsilon'',\epsilon''') \neq (\epsilon,\epsilon'), \ \neq (\epsilon',\epsilon), \quad (46a)$$

$$\beta = \tfrac{1}{2} v N_\epsilon N_{\epsilon'} (1 - x_\epsilon) x_{\epsilon'}, \quad (46b)$$

and

$$\gamma = \tfrac{1}{2} v N_\epsilon N_{\epsilon'} (1 - x_{\epsilon'}) x_\epsilon. \quad (46c)$$

The "boundary conditions" are

$$p_{\epsilon,\epsilon'} = \alpha/\beta \quad \text{when} \quad \gamma = 0, \quad (47)$$

$$p_{\epsilon,\epsilon'} = \gamma/\alpha \quad \text{when} \quad \beta = 0, \quad (48)$$

whereas the general solution is

$$p_{\epsilon,\epsilon'} = \frac{\alpha}{2\beta} \pm \left[\left(\frac{\alpha}{2\beta} \right)^2 - \frac{\gamma}{\beta} \right]^{\frac{1}{2}}. \quad (49)$$

Continuity might require that at some point $p_{\epsilon,\epsilon'}$ have a cusp. That is, the discriminant must vanish at some point, and

$$\alpha = 2(\beta\gamma)^{\frac{1}{2}}. \quad (50)$$

The reality condition requires that

$$\frac{\partial}{\partial t} \left\{ \left(\frac{\alpha}{2\beta} \right)^2 - \frac{\gamma}{\beta} \right\} = 0, \quad (51)$$

where t is representative of any variable in the problem. This is already quite similar to the one-step problem, and suggestive of the BCS equations, but the derivation is not yet complete. Anyhow, for each pair (ϵ,ϵ'), there exists a value of x_ϵ and $x_{\epsilon'}$ (which we shall denote h_ϵ and $h_{\epsilon'}$) for which

$$p_{\epsilon,\epsilon'}=\frac{\alpha}{2\beta}=\left(\frac{\gamma}{\beta}\right)^{\frac{1}{2}}=\left(\frac{(1-h_{\epsilon'})h_\epsilon}{(1-h_\epsilon)h_{\epsilon'}}\right)^{\frac{1}{2}} \tag{52}$$

by Eqs. (49) and (50). At this point, Eq. (44) reads

$$E=2\sum^{\epsilon}N_\epsilon\epsilon h_\epsilon -v\sum^{\epsilon}N_\epsilon\sum^{\epsilon'}N_{\epsilon'}[h_\epsilon h_{\epsilon'}(1-h_\epsilon)(1-h_{\epsilon'})]^{\frac{1}{2}}. \tag{53}$$

We also investigate Eq. (44) in the neighborhood of this point. Let $n_\epsilon=N_\epsilon\cdot h_\epsilon+\delta n$, $n_{\epsilon'}=N_{\epsilon'}\cdot h_\epsilon-\delta n$, and all other occupations remain fixed. For infinitesimal δn, one finds a differential equation, which after some simplification reduces to

$$2\epsilon-\frac{v}{2}\sum^{\epsilon''}N_{\epsilon''}\sum^{\epsilon'''}N_{\epsilon'''} \times\left\{(1-h'')h'''-\frac{(1-h''')h''}{p^2_{\epsilon'',\epsilon'''}}\right\}\frac{1}{N_\epsilon}\frac{d}{dx_\epsilon}p_{\epsilon'',\epsilon'''}\Big|_{x_\epsilon=h_\epsilon} -v\sum_{\epsilon''\neq\epsilon}^{\epsilon''}N_{\epsilon''}\left\{-h''p_{\epsilon,\epsilon''}+\frac{(1-h'')}{p_{\epsilon,\epsilon''}}\right\} -vN_\epsilon(1-2h_\epsilon)=\text{const.} \tag{54}$$

In general, we don't know the value of $d/dx_\epsilon(p_{\epsilon'',\epsilon'''})$, not even at the point in question. However, it is finite, and by Eq. (52), its coefficient vanishes.

$$2\epsilon-v\left(\frac{1-2h_\epsilon}{[h_\epsilon(1-h_\epsilon)]^{\frac{1}{2}}}\right) \times\sum^{\epsilon''}N_{\epsilon''}[h_{\epsilon''}(1-h_{\epsilon''})]^{\frac{1}{2}}=\text{const.} \tag{55}$$

Following BCS, this is solved by defining the gap parameter ϵ_0

$$\epsilon_0=v\sum^{\epsilon''}N_{\epsilon''}[h_{\epsilon''}(1-h_{\epsilon''})], \tag{56}$$

from which it follows that

$$h_\epsilon=\tfrac{1}{2}[1-\tilde{\epsilon}/(\tilde{\epsilon}^2+\epsilon_0^2)^{\frac{1}{2}}], \tag{57}$$

where $\tilde{\epsilon}=\epsilon-\frac{1}{2}$ const. To determine this constant, we refer back to Eqs. (38) and (39) which, upon being combined, yield the condition

$$\sum^{\epsilon}N_\epsilon\tilde{\epsilon}/(\tilde{\epsilon}^2+\epsilon_0^2)^{\frac{1}{2}}=0. \tag{58}$$

It is easy to see that this constant is the chemical potential for a pair 2μ, which is conventionally determined by the condition that the total number of particles be fixed, as here. If N_ϵ is approximately a constant function of ϵ, then Eq. (58) can be written as

$$\int_{E_F-\hbar\omega}^{E_F+\hbar\omega}d\epsilon\frac{\tilde{\epsilon}}{(\tilde{\epsilon}^2+\epsilon_0^2)^{\frac{1}{2}}}=0, \tag{59}$$

and it is seen that μ is independent of ϵ_0 and is equal to its unperturbed value which we denote by E_F. Otherwise, one defines

$$\phi(\epsilon)=N_\epsilon/N_{2\mu}, \tag{60}$$

and Eq. (58) becomes

$$\int_{E_F-\hbar\omega}^{E_F+\hbar\omega}d\epsilon\phi(\epsilon)\frac{\tilde{\epsilon}}{(\tilde{\epsilon}^2+\epsilon_0^2)^{\frac{1}{2}}}=0. \tag{61}$$

This is an implicit equation for the chemical potential and, in general, μ can be a function of ϵ_0.

The ground-state energy is simply obtained by substituting the values of h_ϵ determined by Eqs. (57) and (58) into Eq. (53), as in reference 1.

This concludes our derivation of the equations of superconductivity based on an analysis of the properties of the exact eigenfunction of the reduced Hamiltonian (1). In the following section, we conclude our verification of the BCS theory by showing that the point $\{x_\epsilon\}=\{h_\epsilon\}$ is a stationary point, in the sense that as $n\to\infty$, the contribution of the various configurations to the wave function becomes essentially a delta function centered about this point, and that, therefore, the BCS trial function (or any other trial function which is correct in the neighborhood of this point) becomes asymptotically exact in this limit, and not just the variational energy.

V. THE STATIONARY POINT

In the limit of infinite volume, only certain configurations contribute significantly to the wave-function normalization integral, and also in the calculation of matrix elements to the low-lying excited states. We have seen that the BCS equations are exact in the neighborhood of a certain point in occupation-number-space. We shall now show that this is also the stationary point, and that the BCS wave function correctly weights the relative amplitudes of different configurations in the neighborhood of this point, provided care is exercised in conserving particles.

We investigate the one-step model,[4] for which the wave-function normalization requires

$$1=\sum_{n_\epsilon=0}^{n}\left[\frac{n!}{n_\epsilon!(n-n_\epsilon)!}\right]^2 f^2(n_\epsilon). \tag{62}$$

The first factor is the number of ways we can have the occupation number n_ϵ, i.e., the number of distinct configurations belonging to the same value of n_ϵ. In

[4] The generalization to the model of Sec. IV would be repetitious and will be omitted.

the limit $n\to\infty$, both this factor and $f^2(n_\epsilon)$ are very rapidly varying functions of n_ϵ, and most of the contribution comes from a neighborhood of the point where the summand has a maximum. (The sum could be replaced by an integral at this point and evaluated by the method of steepest descents.)

Let the stationary point be at $\bar{n}_\epsilon$, and let us factor from the sum the value of the summand at this point.

$$1=\left(\frac{n!f(\bar{n}_\epsilon)}{\bar{n}_\epsilon!(n-\bar{n}_\epsilon)!}\right)^2\left(1+\frac{(n-\bar{n}_\epsilon)^2}{(\bar{n}_\epsilon+1)^2}\frac{f^2(\bar{n}_\epsilon+1)}{f^2(\bar{n}_\epsilon)}+\frac{\bar{n}_\epsilon^2}{(n-\bar{n}_\epsilon+1)}\right.$$
$$\times\frac{f^2(\bar{n}_\epsilon-1)}{f^2(\bar{n}_\epsilon)}+\frac{(n-\bar{n}_\epsilon)^2(n-\bar{n}_\epsilon-1)^2}{(\bar{n}_\epsilon+1)^2(\bar{n}_\epsilon+2)^2}\frac{f^2(\bar{n}_\epsilon+2)}{f^2(\bar{n}_\epsilon)}$$
$$\left.+\frac{\bar{n}_\epsilon^2(\bar{n}_\epsilon-1)^2}{(n-\bar{n}_\epsilon+1)^2(n-\bar{n}_\epsilon+2)^2}\frac{f^2(\bar{n}_\epsilon-2)}{f^2(\bar{n}_\epsilon)}+\cdots\right). \quad (63)$$

In our limit,

$$1=\left(\frac{n!f(\bar{n}_\epsilon)}{\bar{n}_\epsilon!(n-\bar{n}_\epsilon)!}\right)^2\left(1+\frac{(1-\bar{x}_\epsilon)^2}{\bar{x}_\epsilon^2}p^2(\bar{x}_\epsilon)-O\left(\frac{1}{n}\right)\right.$$
$$+\frac{\bar{x}_\epsilon^2}{(1-\bar{x}_\epsilon)^2}\frac{1}{p^2(\bar{x}_\epsilon)}-O\left(\frac{1}{n}\right)+\frac{(1-\bar{x}_\epsilon)^4}{\bar{x}_\epsilon^4}p^4(\bar{x}_\epsilon)$$
$$\left.-O\left(\frac{1}{n}\right)+\frac{\bar{x}_\epsilon^4}{(1-\bar{x}_\epsilon)^4}\frac{1}{p^4(\bar{x}_\epsilon)}-O\left(\frac{1}{n}\right)+\cdots\right). \quad (64)$$

To order $1/n$, all the terms in the neighborhood of $\bar{x}_\epsilon\equiv\bar{n}_\epsilon/n$ must contribute equally, therefore,

$$p(\bar{x}_\epsilon)=\bar{x}_\epsilon/(1-\bar{x}_\epsilon). \quad (65)$$

However, comparing this with Eq. (27), we see that

$$\bar{x}_\epsilon=h_\epsilon, \quad (66)$$

and indeed the stationary point is the same as the turning point at which the discriminant of Sec. III vanished. As this is the only point of interest in the calculation of the normalization integral (and of low-lying matrix elements), we must verify that the trial function has the right amplitudes at and near this point.

The BCS function is

$$\psi=\prod_k([1-h(\epsilon_k)]^{\frac{1}{2}}+[h(\epsilon_k)]^{\frac{1}{2}}b_k^+)\,|0\rangle, \quad (67)$$

and is evidently normalized. For the one-step model ($\epsilon_k=0$ or ϵ), h_ϵ is the same as in our Eq. (29), and $h_0=1-h_\epsilon$. Decomposing the function (67) into configurations of distinct n_0 and n_ϵ, we find that the trial amplitudes do correctly depend only on these parameters, but that

$$n_0+n_\epsilon\neq n, \quad (68)$$

so that the trial function does not conserve pairs, as has already been noted. For any fixed value of n_0+n_ϵ, the ratio of the trial amplitude for the configuration (n_0+n_ϵ) to the trial amplitude corresponding to $(n_0+q, n_\epsilon-q)$ is

$$\text{BCS ratio of amplitudes}=(h_\epsilon/[1-h_\epsilon])^q, \quad (69)$$

and is correct for any finite positive or negative integer q (in the limit $n\to\infty$). Moreover, the *average* value of n_0+n_ϵ in the trial function is n; therefore, such quantities as the energy, which are insensitive to the exact number of particles, can be accurately computed with the trial function, as we have already discovered in the preceding sections. This ratio is incorrect for very large values of such that $q/n\neq 0$, except in strong-coupling, where the ratio is correctly given as unity for all q. This suggests that the trial function (or the equivalent Bogoliubov transformation) be handled with some care; but because it is correct at the stationary point, this function does asymptotically, and on the average, approach the exact eigenfunction of the problem as $n\to\infty$. Many investigators have already shown that the variational ground-state energy of the reduced Hamiltonian is exact in an asymptotic sense,[5-7] but as the variational theorem does not imply an equivalent accuracy in the wave function, the present analysis has not been in any sense redundant.

APPENDIX

This section is rather mathematical and concerns the intrinsic error in approximating the nonlinear difference equation for the p functions by a quadratic equation such as (23) or (46). Once we establish that the error is of order n^{-1}, we can calculate this error to leading order to see the effect of finite-volume corrections on the theory.

The error analysis proceeds in several steps. We shall show that:

(a) $p(x)$ approaches a limit function as $n\to\infty$ and that this limit function obeys the correct boundary conditions provided the discriminant vanishes at least at one point in the interval (0,1).

(b) The lowest energy is such that the discriminant vanishes only at one point, the "critical point."

(c) The limit function which $p(x)$ approaches is the solution to the quadratic equation, except in the neighborhood of the critical point.

Let the primitive equation be (for simplicity, we depart slightly from the notation in the text)

$$a(y)p(y)p(y+1/n)-2b(y)p(y)+c(y)=0, \quad 0\leq y\leq 1, \quad (A.1)$$

where this equation holds for all $y=\text{integer}/n$ in the interval; and let $g(y)$ be the solution to the quadratic

[5] P. W. Anderson, Phys. Rev. 112, 1900 (1958).
[6] J. Bardeen and G. Rickayzen, Phys. Rev. 118, 936 (1960).
[7] N. N. Bogoliubov, D. N. Zubarev, and Yu. A. Tserkovnikov, Soviet Phys.-JETP 12, 88 (1960).

equation

$$a(y)g^2(y)-2b(y)g(y)+c(y)=0. \tag{A.2}$$

The coefficients have the properties, $c(0)=a(1)=0$, $b(y)\neq 0$. First, we show that if y_0 and y_1 are, respectively, the least and the greatest points at which the discriminant $D(y)$ vanishes,

$$D\equiv b^2-ac=0, \tag{A.3}$$

then $p(y)$ approaches a continuous limit function as $n\rightarrow\infty$, in the regions $(0,y_0)$ and $(y_1,1)$. The proof for the first region is as follows: let

$$g(y+1/n)\equiv g(y)+(1/n)\Omega(y), \tag{A.4}$$

and

$$p(y)\equiv g(y)+(1/n)S(y). \tag{A.5}$$

If we choose the correct solution to (A.2) in this region, namely,

$$g(y)=\frac{b(y)+[D(y)]^{\frac{1}{2}}}{a(y)}, \tag{A.6}$$

it can be directly verified that $S(y)$ is of order unity in the immediate neighborhood of the point $y=0$. We must now show that this function remains finite on the interval $(0,y_0)$. The function $\Omega(y)$ can be obviously calculated and is of order unity if we exclude a neighborhood of the point y_0. It is also of order unity in that neighborhood if

$$\frac{\partial}{\partial y}D\bigg|_{y=y_0}=0 \text{ (as in the ground state).}$$

Now, we calculate $p(y+1/n)$ by two different methods. Using Eqs. (A.4) and (A.5),

$$p(y+1/n)=g(y)+(1/n)[\Omega(y)+S(y+1/n)], \tag{A.7}$$

and using the primitive equation

$$p\left(y+\frac{1}{n}\right)=\frac{2b(y)}{a(y)}-\frac{c(y)}{a(y)p(y)}. \tag{A.8}$$

Eliminating $p(y)$ by Eq. (A.5), we also assume that $S(y)$ is of order unity, and, therefore,

$$p\left(y+\frac{1}{n}\right)=\frac{2b(y)}{a(y)}-\frac{c(y)}{a(y)g(y)[1+\ S(y)/n\cdot g(y)]}$$

$$=g(y)+\frac{1}{n}\frac{c(y)S(y)}{a(y)g^2(y)}+O\left(\frac{1}{n^2}\right). \tag{A.9}$$

Comparing Eqs. (A.7) and (A.9), we find

$$S(y+1/n)=M(y)S(y)-\Omega'(y), \tag{A.10}$$

where

$$0<M(y)=\frac{c(y)}{a(y)g^2(y)}<1 \quad \text{for} \quad y<y_0, \tag{A.11a}$$

and

$$\Omega'(y)=\Omega(y)+\text{order } 1/n. \tag{A.11b}$$

This difference equation is far simpler than the original equation (A.1). Now, we want to show that $S(y+1/n)$ is finite. An upper limit to S is $\bar{S}$,

$$\bar{S}(y+1/n)=M(y)\bar{S}(y)+\omega, \tag{A.12}$$

where $\omega=Max|\Omega'(y)|$, and is known to be finite. The solution to this equation is

$$\bar{S}(y+1/n)$$
$$=\omega(1+M(y)+M(y)M(y-1/n)$$
$$+M(y)M(y-1/n)M(y-2/n)+\cdots), \tag{A.13}$$

and if $\bar{M}(y)$ is the maximum value of M in $(0,y)$,

$$\bar{S}(y+1/n)<\omega/1-\bar{M}(y), \tag{A.14}$$

and is always finite for $y<y_0$.

A similar proof goes through for the other interval, except that one chooses the other root of the quadratic equation to make p and g agree at $y=1$.

Now, if we use the fact that $b(y)$ decreases monotonically with the energy eigenvalue, then we see that, if the energy is too low, the discriminant can never vanish in $(0,1)$; and both boundary conditions cannot be obeyed by a continuous function [which we have shown $p(y)$ to be]. The lowest value of the energy for which $D(y)=0$ in the interval is such that $y_0=y_1$, i.e., the discriminant vanishes only at one point. Then we have shown that as $n\rightarrow\infty$.

$$p(y)=g(y)=\frac{b(y)+[D(y)]^{\frac{1}{2}}}{a(y)} \quad y<y_0, \tag{A.15}$$

and

$$p(y)=g(y)=\frac{b(y)-[D(y)]^{\frac{1}{2}}}{a(y)} \quad y>y_0. \tag{A.16}$$

Our analysis does not include the immediate neighborhood of y_0. If one wished, he could investigate this critical region (which would involve an analysis similar to that of the WKB approximation at a turning point), and would undoubtedly find that a limit function does not exist here. But as this region can be chosen as small as we please, there is no real point to such an analysis. Nevertheless, we should satisfy ourselves that nothing untoward happens in this region, namely, that our assumption is justified that the lowest energy is that which gives one critical point. As we have mentioned, below this energy there is no solution (to order $1/n$) and, hence, our assumption yields a lower bound; but it agrees asymptotically with the BCS variational solution, which is an upper bound. Hence, it is correct asymptotically, and it must indeed be possible to continue our solution for $p(x)$ through the critical region.

Finally, we should like to calculate the lowest-order correction to the energy. We recall

$$p(y)=\exp\left\{n\left[S(y)-S\left(y-\frac{1}{n}\right)\right]\right\}$$
$$=\exp\left(\frac{\partial S}{\partial y}-\frac{1}{2n}\frac{\partial^2 S}{\partial y^2}+\cdots\right), \quad \text{(A.17)}$$

and

$$p\left(y+\frac{1}{n}\right)=\exp\left\{n\left[S\left(y+\frac{1}{n}\right)-S(y)\right]\right\}$$
$$=\exp\left(\frac{\partial S}{\partial y}+\frac{1}{2n}\frac{\partial^2 S}{\partial y^2}+\cdots\right).$$

Define

$$\exp\partial S/\partial y=\tilde{g}(y)\cong g(y), \quad \text{(A.18)}$$

and to order n^{-2},

$$\exp\left(\frac{1}{2n}\frac{\partial^2 S}{\partial y^2}\right)=\exp\left(\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right), \quad \text{(A.19)}$$

where $g(y)$ is given in Eqs. (A.15) and (A.16).

With these substitutions, the primitive equation becomes

$$\left(a\exp\left[\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right]\right)\tilde{g}^2(y)-2b\tilde{g}(y)$$
$$+\left(c\exp\left[\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right]\right)=0, \quad \text{(A.20)}$$

and if we note that both a and c are proportioned to the interaction v, we see that the interactions off the energy shell have been increased from a strength v to an effective strength

$$\tilde{v}\equiv v\exp\left[\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right]\approx v\left[1+\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right], \quad \text{(A.21)}$$

which is greater than v because, in the important region near y_0,

$$(d/dy)\lg(y)>0, \quad y\approx y_0. \quad \text{(A.22)}$$

Consequently, the ground-state energy divided by the number of particles actually must increase as the volume is decreased (always at fixed density). For $n\gg 1$, this correction is quite negligible, and it always vanishes in the strong-coupling limit (in which $g(y)=1$, $\partial/\partial y[\lg(y)]=0$). In the weak-coupling limit, or for the one-step model, this correction has the effect of slightly increasing the critical temperature for very small volume crystals.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 4 JULY-AUGUST, 1961

Some Cluster Size and Percolation Problems

MICHAEL E. FISHER AND JOHN W. ESSAM
Wheatstone Physics Laboratory, King's College, London, England
(Received December 15, 1960)

The problem of cluster size distribution and percolation on a regular lattice or graph of bonds and sites is reviewed and its applications to dilute ferromagnetism, polymer gelation, etc., briefly discussed. The cluster size and percolation problems are then solved exactly for Bethe lattices (infinite homogeneous Cayley trees) and for a wide class of pseudolattices derived by replacing the bonds and/or sites of a Bethe lattice by arbitrary finite subgraphs. Explicit expressions are given for the critical probability (density), for the mean cluster size, and for the density of infinite clusters. The nature of the critical anomalies is shown to be the same for all lattices discussed; in particular, the density of infinite clusters vanishes as $R(p)\approx C(p-p_c)$ $(p\geqslant p_c)$.

I. INTRODUCTION

RECENTLY Domb[1] has drawn attention to the problem of determining the distribution of cluster sizes for particles distributed in a medium in accordance with a statistical law. In the simplest case, the particles occupy at random the sites of a lattice (or, more generally, the vertices of a linear graph). Each site can accommodate one (and only one) particle and is occupied with a constant probability p. A group of particles which can be linked together by nearest-neighbor *bonds* from one occupied lattice site to an adjacent occupied site are said to form a *cluster*. The main theoretical task is to evaluate the mean cluster size and higher moments of the distribution as functions of the density (or concentration) of the particles, this being measured by the probability p.

[1] C. Domb, Conference on "Fluctuation phenomena and stochastic processes" at Birkbeck College, London, March 1959; Nature 184, 509 (1959).

Finally, we should like to calculate the lowest-order correction to the energy. We recall

$$p(y)=\exp\left\{n\left[S(y)-S\left(y-\frac{1}{n}\right)\right]\right\}$$
$$=\exp\left(\frac{\partial S}{\partial y}-\frac{1}{2n}\frac{\partial^2 S}{\partial y^2}+\cdots\right), \quad \text{(A.17)}$$

and

$$p\left(y+\frac{1}{n}\right)=\exp\left\{n\left[S\left(y+\frac{1}{n}\right)-S(y)\right]\right\}$$
$$=\exp\left(\frac{\partial S}{\partial y}+\frac{1}{2n}\frac{\partial^2 S}{\partial y^2}+\cdots\right).$$

Define

$$\exp\partial S/\partial y=\tilde{g}(y)\cong g(y), \quad \text{(A.18)}$$

and to order n^{-2},

$$\exp\left(\frac{1}{2n}\frac{\partial^2 S}{\partial y^2}\right)=\exp\left(\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right), \quad \text{(A.19)}$$

where $g(y)$ is given in Eqs. (A.15) and (A.16).

With these substitutions, the primitive equation becomes

$$\left(a\exp\left[\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right]\right)\tilde{g}^2(y)-2b\tilde{g}(y)$$
$$+\left(c\exp\left[\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right]\right)=0, \quad \text{(A.20)}$$

and if we note that both a and c are proportioned to the interaction v, we see that the interactions off the energy shell have been increased from a strength v to an effective strength

$$\tilde{v}\equiv v\exp\left[\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right]\approx v\left[1+\frac{1}{2n}\frac{\partial}{\partial y}\lg(y)\right], \quad \text{(A.21)}$$

which is greater than v because, in the important region near y_0,

$$(d/dy)\lg(y)>0, \quad y\approx y_0. \quad \text{(A.22)}$$

Consequently, the ground-state energy divided by the number of particles actually must increase as the volume is decreased (always at fixed density). For $n\gg 1$, this correction is quite negligible, and it always vanishes in the strong-coupling limit (in which $g(y)=1$, $\partial/\partial y[\lg(y)]=0$). In the weak-coupling limit, or for the one-step model, this correction has the effect of slightly increasing the critical temperature for very small volume crystals.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 4 JULY-AUGUST, 1961

Some Cluster Size and Percolation Problems

MICHAEL E. FISHER AND JOHN W. ESSAM
Wheatstone Physics Laboratory, King's College, London, England
(Received December 15, 1960)

The problem of cluster size distribution and percolation on a regular lattice or graph of bonds and sites is reviewed and its applications to dilute ferromagnetism, polymer gelation, etc., briefly discussed. The cluster size and percolation problems are then solved exactly for Bethe lattices (infinite homogeneous Cayley trees) and for a wide class of pseudolattices derived by replacing the bonds and/or sites of a Bethe lattice by arbitrary finite subgraphs. Explicit expressions are given for the critical probability (density), for the mean cluster size, and for the density of infinite clusters. The nature of the critical anomalies is shown to be the same for all lattices discussed; in particular, the density of infinite clusters vanishes as $R(p)\approx C(p-p_c)$ $(p\geqslant p_c)$.

I. INTRODUCTION

RECENTLY Domb[1] has drawn attention to the problem of determining the distribution of cluster sizes for particles distributed in a medium in accordance with a statistical law. In the simplest case, the particles occupy at random the sites of a lattice (or, more generally, the vertices of a linear graph). Each site can accommodate one (and only one) particle and is occupied with a constant probability p. A group of particles which can be linked together by nearest-neighbor *bonds* from one occupied lattice site to an adjacent occupied site are said to form a *cluster*. The main theoretical task is to evaluate the mean cluster size and higher moments of the distribution as functions of the density (or concentration) of the particles, this being measured by the probability p.

[1] C. Domb, Conference on "Fluctuation phenomena and stochastic processes" at Birkbeck College, London, March 1959; Nature 184, 509 (1959).

The problem has applications in various physical contexts, the distribution of grain size in sands and in photographic emulsions,[2] the behavior of dilute ferromagnets[3,4] and other diluted cooperative assemblies, the vulcanization of rubber and the formation of crosslinked polymer gels,[5] the clustering of impurities and defects in crystals, etc.

An alternative version of the problem is to consider the occupation of the *bonds* of a lattice (or linear graph). Each bond can be occupied with probability p; occupied bonds which meet at a lattice site are considered as linked together in a cluster. In this form, the cluster size problem is very closely related to the *percolation processes* introduced by Broadbent and Hammersley[6,7] and since discussed by Hammersley[8,9] and Harris.[10] In fact, if for "occupied" one reads "open" and for "unoccupied" reads "closed," a lattice with particles distributed on bonds becomes a *randomly dammed maze* such as considered by Broadbent and Hammersley. The flow of *fluid* through such a maze constitutes a percolation process and serves as a model for the diffusion of gas molecules through a porous solid, the spread of disease in an orchard, etc.[8] Interest centers on determining the subsequent distribution of fluid and the number of wetted "atoms" (i.e., sites). Clearly, a similar transposition can be made when sites rather than bonds are occupied. We shall refer to the two versions of the general problem as the *site problem* and the *bond problem*, respectively.

One of the interesting features of these problems is the existence of a *critical probability* p_c above which *unbounded clusters of infinite size* are formed in the lattice with a definite density. This phenomenon has significant physical implications. Thus, in a dilute ferromagnetic system, p_c represents the minimum concentration of ferromagnetic atoms necessary before long-range order can set in, and so marks the limit of the cooperative phase transition.[3,4] When $p=p_c$ the ferromagnetic Curie point occurs at zero temperature. When considering the gelation of polymers, p_c is proportional to the minimum number of crosslinks per molecule needed to ensure gel formation. The probability of a particle belonging to an infinite cluster then measures the gel fraction of polymer in relation to the sol fraction (finite clusters). For percolation processes, the formation of infinite clusters implies that the medium attains a finite (nonzero) permeabillty so that fluid will percolate indefinitely away from a source instead of being confined to the local neighborhood of the source. If the probability of cross infection in the spread of a disease exceeds the critical value, an epidemic occurs.

Bounds to the critical probability for various lattices have been obtained by Hammersley[7–9] and Harris,[10] but Domb has shown how p_c can be estimated directly from knowledge of the cluster size distributions.[1] For most physical applications, furthermore, it is useful to have more detailed information on the size distribution. In particular, the mean cluster size and the density of infinite clusters are expected to exhibit singularities at $p=p_c$ and consequently the behavior in the critical region is of considerable interest. Progress in the solution of these problems can be made by studying series expansions based on the enumeration of lattice configurations.[1] Unfortunately for the lattices of principal interest, the standard plane and three-dimensional lattices, it seems to be rather difficult to formulate a direct theoretical approach leading to solutions in closed form. Appreciable insight can be obtained, however, by examining pseudolattices such as the *Bethe lattices*[11] (i.e., infinite homogeneous Cayley trees) and, for example, the various (triangular) *cacti*[12] illustrated in Fig. 1.

As we show in this note, the cluster size problem can be solved exactly and in full detail for a wide class of pseudolattices of this general type. The relation between the behavior of these pseudolattices and that of the normal space lattices is quite closely analogous to that between the results of approximations like that of Bethe and of Rushbrooke and Scoins[11] and the consequences of the exact treatment of order-disorder phenomena on the corresponding lattices. Furthermore, the exact results for the Bethe lattices are useful in obtaining the series expansions for the normal lattices since only relatively few configurational corrections have to be made.[13] One might also hope, by examining the form of the exact solutions for various pseudolattices which allow only a limited number of closed configurations (subgraphs), to discover a general development analogous to that of Yvon for the order-disorder problem.[11] As yet, however, we have not been able to achieve this.

The theory for pseudolattices is developed in the remainder of this paper. The general approach via generating functions is outlined in Sec. II. In the following section, the generating functions for the Bethe lattices are derived. Some of these results have been obtained previously in the special context of polymer gelation by Flory,[5] and the theory is related to known results in the theory of branching processes (multiplicative or cascade processes).[14] In Sec. IV, the gener-

[2] F. Kottler, J. Franklin Inst. **250**, 339, 419 (1950); J. Phys. Chem. **56**, 442 (1952).
[3] H. Sato, A. Arrott, and R. Kikuchi, J. Phys. Chem. Solids **10**, 19 (1959).
[4] R. J. Elliott, B. R. Heap, D. J. Morgan, and G. S. Rushbrooke, Phys. Rev. Letters **5**, 366 (1960).
[5] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, New York, 1953), Chap. 9.
[6] S. R. Broadbent and J. M. Hammersley, Proc. Cambridge Phil. Soc. **53**, 629 (1957).
[7] J. M. Hammersley, Proc. Cambridge Phil. Soc. **53**, 642 (1957).
[8] J. M. Hammersley, Ann. Math. Stat. **28**, 790 (1957).
[9] J. M. Hammersley, Proc. 87th Intern. Colloq., "Probabilités et ses Applications" (Paris, 1959), p. 17.
[10] T. E. Harris, Proc. Cambridge Phil. Soc. **56**, 13 (1960).
[11] C. Domb, Advances in Phys. **9** (1960), see pp. 283–284.
[12] F. Harary and G. E. Uhlenbeck, Proc. Natl. Acad. Sci. U. S. **39**, 315 (1953).
[13] M. F. Sykes (private communication, to be published).
[14] T. E. Harris, Ann. Math. Stat. **19**, 474 (1948).

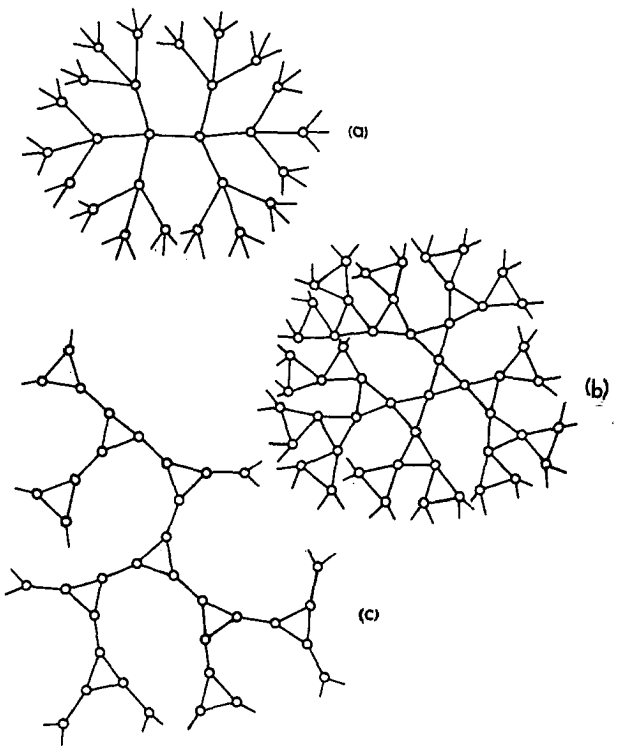


FIG. 1. Various simple pseudolattices: (a) Bethe lattice of coordination number $\sigma+1=4$, (b) simple (triangular) cactus of coordination number 4, (c) expanded cactus of coordination number 3.

ating functions are inverted to yield explicit formulas for the number of Cayley trees on a Bethe lattice and hence, for the other configurational coefficients. We then show how the generating functions can be modified to cover the "decoration" of each bond of a Bethe lattice by an arbitrary finite "bond-graph." In Sec. VI a similar procedure is carried through for the case where, in addition, the sites of a Bethe lattice (of coordination number $\sigma+1=3$) are also replaced by a specified "site-graph." This enables us, for example, to give explicit expressions for the mean cluster size and density of infinite clusters for the bond and site problems on the triangular cacti shown in Fig. 1. Other transformations and the possibility of generalizing the site decoration theory to $\sigma \gtrless 3$ are discussed in Secs. VII and VIII. The variation of the mean cluster size and other properties in the critical region are found to have the same analytic form for all lattices derived from the Bethe lattice, although the standard plane and three-dimensional lattices are expected to exhibit singularities of a different type.

II. GENERATING FUNCTIONS

Following Domb,[1] we approach the site problem on a general lattice by asking for the probability that a given site chosen at random is occupied by a particle belonging to a cluster of exactly s particles. For example, the probability of a given site belonging to a one cluster (isolated particle) is the probability that the chosen site is occupied times the probability that all the nearest-neighbor sites are unoccupied. For a lattice of coordination number $\sigma+1$, this is just $p_1=pq^{\sigma+1}$ where $q=1-p$. More generally, if a_{st} is the *number of distinct clusters of size s and perimeter t which contain the given lattice site* (the *perimeter* being the minimum number of unoccupied sites required to isolate the cluster), then the probability of a site belonging to such a cluster is

$$p_{st}=a_{st}p^sq^t. \tag{1}$$

Complete information on the cluster size distribution is thus contained in the generating function

$$A(x,y)=\sum_{s,t} a_{st}x^sy^t, \tag{2}$$

where the sum is over all possible clusters of finite size and perimeter that can occur on the lattice. In particular, the total probability that a site belongs to a finite cluster is

$$F(p)=A(p,q).$$

For small enough p, infinite clusters will be absent and $F(p)$ is merely the probability that the a site is occupied. Consequently, we obtain the basic identity

$$F(p)=A(p,q)\equiv p \quad \text{for } p<p_c. \tag{3}$$

On the other hand, for particle densities greater than the critical density, infinite clusters will spread through the lattice and the probability that a site belongs to an infinite cluster will be

$$R(p)=p-A(p,q). \tag{4}$$

The vanishing of $R(p)$ defines the critical probability p_c. The mean size density of clusters at a site is defined by

$$\langle s\rangle=\sum_{s,t} sp_{st}, \tag{5}$$

and so

$$\langle s\rangle=[x\partial A/\partial x]_{x=p,y=q}. \tag{6}$$

This relation holds for all p, but above p_c it represents the mean size of *finite* clusters only. Accordingly, it is convenient to normalize $\langle s\rangle$ by dividing by the probability $F(p)$ that a site belongs to a finite cluster. [Below p_c this is just equal to p, but above p_c one must use $A(p,q)$.] Thus,

$$S(p)=[x(\partial/\partial x)\ln A(x,y)]_{x=p,y=q}, \tag{7}$$

which for $p<p_c$ reduces simply to

$$S(p)=[\partial A/\partial x]_{x=p,y=q} \quad (p<p_c). \tag{8}$$

The mean cluster size will exhibit a sharp maximum at $p=p_c$ and this may be used to define the critical point. For all the pseudolattices which are soluble, the maximum in $S(p)$ is an infinite singularity and most probably this is generally true.

Higher moments of the cluster size distribution can be calculated by further differentiation of $A(x,y)$ with

respect to x. Differentiation with respect to y yields moments of the perimeter distribution. Individual cluster contributions may be found by picking out the coefficients of the appropriate powers of x and y.

We remark parenthetically that $S(p)$ represents a "weight average" rather than a "number average," i.e., if there are n_s clusters of size s in the lattice, then $S(p)=\sum s(sn_s)/\sum sn_s$. For some purposes the number average size $S_0(p)=\sum sn_s/\sum n_s$ may be of interest. This can be calculated from $S_0(p)=[x(\partial/\partial x)\ \ln K(x,y)]$ with $x=p$ and $y=q$, where $K(x,y)$ is defined below. For pseudolattices it is found that $S_0(p)$ remains finite and continuous at $p=p_c$ but has a sharp peak as a result of a discontinuous change in the sign of the gradient.

For theoretical purposes, the configurational generating function

$$K(x,y)=\sum_{s,t} k_{st}x^s y^t \tag{9}$$

is more convenient than $A(x,y)$. This may be defined by

$$k_{st}=a_{st}/s \tag{10}$$

so that

$$A(x,y)=x(\partial/\partial x)K(x,y). \tag{11}$$

Alternatively, the coefficient k_{st} is defined as the number of cluster configurations of size s and perimeter t *per site* of the lattice. For a finite lattice of N sites, this is the total number of distinct clusters of a given type that can be placed on the lattice divided by the number of sites. In general, this will depend on N, but for any uniform d-dimensional lattice edge effects fall off relatively as $N^{-1/d}$, so that for large N, k_{st} becomes a *lattice constant*[11] independent of N. In a similar fashion, one may define the number of configurations per bond. If we indicate lattice constants with respect to sites by a superscript S and with respect to bonds by a superscript B, it is clear that

$$k^S=f^S k^B, \tag{12}$$

where f^S is the number of bonds per site. Since $\sigma+1$ bonds radiate from each site and each bond is associated with two sites this is just $f^S=\frac{1}{2}(\sigma+1)$. Other transformation formulas may be written down in the same way.

The Bethe lattices and similar pseudolattices may be regarded as lattices of infinite dimension since in a finite Bethe lattice the relative number of sites in the edge is of order $1(\approx N^{-1/\infty})$. Consequently, the previous definition of a lattice constant breaks down. The definition may be extended in an unambiguous fashion, however, by introducing a convergence factor. If l is the least number of lattice steps from the origin to a characteristic point in an individual configuration, the factor $e^{-\beta l}$ is included in the sums for the total number of configurations and total number of sites (or bonds). These sums then define analytic functions of β, but after taking ratios to evaluate the lattice constant, β may be equated to zero. The lattice constants so defined have just the same transformation properties as on normal lattices.

The analysis of this section has been in terms of the site problem, but the only changes required for the bond problem is the use of lattice constants per bond rather than per site.

III. BETHE LATTICES

A. Bond Problem

Consider now the bond problem on the simple Bethe lattice of coordination number $\sigma+1$ as illustrated in Fig. 1(a). We observe that the perimeter of a cluster of s occupied bonds is given uniquely by

$$t=(\sigma-1)s+\sigma+1. \tag{13}$$

This follows by noting that the perimeter of a single bond is 2σ, and that whenever a new bond is added to a cluster, one bond of the original perimeter is lost but σ new unoccupied bonds must be added to form the new perimeter. It follows that the configurational generating function (9) is

$$K^B(x,y)=y^{\sigma+1}\sum_{s=0}^{\infty} b_s x^s y^{(\sigma-1)s}, \tag{14}$$

where $b_s=k_{s,(\sigma-1)s+\sigma+1}$ is the total number of s clusters (of bonds) per bond of the Bethe lattice. To simplify the analysis of the site problem, we have included a coefficient b_0 equal to the number of sites per bond. Its presence in (14) has no effect on the density of infinite clusters or on the other properties considered. The expression (14) may be rewritten in terms of the fundamental Bethe lattice generating function

$$B_\sigma(Z)=\sum_{s=0}^{\infty} b_s Z^s \tag{15}$$

as

$$K^B(x,y)=y^{\sigma+1}B_\sigma(Z), \tag{16}$$

where

$$Z(x,y)=xy^{\sigma-1}. \tag{17}$$

To calculate $B_\sigma(z)$ explicitly, we use the fundamental identity (3), namely,

$$F(p)=A(p,q)\equiv p \quad (p<p_c).$$

Now

$$A(x,y)=xy^{2\sigma}B_\sigma'(Z), \tag{18}$$

where the prime denotes differentiation with respect to Z. Thus, if

$$z=z(p)=Z(p,q)=p(1-p)^{\sigma-1}, \tag{19}$$

the generating function must satisfy

$$B_\sigma'[z(p)]=G(p)=(1-p)^{-2\sigma} \tag{20}$$

for small enough p. Now $B_\sigma(z)$ is a function only of z, but z is defined by (19) as a function of p for *all* p. To a given value of z, however, correspond *two* values of p, one of which tends to zero with z while the other

tends to unity. Consequently, if we define $p^*(p)$ to be the root of the equation

$$p^*(1-p^*)^{\sigma-1}=p(1-p)^{\sigma-1}=z \tag{21}$$

which vanishes continuously with z (and hence as $p\to 0$ and as $p\to 1$), we may rewrite (20) as

$$G(p)=[1-p^*(p)]^{-2\sigma}. \tag{22}$$

In this form the result is thus valid for all p, and so the probability of a site belonging to a finite cluster is

$$F(p)=p(1-p)^{2\sigma}G(p)=p[(1-p)^{2\sigma}/(1-p^*)^{2\sigma}]. \tag{23}$$

Now z, as a function of p, attains a simple maximum at $p_m=1/\sigma$, which implies that the root of (22) which vanishes with z when $p\leq p_m$ is simply $p^*=p$. For $p>p_m$, however, z decreases again to zero so this root is no longer valid. Consequently, from (23) we have

$$\begin{aligned} F(p)&\equiv p \quad \text{for } p\leq 1/\sigma,\\ &\neq p \quad \text{for } p>1/\sigma. \end{aligned} \tag{24}$$

As explained in the previous section, this establishes that the critical probability is

$$p_c=1/\sigma. \tag{25}$$

The correctness of this result is easily verified by regarding the buildup of a cluster on the Bethe lattice as a branching or cascade process. If the cluster is to spread indefinitely, the expected number of occupied bonds leaving from one end of a given occupied bond must not be less than unity. Conversely, if the expected number exceeds unity an infinite cluster will be formed. Since the probability of traversing a bond is p and σ independent bonds proceed onwards from a given bond, the critical condition is $p_c\sigma=1$ in agreement with (25).

For the complete Bethe generating function, one derives from (20)–(22) the equation

$$dB\sigma/dZ=[1-X(Z)]^{-2\sigma}, \tag{26}$$

where $X(Z)=X(x,y)$ is the root of

$$X(1-X)^{\sigma-1}=Z=xy^{\sigma-1}, \tag{27}$$

which vanishes with Z. [Note that $X(p,q)=p^*(p)$.] This may be integrated to yield

$$B_\sigma(Z)=\frac{1}{\sigma+1}\frac{[2-(\sigma+1)X(Z)]}{[1-X(Z)]^{\sigma+1}}, \tag{28}$$

and with (16) this formally solves the problem.

From (23), the density of infinite clusters is

$$\begin{aligned} R(p)&=p\{1-[(1-p)/(1-p^*)]^{2\sigma}\}\\ &=p\{1-(p^*/p)^{[2\sigma/(\sigma-1)]}\}, \end{aligned} \tag{29}$$

while by differentiating (26) one finds for the mean cluster size

$$S(p)=(1+\sigma p^*)/(1-\sigma p^*). \tag{30}$$

As already noted, Eq. (21) for $p^*(p)$ shows that $p^*=p$ for $p\leq 1/\sigma$. At $p=p_c$ the gradient of $p^*(p)$ changes sign discontinuously, but the magnitude remains unchanged. Thereafter, $p^*(p)$ decreases monotonically and vanishes at $p=1$ as $(1-p)^{\sigma-1}$. Near p_c one has

$$p^*(p)\approx p_c-|p-p_c| \quad (p\simeq p_c), \tag{31}$$

so that the mean cluster size becomes hyperbolically infinite as

$$S(p)\approx 2/|1-(p/p_c)| \quad (p\simeq p_c), \tag{32}$$

and the density of infinite clusters vanishes linearly as

$$R(p)\approx[4\sigma/(\sigma-1)](p-p_c) \quad (p\to p_c+). \tag{33}$$

After removal of the root $p^*=p$, Eq. (21) is of degree $\sigma-1$ and so for the first few values of σ (i.e., small coordination numbers) it may be solved in closed form. One finds

(a) $\sigma=1$ (linear chain) $p_c=1$.

$$p^*(p)=p, \quad R(p)=0 \quad \text{(all } p\text{)}, \tag{34a}$$

$$S(p)=(1+p)/(1-p), \tag{34b}$$

$$B_1(z)=1/(1-z). \tag{34c}$$

(b) $\sigma=2$, $p_c=\frac{1}{2}$,

$$p^*(p)=1-p \quad (p>\tfrac{1}{2}), \tag{35a}$$

$$R(p)=p-p^{-3}(1-p)^4 \quad (p>\tfrac{1}{2}), \tag{35b}$$

$$S(p)=1/|p-\tfrac{1}{2}|-1, \tag{35c}$$

$$B_2(z)=\tfrac{4}{3}\{[1+3(1-4z)^{\frac{1}{2}}]/[1+(1-4z)^{\frac{1}{2}}]^3\}. \tag{35d}$$

(c) $\sigma=3$, $p_c=\frac{1}{3}$,

$$p^*(p)=1-\tfrac{1}{2}p-[p(1-\tfrac{3}{4}p)]^{\frac{1}{2}} \quad (p>\tfrac{1}{3}), \tag{36b}$$

$$R(p)=p-p^{-2}[1-\tfrac{1}{2}p-p^{\frac{1}{2}}(1-\tfrac{3}{4}p)^{\frac{1}{2}}]^3 \quad (p>\tfrac{1}{3}). \tag{36c}$$

The typical behavior of the density of infinite clusters, etc., can be seen from Figs. 2 and 3 which refer to the Bethe lattice of coordination number four ($\sigma=3$).

B. Site Problem

To solve the site problem for the Bethe lattices, it is sufficient to note that the configurations involved are essentially identical with those required for the bond problem. In fact, a one-one correspondence can be set up which is described by

$$k^S{}_{s,t}=\tfrac{1}{2}(\sigma+1)k^B{}_{s-1,t}. \tag{37}$$

It follows immediately that the configurational generating function for the site problem is

$$K^S(x,y)=\tfrac{1}{2}(\sigma+1)xy^{\sigma+1}B_\sigma[Z(x,y)]. \tag{38}$$

The Bethe function $B_\sigma(Z)$ is defined as before by (28), (27), and (15) and, clearly, it determines the critical point and the nature of the critical singularities. These

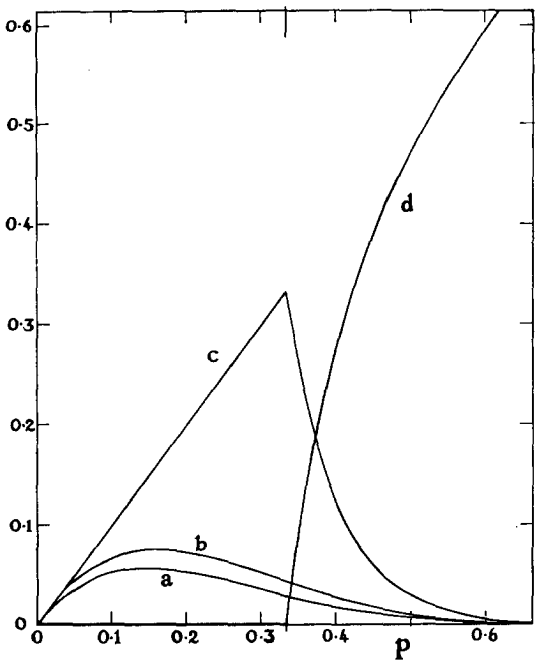


FIG. 2. Cluster distributions for the $\sigma=3$ Bethe lattice: (a) probability of a bond belonging to a cluster of one bond, (b) to a cluster of one or two bonds, (c) $F(p)$ the density of finite clusters, (d) $R(p)$ the density of infinite clusters. (Note the incomplete range of p.)

must, therefore, resemble those for the bond problem. Explicitly, the density of finite clusters is given by

$$F(p)=p^*(1-p)^2/(1-p^*)^2 \tag{39}$$

in place of (23), and the mean size by

$$S(p)=(1+p^*)/(1-\sigma p^*), \tag{40}$$

which exhibits directly the unchanged position of the critical point.

IV. CONFIGURATIONAL COEFFICIENTS

By the previous analysis, the coefficients in the expansions of the various generating functions for the Bethe lattices may all be expressed in terms of the coefficients $b_s(\sigma)$ defined in Eqs. (14) and (15). An explicit expression for these is most readily obtained from (26) which may be written

$$\sum_{s=1}^{\infty} sb_s(\sigma)z^{s-1}=[1-X(z)]^{-2\sigma}, \qquad X(1-X)^{\sigma-1}=z. \tag{41}$$

By Cauchy's theorem, one then has

$$sb_s(\sigma)=\frac{1}{2\pi i}\oint \frac{z^{-s}dz}{[1-X(z)]^{2\sigma}}, \tag{42}$$

where the contour of integration is a small closed loop encircling the origin. Near the origin, the analytic behavior of X is the same as z so that (42) may be transformed to an integral in the X plane, namely,

$$sb_s(\sigma)=\frac{1}{2\pi i}\oint \frac{(1-\sigma X)dX}{X^s(1-X)^{(s+1)(\sigma-1)+3}}. \tag{43}$$

This can be evaluated directly with the aid of the binomial theorem yielding

$$sb_s(\sigma)=\frac{2\sigma[(s+1)\sigma-1]!}{(s-1)![(s+1)\sigma-s+1]!}, \tag{44}$$

which expresses the total number of distinct Cayley trees of s branches passing through a specified bond of a Bethe lattice of coordination number $\sigma+1$.[15] From (44), one obtains for the number of trees *per bond*

$$b_s(\sigma)=\frac{2}{(s+1)(\sigma s+\sigma+1)}\binom{\sigma s+\sigma+1}{s}, \tag{45}$$

which remains valid for $s=0$.

V. BOND DECORATION

We consider now a class of lattices which can be derived from a Bethe lattice by replacing each bond by a replica of a given finite graph of sites and bonds, the *bond graph*. The simplest example of such a decorated

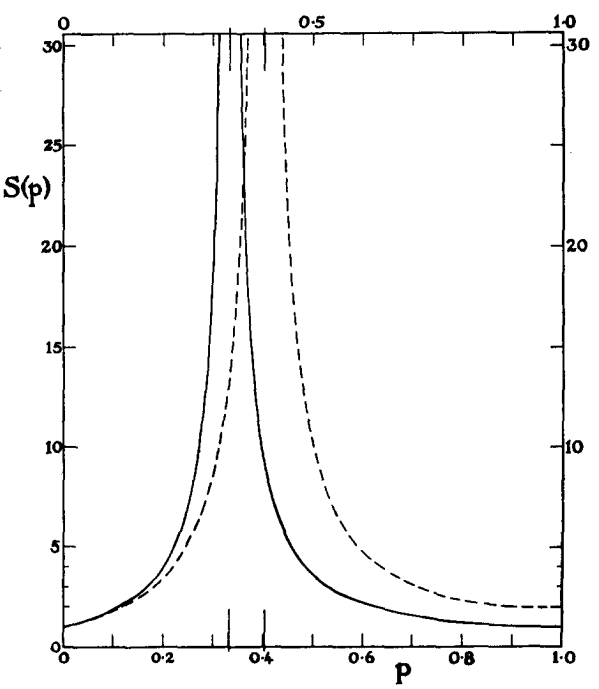


FIG. 3. Variation of $S(p)$ the mean cluster size density (for finite bond clusters): (a) for the $\sigma=3$ Bethe lattice (solid curve), (b) for the triangular cactus of same coordination number (broken curve).

[15] This expression for the number of Cayley trees on a Bethe lattice was originally conjectured by Dr. M. F. Sykes (private communication).

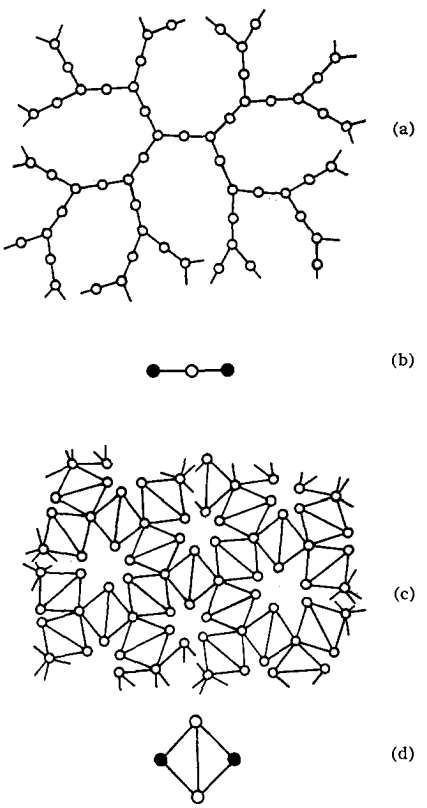


FIG. 4. Decorated Bethe lattices (a) and (c) derived from the $\sigma=2$ Bethe lattice by replacing bonds by the bond graphs (b) and (d).

lattice is that shown in Fig. 4(a) which represents a Bethe lattice of coordination number 3 with an extra site on each bond. The bond graph in this case is simply three sites connected by two bonds [Fig. 4(b)]. A more complicated example derived from the same Bethe lattice is shown in Fig. 4(c). Its bond graph is a square with one diagonal [Fig. 4(d)]. To simplify the general treatment, we will consider only bond graphs which are symmetric with regard to the two terminals or points of attachment [indicated in Figs. 4(b) and 4(d) by solid circles] and that the probability of occupation is the same for all bonds (or sites). Both these restrictions are quite easy to remove.

To construct the configurational generating function for a decorated lattice (from which all its properties may be deduced), we set up a one-many correspondence between configurations on the original (undecorated) Bethe lattice and those on the decorated lattice. Consider a specified bond graph in the decorated lattice, some bonds of which are connected in a cluster of occupied bonds (or sites). If it is possible to cross from one terminal of the bond graph to the other by a connected sequence of occupied bonds (or sites), then the bond graph is to be identified with the corresponding *occupied* bond on the original Bethe lattice. If, on the other hand, it is not possible to cross the bond graph (or it only contains unoccupied perimeter bonds or sites), then it should be identified with an *unoccupied* perimeter bond on the original lattice. On this basis, we define three bond generating functions, namely (for the bond problem),

$$c(x,y)=\sum_{s,t} c_{s,t}x^s y^t, \tag{46}$$

where $c_{s,t}$ is the number of distinct connected configurations of s occupied bonds and t perimeter bonds on the bond graph which join one terminal to the other;

$$d(x,y)=\sum_{s,t} d_{s,t}x^s y^t, \tag{47}$$

where $d_{s,t}$ is the number of distinct connected configurations of s occupied bonds and t perimeter bonds on the bond graph which are connected to one (specified) terminal but not to the other, including the case in which one terminal is joined only to unoccupied perimeter bonds;

$$e(x,y)=\sum_{s,t} e_{s,t}x^s y^t, \tag{48}$$

where $e_{s,t}$ is the number of distinct connected configurations of s occupied bonds and t perimeter bonds on the bond graphs which are not connected to *either* terminal.

By way of example, the bond generating functions for the two decorated lattices of Fig. 4 are

$$c(x,y)=x^2, \quad d(x,y)=xy+y, \quad e(x,y)=0, \tag{49}$$

and

$$\begin{aligned} c(x,y)&=x^5+5x^4y+8x^3y^2+2x^2y^3,\\ d(x,y)&=x^3y^2+3x^2y^3+2xy^3+y^2,\\ e(x,y)&=xy^4, \end{aligned} \tag{50}$$

respectively.

From the definition of $c(x,y)$, it follows that $c(p,q)$ is the probability of reaching one terminal of the bond graph starting from the other terminal. Conversely, $d(p,q)$ is the probability of failing to reach the other terminal from one terminal. Consequently, the identity

$$c(p,q)+d(p,q)\equiv 1 \tag{51}$$

is always valid and $c(p,q)$ is a monotonically increasing function of p.

The configurational generating function for the decorated lattice (indicated by a star) can now be derived by making the transformation $x\rightarrow x^*=c(x,y)$, $y\rightarrow y^*=d(x,y)$, and adding a correction for the clusters which do not span a bond graph. Thus, per bond of the decorated lattice,

$$K^{*B}(x,y)=g_B^{-1}K^B[c(x,y),d(x,y)]+g_B^{-1}e(x,y), \tag{52}$$

where g_B is the number of bonds in the bond graph. By (16), we have the explicit result

$$K^{*B}(x,y)=g_B^{-1}[d(x,y)]^{\sigma+1}B_\sigma[Z^*(x,y)]+g_B^{-1}e(x,y), \tag{53}$$

where

$$Z^*(x,y)=c(x,y)[d(x,y)]^{\sigma-1}. \tag{54}$$

These relations solve the bond problem for any bond-decorated Bethe lattice.

By (53), the critical singularities are determined entirely by the Bethe function $B_\sigma(Z^*)$. It is not immediately apparent, however, that the behavior in the critical region need resemble that for the original lattice. Nevertheless, it follows (51) that $Z^*(p,q)$ has a simple maximum as a function of p of magnitude $\sigma^{-\sigma}(\sigma-1)^{\sigma-1}$ which is exactly the value at which $B_\sigma(Z)$ becomes nonanalytic. Consequently, the critical point of a decorated lattice is determined by

$$c(p_c,q_c)=1/\sigma, \tag{55}$$

and the critical singularities of $R(p)$, $S(p)$, and other variables do have the same functional forms as those for the simple Bethe lattices [compare with Eqs. (32) and (33)]. In view of the probabilistic meaning of $c(p,q)$, the critical equation (55) can also be derived directly by viewing the formation of an infinite cluster as a branching process as in Sec. II.

The site problem on a bond-decorated lattice may be solved in a similar fashion by modifying the definitions of the bond generating functions $c(x,y)$, $d(x,y)$, and $e(x,y)$ so as to refer to sites of the bond graph in place of bonds. In the definition of $c(x,y)$ and $d(x,y)$, it must also be assumed that the first terminal site is already occupied but no factor x should be included for it. As an example, the bond functions for the site problem on the two lattices of Fig. 4 are

$$c(x,y)=x^2, \quad d(x,y)=xy+y, \quad e(x,y)=0, \tag{56}$$

and

$$\begin{aligned} c(x,y)&=x^3+2x^2y,\\ d(x,y)&=x^2y+2xy^2+y^2,\\ e(x,y)&=x^2y^2+2xy^3, \end{aligned} \tag{57}$$

respectively. Equation (51) remains valid in all cases.

The configurational generating function per site of the site-decorated lattice is then

$$K^{*S}(x,y)=\frac{\sigma+1}{(\sigma+1)g_S-2\sigma} \times\{x[d(x,y)]^{\sigma+1}B_\sigma[Z^*(x,y)]+e(x,y)\}, \tag{58}$$

where $Z^*(x,y)$ is defined by (54), and g_S is the number of sites of the bond graph (including the terminal sites). The nature of the critical singularities is unchanged and the critical probability is still determined by (55).

VI. SITE DECORATION

In this section, we endeavor to find generating functions for lattices derived from Bethe lattices by decorating the sites as well as the bonds. Each site of the underlying Bethe lattice will be replaced by a replica of a given *site graph* which we take to be symmetric under interchange of its $\sigma+1$ distinct terminals (points of attachment). Initially, we consider only the first nontrivial case in which the undecorated lattice is a Bethe lattice of coordination number three (i.e., $\sigma=2$). The simplest example of such a lattice is the infinite cactus illustrated in Fig. 1(b) in which the site graph is a triangle and the bond graph is merely a single site. The "expanded cactus" of Fig. 1(c) is obtained when the bond graph is left as a single bond. More generally, the triangles may be replaced by any finite symmetric three-terminal graph.

As before, we aim to solve the problem by setting up a many-one correspondence between configurations on the decorated lattice and those on the original Bethe lattice. In analogy with the three bond generating functions defined in (46)–(48), we thus introduce four site generating functions which, for the bond problem, are

$$t(x,y)=\textstyle\sum_{s,t} T_{st}x^sy^t, \tag{59}$$

where T_{st} is the number of distinct connected configurations of s occupied bonds and t unoccupied perimeter bonds on the site graph in which all three terminals are connected together;

$$u(x,y)=\textstyle\sum_{s,t} U_{st}x^sy^t, \tag{60}$$

where U_{st} is the number of configurations on the site graph in which the first terminal is connected to the second terminal but *not* to the third;

$$v(x,y)=\textstyle\sum_{s,t} V_{st}x^sy^t, \tag{61}$$

where V_{st} is the number of configurations in which the first terminal is connected to neither of the other two terminals (including the configuration of unoccupied perimeter bonds attached to the first terminal); and, finally,

$$w(x,y)=\textstyle\sum_{s,t} W_{st}x^sy^t, \tag{62}$$

where W_{st} is the number of configurations connected to none of the three terminals. For the triangular site graphs of Figs. 1(b) and 1(c), one has simply

$$t=x^3+3x^2y, \quad u=xy^2, \quad v=y^2, \quad w=0, \tag{63}$$

while for the relevant bond graphs

$$c=1, \quad d=0, \quad e=0, \tag{64a}$$

and

$$c=x, \quad d=y, \quad e=0, \tag{64b}$$

respectively.

By these definitions, $t(p,q)$ is the probability of being able to reach the second and third terminal from the first terminal, $u(p,q)$ is the probability that the second terminal only is accessible, and $v(p,q)$ is the probability that the site graph cannot be crossed at all. Consequently, in analogy with (51), the identity

$$t(p,q)+2u(p,q)+v(p,q)\equiv 1 \tag{65}$$

always holds.

In terms of these site generating functions, we may set up a correspondence with configurations on the undecorated Bethe lattice by identifying the decorated

configurations included in $t(x,y)$ with a Bethe lattice at which three occupied bonds meet (a *triple point*), those included in $u(x,y)$ with a site at which only two occupied bonds meet (*double point*), and those in $v(x,y)$ with a site attached to only one occupied bond (*single point*). The problem may then be solved if we can determine the multivariable generating function

$$H(t,u,v',c)=\sum_{\lambda\mu\nu\gamma} h_{\lambda\mu\nu\gamma}t^{\lambda}u^{\mu}v'^{\nu}c^{\gamma}, \tag{66}$$

where t, u, v', and c are regarded simply as generating symbols and where $h_{\lambda\mu\nu\gamma}$ is the number of distinct cluster configurations per bond of a Bethe lattice of coordination number three with λ triple points, μ double points, ν single points, and a total of $\gamma+\nu$ occupied bonds (with γ nonnegative). In terms of this function, the configurational generating function per bond of the decorated lattice is given by

$$lK^{*B}(x,y)=H\{t(x,y),\, u(x,y),\, c(x,y)v(x,y)+d(x,y),\, c(x,y)\}$$
$$+c(x,y)[v(x,y)]^2+d(x,y)v(x,y)+e(x,y)+\tfrac{2}{3}w(x,y), \tag{67}$$

where l is the number of bonds of the decorated lattice per bond of the original lattice. If g_B is the number of bonds in the bond graph and j_B the number in the site graph, one has

$$l=g_B+\tfrac{2}{3}j_B. \tag{68}$$

The replacement of v' in (66) by $cv+d$ in (67) allows for the two possibilities: (a) the cluster on the decorated lattice extends across the bond graph leading to the single site graph and has perimeter bonds in the site graph, (b) the cluster on the decorated lattice does not reach across the bond graph so that its (local) perimeter lies entirely in the bond graph. The four end terms in (67) enumerate the small clusters which extend no further than across one bond graph.

To enumerate the cluster configurations $h_{\lambda\mu\nu\gamma}$ (on the $\sigma=2$ Bethe lattice), consider first those configurations with no triple points. These are just linear chains of bonds and double points, and their contribution to the generating function is simply ($\gamma\geqslant 0$)

$$v'\cdot 2uv'+v'\cdot 2uc\cdot 2uv'+v'(2uc)^2 2uv'+\cdots$$
$$=2uv'^2/(1-2uc). \tag{69}$$

The configurations containing triple points may be enumerated by setting up a many-one correspondence with the standard cluster configurations on the lattice. These latter are described by the configurational generating function

$$K^B(x,y)=y^3B_2(xy)=\tfrac{2}{3}y^3+xy^4+2x^2y^6+\cdots, \tag{70}$$

where the successive terms correspond to a single site with three unoccupied perimeter bonds, two sites connected by an occupied bond and four unoccupied perimeter bonds, three sites connected by two bonds, etc. *Each site* in one of these configurations may now be identified with a triple point of the $h_{\lambda\mu\nu\gamma}$ configurations. Each occupied bond x must then be identified with a linear chain of bonds and double points leading from one triple point to another, while each perimeter bond y corresponds to a linear chain leading from a triple point and terminating in a single point. The generating functions for these chains are

$$c+c\cdot 2uc+c(2uc)^2+\cdots=c/(1-2uc) \tag{71}$$

and

$$v'+2uc\cdot v'+(2uc)^2v'+\cdots=v'/(1-2uc), \tag{72}$$

respectively. By combining (69), (71), and (72) and remembering that each configuration has one more site than it has occupied bonds, we obtain

$$H(t,u,v',c)=tK^B[tc/(1-2uc),\, v'/(1-2uc)]$$
$$+2uv'^2/(1-2uc). \tag{73}$$

Consequently, by (67) and (70), the configurational generating function per bond for a site and bond decorated lattice is finally given in explicit terms by

$$lK^{*B}(x,y)=\frac{t(cv+d)^3}{(1-2uc)^3}B_2\left[\frac{tc(cv+d)}{(1-2uc)^2}\right]+\frac{2u(cv+d)^2}{1-2uc}$$
$$+v(cv+d)+e+w, \tag{74}$$

where c, d, e, t, u, v, and w are defined as functions of x and y by (46)–(48) and (59)–(62).

From this result, it follows as before that the critical point and singularities are determined only by the Bethe function. In view of the identities (51) and (65) which hold when $x=p$ and $y=q$, the argument of B_2 in (74) can be written

$$Z^*(p,q)=\left[\frac{tc}{1-2uc}\left(1-\frac{tc}{1-2uc}\right)\right]_{x=p,y=q},$$

and so exhibits a maximum as a function of p of magnitude $Z_m=\frac{1}{4}=z_c$. Consequently, the critical equation is

$$2u(p,q)c(p,q)+2t(p,q)c(p,q)=1, \tag{75}$$

and the analytic forms of the mean cluster size and other properties in the critical region are the same as for the original Bethe lattice.

The critical equation (75) can be derived directly from the branching process viewpoint by calculating the expected number of paths reaching from one terminal of a site graph through the following bonds to the nearer terminals of the next site graphs. By the probabilistic interpretation of c, u, and t, this is just the left-hand side of (75). (The first term comes from the two cases in which only one through route is open, while the second term represents the cases in which both further terminals are accessible.)

For the normal and expanded cacti of Fig. 1, one obtains with (63) and (64) the critical equations and critical points:

Triangular cactus (bond problem)

$$1-2p-2p^2+2p^3=0, \quad p_c=0.403032, \tag{76}$$

Expanded cactus (bond problem)

$$1-2p^2-2p^3+2p^4=0, \quad p_c=0.637278. \tag{77}$$

These critical probabilities should be compared with the results for the Bethe lattices of the same coordination number, namely, $(\sigma+1=4)$ $p_c=0.333333$ and $(\sigma+1=3)$ $p_c=0.500000$, respectively. As might be expected, the critical probability increases with the increasing connectivity of the lattice. (A similar phenomenon occurs in order-disorder lattice statistics where the increasing connectivity lowers the critical temperature below the Bethe approximation value.) We mention in passing the critical equation

$$1-2p-2p^2-2p^3+3p^4=0, \tag{78}$$

which may be derived for the simplest square Husimi tree[12] which also has coordination number $\sigma+1=4$. The corresponding critical point is

Square Husimi tree (bond problem)

$$p_c=0.353933, \tag{79}$$

which lies between the Bethe lattice and cactus results because of the low connectivity of a square compared with a triangle.

Explicit expressions for the mean cluster size, etc., may be derived by differentiating (74). As an example, we quote the results for the simple cactus (bond problem):

$$R(p)=p-\frac{2(1-p)^6(1+p-p^2)}{p^4(3-2p)^3} \quad (p\geqslant p_c) \tag{80}$$

$$S(p)=\frac{1+4p+2p^2-2p^3+p^5}{1-2p-2p^2+2p^3} \quad (p\leqslant p_c)$$

$$=\frac{4+10q+11q^2+14q^3-3q^4-20q^5+8q^6}{2(1+q-q^2)(1+2q)(1-4q^2+2q^3)} \quad (p\geqslant p_c). \tag{81}$$

The behavior of the cluster size is compared with that for the Bethe lattice of coordination in Fig. 3. As p tends to unity $S(p)$, the mean size of finite clusters on the cactus, approaches the value 2, whereas on the Bethe lattice $S(p)$ approaches 1. The reason for this difference arises from the existence on the cactus of two distinct clusters of minimum perimeter: the single bond and the triangle of three bonds. (Only the former, of course, occurs on the Bethe lattice.) In the limit $p\to 1$, these clusters dominate and appear equally frequently.

VII. BOND-TO-SITE TRANSFORMATION

The site problem on a Bethe lattice with decorated sites and bonds may be solved explicitly for the case $\sigma=2$ by modifying the definitions of the site generating functions (59)–(62) along the lines used to redefine the bond generating functions (see last two paragraphs of Sec. V). The argument then proceeds in parallel with that for the bond problem although the details differ slightly. The results will not be presented here, but it is worth noting that the site problem on the cacti of Figs. 1(b) and 1(c) can be solved immediately without further theoretical development.

To see this, consider a configuration of sites on the simple cactus of Fig. 1(b). Any such configuration can be put in direct one-one correspondence with a configuration of *bonds* on the underlying simple Bethe lattice of coordination number three. Each site on the cactus corresponds to the underlying bond on the Bethe lattice. Neighboring sites correspond to neighboring bonds. Consequently, the configurational generating function

$$xy^4+2x^2y^5+(14/3)x^3y^6+\cdots$$

per site or per bond, respectively, is the same for both problems. The critical probabilities are, therefore, identical as are all other properties.

Clearly this *bond-to-site* transformation can be applied to any suitably related pair of lattices. Thus, the expanded cactus of Fig. 1(c) can be derived from the decorated $(\sigma=2)$ Bethe lattice shown in Fig. 4(a) by identifying bonds on the decorated lattice with sites of the expanded cactus. Hence, the site problem on the expanded cactus is identical with the bond problem on the decorated Bethe lattice. By Eqs. (49) and (55), the critical probability for the two problems is $p_c=2^{-\frac{1}{2}}$.

In the same way, the bond problem on the plane honeycomb lattice, coordination number three, is isomorphic with the site problem on the plane kagomé lattice, coordination number four (reference 11, p. 187). At present, however, neither of these problems is soluble in closed form.

VIII. FURTHER GENERALIZATION

The multivariable generating function $H(t,u,v',c)$ defined in (66) enumerates on the $\sigma=2$ Bethe lattice configurations with a specified number of triple points, double points, single points, and bonds. Now, since the coordination number is 3, each triple point is associated with *no* unoccupied perimeter bonds, each double point is associated with *one* unoccupied bond, and each single point with *two* unoccupied bonds (and *one* occupied bond). Accordingly, if the substitutions $t=1$, $u=y$, $v'=xy^2$, and $c=x$ are made, the generating function $H(t,u,v',c)$ will merely classify cluster configurations by the total numbers of the internal bonds and of the perimeter bonds. But this is just what the configurational generating function for the simple problem does. In

other words,

$$K^B(x,y)=H(1,y,xy^2,x)+\tfrac{2}{3}y^3+xy^4, \tag{82}$$

where the last two terms account for the single site and single bond which are not included in H. If this expression is combined with (73), one obtains

$$H(t,u,v',c)=tH\left(1,\frac{v'}{1-2uc},\frac{tcv'^2}{(1-2uc)^3},\frac{tc}{1-2uc}\right)+\frac{2}{3}\frac{tv'^3}{(1-2uc)^3}+\frac{t^2cv'^4}{(1-2uc)^5}+\frac{2uv'^2}{(1-2uc)} \tag{83}$$

and

$$K^B(x,y)=K^B\left(\frac{x}{1-2xy},\frac{xy^2}{1-2xy}\right)+\frac{2xy^4}{1-2xy}+\tfrac{2}{3}y^3. \tag{84}$$

The first relation re-expresses $H(t,u,v',c)$ in a reduced form independent of its first argument. If we set $t=1$, (83) reduces to a functional equation for $H(1,u,v',c)$ which is equivalent to the functional relation (84) for $K^B(x,y)$. By Eq. (20), this may also be transformed into a functional equation for $B_2(z)$, namely,

$$B_2(z)=\frac{z^3}{(1-2z)^3}B_2\left[\frac{z^2}{(1-2z)^2}\right]+\frac{z}{1-2z}+\frac{2}{3}. \tag{85}$$

This equation does not seem easy to solve directly in closed form although it defines $B_2(z)$ uniquely for small z as may be seen by assuming the power series expansion for $B_2(z)$ and determining successive coefficients by comparing like powers of z. It is easily verified, however, that the correct solution is provided by (35d).

When one attempts to generalize the foregoing approach to the site decoration of a Bethe lattice of coordination number 4, one is lead to introduce a generating function which enumerates configurations by number of quadruple points as well as by triple points, double points, etc. The terms in this generating function may be set in correspondence with configurations on the $\sigma=2$ and $\sigma=3$ Bethe lattices. The principle of the argument is the same as in Sec. VI, but is more involved and will not be presented. If s and t are the enumerating symbols for quadruple points and triple points, respectively, the result may be expressed most compactly by introducing the generating function $J_3(s,t;z)$ which reduces to $B_3(z)$ when $s=1$ and $t=1$. For this function we obtain in analogy to (83) the relation

$$J_3(s,t;z)=sz_1^4J_3(1,t^*;z^*)+z_0^3B_2(3tz_0^2)+z_0+\tfrac{1}{2}, \tag{86}$$

where

$$\begin{aligned} z_0&=z/(1-3z),\\ z_1&=[1-(1-12tz_0^2)^{\frac{1}{2}}]/6tz_0,\\ z^*&=sz_0z_1^2/(1-6tz_0z_1),\\ t^*&=(t+sz_1)/sz_1. \end{aligned} \tag{87}$$

When one sets $s=1$, the relation (86) reduces to a functional equation in *two* variables for $J_3(1,t;z)$. This equation may be solved in a double power series, but owing to its complexity, we have not been able to solve it in closed form. Consequently, it is not possible, for example, to give explicit expressions for the cluster size on such pseudolattices as the square Husimi tree.

Although we cannot give explicit formulas for all lattices derivable from the Bethe lattices by site and bond decorations, there seems no reason to doubt that the nature of the critical singularities will be the same in all cases. This seems to be connected with the infinite-dimensional and multiplicative properties of the Bethe lattices and their derivatives. By analogy with the behavior of other statistical lattice problems, such as the Ising model, one would expect the critical singularities to be sharper for the normal lattices in two and three dimensions than for the Bethe lattices. Thus, at $p=p_c$ the gradient dR/dp might well be infinite and the mean cluster size density might diverge as $|p-p_c|^{-\alpha}$ with α greater than unity. The sharpness would be expected to fall off with increasing dimension and approach the present results in the limit. Rigorous confirmation of these conjectures must await a comprehensive attack on the problem for the standard lattices, but some indication of their validity can be obtained from a numerical study of the initial terms of the configurational series.[13]

ACKNOWLEDGMENT

We are grateful to Professor C. Domb for having drawn our attention to the subject and for his comments on the manuscript.

Critical Probabilities for Cluster Size and Percolation Problems

MICHAEL E. FISHER
Wheatstone Physics Laboratory, King's College, London, England
(Received February 17, 1961)

When particles occupy the sites or bonds of a lattice at random with probability p, there is a critical probability p_c above which an infinite connected cluster of particles forms. Rigorous bounds and inequalities are obtained for p_c on a variety of lattices and compared briefly with previous numerical estimates. In particular, by extending Harris' work, it is proved that $p_c(s,L_2) \geqslant \frac{1}{2}$ for the site problem on a plane lattice L_2 (without crossing bonds), while for the bond problem $p_c(b,L_2)+p_c(b,L_2^D) \geqslant 1$ where L_2^D is the dual lattice to L_2. Simple arguments demonstrate that the bond problem is a special case of the site problem and that the critical probabilities for the bond problem on the plane square and triangular lattice cannot exceed those for the corresponding site problems.

1. INTRODUCTION

SEVERAL authors[1-7] have recently drawn attention to the importance of *cluster size* and *percolation* problems in various branches of physics, physical chemistry, and statistics. These problems may be stated in mathematical terms as follows: *particles* are distributed at random over an infinite linear graph L composed of *sites* (vertices) linked together by *bonds* (lines). In the case of principal physical interest, the linear graph L forms an infinite regular space lattice in two or three dimensions. Two possibilities arise naturally: (a) the *site problem* in which the particles occupy only the sites of L; (b) the *bond problem* in which the particles occupy only the bonds. Each site (or bond) is supposed to be occupied independently of all others with a fixed probability p. The site problem describes, for example, a random binary alloy or dilute ferromagnetic crystal,[4-7] while the bond problem corresponds to a randomly blocked maze through which the percolation of a fluid may be envisaged. A group of n particles on the lattice linked to one another through adjoining bonds and sites is said to form a (connected) cluster of size n. The task of the mathematical theory is to calculate the statistical properties of these clusters on a given lattice L as a function of the probability p which measures the particle density.

An important feature of the problem is the existence of a *critical probability* p_c. When p is less than p_c all clusters are finite in size, but when p exceeds p_c a cluster of infinite size spreads through the lattice. In the latter case there is a nonzero probability $R(p)$ that a given site (or bond) is occupied by a member of an infinite connected cluster of particles on the sites (or bonds). As yet, critical probabilities have been calculated exactly only for certain pseudolattices[7] (infinite Cayley trees and their generalizations). Numerical estimates of p_0 for the bond problem on the plane square and simple cubic lattices have, however, been obtained by Hammersley[8] who used a Monte Carlo method, and Domb[4] and others[5,6] have shown how estimates may also be obtained from series expansions for the mean cluster size. The validity of such estimates may be gauged by comparing them with rigorous upper and lower bounds. General methods for obtaining such bounds (for the bond probable) have been developed by Hammersley,[3] but unfortunately these do not yield close limits. The strongest result yet obtained is due to Harris[9] who proved that for the *bond problem on the plane square lattice* $p_c \geqslant \frac{1}{2}$. The numerical evidence[3,5,6] suggests quite strongly that $p_c = \frac{1}{2}$ so that Harris' result is probably the best possible lower bound.

In this article we consider the critical probabilities of a range of lattice and obtain bounds and inequalities connecting different lattices and different problems on the same lattice. In particular, we show how the general approach used by Harris can be extended rigorously to cover the site problem on a plane lattice and the bond problem on a pair of dual plane lattices. Our main conclusions for plane lattices are summarized in the following relations in which s and b denote the site and bond problem, respectively, H, S and T denote the plane honeycomb, square, and triangular lattices, respectively, and L_2 and L_2^D denote a regular two-dimensional lattice with no crossing bonds and its dual lattice.

Site and bond problems

$$p_c(s,H) \geqslant p_c(s,S) \geqslant p_c(s,T), \tag{1.1}$$

$$p_c(b,H) \geqslant p_c(b,S) \geqslant p_c(b,T). \tag{1.2}$$

$$p_c(s,S) \geqslant p_c(b,S), \tag{1.3}$$

$$p_c(s,T) \geqslant p_c(b,T). \tag{1.4}$$

[1] S. R. Broadbent and J. M. Hammersley, Proc. Cambridge Phil. Soc. **53**, 629 (1957).
[2] J. M. Hammersley, Proc. Cambridge Phil. Soc. **53**, 642 (1957).
[3] J. M. Hammersley, (a) Ann. Math. Stat. **28**, 790 (1957); (b) Proc. 87th Intern. Colloq. "Probabilités et ses Applications," Paris (1959), p. 17.
[4] C. Domb, "Fluctuation Phenomena and Stochastic Processes," Conference at Birkbeck College, London, March, 1959, reported in Nature **184**, 509 (1959).
[5] R. J. Elliott, B. R. Heap, D. J. Morgan, and G. S. Rushbrooke, Phys. Rev. Letters **5**, 366 (1960).
[6] C. Domb and M. F. Sykes, Phys. Rev. (to be published).
[7] M. E. Fisher and J. W. Essam, J. Math. Phys. **2**, 609 (1961), this issue.
[8] J. M. Hammersley, lecture to the British Association, Dublin, September 5, 1957 (unpublished).
[9] T. E. Harris, Proc. Cambridge Phil. Soc. **56**, 13 (1960).

Site problems

$$p_c(s,H)>0.52. \tag{1.5}$$

$$p_c(s,L_2)\geqslant\tfrac{1}{2}. \tag{1.6}$$

Bond problems

$$p_c(b,L_2)+p_c(b,L_2^D)\geqslant 1. \tag{1.7}$$

Since the square lattice S is self-dual, this last result includes Harris' theorem $p_c(b,S)\geqslant\frac{1}{2}$ which, together with (1.4), implies $p_c(s,S)\geqslant\frac{1}{2}$ independently of (1.6). The relation (1.7) also shows, for example, that the critical probability $p_c(b,H)$ for the honeycomb cannot be less than unity minus the corresponding probability $p_c(b,T)$ on the triangular lattice.

In the proof of the two results (1.6) and (1.7) [see Appendixes], it is assumed that (a) the lattice has only finitely many sites and bonds per unit cell, and (b) the lattice possesses two perpendicular symmetry axes about which it may be reflected into itself or, more generally, that it may be topologically distorted into such a symmetric form. The assumption (a) suffices to exclude certain pathological lattices for which the theorems fail owing to the growth of infinite cluster *within* a unit cell. From the physical viewpoint, (a) is no restriction. The assumption (b) is probably not necessary for the general validity of (1.6) and (1.7), but it facilitates proof and is not, in any case, a serious restriction for physical applications since it excludes none of the standard plane lattices. [Many asymmetric lattices apparently excluded can be included by converting them into symmetric lattices through the addition of sufficient extra bonds and sites and then using the *containment theorem* (2.4) of Sec. 2. Indeed (1.6) and (1.7) are probably valid for an arbitrary planar graph provided it is "finite in the small" and reasonably "homogeneous in the large," although these conditions are not easy to formulate precisely.]

Inequalities similar to (1.1) and (1.2) also hold between the three-dimensional tetrahedral, simple cubic, body-centered cubic and face-centered cubic lattices, respectively.

Estimates of the critical probabilities for the site problem on the plane lattices have been made by Elliott, Heap, Morgan, and Rushbrooke[5] who obtained from the first few terms of a series

$$p_c(s,T)\simeq 0.36,\quad p_c(s,S)\simeq 0.48,\quad p_c(s,H)\simeq 0.49.$$

These estimates are inconsistent with the strict bounds (1.5) and (1.6). In fact, the errors must be at least 39, 4, and 6%, respectively. For the square and honeycomb lattices, the errors are within the 10% confidence limit quoted,[5] although the true figure for the honeycomb seems to lie even outside this range.[4,6] The estimates of Domb and Sykes, based on longer series and a more sensitive extrapolation procedure, are in accord with our conclusions. They find, for the site problem,

$$p_c(s,T)\simeq 0.51,\quad p_c(s,S)\simeq 0.55,\quad p_c(s,H)\simeq 0.7,$$

and for the bond problem

$$p_c(b,T)\simeq 0.34,\quad p_c(b,S)\simeq 0.50,\quad p_c(b,H)\simeq 0.67.$$

For the three-dimensional cubic lattices Elliott *et al.*[5] estimated

$$p_c(s,\text{F.C.C.})\simeq 0.18,\quad p_c(s,\text{B.C.C.})\simeq 0.22,$$
$$p_c(s,\text{S.C.})\simeq 0.28.$$

The last result for the simple cubic lattice has been confirmed by Domb and Sykes[6] who also considered the bond problem for which they found

$$p_c(b,\text{F.C.C.})\simeq 0.12 \quad\text{and}\quad p_c(b,\text{S.C.})\simeq 0.24.$$

All these estimates are consistent with the rigorous inequalities and bounds which follow from the present analysis.

2. CONNECTIVITY RELATIONS

Consider the plane triangular lattice T in which six bonds radiate from each site to the nearest neighboring sites. If all the bonds parallel to a fixed direction, say the horizontal, are removed, a lattice remains which is topologically identical with the plane square lattice S in which four bonds radiate from each site. In a similar way, if the squares of the plane square lattice are identified with the squares of a chess board and the bonds forming the right-hand edge of each black square are removed, the remaining lattice is topologically identical with the plane honeycomb lattice H of coordination number three. In short, the triangular lattice T *contains* the square lattice S which in turn contains the honeycomb lattice H. Loosely, one might say that the "connectivity" of the triangular lattice exceeds that of the square lattice which exceeds that of the honeycomb lattice.

When an infinite cluster of particles exists in a lattice, a connected path may be drawn from particle to particle "right across" the lattice. One feels intuitively that the greater the connectivity of a lattice, the easier it should be to find a path across for given p. This in turn suggests that a lattice of greater connectivity should have a lower critical probability.

To put the argument in more formal terms, consider the site problem on the triangular lattice T and suppose p exceeds the critical probability $p_c(s,T)$ so that an infinite cluster of particles is present in the lattice. The removal of a given bond may leave all particles in the infinite cluster still attached or it may disconnect one or more particles from the infinite cluster. Removal of a fraction of all bonds must, therefore, reduce the probability $R_s(p)$ that a given site belongs to an infinite cluster or, at best, leave it unchanged. Thus between the square and triangular lattice we have the inequality

$$R_s(s,S;p)\leqslant R_s(s,T;p). \tag{2.1}$$

(The symbol s denotes sites as before.) Now the critical probability is defined by the vanishing of the density of

infinite clusters, that is by

$$R_s(p_c-0)=0, \quad R_s(p_c+0)>0. \tag{2.2}$$

Consequently, the critical probabilities are related by the inequality

$$p_c(s,S)\geqslant p_c(s,T). \tag{2.3}$$

Clearly, the argument applies irrespective of dimensions or structural details to any pair of lattices, one of which contains the other, so we may write generally

$$p_c(s,L)\geqslant p_c(s,L^\dagger), \tag{2.4}$$

where $L^\dagger$ contains L. The result (1.1) follows as a special case and so does the corresponding chain of inequalities for the three-dimensional tetrahedral, simple cubic, body-centered cubic and face-centered cubic lattices. If all the bonds linking a site to the lattice are removed, the site itself may be deleted without affecting the argument. With this convention, the kagomé lattice,[10] for example, is contained in the triangular lattice and the inequality (2.4) applies.

For the bond problem, the argument goes through when it is noted that the removal of a set of bonds together with the particles that occupy them (at random) alters the total number of particles and bonds on the lattice but does not alter the probability p of occupation of a bond. Alternatively, one may consider the addition of extra bonds to, for example, the honeycomb lattice so as to form the square lattice. If an infinite cluster is present initially, then an infinite cluster must still exist in the resulting square lattice. Consequently, the critical probability for the square lattice cannot exceed that for the honeycomb lattice. This establishes the validity of (2.4) for the bond problem as well as for the site problem and hence justifies (1.2).

Experience with configurational problems on lattices indicates that any *over-all* change, however slight, in the structure of the lattice, results in a change of the corresponding critical parameter.[11] Consequently, it is highly probable that the equalities in the connectivity relations for the critical probabilities can only occur if the total density of deleted bonds is zero. To prove rigorously the strict inequality calls, however, for a more subtle analysis.

The foregoing general result may also be used in certain cases to derive a relationship between the bond and site problems on the same lattice. To show this we note that the *bond* problem on any lattice L is isomorphic with the *site* problem on a suitably defined *covering* lattice L^c. The covering lattice is constructed by replacing each bond of L by a site (placed at its center) and linking these sites together by sufficient new bonds to ensure that if two bonds of L meet (at a vertex), then the corresponding two sites of L^c are joined by a direct bond and vice versa. The covering lattice of a linear chain, for example, is also a linear chain, while the covering lattice of the plane honeycomb is the kagomé lattice. Any configuration of occupied and vacant (unoccupied) *bonds* on L will be in one-one correspondence with a configuration of occupied and vacant *sites* on L^c, and all topological relations between bonds on L hold between the corresponding sites on L^c. Consequently, the configurational generating function[7] for the bond problem on L is identical with that for the site problem on L^c and the two critical probabilities are equal, i.e.,

$$p_c(b,L^c)=p_c(s,L). \tag{2.5}$$

It is evident that not every lattice can be the covering lattice for another lattice. Consequently, although any bond problem is equivalent to a suitable site problem, the reverse is not true. In other words, the site problem is of greater generality.

Now the covering lattice S^c of the plane square lattice S is a chess board of alternate squares and tetrahedra, i.e., it is a normal square lattice of sites and bonds with two additional diagonal bonds (crossing) in each alternate square. If these diagonal bonds are removed, one regains the square lattice: in other words, S^c contains S. From (2.4) and (2.5) we thus obtain the inequality

$$p_c(s,S)\geqslant p_c(b,S) \tag{2.6}$$

which states that the critical probability for the site problem on the square lattice exceeds (or possibly equals) that for the bond problem on the same lattice. In combination with Harris' theorem,

$$p_c(b,S)\geqslant \tfrac{1}{2}. \tag{2.7}$$

This proves rigorously as mentioned in the Introduction, that

$$p_c(s,S)\geqslant \tfrac{1}{2}. \tag{2.8}$$

In accord with the comments above, it seems very probable that the strict inequality holds in (2.6) and (2.8).

For the plane triangular lattice, we can prove in a similar way that

$$p_c(s,T)\geqslant p_c(b,T). \tag{2.9}$$

The covering lattice T^c can be regarded as a kagomé lattice with nine additional diagonal bonds in each hexagon so that each of the six sites around a hexagon is connected directly to the other five. If these sites are numbered 1 to 6 in cyclic order and the six bonds (1,4), (2,5), (3,6), (1,3), (3,5), and (5,1) are deleted leaving the bonds (2,4), (4,6), and (6,2), the hexagon is decomposed into four triangles. If the corresponding bonds are deleted throughout the covering lattice T^c, one is left with a lattice which is topologically identical to the original triangular lattice T (although because of the distortion this is not easy to see at a glance).

[10] The kagomé lattice, which has coordination number four, can be constructed by placing a site at the center of each bond of the regular honeycomb lattice and joining these sites by nearest neighbor bonds (deleting the honeycomb bonds and sites).

[11] M. E. Fisher and M. F. Sykes, Phys. Rev. **114**, 45 (1959).

Intuitively, it is clear that the "connectivity" of a covering lattice is always greater than the connectivity of the original lattice. Thus, for example, the coordination number of L^c is larger than that of L (excluding the trivial case of a linear chain). Consequently, although we cannot present a general proof, it seems highly probable that the critical probability for the site problem always exceeds that for the bond problem on the same lattice. This is borne out by exact calculations for a range of pseudolattices[7] and by numerical estimates.[6]†

3. GENERAL BOUNDS

Hammersley[1–3] has shown how to obtain certain upper and lower bounds for the bond problem on a general lattice and, in particular, Broadbent and Hammersley[1] showed that if c_n is the number of distinct non-self-intersecting walks of n steps leaving the origin of a lattice L, then

$$p_c(b,L)\geqslant 1/\mu(L) \tag{3.1}$$

where

$$\ln\mu(L)=\lim_{n\to\infty} n^{-1}\ln c_n. \tag{3.2}$$

The existence of the walk limit $\mu(L)$ was proved by Hammersley and Morton,[12] and Fisher and Sykes[11] have shown how to obtain accurate estimates and rigorous upper and lower bounds for it. Hammersley's methods for the bond problem are easily extended to deal with the site problem, and one may prove, for example, the inequality corresponding to (3.1), namely,

$$p_c(s,L)\geqslant 1/\mu(L). \tag{3.3}$$

Application of this to the honeycomb lattice using the bound

$$\mu(H)<1.928 \tag{3.4}$$

obtained by Fisher and Sykes[11] (who estimated $\mu\simeq 1.845$) yields

$$p_c(s,H)>0.519. \tag{3.5}$$

This serves to disprove the conjecture that the critical probabilities for the site problem should be exactly $\frac{1}{2}$ for all plane lattices. (Initial numerical estimates lent some support to this conjecture.[4,5])

The result (3.5) can easily be improved, but in view of the *lower* bound[11] 1.787 for $\mu(H)$, one could not obtain a lower bound for $p_c(s,H)$ or $p_c(b,H)$ higher than 0.560 on the basis of the theorems (3.3) and (3.1). The numerical evidence of Domb and Sykes[6] suggests that these critical probabilities are likely to exceed 0.65 so that the theorems are rather weak.

To establish (3.3), we follow closely Broadbent and Hammersley's proof of (3.1). An n-step walk in L is said to be *allowed* if all the sites it passes through and reaches after leaving the origin are occupied by particles. If p_{jn} is the probability that exactly j of the c_n non-self-intersecting walks of n steps are allowed then the probability that at least one is allowed is simply

$$P_n=\sum_{j=1}^{c_n} p_{jn}. \tag{3.6}$$

When an infinite cluster is present there must be a nonzero probability that at least one non-self-intersecting walk is allowed for all n, i.e., $P_\infty>0$. Conversely, if $P_n\to 0$ as n becomes infinite, an infinite cluster cannot exist. Now by (3.6)

$$P_n\leqslant\sum_{j=1}^{c_n} jp_{jn}=\langle j\rangle_n \tag{3.7}$$

where $\langle j\rangle_n$ is the expected number of allowed walks of n steps. Since the walks are non-self-intersecting, the probability of a given walk being allowed is p^n, and so

$$\langle j\rangle_n=p^nc_n \tag{3.8}$$

which, in virtue of (3.2), behaves asymptotically like $[p\mu(L)]^n$. Consequently, if $p<1/\mu(L)$, $\langle j\rangle_n$ vanishes as n become infinite and by (3.7) so must P_n. This establishes that the critical probability must be at least as great as $1/\mu(L)$ which is the result (3.3).

4. HARRIS' THEOREM AND ITS EXTENSIONS

Harris[9] considered the bond problem on the plane square lattice and gave a rigorous formal mathematical proof of the inequality,

$$p_c(b,S)\geqslant\tfrac{1}{2}.$$

Owing to the complexity of the mathematical details, Harris' paper is not easy to follow and so we now present an outline of his proof with emphasis on the underlying principle which can then be applied to the site problem and to other lattices.

We first observe that for any lattice in which bonds (or sites) are occupied with probability p, one can associate a *complementary* situation in which the probability of occupation is $p'=q=1-p$. Clearly, there is complete symmetry between the two situations, and the complementary problem merely describes the clustering of *vacant* bonds (or sites) in the original case. Thus if the density of infinite clusters of occupied bonds (or sites) is $R(p)$, the density of infinite clusters of vacant bonds (or sites) is $R(1-p)$. Consequently, if $p_c<\frac{1}{2}$, there is a range of values of p (from p_c to $1-p_c$) in which an infinite cluster of *occupied* bonds (or sites) and an infinite cluster of *vacant* bonds exist *simultaneously*.

One may also associate with any L_2, which denotes a plane lattice without crossing bonds, a *dual lattice* L_2^D constructed by placing a site at the center of each elementary polygon in the lattice L_2 and joining these

† *Note added in proof.* This general theorem has since been established rigorously by Hammersley (to be published).

[12] J. M. Hammersley and K. W. Morton, J. Roy. Stat. Soc. **B16**, 23 (1954).

sites by nearest neighbor bonds so that one bond of the dual lattice crosses each bond of the original lattice and vice versa. The dual of a dual lattice is thus the original lattice. The triangular and honeycomb lattices are typical dual pairs, and the plane square lattice is evidently self-dual. For fixed p any configuration of occupied and vacant bonds on L_2 corresponds to a unique configuration on the dual lattice L_2^D.

As an introduction to Harris' methods, we show why an infinite cluster cannot exist in a two-dimensional lattice L_2 of finite width. To be more precise we will prove that the probability of such an event is zero. (Trivial cases are avoided by assuming throughout that $0<p<1$.) Consider firstly the site problem on a horizontal strip, m sites wide. If the probability $R_s(s,L_2;p)$ that a particular site belongs to an infinite cluster of occupied sites is nonzero, then there is a nonzero probability that an infinite allowed walk leaves the origin and continues indefinitely along the strip. Such a walk would be *blocked* however by the occurrence of a column of m vacant sites extending across the strip. (Remember that crossing bonds have been excluded.) The probability that a specified column is vacant is

$$q^m=(1-p)^m>0.$$

Now it is a well-known result of probability theory that an event which has a fixed nonzero probability will with probability one occur at least once in an infinite series of independent trials (and, in fact, will recur infinitely often).[13] We conclude that a column of vacant sites will eventually be encountered and consequently, the walk is certain to be blocked. Thus with with probability one no infinite allowed walk exists in the strip and so

$$R_s(s,L_2;p)=0 \quad (m \text{ finite}). \tag{4.2}$$

For the corresponding bond problem it is convenient (but not essential) to avoid edge effects by wrapping the lattice on a cylinder of circumference m sites. In this case the allowed walk proceeds via occupied *bonds* and so it is appropriate to consider the *complementary dual lattice* L_2^D of width m' bonds. A closed chain of occupied bonds stretching directly around the dual cylinder will occur at a specified position with probability $p'^{m'}=q^{m'}>0$ and hence is certain to occur at some position. Such a chain on L_2^D corresponds to a series of vacant bonds on L_2 which completely block any walk along the strip. (These blocking bonds on L_2 are not necessarily connected.) We may conclude as before that the probability of an infinite cluster of occupied bonds is zero.

Harris uses essentially the same principle for the infinite square lattice to prove that $R_s(b,S;\frac{1}{2})=0$. Since $R(p)$ is a nondecreasing function of p, this implies that $p_c(b,S)\geqslant\frac{1}{2}$ as stated before. When $p=\frac{1}{2}$, the complementary dual to the square lattice is identical to the square lattice itself. Harris makes the hypothesis that $R_s(b,S;\frac{1}{2})>0$, i.e., that an infinite cluster of occupied bonds does exist[14] when $p=\frac{1}{2}$. He then shows that

(A) *with probability one there exists a closed (connected) chain of occupied bonds which encloses the origin site O (and any specified finite region surrounding O).*

Consequently, such a chain also exists on the complementary dual lattice. But, as for the strip, this implies the existence of a series of vacant bonds in S which completely surround O and so block any infinite walk from O. Thus with probability one no infinite allowed walk leaves O and there is no infinite cluster in S. This contradicts the hypothesis and so we conclude $R_s(b,S;\frac{1}{2})=0$ as desired.

The complexity of Harris' proof arises from the difficulty of establishing the assertion (A) rigorously. It is necessary, in particular, to prove similar intermediate results about the probabilities, $R(p,\pi/2)$ and $R(p,\pi)$, that the origin is connected to a cluster with an infinite component lying entirely within one quarter plane and within one half-plane, respectively.

Before discussing the details of the proof and its extensions, however, we note that if (A) is proved for the site problem on a lattice L_2 on the hypothesis $R_s(s,L_2;\frac{1}{2})>0$, we may conclude generally that

$$p_c(s,L_2)\geqslant\tfrac{1}{2} \tag{4.3}$$

as stated in (1.6). In this case it is not necessary to introduce the dual lattice since at $p=\frac{1}{2}$ the complementary lattice situation is identical to the original. The assertion (A) thus shows that a connected ring of vacant sites surrounds the origin and hence with probability one there is no infinite cluster of occupied sites. This is a contradiction and the result (4.3) follows immediately. The gist of the argument may be stated picturesquely as showing the impossibility of a two-dimensional random sponge, or the fact that all seas in an infinite continent are lakes, that is, inland seas having finite area and shoreline![15]

For the bond problem on a pair of dual lattices one may formulate in analogy to (A) the assertion:

(B) *the hypothesis* $R(b,L_2;p)>0$ *implies* $R(b,L_2^D;1-p)=0$.

In other words, one cannot simultaneously have infinite clusters on a lattice and on its complementary dual.

[13] See, for example, W. Feller, *Probability Theory and its Applications* (John Wiley & Sons, New York, 1950), Sec. 8.3.

[14] Harris refers to *active* and *passive* bonds rather than occupied and vacant bonds. He calls an infinite connected cluster of occupied bonds a CISAL and of vacant bonds a CISPL. His $R(p)$ denotes our $R_s(b,S;p)$, i.e., the probability that a *site* is connected to an infinite cluster of bonds (see Appendexes A and B).

[15] I am indebted to Dr. M. F. Sykes for these metaphors.

Since $R(p)$ is a nondecreasing function of p which reaches a maximum at $p=1$, this implies

$$p_c(b,L_2)+p_c(b,L_2^D)\geqslant 1 \tag{4.4}$$

as stated in (1.7). Because infinite clusters of vacant and occupied bonds could apparently intersect one another freely at lattice sites, this result is less obvious intuitively than (4.3) and draws attention to the overwhelming importance of the assumption of complete randomness.

Rigorous proof of assertion (A) for the site problem and hence of (4.3) can quite easily be achieved by modifying the details of Harris' proof provided it is assumed (as mentioned in the Introduction) that the lattice has two orthogonal symmetry axes. The main feature is the repeated use of the complementary lattice situation rather than the dual lattice. The necessary changes are presented in detail in Appendix A.

Proof of (B) and hence of the theorem (4.4) can also be based on Harris' work, but is slightly more tricky owing to the need to keep account of the possibilities both on L_2 and on its (distinct) dual. The arguments needed for a strict proof are given in Appendix B.

ACKNOWLEDGMENTS

I am grateful to Dr. M. F. Sykes and Professor C. Domb for discussions and am indebted to Dr. J. M. Hammersley for his valuable criticisms of a draft version of the paper.

APPENDIX A. SITE PROBLEM ON A PLANE LATTICE

In this Appendix we consider only the site problem on a plane lattice with no crossing bonds and prove

$$p_c(s,L_2)\geqslant \tfrac{1}{2}. \tag{A1}$$

It is sufficient to prove the result for a *close packed lattice* L_{2c}, that is, a lattice in which all the basic polygons are triangles so that the nearest-neighbor sites of any site form a closed connected chain. For if the lattice L_2 in question is not close packed, it can be made so by the addition of sufficient extra (noncrossing) diagonal bonds,[16] and if L_{2c} is the resulting close-packed lattice, one has by (2.4)

$$p_2(s,L_2)\geqslant p_2(s,L_{2c}). \tag{A2}$$

We follow as closely as possible Harris' proof[9] and retain his numbering of sections, lemmas, etc., but with the prefix H. The appropriate transcriptions of Harris' terminology and definitions are as follows:

vertex: site
link: bond
active link: bond with both terminal sites occupied
passive link: bond with both terminal sites vacant
$R(p)$: $R_s(s,L_{2c};p)$
p_d: $p_c(s,L_{2c})$
CISAL (CISPL): infinite cluster of occupied (vacant) sites
T: translation of a lattice configuration through a distance corresponding to one unit cell.

The proof commences in Sec. H3 where the ergodic properties of T are established. This requires no further modification. Section H4 is a probabilistic lemma on the combination of independent events which is required in lemma H8.1 and also need no changes.

In Sec. H5 the result is proved for one quadrant. At this point it is convenient to introduce the symmetry assumption for the lattice. We suppose that, possibly after a suitable topological distortion of the lattice, it is possible to introduce a rectilinear coordinate system with coordinates x and y such that the lattice is invariant separately under the two transformations $x\rightarrow -x$ and $y\rightarrow -y$. Without loss of generality we may suppose the x and y axes are orthogonal. The transformation T will then stand for translation through one unit cell along either the x axis (taking x into $x+1$) or the y axis (taking y into $y+1$) as required from the context. To avoid undue complication of detail, we will omit the essentially straightforward modifications needed to allow for the fact that the sites of a general symmetric L_{2c} are not all equivalent and do not necessarily lie on the integral points of the Cartesian plane.[17]

An immediate consequence of the symmetry assumption is that the probability $R(p,\pi/2)$ that a vertex at (or adjacent to) the origin is connected to a cluster with an infinite component lying entirely in one quadrant is (for corresponding points) independent of the quadrant concerned. For the half-planes similarly the probability $R_Y(p,\pi)$ is the same for $L(y\geqslant 0)$ and $L(y\leqslant 0)$, but it is not necessarily equal to $R_X(p,\pi)$ the probability for the right and left half-planes $L(x\geqslant 0)$ and $L(x\leqslant 0)$. With these preliminaries, lemmas H5.1 and H5.2 which establish the existence of an active chain surrounding the origin in one quadrant when $R(\frac{1}{2},\pi/2)>0$ stand, but definition H5.1 is superfluous.

[16] By considering paths from one point of the plane to the corresponding points in the adjacent unit cells, one can prove that either the plane is entirely divided up into finite polygons or else that the lattice can be decomposed into separate one-dimensional strips. In the latter case the theorem is immediate. In the former, each finite polygon can be triangulated with a finite number of additional bonds.

[17] To remedy these omissions, one shows (along the lines of footnote 16) that either the theorem is immediate or that (C) any two sites in a cell may be connected together through a finite number of bonds. With (C) Broadbent's and Hammersley's general treatment of the bond problem with a finite number of different classes of site, etc. (footnote 1) shows that for all sites and bonds, $R(p)$ vanishes at the same value of $p(=p_c)$. The assertion (C) is also directly useful in generalizing lemmas H6.1, H7.2, and H8 when no sites lie on the symmetry axes.

Lemma H5.3 is still valid but its proof should be modified as follows: (The dual lattice is not used.)

Lemma 5.3. We have $R(\frac{1}{2},\pi/2)=0$.

Proof. Take $p=\frac{1}{2}$. Suppose $R(\frac{1}{2},\pi/2)>0$. Then with probability 1 there is a chain C of active links (i.e., occupied sites) satisfying the condition of lemma H5.2. If $R(\frac{1}{2},\pi/2)>0$, the probability is positive that in the complementary lattice there is an infinite cluster of occupied sites (CISAL) in the quadrant $L(x>0, y>0)$ containing the origin. But such a cluster represents a CISPL on L_{2c} which must include a chain C' of passive links (i.e., vacant sites) crossing C at some site. But this is impossible and so $R(\frac{1}{2},\pi/2)=0$.

In Sec. H6 it is proved with aid of lemma H5.3 that at $p=\frac{1}{2}$ no infinite cluster exists in the half-plane. For the present case it is necessary to modify lemma H6.1 to avoid assuming the full symmetry of the square lattice. In fact with two slight changes, the existing proof establishes:

Lemma 6.1′. If $R_Y(\frac{1}{2},\pi)>0$, then $R_X(\frac{1}{2},\pi)=0$ so that without loss of generality we may assume $R_X(\frac{1}{2},\pi)=0$ always.

The modifications are as follows: In the first paragraph on the hypothesis $R_Y(\frac{1}{2},\pi)=r>0$, the probability $P(W_i)$ is r^2 (rather than $r^2/2$). In the second paragraph one must, as before, consider the complementary lattice rather than the dual to prove

$$R_X(\tfrac{1}{2},\pi)=0.$$

The next stage is to establish that if $R(\frac{1}{2})>0$, there is a positive probability than an active chain lying in a finite part of the plane connects two specified points on one axis. Lemma H7.1 is valid but the first section of its proof leading to the conclusion "*there is with probability one an active half-circuit lying in* $L(x\geqslant 0)$" must be adapted to avoid consideration of the dual lattice. The assertion may be proved as follows: By lemma H6.1′ there is probability zero that there exists within $L(x\geqslant 0)$ either an infinite cluster of vacant sites containing the origin or an infinite cluster of occupied sites containing the origin. Hence with probability one the origin is connected either (a) to a finite cluster of vacant sites in $L(x\geqslant 0)$, or (b) to a finite cluster of occupied sites in $L(x\geqslant 0)$. In case (a) the *external perimeter*[7,3b] of this cluster is a set of occupied sites which, since the lattice is close packed (by hypothesis), must form a (connected) active half-circuit lying in $L(x\geqslant 0)$. In case (b) the external perimeter must belong to a finite cluster of vacant sites in $L(x\geqslant 0)$ and the outermost external perimeter of this second cluster will then be the required active half-circuit. The remainder of the proof of lemma H7.1 and lemma H7.2 and its proof (where the use made of symmetry is justified by our assumptions) require no modification.

The final result for the whole plane is deduced in Sec. H8. Lemma H8.1 shows that if $R(\frac{1}{2})>0$, there is a positive probability that a pair of active half-circuits link $(0,i)$ to $(0,-i)$ and together form a closed circuit around the origin. (The proof again relies on symmetry.) The proof of theorem H1 then establishes assertion (A) for the site problem and consideration of the complementary lattice then proves $p_c(s,L_{2c})\geqslant\frac{1}{2}$ as required.

APPENDIX B. BOND PROBLEM ON DUAL LATTICES

In this Appendix we consider the bond problem on a pair of dual lattices L_2 and L_2^D and prove that

$$p_c(b,L_2)+p_c(b,L_2^D)\geqslant 1. \tag{B1}$$

As in Appendix A, we follow Harris as closely as possible. Appropriate transcriptions are now

active link: occupied bond
passive link: vacant bond
$R(p)$: $R_s(b,L_2;p)$
p_d: $p_c(b,L_2)$
CISAL (CISPL): infinite cluster of occupied (vacant) bonds.

For brevity, the superscript D will be used to denote properties of the dual lattice. The transformation T is defined as in Appendix A. As before, we assume L_2 (and hence L_2^D) has two orthogonal axes of symmetry and we ignore, as before, the inessential complications that might arise from the existence of different classes of bonds, etc.[17]

The preliminary lemmas H3.1 and H4.1 require no modification. Lemmas H5.1 and H5.2 which prove the existence of a chain circling the origin in the quarter plane when $R(p,\pi/2)>0$ now hold separately for L_2 and for L_2^D [in the latter case on the hypothesis $R^D(p,\pi/2)>0$]. Consideration of the relationship between the complementary dual lattices as in lemma H5.3 now leads to the modified lemma

Lemma 5.3′. If $R(p,\pi/2)>0$, then $R^D(1-p,\pi/2)=0$ and vice versa.

Since $R(p,\pi/2)$ is a nondecreasing function of p, this implies:

Lemma 5.4′. There exist probabilities p_0 and p_0^D with $p_0+p_0^D\geqslant 1$, such that

$$R(p,\pi/2)=0,\quad p<p_0 \quad\text{and}\quad R^D(p,\pi/2)=0,\quad p<p_0^D$$
$$>0,\quad p>p_0 \qquad\qquad >0,\quad p>p_0^D.$$

If $R(p,\pi/2)$ is nonzero, then $R(p,\pi)$ and $R(p)$ must also be nonzero and, consequently, p_c and p_c^D certainly do not exceed p_0 and p_0^D, respectively. To obtain a result for the half-plane as in lemma H6.1, we suppose that $R(p,\pi)>0$ for some $p<p_0$. With the aid of lemma 5.4′ we can then prove, using Harris' argument, that there is an active chain connecting a site on the positive y axis to a site on the negative y axis. Consideration of the complementary dual lattice then leads to

Lemma 6.1′. If $R_Y(p,\pi)>0$ when $p<p_0$, then

$$R_X^D(1-p,\pi)=0.$$

The corresponding theorem holds for the dual lattice so that, as for the quarter plane, we deduce

Lemma 6.2′. There exist probabilities $p_1(\leqslant p_0)$ and $p_1^D(\leqslant p_0^D)$ with $p_1+p_1^D \geqslant 1$ such that

$$R_Y(p,\pi)=0, \quad p<p_1 \quad \text{and} \quad R_X^D(p,\pi)=0, \quad p<p_1^D$$
$$>0, \quad p>p_1 \qquad\qquad >0, \quad p>p_1^D.$$

We now need to establish to existence of active half-circuits between specified points. Following Harris' argument, one proves

Lemma 7.1′. If $R_X^D(p^D,\pi)=0$ (as when $p^D<p_1^D$) and if V is any box in L_2, then there exists a box V', $|V'|>|V|$ such that with probability exceeding $\frac{1}{2}$ (say), there is an active half-circuit in $L_2(x\geqslant 0)\bigcap(V'-V)$ when $p=1-p^D$ $(>1-p_1^D)$.

This result and the corresponding dual lemma can now be used to show the existence of an active half-circuit with predetermined end points on the hypothesis $R(p)>0$ when $R_X^D(1-p,\pi)=0$, i.e., when $1-p<p_1^D$. Thus:

Lemma 7.2′. Suppose $R(p)=r>0$ and $R_X^D(1-p,\pi)=0$ (fulfilling conditions of lemma 7.1′). Let V be any box in L_2. Then for sufficiently large i there is a box V'' in L_2, $|V|<i<|V''|$ such that the probability exceeds $r^2/32$ that there is an active chain in $L_2\bigcap(V''-V)$ connecting $(0,i)$ to $(0,-i)$.

We may now proceed without essential modification to lemma H8.1 which proves that under the same conditions there is with fixed probability a *closed* chain of active links through $(0,i)$ and $(0,-i)$ which sourrounds $(0,0)$ in a finite part of L_2. As in theorem H1, this is sufficient to prove the existence with probability one of a set of passive links on the complementary dual which block any active chains from the origin. Thus if $R(p)>0$ where $p>1-p_1^D$, then $R^D(1-p)=0$ and correspondingly for the dual situation. Combined with lemma 6.2′ (which states that $p_1+p_1^D\geqslant 1$), this proves that we cannot have both $R(p)>0$ and $R^D(1-p)>0$. The theorem (B1) follows at once.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 4 JULY-AUGUST, 1961

Properties of 2-Semi-Isomorphic Graphs and Their Applications: Active Network Analysis*

IVAN T. FRISCH AND WAN H. KIM
Department of Electrical Engineering, Columbia University, New York, New York
(Received February 17, 1961)

Two-semi-isomorphic graphs are defined by extending Whitney's concept of 2-isomorphic graphs. The properties of these newly defined graphs are then investigated. The discussion leads to the derivation of a simple formula for the determination of the sign of a tree product which is common to two 2-semi-isomorphic graphs. The formula is then applied to the topological analysis of mutually coupled active networks.

INTRODUCTION

THE application of the theory of linear graphs to the analysis of an ordinary network, i.e., a network consisting entirely of R, L and C elements, has been found very useful, particularly for computer analysis of a network.[1] However, because of dependent node pairs, the graphical characterization of a network containing active or mutually coupled devices is not as simple as it would be for an ordinary network. Percival[2] introduced the artificial two-terminal elements, the so-called "current and voltage elements," in order to characterize a mutually dependent node pair and defined the "mathematical equivalent circuit" of a network containing tubes and transformers. Following Percival's work, Coates[3] suggested the use of two graphs which were named the "current and voltage graphs" of an active network by Mayeda.[4] Kim[5,6] extended the use of the current and voltage graphs to the analysis of linear multipoles and derived the formulas for the topological transformation of active and mutually coupled networks. The use of a pair of graphs corresponding to an active network, instead of one graph as in the analysis of an ordinary network, raises a number of new problems. Among these, the most important one is "how to determine the sign of a tree product which is common to both graphs." Coates and Mayeda have previously considered this problem.

In this paper 2-semi-isomorphic graphs are defined

* This work is supported by National Science Foundation Grants.

[1] W. H. Kim, D. H. Younger, C. V. Freiman, and W. Mayeda, "On iterative factorization in network analysis by digital computer," Proceedings of 1960 East Joint Computer Conference (December 1960), pp. 241–253.

[2] W. W. Percival, Proc. Inst. Elec. Engrs. (London) **102, PtC**, 270 (1955).

[3] C. L. Coates, "General topological formulas for linear network functions," General Electric Research Laboratory, Rept. No. 57-RL-1746, Schenectady, New York.

[4] W. Mayeda, "Topological formulas for active networks," Interim Tech. Rept. No. 8, University of Illinois, Urbana, Illinois, January, 1958.

[5] W. H. Kim, Matrix and Tensor Quart. (London) **11**, 33 (1960).

[6] W. H. Kim, J. Franklin Inst. **271**, 200 (1961).

Lemma 6.2′. There exist probabilities $p_1(\leqslant p_0)$ and $p_1^D(\leqslant p_0^D)$ with $p_1+p_1^D \geqslant 1$ such that

$$R_Y(p,\pi)=0, \quad p<p_1 \quad \text{and} \quad R_X^D(p,\pi)=0, \quad p<p_1^D$$
$$>0, \quad p>p_1 \qquad\qquad >0, \quad p>p_1^D.$$

We now need to establish to existence of active half-circuits between specified points. Following Harris' argument, one proves

Lemma 7.1′. If $R_X^D(p^D,\pi)=0$ (as when $p^D<p_1^D$) and if V is any box in L_2, then there exists a box V', $|V'|>|V|$ such that with probability exceeding $\frac{1}{2}$ (say), there is an active half-circuit in $L_2(x\geqslant 0)\bigcap(V'-V)$ when $p=1-p^D$ $(>1-p_1^D)$.

This result and the corresponding dual lemma can now be used to show the existence of an active half-circuit with predetermined end points on the hypothesis $R(p)>0$ when $R_X^D(1-p,\pi)=0$, i.e., when $1-p<p_1^D$. Thus:

Lemma 7.2′. Suppose $R(p)=r>0$ and $R_X^D(1-p,\pi)=0$ (fulfilling conditions of lemma 7.1′). Let V be any box in L_2. Then for sufficiently large i there is a box V'' in L_2, $|V|<i<|V''|$ such that the probability exceeds $r^2/32$ that there is an active chain in $L_2\bigcap(V''-V)$ connecting $(0,i)$ to $(0,-i)$.

We may now proceed without essential modification to lemma H8.1 which proves that under the same conditions there is with fixed probability a *closed* chain of active links through $(0,i)$ and $(0,-i)$ which sourrounds $(0,0)$ in a finite part of L_2. As in theorem H1, this is sufficient to prove the existence with probability one of a set of passive links on the complementary dual which block any active chains from the origin. Thus if $R(p)>0$ where $p>1-p_1^D$, then $R^D(1-p)=0$ and correspondingly for the dual situation. Combined with lemma 6.2′ (which states that $p_1+p_1^D\geqslant 1$), this proves that we cannot have both $R(p)>0$ and $R^D(1-p)>0$. The theorem (B1) follows at once.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 4 JULY-AUGUST, 1961

Properties of 2-Semi-Isomorphic Graphs and Their Applications: Active Network Analysis*

IVAN T. FRISCH AND WAN H. KIM
Department of Electrical Engineering, Columbia University, New York, New York
(Received February 17, 1961)

Two-semi-isomorphic graphs are defined by extending Whitney's concept of 2-isomorphic graphs. The properties of these newly defined graphs are then investigated. The discussion leads to the derivation of a simple formula for the determination of the sign of a tree product which is common to two 2-semi-isomorphic graphs. The formula is then applied to the topological analysis of mutually coupled active networks.

INTRODUCTION

THE application of the theory of linear graphs to the analysis of an ordinary network, i.e., a network consisting entirely of R, L and C elements, has been found very useful, particularly for computer analysis of a network.[1] However, because of dependent node pairs, the graphical characterization of a network containing active or mutually coupled devices is not as simple as it would be for an ordinary network. Percival[2] introduced the artificial two-terminal elements, the so-called "current and voltage elements," in order to characterize a mutually dependent node pair and defined the "mathematical equivalent circuit" of a network containing tubes and transformers. Following Percival's work, Coates[3] suggested the use of two graphs which were named the "current and voltage graphs" of an active network by Mayeda.[4] Kim[5,6] extended the use of the current and voltage graphs to the analysis of linear multipoles and derived the formulas for the topological transformation of active and mutually coupled networks. The use of a pair of graphs corresponding to an active network, instead of one graph as in the analysis of an ordinary network, raises a number of new problems. Among these, the most important one is "how to determine the sign of a tree product which is common to both graphs." Coates and Mayeda have previously considered this problem.

In this paper 2-semi-isomorphic graphs are defined

* This work is supported by National Science Foundation Grants.

[1] W. H. Kim, D. H. Younger, C. V. Freiman, and W. Mayeda, "On iterative factorization in network analysis by digital computer," Proceedings of 1960 East Joint Computer Conference (December 1960), pp. 241–253.

[2] W. W. Percival, Proc. Inst. Elec. Engrs. (London) **102, PtC**, 270 (1955).

[3] C. L. Coates, "General topological formulas for linear network functions," General Electric Research Laboratory, Rept. No. 57-RL-1746, Schenectady, New York.

[4] W. Mayeda, "Topological formulas for active networks," Interim Tech. Rept. No. 8, University of Illinois, Urbana, Illinois, January, 1958.

[5] W. H. Kim, Matrix and Tensor Quart. (London) **11**, 33 (1960).

[6] W. H. Kim, J. Franklin Inst. **271**, 200 (1961).

by an extension of Whitney's definition of 2-isomorphic graphs.[7] The properties of these newly defined 2-semi-isomorphic graphs are then investigated and a simple formula is obtained for the sign of the common tree product of the two graphs. The current and voltage graphs for a network are shown to be special cases of 2-semi-isomorphic graphs and the application of the properties of 2-semi-isomorphic graphs to their analysis is illustrated.

1. 2-SEMI-ISOMORPHIC GRAPHS AND THEIR PROPERTIES

Definition 1

The principal node of an edge[8] of a tree of a connected linear graph is the one of the two terminal nodes of the edge that is located furthest from a fixed reference node in a path of the tree containing the edge and the reference node. The other node of the edge is called the "minor node" of the edge in the tree. Since there exists a unique path in a tree from a node to every other node, the principal and minor nodes of an edge of a tree are always distinguishable. Note that the principal and minor nodes of an edge are determined with respect to a chosen tree of a connected graph, and, if there exists another tree of the graph which contains the edge, then the principal and minor nodes of the edge in this tree may not be the same as found in the first tree. This is illustrated in Example 1. It is also clear that a reference node of a graph cannot be the principal node of any edge of the graph. Then lemma 1 follows directly from definition 1.

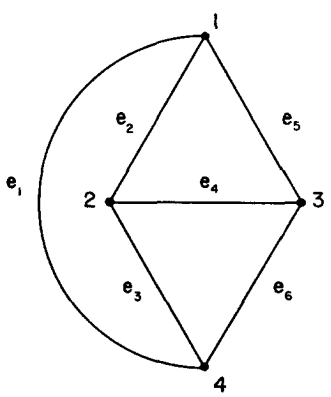


FIG. 1. A graph G.

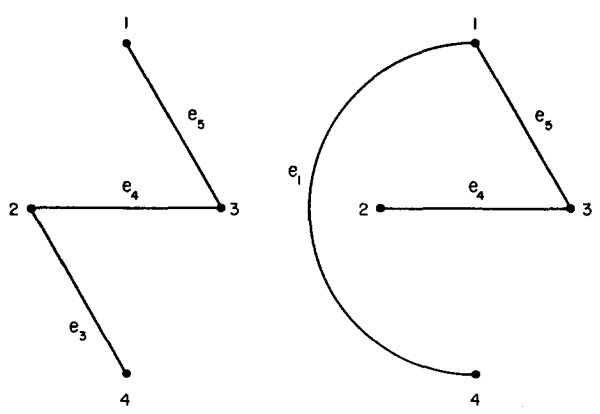


FIG. 2. Trees (a) t_1 and (b) t_2 of G.

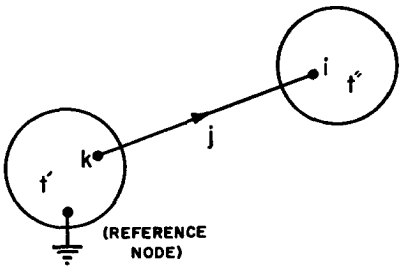


FIG. 3. Tree t divided into disjoint subgraphs t' and t''.

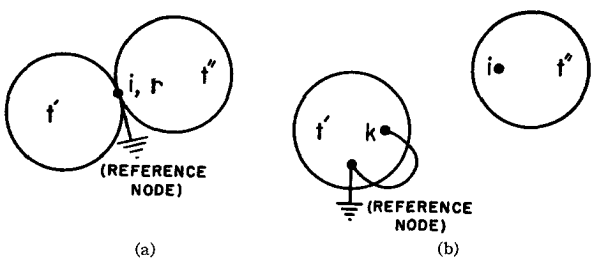


FIG. 4. Subgraphs represented by (a) At_{ij} and (b) At_{kj}.

Lemma 1

In a graph or subgraph of a graph which contains no circuits (loops), no node can be the principal node of more than one edge of the graph or of the subgraph of a graph.

Example 1. Given the graph G in Fig. 1 where node 4 is the reference node. Let us choose a pair of trees of G, t_1 and t_2, as shown in Fig. 2. Then, in t_1, the principal nodes of edges e_4 and e_5 are nodes 3 and 1, respectively. However, in t_2, nodes 2 and 3 are the principal nodes of e_4 and e_5. It is noted that the principal and minor nodes of e_4 in t_1 are interchanged in t_2 as are the principal and minor nodes of e_5.

Let us denote the incidence matrix[9] of a connected and oriented graph G with v vertices and e edges by A.

The incidence matrix $A=[a_{ij}]$ is a matrix with v rows and e columns such that

$a_{ij}=1$ if the element j is incident at vertex i and is oriented away from the vertex,
$=-1$ if the element j is incident at vertex i and is oriented towards the vertex,
$=0$ if the element j is not incident at vertex i.

[7] H. Whitney, J. Math., **55**, 236 (1933).

[8] The term "branch" may be substituted occasionally for "edge," but the term "element" will only be used in the sense of "element of a network" or "element of a matrix."

[9] S. Seshu, IRE Trans. PGCT **CT-2**, 356 (1955).

TABLE I. Relationships between G_1 and G_2 for operations performed on G_1.

Operations on G_1	Relationships between G_1 and G_2
No operation	isomorphic
Operation 1	1-isomorphic
Operations 1 and/or 2	2-isomorphic
Operations 1 and/or 2 and/or 3	2-semi-isomorphic

Let us denote a square submatrix of the incidence matrix corresponding to a tree of the graph by A_t. $A_{t_{ij}}$ is the matrix obtained by deleting row i and column j from A_t. $|A_t|_{ij}$ is the minor of the determinant of A_t of (ij) position. Then

Lemma 2

$$|A_t|_{ij} \text{ is } \begin{cases} \text{nonzero, if } i\text{th node is the principal node of } j\text{th edge in } t, \text{ and} \\ \text{zero, if } i\text{th node is the minor node of } j\text{th edge in } t. \end{cases}$$

Proof. Consider an edge j of a tree t of a connected graph. Denote the principal and minor nodes of edge j by nodes i and k, respectively. Since a tree is a circuitless graph or subgraph of a connected graph, it is possible to cut tree t by removing edge j into two disjoint subgraphs t' and t'', such that t' contains the reference node of t, as shown in Fig. 3. Now expand the determinant of $A_t=[a_{pq}]$ about column j. Then, one gets[10]

$$|A_t| = (-1)^{i+j}a_{ij}|A_t|_{ij} + (-1)^{k+j}a_{kj}|A_t|_{kj} \neq 0, \quad (1)$$

where a_{ij} and a_{kj} are nonzero elements of A.

The submatrix $A_{t_{ij}}$ is the incidence matrix of the subgraph of t obtained by removing edge j and identifying node i with the reference node r as shown in Fig. 4(a), while $A_{t_{kj}}$ represents the subgraph of t obtained by removing edge j and identifying node k with the reference node as shown in Fig. 4(b). Since the subgraph corresponding to $A_{t_{kj}}$ contains a circuit and is not connected, $|A_t|_{kj}=0$. Therefore, from (1), $|A_t|_{ij}\neq 0$. It is also clear that the subgraph of $A_{t_{ij}}$ is a connected subgraph of t with no circuits. Hence lemma 2.

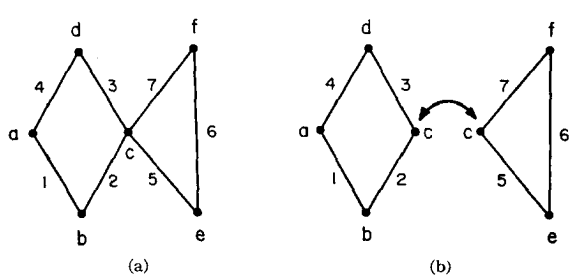


FIG. 5. 1-Isomorphic graphs (a) G_1 and (b) G_2.

[10] It is a well-known theorem that a square submatrix, of order $(n-1)$, of the incidence matrix of a connected graph with n nodes is nonzero if and only if the edges corresponding to the columns of the submatrix constitute a tree of the graph (see reference 9).

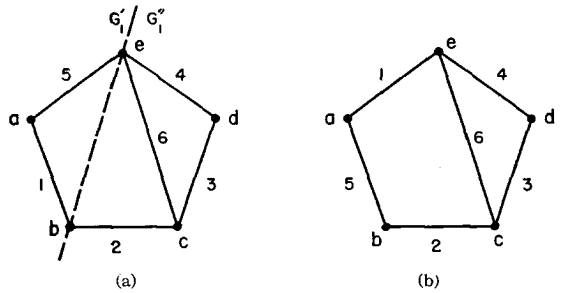


FIG. 6. 2-Isomorphic graphs (a) G_1 and (b) G_2.

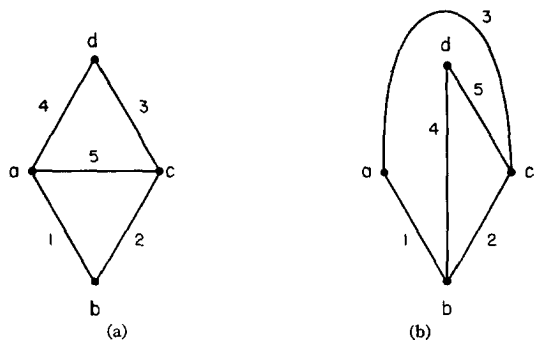


FIG. 7. 2-Semi-isomorphic graphs (a) G_1 and (b) G_2.

Let us now define the operations to be performed on a graph G_1 containing e edges and n nodes.

Operations. 1. Break G_1 at a single node into two disjoint connected pieces, or join two connected pieces at a node. 2. Cut G_1 at two nodes into two connected pieces and turn one of the pieces around at the two nodes. 3. Remove one or more edges in G_1 and put them back between any two nodes in G. Operation 3 is proposed as an addition to operations 1 and 2 defined by Whitney.[7]

If we let the graph resulting from any of the operations defined, or combination of the three operations, be denoted by G_2, then Table I gives the relationships between G_1 and G_2.

The following example illustrates these relationships between G_1 and G_2.

Example 2. (a) 1-isomorphic graphs: If we are given a graph G_1 in Fig. 5(a), cutting G_1 at c results in G_2 as shown in Fig. 5(b). G_1 and G_2 are 1-isomorphic with each other. (b) 2-Isomorphic graphs: We are given a graph G_1 in Fig. 6(a). Break G_1 at nodes b and e into subgraphs G_1' and G_1'', G_1' containing edges 1 and 5, and G_1'', edges 2, 3, 4, and 6. Next, turn G_1' around nodes b and e to obtain G_2 as shown in Fig. 6(b). G_1 and G_2 are said to be 2-isomorphic with each other. (c) 2-semi-isomorphic graphs: We are given a graph G_1 in Fig. 7(a). Remove edges 3, 4, and 5 in G_1 and reconnect them between node pairs (ac), (bd), and (cd). This results in G_2 as shown in Fig. 7(b). G_1 and G_2 are 2-semi-isomorphic with each other. From the definition

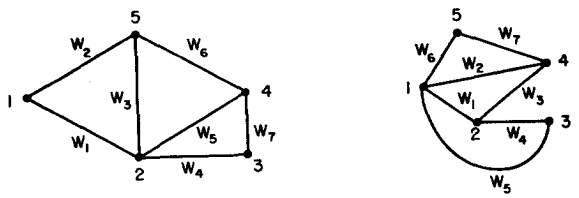


FIG. 8. 2-Semi-isomorphic graphs (a) G_1 and (b) G_2.

of operation 1 and the example, it is noted that operation 1 is valid on G_1 only when G_1 is a separable connected graph or collection of disconnected graphs. However, a graph corresponding to an electrical network is, in general. not separable. Therefore, we eliminate operation 1 from our further discussion. That is, we assume that a graph G_2 which is 2-semi-isomorphic to G_1 is obtained by operations 2 and/or 3.

Definition 2

In 2-semi-isomorphic graphs G_1 and G_2, edge e_i in G_1 and the same edge in G_2 together are called the "edge-pair e_i" of G_1 and G_2.

Definition 3

Let the terminal nodes of the edges in an edge pair of 2-semi-isomorphic graphs G_1 and G_2 be (i,j) and (p,q), respectively. If $i=p$ and $j=q$, and the orientation of both edges in the edge-pair is the same, then the edge pair is said to be ordinary; otherwise, it is active. Each edge of an ordinary and active edge pair is called an ordinary edge and active edge, respectively. The definition of the classification of edges in 2-semi-isomorphic graphs will become evident in a later section in which their application is illustrated.

Definition 4

A pair of trees of 2-semi-isomorphic graphs G_1 and G_2 that contain the same edges are called a tree pair, and the product of weights of the edges constituting a tree pair is a common tree product of G_1 and G_2.

Definintions 1–4 are illustrated in the following example.

Example 3. Consider the 2-semi-isomorphic graphs G_1 and G_2 as shown in Fig. 8.

The edge-pairs w_1 and w_4 are the ordinary edge pairs, but w_3, w_2, w_5, w_6, and w_7 are the active edge pairs. If we choose a tree pair of G_1 and G_2 as shown in Fig. 9, where node 2 is picked as the reference node, then, the principal and minor nodes of each edge in the trees are given in Table II.

Let us now introduce the orientation of edges in 2-semi-isomorphic graphs G_1 and G_2 of $(n+1)$ nodes and e edges. The incidence matrix of a tree T of an oriented and connected graph, $A_T=[a_{ij}]$, has $(+1)$ as an element of (i,j) position if edge j is incident at node i in T and the orientation of the edge is directed away from node i. If the orientation of edge j is directed toward node i, then $a_{ij}=-1$. The trees of a tree pair of G_1 and G_2 are denoted by T_1 and T_2, respectively. Then, the sign of the common tree product of tree pair T_1 and T_2, ϵ, is given by

$$\epsilon=|A_{T_1}\cdot A_{T_2}{}^t|,$$

where A^t is the transpose of A.

If $A_{T_1}=A_{T_2}$, which is true for a graph corresponding to ordinary network, the sign of the common tree product ϵ is always positive.[7] However, when $A_{T_1}\neq A_{T_2}$, the sign of the common tree product should be determined.

Lemma 3. The sign of the common tree product of a tree pair of 2-semi-isomorphic graphs is determined only by the active edge pairs in the tree pair.

Proof. Consider a tree pair T_1 and T_2 of $(n+1)$ nodes, containing k active edge pairs and $(n-k)$ ordinary edge pairs. Then, arrange the columns of the incidence matrices of T_1 and T_2, A_{T_1} and A_{T_2}, such that the first $(n-k)$ columns correspond to the ordinary edge pairs and the last k columns to the active edge pairs. Thus, we get

$$\begin{aligned} A_{T_1}&=\left[\begin{array}{ccc|ccc} a_{11}&\cdots&a_{1,n-k} & a_{1,n-k+1}&\cdots&a_{1n}\\ &&& &&\\ a_{n1}&\cdots&a_{n,n-k} & a_{n,n-k+1}&\cdots&a_{nn}\end{array}\right]=[P_{11}P_{12}]\\ A_{T_2}&=\left[\begin{array}{ccc|ccc} b_{11}&\cdots&b_{1,n-k} & b_{1,n-k+1}&\cdots&b_{1n}\\ &&& &&\\ b_{n1}&\cdots&b_{n,n-k} & b_{n,n-k+1}&\cdots&b_{nn}\end{array}\right]=[Q_{11}Q_{12}], \end{aligned} \tag{3}$$

where

$$\begin{aligned} P_{11}&=\begin{bmatrix} a_{11}&\cdots&a_{1,n-k}\\ &&\\ a_{n1}&\cdots&a_{n,n-k}\end{bmatrix} \quad P_{12}=\begin{bmatrix} a_{1,n-k+1}&\cdots&a_{1n}\\ &&\\ a_{n,n-k+1}&\cdots&a_{nn}\end{bmatrix}\\ Q_{11}&=\begin{bmatrix} b_{11}&\cdots&b_{1,n-k}\\ &&\\ b_{n1}&\cdots&b_{n,n-k}\end{bmatrix} \quad Q_{12}=\begin{bmatrix} b_{1,n-k+1}&\cdots&b_{1n}\\ &&\\ b_{n,n-k+1}&\cdots&b_{nn}\end{bmatrix}, \end{aligned} \tag{4}$$

TABLE II. Principal and minor nodes for edges in trees of Fig. 9.

	Tree of G_1				Tree of G_2			
Edges	w_1	w_2	w_4	w_6	w_1	w_2	w_4	w_6
Principal nodes	1	5	3	4	1	4	3	5
Minor nodes	2	1	2	5	2	1	2	1

and since P_{11} and Q_{11} represent the incidence relationship of ordinary edge pairs, due to definition 3 $P_{11}=Q_{11}$, i.e., $a_{ij}=b_{ij}$ for $i=1, \cdots n$ and $j=1, \cdots n-k$.

Now, assume that the edge 1 corresponding to the first column of A_{T_1} and A_{T_2} is incident at nodes r and s, i.e., a_{r1}, a_{s1}, b_{r1}, and b_{s1} are nonzero and $a_{r1}=b_{r1}=-a_{s1}=-b_{s1}$, and expand the determinants of A_{T_1} and A_{T_2} about the first column. If node r is the principal node of edge 1 of T_1, node s is the principal node of edge 1 of T_2, and $s\neq r+1$; then, interchange the sth and $(r+1)$th rows in the matrices A_{T_1} and A_{T_2} before expanding the determinants of A_{T_1} and A_{T_2}. If $s=r+1$, then no interchanges of rows are necessary. Thus we get, using lemmas 1 and 2,

$$\begin{aligned} |A_{T_1}| &= (-1)^{r+1+u} a_{r1} |M_{T_1}|_{r1} \\ |A_{T_2}| &= (-1)^{r+1+1+u} b_{s1} |M_{T_2}|_{s1}, \end{aligned} \tag{5}$$

where u is the number of interchanges of rows (equal to 1 or 0), $|M_T|_{pq}$ is the minor of the determinant of A_T of (p,q) position, and $[M_T]_{pq}$ is the matrix corresponding to the minor. Since edge 1 is an ordinary edge, $a_{r1}=-b_{s1}$ for $r\neq s$, and (5) is rewritten as

$$\begin{aligned} |A_{T_1}| &= (-1)^{r+1+u} a_{r1} |M_{T_1}|_{r1} \\ |A_{T_2}| &= (-1)^{r+1+u} a_{r1} |M_{T_2}|_{s1}. \end{aligned} \tag{6}$$

Therefore, the sign prefixing minors $|M_{T_1}|_{r1}$ and $|M_{T_2}|_{s1}$ are the same.

When edge 1 has node r as its principal node both in T_1 and T_2, i.e., $a_{r1}=b_{r1}$, then the determinants of the incidence matrices are expanded about the element a_{r1} and we get

$$\begin{aligned} |A_{T_1}| &= (-1)^{r+1} a_{r1} |M_{T_1}|_{r1}, \\ |A_{T_2}| &= (-1)^{r+1} b_{r1} |M_{T_2}|_{r1} = (-1)^{r+1} a_{r1} |M_{T_2}|_{r1}. \end{aligned} \tag{7}$$

Thus, the minors in (7) have the same sign prefixing them.

Next, expand the determinants of $[M_{T_1}]_{r1}$ and $[M_{T_2}]_{s1}$ (or $[M_{T_2}]_{r1}$) about the first column of the submatrices. If edge 2 of the tree pair, which corresponds to the first column of the submatrices, corresponds to the minors and has the same node as its principal node both in T_1 and T_2, the process described by Eq. (7) is repeated. If the principal node of edge 2 in T_1 is different from that of edge 2 in T_2, then the process described by Eq. (6) is repeated. Therefore, one continues to expand the determinants of the incidence matrices of T_1 and T_2 by the Laplace expansion as described by the Eqs. (6) and (7) until the remaining minors of the determinants contain only columns corresponding to active edge pairs in the tree pairs. It is clear that the sign of the remainders are the same.

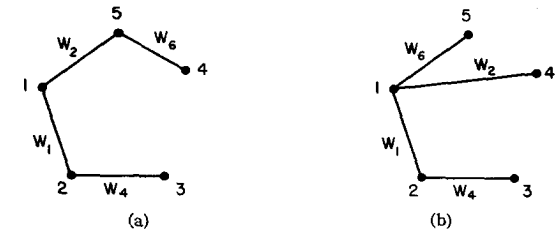


FIG. 9. A tree pair of (a) G_1 and (b) G_2 of Fig. 3.

Furthermore, suppose we relabel row s as row (r,s) when a_{ri} is used as a complementary minor in expanding about a column, where s is the minor vertex of the edge corresponding to the column. Then, the final resulting submatrix will be the incidence matrix of the reduced graph of the tree in the tree pair 1. The sign of the tree product then depends only on this final reduced tree. Hence lemma 3.

Definition 5

The reduced tree pair of a tree pair of 2-semi-isomorphic graphs are a pair of the subgraphs derived from the tree pair by removing all ordinary edge pairs and identifying their terminal nodes in the pair.

Example 4. Given a tree pair T_1 and T_2 as shown in Fig. 10, where node 0 is the reference node. The incidence matrices of T_1 and T_2, A_{T_1} and A_{T_2}, are found as

$$A_{T_1}=\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} +1 & +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 \end{bmatrix} \end{array}$$

$$A_{T_2}=\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} +1 & +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & -1 & 1 \\ 0 & -1 & 0 & 0 & 0 \end{bmatrix} \end{array}, \tag{8}$$

where the first three columns of A_{T_1} and A_{T_2} correspond to the ordinary edge pairs 1, 2, and 3.

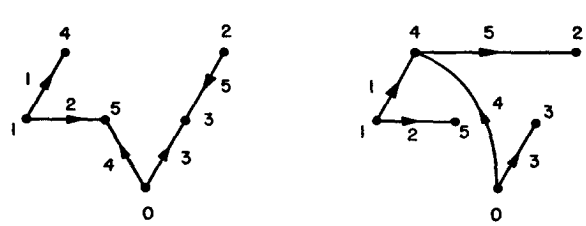


FIG. 10. A tree pair, (a) T_1 and (b) T_2.

First, interchange rows 2 and 4 in the matrices, then expand the determinants. Thus, we get

$$|A_{T_1}| = (-1)^3 \begin{array}{c|ccccc} & w_1 & w_2 & w_3 & w_4 & w_5 \\ 1 & +1 & +1 & 0 & 0 & 0 \\ 4 & -1 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & -1 & 0 & -1 \\ 2 & 0 & 0 & 0 & 0 & 1 \\ 5 & 0 & -1 & 0 & -1 & 0 \end{array} = (-1) \begin{array}{c|cccc} & w_2 & w_3 & w_4 & w_5 \\ (1,4) & +1 & 0 & 0 & 0 \\ 3 & 0 & -1 & 0 & -1 \\ 2 & 0 & 0 & 0 & +1 \\ 5 & -1 & 0 & -1 & 0 \end{array},$$

$$|A_{T_2}| = (-1)^3 \begin{array}{c|ccccc} & w_1 & w_2 & w_3 & w_4 & w_5 \\ 1 & +1 & +1 & 0 & 0 & 0 \\ 4 & -1 & 0 & 0 & -1 & 1 \\ 3 & 0 & 0 & -1 & 0 & 0 \\ 2 & 0 & -1 & 0 & 0 & 0 \\ 5 & 0 & -1 & 0 & 0 & 0 \end{array} = (-1) \begin{array}{c|cccc} & w_2 & w_3 & w_4 & w_5 \\ (1,4) & 0 & 0 & -1 & +1 \\ 3 & 0 & -1 & 0 & 0 \\ 2 & 0 & 0 & 0 & -1 \\ -5 & -1 & 0 & 0 & 0 \end{array}. \tag{9}$$

Interchanging rows 5 and 3 we get

$$|A_{T_1}| = (-1)^2 \begin{array}{c|cccc} & w_2 & w_3 & w_4 & w_5 \\ (1,4) & +1 & 0 & 0 & 0 \\ 5 & -1 & 0 & -1 & 0 \\ 2 & 0 & 0 & 0 & +1 \\ 3 & 0 & -1 & 0 & -1 \end{array} = \begin{array}{c|ccc} & w_3 & w_4 & w_5 \\ (1,4,5) & 0 & -1 & 0 \\ 3 & 0 & 0 & +1 \\ 2 & -1 & 0 & -1 \end{array},$$

$$|A_{T_2}| = (-1)^2 \begin{array}{c|cccc} & w_2 & w_3 & w_4 & w_5 \\ (1,4) & 0 & 0 & -1 & +1 \\ 5 & -1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & -1 \\ 3 & 0 & -1 & 0 & 0 \end{array} = \begin{array}{c|ccc} & w_3 & w_4 & w_5 \\ (1,4,5) & 0 & -1 & +1 \\ 3 & 0 & 0 & -1 \\ 2 & -1 & 0 & 0 \end{array}. \tag{10}$$

If we expand about the element a_{31} in each determinant, we have

$$|A_{T_1}| = (-1) \begin{array}{c|cc} & w_4 & w_5 \\ (1,4,5) & -1 & 0 \\ 3 & 0 & +1 \end{array},$$

$$|A_{T_2}| = (-1) \begin{array}{c|cc} & w_4 & w_5 \\ (1,4,5) & -1 & +1 \\ 3 & 0 & -1 \end{array}. \tag{11}$$

FIG. 11. Reduced trees.

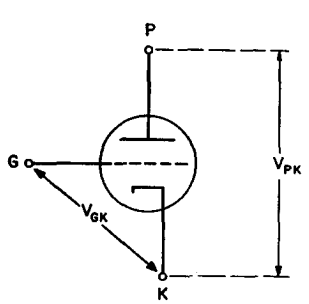


FIG. 12. Triode.

It is noted that the matrices corresponding to the resulting minors are the incidence matrices of the reduced trees shown in Fig. 11.

Definition 6

The sign of an active edge pair is defined as $+1$ if both edges in the pair are directed away or toward their principal nodes: -1, otherwise.

We therefore have

Theorem 1. The sign of the common tree product of a tree pair, is given by

$$\epsilon = (-1)^{\gamma} \prod_k [\text{sign of active edge pairs of the reduced tree pair of a tree pair}], \tag{12}$$

where γ is the number of interchanges of edges needed to give all active edge pairs in the reduced tree pair the same principal nodes, and k is the number of active edge pairs in the reduced tree pair.

Proof. By lemma 3, we need only consider the

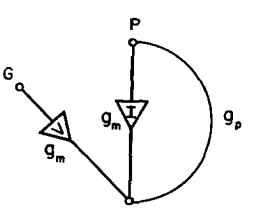


FIG. 13. Mathematical equivalent circuit of triode G.

reduced graphs of the trees in the tree pair to determine the sign of the common tree product. Let t_1' and t_2' denote the reduced trees and $A_{T_1}'=[a_{ij}']$ and $A_{T_2}'=[b_{ij}']$ their respective incidence matrices. Now, interchange columns in A_{T_1}' until all edges in t_1' and t_2' have the same principal vertices, the resulting matrix being A_{T_1}''. By lemma 1, this can always be done.

Suppose that, in column 1 of A_{T_1}'' and A_{T_2}'', the respective principal nonzero entries corresponding to principal vertices are a_{1s}' and b_{1s}'. Then,

$$\begin{aligned} |A_{T_1}'| &= (-1)^{1+s}a_{1s}'|M_{T_1}'|_{1s}(-1)^{\gamma}, \\ |A_{T_2}'| &= (-1)^{1+s}b_{1s}'|M_{T_2}'|_{1s} \quad , \end{aligned} \tag{13}$$

where $|M_T'|_{ij}$ is the minor of the determinant of A_T' of (i,j) position and γ is the number of interchanges of columns in A_{T_1}' to give all edge pairs in t_1' and t_2' the same principal vertices.

If the sign of edge pair 1 is positive, then $b_{1s}'=a_{1s}'$ and

$$\begin{aligned} \epsilon &= |A_{T_1}'|\,|A_{T_2}'| \\ &= [(-1)^{(1+s)}a_{1s}']^2|M_{T_1}'|\,|M_{T_2}'|(-1)^{\gamma} \\ &= |M_{T_1}'|\,|M_{T_2}'|(-1)^{\gamma}. \end{aligned} \tag{14}$$

If the sign of edge pair 1 is negative, then $b_{1s}'=-a_{1s}'$ and

$$\begin{aligned} \epsilon &= |A_{T_1}'|\,|A_{T_2}'| \\ &= [(-1)^{(1+s)}a_{1s}']^2|M_{T_1}'|\,|M_{T_2}'|(-1)^{\gamma+1} \\ &= |M_{T_1}'|\,|M_{T_2}'|(-1)^{\gamma+1}. \end{aligned} \tag{15}$$

We next expand the minors $|M_{T_1}'|$ and $|M_{T_2}'|$ about edge 2 which is the first edge in the minors. The process described by Eqs. (14) and (15) is thus repeated until we have expanded about all columns of A_{T_1}' and A_{T_2}'. The final result is $\epsilon=(-1)^{\gamma}(-1)^{\beta}$, where β is the number of active edge pairs with a negative sign. Hence theorem 1.

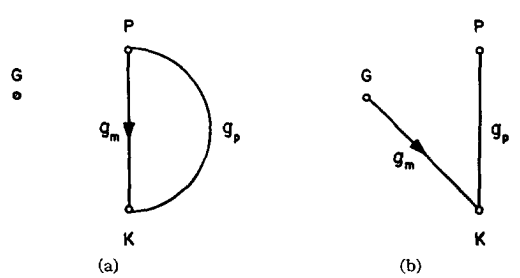


FIG. 14. Current and voltage graphs. (a) Current graph G_I. (b) Voltage graph G_V.

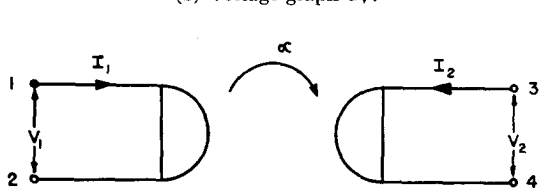


FIG. 15. A gyrator.

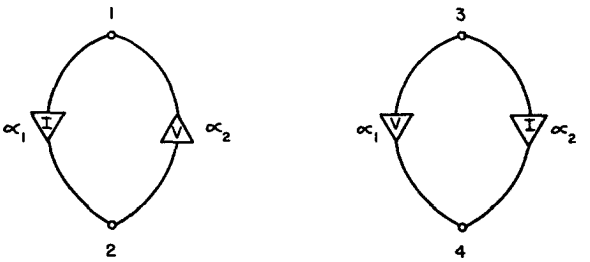


FIG. 16. Mathematical equivalent circuit of gyrator.

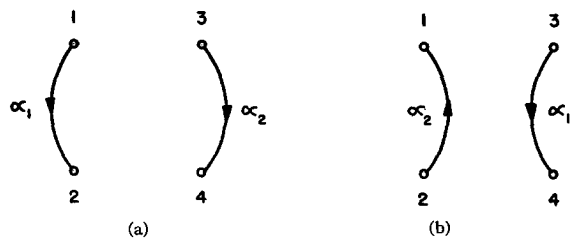


FIG. 17. (a) Current and (b) voltage graphs.

2. APPLICATIONS OF 2-SEMI-ISOMORPHIC GRAPHS

There are a number of engineering problems in which one may find applications of the properties of 2-semi-isomorphic graphs. In this section we shall discuss the analysis of active and mutually coupled and non-reciprocal linear networks. Let G be a graph representing a linear network, and G_I and G_V denote the subgraphs of G, the so-called "current and voltage graphs," respectively.

These concepts are illustrated in examples 5 and 6.

Example 5. A triode of Fig. 12 is characterized as a linear active device by

$$I_p=g_mV_{GK}+g_pV_{pK}, \tag{16}$$

where g_m is the transconductance and g_p is the plate conductance. The "mathematical equivalent circuit G" of the triode characterized by (16) is given in Fig. 13. Note that since g_p relates the voltage and current between the same pair of nodes in (16), it is represented by a single edge between those nodes in Fig. 13. On the other hand, g_m relates the voltage and current between two different pairs of nodes in (16) and, hence, appears as an edge between both pairs of nodes in Fig. 13. The current and voltage graphs of the graph in Fig. 13 are shown in Figs. 14(a) and 14(b), respectively. Note that the passive edge g_p appears between the same nodes in G_I and G_V, whereas the active edge g_m does not.

Example 6. A two-channel gyrator, shown in Fig. 15, may be characterized as a linear nonreciprocal device by the following set of equations:

$$\begin{aligned} I_1 &= \alpha_1 V_2 \\ I_2 &= -\alpha_2 V_2, \end{aligned} \tag{17}$$

where $\alpha_1=\alpha_2=\alpha$ is the gyrator admittance. The

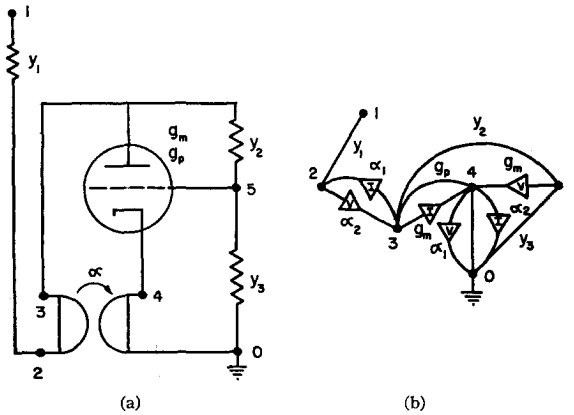


FIG. 18. An active and mutually coupled network and its mathematical equivalent circuit. (a) An active network. (b) Mathematical equivalent circuit of a where $\alpha_1=\alpha_2=\alpha$.

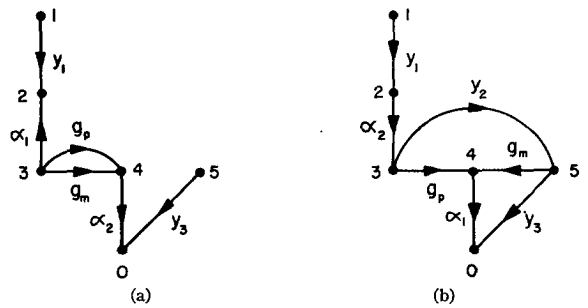


FIG. 19. Current and voltage graphs. (a) Current graph G_I. (b) Voltage graph G_V.

"mathematical equivalent circuit" G of the gyrator is shown in Fig. 16. The current and voltage graphs are shown in Figs. 17(a) and 17(b), respectively.

With these concepts available, the node determinant of the admittance matrix of the network is found to be

$$\Delta=|A_I Y_e A_V{}^t|=\sum_i \epsilon_i x(\text{common tree product of a tree pair } v_i \text{ of } G_I \text{ and } G_V), \quad (18)$$

where A_I and A_V are the incidence matrices of G_I and G_V, respectively, and A^t is the transpose of A. ϵ_i is the sign of the common tree product of tree pair v_i determined by theorem 1. The summation is for all possible tree pairs of G_I and G_V. It is understood that a tree product is the product of admittances of a set of edges constituting a tree.

When a network contains only the ordinary elements, the current and voltage graphs of the network are identical, i.e., $A_I=A_V$, and the sign of a tree product is always positive. However, if a network includes dependent node pairs, its current and voltage graphs are *not* identical but they are 2-semi-isomorphic. Hence, the common tree product of G_I and G_V may *not* always be positive.

Example 7. Given a network containing a tube and a gyrator as shown in Fig. 18(a), where the transconductance and the plate conductance of the tube are g_m and g_p, respectively. β denotes the gyrator conductance. The mathematical equivalent circuit of the network in terms of "current and voltage elements" is then found as shown in Fig. 18(b). Then, the current

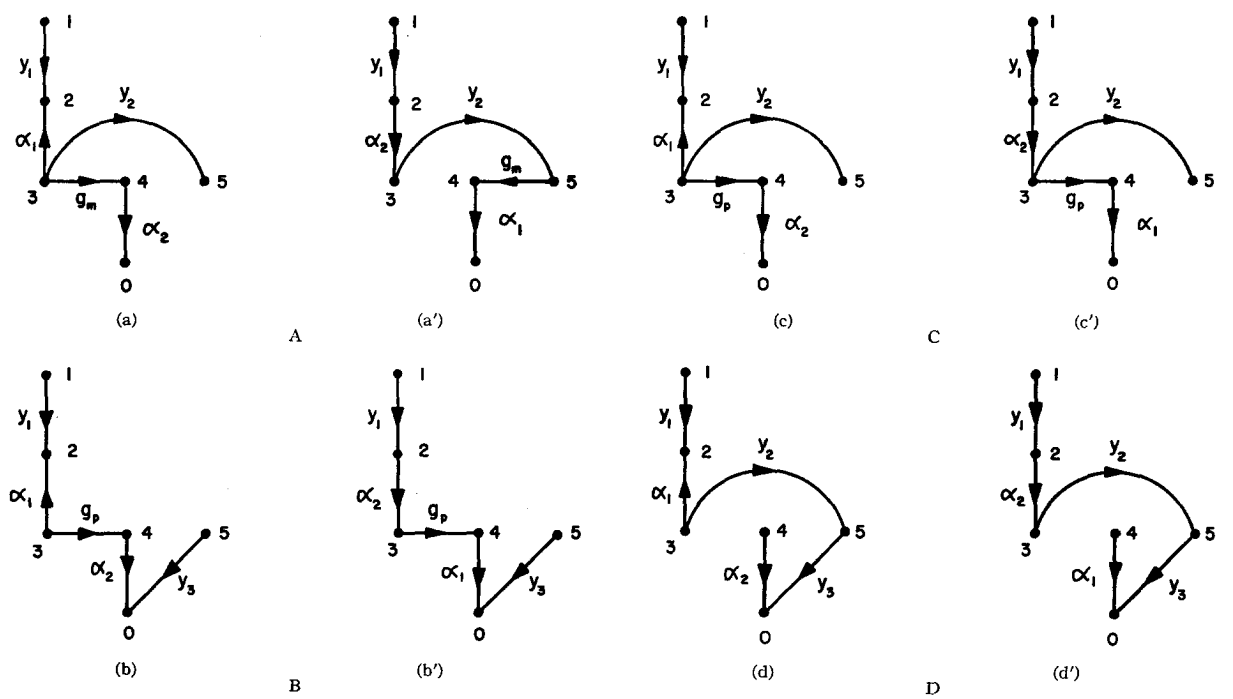


FIG. 20. Tree pairs of G_I and G_V of Fig. 19. A. Tree pair 1: (a) Tree 1 of G_I. (a') Tree 1 of G_V. B. Tree pair 2: (b) Tree 2 of G_I. (b') Tree 2 of G_V. C. Tree pair 3: (c) Tree 3 of G_I. (c') Tree 3 of G_V. D. Tree pair 4: (d) Tree 4 of G_I. (d') Tree 4 of G_V.

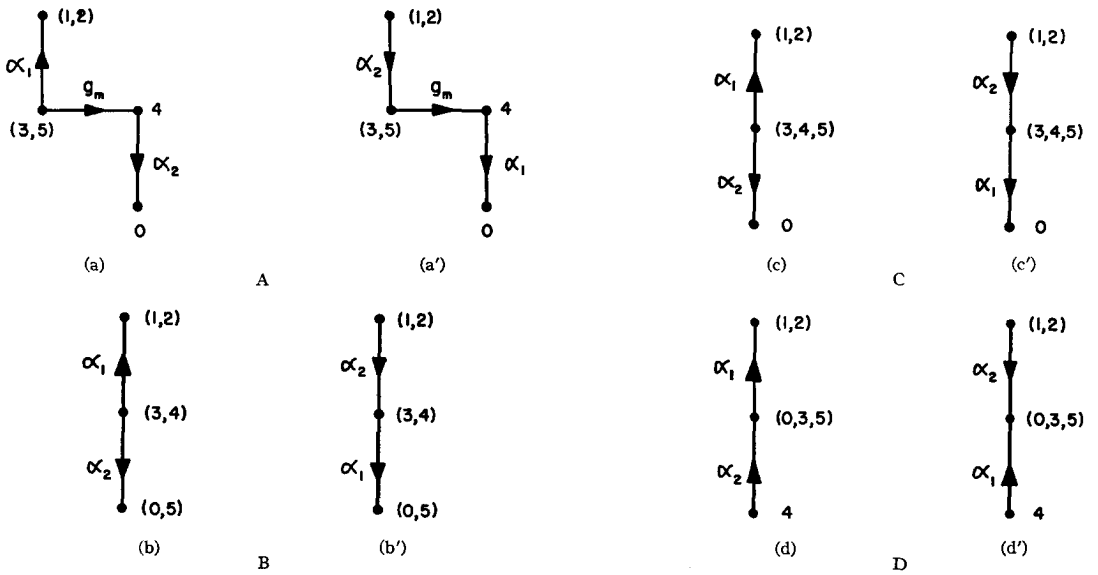


FIG. 21. Reduced tree pairs of Fig. 20. A. Reduced tree pair 1: (a) Of tree 1 of G_I. (a') Of tree 1 of G_V. B. Reduced tree pair 2 (b) Of tree 2 of G_I. (b') Of tree 2 of G_V. C. Reduced tree pair 3: (c) Of tree 3 of G_I. (c') Of tree 3 of G_V. D. Reduced tree pair 4: (d) Of tree 4 of G_I. (d') Of tree 4 of G_V.

graph G_I of the equivalent circuit, which contains the current elements and the ordinary elements, is given in Fig. 19(a); and its voltage graph G_V with the voltage elements and the ordinary elements is shown in Fig. 19(b). The orientations of the ordinary edges y_1, y_2, y_3, and g_p are assigned arbitrarily. It is noted that the current and voltage graphs in Fig. 19 are 2-semi-isomorphic. We, therefore, use the properties of 2-semi-isomorphic graphs described in the previous section in order to evaluate the node determinant of the network of Fig. 18(b).

By inspection of Fig. 19, it is clear that there exist four tree pairs of the two graphs as shown in Fig. 20.

In order to find the sign of each common tree product of the tree pairs by the formula of theorem 1, the ordinary edge pairs in the tree pairs are reduced, and the reduced tree pairs are shown in Fig. 21.

For the reduced tree pair 1, the number of active edge pairs is three. However, edge pair g_m becomes an ordinary edge pair in the reduced tree. Therefore, the reduced tree pair 1 of Fig. 21 is gain reduced as shown in Fig. 22.

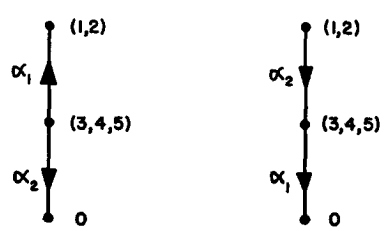


FIG. 22. Reduced tree pairs of Fig. 21A.

In Fig. 22, due to definition 6, the signs of edge pairs α_1 and α_2 are

$$\begin{aligned} &\text{sign of edge pair } \alpha_1 = -1, \\ &\text{sign of edge pair } \alpha_2 = +1. \end{aligned} \tag{19}$$

The number of interchanges of the nodes needed to give edge pairs α_1 and α_2 the same principal nodes is one. Therefore, from theorem 1, we have

$$\begin{aligned} &\text{the sign of the common tree product of tree pair 1} \\ &= (-1)\cdot(\text{sign of edge pair } \alpha_1)\ (\text{sign of edge pair } \alpha_2) \\ &= (-1)(1)(-1) = +1. \end{aligned} \tag{20}$$

Similarly, one gets:

$$\begin{aligned} &\text{The sign of the common tree product of} \\ &\qquad \text{tree pair } 2 = +1; \\ &\text{The sign of the common tree product of} \\ &\qquad \text{tree pair } 3 = +1; \\ &\text{The sign of the common tree product of} \\ &\qquad \text{tree pair } 4 = +1. \end{aligned} \tag{21}$$

Thus, the node determinant of the network Δ is found by

$$\Delta = y_1y_2\alpha_1\alpha_2g_m + y_1y_3g_p\alpha_1\alpha_2 + y_1y_2g_p\alpha_1\alpha_2 + y_1y_2y_3\alpha_1\alpha_2. \tag{22}$$

Electrical Conduction in a Stretched and Twisted Tube

A. C. PIPKIN AND R. S. RIVLIN
Brown University, Providence, Rhode Island
(Received December 8, 1960)

The constitutive equation derived in a previous paper relating electric current and field in an isotropic material which is subjected to a time-independent deformation is applied to the solution of the problem of electrical conduction in a twisted tube of circular cross section to which a longitudinal electric field is applied. It is shown that the current follows a helical path and that an axial magnetic field is produced.

1. INTRODUCTION

IN a previous paper,[1] the constitutive equation for electrical conduction in a deformed isotropic material has been derived. In the deformation of a body, a particle initially at X_i in the rectangular Cartesian coordinate system x moves to x_i in the same coordinate system. It is assumed that the components J_i of the current density vector $\mathbf{J}$ are polynomials in the deformation gradients $\partial x_p/\partial X_q$ and the components E_p of the electric field $\mathbf{E}$. It is shown in the previous paper that the constitutive equation for the current must be expressible, with the notation

$$g_{ij}=(\partial x_i/\partial X_p)(\partial x_j/\partial X_p), \quad \mathbf{g}=\|g_{ij}\|, \tag{1.1}$$

in the form

$$J_i=(A\delta_{ij}+Bg_{ij}+Cg_{ik}g_{kj})E_j, \tag{1.2}$$

where A, B, and C are expressible as polynomials in the quantities

$$\text{Tr } \mathbf{g}, \quad \text{Tr } \mathbf{g}^2, \quad \text{Tr } \mathbf{g}^3, \tag{1.3}$$

and

$$E_iE_i, \quad E_ig_{ij}E_j, \quad E_ig_{ij}g_{jk}E_k, \quad g^{-\frac{1}{2}}=|g_{ij}|^{-\frac{1}{2}}. \tag{1.4}$$

The assumption of polynomial dependence was introduced in order to simplify the derivation of Eq. (1.2). If A, B, and C are general single-valued functions of the quantities (1.3) and (1.4), the invariance requirements which led to the form (1.2) are still satisfied. The argument $g^{-\frac{1}{2}}$ can then be eliminated from consideration since g can be expressed in terms of the quantities (1.3) [see Eq. (3.1)].

We now assume that the electrical conduction is ohmic, as it will generally be if the electric field is sufficiently small. The coefficients A, B, and C are then independent of $\mathbf{E}$, and they are, therefore, functions of the quantities (1.3) alone.

We shall use the constitutive equation (1.2) to solve the problem of the electrical conduction in a stretched and twisted tube of conducting material when a potential difference is applied to its ends. It will be shown that the current in the tube flows in helices and gives rise to an axial magnetic field. The conductivity of the deformed tube and the magnitude of this magnetic field are calculated.

2. GENERAL CASE

Let us suppose that initially the tube has inner radius R_0 and outer radius R_1. Let the tube be twisted through an angle ψ radians per unit length and then stretched to λ times its original length. If the X_3 axis is taken along the axis of the tube, the deformation may be described by the equations

$$x_1=r(R)\cos(\theta+\psi X_3), \tag{2.1}$$

$$x_2=r(R)\sin(\theta+\psi X_3), \tag{2.2}$$

and

$$x_3=\lambda X_3, \tag{2.3}$$

where

$$R^2=X_1^2+X_2^2 \tag{2.4}$$

and

$$\theta=\arctan(X_2/X_1). \tag{2.5}$$

The precise form of the function $r(R)$ depends on the rheological properties of the material and on the system of forces applied to the tube. Since $r(R)$ has not been specified, the deformation may also include inflation under internal pressure.

From (1.1) we obtain, at the particle which is at $(r, 0, x_3)$ in the deformed state,

$$\|g_{ij}\|=\begin{Vmatrix} (r')^2 & 0 & 0 \\ 0 & (r/R)^2(1+R^2\psi^2) & r\psi\lambda \\ 0 & r\psi\lambda & \lambda^2 \end{Vmatrix}. \tag{2.6}$$

These are the components of $\mathbf{g}$ in a local Cartesian coordinate system with axes in the radial, circumferential, and axial directions, not only at $(r, 0, x_3)$ but also at any other point. Let the components of $\mathbf{E}$ and $\mathbf{J}$ in such a system be E_r, E_θ, E_z, and J_r, J_θ, J_z. Then introducing (2.6) into (1.2), we obtain

$$J_r=[A+B(r')^2+C(r')^4]E_r, \tag{2.7}$$

$$J_\theta=AE_\theta+f^2[B(1+\rho^2)+Cf^2(1+\rho^2)^2+C\rho^2\lambda^2]E_\theta + f\rho\lambda[B+Cf^2(1+\rho^2)+C\lambda^2]E_z, \tag{2.8}$$

$$J_z=AE_z+f\rho\lambda[B+Cf^2(1+\rho^2)+C\lambda^2]E_\theta + \lambda^2[B+C(f^2\rho^2+\lambda^2)]E_z, \tag{2.9}$$

where we have used the notation

$$f=r(R)/R, \quad \rho=\psi R. \tag{2.10}$$

A, B, and C are functions of the quantities (1.3), which

[1] A. C. Pipkin and R. S. Rivlin, J. Math. Phys. 1, 127 (1960).

take the following values:

$$\text{Tr } \mathbf{g}=(r')^2+f^2(1+\rho^2)+\lambda^2, \tag{2.11}$$

$$\text{Tr } \mathbf{g}^2=(r')^4+f^4(1+\rho^2)^2+2f^2\rho^2\lambda^2+\lambda^4, \tag{2.12}$$

$$\text{Tr } \mathbf{g}^3=(r')^6+f^6(1+\rho^2)^3+3f^4(1+\rho^2)\rho^2\lambda^2 +3f^2\rho^2\lambda^4+\lambda^6. \tag{2.13}$$

In the particular case when $E_r=E_\theta=0$ and $E_z=E$, with E a constant, we have

$$J_r=0, \tag{2.14}$$

$$J_\theta/E=f\rho\lambda[B+Cf^2(1+\rho^2)+C\lambda^2], \tag{2.15}$$

$$J_z/E=A+\lambda^2[B+C(f^2\rho^2+\lambda^2)]. \tag{2.16}$$

No current flows through the sides of the tube since J_r is zero, and the steady-state condition $\nabla\cdot\mathbf{J}=0$ is satisfied identically.

The total current J passing through the ends of the tube is found from Eq. (2.16):

$$J=2\pi\int_{r_0}^{r_1} J_z(r)r\,dr. \tag{2.17}$$

Here $r_0=r(R_0)$ and $r_1=r(R_1)$, i.e., r_0 and r_1 are the inner and outer radii of the tube in its deformed state. The average axial conductivity is

$$\frac{J}{\pi(r_1^2-r_0^2)E}=\frac{2}{r_1^2-r_0^2}\int_{R_0}^{R_1}[A+\lambda^2B +\lambda^2(f^2\rho^2+\lambda^2)C]fr'R\,dR, \tag{2.18}$$

where the integration variable has been changed to R.

Besides the current in the axial direction, there is the circumferential component J_θ. The current flows in helices, unless $B=C=0$. The component J_θ produces an axial magnetic field H inside the tube which, neglecting end effects, is given by

$$H=(4\pi/c)\int_{r_0}^{r_1} J_\theta(r)dr$$

$$=E(4\pi/c)\int_{R_0}^{R_1} f\rho\lambda[B+C(f^2+f^2\rho^2+\lambda^2)]r'dR, \tag{2.19}$$

where c is the velocity of light in free space.

3. INCOMPRESSIBLE MATERIALS

The preceding results can be brought into a more explicit form if the material is incompressible. In this case $|\partial x_i/\partial X_j|=1$, and it follows from Eq. (1.1) that $|g_{ij}|=1$ as well. The latter determinant can be expressed in terms of the quantities (1.3):

$$g=|g_{ij}|=\tfrac{1}{6}[2\text{ Tr }\mathbf{g}^3-3\text{ Tr }\mathbf{g}\text{ Tr }\mathbf{g}^2+(\text{Tr }\mathbf{g})^3]. \tag{3.1}$$

Thus, with $g=1$, Tr $\mathbf{g}^3$ can be expressed in terms of Tr $\mathbf{g}$ and Tr $\mathbf{g}^2$, hence A, B, and C are functions of Tr $\mathbf{g}$ and Tr $\mathbf{g}^2$ alone.

With regard to the deformation described by Eqs. (2.1)–(2.5), the condition $|\partial x_i/\partial X_j|=1$ implies that

$$\lambda r(R)r'(R)=R, \tag{3.2}$$

and therefore,

$$r^2=r_0^2+\lambda^{-1}(R^2-R_0^2). \tag{3.3}$$

Only r_0 is left to be determined by the properties of the material and the forces applied. In any particular experiment, r_0 can be measured, and $r(R)$ is thus determined without reference to any knowledge of the rheological properties of the material.

By making use of Eq. (3.3), we obtain

$$f^2=(r/R)^2=\lambda^{-1}R^{-2}[R^2-R_0^2+\lambda r_0^2] =\lambda^{-1}\rho^{-2}[\rho^2-(\psi R_0)^2+\lambda(\psi r_0)^2] \tag{3.4}$$

and

$$(r')^2=\lambda^{-1}R^2[R^2-R_0^2+\lambda r_0^2]^{-1} =\lambda^{-1}\rho^2[\rho^2-(\psi R_0)^2+\lambda(\psi r_0)^2]^{-1}. \tag{3.5}$$

By using Eqs. (3.2) and (3.3), the average axial conductivity (2.18) can be written as

$$\frac{J}{\pi(r_1^2-r_0^2)E}=\frac{1}{(\psi R_1)^2-(\psi R_0)^2}\int_{(\psi R_0)^2}^{(\psi R_1)^2}[A+\lambda^2B +\lambda^2(f^2\rho^2+\lambda^2)C]d(\rho^2). \tag{3.6}$$

When Eqs. (3.4) and (3.5) are used in Eqs. (2.11) and (2.12), it is seen that Tr $\mathbf{g}$ and Tr $\mathbf{g}^2$ are functions of ρ^2, depending on the parameters $(\psi R_0)^2$, $(\psi r_0)^2$, and λ. The same is then true of A, B, and C. The average axial conductivity is, therefore, a function of $(\psi R_0)^2$, $(\psi R_1)^2$, $(\psi r_0)^2$, and λ.

Similarly, $\psi H/E$ is a function of $(\psi R_0)^2$, $(\psi R_1)^2$, $(\psi r_0)^2$, and λ:

$$\psi H/E=(2\pi/c)\int_{(\psi R_0)^2}^{(\psi R_1)^2}[B+C(f^2+f^2\rho^2+\lambda^2)]d(\rho^2). \tag{3.7}$$

H is an odd function of ψ; hence, a small twist may be expected to produce a proportional magnetic field.

4. SMALL DEFORMATIONS

It has been shown[1] that when the deformation is of the form $x_i=X_i+u_i(X_p)$, where the displacement gradients $\partial u_i/\partial X_j$ are of infinitesimal order, the constitutive equation can be written in the form

$$J_i=[(S_1+S_2e_{kk})\delta_{ij}+S_4e_{ij}]E_j, \tag{4.1}$$

provided that the conduction is ohmic. S_1, S_2, and S_4 are constants, and e_{ij} is defined by

$$e_{ij}=\tfrac{1}{2}[(\partial u_i/\partial X_j)+(\partial u_j/\partial X_i)]. \tag{4.2}$$

If in Eqs. (2.1)–(2.3) we set $\lambda=1+\epsilon$ and $r(R)=R+u(R)$, and treat ϵ, $u(R)/R$, and ψR as infinitesimals,

then at the point $(r, 0, x_3)$, we obtain

$$\|e_{ij}\| = \begin{Vmatrix} u'(R) & 0 & 0 \\ 0 & u(R)/R & \frac{1}{2}\psi R \\ 0 & \frac{1}{2}\psi R & \epsilon \end{Vmatrix}. \tag{4.3}$$

On taking $E_r = E_\theta = 0$ and $E_z = E$, and making use of Eq. (4.3), we obtain

$$J_r = 0, \tag{4.4}$$

$$J_\theta/E = \tfrac{1}{2}\psi R S_4, \tag{4.5}$$

and

$$J_z/E = S_1 + S_2 e_{kk} + S_4\epsilon. \tag{4.6}$$

From Eq. (4.5) the axial magnetic field can be found immediately:

$$H = (4\pi/c)\int_{R_0}^{R_1} \tfrac{1}{2}\psi R S_4 E dR = (E/c)\pi(R_1^2 - R_0^2)\psi S_4. \tag{4.7}$$

In the case of an elastic material with no forces applied to the sides of the tube, $u(R)$ is equal to $-\sigma\epsilon R$, where σ is Poisson's ratio. In this case, the average axial conductivity is the constant conductivity

$$J_z/E = S_1 + S_2(1-2\sigma)\epsilon + S_4\epsilon. \tag{4.8}$$

In an incompressible material, whether elastic or not, e_{kk} vanishes, hence

$$J_z/E = S_1 + S_4\epsilon. \tag{4.9}$$

ACKNOWLEDGMENT

The results presented in this paper were obtained in the course of research carried out under a grant from the National Science Foundation.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 4 JULY-AUGUST, 1961

Errata: Numerical Integration of the Transport Equation with No Angular Truncation

[J. Math. Phys. **1**, 225 (1960)]

HERBERT S. WILF

University of Illinois, Urbana, Illinois

Equation (8)′ on p. 226 should read

$$K_{ls}^*(t) = \int_0^1 P_l(\omega)\left[\frac{e^{-t/\omega} + (-1)^{l+s}e^{-\frac{\tau_0 - t}{\omega}}}{\omega(1-e^{-\tau_0/\omega})}\right] P_s(\omega)d\omega. \tag{8}'$$

Equation (34) on p. 229 is modified accordingly.

then at the point $(r, 0, x_3)$, we obtain

$$\|e_{ij}\| = \begin{Vmatrix} u'(R) & 0 & 0 \\ 0 & u(R)/R & \frac{1}{2}\psi R \\ 0 & \frac{1}{2}\psi R & \epsilon \end{Vmatrix}. \tag{4.3}$$

On taking $E_r = E_\theta = 0$ and $E_z = E$, and making use of Eq. (4.3), we obtain

$$J_r = 0, \tag{4.4}$$

$$J_\theta/E = \tfrac{1}{2}\psi R S_4, \tag{4.5}$$

and

$$J_z/E = S_1 + S_2 e_{kk} + S_4\epsilon. \tag{4.6}$$

From Eq. (4.5) the axial magnetic field can be found immediately:

$$H = (4\pi/c)\int_{R_0}^{R_1} \tfrac{1}{2}\psi R S_4 E dR = (E/c)\pi(R_1^2 - R_0^2)\psi S_4. \tag{4.7}$$

In the case of an elastic material with no forces applied to the sides of the tube, $u(R)$ is equal to $-\sigma\epsilon R$, where σ is Poisson's ratio. In this case, the average axial conductivity is the constant conductivity

$$J_z/E = S_1 + S_2(1-2\sigma)\epsilon + S_4\epsilon. \tag{4.8}$$

In an incompressible material, whether elastic or not, e_{kk} vanishes, hence

$$J_z/E = S_1 + S_4\epsilon. \tag{4.9}$$

ACKNOWLEDGMENT

The results presented in this paper were obtained in the course of research carried out under a grant from the National Science Foundation.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 4 JULY-AUGUST, 1961

Errata: Numerical Integration of the Transport Equation with No Angular Truncation

[J. Math. Phys. **1**, 225 (1960)]

HERBERT S. WILF

University of Illinois, Urbana, Illinois

Equation (8)′ on p. 226 should read

$$K_{ls}^*(t) = \int_0^1 P_l(\omega)\left[\frac{e^{-t/\omega} + (-1)^{l+s}e^{-\frac{\tau_0 - t}{\omega}}}{\omega(1-e^{-\tau_0/\omega})}\right] P_s(\omega)d\omega. \tag{8}'$$

Equation (34) on p. 229 is modified accordingly.